Horst Völz

Elektronik

Grundlagen, Prinzipien, Zusammenhänge

5. Auflage

Mit 672 Abbildungen und 161 Tabellen

AKADEMIE-VERLAG BERLIN
1989

ISBN 3-05-500046-3

Erschienen im Akademie-Verlag Berlin, DDR-1086 Berlin, Leipziger Straße 3—4
© Akademie-Verlag Berlin 1986
Lizenznummer: 202 · 100/437/89
Printed in the German Democratic Republic
Gesamtherstellung: VEB Druckerei „Thomas Müntzer", 5820 Bad Langensalza
Einband und Schutzumschlag: Rolf Kunze
LSV 1154
Bestellnummer: 763 249 4 (6118)

09500

Aus dem Vorwort der ersten Auflage

Jeder experimentell tätige Wissenschaftler, insbesondere Naturwissenschaftler, hat heute mit einer größeren Zahl z.T. sehr komplizierter elektronischer Geräte umzugehen. Seit geraumer Zeit zählen auch die elektronischen Rechenanlagen zu seinem Handwerkszeug. Bei den modernsten Meßgeräten wird schon ein gewisser Rechenteil zur Vorverdichtung der Information eingebaut. Um die elektronischen Geräte optimal auszunutzen, ist es sowohl notwendig, ihre Funktionsweise zu verstehen, als auch, kleine Zusatzgeräte entwerfen und aufbauen zu können. Zuweilen ist gerade jener Teil, der unmittelbar mit dem Forschungsobjekt zusammengeschaltet wird, eine spezielle, für die Lösung des Problems typische Schaltung. Auch der Mathematiker möchte häufig mehr als nur die mathematischen Gesetze der elektronischen Rechenanlagen kennen.

Für die genannten Wissenschaftler und Techniker fehlt seit langem ein Buch, das einerseits das innere Verstehen der wesentlichen Gesetzmäßigkeiten der Elektronik und andererseits eine gewisse Anleitung zum Aufbau von elektronischen Schaltungen vermittelt, und zwar soweit, daß der Naturwissenschaftler zumindest schnell eine gemeinsame Sprache mit den speziell ausgebildeten Elektronikern findet. Ferner sollte eine wenn auch geringe Anleitung zum elektronischen Arbeiten gegeben werden.

Die Mehrzahl der vorhandenen Elektronikbücher kann in zwei Gruppen geteilt werden: einmal die Literatur der Ingenieurschulen und technischen Hochschulen, deren Zweck vor allem in der Entwicklung von Schaltungen und ihrer Dimensionierung liegt, und andererseits die mehr theoretische Literatur, welche die mathematisch-physikalischen Zusammenhänge vermittelt. Beide Buchgruppen sind für den vorgenannten Zweck zu spezialisiert und umfangreich. Sie erfordern außerdem Vorkenntnisse, die gerade die eingangs erwähnten Berufsgruppen erst erwerben müssen. Mit dem vorliegenden Werk wurde ein Kompromiß angestrebt. Wie weit das gelungen ist, kann nur die Reaktion auf dieses Buch zeigen. Mit dem Buch können daher auch keine neuen wissenschaftlichen Ergebnisse vorgelegt werden. Es ist aber versucht worden, soweit es der Linie des Buches entsprach, die neuesten Ergebnisse zu berücksichtigen. Bei der Auswahl des Stoffes konnte und sollte keine auch nur annähernde Vollständigkeit angestrebt werden. Vorrangig war das Ziel, didaktisch (auch für Leser ohne Vorkenntnisse) bei geringem mathematischen Aufwand die wichtigsten Fakten so zu vermitteln, daß der Leser selbständig weiter denken kann. Die didaktische Linie wurde nicht nur von der Reihenfolge des Stoffes her aufgebaut, es wurde auch versucht, durch möglichst viele Bilder anschauliches und assoziatives Denken anzuregen sowie verdichtete Information zu schaffen. Ferner sollte durch die Bilder ein schnelles Wiederfinden von Gelesenem erleichtert werden ... In ähnlicher Weise sind auch die Tabellen angelegt. Ihr Zahlenmaterial ist bewußt mehr beispielhaft als umfangreich gewählt.

Im allgemeinen wurde versucht, und hierzu ist die Elektronik ja relativ gut geeignet, das Modell- und Systemdenken zu fördern. Ferner wurde durch umfangreiche

Rückzitate und möglichst wenig Zitate zum noch nicht gelesenen Text angestrebt, durch Wiederholung des Durchgearbeiteten die erworbenen Kenntnisse zu vertiefen. So sollte das Buch auch für das Selbststudium geeignet werden . . .

Dieses Buch geht im wesentlichen auf Vorlesungen zurück, die ich seit mehr als zehn Jahren an der Humboldt-Universität in Berlin halte. Dabei zeigte sich regelmäßig, daß die Elektronik sehr schnellebig ist. Deshalb waren immer wieder Änderungen in der Stoffauswahl und in der didaktischen Anordnung notwendig. Das spiegelt sich gewiß auch hier in einigen nicht zu vermeidenden Mängeln wider. Soweit wie möglich wurde versucht, dem durch Trendbetrachtungen abzuhelfen . . .

Berlin, Dezember 1973 H. Völz

Vorwort zur vierten Auflage

Das Anliegen der vorangegangenen Auflagen, wie es sich in dem Auszug aus dem Vorwort der ersten Auflage ergibt, wurde beibehalten. Die fast vollständige Überarbeitung war aufgrund der Fortschritte auf dem Gebiet notwendig. Autor und Verlag hoffen, daß so erneuert das Buch dem Leser wieder die notwendige Unterstützung gibt.
Neu aufgenommen wurden insbesondere folgende Gebiete:

- *Mikrorechentechnik* von den Grundbauelementen über Peripheriebausteine bis zur Software,
- *Optoelektronik* als ein Zweig, der künftig hohe Bedeutung erlangen wird,
- *Filtertechnik* (aus didaktischen Gründen zwar in mehreren Kapiteln enthalten, aber in allen ihren Varianten bis hin zum Analogprozessor, also den Digitalfiltern),
- *Neue Sonderbauelemente*, wie z.B. Oberflächenwellen, Piezomaterialien, einige Sensoren und Tieftemperaturelektronik.

Wesentlich verkürzt bzw. auf Spezialgebiete eingeschränkt wurden

- Röhrentechnik (nicht Bild- und Oszillographenröhre)
- Halbleitertechnologie
- Analogrechentechnik
- Literaturverzeichnis, Bild- und Tabellennachweis.

Darüber hinaus wurde es notwendig, z.T. die Zusammenfassungen zu Abschnitten neu zu wählen. So entstanden deutlich betont u. a. die Gebiete:

- Lineare Wechselstromschaltungen (in Richtung Filter)
- Operationsverstärker
- Standard-Interfaces
- Dick- und Dünnfilmtechnik
- Zuverlässigkeit.

Dem Bildmaterial wurde eine erhöhte Aufmerksamkeit gewidmet. Für die einzelnen Teilkomplexe wurden meist jeweils mehrere Bilder so zu einer Bildtafel zusammengefügt, daß die gewünschten Aussagen als Einheit erscheinen.

Bei der Ausarbeitung erfolgte eine vorbildliche Zusammenarbeit mit dem Akademie-Verlag. Besonders hervorzuheben ist die zuständige Lektorin Frau U. Heilmann. Sie ging auf alle Wünsche von mir ein und gab stets hilfreiche Unterstützung. Mit starkem Einsatz und Einfühlungsvermögen betreute Frau R. Gelbrich als Außenlektorin das Manuskript und die Korrekturarbeiten für die Drucklegung. Auch Frau H. Winkler aus der Herstellung förderte schon in der Entstehungsphase durch mehrere Beratungen mit nützlichen Hinweisen die Arbeit.

Das Manuskript schrieb mit großer Umsicht und einem hohen Einfühlungsvermögen Frau C. Fröhlich. Die Reinzeichnungen konnten in mehrfacher direkter Rücksprache mit Frau G. Haevecker auf das oben genannte Niveau gebracht werden. Die Bibliothek und die Vervielfältigung des Zentralinstituts für Kybernetik und Informationsprozesse gaben hilfreiche Unterstützung bei der Ausarbeitung. Allen diesen Kolleginnen sei herzlich für das hohe Engagement bei der Fertigstellung gedankt.

Das Manuskript wurde vom Juni 1982 bis Mai 1983 erarbeitet.

H. Völz

Inhaltsverzeichnis

0. Einleitung

1. Elektrotechnische Grundlagen

1.1. Stationäre Ströme 4

1.1.1. Bewegungen der Elektronen 4
Elektronengas • Stromtransport • Stromstärke, Stromdichte • Mikro-Ohmsches Gesetz • Strömungsfeld

1.1.2. Ohmscher Widerstand 9
Ohmsches Gesetz • Zusammenschaltung von Widerständen • Widerstandsnetzwerke

1.1.3. Strom-Spannungs-Quellen 12
Wesen der Spannungsquelle • Wesen der Stromquelle • Zweipole • Nutzbare Spannung • Anpassungen • Verlauf der Leistungsanpassung

1.1.4. Vermaschte Stromkreise 16
Ersatzspannungsquellen • Überlagerungsgesetz • Maschenregel • Knotenregel

1.1.5. Vierpole 18

1.1.6. Technische Widerstände 21

1.1.7. Technische Spannungsquellen 26
Chemische Grundlagen • Primärelemente • Sekundärelemente

1.2. Magnetostatik 35

1.2.1. Ursachen des Magnetismus 35

1.2.2. Feldgrößen 39

1.2.3. Hysteresekurve 40

1.2.4. Permeabilitäten 41

1.2.5. Magnetischer Kreis 42

1.2.6. Energieprodukt 42

1.2.7. Kreis mit Luftspalt 44

1.2.8. Kraftwirkungen 45

1.2.9. Magnetische Werkstoffe 46

1.3. Elektrostatik 49

1.3.1. Isolierstoffe 51

1.4.	**Wechselstrom**	52
1.4.1.	Definition	52
1.4.2.	Periodische Wechselströme	52
1.4.3.	Harmonische Schwingung	53
1.4.4.	Zusammenschaltung von Wechselquellen	54
1.4.5.	Drehstromnetz	55
1.4.6.	FOURIER-Analyse	56
1.4.7.	Beschreibung von Schwingungen	58
1.4.8.	Entstehung von Oberwellen und Gleichrichtung	60
1.4.9.	Messung von Wechselgrößen	63
1.5.	**Grundbauelemente**	66
1.5.1.	Definition	66
1.5.2.	Einschaltvorgänge	67
1.5.3.	Rechteckströme	69
1.5.4.	Harmonische Ströme	71
1.5.5.	Technische Kapazitäten Allgemeines • Keramische Kondensatoren • Folienkondensatoren • Elektrolytkondensatoren (Elko) • Vergleiche und Ergänzungen	72
1.5.6.	Technische Induktivitäten Grundlagen • Blechkernspulen • Ferritkernspulen • Offene Spulen	80
1.6.	**Transformatoren und Übertrager**	88
1.6.1.	Gegeninduktivität	88
1.6.2.	Idealer Transformator	90
1.6.3.	Verbessertes Ersatzschaltbild	91
1.6.4.	Frequenzgang	92
1.6.5.	Leistungstransformator	93
1.6.6.	Spartransformator	93
1.6.7.	Variometer	94
2.	**Lineare Wechselstromschaltungen**	
2.1.	**Einführung**	95
2.2.	**Einfache Zweipole**	96
2.2.1.	Grundbauelemente	96
2.2.2.	Reihenschwingkreis	97
2.2.3.	Einschwingvorgänge	99
2.2.4.	Pole und Nullstellen	101
2.2.5.	Kreisdiagramm	102
2.2.6.	Relative Bandbreite	103
2.2.7.	Ergänzungen zu realen Bauelementen	105

2.3.	**Komplexe Spannungsteiler**	106
2.3.1.	Einfache frequenzselektive Teiler	106
2.3.2.	Schwingungskreisspannungsteiler	108
2.3.3.	Frequenzunabhängige Teiler	109
2.3.4.	*RC*-Siebglieder	109
2.3.5.	Brückenschaltungen	112
2.4.	**Systemtheorie**	116
2.4.1.	Allgemein komplexe Frequenzen	116
2.4.2.	Allgemeine Systeme	116
2.4.3.	Pol-Nullstellen-Plan	118
2.4.4.	Einteilung der Systeme	119
2.4.5.	Potentialanalogien	120
2.4.6.	Zulässige Zweipolfunktionen	121
2.5.	**Theorie der Filter**	123
2.5.1.	Allgemeine Einführung	123
2.5.2.	Wellenparametertheorie	124
2.5.3.	Betriebsparametertheorie	129

Potenzfilter • TSCHEBYSCHEFF-Filter • CAUER-Filter und Vergleiche • HF-Bandfilter

3. Grundlagen der Halbleiterbauelemente

3.1.	**Eigenschaften von Halbleitermaterialien**	137
3.1.1.	Bändermodell	137
3.1.2.	Leitungsmechanismen	139
3.1.3.	Leiter, Halbleiter, Isolatoren	141
3.1.4.	FERMI-Verteilung	141
3.1.5.	Zustandsdichte	143
3.1.6.	Leitungsmechanismen	147
3.2.	**Der *pn*-Übergang**	149
3.2.1.	Stromloser *pn*-Übergang	149
3.2.2.	Stromdurchflossener *pn*-Übergang	152
3.2.3.	Durchlaßbetrieb	153
3.2.4.	Sperrbetrieb	154
3.2.5.	Raumladekapazität	156
3.2.6.	ZENER- und Avalanche-Effekt	156
3.2.7.	Trägheit von *pn*-Übergängen	158
3.2.8.	*psn*-Struktur	159
3.2.9.	Metall-Halbleiter-Kontakte	160

3.3.	**Feldeffekttransistoren**	161
3.3.1.	Sperrschicht-FET	163
3.3.2.	MOSFET	165
3.3.3.	Bändermodell beim MOSFET	167
3.3.4.	Anreicherungs- und Verarmungstyp	167
3.3.5.	Einfluß der Substratspannung und punch-through	169
3.3.6.	Theorie der Ladungssteuerung	170
3.3.7.	Kleinsignalverhalten und Ersatzschaltbild	171
3.3.8.	Potentialtopfmodell	173
3.4.	**Bipolartransistoren**	174
3.4.1.	Grundmodell	174
3.4.2.	Mathematisches Modell	176
3.4.3.	Die vier Betriebszustände	178
3.4.4.	Physikalisches Modell	178
3.4.5.	Effektive Basisweite und punch-through	179
3.4.6.	Kennlinien	180
3.4.7.	Kleinsignalparameter	181

4. Technologiegrundlagen

4.1.	**Reinstsilicium und Einkristalle**	185
4.1.1.	Herstellung des Rohsiliciums	185
4.1.2.	Physikalische Reinigung	186
4.1.3.	Herstellung von Einkristallen	187
4.1.4.	Herstellung der Scheiben	190
4.2.	**Oberflächenschichten**	193
4.2.1.	Epitaxie	193
4.2.2.	Diffusion	194
4.2.3.	Isolierschichten	196
4.2.4.	Metallschichten	198
4.2.5.	Ionenimplantation	199
4.2.6.	Vergleich der Dotierungstechniken	201
4.3.	**Strukturierung**	202
4.3.1.	Fotolithografie	202
4.3.2.	Standardmethode für Masken	204
4.3.3.	Leicht abgewandelte Techniken	205
4.3.4.	Elektronenstrahllithografie	206
4.3.5.	Röntgenstrahllithografie	208
4.3.6.	Zu physikalischen Grenzen	208

4.4.	**Epitaxie-Planartechnik**	208
4.4.1.	Strukturierung eines bipolaren Transistors	209
4.4.2.	Strukturierung eines MOSFET	211
4.5.	**Herstellen der Bauelemente**	212
4.5.1.	Übergang zu den Chips	212
4.5.2.	Gehäuseeinbau	213
4.5.3.	Drahtbonden	214
4.5.4.	Plastverkapselung	215

5. Bauelemente — Eigenschaften und Grundschaltungen

5.1.	**Halbleiterwiderstände**	217
5.1.1.	Halbleitermaterial	217
5.1.2.	FET-Widerstände	220
5.1.3.	Heißleiter	220
5.1.4.	Kaltleiter	223
5.1.5.	Varistoren	224
5.2.	**Kapazitäten**	225
5.3.	**Induktivitäten**	227
5.4.	**Dioden**	228
5.4.1.	Bauarten	230
5.4.2.	Kennlinien und Daten	232
5.4.3.	Grundlagen der Gleichrichtung	234
5.4.4.	Gleichrichterschaltungen	236
5.4.5.	Stabilisation von Spannungen	239
5.4.6.	Spannungsbegrenzung	240
5.4.7.	Demodulation von Amplitudenmodulation	241
5.4.8.	Zwei Frequenzen an nichtlinearer Kennlinie	241
5.4.9.	Ringmodulator	243
5.5.	**Feldeffekttransistoren**	247
5.5.1.	Verstärkung und Ersatzschaltbilder	249
5.5.2.	Arbeitspunkt und Arbeitskennlinie	250
5.5.3.	*RC*-Verstärker	251
5.5.4.	Güte	254
5.5.5.	Arbeitskennlinie bei komplexem Arbeitswiderstand	255
5.5.6.	Resonanz- und Breitbandverstärker	256
5.5.7.	Kettenverstärker	257
5.5.8.	Technologien spezieller FET	258

5.6. Bipolartransistor 262

5.6.1. Die drei Schaltungen 262

5.6.2. Betriebswerte 264

5.6.3. Kennlinien und Arbeitspunkt 267
Näherungsbetriebswerte • Arbeitspunkt im Kennlinienfeld • Einstellung des Arbeitspunktes

5.6.4. Ersatzschaltbilder 269
Vielzahl der Ersatzschaltbilder • Frequenzgang

5.6.5. Transistortechnologien 272
Einteilungen • Isoliertechnologien • Struktur der *pn*-Übergänge • Flächenstruktur

5.7. Grenzdaten 279

5.7.1. Rauschen 281
Allgemeine Grundlagen • Rauschursachen • Rauschkennzahlen • Diode • FET • Bipolarer Transistor

5.7.2. Restströme 287
Restströme des bipolaren Transistors • Temperatureinflüsse

5.7.3. Minimalabmessungen 289

5.7.4. Beseitigung äußerer Störungen 291
Erdung • Schirmung • Symmetrierung • Netzstörungen

5.7.5. Obere Grenzen, Schaltverhalten 295
Spannungsgrenzen • Schaltverhalten • Schalten von Blindlast • Zweiter Durchbruch • Sicherer Arbeitsbereich

5.7.6. Wärmeableitung 300
Thermische Grundlagen • Elektrische Analogie • Beispiele für Wärmewiderstände • Wärmeleitrohr • Impulswärmewiderstand • Thermische Stabilität

5.7.7. Leistungsgrenzen 311

6. Bauelemente spezieller Gebiete

6.0. Überblick 312

6.1. Optoelektronik 314

6.1.1. Lichttechnische Größen und allgemeine Grundlagen 314
Maßeinheiten • Schwarzer Strahler • Farbmetrik • Optische Darstellungen • Ergonomische Grundlagen

6.1.2. Halbleiter-Fotoempfänger 325
Allgemeine Grundlagen • Detektivität • Einteilung der Fotoeffekte • Fotowiderstände • Fotoelemente • Fotodioden • Weitere Fotoempfänger

6.1.3. Leucht- und Laserdioden 334
Grundlagen der Lumineszenzdioden • Aufbau und Daten von Lumineszenzdioden • Laserprinzip • Halbleiterlaser

6.1.4. Optokoppler und Lichtschranken 346

6.1.5. Bauelemente mit Ladungskopplung 347
Drei- und Zweiphasenbetrieb • Verluste, Grenzen und Anwendungen • Lichtsensoren

6.1.6. Flüssigkristalle 353
Aufbau und Arbeitsweise von TN-Zellen • Fluoreszenzaktiviertes Display

6.1.7. Weitere Anzeigetechniken 361
Lumineszenzzellen • Elektrochemische Anzeigen • Ferroelektrische Keramik • Magnetische Anzeigen

6.1.8. Vergleich von Anzeigen 365

6.1.9. Lichtleitfasern 366
Dämpfung • Typen von Fasern

6.2. Vakuumelektronik 370

6.2.1. Grundlagen 372
Emission von Elektronen • Fotoemission • Thermoemission • Ströme einer Röhre • Bewegung und Ablenkung von Elektronen • Leuchtstoffe

6.2.2. Lichtempfindliche Röhren 386
Fotozellen und SEV • Bildwandler, Bildverstärker • Bildlaufnahmeröhren

6.2.3. Anzeige- und Bildröhren 390
Digitron (Fluoreszenzanzeigeröhre) • Oszillographenröhre • Sichtspeicherröhre • Bildröhre • Farbbildröhre

6.2.4. Plasmabauelemente 400
Grundlagen • Ziffernanzeigeröhren und Plasmapanels

6.3. Mechano- und Akustoelektronik 404

6.3.1. Überblick 404

6.3.2. Relais 406

6.3.3. Meßinstrumente 410
Drehspulinstrument • Dynamometer • Dreheiseninstrument • Elektrostatische Instrumente • Zungenfrequenzmesser • Registrierinstrumente, XY-Schreiber und Plotter

6.3.4. Wandler, Resonanzen und Filter 413
Magnetostriktion • Piezoeffekt • Mechanische Resonanzen, Quarze • Mechanische Filter

6.3.5. Oberflächenwellen (OFW) 428
Eigenschaften von OFW-Filtern • Bauelemente mit OFW

6.4. Leistungselektronik 432

6.4.1. Geschichtliche Entwicklung und Anwendungsgebiete 432

6.4.2. Vierschichtdiode als Grundlage 435

6.4.3. Thyristor 437

6.4.4. Familie der Thyristorbauelemente 439

6.4.5. Sonderbauelemente für Thyristoren 441

6.4.6. Ausgewählte Schaltungen 444

6.4.7. Vergleich mit anderen Leistungsbauelementen 445

6.5. Sensoren 446
6.5.1. Überblick 446
6.5.2. Magnetische Sensoren 448
HALL-Effekt • Feldplatte • Magnetdioden • WIEGAND-Sensor
6.5.3. Dehnungsmeßstreifen (DMS) 453
6.5.4. Tasten 454

6.6. Mikrowellenbauelemente 456
6.6.1. Tunneldioden 456
6.6.2. Backwarddioden 457

6.7. Tieftemperaturelektronik 458

7. Analoge Schaltungen

7.1. Rückkopplung 461
7.1.1. Phasenreine Rückkopplung 463
Kennwerte • Änderungen im Kennlinienfeld • Negativer Innenwiderstand • Verstärkung und Stabilität • Nichtlinearitäten
7.1.2. Frequenzabhängige Rückkopplungen 470
Frequenzselektive Verstärker • Schwingungserzeugung • Grenzen der Gegenkopplung
7.1.3. Pendelrückkopplung 474
7.1.4. Elektronisch erzeugte Bauelemente 475
Elektronische Eingangswiderstände • Elektronische Ausgangswiderstände • Impedanzkonverter, Gyrator • Vollständigkeit der passiven Bauelemente
7.1.5. Zusammenfassung 480

7.2. Modulationen 481
7.2.1. Stetige Modulationen 482
Amplitudenmodulation • Frequenz- und Phasenmodulation
7.2.2. Pulsmodulationen 486
7.2.3. Quantisierte Signale 488

7.3. Einfache Analogschaltungen 489
7.3.0. Allgemeiner Überblick 489
7.3.1. Einstufige Schaltungen 490
Einstufiger Verstärker • Konstant-Stromquellen • Konstant-Spannungsquellen und Siebung • Einstellbare Z-Diode und Umkehrung von NTC-Widerständen
7.3.2. Mehrstufige Gleichstromverstärker 498
DARLINGTON-Transistoren • Kaskodeverstärker • WHITE-Folger und ähnliche Schaltungen • Schaltungen mit großer Verstärkung • Stabiler Innenwiderstand Null
7.3.3. Differenzverstärker 504
Einführung • Grundschaltung • Gleichphasiger Ausgang • Schaltungsvarianten • NORTHON-Verstärker

7.3.4. Grundschaltungen für IC 509
Prinzipien der analogen IC • Stromspiegel • Potentialverschiebestufen • Spannungsquellen

7.3.5. Weitere Stromquellen 514

7.3.6. Transmissionsgatter, Multiplexer, Sample-hold 515

7.3.7. Oszillatorschaltungen 516
Sinusoszillatoren • Rechteckgeneratoren • Sägezahn- bzw. Dreieckgeneratoren

7.3.8. Impulsschaltungen 521
Überblick • Monostabiler Multivibrator • Begrenzer, Komparator, SCHMITT-Trigger • Pulslängenmodulator

7.4. Ausgangsstufen und Leistungsverstärker 526

7.4.1. Allgemeine Grundlagen 526
Grenzen der Ausgangsarbeitsfläche • Bedeutung der Ausgangsfläche • Leistungsumsatz

7.4.2. Endstufe mit nur einem Transistor 528
Optimaler Arbeitswiderstand • Kollektor- und Emitterschaltung • Leistungsbilanzen • Verschiedene Betriebsfälle • Induktiv überbrückter Arbeitswiderstand • Gleitende Arbeitspunktverlagerung

7.4.3. Gegentaktendstufen 534
Prinzip; Gegentakt-A-Betrieb • Idealer B-Betrieb • Grafische Dimensionierung • Übernahmeverzerrungen; AB-Betrieb • Verzerrungen bei Gegentaktschaltungen • Nichtsinusförmige Steuersignale

7.4.4. Praktischer Aufbau von Gegentaktendstufen 540
Einstellung des Ruhestromes • Steuerstufen • Mittenspannung der Endstufe • Quasikomplementäre Endstufen • Schutzschaltungen

7.4.5. Ausgewählte integrierte Endstufen 545
A 211 D • A 210 D bzw. K • MDA 2010 bzw. MDA 2020

7.4.6. Weitere Prinzipien 551
C-Endstufe • D-Verstärker • Umschaltung zwischen zwei Endstufen • Spezielle Transistorverzerrungen

7.5. Operationsverstärker 554

7.5.1. Einführung 554
Zur Geschichte • Der ideale Operationsverstärker

7.5.2. A 109 555
Die Innenschaltung • Layout • Statische Kenndaten und Eigenschaften

7.5.3. Frequenzgang und -korrektur 559
Äußere Beschaltung • Äußere Beschaltung mit R, C, R • Innere Korrektur • Sprungantwort • Optimale Anschaltung der Korrekturglieder • Slew-Rate • Beispiel A 109

7.5.4. Ausgewählte Operationsverstärker 564
µA 741/748/777 • Programmierbare Typen B 176 D, B 177 D und µA 776 • Maskenprogrammierbare Typen • BIFET-Operationsverstärker

7.5.5. Die beiden Grundschaltungen des Operationsverstärkers 569

7.5.6. Einfache lineare Schaltungen 573

7.5.7. Nichtlineare Schaltungen 574
Logarithmische und exponentielle Verstärker • Funktionsgeneratoren mit Beispiel für Sinus • Multiplikatoren • Gleichrichterschaltungen • Ringmodulatorschaltungen • Begrenzer, Komparator und SCHMITT-Trigger • Modulations- bzw. Chopperverstärker • Multivibratoren und Funktionsgeneratoren

7.5.8. Aktive Filter 586
Einsverstärker • Angefachte Einsverstärkerschaltungen • Mit invertierendem Verstärker • Allpässe und Breitbandphasenschieber

7.6. Stabilisierte Stromversorgungen 591

7.6.1. Grundprinzipien 591
Kenngrößen • Störwertaufschaltungen

7.6.2. Spannungsregler 594

7.6.3. Schutzschaltungen 596

7.6.4. Integrierte Schaltkreise 598
Spannungsregler MAA 723 • Festspannungsregler MA 78XX

7.6.5. Stromregler 603

7.6.6. Geschaltete Stromversorgungen 604
Allgemeine Grundlagen • Prinzip des Durchflußwandlers • Prinzip des Sperrwandlers • Weitere Prinzipien • Vor- und Nachteile • Eisenloser Gleichspannungswandler • Integrierte Ansteuerschaltkreise

7.7. Ausgewählte Gebiete zur Meßtechnik 610

7.7.1. Grundlagen des Messens 610
Grundbegriffe • Meßkette • Maßeinheiten • Logarithmische Maße • Meßmethoden • Dem Messen verwandte Methoden

7.7.2. Einfache Meßschaltungen 616
Messung von Gleichwerten • Korrektur des Rückwirkungsfehlers • Widerstandsmessung • Wechselstrommessung und Vielfachinstrumente • Meßverstärker

7.7.3. Ausgewählte komplexe Meßschaltungen 624
Allgemeines Meßprinzip • Frequenzbereichfilterung • Beispiele aus der Audiotechnik • Trägerfrequenzverfahren und phasenempfindlicher Gleichrichter • Periodische Eingangssignale und Akkumulation • Auto- und Kreuzkorrelation • Phasenregelkreis (PLL) • Impulsverdichtung

7.7.4. Oszillographentechnik 630
Ellipsenauswertung • Untersuchung von Nichtlinearitäten • Prinzipaufbau von Oszillographen • Beispiel: gedämpfte Schwingung • Dreidimensionale Oszillographie • Mehrstrahl-, Sampling- und Projektionsoszillograph • Digital- und Speicheroszillograph

7.7.5. Meßwert- und Datenspeicherung 637
Magnetbandspeicherung • Folienspeicher, Floppy disc

7.8. Ladung verschiebende Schaltungen 639

7.8.1. Eimerkettenspeicher 641

7.8.2. Ladung rückwärts, Information vorwärts 641

7.8.3. Störeinflüsse 643

7.8.4. FET-Eimerkette 643

7.9. Nichtklassische Filter 644

7.9.1. Abgrenzung und Überblick 644

7.9.2. Grundlagen der Verzweigungsnetzwerke 646

7.9.3. Grundgleichungen 647

7.9.4. Prinzip der Abtastfilter 647

7.9.5. Beschreibung mittels z-Transformation 649

7.9.6. Beispiel: CCD-Filter 651

7.9.7. SC-Filter 651

7.9.8. Grober Vergleich der Filtertypen 653

8. Digitale Schaltungen

8.1. Kodes und Zahlendarstellungen 654

8.1.1. Zum Begriffspaar analog — digital 654

Analogie, analog • Kontinuierlich, stetig • Diskret • Quantisiert • Digital, dual, binär • Struktur der Begriffe • Vorteile analoger und digitaler Technik

8.1.2. Zahlensysteme 659

8.1.3. Bezeichnungsaufwand 660

8.1.4. Kodierungen 661

BCD-Zahlen • Tetradenkodes • Kodeeigenschaften • Prüfbare Kodes • Fehlerkorrigierende Kodes • Byte, Kbyte • Zahlen mit Vorzeichen • Gleitkommazahlen • Kodierung von Buchstaben, Befehlen usw.

8.2. Kombinatorische Schaltungen 668

8.2.1. Mathematische Grundlagen 669

Aussagenlogik • Abstrakte Algebra • Mengentheorie

8.2.2. Boolesche Algebra 672

Mögliche Funktionen • Vielzahl der Darstellungsformen • Mögliche Grundfunktionen • Höherstellige Funktionen • Schwellenwertlogik • Rechenregeln • Normalformen • Minimierung • Schaltungen mit mehreren Ausgängen

8.2.3. Prinzipien technischer Realisierungen 683

Diodenschaltungen • Transistor als Negator • Positive und negative Logik • Pegelbereiche und Fan-out • Innere und äußere Logik

8.2.4. TTL-Technik 688

Die typische Eingangsstufe • Ausgangsstufe der Standard-TTL • Kennwerte der Standard-TTL • Weitere Ausgangsstufen • Teilschaltungen zur Verbesserung • Schaltzeit-Leistungsprodukt; TTL-Familien

8.2.5. ECL-Technik 698

8.2.6. I^2L-Technik 700

Das Wirkprinzip • Logische Verknüpfungen • Injektionsquellen • Schaltzeiten und Injektorstromstärke • Weiterentwicklungen

8.2.7. MOS-Techniken 704

p- und n-MOS-Technik • Prinzipieller Aufbau der Gatter und Logik • CMOS-Schaltungen • CMOS-Technologien • SOS-Technik

8.2.8. Ausgewählte Beispiele 709
Koder — Dekoder • BCD-7-Segment-Dekoder • Multiplexer — Demultiplexer • Vergleicher • Adder • Arithmetisch-logische Einheiten: ALU • Programmierbare Felder: PLA

8.2.9. Ergänzungen und Vergleiche 721
Kontaktprellen • Hazards • Entwicklungsgeschichte und Vergleich

8.3. Speicherschaltungen 723

8.3.1. Allgemeine Grundlagen 723
Speicherfunktionen • Eigenschaften von Speichern

8.3.2. Organisationsformen 730

8.3.3. Bipolare Speicher 732
Latch, Speicherzelle • Statisches RAM, 16 bit, Typ D 181 • I^2L-Speicher • Weitere Speicherzellen, minimaler Strom

8.3.4. MOS-Speicher 738
Statische Speicherzellen • Statischer (1 K × 1 bit)-MOS-Speicher 2102 • Ein VMOS-(2 K × 8 bit)-Speicher • Absenkung der Verlustleistung bei statischen RAM • Dynamische Speicherzellen • Dynamisches 16-K-RAM U256 D bzw. 4116 • Entwicklung zu hohen Kapazitäten, Redundanz • Soft-Errors

8.3.5. Programmierbare Speicher — Festwertspeicher 749
Überblick, Definition • ROM • PROM • Floating gate = FG • EPROM • E^2PROM • EAROM bzw. MNOS • NVRAM

8.3.6. Umlaufspeicher 757
CCD-Speicher • Bubbles

8.3.7. Assoziativspeicher 761
Grundprinzip • Typische assoziative Speicherzellen • Assoziative Festwertspeicher

8.3.8. Kellerspeicher 764

8.4. Grundlagen sequentieller Schaltungen 765

8.4.1. Zwei Modelle zur Analyse 765
Erstes Versuchsgerät • MEALY- und MOORE-Automat

8.4.2. Beschreibungsmethoden 768

8.4.3. PETRI-Netze 769

8.4.4. Formale Beschreibung 771

8.4.5. Typen von Automaten 772

8.4.6. Beispiel: Steuerung für ein Magnetbandgerät 773

8.4.7. Entwurfsprinzipien 775

8.4.8. Fehler: Race und Rückkopplungshazards 776

8.5. Einfache sequentielle Schaltungen 777

8.5.1. Impulsschaltungen 777

8.5.2. Flipflop 779
Funktionen von FF • Universelles FF, D-FF und T-FF • Anzahl der möglichen FF • Takt- und Vorbereitungseingang • JK-FF • Master-Slave-FF • Realisierte Beispiele für DV-FF und JK-MS-FF

8.5.3. Frequenzteiler 785
Asynchrone Frequenzteiler • Synchrone Frequenzteiler

8.5.4. Zählerschaltungen 788
Grundprinzip • Ringzähler • JOHNSON-Zähler • Dualzähler • BCD-Zähler

8.5.5. Register 794
Überblick • Pufferregister • Betriebsarten statischer Schieberegister • Beispiele statischer Schieberegister • Realisierungstechniken • Quasistatische Register • Dynamische Register • Erzeugung pseudozufälliger Folgen

8.5.6. Einfache Rechnerschaltungen 802
Vielfalt der Schaltungen für eine Funktion • Register für Rechenschaltungen • Addition mit Schieberegister • Multiplikation und Division mit 2 • Multiplikation: Beispiel

8.5.7. Zur Datenübertragung 806
Handshake-Prinzip • Bussysteme

8.6. Mikrorechentechnik 808

8.6.1. Einführung und Abgrenzung 808
Mikroelektronik, -rechner, -prozessor • Anwendung der Rechentechnik • Klassischer Rechenautomat • Ablauf einer Rechnung • Hard- und Software • Größeneinteilung von Rechnern • Grenzen der Rechentechnik • Zur Geschichte der Rechentechnik

8.6.2. Mikroprozessoren 824
Zusammenwirken mit anderen Baugruppen • Innerer Aufbau • Gewinnung und Aufteilung des Befehlswortes: Befehlstypen • Transportbefehle • ALU-Befehle • Ein-, Ausgabe • Steuerbefehle • Adressierungstypen • Registeranordnungen und -typen

8.6.3. U 880—Z 80 843
Aufbau der Hardware • Zeitverhalten • Verschiedene Darstellungen des Befehlssatzes • Alphabetische Ordnung der Befehlsgruppen • Ordnung nach Befehlsgruppen • Ordnung nach Hex-Kode • Zusätzliche Befehle

8.6.4. Interrupt-Prinzipien 863
Einfaches Interrupt • Mehrfache Interrupts • Realisierungstechniken • Interrupt-System des U 880

8.6.5. Peripherieschaltungen 867
PIO U 855 • SIO U 856 • CTC U 857 • U 880-DMA • Arithmetikprozessoren • Tastatur- und Anzeigenanschlüsse • Fernseh-Interface • Speicher-Interface

8.6.6. Bedeutung der Informationen 885

8.6.7. Geschichte der Mikroprozessoren 886
Vom 4004 zum Z 80 und 6502 • Marktanteile und Tendenzen

8.7. Standard-Busse; Standard-Interface 888

8.7.1. Überblick 888

8.7.2. Kettenförmiges System SI 1.2 891

8.7.3. Sternförmige Systeme 892
Sif 1000 • S-100-Bus; IEEE 696 • Weitere Mikrorechnerbusse

8.7.4. Linienförmige Systeme 893
IEC-Bus • SI 2.2; SIAL • CAMAC

8.7.5. Serielle Systeme 901
V.24 und RS 232 • PDV-Bus • HP-IL-Bus

8.7.6. Vergleich von Bussen 908

8.8. Software 908

8.8.1. Allgemeine Grundlagen der Programmierung 908
Effizienz und Top-down • Strukturierte Programmierung und deren Grundelemente • Verzweigungen • Schleifen • Datenstrukturierung • Unterprogramme

8.8.2. Einteilung der Software 914
Beispielprogramm Addition • Arten von Programmen

8.8.3. Maschinennahe Programmierung 917
Assembler • Testhilfen

8.8.4. Grundsoftware 920

8.8.5. Höhere Programmiersprachen 921
Compiler und Interpreter • Verschiedene Programmiersprachen

8.8.6. Programmieren auf anderen Rechnern 924

8.8.7. Zusammenfassung 925

8.9. Speicherschaltungen und Sonderrechner 926

8.9.1. Systematik digitaler Schaltungen; Anwendungen des Mikroprozessors 926

8.9.2. Mikroprogrammierung 926

8.9.3. Gegenseitige Ersetzbarkeit binärer Schaltungen 928

8.9.4. Kundenwunschschaltungen 930

8.9.5. Vergleich von Anwendungen 932

8.9.6. Taschenrechner 934

8.9.7. Spezialrechner 936
Problemstellung • Pipelining • Feldrechner • Baumstruktur • Parallelität auf höherer Ebene

9. Spezielle Gebiete

9.1. Digital-Analog-Wandler 940

9.1.1. Schalterprinzip 941

9.1.2. Prinzip der Kondensatorentladung 942

9.1.3. Impulsdauerprinzip 943

9.1.4. Multiplizierende DA-Wandler 944

9.2. Analog-Digital-Wandler 944

9.2.1. Grundlagen 944

9.2.2. AD- mit DA-Wandlern 945

9.2.3. Direkte Wandlung 946

9.2.4. Wegumsetzer 947

9.2.5. Wandler auf Zeitbasis 947

9.2.6. Wandler mittels Spannungs-Frequenz-Wandler 948
9.2.7. Wägeprinzip 949
9.2.8. Prinzip der Ladungsübertragung 950
9.2.9. Vergleich von ADU 951

9.3. Digitalfilter 952
9.3.1. Prinzip 952
9.3.2. Beispiel 953
9.3.3. Vor- und Nachteile 954
9.3.4. Signalprozessor 2920 955

9.4. Sondertechnologien 958
9.4.1. Überblick 958
9.4.2. Substrate 964
9.4.3. Spezialbauelemente 965
Überblick und Standardmethoden • Beam-lead • TFT
9.4.4. Dickschichttechnik 968
Pastensysteme • Herstellungsprozeß • Eigenschaften der hergestellten Bauelemente
9.4.5. Dünnschichttechnik 971
Aufdampftechnik • Sputtern • NiCr-System • Tantalsystem
9.4.6. Abgleich von R und C 973
9.4.7. Kapselung der Schaltungen 974
9.4.8. Vergleich mit anderen Techniken 975
9.4.9. Bedeutung diskreter Bauelemente 975

9.5. Zuverlässigkeit 976
9.5.1. Problembereich 976
9.5.2. Grundlagen 977
Fehlerverteilung, Ausfallrate • Bauteile und Systeme • Wartung und Reserve
9.5.3. Ursachen und Einflüsse 982
Analyse der Ursachen • Auswirkungen äußerer Einflüsse • Betriebsbelastungen
9.5.4. Werte der Ausfallrate 987
9.5.5. Zeitraffende Methoden und Frühausfälle 989
9.5.6. Strahlung, EMV, EMP 990
Korpuskel- und γ-Strahlung • EMV • Geschichte des EMP • Wirkungsmechanismus des EMP • Schutz gegen EMP
9.5.7. Software 994
Einteilung möglicher Fehler • Qualitätsmerkmale • Zuverlässigkeit

9.6. Einige Grundlagen der Informationstheorie 997
9.6.1. Kanalmodell 997

9.6.2. Entropie 998
Gleichverteilung • Unterschiedliche Wahrscheinlichkeiten • Eine anschauliche Herleitung • Verlauf von $-p \operatorname{ld} p$ • Binäre Quelle • Ternäre Quelle • Kontinuierliche Quelle • Verbundentropie

9.6.3. Abläufe in der Zeit 1005
Abtasttheorem • Kanalkapazität • Information und Energie

10. Verzeichnisse

10.1. Bücherverzeichnis 1008

10.2. Abkürzungsverzeichnis 1020

10.3. Verzeichnis der Bauelemente 1024

10.4. Sachwortverzeichnis 1025

0. Einleitung

Vor mehreren Jahren waren nur Elektrik und Elektronik zu unterscheiden. Seit wenigen Jahren ist als drittes Gebiet die Mikroelektronik hinzugekommen. Alle drei Gebiete beruhen auf einer gezielten Bewegung von elektrischen Ladungsträgern, also Elektronen, Defektelektronen und Ionen. Dieser Prozeß wird in seiner Gesamtheit Leitung genannt.

Elektrik liegt dann vor, wenn die Leitung bevorzugt in Metallen oder Elektrolyten erfolgt.

Elektronik benutzt für die Leitung vor allem Halbleiter, Gase und Vakuum. Dabei ist aber bedeutsam, daß die Ladungsträger leicht zu beeinflussen sind. Zu unterscheiden sind innere Beeinflussung (z.B. durch Raumladung und Steuerelektroden) und äußere Beeinflussung (z.B. durch magnetische, elektrische Felder, Wärme, Licht radioaktive Strahlung und mechanische Einwirkungen).

Mikroelektronik liegt dann vor, wenn eine Vielzahl von aktiven Bauelementen untrennbar auf einem Halbleitersubstrat zusammengefaßt sind und funktionell als eine Einheit arbeiten.

Es ist offensichtlich, daß diese drei Formen aufeinander aufbauen und die Mikroelektronik vor allem ein hochentwickelter und immer mehr Selbständigkeit erreichender Sonderzweig der Elektronik ist. Diese Entwicklung erfolgte zwischen Elektrik und Elektronik vor ca. 30 bis 40 Jahren. Jedes Gebiet besitzt heute aber bereits seine typischen Bauelemente, aus denen Baugruppen, Geräte, Einrichtungen und Anlagen aufgebaut werden. Zu den Bauelementen der *Elektrik* gehören also

Widerstände, Kondensatoren, Spulen, Ringkerne, Transfluxoren, Quarze, Transformatoren, Übertrager, Relais, Motoren, Schalter, Stecker, Fassungen, Antennen, Hohlleiter, Batterien, Elemente, Generatoren (Motor).

Elektronische Bauelemente sind dagegen

Dioden, Fotodioden, Fotowiderstände, Fotozellen, Feldplatten, Transistoren, Thyristoren, HALL-Elemente, Vakuumröhren, Thyratrons, Ignotrons, Quecksilberdampfröhren, Elektronenstrahlröhren, Bildwandler, SEV, Wanderfeldröhren, Laser, Maser, Glimmlampen, Leuchtstoffröhren, Sonnenbatterien, Gleichrichter (Selen, Silicium), magneto-hydronischer Generator.

Mikroelektronische Bauelemente sind schließlich funktionell orientiert und sind nicht zerlegbare Einheiten:

Operationsverstärker, Leistungsverstärker, Spannungsstabilisatoren, logische Gatter, Speicherschaltkreise, Mikroprozessoren, Widerstandsnetzwerke, CCD-Matrizen.

Eine weitere wichtige Einteilung der Bauelemente erfolgt nach *aktiven* oder *passiven*. Aktive Bauelemente sind imstande, Energie abzugeben, oder lassen sich in ihrer Energieabgabe (verstärkend) steuern. Zu ihnen gehören u. a.

Relais, Motoren, Batterien, Elemente, Generatoren, Transistoren, Vakuumröhren, Thyratrons, Ignotrons, Bildwandler, SEV, Wanderfeldröhren, Laser, Maser, Sonnenbatterien.

Den relativen Anteil verschiedener Bauelementgruppen an einem elektronischen Gerät in Abhängigkeit von der Zeit zeigt Bild 0.1. Hieraus geht deutlich der relativ hohe Anteil jener Bauelemente hervor, die nicht zur Elektronik gezählt werden. Dies ist neben theoretischen Bezügen ein weiterer Grund, weshalb sie in diesem Buch zumindest teilweise berücksichtigt werden.

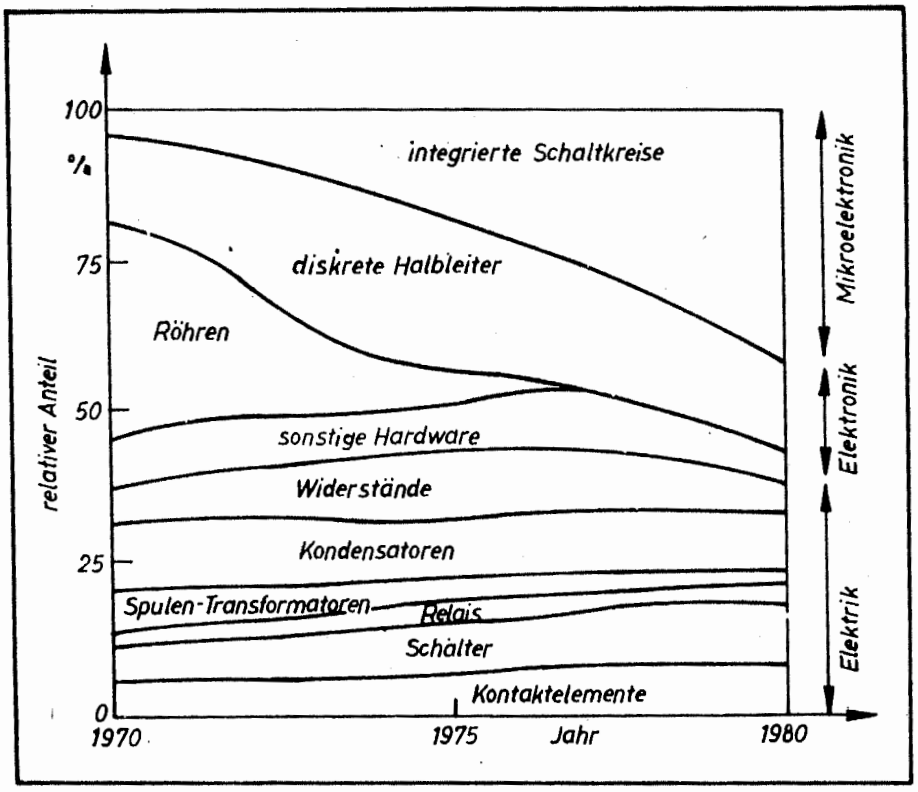


Bild 0.1. Der relative Anteil von verschiedenen Bauelementen an einem typischen elektronischen Gerät im Laufe der Zeit

Eine weitere Untergliederung von Elektrik und Elektronik/Mikroelektronik läßt sich schließlich nach dem *Wirkungsziel* der Geräte und Anlagen vornehmen. Hierzu ist es notwendig, drei Klassen von Objekten einzuführen, nämlich Stoff, Energie und Information. Auf stoffliche Prozesse hat im Grunde nur die *Elektrochemie* Einfluß. *Energetische Aspekte* spielen in der Starkstromtechnik, Kraftfahrzeugelektrik usw. eine bedeutsame Rolle. Sie werden entsprechend dem hohen Rang, welchen Energie-

Tabelle 0.1. Einteilung einiger Untergebiete von Elektrik und Elektronik einschließlich Mikroelektronik

Aspekt	Elektrik	Elektronik	Typische Stromstärke
Energie	Starkstromtechnik	Leistungselektronik	kA
Information	Nachrichtentechnik	Informationselektronik	nA — mA

fragen annehmen, gewiß immer wichtiger. Je nach dem Umfang, welchen hierbei die Elektronik und Mikroelektronik einnimmt, wird zwischen *Starkstromtechnik* und *Leistungselektronik* unterschieden.

Bei informationellen Prozessen ist die Leistungsbilanz von untergeordneter Bedeutung. Hier kommt es nur darauf an, daß ein elektronisches Signal den ausreichenden Störabstand besitzt. Viel wichtiger ist die funktionelle Wirkung, welche das Signal im Empfangssystem auslöst. Je nach dem Elektronik-/Mikroelektronikanteil werden *Nachrichtentechnik* und *Informationselektronik* (-technik) unterschieden (vgl. Tab. 0.1).

Auf die Einteilung von analoger und digitaler Technik wird in Abschnitt 8.1.1. eingegangen. Das sich z. Z. stark entwickelnde Gebiet der *Optoelektronik* wurde bei diesen Einteilungen noch nicht besonders herausgestellt.

1. Elektrotechnische Grundlagen

1.1. Stationäre Ströme

1.1.1. Bewegungen der Elektronen

1.1.1.1. Elektronengas

Als Grundlage stark vereinfachter Betrachtungen kann angenommen werden, daß in einem elektrischen Leiter frei bewegliche Elektronen existieren. Sie verhalten sich dort ähnlich wie ein Gas in einem abgeschlossenen Gefäß. Deshalb wird auch von der *Elektronengastheorie* gesprochen. Sie geht im wesentlichen auf DRUDE und LORENTZ zurück und wurde später von SOMMERFELD quantentheoretisch untermauert. Hier soll zunächst die DRUDEsche Theorie in ihrer einfachsten Form benutzt werden. Im Abschnitt 3.1. wird das allgemein gültige *Bändermodell* eingeführt.

Im dicht gepackten Material durchdringen sich die äußeren Elektronenschalen so stark, daß es für einige Elektronen nicht mehr möglich ist, sie einem bestimmten Atom zuzuordnen. Diese Elektronen sind also *frei bewegliche Ladungsträger*. Ihre Anzahl im Vergleich zur Anzahl der Atome weist für einige Materialien Tab. 1.1.1 aus. Für gute Leiter existiert also grob ein freies Elektron je Atom.

Tabelle 1.1.1. Beispiele für die Anzahl von Ladungsträgern und Atomen in Leitern

Material	Ladungsträger je cm³	Atome je cm³
Silber	$3{,}6 \cdot 10^{22}$	$5{,}9 \cdot 10^{22}$
Kupfer	$3{,}4 \cdot 10^{22}$	$8{,}4 \cdot 10^{22}$
Aluminium	$2{,}2 \cdot 10^{22}$	$8{,}3 \cdot 10^{22}$
Platin	$1{,}4 \cdot 10^{22}$	$6{,}6 \cdot 10^{22}$
Zink	$5{,}8 \cdot 10^{21}$	$6{,}6 \cdot 10^{22}$
Blei	$4{,}5 \cdot 10^{21}$	$3{,}3 \cdot 10^{22}$
Quecksilber	$4{,}3 \cdot 10^{20}$	$4{,}1 \cdot 10^{22}$

Die freien Elektronen nehmen die gleiche thermische Energie wie das Material an. Dadurch befinden sie sich *in ständiger Bewegung*, wobei ihre Geschwindigkeit und Richtung statistischer Natur sind. Aus der Energiegleichung können die zugehörigen Mittelwerte bestimmt werden:

$$\tfrac{1}{2} m_0 v_{\mathrm{th}}^2 = \tfrac{3}{2} kT \,. \tag{1}$$

Hierbei bedeuten

m_0 Masse des Elektrons ($9{,}11 \cdot 10^{-24}$ g),
v_{th} mittlere Temperaturgeschwindigkeit der Elektronen,
k BOLTZMANN-Konstante ($1{,}381 \cdot 10^{-23}$ JK^{-1}),
T absolute Temperatur (K).

Für die Geschwindigkeit folgt also:

$$v_{\mathrm{th}} = \sqrt{\frac{3kT}{m_0}} \,. \tag{2}$$

Sie beträgt bei Zimmertemperatur etwa 10^5 m/s.

Infolge der großen Anzahl von freien Elektronen und der relativ hohen Geschwindigkeit erfolgen *Zusammenstöße*. Je Elektron und Sekunde treten etwa 10^{15} auf. Zwischen zwei Zusammenstößen fliegt ein Elektron geradlinig. Diese Strecke heißt *freie Weglänge* λ. Sie ist wie die Geschwindigkeit von der Temperatur abhängig und beträgt bei Zimmertemperatur etwa 10^{-10} m. Die zugehörige Zeit heißt *Stoßzeit*. Sie folgt aus der freien Weglänge und der Geschwindigkeit gemäß

$$\tau = \lambda / v_{\mathrm{th}} \,. \tag{3}$$

Im Innern eines Volumens von 1 cm³ finden daher etwa 10^{34} ... 10^{38} Stöße je Sekunde statt. Die an die Oberfläche des Leiters gelangenden Elektronen bewirken dort Spannungsimpulse. Sie verursachen das später zu behandelnde Wärmerauschen.

Da die thermische Elektronenbewegung gleichberechtigt in allen Richtungen erfolgt, führt sie im zeitlichen Mittel zu keinem Elektronentransport.

1.1.1.2. Stromtransport

Innerhalb eines Leiters existiere jetzt ein elektrisches Feld $\boldsymbol{E}$. Hierdurch werden die einzelnen Elektronen zusätzlich zur Wärmebewegung beschleunigt. Die Größe dieser *Beschleunigung* $\boldsymbol{b}$ kann aus der Kräftegleichung

$$m_0 \boldsymbol{b} = e_0 \boldsymbol{E} \tag{4}$$

berechnet werden. Dabei ist e_0 die Ladung des Elektrons mit $1{,}60 \cdot 10^{-11}$ As.

Es sei weiter angenommen, das elektrische Feld wirke nur kurzzeitig ein. Infolge der Elektronenstöße wird dann die aus der Beschleunigung $\boldsymbol{b}$ gewonnene Energie gleichmäßig auf alle Richtungen verteilt und als zusätzliche Wärme dem Leiter zugeführt. Die Stöße der Elektronen überführen also die äußere elektrische Energie im Leiter in *Wärme*. So wird die maximale Stromstärke mit der zulässigen Erwärmung des Leiters begrenzt. Als grober Richtwert für die meisten Geräte kann etwa 3 A/mm² angenommen werden. Er ist im wesentlichen durch die mögliche Wärmeabfuhr bestimmt.

Auf Grund der Stöße wird aber auch die durch das äußere elektrische Feld bewirkte gerichtete Geschwindigkeit der Elektronen nur während der Einwirkungsdauer des Feldes wirksam. Die entstehende *Geschwindigkeit*

$$\boldsymbol{v}_{\mathrm{tr}} = \frac{1}{2} \frac{e_0 \boldsymbol{E}}{m_0} \tag{5}$$

ist also direkt proportional der angelegten Feldstärke. Die eigentlich ständig vorhandene Beschleunigung hält so gerade diese Geschwindigkeit aufrecht. Es tritt ein *Reibungsverlust* durch Stöße auf, der proportional der Geschwindigkeit ist.

Ein mechanisches Analogon ist das Stokessche Gesetz. In einer zähen Flüssigkeit fallen Körper nach einer gewissen Anfangsbeschleunigung mit konstanter Geschwindigkeit weiter. Die auf sie ständig einwirkende Erdbeschleunigung wird durch die Reibung in der zähen Flüssigkeit verbraucht.

1.1.1.3. Stromstärke, Stromdichte

Die Intensität des elektrischen Stromes wird durch die *Anzahl der Elektronen n*, die je Sekunde einen Leitungsquerschnitt durchfließen, gemessen:

$$\boxed{I = \frac{ne_0}{t}}\,. \tag{6}$$

Ein Elektron je Sekunde entspricht daher einer Stromstärke von $1{,}60 \cdot 10^{-19}$ A. Prinzipiell kann deshalb die Stromstärke durch Abzählen der Elektronen gemessen werden. Für sehr kleine Ströme wird diese Technik auch benutzt. Die dabei entstehende mittlere Frequenz ist ein direktes Maß der Stromstärke. So entsprechen 10^{-15} A etwa 6 kHz und 1 pA etwa 6 MHz. Stärkere Ströme werden leichter mit Methoden gemessen, die mittelbare Wirkungen des Stromes (Magnetfeld, Erwärmung usw.) ausnutzen.

Neben der Stromstärke hat die *Stromdichte*

$$i = \frac{I}{A} \tag{7}$$

gewisse Bedeutung (A = Fläche, Querschnitt). Sie ist gegeben durch die Anzahl der Elektronen, die sich pro Zeit durch die Flächeneinheit bewegen.

1.1.1.4. Mikro-Ohmsches Gesetz

Es möge die relativ große Stromstärke von 1 A in einen Leiter mit 1 mm² Querschnitt fließen. Dies bedeutet, daß $6 \cdot 10^{18}$ Elektronen in der Sekunde den definierten Querschnitt passieren. Bei Kupfer mit $3{,}4 \cdot 10^{11}$ Elektronen je mm³ ergibt sich eine Geschwindigkeit von 0,18 mm/s. Bei allen Leitern ist damit diese *Geschwindigkeit so gering*, daß sie gegenüber der statistisch thermodynamischen vernachlässigbar ist. Bei ihnen stellt die Elektronenbewegung durch ein elektrisches Feld also nur eine vernachlässigbare Korrektur gegenüber der thermischen Bewegung dar, und somit bleiben für alle technisch anwendbaren Ströme *Stromdichte* $\boldsymbol{i}$ und *elektrische Feldstärke* $\boldsymbol{E}$ einander *proportional*. Diese Proportionalität stellt das eigentliche von Ohm entdeckte Gesetz dar:

$$\boxed{\boldsymbol{E} = \varrho \boldsymbol{i}}\,. \tag{8}$$

Es soll hier als mikro-Ohmsches Gesetz bezeichnet werden. Der Proportionalitätsfaktor zwischen beiden Größen ist der spezifische Widerstand ϱ. Er liegt bei Metallen zwischen 0,016 und 2 Ω mm²/m. Die Tab. 1.1.2 zeigt einige Beispiele. Um in Kupfer die obige Stromdichte von 1 A/mm² zu erreichen, ist eine Feldstärke von etwa 178 μV/cm notwendig.

Es sei erwähnt, daß es Materialien gibt, bei denen die Ladungsträgerzahl so gering ist, daß selbst bei kleinen Strömen das mikro-Ohmsche Gesetz nicht mehr gilt. Es gilt auch dann nicht, wenn die thermische Geschwindigkeit und die durch die Feldstärke hervorgerufene Geschwindigkeit in die gleiche Größenordnung kommen. In solchen Fällen wird das mikro-Ohmsche Gesetz günstig durch eine Potenzreihe ersetzt. Dies kann z. B. bei tiefen Temperaturen auftreten. Die Supraleitung beruht jedoch auf anderen physikalischen Ursachen.

1.1.1.5. Strömungsfeld

In einem ausgedehnten Leitermaterial kann die Feldstärke in Größe und in Richtung vom Ort abhängig sein. Ein vereinfachtes, in die Ebene projiziertes Beispiel zeigt

Tabelle 1.1.2. Spezifischer Widerstand einiger Materialien in $\Omega mm^2/m$ bzw. $10^{-6}\ \Omega m$

Reine Metalle	ϱ	Halbleiter	ϱ
Silber	0,016	Graphit	15 ... 100
Kupfer	0,018	Dotiertes Ge bzw. Si	10^2 ... 10^6
Gold	0,022	Germanium reinst	$5 \cdot 10^8$
Chromium	0,026	Silicium reinst	$2{,}3 \cdot 10^{10}$
Aluminium	0,029	Cu_2O	$2 \cdot 10^{11}$
Wolfram	0,053	FeO	10^9
Zink	0,059		
Eisen	0,10	**Isolierstoffe**	ϱ
Platin	0,105		
Zinn	0,11	Plaststoffe	10^{13} ... 10^{16}
Tantal	0,16	Glas	10^{16} ... 10^{19}
Blei	0,21	Porzellan	10^{17} ... 10^{18}
Quecksilber	0,94	Silikone	10^{18}
Bismut	1,2	Diamant	10^{21}
		Paraffin	10^{21}
Legierungen	ϱ	Glimmer	10^{22}
		Quarz	10^{22} ... 10^{24}
Duraluminium (4 Cu, 1 Mn)	0,05		
Messing	0,06 ... 0,07	**Sonstiges**	ϱ
Eisendraht	0,12 ... 0,14		
Bronze	0,18	Elektrolyte	10^5 ... 10^7
(53, Cu, 47 Ni)		Seewasser	$3 \cdot 10^5$
Konstantan	0,50	Flußwasser	10^7 ... 10^8
Chromnickel	1,0	Destill. Wasser	$(1 \ldots 4) \cdot 10^{10}$
Kanthal	1,3 ... 1,5	Reinstwasser	$2{,}5 \cdot 10^{11}$
		Erde	10^8 ... 10^{10}
		Holz	ca. 10^{14}

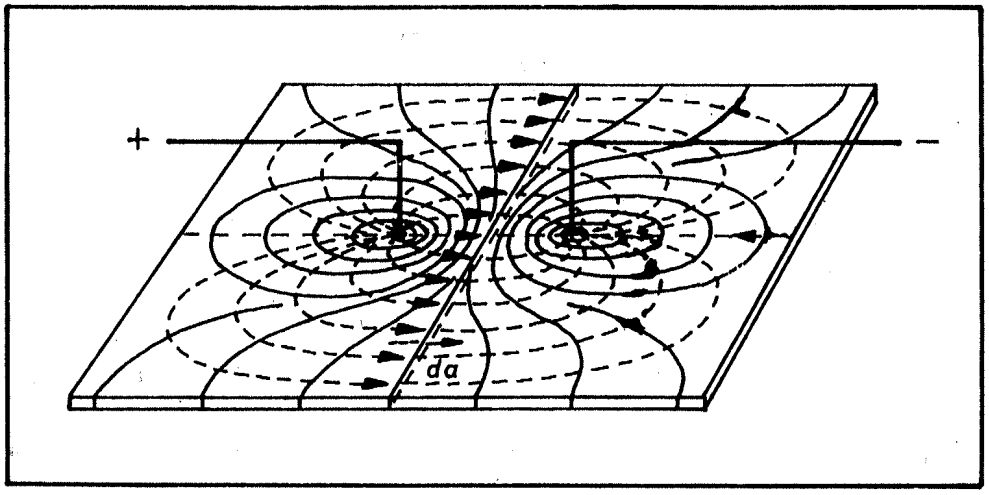


Bild 1.1.1. Einer dünnen leitenden ebenen Folie wird mittels der beiden Elektroden + und − Strom zu- bzw. abgeführt. In der Folie fließt dann der Strom längs der gestrichelten Linien. Er ist in Richtung und Stärke vom Ort in der Folie abhängig. Die senkrecht dazu stehenden ausgezogenen Linien besitzen das gleiche Potential und heißen Äquipotentiallinien

Bild 1.1.1. Zur Veranschaulichung wurde hierbei von der thermischen Bewegung der Elektronen abgesehen. Weiter wurde gedanklich eine gewisse Anzahl benachbarter Elektronen zusammengefaßt. Ihr gemeinsamer Weg durch das Leitermaterial sei als *Stromröhre* bezeichnet. Je nach der jeweiligen Stromdichte ändert sich ihr Durchmesser. Das Zentrum solcher Stromröhren sind die Stromlinien in Bild 1.1.1. Ihr

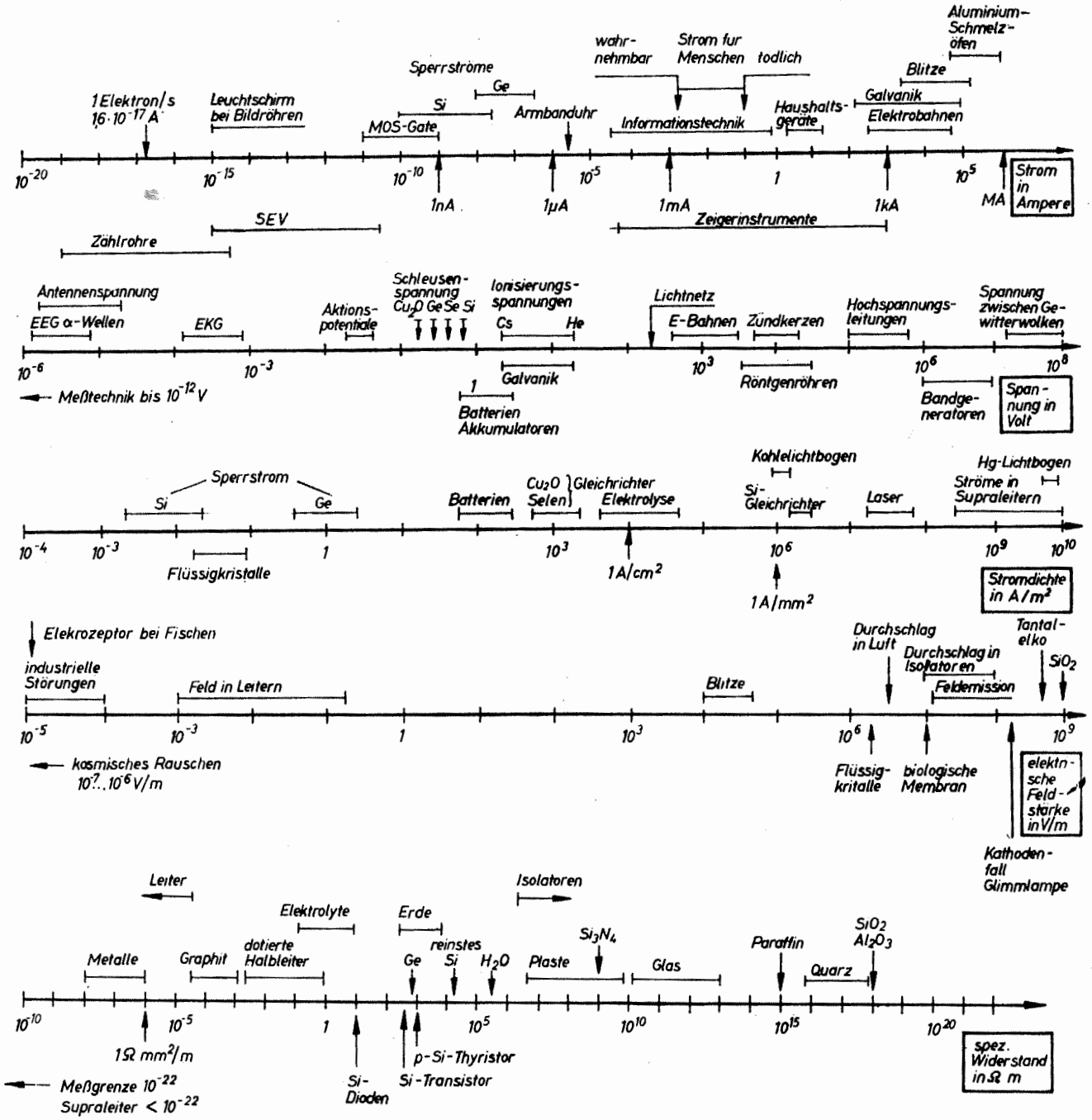


Bild 1.1.2. Überblick zu den Größenordnungen von Strom, Spannung, Stromdichte, Feldstärke und spezifischem Widerstand

Abstand entspricht dem Querschnitt der Stromröhren. Die mittlere Geschwindigkeit der Elektronen ist ihm umgekehrt proportional. Längs einer Stromröhre läßt sich dann die *elektrische Spannung* definieren,

$$U = \int_{P_1}^{P_2} \boldsymbol{E}\,\mathrm{d}\boldsymbol{s}\,. \tag{9}$$

Hierbei bedeuten P_1 und P_2 die beiden Endpunkte, zwischen denen die Stromröhre betrachtet wird.

Die Spannung ist also ein Differenzwert zwischen zwei Punkten. Wird nun ein Punkt willkürlich festgehalten, so kann die Spannung an den damit verglichenen Punkten als *Potential* betrachtet werden. Sehr oft wird der gemeinsam verwendete Bezugspunkt ins Unendliche verlegt. Allgemein ist das Potential jedoch nur eine relative Größe im Gegensatz zur absoluten Größe der Spannung.

Werden die verschiedenen Stromlinien in Bild 1.1.1 bezüglich eines Punktes betrachtet, so läßt sich auf vielen ein gleich großes Potential finden. Werden die Punkte gleichen Potentials verbunden, so entstehen *Äquipotentiallinien*, die im Bild ausgezogen sind. Äquipotential- und Stromlinien stehen immer aufeinander senkrecht.

Die Zusammenfassung von Stromlinien in einem Leiter heißt *Strömungsfeld*. Wird von den Stromzuführungen (Quellen +) und den Stromabführungen (Senken —) abgesehen, so können in einem Strömungsfeld Elektronen weder entstehen noch verschwinden. Deshalb ist jedes Flächenintegral über die Stromdichte, das nicht Quellen oder Senken einschließt, gleich Null.
Dies ist der *erste* KIRCHHOFF*sche Satz*:

$$\oint \boldsymbol{i}\, \mathrm{d}\boldsymbol{a} = 0\,. \tag{10}$$

Der *zweite* KIRCHHOFF*sche Satz* folgt unter der Voraussetzung, daß auf dem Integrationsweg keine Quellen und Senken vorhanden sind:

$$\oint \boldsymbol{E}\, \mathrm{d}\boldsymbol{s} = 0\,. \tag{11}$$

Entsprechend (9) bedeutet dies auch, daß jedes Potential unabhängig vom Weg ist.

Bild 1.1.2 zeigt einen Überblick zu den möglichen Größenordnungen von Spannungen, Strömen, Stromdichten, Feldstärken und spezifischen Widerständen.

1.1.2. Ohmscher Widerstand

1.1.2.1. Ohmsches Gesetz

Es liege ein homogener Leiter mit konstantem Querschnitt A vor. Die einzelnen Stromröhren besitzen dann einen unveränderlichen Querschnitt $\mathrm{d}\boldsymbol{a}$. Sie sind damit durch die zugehörige, konstante Stromdichte beschreibbar. So kann von der Stromdichte zur *Stromstärke* übergegangen werden:

$$I = \int \boldsymbol{i}\, \mathrm{d}\boldsymbol{a}\,. \tag{12}$$

Da außerdem die Beziehung für den Spannungsabfall (9) gilt, folgt mit dem mikro-Ohmschen Gesetz (8) die Beziehung des üblichen *Ohmschen Gesetzes*:

$$\boxed{U = RI}\,. \tag{13}$$

Der Proportionalitätsfaktor R heißt *Ohmscher Widerstand*:

$$R = \frac{U}{I} = \frac{\int \boldsymbol{E}\, \mathrm{d}\boldsymbol{s}}{\int \boldsymbol{i}\, \mathrm{d}\boldsymbol{a}}\,. \tag{14}$$

Bei konstantem ϱ und einem langen dünnen Draht der Länge l folgt hieraus

$$\boxed{R = \varrho \frac{l}{A}}\,. \tag{15}$$

Ein Kupferdraht mit 1 mm² Querschnitt und 1 km Länge hat somit einen Widerstand von etwa 17 Ω. Eine Stromstärke von 3 A führt zu einem Spannungsabfall von 52 mV/m. Das Symbol für den elektrischen Widerstand zeigt Bild 1.1.3a). Zuweilen wird statt des elektrischen Widerstandes sein reziproker Wert, der *Leitwert* benutzt. Er wird in Siemens, $S = \Omega^{-1}$, gemessen.

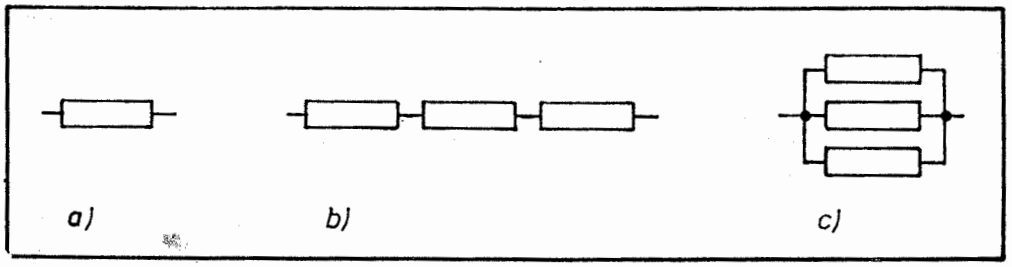


Bild 1.1.3. a) Symbol eines Widerstandes
b) Reihenschaltung und
c) Parallelschaltung von Widerständen

1.1.2.2. Zusammenschaltung von Widerständen

Mehrere Widerstände können gemäß Bild 1.1.3 in Reihe oder parallel zusammengeschaltet werden. Bei der *Reihenschaltung* muß der Strom alle Widerstände einzeln nacheinander durchlaufen. Sie können also durch einen Widerstandswert ersetzt werden, der gleich der Summe der einzelnen Widerstände ist:

$$\boxed{R_{\text{ges}} = R_1 + R_2 + R_3 + \dots}\,, \tag{16}$$

Eine *Parallelschaltung* von Widerständen wirkt dagegen so, als ob der Querschnitt für den Stromdurchfluß vergrößert wird. Gemäß Gl. (5) folgt daher für die Parallelschaltung der Widerstände die Beziehung

$$\boxed{\frac{1}{R_{\text{ges}}} = \frac{1}{R_1} + \frac{1}{R_2} + \frac{1}{R_3} + \dots}\,. \tag{17}$$

Wird anstelle der Widerstände ihr Leitwert benutzt, so ergibt sich

$$G_{\text{ges}} = G_1 + G_2 + G_3 + \dots\,. \tag{18}$$

Aus (16) bis (18) und dem Verhältnis von Widerstand und Leitwert kann ein einfaches *Dualitätsprinzip* abgeleitet werden. Es besagt, daß beim Übergang von Parallel- in Reihenschaltung die Berechnungsmethoden beibehalten werden können, wenn gleichzeitig vom Widerstand zum Leitwert übergegangen wird. Auch die umgekehrte Aussage gilt. Symbolisch gilt also

$$\begin{array}{ll} R & \text{Parallelschaltung} \\ \updownarrow & \updownarrow \\ G & \text{Reihenschaltung}\,. \end{array}$$

Wegen ihres häufigen Gebrauchs sei noch eine einfache Formel für die Parallelschaltung von zwei Widerständen angegeben:

$$\boxed{R_{\text{ges}} = \frac{R_1 \cdot R_2}{R_1 + R_2}}\,. \tag{19}$$

1.1.2.3. Widerstandsnetzwerke

In der Praxis existieren häufig komplizierte Zusammenschaltungen von Widerständen. Sie werden vielfach als Netzwerke bezeichnet. Einen möglichen Ausschnitt zeigt Bild 1.1.4a). Der Widerstand zwischen zwei Anschlüssen kann prinzipiell immer wenn auch z. T. recht aufwendig berechnet werden, indem der Reihe nach mittels Parallel-

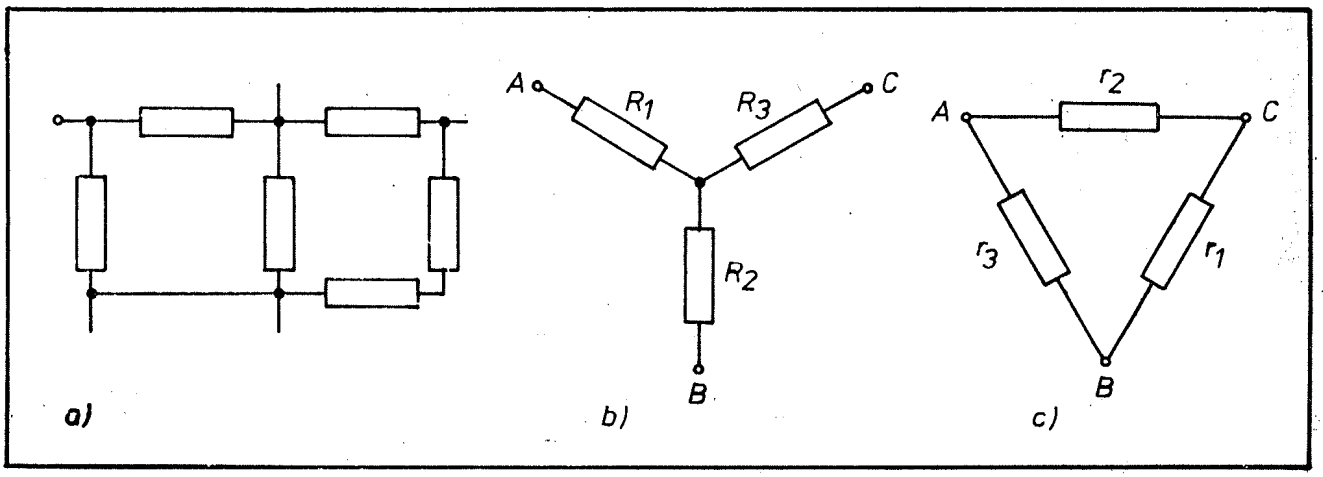


Bild 1.1.4. Beispiel eines Widerstandsnetzes (a) und zwei äquivalente Beispiele in Stern- (b) bzw. Dreiecksform (c)

und Reihenschaltung zur Vereinfachung Zwischenwerte berechnet werden. In der Elektrotechnik wurden vielfältige Regeln und Methoden entwickelt, die eine zielstrebige und einfache Berechnung ermöglichen. Hierzu sei auf die entsprechenden Standardwerke der Elektrotechnik verwiesen. Ein wichtiges Beispiel ist die sogenannte *Dreieck-Stern-Umwandlung*. Hierbei kann prinzipiell durch Messungen an den Klemmen A, B und C nicht entschieden werden, welche der beiden Schaltungen des Bildes 1.1.4 vorliegt. Wenn beide Schaltungen bezüglich der Punkte A, B, C auch gleichartige Widerstandswerte ergeben sollen, so können die Widerstände R_{AB}, R_{AC} und R_{CA} für beide Schaltungen jeweils getrennt berechnet werden. Aus dem Gleichsetzen ergeben sich drei Gleichungen mit jeweils drei Unbekannten, und als Lösung folgen die Werte der Tab. 1.1.3.

Tabelle 1.1.3. Umrechnung einer Stern- in eine Dreieckschaltung und umgekehrt
Die Bezeichnungen entsprechen Bild 1.1.4.

$R_1 = \dfrac{r_2 \cdot r_3}{r_1 + r_2 + r_3}$	$r_1 = \dfrac{R_1R_2 + R_2R_3 + R_1R_3}{R_1}$
$R_2 = \dfrac{r_1 \cdot r_3}{r^3 + r_2 + r_3}$	$r_2 = \dfrac{R_1R_2 + R_2R_3 + R_1R_3}{R_2}$
$R_3 = \dfrac{r_1 \cdot r_2}{r_1 + r_2 + r_3}$	$r_3 = \dfrac{R_1R_2 + R_2R_3 + R_1R_3}{R_3}$

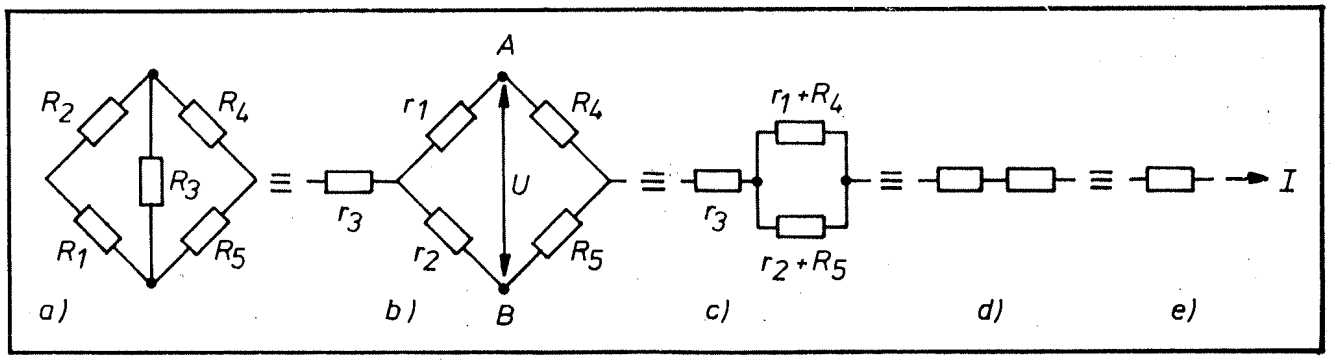


Bild 1.1.5. Berechnungsmethode der belasteten WHEATSTON-Meßbrücke (a). Die Umwandlung in eine formal unbelastete (b) erfolgt mit der Dreieck-Stern-Umrechnung (Bild 1.1.6, Tab. 1.1.3). Hier ist die Spannungsdifferenz A — B leicht angebbar. Für den Gesamtstrom werden nur die Reihen- und Parallelschaltung von Widerständen angewendet (c, d, e).

Bild 1.1.5 demonstriert eine Anwendung der Dreieck-Stern-Umwandlung an der WHEATSTONE*schen Brücke*. Die sonst schwierige Berechnung der Spannung an dem Widerstand R_3 bzw. des durch ihn hindurchfließenden Stromes wird mittels der Dreieck-Stern-Umwandlung wesentlich vereinfacht. Aus der durch R_3 belasteten Brücke wird eine unbelastete Brücke, und so ist der Spannungsabfall von A nach B einfach zu bestimmen. Danach ist es auch relativ leicht, den Gesamtwiderstand der Brücke und damit den Gesamtstrom zu berechnen.

1.1.3. Strom-Spannungs-Quellen

1.1.3.1. Wesen der Spannungsquelle

Um im Leiter eine Feldstärke zu erzeugen, sind Energien notwendig. Um diese Feldstärke ständig aufrecht zu erhalten, ist eine *Leistung* erforderlich, die den Verlust durch die *Stromwärme zu decken* hat. Beide Aufgaben kann eine Spannungsquelle übernehmen. Es sei angenommen, daß in ihrem Inneren Energien nichtelektrischer Art vorhanden sind und in elektrische umgewandelt werden. Bei den Batterien und Elementen wird hierzu *chemische Energie* verbraucht, während bei Dynamomaschinen *mechanische Energie* ausgenutzt wird. Im Inneren der Quelle werden die Elektronen dem Potentialunterschied entgegen bewegt. Es sei nun angenommen, daß innerhalb der Spannungsquelle die Ladung

$$Q = ne_0 \tag{20}$$

transportiert wird. Die dabei geleistete Arbeit ist dem zu überwindenden Spannungsunterschied proportional:

$$W = U \cdot Q\,. \tag{21}$$

Aus den beiden Gleichungen ergibt sich die Spannung der Quelle

$$U = \frac{W}{e_0} \tag{22}$$

als jene Arbeit, die in der Quelle je Elektron geleistet werden muß. Die Symbole für eine Spannungsquelle zeigt Bild 1.1.6a).

1.1.3.2. Wesen der Stromquelle

Neben der Spannungsquelle ist es in einigen Fällen auch sinnvoll, eine innere Stromquelle anzunehmen. Sie kann durch eine „Förderungseinrichtung" erklärt werden, die ständig Elektronen gemäß

$$I_K = \frac{ne}{t} \tag{23}$$

entgegen den äußeren Einflüssen transportiert. Das Symbol einer Stromquelle zeigt Bild 1.1.6.

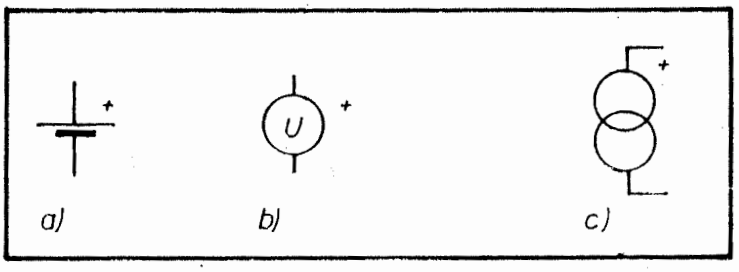


Bild 1.1.6. Symbole für Spannungs- und Stromquellen

a) für Primär- und Sekundärzellen; der lange Strich entspricht dem Pluspol. Die Bezeichnung ist also genau reziprok zu dem dünnen kurzen Stutzen der Monozelle

b) Allgemeines Spannungssymbol. In den Kreis wird vielfach der Spannungswert eingeschrieben, dann entfällt „U"

c) Symbol der Stromquelle. In die beiden verschlungenen Kreise wird zuweilen die Stromstärke eingetragen. Sie kann aber auch daneben geschrieben werden

Wird an diese Stromquelle kein Widerstand angelegt, so erzeugt sie einen ständig zunehmenden Ladungsdruck an ihren Klemmen. Ihre Klemmenspannung müßte also ständig zunehmen. Dies ist physikalisch nicht zulässig. Deshalb besitzt jede reale Stromquelle in ihrem Inneren Verluste, so daß die erreichbare Spannung begrenzt ist.

1.1.3.3. Zweipole

Es liege eine Schaltung vor, auf deren inneren Aufbau nur über zwei Anschlüsse zu folgern ist. Dies entspricht der „Black-box"-Methode der Kybernetik. Die Schaltung könnte u. a. aus einem einzelnen Widerstand, einem Widerstandsnetzwerk, einer Spannungsquelle, einer Zusammenschaltung von Spannungs- oder Stromquellen bzw. einer Zusammenschaltung von Spannungsquellen und Widerständen bestehen. Beim Zweipol interessieren jedoch nur alle möglichen Erkenntnisse, die von außen mittels der beiden Anschlußdrähte erschlossen werden können. In diesem Sinne kann auch eine übliche Batterie, z.B. eine Trockenbatterie, als Zweipol aufgefaßt werden. Sofern außen kein Widerstand angelegt ist, befindet sich an ihren Klemmen eine Leerlaufspannung U_L. Für das Auftreten von U_L gilt also $R_a \to \infty$, $I \to 0$. Werden dagegen die beiden äußeren Anschlüsse direkt verbunden, d. h. die Batterie kurzgeschlossen, so liegt an diesen Drähten keine Spannung, und es fließt der sogenannte Kurzschlußstrom I_K. Hierzu gehören die Bedingungen $R_a \to 0$, $U \to 0$. In keinem praktischen Fall wird I_K unendlich groß. Dies muß erklärt werden. Dazu kann von *drei unterschiedlichen Annahmen* ausgegangen werden:

— Die *Klemmenspannung ist eine Funktion des entnommenen Stromes.* Dann wäre die in der Spannungsquelle pro Elektron aufgewendete Arbeit vom Strom, d. h. von der Elektronenzahl je Sekunde abhängig. In Sonderfällen kann dieses zwar physikalisch richtig sein, für die Elektrotechnik ist die Annahme jedoch höchst unpraktisch.

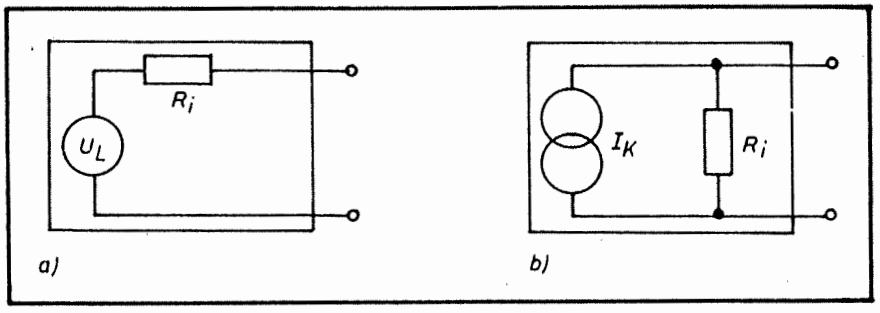


Bild 1.1.7. Elektrische Quelle in Form von Ersatzschaltbildern gemäß innerer a) Spannungs-, b) Stromquelle

— In der Batterie existiere eine *Spannungsquelle und ein Widerstand.* Dann ist die Spannung an der Spannungsquelle unabhängig vom Strom, und das äußere Verhalten der Spannungsquelle muß mit einem Innenwiderstand R_i beschreibbar sein. Mit der Zusatzannahme, daß R_i ebenfalls vom Strom unabhängig ist, ergeben sich besonders übersichtliche Verhältnisse. So entsteht eine Ersatzschaltung gemäß Bild 1.1.7a). Sie hat sich für die meisten praktischen Fälle bewährt, wenngleich sie auch nichts über die realen Verhältnisse in der Batterie aussagt. Entsprechend den Voraussetzungen kann der Innenwiderstand gemäß

$$\boxed{R_i = \frac{U_L}{I_K}} \tag{24}$$

berechnet werden.

— Im Innern der Batterie existiere eine *Stromquelle*. Hierbei wird die maximale Leerlaufspannung wiederum durch einen *Innenwiderstand* begrenzt. Er liegt jetzt parallel zur Stromquelle entsprechend Bild 1.1.7b). Für die Größe des Innenwiderstandes gilt wieder (24).

1.1.3.4. Nutzbare Spannung

Eine Strom- oder Spannungsquelle werde gemäß Bild 1.1.8 durch einen äußeren Lastwiderstand R_a belastet. An diesem Widerstand tritt dann ein Spannungsabfall auf, der dem hindurchfließenden Strom entspricht. Bei der Spannungsquelle berechnet sich der Strom wegen (13) (Ohmsches Gesetz) zu

$$I = \frac{U_L}{R_i + R_a}. \tag{25}$$

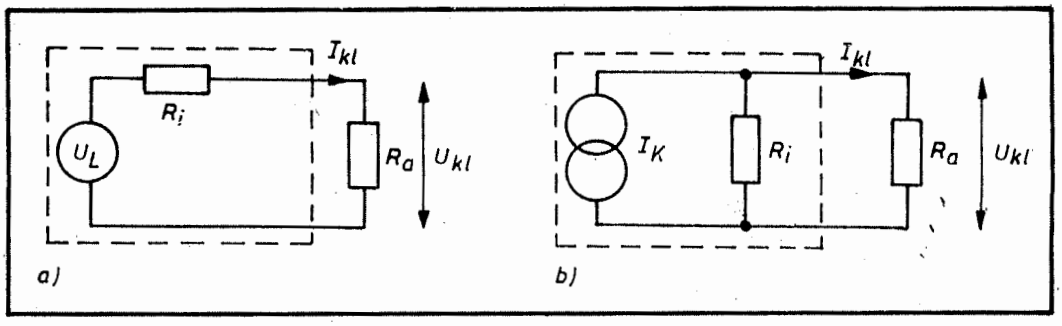


Bild 1.1.8. Belastung einer a) Spannungs- bzw. b) Stromquelle mit einem Widerstand R_a

Andererseits folgt für die Spannung an R_a (Klemmenspannung) gemäß $U = RI$ auch:

$$U_{kl} = U_L \frac{R_a}{R_i + R_a}. \tag{26}$$

Für die Stromersatzquelle ergibt sich die Klemmenspannung am einfachsten aus der Wirkung des Kurzschlußstromes auf die Parallelschaltung von Innen- und Lastwiderstand:

$$U_{kl} = I_K \frac{R_i R_a}{R_i + R_a}. \tag{27}$$

Da wegen (24) andererseits gilt

$$U_L = I_K R_i, \tag{28}$$

folgt wiederum (26). Damit ist bewiesen, daß beide Ersatzschaltungen von Bild 1.1.7 vollkommen gleichwertig sind. Sie stellen nur zwei unterschiedliche Betrachtungsweisen desselben Sachverhaltes dar.

Die Ersatz-Spannungsquelle ist sinnvoll dann anzuwenden, wenn $R_a \gg R_i$ gilt. Dann ergibt sich nämlich

$$U_{kl} = \frac{U_L}{1 + \frac{R_i}{R_a}} \approx U_L. \tag{29}$$

Die Stromquelle ist hingegen sinnvoll, wenn $R_i \gg R_a$ gilt. Dann läßt sich aus (27) der Ausdruck $I_K R_a$ betonen, und es folgt

$$U_{kl} = \frac{I_K R_a}{1 + \frac{R_a}{R_i}} \approx I_K R_a. \tag{30}$$

Bei der Spannungsquelle wird unter den oben genannten Bedingungen die nutzbare Spannung also in der Größenordnung der Klemmenspannung sein, während bei der Ersatz-Stromquelle sich die Klemmenspannung näherungsweise aus dem Produkt Kurzschlußstrom mal Lastwiderstand berechnen läßt. Mit diesen angeführten Daten kann das *Dualitätsprinzip* erweitert werden:

U	R	Parallelschaltung
↕	↕	↕
I	G	Reihenschaltung .

1.1.3.5. Anpassungen

Für die Anpassung einer Strom- bzw. Spannungsquelle an einen Lastwiderstand sind *drei Fälle* zu unterscheiden:

1. *höchste Spannung*: Sie ergibt sich näherungsweise durch den Leerlaufbetrieb $R_a \to \infty$.
2. *höchster Strom*: Er ergibt sich näherungsweise im Kurzschlußbetrieb, also für $R_a \to 0$.

In beiden Fällen bleibt das Produkt UI, d. h. die Leistung Null.

3. *Leistungsanpassung*: Sie muß ersichtlich zwischen den beiden vorgenannten Fällen für ein bestimmtes R_a gegeben sein und soll durch die Extremwertberechnung bestimmt werden.

Es gilt

$$P_a = U_{kl} I_{kl} = \frac{U_{kl}^2}{R_a}. \tag{31}$$

Unter Verwendung von (26) folgt hieraus

$$P_a = U_L^2 \frac{R_a}{(R_a + R_i)^2}. \tag{32}$$

Das Maximum folgt aus $\mathrm{d}P_a/\mathrm{d}R_a = 0$ und existiert bei

$$\boxed{R_i = R_a}. \tag{33}$$

Die hierbei am Lastwiderstand vorhandene Leistung beträgt

$$P_{max} = \frac{U_L^2}{4R_i} = \frac{I_K}{2}\frac{U_L}{2}. \tag{34}$$

Sie ist genau ein Viertel des Produktes $I_K U_L$. Diese prinzipiell mögliche Leistung wird bei Kurzschluß in der Batterie selbst in Wärme umgewandelt. Andererseits folgt aus (33), daß bei Leistungsanpassung dieselbe Leistung am Lastwiderstand wie in der Spannungs- bzw. Stromquelle in Wärme umgesetzt wird. Die dann insgesamt in Wärme umgesetzte Leistung ist die Hälfte von $I_K U_L$.

1.1.3.6. Verlauf der Leistungsanpassung

Die Beziehung (32) wird jetzt auf die maximale Leistung von (34) bezogen und ergibt so einen normierten, relativen Leistungswert

$$\frac{P_a}{P_{max}} = \frac{4P_a}{U_L I_K} = \frac{4R_i R_a}{(R_i + R_a)^2}. \tag{35}$$

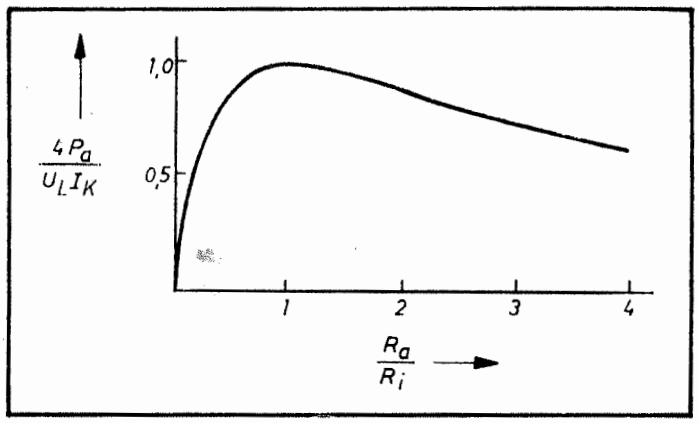


Bild 1.1.9. Verlauf der an einem Lastwiderstand erreichbaren normierten Leistung. Die Kurve stellt die Auswertung von Gl. (35) dar

Dieser Verlauf ist in Abhängigkeit vom Lastwiderstand in Bild 1.1.9 dargestellt. Daraus folgt, daß sich eine Fehlanpassung unterschiedlich stark auf die erreichbare Leistung auswirkt. Für R_a kleiner als R_i, Unteranpassung, fällt die Kurve relativ steil gegen Null ab. Dagegen erfolgt der Abfall bei einer Überanpassung, also R_a größer R_i, wesentlich langsamer. Das bedeutet, daß eine Überanpassung relativ weniger schädlich als eine Unteranpassung ist.

1.1.4. Vermaschte Stromkreise

Die meisten praktischen Schaltungen bestehen aus einer Vielzahl von Widerständen und Spannungs- bzw. Stromquellen. Häufig sind dann die Ströme und Spannungen in diesen Schaltungen zu berechnen. Hierzu gibt es eine Vielzahl von Methoden. Sie führen je nach dem vorliegenden Fall und den vorhandenen Erfahrungen mehr oder weniger schnell zum Ergebnis. Hier seien nur einige besonders übersichtliche Methoden mitgeteilt.

1.1.4.1. Ersatzspannungsquelle

Aus einem komplizierten Netzwerk seien entsprechend Bild 1.1.10 zwei Anschlußpunkte A und B herausgegriffen. An ihnen können die Leerlaufspannung U_L und der Kurzschlußstrom I_K bestimmt werden. Beide Anschlüsse bilden dann, von außen

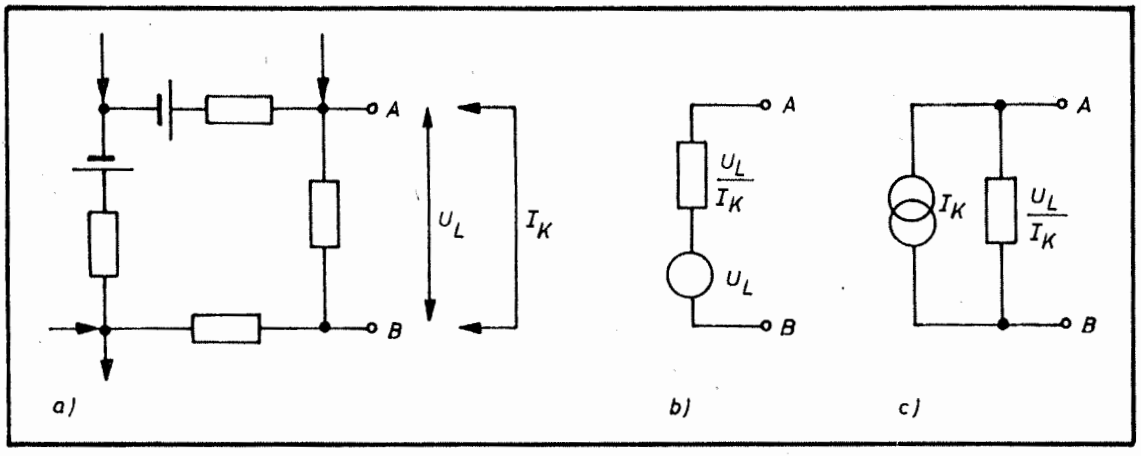


Bild 1.1.10. Bestimmung der Ersatzspannungs- (b) bzw. Stromquelle (c) eines Netzwerkes. Der Innenwiderstand $R_i = U_L/I_K$ ist also lediglich durch zwei Messungen für die Leerlaufspannung U_L und den Kurzschlußstrom I_K zu bestimmen

gesehen, einen Zweipol. Seine typischen Größen sind gemäß Abschnitt 1.1.3.3. durch eine Spannungs- bzw. Stromquelle und den zugehörigen Innenwiderstand

$$R_i = \frac{U_L}{I_K} \tag{36}$$

beschreibbar. Dieses Verfahren ist prinzipiell auf alle möglichen Punkte des Netzwerkes anwendbar.

1.1.4.2. Überlagerungsgesetz

Zur Berechnung werden in Gedanken alle Spannungsquellen des Netzwerkes bis auf eine kurzgeschlossen. Dann liegt für diese Spannungsquelle ein reines Widerstandsnetzwerk vor, das nach den dafür geltenden Regeln zu berechnen ist. In genau der gleichen Weise kann dieses Verfahren nacheinander für alle anderen Quellen durchgeführt werden. Anschließend werden die Einzelströme durch jeden einzelnen Widerstand zum realen Strom aufaddiert. Mit diesen realen Strömen lassen sich auch die Spannungen an den einzelnen Widerständen berechnen.

1.1.4.3. Maschenregel

In jedem Netzwerk existieren Maschen. Eine Masche ist dadurch gekennzeichnet, daß sie durch einen umfließenden Strom, der keine Kreuzungen aufweisen darf und bei den Verzweigungen vernachlässigt werde, gegeben ist (Bild 1.1.11). In dieser

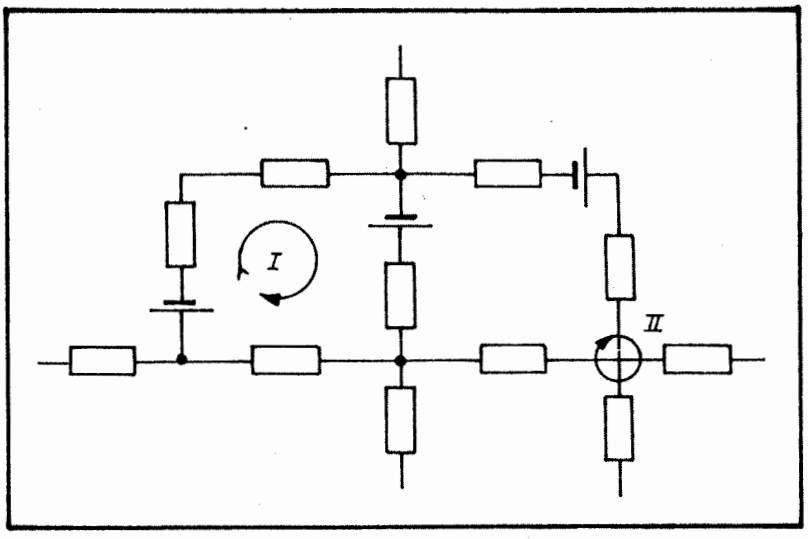


Bild 1.1.11. Zur Erklärung von Masche (Kreis I) und Knoten (Kreis II)

Masche ist ein Strompfeil als Umlaufrichtung vorzugeben. Spannungen und Ströme der gleichen Richtung werden positiv, entgegengesetzt gerichtete negativ gezählt. Unter diesen Bedingungen muß für die Summe der Spannungen und der Spannungsabfälle

$$\Sigma U = \Sigma IR \tag{37}$$

gelten. Diese Betrachtung wird der Reihe nach für alle Maschen des Netzwerkes durchgeführt, und aus der Überlagerung lassen sich die wirklichen Ströme und Spannungen bestimmen. Die Maschenregel entspricht dem zweiten KIRCHHOFFschen Satz von Abschnitt 1.1.1.5., insbesondere (11).

1.1.4.4. Knotenregel

Ein Knoten ist ein Verzweigungspunkt mit zu- und abfließenden Strömen (Bild 1.1.11). Da in ihm keine Ströme verloren gehen und entstehen können, muß entsprechend dem ersten KIRCHHOFFschen Satz (10)

$$\sum I = 0 \tag{38}$$

gelten. Auch diese Regel ist der Reihe nach für alle Knoten anzuwenden.

1.1.5. Vierpole

Ein Vierpol ist ein Netzwerk mit je zwei Eingangs- und Ausgangsklemmen (Bild 1.1.12). Für allgemeine Betrachtungen ist es zunächst uninteressant, was im Innern des Vierpols für eine Schaltung vorliegt. Bedeutsam sind vielmehr die *Zu-*

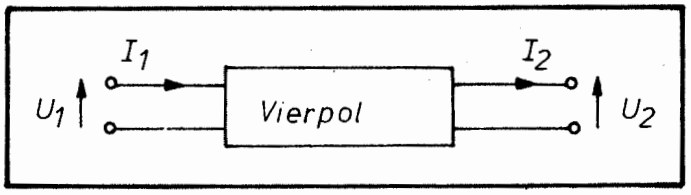


Bild 1.1.12. Vierpol als Black-Box mit zwei Ein- und zwei Ausgangsleitungen

sammenhänge zwischen den Eingangs- und Ausgangsklemmen. Allgemein können von den vier Größen I_1, U_1, I_2 und U_2 zwei als unabhängige Variable X_1 und X_2 und die beiden anderen als abhängige Y_1 und Y_2 betrachtet werden. Dann gilt das Gleichungssystem

$$\begin{matrix} Y_1 \\ Y_2 \end{matrix} = \boldsymbol{K} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \hat{=} \begin{matrix} Y_1 = k_{11}X_1 + k_{12}X_2 \\ Y_2 = k_{21}X_1 + k_{22}X_2 \,. \end{matrix} \tag{39}$$

Insgesamt bestehen *12 mögliche Zuordnungen* der Ströme und Spannungen zu den Variablen X, Y, wobei jeweils zwei immer nur einer Vertauschung der Zeilen ent-

Tabelle 1.1.4. Grundgleichungen für Vierpole

Nr.	Name	Formel	Einheit der Koeffizienten
1	Widerstandsmatrix	$\begin{pmatrix} U_1 \\ U_2 \end{pmatrix} = \mathbf{Z} \begin{pmatrix} I_1 \\ I_2 \end{pmatrix}$	Widerstand
2	Leitwertmatrix	$\begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = \mathbf{Y} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix}$	Leitwert
3	Kettenmatrix	$\begin{pmatrix} U_1 \\ I_1 \end{pmatrix} = \mathbf{U} \begin{pmatrix} U_2 \\ I_2 \end{pmatrix}$	gemischt
4	Hybridmatrix	$\begin{pmatrix} U_1 \\ I_2 \end{pmatrix} = \mathbf{G} \begin{pmatrix} I_1 \\ U_2 \end{pmatrix}$	gemischt
5	Parallel-Reihenmatrix	$\begin{pmatrix} I_1 \\ U_2 \end{pmatrix} = \mathbf{C} \begin{pmatrix} U_1 \\ I_2 \end{pmatrix}$	gemischt
6	Kettenmatrix rückwärts	$\begin{pmatrix} U_2 \\ I_2 \end{pmatrix} = \mathbf{B} \begin{pmatrix} U_1 \\ I_1 \end{pmatrix}$	gemischt

Tabelle 1.1.5. Beziehungen zur Umrechnung der Kenngrößen von Vierpolen
Die Dreiecke bedeuten Determinante, also z. B. $\triangle^Y = Y_{11}\,Y_{22} - Y_{12}\,Y_{21}$.

	Y	Z	A	B	H	C
Y	$\begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix}$	$\frac{1}{\triangle^Z}\begin{bmatrix} Z_{22} & -Z_{12} \\ -Z_{21} & Z_{11} \end{bmatrix}$	$\frac{1}{A_{12}}\begin{bmatrix} A_{22} & -\triangle^A \\ 1 & -A_{11} \end{bmatrix}$	$\frac{1}{B_{12}}\begin{bmatrix} -B_{11} & 1 \\ -\triangle^B & B_{22} \end{bmatrix}$	$\frac{1}{H_{11}}\begin{bmatrix} 1 & -H_{12} \\ H_{21} & \triangle^H \end{bmatrix}$	$\frac{1}{C_{22}}\begin{bmatrix} \triangle^C & C_{12} \\ -C_{21} & 1 \end{bmatrix}$
Z	$\frac{1}{\triangle^Y}\begin{bmatrix} Y_{22} & -Y_{12} \\ -Y_{21} & Y_{11} \end{bmatrix}$	$\begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}$	$\frac{1}{A_{21}}\begin{bmatrix} A_{11} & -\triangle^A \\ 1 & -A_{22} \end{bmatrix}$	$\frac{1}{B_{21}}\begin{bmatrix} -B_{22} & 1 \\ -\triangle^B & B_{11} \end{bmatrix}$	$\frac{1}{H_{22}}\begin{bmatrix} \triangle^H & H_{12} \\ -H_{21} & 1 \end{bmatrix}$	$\frac{1}{C_{11}}\begin{bmatrix} 1 & -C_{12} \\ C_{21} & \triangle^C \end{bmatrix}$
A	$\frac{1}{Y_{21}}\begin{bmatrix} -Y_{22} & 1 \\ -\triangle^Y & Y_{11} \end{bmatrix}$	$\frac{1}{Z_{21}}\begin{bmatrix} Z_{11} & -\triangle^Z \\ 1 & -Z_{22} \end{bmatrix}$	$\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$	$\frac{1}{\triangle^B}\begin{bmatrix} B_{22} & -B_{12} \\ -B_{21} & B_{11} \end{bmatrix}$	$\frac{1}{H_{21}}\begin{bmatrix} -\triangle^H & H_{11} \\ -H_{22} & 1 \end{bmatrix}$	$\frac{1}{C_{21}}\begin{bmatrix} 1 & -C_{22} \\ C_{11} & -\triangle^C \end{bmatrix}$
B	$\frac{1}{Y_{12}}\begin{bmatrix} -Y_{11} & 1 \\ -\triangle^Y & -Y_{22} \end{bmatrix}$	$\frac{1}{Z_{12}}\begin{bmatrix} Z_{22} & -\triangle^Z \\ 1 & -Z_{11} \end{bmatrix}$	$\frac{1}{\triangle^A}\begin{bmatrix} A_{22} & -A_{12} \\ -A_{21} & -A_{11} \end{bmatrix}$	$\begin{bmatrix} B_{11} & B_{12} \\ -B_{21} & B_{22} \end{bmatrix}$	$\frac{1}{H_{12}}\begin{bmatrix} 1 & -H_{11} \\ H_{22} & -\triangle^H \end{bmatrix}$	$\frac{1}{C_{12}}\begin{bmatrix} -\triangle^C & C_{22} \\ -C_{11} & 1 \end{bmatrix}$
H	$\frac{1}{Y_{11}}\begin{bmatrix} 1 & -Y_{12} \\ Y_{21} & \triangle^Y \end{bmatrix}$	$\frac{1}{Z_{22}}\begin{bmatrix} \triangle^Z & Z_{12} \\ -Z_{21} & 1 \end{bmatrix}$	$\frac{1}{A_{22}}\begin{bmatrix} A_{12} & \triangle^A \\ 1 & -A_{21} \end{bmatrix}$	$\frac{1}{B_{11}}\begin{bmatrix} -B_{12} & 1 \\ \triangle^B & B_{21} \end{bmatrix}$	$\begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}$	$\frac{1}{\triangle^C}\begin{bmatrix} C_{22} & -C_{12} \\ -C_{21} & C_{11} \end{bmatrix}$
C	$\frac{1}{Y_{22}}\begin{bmatrix} \triangle^Y & Y_{12} \\ -Y_{21} & 1 \end{bmatrix}$	$\frac{1}{Z_{11}}\begin{bmatrix} 1 & Z_{12} \\ Z_{21} & \triangle^Z \end{bmatrix}$	$\frac{1}{A_{11}}\begin{bmatrix} A_{21} & \triangle^A \\ 1 & -A_{12} \end{bmatrix}$	$\frac{1}{B_{22}}\begin{bmatrix} -B_{21} & 1 \\ \triangle^B & B_{12} \end{bmatrix}$	$\frac{1}{\triangle^H}\begin{bmatrix} H_{22} & -H_{12} \\ -H_{21} & H_{11} \end{bmatrix}$	$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$

sprechen. Dadurch gibt es sechs verschiedene Gleichungssysteme, die mehr oder weniger große praktische Bedeutung für unterschiedliche Anwendungen besitzen. Sie sind in Tab. 1.1.4 dargestellt. Grundsätzlich lassen sich die *Parameter aller Gleichungen ineinander überführen.* Dies zeigt Tab. 1.1.5. Weiter existieren z. T. einfache Formeln für die Zusammenschaltung von Vierpolen. Dies zeigt für die fünf wichtigsten Fälle Bild 1.1.13. Weiter sind darin die *elementaren Vierpole* enthalten, aus denen sich alle Vierpole zusammensetzen lassen. Für die oft vorkommende *Kettenschaltung* von Abzweigschaltungen ist unten das typische Spannungsteilerglied in der A-Matrix dargestellt.

Für die *Systematik* der Vierpole haben vier Bezeichnungen besondere Bedeutung. Der *aktive Vierpol* enthält in seinem Inneren Quellen und ist damit imstande, nach außen Leistung abzugeben. Der *passive Vierpol* enthält keine Quellen und verbraucht damit in jedem Fall Energie. Bei einem *symmetrischen Vierpol* können Eingang und Ausgang vertauscht werden, ohne daß sich am Verhalten des Vierpols in der vor-

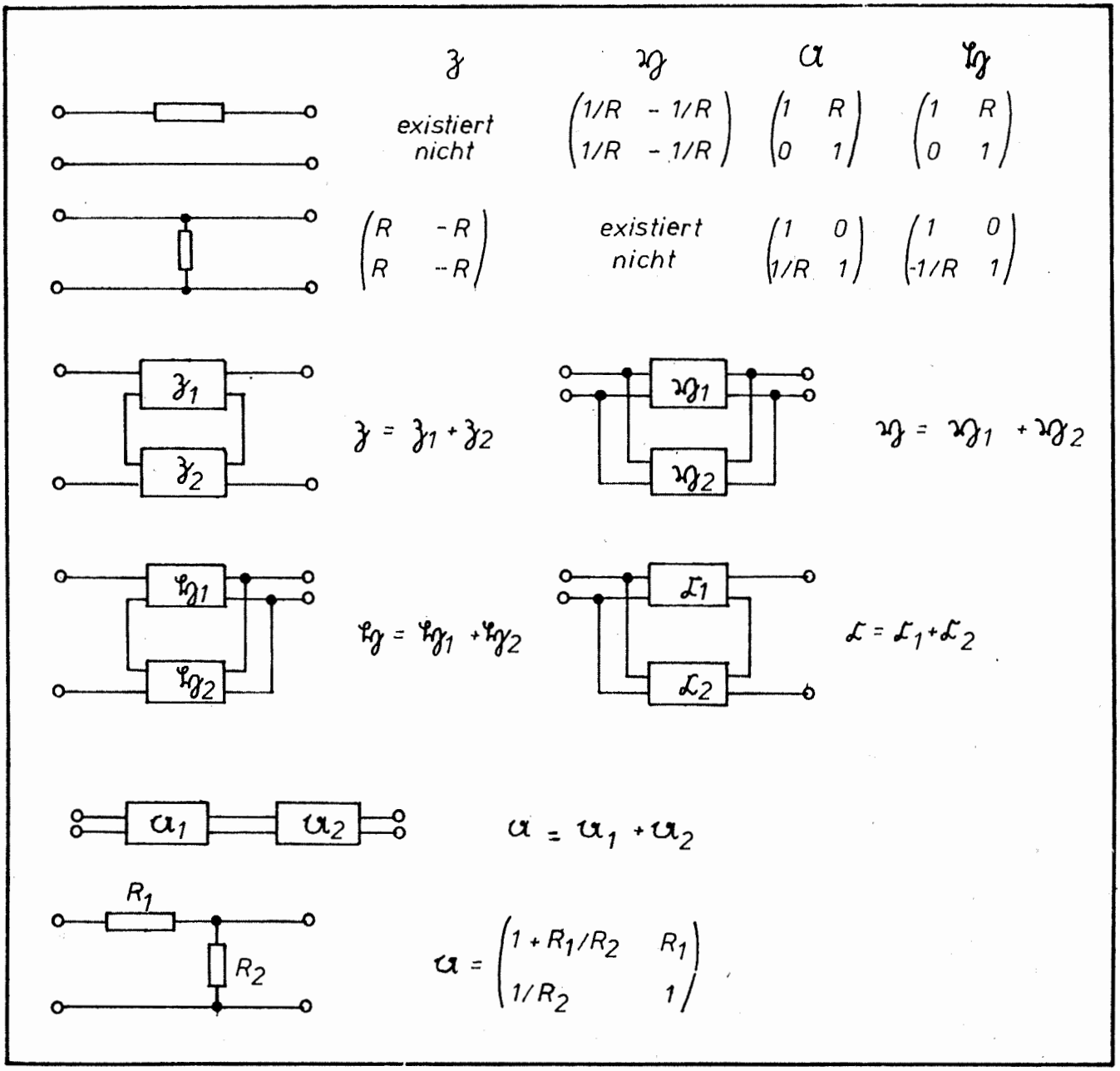


Bild 1.1.13. Beispiele von Kenngrößen einfacher Vierpole sowie Zusammenschaltungen von Vierpolen. Für **A** muß gelten $\mathbf{A} = \mathbf{A}_1 * \mathbf{A}_2$

gegebenen Schaltung etwas ändert. Der *lineare Vierpol* enthält keine strom- und spannungsabhängigen Größen. Für ihn gelten also die KIRCHHOFFschen Gesetze.

Als weitere typische Kenngrößen für einen Vierpol seien u. a. die *Ersatzquelle* für den Ein- und Ausgang, seine *Spannungsübertragung* U_2/U_1, seine *Stromübertragung* I_2/I_1 und der Kernwiderstand genannt. Über den *Kernwiderstand* vorwärts M_{21} erzeugt ein Eingangsstrom I_1 am Ausgang die Spannung U_2 gemäß

$$U_2 = M_{21} I_1 . \tag{40}$$

Beim Kernwiderstand rückwärts sind in (40) nur die Indizes zu vertauschen. Der Kernwiderstand entspricht dem Koeffizienten Z_{21} bzw. Z_{12} der Widerstandsmatrix. Bei all diesen Größen ist weiter der *Abschluß des Vierpols am Eingang bzw. Ausgang* zu beachten. Besonders wichtige Fälle sind der *Leerlauf* und der *Kurzschluß*.

Neben den bisher dargestellten Matrizengleichungen auf der Basis von Strömen und Spannungen existieren noch zwei andere Vierpolgleichungen. Die *Streumatrix*, zuweilen auch *Betriebsmatrix* genannt, hat vor allem bei Höchstfrequenzen Bedeutung. Sie benutzt die verschiedenen *Reflexionsfaktoren* von Wellen. Für homogen verteilte Bauelemente, z.B. in Kabeln und Leitungen, hat die *Leitungs- oder Betriebsparametergleichung* Bedeutung. In ihr gehen u. a. *Wellenwiderstände*, Dämpfungsmaße und auf die Längeneinheit bezogene Bauelementegrößen ein. Für die Rechnung werden bevorzugt hyperbolische Funktionen verwendet. Beide Matrizendarstellungen lassen sich natürlich auf die hier benutzten Gleichungen überführen und umgekehrt. Für weitere Einzelheiten sei auf [R 12, Bd. 2, S. 164, P 13 und V 7, Bd. 1, S. 224] verwiesen.

1.1.6. Technische Widerstände

In elektronischen Geräten befinden sich generell viele Widerstände. Ihre Anzahl wird zwar durch die Mikroelektronik je Gerät gesenkt. So fiel sie für ein Farbfernsehgerät in den letzten 10 Jahren von ca. 600 auf 200. Mit der steigenden Anzahl elektronischer Geräte nehmen dennoch ihre Produktionszahlen zu. 1978 wurden in der Welt ca. $4 \cdot 10^{10}$ Widerstände eingesetzt. Infolge der großen Anwendungsbreite der Widerstände haben sich mehrere Technologien, Ausführungsformen und Qualitätsparameter entwickelt. Von zwei Ausnahmen abgesehen, wird als Grundmaterial für die Widerstände ein länglicher, alkalifreier Porzellankörper verwendet. Auf dieses Substrat wird der Widerstandsdraht aufgewickelt bzw. eine homogene Widerstandsschicht aufgebracht. Die Widerstandsschicht wird dann wendelförmig getrennt. Dadurch wird der notwendige Widerstandswert eingestellt. Dieser Prozeß erfolgt mittels Schleifen, Sandstrahlen oder Laser. Zum Schutz werden sowohl Draht- als auch Schichtwiderstände mit Schutzschichten, z.B. aus Lacken, Glasierungen oder Zementen versehen. Die Drahtanschlüsse der Widerstände sind heute fast ausnahmslos axial ausgeführt.

Der *Kohleschichtwiderstand* wird z. Z. noch am meisten verwendet. Mittels thermischer Spaltung von z.B. Benzin oder Heptan bei ca. 1000 °C wird die Kohleschicht erzeugt. Immer mehr Bedeutung gewinnt der mit weitaus besseren Eigenschaften ausgestattete *Metallschichtwiderstand*. Hierbei wird die Schicht, meist NiCr, in Ausnahmefällen auch Phosphor-dotiertes Nickel oder, für höchste Ansprüche Gold-Platin 80:20, auf den Porzellankörper aufgedampft oder gesputtert. Bei den *Metalloxidwiderständen* wird Zinnchlorid unter Zusatz von Antimonchlorid bei 700 bis 1000 °C auf den Isolierkörper gesprüht. Dabei bilden sich halbleitende Schichten aus Zinn- und Antimonoxiden. Bei den Metallglasur- bzw. *Dickschichtwiderständen* wird wie bei der Dickschichttechnik (vgl. 9.4.4.) der Isolierkörper mit einer Widerstandsschicht aus Glaspulver, Metallteilchen und einem organischen Lösungsmittel überzogen. Dieses Material wird dann bei hohen Temperaturen fest aufgeschmolzen. Es entsteht eine

Tabelle 1.1.6. International genormte Widerstandsreihen

Reihe E 6 ± 20% Tol.	Reihe E 12 ± 10% Tol.	Reihe E 24 ±5% Tol.	Reihe E 48 ± 1% Tol.	Reihe E 96 ± 1% Tol.	Reihe E 6 ± 20% Tol.	Reihe E 12 ± 10% Tol.	Reihe E 24 ± 5% Tol.	Reihe E 48 ± 2% Tol.	Reihe E 96 ± 1% Tol.
1,0	1,0	1,0	1,00	1,00				3,16	3,16
				1,02					3,24
			1,05	1,05	3,3	3,3	3,3	3,32	3,32
				1,07					3,40
		1,1	1,10	1,10				3,48	3,48
				1,13					3,57
			1,15	1,15			3,6	3,65	3,65
				1,18					3,74
	1,2	1,2	1,21	1,21				3,83	3,83
				1,24		3,9	3,9		3,92
			1,27	1,27				4,02	4,02
		1,3		1,30					4,12
			1,33	1,33				4,22	4,22
				1,37			4,3		4,32
			1,40	1,40				4,42	4,42
				1,43					4,53
			1,47	1,47				4,64	4,64
1,5	1,5	1,5		1,50	4,7	4,7	4,7		4,75
			1,54	1,54				4,87	4,87
				1,58					4,99
		1,6	1,62	1,62			5,1	5,11	5,11
				1,65					5,23
			1,69	1,69				5,36	5,36
				1,74					5,49
			1,78	1,78		5,6	5,6	5,62	5,62
	1,8	1,8		1,82					5,76
			1,87	1,87				5,90	5,90
				1,91					6,04
			1,96	1,96			6,2	6,19	6,19
		2,0		2,00					6,34
			2,05	2,05				6,49	6,49
				2,10					6,65
			2,15	2,15	6,8	6,8	6,8	6,81	6,81
2,2	2,2	2,2		2,21					6,98
			2,26	2,26				7,15	7,15
				2,32					7,32
			2,37	2,37			7,5	7,50	7,50
		2,4		2,43					7,68
			2,49	2,49				7,87	7,87
				2,55					8,06
			2,61	2,61		8,2	8,2	8,25	8,25
				2,67					8,45
	2,7	2,7	2,74	2,74				8,66	8,66
				2,80					8,87
			2,87	2,87			9,1	9,09	9,09
				2,94					9,31
		3,0	3.01	3,01				9,53	9,53
				3,09					9,76

leitende Glasurschicht, die auch *Cermet* genannt wird. Als Metallteilchen werden u. a. Tantal, Tantalkarbid, Titan. Titannitrid oder Ruthenium verwendet.

Nicht auf einen Isolierkörper aufgebracht sind die vor allem in England und früher stark gebräuchlichen *Kohlemassewiderstände*. Sie werden aus Ruß bzw. Graphitpulver und Quarzmehl in Verbindung mit einem härtbaren Kunstharz hergestellt.

Für hohe Anforderungen sind in der letzten Zeit *Kompaktmetallwiderstände* entwickelt worden. Sie werden in Dünnschichttechnik (vgl. 9.4.5.) auf ein ebenes Glassubstrat aufgedampft oder gesputtert. Mittels Photolithografie oder anderer Techniken werden mäanderförmige Bahnen erzeugt. Die beiden Anschlüsse werden einseitig im Rastermaß der Leiterplatten herausgeführt. Der Widerstand wird dann quaderförmig eingegossen. Seine Abmessungen betragen etwa 6,4 × 5,8 mm².

Ebenfalls in der letzten Zeit werden immer häufiger *Widerstandsnetzwerke* eingesetzt. Sie enthalten in einem Gehäuse (meist auch auf einem Substrat) gleichzeitig mehrere Widerstände. Es gibt Ausführungen mit Anschlüssen in einer Reihe, aber auch in DIL-Gehäuse (vgl. 4.5.2.).

Schichtwiderstände sind durch eine größere Anzahl von *Eigenschaftsparametern* gekennzeichnet. Von entscheidender Bedeutung ist der *Widerstandswert*. Infolge der Vielzahl der Anforderungen haben sich hier Normen entwickelt, welche den Streuungen im Herstellungsprozeß angepaßt sind. Es werden keine ganzzahligen Werte, sondern gleichmäßige Abstufungen entsprechend den geforderten *Toleranzen* verwendet, z.B. $\sqrt[12]{10}$. Die wichtigsten *internationalen Normreihen* zeigt Tab. 1.1.6.

Die für einen Widerstand *zulässige Leistung* hängt von mehreren Parametern ab, u. a. dem Volumen, dem Material, der Wärmeableitung und der Umgebungstemperatur. Es wird immer ein Nennwert der Leistung P_N angegeben. Dieser Wert muß in Abhängigkeit von der auftretenden Temperatur reduziert werden. Da die wirklichen physikalischen Zusammenhänge recht kompliziert sind, wurden *Unterlastkurven* (derating curves) entwickelt. Sie verwenden zwei Temperaturen: eine typische Umgebungstemperatur T_N (meist zwischen 40 und 70 °C) und eine Maximaltemperatur T_{max}, die für den Widerstand gerade noch zulässig ist (vgl. Tab. 1.1.7). Zwischen den beiden Temperaturen wird ein linearer Abfall der zulässigen Leistung festgelegt (Bild 1.1.14).

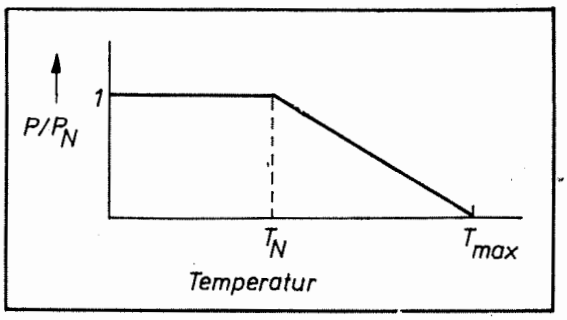


Bild 1.1.14. Schematischer Verlauf der Unterlastkurve (derating curve)

Weitere wichtige Werte der Widerstände sind die *Toleranz*, der *Temperaturkoeffizient*, sein *Rauschen*, der *Spannungskoeffizient* (gibt an, um wieviel sich der Widerstandswert mit der angelegten Spannung ändert). Sie sind als Richtwerte in Tab. 1.1.7 eingetragen.

Das *Rauschen bei Widerständen* hat mehrfache Ursachen: einmal das thermodynamische Rauschen, dann ein Rauschen, welches von dem Strom an den inneren Kontaktstellen des Widerstandsmaterials verursacht wird, und, in Sonderfällen, ein Rauschen, das an den Übergangsstellen zu den Anschlüssen entsteht. Während das

Tabelle 1.1.7. Richtwerte einiger Widerstandstypen
Präzisionswiderstände lassen z. T. wesentlich bessere Werte erreichen. Sie sind jedoch nicht für Kohlemasse- und Kohleschichtwiderstände möglich.

	Kohlemasse	Kohleschicht	Metalloxid	Metallglasur	Metallschicht	Kompaktmetall	Draht
Widerstandsbereich in Ohm	$1 \dots 10^8$	$1 \dots 10^8$	$12 \cdot 10^6$	$10 \dots 10^{10}$	$1 \dots 10^7$	$10 \dots 10^6$	$10^2 \dots 10^6$
Belastung Nennwert in Watt	0,1 ... 4	0,1 ... 2	0,1 ... 10	0,1 ... 10	0,1 ... 2	0,3	1 ... 1000
Toleranzbereich (normal)	5 ... 20	5	1 ... 5	0,5 ... 1	0,2 ... 1	0,005	0,1 ... 10
Max. Betriebstemp. in °C	110	150	250	100 ... 250	150 ... 300	200	150 ... 450
Temperaturkoeff. in 10^{-6}/K	$-2000 \dots -3000$	$-100 \dots -1500$	± 350	$\pm 200 \dots \pm 500$	± 200	± 1	$-40 \dots +600$
Rauschen in μV/V	2 ... 6	0,05 ... 1	0,03	0,4 ... 1	0,015 ... 0,1	vernachlässigbar	
Spannungskoeffizient in 10^{-6}/V	3000	100	2	vernachlässigbar			
Materialwiderstand in Ωcm	$2 \cdot 10^2 \dots 2 \cdot 10^5$	$4 \cdot 10^{-3}$	10^{-3}	0,2 ... 200	$10^{-6} \dots 10^{-4}$	$10^{-6} \dots 10^{-4}$	$10^{-6} \dots 10^{-4}$
Schichtdicke in m	massiv	$2 \cdot 10^{-8} \dots 10^{-7}$	$(2 \dots 8) \cdot 10^{-7}$	$\approx 10^{-6}$	$\approx 10^{-6}$	$\approx 10^{-8}$	Draht

thermische Rauschen von Null bis zu mehreren GHz konstant, also weiß ist, zeigen die anderen beiden Effekte eine starke Zunahme zu tiefen Frequenzen.

Genormte *Soll-Temperaturkoeffizienten* existieren mit den Grenzen 15; 25; 50; 100 und $200 \cdot 10^{-6}$/K.

Infolge der oft vorhandenen Kleinheit der Widerstände wird der Widerstandswert und die Toleranz mit *Farbpunkten* bzw. *Farbringen* codiert. Hierfür gilt Tab. 1.1.8.

Tabelle 1.1.8. Farbreihe zur Kennzeichnung von Widerständen
Toleranz ohne Kennzeichnung: $\pm$ 20%

Farbe	1. bis 2. bzw. 3. Ring	3. bis 4. Ring	letzter Ring
	Ziffernfolge	Multiplikator	Toleranz
Silber	—	10^{-2}	$\pm$10%
Gold	—	10^{-1}	$\pm$ 5%
Schwarz	0	1	—
Braun	1	10	$\pm$ 1%
Rot	2	10^2	$\pm$ 2%
Orange	3	10^3	—
Gelb	4	10^4	—
Grün	5	10^5	$\pm$0,5%
Blau	6	10^6	$\pm$0,25%
Violett	7	10^7	$\pm$0,1%
Grau	8	10^8	—
Weiß	9	10^9	—

Neben den *Festwiderständen* werden auch vielfältige veränderbare Widerstände benötigt. Sie können je nach ihrem Anwendungszweck eingeteilt werden in Widerstände, die ständig verstellbar sein müssen und meist als *Potentiometer* oder Schiebewiderstände bezeichnet werden, sowie in Widerstände, die nur einmal oder sehr selten einzustellen (abzugleichen) sind. Sie werden meist als *Trimm-Potentiometer* (Trimmer) oder Einstellregler bezeichnet.

Zwischen dem Drehwinkel bzw. Weg und dem zugehörigen Widerstand lassen sich verschiedene Kurvenformen realisieren. Die entsprechende Abhängigkeit zeigt schematisch Bild 1.1.15. Zuweilen werden außer den drei Anschlüssen auch noch einzelne Anzapfungen herausgeführt.

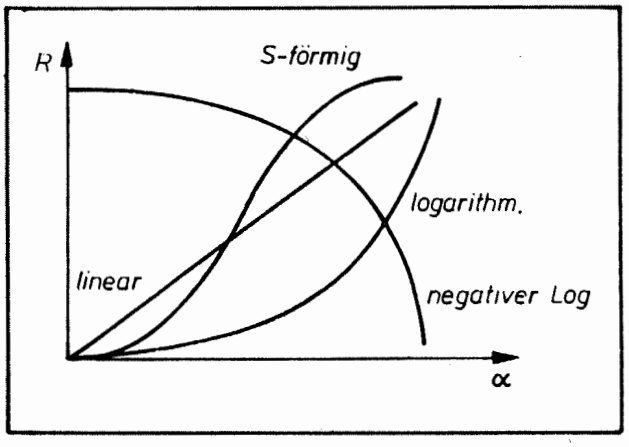


Bild 1.1.15. Zusammenhang zwischen Widerstandswert und Einstellwinkel bei verschiedenen Potentiometern

Ähnlich wie es bei den Widerständen Präzisionswiderstände enger Toleranz gibt, existieren auch bei den Potentiometern *Präzisionspotentiometer*. Bei ihnen wird eine sehr große Linearität zwischen Drehwinkel und Widerstandsänderung gefordert. Forderungen von besser als 1% sind nicht mit einem Drehwinkel unter 360° zu realisieren. Deshalb kommen Wendeln mit mehreren Gängen zur Anwendung. Derartige *Wendelpotentiometer* haben bis zu 20 Gänge und erreichen Einstellgenauigkeiten von 0,05%.

1.1.7. Technische Spannungsquellen

Zur Energieversorgung elektronischer Geräte stehen heute folgende *Energiequellen* zur Verfügung:

— das Starkstromnetz (vgl. auch 5.4.4.) und 7.6.6.
— chemische Energie aus: Primärelementen
Sekundärelementen
Füllelementen
Brennstoffzellen

— Energie des Sonnenlichtes, u. a. über Photoelemente (vgl. 6.1.2.)
— sonstige Energie, wie Wärme (Thermoelemente), nukleare Energie usw.

Hier werden nur die chemischen Elemente behandelt und von ihnen genau genommen nur die *Primär- und Sekundärelemente*.

Die *Füllelemente* sind spezielle Primärelemente. Sie enthalten im Ruhezustand keinen Elektrolyten oder zumindest nicht in der wirksamen Phase. Er liegt dann z. B. kristallin vor und geht kurz vor der Verwendung der Füllelemente durch Einfüllen von z. B. destilliertem Wasser in seine ionische und damit aktive Form über. Füllelemente werden in speziellen kommerziellen Bereichen, z. B. in der Meteorologie bei Wettersonden, oder im militärischen Bereich verwendet. Füllelemente in Seenotleuchten sind so gestaltet, daß Meerwasser sie aktiviert.

Brennstoffelemente unterscheiden sich von den anderen chemischen Elementen dadurch, daß sie den eigentlichen Energieträger nicht „eingebaut in der Zelle" enthalten, sondern daß dieser beim Betrieb ständig zugeführt wird, z. B. als H_2 und O_2, und die entstehenden Abprodukte dann auch abgeführt werden müssen. Brennstoffelemente haben noch nicht jene Reife erreicht, die ein breiter Einsatz verlangt. Im kommerziellen Bereich, z. B. im Raumflug, haben sie sich aber schon durchaus bewährt.

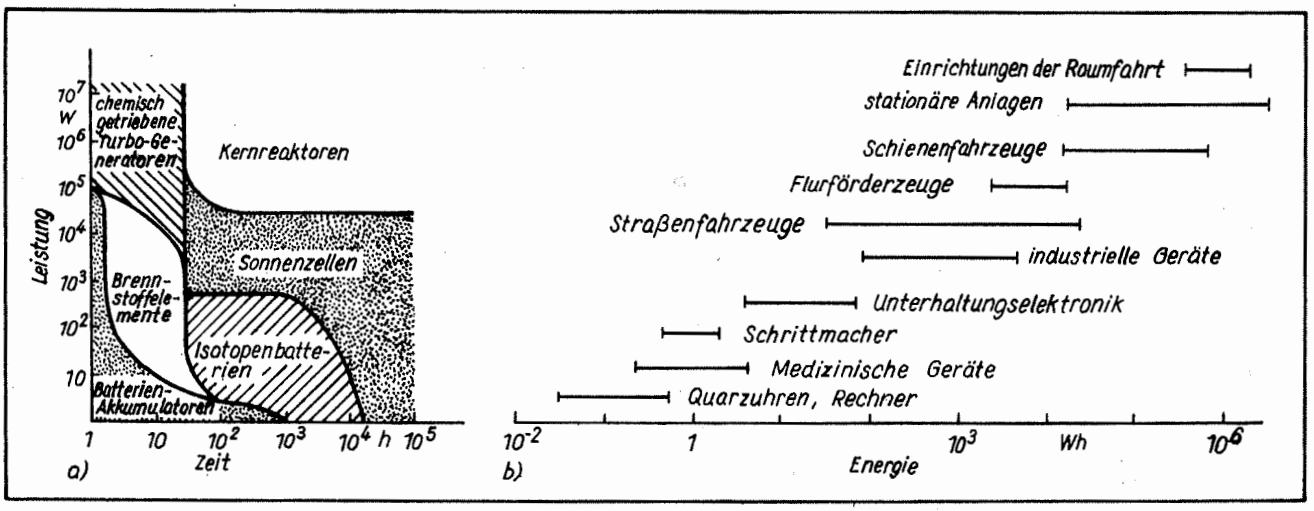


Bild 1.1.16. Überblick a) zu den Einsatzgebieten der verschiedenen Techniken zur Erzeugung elektrischer Energie und b) zu den Leistungsdaten wichtiger Gerätetechnik

Bild 1.1.16 zeigt einige Zusammenhänge zu den *Einsatzgebieten elektrischer Energieversorgungen.* Vor mehreren Jahren hatte es den Anschein, als ob die überwiegende Stromversorgung elektronischer Geräte aus dem überall vorhandenen elektrischen Netz erfolgen würde. Mit der Entwicklung der Mikroelektronik trat jedoch eine erhebliche Miniaturisierung ein, und immer mehr Geräte wurden dadurch transportabel. Sie mußten deshalb aus internen Stromquellen versorgt werden. So setzte eine beachtliche Weiterentwicklung der chemischen Stromquellen ein. Noch heute ist dieses Gebiet stark im Fluß. Dies zeigt sich u. a. in der z. T. noch unausgewogenen Vielfalt der Typen, sowohl was die verwendeten chemischen Substanzen als auch die Gehäuseformen, zumindest bei den kleinsten Zellen, betrifft. Andererseits wird z. Z. zumindest bei den Primärelementen auch die Materialökonomie untersucht. Werden verbrauchte Batterien doch einfach weggeworfen, obwohl in ihnen noch ein hoher Materialwert enthalten ist.

Der Umfang des *weltweiten Batterieverbrauchs* wird für 1980 auf rund 10^{10} Stück eingeschätzt, wovon bereits ca. 15% die kleinsten Knopfzellen betreffen. Mit der Miniaturisierung der Geräte erfolgt auch eine fortschreitende Verkleinerung der Batterien.

Weltweit wurde für 1970 ein Umsatz an Batterien aller Art von $1{,}6 \cdot 10^9$ Dollar geschätzt. Anteilmäßig gilt in Prozenten:

Bleiakkumulatoren	61
davon Starterbatterien	46 (75% relativ)
Alkalische Akkumulatoren	5
Primärelemente	34
davon Trockenelemente	24 (71% relativ).

1.1.7.1. Chemische Grundlagen

In allen chemischen Zellen müssen wir den Katoden- und Anodenraum unterscheiden. Die *Anode* ist dabei der *Minuspol* der Zelle. Beide stehen nur über den Elektrolyten, eventuell durch spezifische Trennwände auch nur für bestimmte Ionenarten in Verbindung. Es werden wäßrige und organische Elektrolyten unterschieden. Bei der *Entladung* wird in den Zellen ein Metall (Me) bzw. Wasserstoff oxydiert. Der chemische Prozeß wird durch die zumindest elektrische (nicht ionische) Trennung von Katoden- und Anodenraum so geführt, daß die dabei frei werdenden Elektronen sich nur über den Außenraum der Zelle, also über den gewünschten Leiterweg ausgleichen können. Aus diesem Grunde müssen in einem Element alle spontanen chemischen Reaktionen unterbunden werden. *Reaktionen dürfen erst dann eintreten, wenn der äußere Stromweg geschlossen wird.* Einen Überblick zu theoretisch erreichbaren Energiedichten von Anodenmaterialien gibt Tab. 1.1.9.

Tabelle 1.1.9. Theoretische Energiedichten und reversible Spannungen von chemoelektrischen Zellen bez. des Anodenmaterials

Element	Li	Mg	Na	Al	Zn	Fe^{2+}	H_2	Cd	Fe^{3+}	Cu^{2+}
Spannung in V	3,4	3,1	2,7	2,7	1,7	1,3	1,2	1,2	1,1	0,62
Energiedichte in Wh/kg	3200	2800	1800	2800	890	760	3700	440	880	290

Die allgemeine chemische Reaktion einer Zelle lautet

$$\boxed{\mathrm{Me} + \frac{n}{4}\,\mathrm{O_2} + \frac{n}{2}\,\mathrm{H_2O} \rightleftarrows \mathrm{Me(OH)}_n} \,. \tag{41}$$

Der Katodenraum stellt dabei in der Regel den Sauerstoff und die Anode das Metall. In Primärzellen verläuft diese Reaktion nur von links nach rechts, in Sekundärzellen

beim *Laden* auch von rechts nach links. Im Prinzip müßte dies in fast allen Primärelementen auch möglich sein. Hierauf weist die Literatur zum Regenerieren von Primärzellen hin, die in der Regel dann mit einem Gleichstrom erfolgt, der zu einem erheblichen Teil mit Wechselstrom überlagert ist. Die Praxis zeigt jedoch, daß sich hierbei leicht Dendriten an der Anode bilden, die zu einem Kurzschluß in der Zelle und damit zur Unbrauchbarkeit führen. Zum Teil tritt dies trotz vieler konstruktiver Gegenmaßnahmen auch bei Sekundärelementen auf. Schließlich sei noch darauf hingewiesen, daß der Sauerstoff auch aus der Luft entnommen werden kann. Dies führt zu den *Luftsauerstoffelementen*.

1.1.7.2. Primärelemente

Das älteste und heute noch am meisten verwendete Primärelement geht auf LECLANCHÉ zurück. Es enthält Zn als Anode, NH_4Cl (Salmiak) als Elektrolyten und MnO_2 (*Braunstein*) als Katode. Braunstein ist also der *Sauerstoff* abgebende Teil und wird dabei zu Mn_2O_3 reduziert. Im Zentrum der Zelle befindet sich, vom Braunstein umhüllt, ein Kohlestab als Leiter. Deshalb ist die Bezeichung Kohle-Zink-Element falsch. Sie müßte Braunstein-Zink-Element lauten.

Das LECLANCHÉ-Element kann auf verschiedene Weise verbessert werden, unter anderem durch den Elektrolyten $MgCl_2$ bzw. durch Zusätze von $ZnCl_2$ und $HgCl_2$ zum Elektrolyten. Der Elektrolyt ist im LECLANCHÉ-Element in Mehl, Stärke, Zellulose-Derivaten oder, selten, in Papier (paper-lined) gebunden. Für Sonderanwendungen wird die LECLANCHÉ-Zelle auch durch einen äußeren Stahlmantel auslaufsicher (*leak-proof*) gestaltet.

Die beiden typischen Aufbauvarianten von LECLANCHÉ-Elementen zeigt Bild 1.1.17. Die *Rundzellen* werden in verschiedenen genormten Abmessungen hergestellt, während für die *Flachzelle* nur noch eine Größe existiert.

In Tab. 1.1.10 sind neben den Abmessungen weitere wichtige Daten enthalten. Die *Kapazitätsangaben* können bei chemischen Elementen generell nur als Richtwerte angesehen werden. Sie hängen u. a. vom Lagerzustand (Selbstentladung), von der Technologie der Zelle und vor allem von der Art der Entladung ab. Insbesondere werden Dauerbetrieb und intermittierender Betrieb unterschieden. Hierzu gibt Bild 1.1.18a) einen Überblick. Aber auch die Größe des Entladungsstromes selbst geht wesentlich ein. Höhere Ströme bewirken eine kleinere Kapazität. Deshalb ist in Tab. 1.1.10 der Entladestrom und die Entladebedingung angegeben. Zur *genormten Prüfung* wird übrigens ein Widerstand festgelegt.

Die *Einsatzgebiete* der LECLANCHÉ-Zelle liegen dort, wo an die Spannungskonstanz keine hohen Ansprüche gestellt werden, wo das Last-Pausen-Verhältnis klein ist und wo Kosten gespart werden sollen. Einige Hersteller bieten sie in *drei* verschiedenen *Leistungsklassen* an: Die billigste ist für elektronische Geräte weniger zu empfehlen, sondern höchstens für Taschenlampen, da ihre Spannung sehr schnell absinkt (Bild 1.1.18b), Kurve 1); die mittlere Sorte (Kurve 2) eignet sich schon für einfache Geräte der Unterhaltungselektronik, z. B. Kofferradios, die teure ist für Geräte mit höherer Stromaufnahme wie etwa Elektronenblitzgeräte, Kassettenrecorder usw. vorgesehen (Kurve 3). Zellen der letzteren beiden Klassen können auch zur Speisung von aufwendigeren elektronischen Schaltungen verwendet werden, sofern diese keine hohen Ströme benötigen und nicht dauernd in Betrieb sind, also etwa für Digitalmultimeter in CMOS-Technik mit LCD-Anzeigen, nicht jedoch für Uhren.

Bei der *alkalischen Zelle* ist der saure Elektrolyt des LECLANCHÉ-Elements durch KOH ersetzt. *Kalilauge* hat die Tendenz, an Oberflächen und durch kleine Spalte zu kriechen und ist darüber hinaus sehr aggressiv. Deshalb mußte hierfür die völlige Abdichtung konstruktiv gelöst werden. Obwohl diese Zellen generell nur als leak-proof

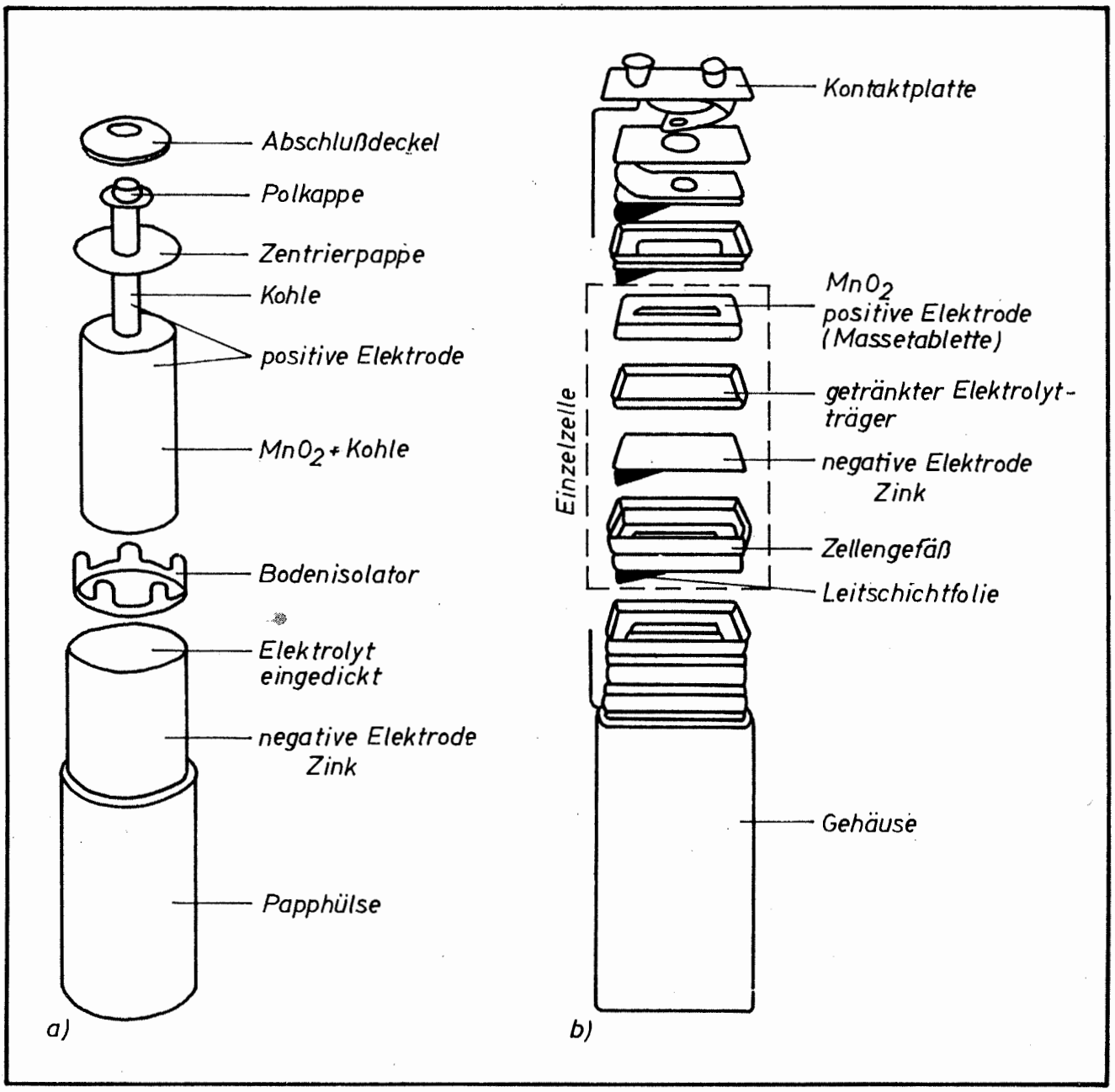


Bild 1.1.17. Aufbau a) der Rundzelle und b) der Flachzelle von LECLANCHÉ-Elementen

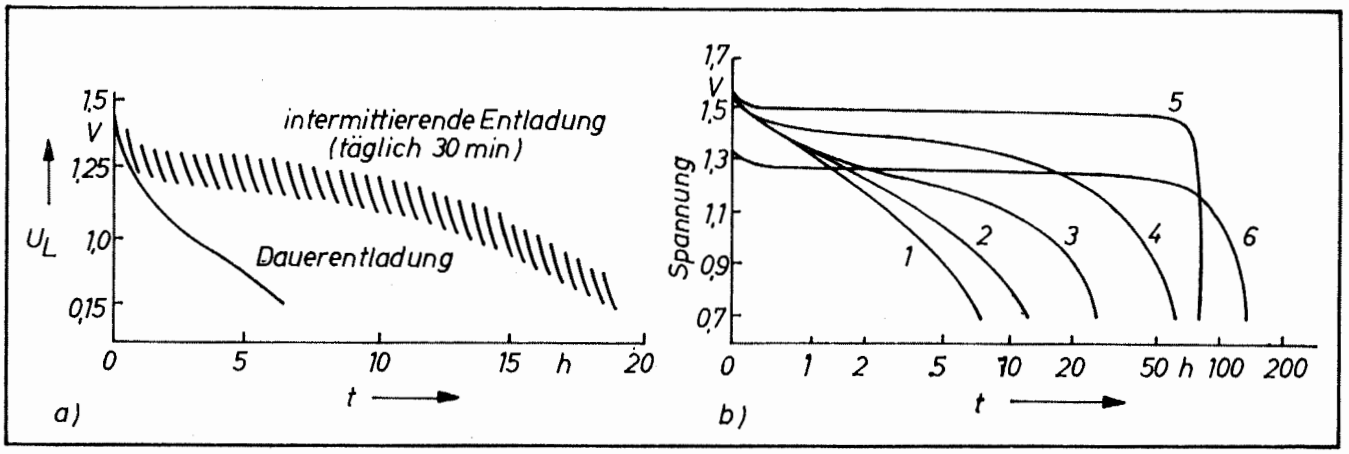


Bild 1.1.18. a) Prinzipielle Entladekurve einer R20-Zelle mit Dauer- und intermittierendem Strom, b) Entladekurve für unterschiedliche Primärzellen bei vergleichbaren Abmessungen

1 bis 3 verschiedene Ausführungen von LECLANCHÉ-Elementen
4 Alkalische Braunstein-Zink-Zelle 5 Silberoxid-Zink-Zelle 6 Quecksilberoxid-Zink-Zelle

Tabelle 1.1.10. Überblick zu den wichtigsten genormten Primärelementen

	Handelsbezeichnung							
	Lady	Micro	Mignon	Baby	Mono	Stab-batterie	Flach-batterie	Energieblock
IEC-Norm	R1	RO3	R6	R14	R20	2R10	3R12	6F22
ASA-Norm	N	AAA	AA	C	D	—	—	—
IIS-Norm	UM5	UM4	UM3	UM2	UM1	—	—	—
Nennspannung in Volt	1,5	1,5	1,5	1,5	1,5	3	4,5	9
Abmessungen max. ∅ × Höhe in mm × mm bzw. Länge/Breite/Höhe	12 × 30	10,5 × 40,5	14,5 × 50,5	26 × 50	34 × 61,5	21 × 74	62/22/67	26,5/17,5/48,5
Kurzschlußstrom in A	1	1	6	10	7—18	3	6	0,6
Strom in mA mit 2—4h/d ergibt Kapazität in Ah	4	4	30	30	60	30	50	9
bei LECLANCHÉ	0,4	0,37	1	2,6	5,8	0,6	1,6	0,38
bei Alkali	0,6	0,76	1,6	4,8	9,8	—	—	0,48
Kapazität bei gleichgroßem NiCd-Sammler in Ah	0,15	0,18	0,5	1,8	4,0	—	—	0,11
Ca. Verkaufsanteile 1980 in %, 5% sonstige (Klammerwerte galten 1960)	1	3	35(27)	26(21)	20(52)	2	4	4

im Stahlmantel hergestellt werden, empfiehlt sich dennoch bei Lagerung im Gerät eine regelmäßige Überprüfung ihrer Dichtheit.
Der Innenwiderstand dieser Zellen ist viel geringer und die Energiedichte 2- bis 3mal größer als beim LECLANCHÉ-Element. Sie besitzen, wie die Kurve 4 in Bild 1.1.18 zeigt, eine erheblich flachere Entladecharakteristik. Weiter ist ihr Temperaturbereich von 0 °C bei LECLANCHÉ auf −40 °C erweitert. Im inneren konstruktiven Aufbau ist bei der alkalischen Zelle innen das Zink und außen das MnO_2 angeordnet (vgl. Bild 1.1.19).

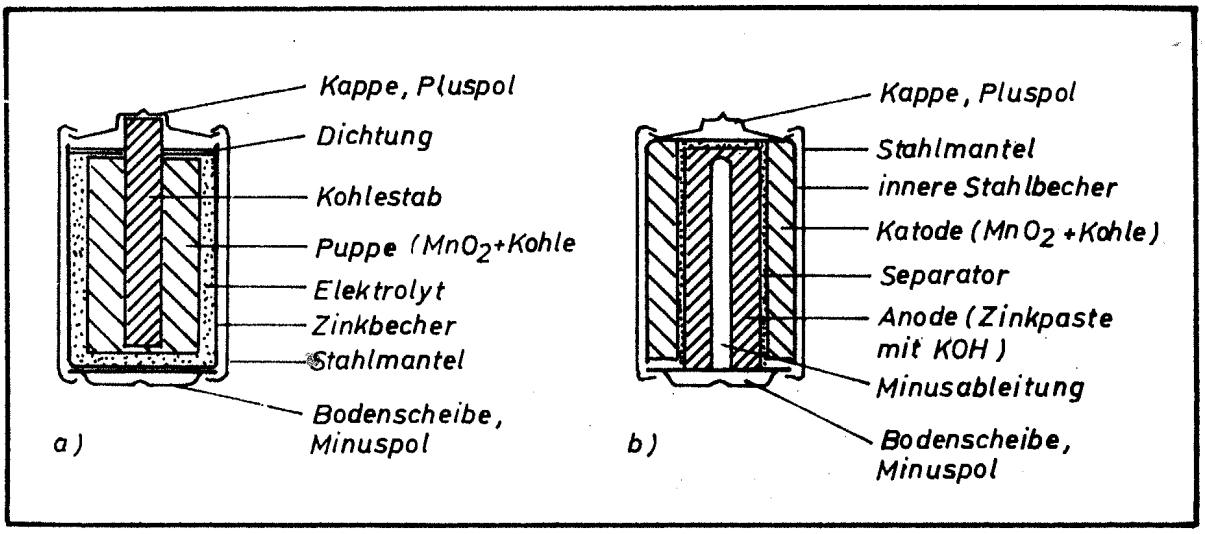


Bild 1.1.19. Vergleich von LECLANCHÉ- (a) und alkalischem Zinkelement (b)

Als weitere Abwandlung des LECLANCHÉ-Elements sei das *Luftsauerstoffelement* genannt. Es besitzt kein MnO_2, sondern nur Aktivkohle, die sich mit Luftsauerstoff ständig neu vollsaugt. Dadurch ist dieses Element nur für kleine Ströme geeignet, besitzt aber eine hohe Gesamteffektivität. Es stellt bereits den halben Übergang zu den Brennstoffelementen dar.

Eine noch höhere Qualität der Primärelemente mit der Zinkanode wurde insbesondere in der letzten Zeit mit den Katoden *Quecksilber-* bzw. *Silberoxid* erreicht. Bei der *Quecksilber-Zink-Zelle* kommen als Vorteile hinzu: höhere Kapazität, sehr kleiner Innenwiderstand (die Katode wird zum gut leitenden Quecksilber reduziert) und ein Temperaturbereich von −20 bis +100 °C. Der Elektrolyt ist KOH. Die *Silberzelle* existiert in zwei Arten: mit Silber-I- und Silber-II-Oxid. Der Elektrolyt ist KOH bzw. NaOH.

Die zuletzt genannten Hochleistungszellen haben größere Bedeutung durch die Mikroelektronik als *Knopfzellen* erreicht. Nur auf ihrer Basis sind in den entsprechend kleinen Volumen ausreichende Energiemengen zu speichern. Bei den Knopfzellen steht zur Zeit noch die Normung weitgehend aus. Unter Beachtung dieser Tatsache gibt Tab. 1.1.11 einen brauchbaren Überblick. Den Aufbau einiger Knopfzellen zeigt Bild 1.1.20.

Schließlich seien noch die *Lithiumzellen* erwähnt. Lithium hat als Anodenmaterial gleich zwei Vorteile. Es erzeugt das höchste Potential von ca. 3 V und besitzt eine extrem niedrige Dichte (0,54 g/cm³). Da Lithium spontan mit Wasser reagiert, müssen organische Elektrolyten verwendet werden. Wie bei den zuvor behandelten Zinkelementen gibt es auch bei Lithium eine Vielzahl von Varianten. In den meisten Fällen besitzen die Lithiumzellen einen sehr hohen Innendruck von etwa 10^4 Pa.

Einen Vergleich wichtiger Eigenschaften der Primärelemente gibt Tab. 1.1.12.

Tabelle 1.1.11. Abmessungen und genormte Bezeichnungen wichtiger Primärknopfzellen
Der R-Nr. nach IEC sind oft noch Buchstaben vorangestellt. Deren Bedeutung geht aus der Aufstellung rechts hervor. Vieles spricht dafür, daß sich folgende Typen bevorzugt durchsetzen werden: R41 für Damenuhren, R54 für Herrenuhren und R44 und R54 für Rechner. Neben den in der Tabelle vorgestellten Abmessungen existieren noch Durchmesser von ca. 6,8 sowie Höhen von 1,6 mm. Größere Knopfzellen existieren mit Durchmessern von 15 bis 50 mm und Höhen von 6 bis 10 mm, mit Kapazitäten von 50 bis 1000 mAh. Beispielhafte Kapazitätswerte sind etwa für SR44 ≈ 180 und SR54 ≈ 80 mAh.

Höhe in mm	Durchmesser in mm		
	7,9	9,5	11,6
5,4	R48	R46	R44
4,2			R43
3,6	R41	R45	R42
3,1			R54
2,6	R59	R57	R56
2,1	R58		R55

U = Universal
PX = Film/Foto
H = Hörgeräte
G = Computer
P = Braunstein/Zink
M = Quecksilberoxid/Zink
S = Silberoxid/Zink
A = Braunstein/Zink
L = Luft/Zink

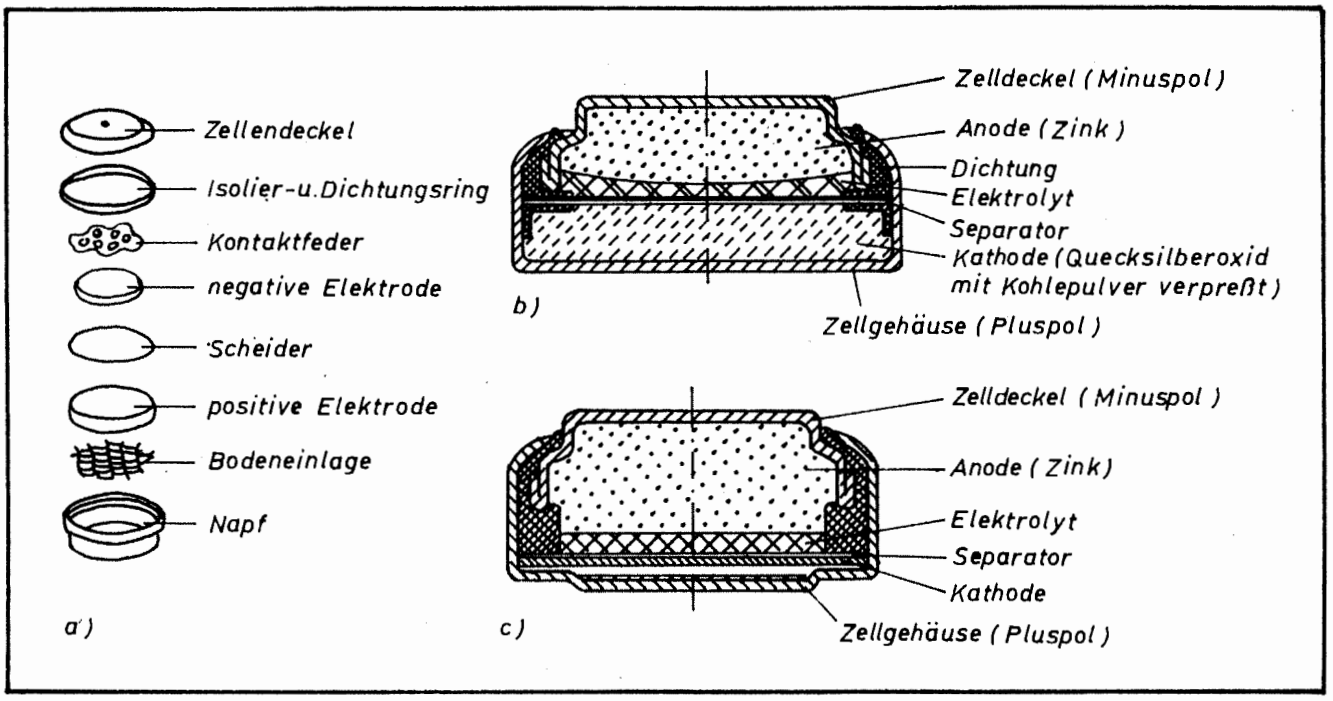


Bild 1.1.20. Prinzipaufbau einer Knopfzelle (a) sowie die Beispiele Quecksilberoxid- (b) und Luftsauerstoffelement (c). Beim Luftsauerstoffelement wird kurz nach Inbetriebnahme eine Folie am Pluspol entfernt, wodurch dann der Luftsauerstoff Zutritt erhält.

1.1.7.3. Sekundärelemente

In der Hauptsache existieren heute *drei Typen* von Sekundärelementen. Mit der zusätzlichen Unterscheidung von offenen und gasdichten NiCd-Akkumulatoren sind die wichtigsten Daten in Tab. 1.1.13 zusammengefaßt. Daher genügen hier wenige ergänzende Bemerkungen. Alle Sekundärelemente werden im Normalfall mit etwa 0,1 A/Ah etwa 12 Stunden geladen. Der *Ladeschluß* ist günstig aus der *Ladespannung* abzuleiten. Sie zeigt vor dem voll geladenen Zustand einen deutlichen Spannungsanstieg gemäß Bild 1.1.21a). Aus diesem Grunde wird auch die Ladeschlußspannung angegeben. Für den NiCd-Sammler ist in vielen Fällen zumindest teilweise eine

Tabelle 1.1.12. Eigenschaften von Primärzellen
Es bedeuten K Knopfzellen, R Rundzellen, F Flachzellen.

System	Bauformen	Kapazitäts-bereich in Ah	Energiedichten		Spannung in V	Selbstentla-dung im Jahr	relativer Preis
			mWh/g	mWh/cm³			
LECLANCHÉ	K R F	0,06—300	50—80	120—190	1,2	20%	1
Alkalizelle	K R F	0,06—10	70—100	200—300	1,3	10%	2
Luft-Zink	K	0,18—0,3	180—350	650—800	1,1		
	R F	50—180					
Quecksilberoxid	K	0,04—0,35	90—120	400—520	1,2	5%	5
Silber-I-Oxid	K	0,015—0,25	70—100	350—430	1,3	5%	10
Silber-II-Oxid	K	0,04—0,25	130—170	550—650	1,3	5%	10
Lithium	K R F	0,05—30	150—300	500—800	1,5—3	1%	10

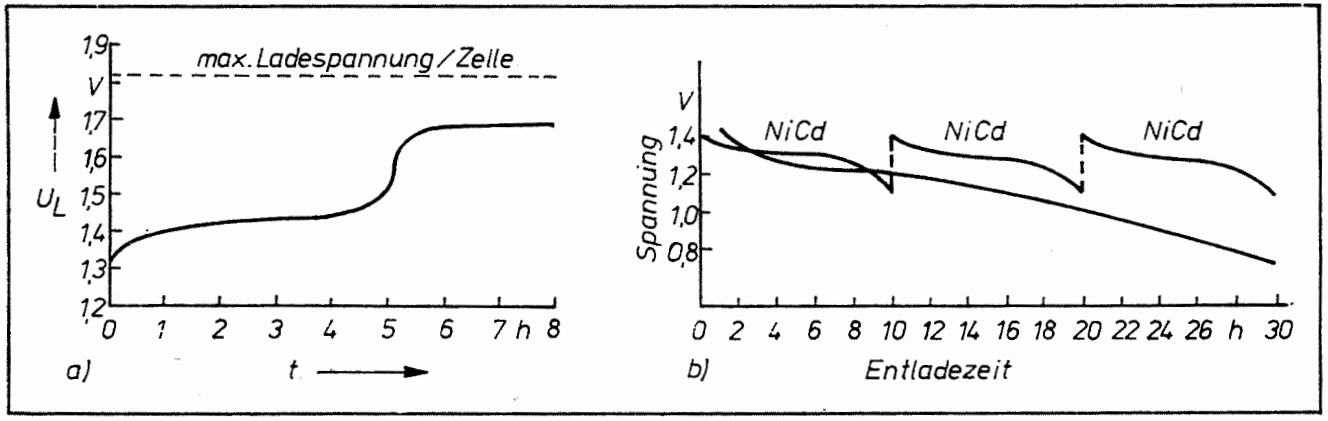


Bild 1.1.21. a) Ladekurve und b) Entladekurve eines NiCd-Sammlers im Vergleich zu einem etwa gleich großen alkalischen Element

Tabelle 1.1.13. Wichtige Eigenschaften von Sekundärelementen
Es bedeuten R Rund-, F Flach- und K Knopfzellen.

System	Pb	NiCd	NiCd gasdicht	AgZn
Geladener Zustand	PbO_2 + H_2SO_4 + Pb	$2NiO(OH) + \frac{Fe}{Cd} + 2H_2O$		$Ag + Zn(OH)_2$
Entladener Zustand	$PbSO_4 + 2H_2O$ + $PbSO_4$	$2Ni(OH)_2 + \frac{Cd}{Fe}$ Fe (OH)		$AgO + Zn + H_2O$
Elektrolyt	H_2SO_4 Dichte 1,18 ... 1,28	KOH mit Zusatz Dichte 1,17		40% KOH
Bauformen	R,F	K,R,F		K,F
Kapazitätsbereich Ah	1—25	0,01—50		0,05—0,5
Energiedichten				
mWh/g	20—35	20—30		30—115
mWh/cm³	50—100	40—80		100—240
Spannungen				
Nenn-	2	1,2	1,2	1,5
Entladung Ende	1,80	1,0	1,10⁺)	1,0⁺)
Ladung Ende	2,78	1,85	1,55	2,1
Zyklen	1000 (200—400)*)	2000	200	geringe Lebensdauer
Innenwiderstand, Ω/Ah	0,06	0,2	0,3	0,002
Selbstentladung 1 Monat bei 20 °C	40%	18%	28%	1,6%
Spitzenentladestrom kurzzeitig, A/Ah	3,5	0,67	0,5	20, Sonderfälle 1000
Wirkungsgrad				
Ah	0,9	0,72		0,95
Wh	0,7	0,5		0,90
Temperaturbereich, °C	5—55	—20 ... +40		

⁺) Unterschreiten dieser Werte schadet den Zellen erheblich.
*) Gilt für geschlossene Systeme.

Schnelladung mit ca. 0,5 A/Ah zulässig. Insbesondere der gasdichte NiCd-Sammler ermöglicht auch eine ständige Ladung zum Aufrechterhalten der vollen Kapazität mit etwa 0,002 A/Ah. Dabei ist aber zu beachten, daß die *Selbstentladung* aller Sammler stark von der Temperatur abhängt, sie ist z.B. bei 60 °C etwa 8mal so groß wie in der Tabelle und fällt beträchtlich zu tiefen Temperaturen.

Für die Mikroelektronik haben im wesentlichen nur die geschlossenen Systeme Bedeutung. In der DDR stehen auf Pb-*Basis* die *Kleinstzellen* mit 0,3 und 0,5 Ah zur Verfügung. Da es bei den Pb-Zellen keinen Mechanismus der Gasaufzehr gibt, wird hier die Ladung bei geringem Aufblähen des Kunststoffgehäuses festgestellt.

Bei den gasdichten NiCd-Sammlern gibt es dagegen einen Mechanismus der Gasabsorption. Er beruht im wesentlichen darauf, daß der Sauerstoff an überproportional vorhandenes Cadmium gebunden wird. Die Bildung von Wasserstoff wird so vermieden. Für mikroelektronische Geräte muß zuweilen entschieden werden, ob Primär- oder Sekundärelemente eingesetzt werden sollen. Als *Pufferung* für Netzausfall kommen nur Sekundärelemente mit Ladestrom zur Aufrechterhaltung der Kapazität in Betracht. Für bewegliche Geräte kann etwa eingeschätzt werden, daß Sekundärelemente rentabler sind, wenn alle zwei bis drei Wochen ein Batteriesatz benötigt wird. Bild 1.1.21b) zeigt das Verhältnis der Entladungen bei etwa gleichen Bedingungen.

Schließlich zeigt Tab. 1.1.14 einen geschichtlichen Überblick zu den chemischen Energiequellen.

Tabelle 1.1.14. Geschichtliche Daten zu den chemischen Stromquellen

1789	GALVANI: Froschschenkelversuche
1800	VOLTA-Säule
1854	GINSTEDEN: Prinzip des Bleisammlers
1859	PLANTÈ: Berauchbarer Bleisammler
1868	LECLANCHÉ-Element
1881	Großoberflächenplatte für Bleisammler
1899	JUNGNER: Grundlagen für NiCd-Sammler
1900	P. SCHMIDT: Trockenelement (Eindicken des Elektrolyts)
1901	EDISON: Nickel-Eisen-Sammler
1933	Patent für gasdichte Zellen
1939	Leak-Proof-Zelle
1940	Quecksilberoxidzelle
1945	Anwendung des NiCd-Sammlers
1950	Gasdichte NiCd-Zelle
1950	Ag_2O-Zelle
1960	Li-Zelle
1970	AgO-Zelle

1.2. Magnetostatik

1.2.1. Ursachen des Magnetismus

Jeder elektrische Strom I bewirkt ein magnetisches Feld. Hierfür gilt das BIOT-SAVART-*Gesetz*

$$\boldsymbol{H} = \frac{1}{4\pi} \int \frac{I \sin \varphi}{r^2} \mathrm{d}\boldsymbol{s} \,. \tag{1}$$

Darin sind gemäß Bild 1.2.1a) r die Entfernung zum Punkt P, in dem die Feldstärke $\boldsymbol{H}$ gemessen wird, $\boldsymbol{s}$ der Weg auf dem Stromleiter und φ der Winkel zum Aufpunkt P. Für einen *Kreisstrom* mit dem Radius R (Bild 1.2.1b)) gilt folglich auf der x-Achse

$$H_x = \frac{I}{2} \frac{R^2}{(R^2 + x^2)^{3/2}} \,. \tag{2}$$

Genau im Innern des Ringstromes $x = 0$ beträgt die Feldstärke daher

$$H_{x,0} = \frac{I}{2R} \,. \tag{3}$$

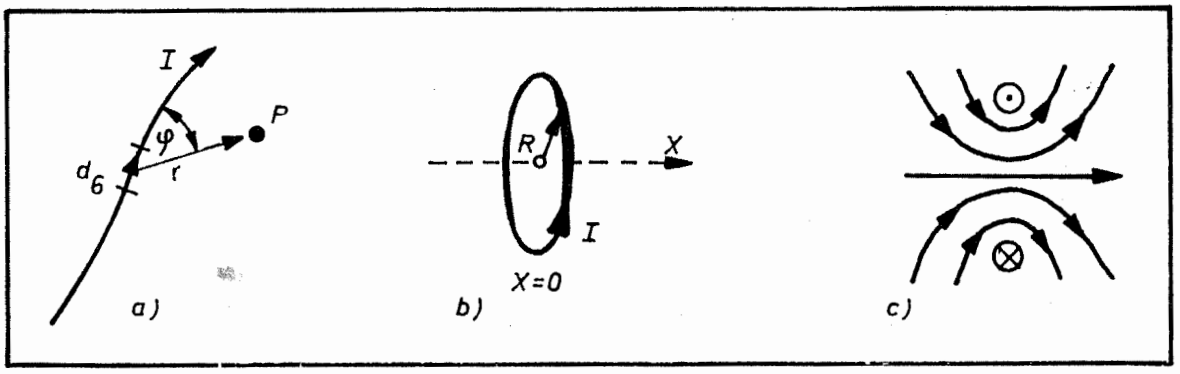


Bild 1.2.1. Zum Zusammenhang von Strom und Magnetfeld
a) Zur Ableitung des BIOT-SAVARTschen Gesetzes
b) Kreisstrom auf den Radius R für Feldstärke gemäß Gl. (2)
c) Dipolfeld des Kreisstromes

Die Gesamtheit des Magnetfeldes eines Kreisstromes stellt einen *magnetischen Dipol* dar, dessen Feldlinien in Bild 1.2.1c) dargestellt sind. Liegt eine lange Spule mit n Windungen bei einer Länge l vor, so beträgt die Feldstärke im Inneren der Spule

$$H_x = I \frac{n}{l}. \tag{4}$$

Diese Beziehung macht verständlich, warum die Feldstärke in A/cm gemessen wird. Eine alte Maßeinheit ist das Oersted (Oe). Es gilt

$$1\ \text{Oe} = \frac{10}{4\pi}\ \text{A/cm}\ . \tag{5}$$

Der Zusammenhang von magnetischer Feldstärke und Strom kann auch durch ein geschlossenes Integral erfaßt werden:

$$\boxed{\oint \boldsymbol{H}\,\mathrm{d}\boldsymbol{s} = \Theta = \sum I} \tag{6}$$

Hierin bedeutet Θ die *Durchflutung* des geschlossenen Weges mit der Stromsumme. Deshalb heißt dieses Gesetz auch Durchflutungsgesetz.

Der Magnetismus als Stoffeigenschaft gehorcht quantenmechanischen Gesetzen. Deshalb ist er schwer klassisch darzustellen. Dennoch sei dieser Weg hier gewählt. Jedes Atom wird von Elektronen umkreist. Diese Bahnbewegung entspricht einem Strom und muß daher ein magnetisches Feld hervorrufen. Außerdem rotiert jedes Elektron um seine eigene Achse. Dies ist der Spin. Beide bewirken einen magnetischen Dipol, der als BOHRsches *Magneton* bzw. als Elementareinheit des Magnetismus bezeichnet wird:

$$\mu_B = \frac{h}{4\pi c} \cdot \frac{e_0}{m} \approx 9{,}274 \cdot 10^{-24}\ \text{As/T}\ . \tag{7}$$

Darin bedeuten c die Lichtgeschwindigkeit ($2{,}9979 \cdot 10^8$ m/s), h das PLANCKsche Wirkungsquantum ($6{,}625 \cdot 10^{-34}$ Js), e_0 die Ladung des Elektrons ($1{,}602 \cdot 10^{-19}$ As) und m die Masse des Elektrons ($9{,}109 \cdot 10^{-28}$ g). (Die Maßeinheit T = Tesla wird später erklärt.)

In den meisten Atomen kompensieren sich die Momente der einzelnen Elektronen nach außen. Wenn jedoch eine innere Schale nicht ausgefüllt ist, besteht die Möglichkeit, daß Spinmomente nach außen wirksam werden. Diese Wirkung verstärkt sich auf makroskopische Größen, wenn die Atome in einem Festkörper so angeordnet sind, daß sich die nach außen wirksamen Spinmomente in die gleiche Richtung stellen (Bild 1.2.2b)). Hierzu muß ein genau definierter Abstand auftreten. Er wird mittelbar durch das quantentheoretische *Austauschintegral* beschrieben. Es ist in Bild 1.2.3a) für wichtige Materialien angegeben. Nur die Materialien mit positivem Austauschintegral sind ferromagnetisch. Bild 1.2.3b) ergänzt hierzu das prinzipiell mögliche magnetische

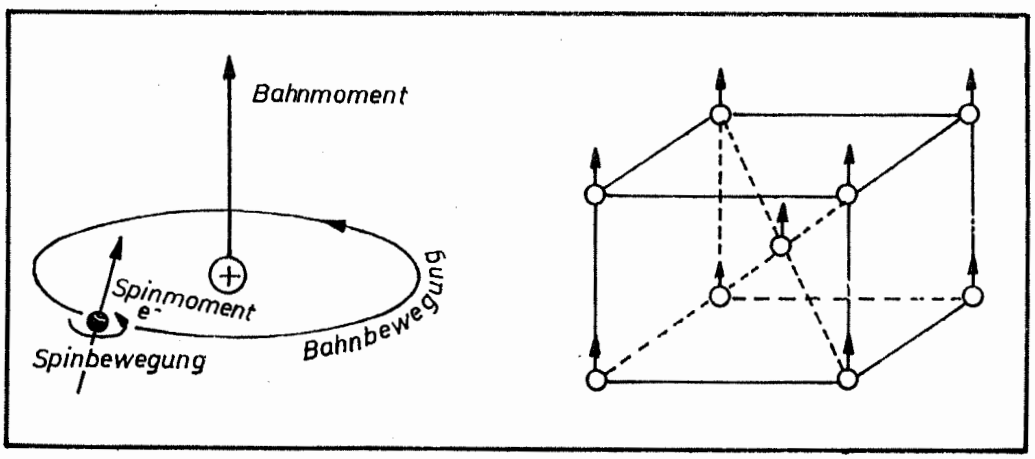


Bild 1.2.2. a) Bewegung eines Elektrons um den Kern und Eigenrotation des Elektrons bewirken je ein magnetisches Moment
b) Beispiel der gleichen Ausrichtung aller Spins einer Kristallzelle

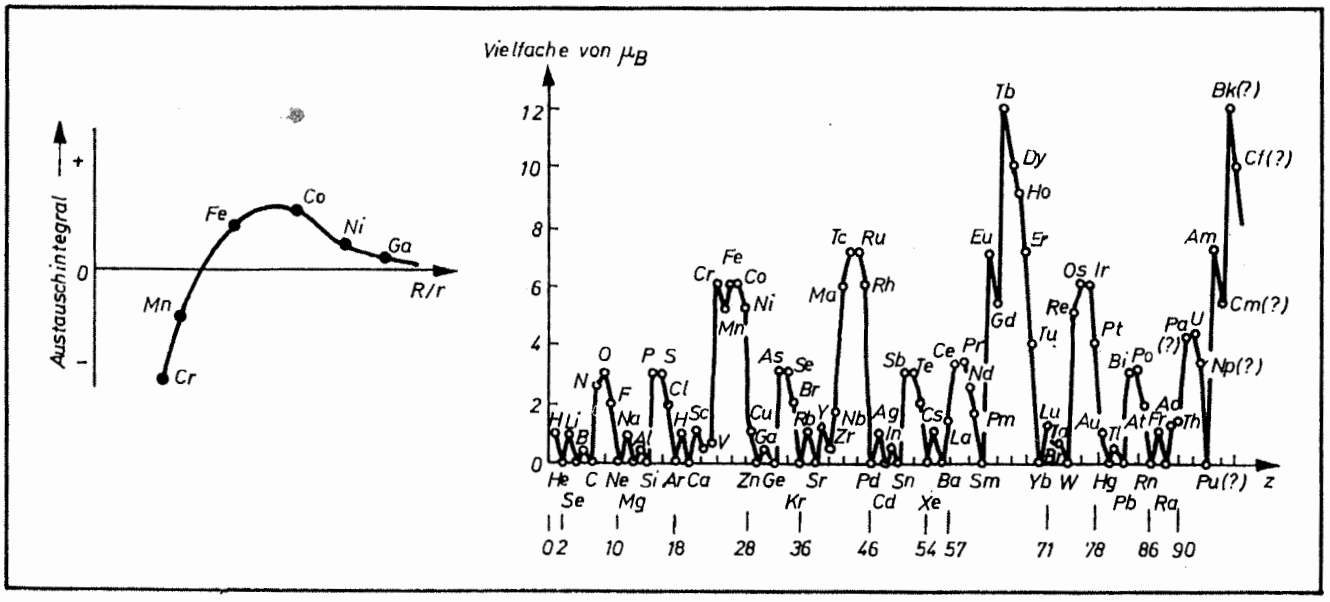


Bild 1.2.3. a) Verlauf des Austauschintegrals als Funktion von Atomabstand R zum Radius der unabgeschlossenen $3d$-Schale r
b) Anzahl der unabgesättigten magnetischen Momente als Funktion der Kernladungszahl

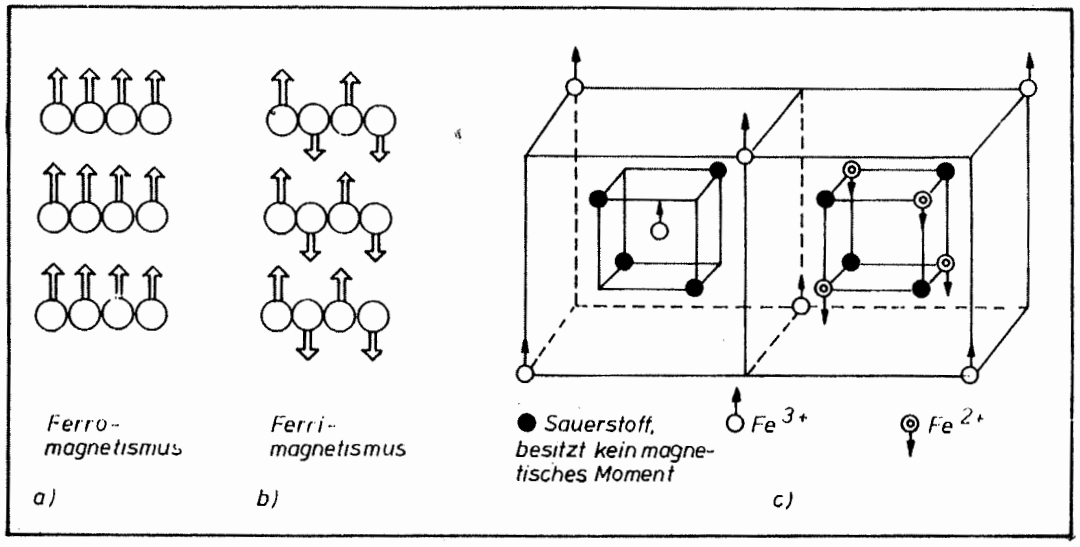


Bild 1.2.4. Zuordnung der magnetischen Momente bei a) Ferro- und b) Ferrimagnetismus sowie c) Lage der Momente im Gitter des Magnetits Fe_2O_3

Moment über das ganze Periodensystem. Es läßt erkennen, daß auch die seltenen Erden magnetisch sind.

Neben dem bisher behandelten *Ferromagnetismus* gibt es noch mehrere Arten von Magnetismus (Dia-, Para-, Antiferro-, Meta- und Superparamagnetismus). Hierfür muß auf die Spezialliteratur verwiesen werden. Praktisch wichtig ist der *Ferrimagnetismus*. Er entsteht, wenn zwei Materialien im Festkörper enthalten sind und ihre magnetischen Momente unterschiedlich, aber entgegengesetzt gerichtet sind. Im pauschalen Verhalten ist er vor allem über den Temperaturgang vom Ferromagnetismus zu unterscheiden. (Vgl. Bild 1.2.4.)

Das magnetische Material müßte nach den bisherigen Betrachtungen als einheitlicher Magnet auftreten. In der Regel ist das Material nach außen aber fast ohne magnetische Wirkungen. Dies hat zwei Ursachen. Einmal bestehen die meisten Magnetmaterialien aus mehreren *kleinen kristallinen Bezirken*. Sie sind eben kein Einkristall (s. 4.1.). Zum anderen würde selbst ein Einkristall gemäß Bild 1.2.5a) ein starkes Feld im Außenraum erzeugen. Es ist energetisch günstiger, daß sich im Material magnetische Bloch-*Wände* bilden (Bild 1.2.5b) und c)). Dann ist das Außenfeld

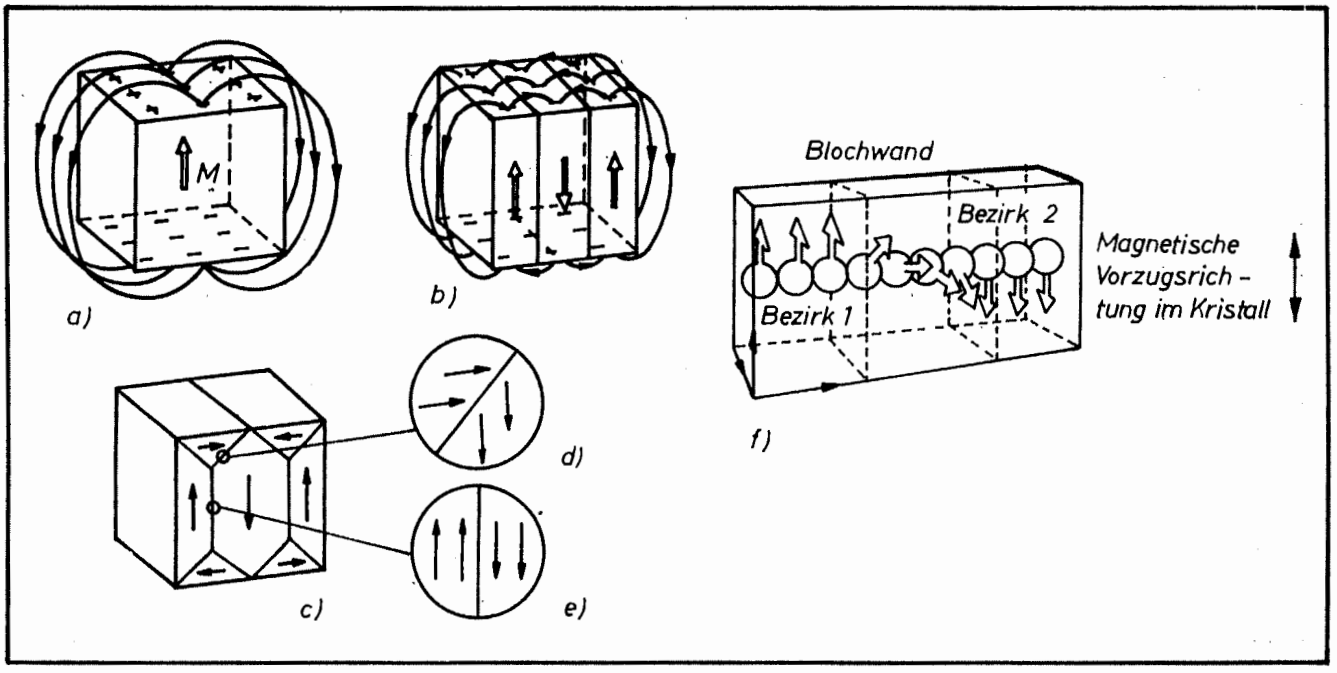


Bild 1.2.5. Magnetisierungsstrukturen im magnetischen Material

a) Einheitlicher magnetisierter Bereich hat viel Feldenergie im Außenraum
b) Aufteilung des Blockes durch zwei Block-Wände
c) bis e) Weitere Unterteilung des Kristallits. Hier treten 90°- und 180°-Bloch-Wände auf
f) Drehung des Magnetisierungsvektors bei einer 180°-Bloch-Wand. Die Bloch-Wand ist ca. 100 bis 1000 Atomabstände dick. Sie besitzt eine Energie von ca. 10^{-4} bis 10^{-2} J/m²

bereits wesentlich geringer. Noch günstiger liegen die energetischen Verhältnisse, wenn sich 90°- und 180°-Bloch-Wände gemäß Bild 1.2.5d) bis f) bilden. Die Bezirke mit einheitlicher Magnetisierung heißen Weiss*sche Bezirke* oder *Domänen*. Die in ihnen gegebene Magnetisierung heißt *spontane Magnetisierung*. Die Größe der Weissschen Bezirke kann sehr verschieden sein. Übliche Werte liegen bei 0,001 bis 0,1 mm³ bzw. 10^8 bis 10^{11} Atomen. Die Richtung der spontanen Magnetisierung ist an eine bestimmte Richtung zum Kristallgitter der Körper gebunden. Die Wechselwirkung zwischen Gitter und spontaner Magnetisierung ist die Kristallanisotropie. Sie besitzt etwa folgende Werte:

Material	Selt. Erden/ Cobalt	MnBi	Cobalt	Eisen	Nickel	80 Ni 20 Co
Wert in J/m³	$1{,}2 \cdot 10^7$	$1{,}5 \cdot 10^6$	$4 \cdot 10^5$	$4 \cdot 10^4$	$4 \cdot 10^3$	400

Bei *dünnen Schichten* treten etwas abweichende Bedingungen auf. Der Magnetisierungsvektor dreht sich bei ihnen z. T. nur in der Schichtebene. Dann liegen NEEL-*Wände* vor. Außerdem können sich spezielle Magnetisierungsstrukturen, z. B. Tannenbäume und NEEL-Spieße, ausbilden.

1.2.2. Feldgrößen

Das magnetische Feld kann in *Analogie zum elektrischen Strom* betrachtet werden, nur daß bei ihm keine Ladungen fließen. Es ist primär vielmehr eine statische Erscheinung. Der Spannung U bzw. dem Potential φ entspricht das magnetische Potential Θ, dem Strom I der magnetische Fluß Φ und der Stromdichte i die *magnetische Flußdichte* oder *Induktion*

$$\boldsymbol{B} = \frac{\mathrm{d}\Phi}{\mathrm{d}A}. \tag{8}$$

Die Induktion wird in Tesla (T) gemessen. Die alte Maßeinheit ist das Gauß (G), wobei gilt 10^4 G = 1 T = 1 Vs/m². Der magnetische Fluß wird in Weber (Wb) gemessen. Eine alte Maßeinheit ist das Maxwell. Es gilt 10^8 M = 1 Wb = 1 Vs.

Damit ist das magnetische Feld wie ein Stromfluß i interpretierbar, obwohl es einen Zustand darstellt, bei dem eigentlich nichts fließt. Ursache des elektrischen Stromes war die elektrische Feldstärke $\boldsymbol{E}$. Ihr entspricht jetzt die magnetische Feldstärke $\boldsymbol{H}$. Sie kann mittels der Kraftwirkung auf einen magnetisierten Kristalliten gemessen werden.

Entsprechend

$$U = \int \boldsymbol{E}\,\mathrm{d}\boldsymbol{s} \tag{9}$$

folgt daher jetzt

$$\Theta = \int \boldsymbol{H}\,\mathrm{d}\boldsymbol{s}\,, \tag{10}$$

und das mikro-Ohmsche Gesetz

$$\boldsymbol{i} = \frac{1}{\varrho}\boldsymbol{E} \tag{11}$$

führt zu der Beziehung

$$\boxed{\boldsymbol{B} = \mu \boldsymbol{H}}\,. \tag{12}$$

Dabei wird μ in zwei Größen zerlegt; die *Permeabilität* des Vakuums μ_0 und die relative Permeabilität μ_r des Materials,

$$\mu = \mu_0 \mu_r\,. \tag{13}$$

μ_0 ist von dem Maßsystem abhängig. Im „Système International" gilt

$$\mu_0 = 4\pi\ 10^{-7}\ \mathrm{Vs\ A^{-1}\ m^{-1}}\,. \tag{14}$$

Im Gegensatz zum elektrischen Feld hängt μ_r von vielen Einflüssen einschließlich der Vorgeschichte ab. Neben der magnetischen Flußdichte hat in vielen Fällen noch

die *Magnetisierung*

$$\boldsymbol{M} = \boldsymbol{B} - \mu_0 \boldsymbol{H} \tag{15}$$

Bedeutung.

1.2.3. Hysteresekurve

Auf ein nach außen unmagnetisch wirkendes Material gemäß Bild 1.2.5c) wirke ein äußeres Magnetfeld ein. Dann bestehen nacheinander *drei Möglichkeiten* der Auswirkungen im Material.

1. *Wandverschiebungen*: Durch Verlagerung der magnetischen Wände wachsen jene Bereiche, deren Magnetisierungsvektor zum äußeren Feld gelagert ist.
2. *Irreversible Sprünge*: Bei ihnen wechselt plötzlich und nicht reversibel der Magnetisierungsvektor eines Bereiches sprungartig in eine andere Richtung. Diese Magnetisierungssprünge hat BARKHAUSEN mittels des BARKHAUSEN-Effekts nachgewiesen. Um eine magnetisierbare Probe liegt dabei eine Spule, die über einen Verstärker zum Lautsprecher führt. Wird die Probe magnetisiert, so ist im Lautsprecher deutlich das Prasseln der BARKHAUSEN-Sprünge zu hören.
3. *Reversible Drehprozesse*: Infolge eines großen äußeren Feldes drehen sich die Magnetisierungsvektoren ohne Verlagerung der Wände in den einzelnen Bereichen so, daß sie möglichst parallel zum äußeren Feld zu stehen kommen.

Diese drei Effekte treten der Reihe nach mit steigender Feldstärke auf und erzeugen so die *Hysteresekurve*. Bild 1.2.6a) macht dies anschaulich. Dabei ist nur der Verlauf von pauschal unmagnetischem Material bis zum voll gesättigten Material gezeigt. Infolge der irreversiblen Magnetisierungssprünge verläuft die Kurve bei Rücknahme des Feldes völlig anders. In Bild 1.2.6a) ist dies gestrichelt angedeutet, während Bild b) die vollständige *Grenzhysterese* und die *Neukurve* zeigt. Darin sind darüber hinaus die wichtigsten Größen der *Koerzitivfeldstärke* und *Remanenz*, genauer remanente magnetische Induktion, eingetragen.

Die Neukurve eines Materials wird auch erhalten, wenn unterschiedlich stark ausgesteuerte Magnetisierungskurven auf dem Oszillographen sichtbar gemacht werden. Die Spitzenwerte bestimmen dann genau den Verlauf der Neukurve, dies ist

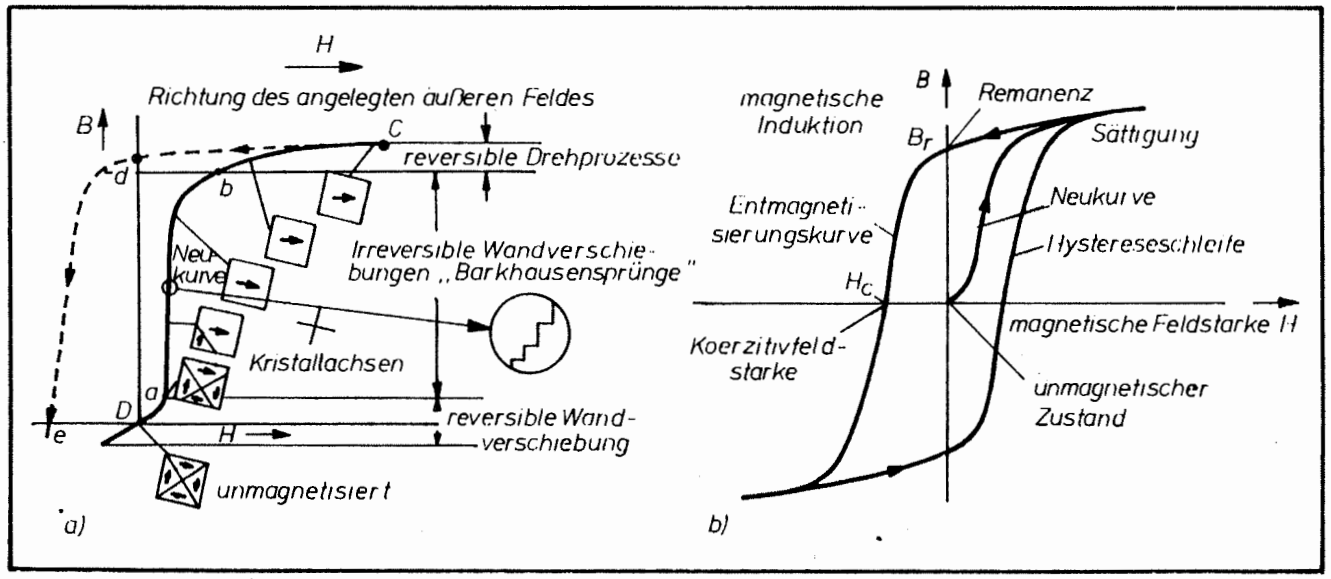


Bild 1.2.6. a) Prozesse und Zusammenhänge beim Magnetisieren eines Materials, b) Grenzhysterese und Neukurve

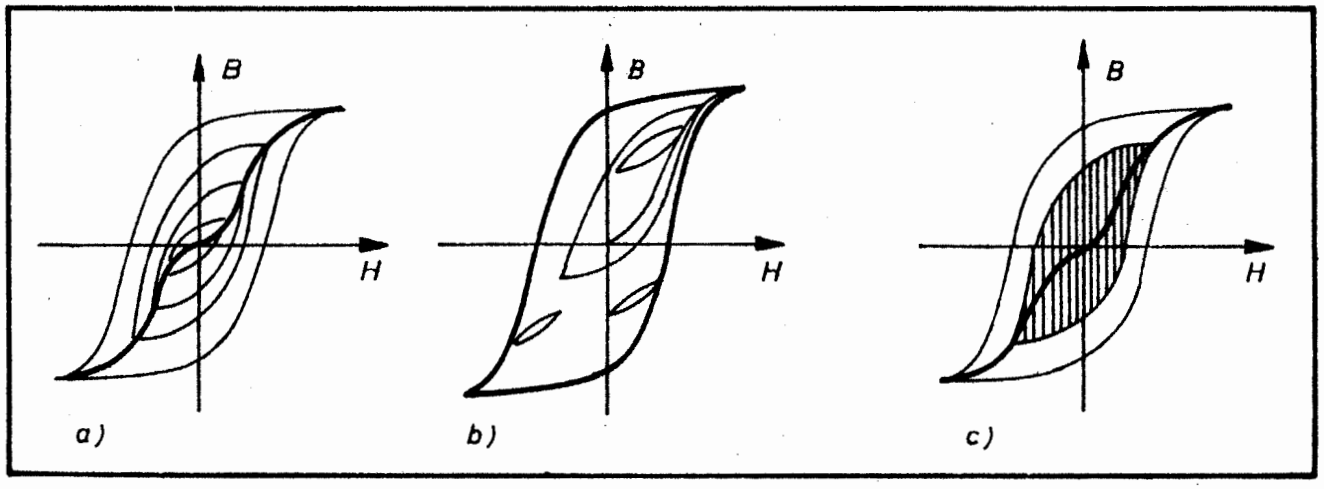


Bild 1.2.7. Ausschnitte aus Hysteresekurven

a) Symmetrische, d. h. gleichstromfreie Aussteuerungen aus dem Nullpunkt
b) Aussteuerungen um unterschiedliche Gleichwerte
c) Die Verluste der Ummagnetisierung entsprechen der eingeschlossenen Fläche

in Bild 1.2.7a) angedeutet. Bild 1.2.7b) zeigt eine Hystereseunterkurve, welche mit Gleich- und Wechselfeldaussteuerung entsteht. Bild 1.2.7c) soll schließlich demonstrieren, daß die in Wärme umgewandelte Energie der Ummagnetisierung der Fläche der entsprechenden Teilkurve entspricht. Dies wird im Abschnitt 1.5. noch genauer betrachtet.

1.2.4. Permeabilitäten

Entsprechend der Vielgestaltigkeit der Hysteresekurve ist es nicht möglich, von einer Permeabilität des Materials zu sprechen. Sie hängt von vielen Parametern, wie Arbeitspunkt in der Hysterese, Aussteuerung und Temperatur, ab. Hier sollen einige spezielle Permeabilitäten zusammengestellt werden. In einem Punkt auf der Hysterese können gemäß Bild 1.2.8a) zunächst die *relative* Permeabilität

$$\mu_r = \frac{1}{\mu_0}\frac{B}{H} \tag{16}$$

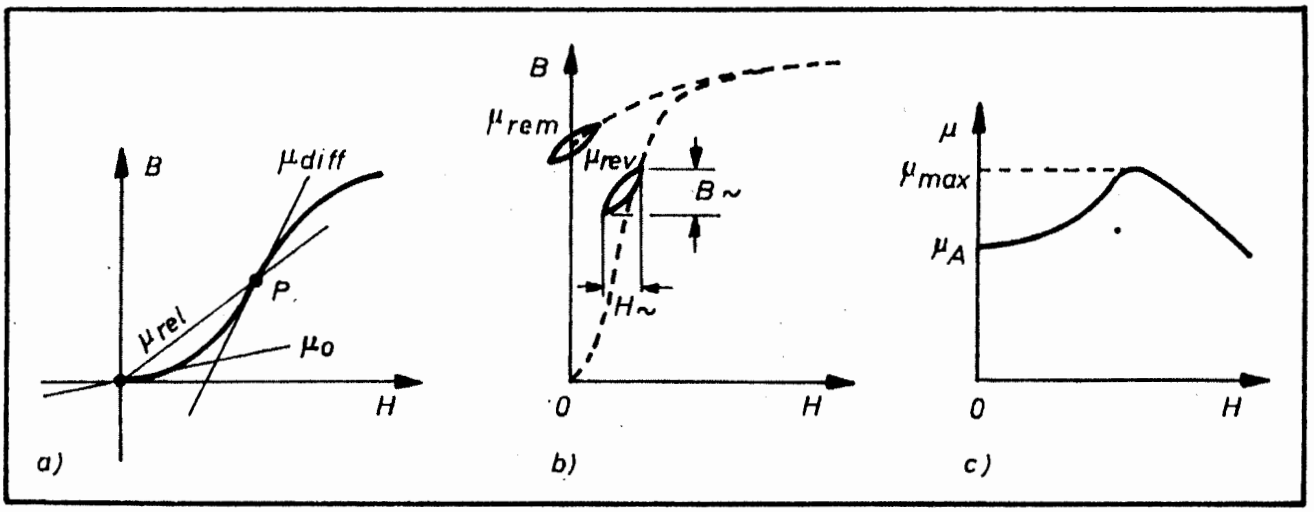


Bild 1.2.8. Zur Definition verschiedener Permeabilitäten

und die *differentielle* Permeabilität als Tangente an die Kurve unterschieden werden,

$$\mu_{\text{diff}} = \frac{1}{\mu_0} \frac{\mathrm{d}B}{\mathrm{d}H}. \tag{17}$$

Die differentielle Permeabilität ist nicht wesentlich verschieden von der *reversiblen*, die bei kleiner Aussteuerung um den Arbeitspunkt erhalten wird (Bild 1.2.8b)):

$$\mu_{\text{rev}} = \frac{1}{\mu_0} \frac{B_\sim}{H_\sim}. \tag{18}$$

Für den Spezialfall der Remanenz geht die reversible Permeabilität in die *remanente* über. Für den Spezialfall des Ursprungs heißt sie *Anfangspermeabilität*. Für größere Aussteuerung heißt die reversible Permeabilität auch *Überlagerungspermeabilität*.

Wie aus Bild 1.2.8b) zu ersehen, nimmt die Permeabilität zunächst mit der Gleichfeldstärke zu. Etwa im Wendepunkt der Neukurve wird so die *Maximalpermeabilität* erreicht (Bild 1.2.8c)).

1.2.5. Magnetischer Kreis

Die magnetischen Flußlinien besitzen im Gegensatz zu den elektrischen Flußlinien weder Quellen noch Senken. Daher läßt sich jede Flußlinie als geschlossener Weg auch durch verschiedene Materialien verfolgen. Auf den einzelnen Teilabschnitten n kann dabei die Permeabilität des Materials und die zugehörige Feldstärke unterschiedliche Werte annehmen, der Fluß Φ muß jedoch konstant sein,

$$\Phi = B_n A_n = \mu_0 \mu_n H_n A_n = \text{const}. \tag{19}$$

Besitzen die einzelnen Materialien jeweils eine Flußweglänge l_n, so gilt mit dem *Durchflutungsgesetz*

$$\Theta = \sum_n H_n l_n, \tag{20}$$

und mit (19) ergibt sich

$$\boxed{\Theta = \sum_n \frac{l_n}{\mu_0 \mu_n A_n} \Phi} \tag{21}$$

Da Θ der elektrischen Spannung und Φ dem elektrischen Strom entsprechen, ist die Größe

$$\boxed{P = \frac{l_n}{\mu_0 \mu_n A_n}} \tag{22}$$

dem elektrischen Widerstand zuzuordnen. Sie heißt *magnetischer Widerstand*.

1.2.6. Energieprodukt

Es liege ein Dauermagnet der Länge l_d und mit der magnetischen Feldstärke H_d vor. Dieser Magnet wirke über Leitbleche entsprechend Bild 1.2.9a) auf einen magnetischen Arbeitsraum mit der Länge l_a. An den Leitblechen soll kein Verlust auftreten. Sie führen also die gesamte magnetische Energie in den *Arbeitsraum* über. Ihre Permeabili-

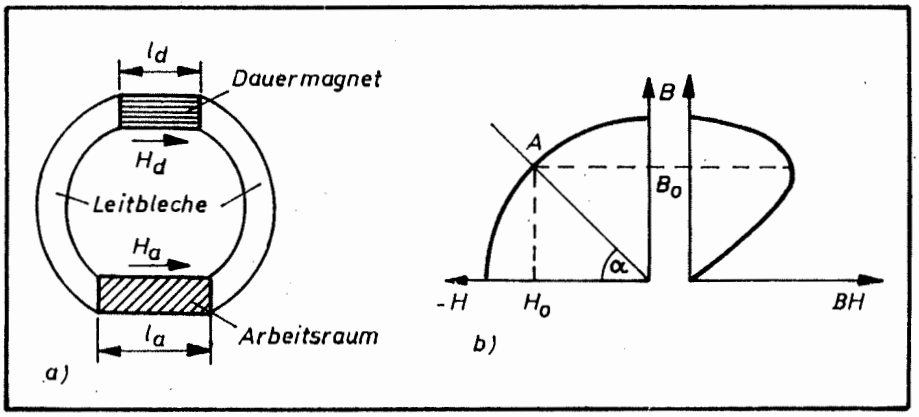


Bild 1.2.9. Zum Dauermagnetkreis
a) Aufbau
b) Zur Bestimmung des Energieproduktes

tät wird dazu unendlich groß angenommen. Dann gilt

$$H_d l_d + H_a l_a = 0 . \tag{23}$$

Andererseits muß der *Gesamtfluß* sowohl im Dauermagneten als auch im Arbeitsraum gleich groß sein:

$$\Phi = B_d A_d = B_a A_a . \tag{24}$$

Werden in beiden Räumen die Volumina $V = A \cdot l$ erfaßt, so entsteht nach kurzer Umrechnung eine Beziehung, die das Energieprodukt enthält:

$$2W_a = V_d B_d H_d = -V_a B_a H_a . \tag{25}$$

Da in ihr sowohl die magnetische Induktion als auch die magnetische Feldstärke stehen, ist das Energieprodukt vom Arbeitspunkt des magnetischen Materials abhängig. Es kann am besten entsprechend Bild 1.2.9b) aus der Hysteresekurve des magnetischen Materials im 2. Quadranten bestimmt werden. Danach wird die maximale Energie von einem Dauermagneten in einen Arbeitsraum genau dann überführbar, wenn ein bestimmtes entmagnetisierendes Feld H_0 auftritt. Dieses

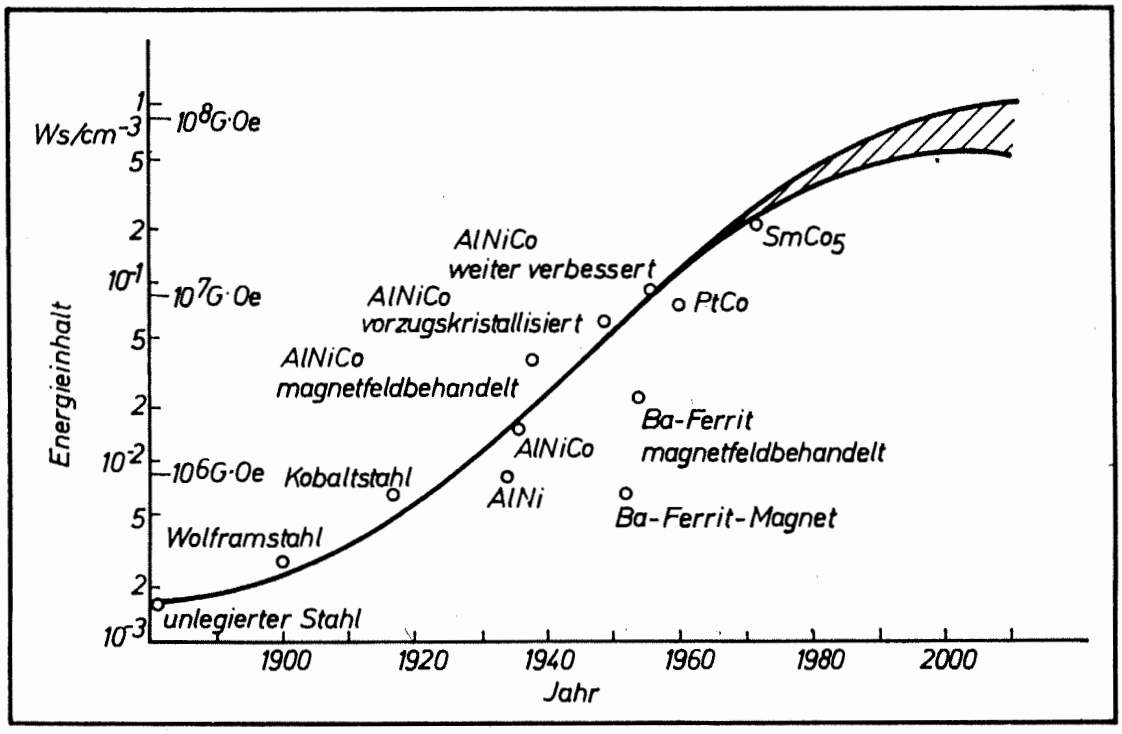


Bild 1.2.10. Zur Entwicklung der Energiedichte magnetischer Materialien

Feld ist entsprechend den vorangegangenen Gleichungen stets durch eine geschickte Geometrie des Magneten und Arbeitsraumes und damit durch die Gestaltung der Leitbleche zu erreichen.

Unter zusätzlicher Berücksichtigung des nichtlinearen Verlaufs der Magnetisierungskennlinien liegt hier also eine Anpassung auf das maximale Energieprodukt vor, die der Anpassung bei der Stromquelle auf maximale Leistung ähnlich ist. Einen Überblick zur Entwicklung der Energiedichte gibt Bild 1.2.10.

Eine Zusammenstellung der wichtigsten magnetischen Größen zeigt Tab. 1.2.1. Eine Zusammenstellung der möglichen Typen von Hysteresekurven zeigt Bild 1.2.11.

Tabelle 1.2.1. Zusammenstellung der wichtigsten magnetischen Einheiten

Bezeichnung	Formel Symbol	Einheit	Alte Einheit und Umrechnung
Feldstärke	H	A/m	Oersted: 1 Oe $= \frac{10^3}{4\pi}$ A/m
Fluß	Φ	1 Vs = 1 Wb	Maxwell: 1 M $= 10^{-8}$ Wb
Flußdichte = Induktion	B	1 Vs/m^2 = 1 T	Gauß: 1 G $= 10^{-4}$ T
Durchflutung	Θ	A[1])	
Magnetischer Widerstand	P	1 A/Vs $= \frac{1}{\Omega s} = \frac{1}{H}$	
Magnetische Spannung	$U_m = P\Phi$	A	Gilbert: 1 Gb $= \frac{10}{4\pi}$ A

[1]) oder in Ampere × Windungen

1.2.7. Kreis mit Luftspalt

In vielen Fällen ist es nicht möglich oder nicht erwünscht, einen magnetischen Kreis aufzubauen, der vollständig im magnetischen Material verläuft. Dies tritt z.B. auf, wenn kleine Proben im Magnetfeld gemessen werden sollen oder der Luftspalt gezielt genutzt werden soll, u. a. für Kraftwirkungen oder zur Linearisierung der Hysteresekennlinien. In diesen Fällen wird auch davon gesprochen, daß der Magnetkreis offen bzw. nicht geschlossen ist. Die Feldlinien müssen aber in jedem Fall über den vorhandenen Luftweg geschlossen sein. In (19) bedeutet dies, daß dann im Luftweg der Feldlinien $\mu_n = 1$ ist. Der entsprechende magnetische Widerstand ist folglich selbst bei kleinen Spalten vergleichsweise gegenüber dem anderen Teil des Feldlinienweges groß. Als Folge davon ist das Magnetfeld im Material H_i z. T. wesentlich kleiner als im Luftweg H_a (äußeres Feld). Dies betrachtet man formal so, als ob der Luftweg ein *entmagnetisierendes Feld* H_N bewirkt:

$$H_i = H_a - H_N . \tag{26}$$

Das entmagnetisierende Feld ist der Magnetisierung M des Materials mit einem *Entmagnetisierungsfaktor* N, der nur von den geometrischen Abmessungen des Materials und dem damit bedingten Luftspalt abhängt, proportional,

$$H_N = NM . \tag{27}$$

Für eine Kugel aus magnetischem Material gilt z. B. $N = 1/3$. Die Auswirkungen der Entmagnetisierung auf die Hysteresekurve zeigt Bild 1.2.12. Die Gl. (27) führt zur *Scherungsgeraden* und verschiebt die Hysteresekennlinie entsprechend verflachend nach rechts. Die Werte der Koerzitivfeldstärke bleiben dabei erhalten. Kleiner wird vor allem die Remanenz.

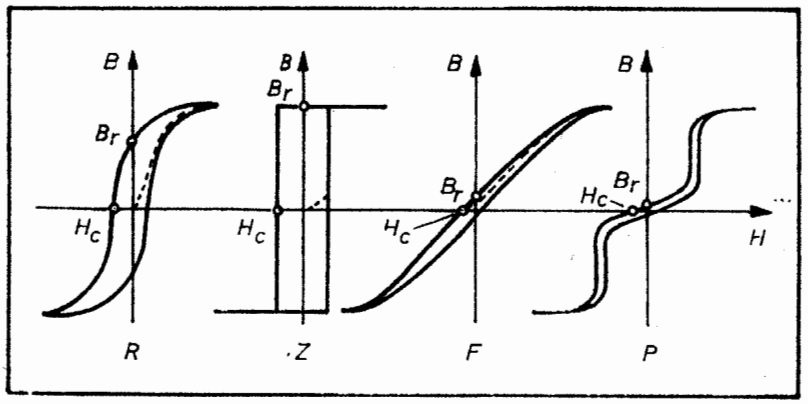


Bild 1.2.11.

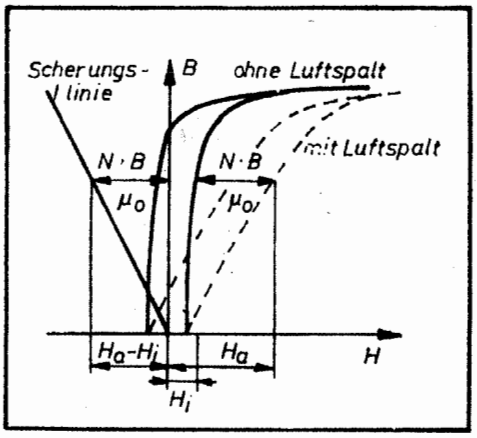


Bild 1.2.12.

Bild 1.2.11. Die vier typischen Hysteresekurven: *R* — die runde (hier nur behandelte); *Z* — die rechteckförmige, welche bei der Speicherung und näherungsweise bei Materialien mit hoher Anfangspermeabilität erwünscht ist; *F* — die flache, mit kleinen Verlusten; und *P*, welche insbesondere in perminvar-Materialien auftritt

Bild 1.2.12. Zur Scherung der Hysteresekurve durch Luftspalt bzw. Entmagnetisierung

1.2.8. Kraftwirkungen

In Magnetfeldern oder zwischen Magnetfeldern treten *drei Fälle von Kraftwirkungen* auf:

- Der *Luftspalt* versucht sich zu *verkleinern*. Dies hat seine Anwendung z.B. bei Relais und Anzugsmagneten.
- Zwischen Magnetpolen wirken *Kräfte des Anziehens bzw. Abstoßens*.
- Das Magnetfeld *wirkt* auf einen stromdurchflossenen Leiter bzw. *auf bewegte Ladungen*, z.B. einen Elektronenstrahl.

Im Prinzip sind alle diese Wirkungen über die Feldenergie zu berechnen. Für sie gilt

$$W_\mathrm{m} = \int_0^V \int_0^B \boldsymbol{H}\,\mathrm{d}\boldsymbol{B}'\,\mathrm{d}V' . \tag{28}$$

Zwischen Kraft F und Energie besteht über eine differentielle Wegänderung ds die Beziehung

$$F = \frac{\mathrm{d}W}{\mathrm{d}s} . \tag{29}$$

Hiermit folgt im homogenen Magnetfeld des Luftspaltes mit dem Querschnitt A

$$F = \frac{B_\mathrm{L}^2 \cdot A}{2\mu} \tag{30}$$

oder, auf den magnetischen Fluß bezogen,

$$F = \frac{1}{2}\,\Phi_\mathrm{L} \cdot H_\mathrm{L} . \tag{31}$$

Im homogenen Magnetfeld wird auf einen Leiter der Länge l, mit dem Strom I und dem Winkel α gegenüber dem Feld die Kraft

$$F = BIl \sin\alpha \tag{32}$$

ausgeübt.

1.2.9. Magnetische Werkstoffe

Die beiden *Hauptkenngrößen* der magnetischen Werkstoffe sind die *Sättigungsmagnetisierung* und die *Koerzitivfeldstärke*. Hieraus läßt sich eine Klassifikation gemäß Bild 1.2.13a) aufstellen. Die drittwichtigste Größe dürfte dann die *Anfangspermeabilität* sein. Sie steht in einem locker proportionalen Zusammenhang zur Koerzitivfeldstärke (Bild 1.2.13b)). Aus dem Bild ist weiterhin die Einteilung in weich- und hartmagnetische Werkstoffe zu erkennen. Der Einsatz der *hartmagnetischen Materialien* beruht im wesentlichen auf ihrem *Energieinhalt*, während die *weichmagnetischen* als gute *magnetische Leiter* verwendet werden. Sie führen die magnetischen Felder dorthin, wo sie benötigt werden, und gestalten sie so, daß der erwünschte Zweck erreicht wird. In diesem Sinne gibt es in Analogie zu Bild 1.2.10 auch eine Entwicklung bezüglich der Permeabilität. Leider gibt es hierzu noch keine Kurve.

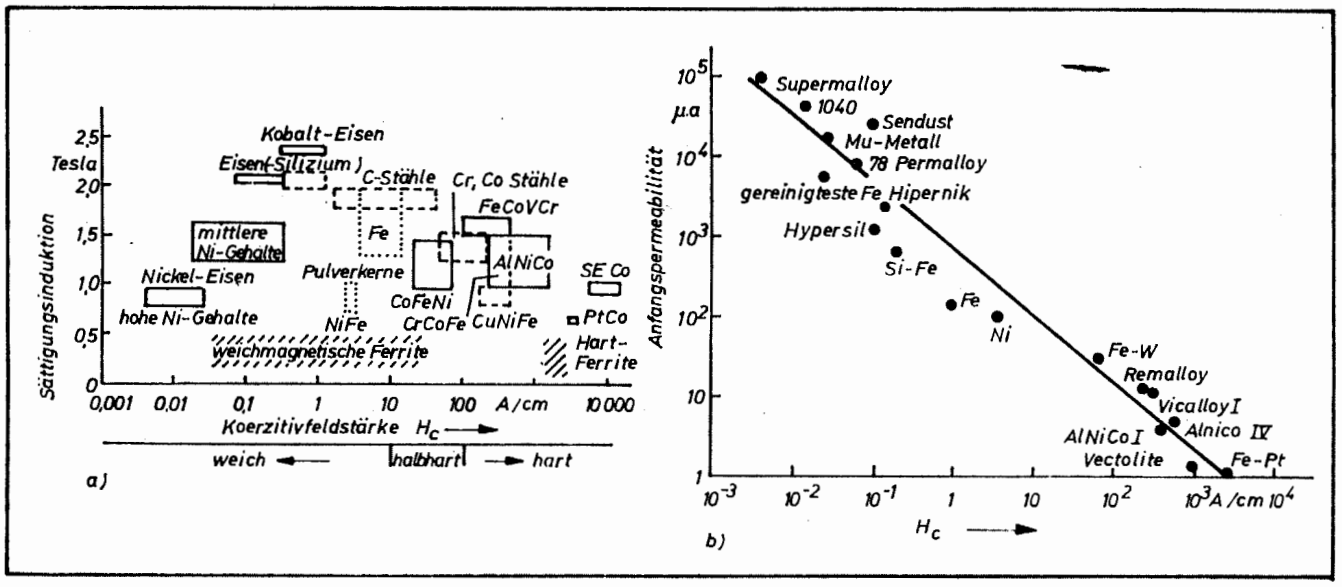


Bild 1.2.13. a) Einteilung der magnetischen Werkstoffe nach Koerzitivfeldstärke und Sättigungsinduktion
b) Zusammenhang von Permeabilität und Koerzitivfeldstärke

Nach der *Materialgrundlage* können heute zwei Arten von Magnetmaterialien unterschieden werden, *Metalle* und *Ferrite*. Ihre Herstellungsprinzipien sind in Bild 1.2.14 dargestellt. Die vielfältigen Prozesse sind entscheidend für die spezifischen Eigenschaften der Materialien. Fast immer sind daher hochwertige magnetische Materialien gegenüber *mechanischen Bearbeitungen* und Beanspruchungen *empfindlich*. Sie müssen entweder sofort in der richtigen Form hergestellt oder kompliziert nachbehandelt werden.

Einen Überblick zu den wichtigsten magnetischen Materialien gibt Tab. 1.2.2. Die *Ferrite* haben allgemein die chemische Zusammensetzung $MeO \cdot Fe_2O_3$ bzw. $Me^1O \cdot Me^2O \cdot Fe_2O_3$, wobei Me ein Metall, z.B. Barium, Strontium, Mangan usw., bedeutet. In diesem Sinne gehört auch $FeO \cdot Fe_2O_3$ zu den Ferriten.

Bei den *metallischen Gläsern* (oder amorphen Metallen) wird die Schmelze direkt auf eine schnell rotierende Walze gespritzt und dabei mit etwa 10^6 K/s extrem schnell abgekühlt. Als Material wird dabei verwendet $T_{70\ldots88} \cdot M_{15\ldots30}$, wobei T *Metalle* wie Fe, Co, Ni, Cr und M *Metalloide* (Glasbildner) wie B, P, C, Si sind. Dadurch entstehen sehr kleine Weisssche Bezirke, und es lassen sich spezielle Eigenschaften, z. T. auch durch Nachbehandlung, erreichen, z.B. für Magnetköpfe Materialien, die zugleich magnetisch weich und mechanisch hart sind.

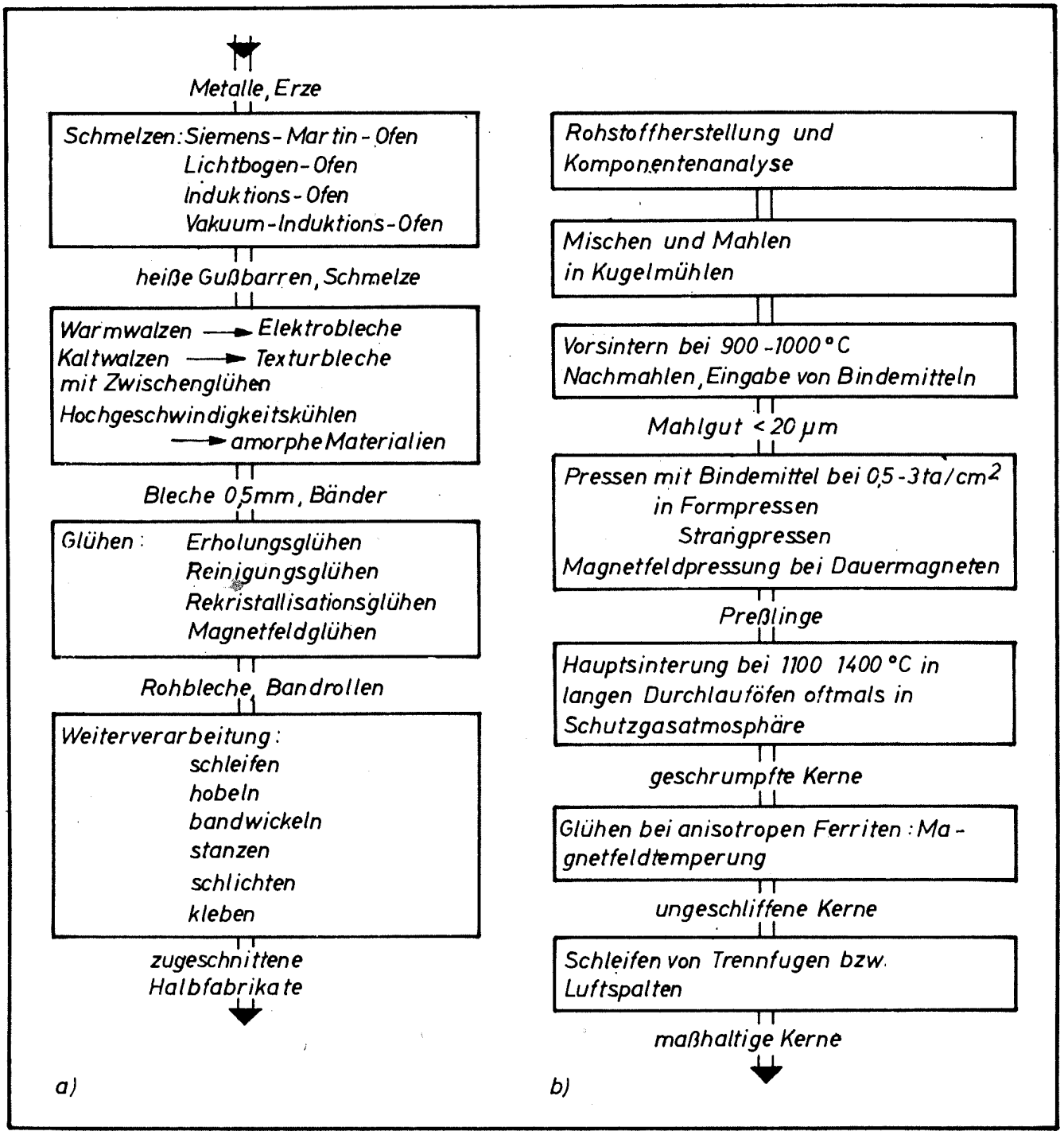


Bild 1.2.14. Die wesentlichen Herstellungsprozesse für a) metallische Magnetmaterialien und b) Ferrite

Erwähnt sei schließlich noch die *Anisotropie* von magnetischen Materialien. Sie besitzen in verschiedenen Richtungen unterschiedliche Eigenschaften. Dies wird unter anderem durch *Kaltwalzen* (es entsteht eine *Textur*) oder durch Magnetfeldglühung erreicht. Hierdurch werden in einer bestimmten Richtung auf Kosten der anderen bessere magnetische Werte erreicht.

Für verschiedene Anwendungen ist es schließlich nicht günstig oder möglich, kompaktes magnetisches Material zu verwenden. Dann werden kleine Magnetpartikelchen in ein Bindemittel (Matrix) eingefügt. So entstehen u. a. die *Pulverkerne*, der *magnetische Gummi*, die *magnetischen Flüssigkeiten* und die Schichten auf dem *Magnetband*.

Tabelle 1.2.2. Zusammensetzungen und Eigenschaften der wichtigsten magnetischen Materialien

a) Weichmagnetische Materialien

	%-Anteile	ϱ $\frac{\Omega \cdot mm^2}{m}$	T_c °C	B T	H_c A/cm	μ_a	μ_{max}	P W/kg
	1	2	3	4	5	6	7	8
Reine Metalle	99,9 Fe	0,12	770	2,2	0,1 ... 0,8	20000	680000	
	99% Ni	0,068	358	0,6	0,1 ... 1,6	200	1000	
	99% Co		1120	1,8	8	70	240	
Legierungen	Fe + 4 Si	0,4 ... 0,6	750	2,0	0,2 ... 0,6	500	10000	0,5 ... 1,2
	Fe + 36 ... 40 Ni	0,65	250	1,3	0,15 ... 0,6	4500	30000	0,15 .. 0,6
	Fe + 50 Ni	0,45	470	1,5	0,1	2300	20000	0,07 ... 0,3
	Fe + 75 ... 80 Ni	0,5	400	0,6 ... 0,8	0,01 ... 0,08	30000	100000	0,05 ... 0,17
	Fe + 35 ... 50 Co	0,2	970	2,4	0,25 ... 2,5	8000	40000	
	Fe + 16 Al	1,45	300	0,75	0,04	30000	100000	
	Fe + 79 Ni + 5 Cr	0,55	400	0,8	0,004	100000	900000	
	Fe + 9,6 Si + 5,4 Al	0,9	500	1,1	0,01 ... 0,04	40000	150000	

	1	2	3	4	5	6
Amorph	80 Fe, 16 P, 3 C, 1 B			1,7	0,04	
	40 Fe, 40 Ni, 14 P, 6 B	1,6	230	0,86	0,016	12000
Ferrite	Pulverkerne					
	Kerngröße 1 ... 10 μm	1 ... 10 · 10^4		1,2 ... 2	4 ... 15	3 ... 75
	25 ... 40 MnO, 10 ... 20 ZnO	10^6	120 ... 220	0,4 ... 0,5	0,1 ... 0,6	500 ... 3000
	15 ... 40 NiO, 5 ... 35 ZnO, 1 CoO	10^7 ... 10^9	150 ... 500	0,2 ... 0,4	0,2 ... 5	100 ... 2000
	20 MgO, 72 ZnO	10^7	125	0,3	0,3	500
	10 ... 75 MgO, 5 ... 20 MnO	10^7	150 ... 300	0,2 ... 0,3	0,5 ... 1,5	40 ... 80
	18 BaO, 12 CoO	10^9	250	0,25	8	10 ... 20
	Fe-Y-Granat	10^{13}	100 ... 300	0,01 ... 0,2	1	<10

(Fortsetzung Tabelle 1.2.2)

b) Hartmagnetische Materialien

%-Anteile	T_c °C	$B \cdot M_{max}$ kJ/m³	B_r T	H_c A/cm
Fe + 12 Al + 25 Ni	720	4 ... 10	0,7	400
Fe + 8 Al + 14 Ni + 23 Co	830	15 ... 60	1,2	500
76,7 Pt + Co		60 ... 100	0,8	2000
Cobalt und seltene Erden	ca. 700	120 ... 200	0,9 ... 1,1	10000
Mn Al	—	50	0,58	2600
$BaO \cdot 6\,Fe_2O_3$	450	7,5	0,2	135
$PbO \cdot 4\,Fe_2O_3$	455	5	0,25	120
$Sr \cdot 6\,Fe_2O_3$		30	0,3	195

1.3. Elektrostatik

Die Grundgröße der Elektrostatik ist die *freie Ladung*, das ruhende Elektron. Von jeder Ladung breitet sich ein elektrostatisches Feld aus. Es ist dasselbe Feld, unter dessen Einfluß sich die Elektronen beim stationären Strom bewegen. Dieses Feld breitet sich mit Lichtgeschwindigkeit, also mit rund 300000 km/s, und damit im Vergleich zum stationären Strom extrem schnell aus. Formal bestehen zwischen den drei elektrischen Gegebenheiten Analogien. Sie sind in Tab. 1.3.1 zusammengestellt und bedürfen nur geringer Erklärung. Entsprechend dem mikro-Ohmschen Gesetz Gl. (1.1.8) und dem Zusammenhang beim Magnetismus gemäß Gl. (1.2.12) gilt jetzt der Zusammenhang zwischen dem *elektrischen Feld* $\boldsymbol{E}$ und der *dielektrischen Verschiebung* $\boldsymbol{v}$

$$\boxed{\boldsymbol{v} = \varepsilon_0 \varepsilon_r \boldsymbol{E}} \,. \tag{1}$$

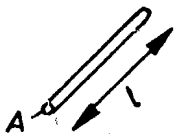
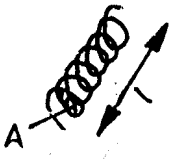
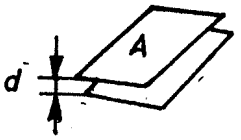
Darin ist ε_r die relative (materialabhängige) Dielektrizitätskonstante (DK). Für die absolute *Dielektrizitätskonstante* gilt

$$\varepsilon_0 = 8{,}854185 \cdot 10^{-12}\ \mathrm{As/Vm} \,. \tag{2}$$

In der Elektrostatik hat der *elektrostatische Widerstand* P_C kaum eine Bedeutung. Wichtig ist dagegen der reziproke Wert, die *Kapazität* C.

Die Elektrostatik ist in allen Lehrbüchern das Musterbeispiel für die *Äquipotentiallinien* und *Verschiebelinien*. Deshalb seien auch hier die beiden Grundbilder mit gleicher und entgegengesetzter Ladung in Bild 1.3.1 vorgestellt. Bei den entgegengesetzten Ladungen fällt die Symmetrielinie mit dem Potential O auf. Hier münden alle Verschiebungslinien senkrecht ein. Deshalb kann an diese Stelle auch ein geradliniger ebener Leiter gesetzt werden, ohne daß sich etwas ändert. Insofern ist es berechtigt, diesen Leiter als Spiegel und die dahinterliegende Ladung als das Spiegelbild der davorliegenden aufzufassen. Umgekehrt gilt genauso, daß eine Ladung, die einem Leiter gegenübersteht, fiktiv ein Spiegelbild ihrer Ladung bewirkt. Dies kann u. a. für die Berechnung der Linien von großem Vorteil sein. Von hier aus ist es auch leicht, zur *Influenz* zu gelangen. Sie bewirkt infolge von Kraftwirkungen in den elektrischen Feldern eine Ladungstrennung auf Leitern. Dies ist im Bild 1.3.2a) am Beispiel einer metallischen Kugel dargestellt. Wird jetzt die Kugel zerteilt, so erfolgt eine *Ladungstrennung* (Bild 1.3.2b)).

Tabelle 1.3.1. Vergleich von stationärem Strom, Elektrostatik und Magnetostatik

Stationärer Strom	Magnetostatik	Elektrostatik
		
Ursache für Strom ist Feldstärke	Ursache für magnetischen Fluß ist Feldstärke	Ursache für Verschiebung ist Feldstärke
Spannungs-Stromquelle	magnetischer Dipol Kreisstrom	freie Ladung
Äquipotentiallinien	Linien für konstantes magnetisches Potential Θ	Äquipotentiallinien
Stromlinien	Flußlinien	Verschiebungslinien
i; Strom	$\boldsymbol{B}$; Induktion	$\boldsymbol{D}$; Verschiebung
$I = \int i \, \mathrm{d}A$	$\Phi = \int \boldsymbol{B} \, \mathrm{d}A$	$Q = \int \boldsymbol{D} \, \mathrm{d}A$
$\boldsymbol{E} = \varrho i$	$\boldsymbol{H} = \frac{1}{\mu} \boldsymbol{B}$	$\boldsymbol{E} = \frac{1}{\varepsilon} \boldsymbol{D}$
$U = RI$	$\Theta = P\Phi$	$U = \frac{1}{C} Q$
$R = \frac{\varrho l}{A}$	$P_\mathrm{m} = \frac{l}{\mu \mu_\mathrm{r} A}$; $L = n^2/P$	$P_\mathrm{C} = \frac{\mathrm{d}}{\varepsilon/\varepsilon_\mathrm{r} A}$; $C = 1/P_\mathrm{C}$
$U = \int \boldsymbol{E} \, \mathrm{d}s$	$\Theta = \oint \boldsymbol{H} \, \mathrm{d}s = \Sigma I$	$U = \int \boldsymbol{E} \, \mathrm{d}s$
$P = UI = I^2 R$	$W = \Phi L = I^2 L$	$W = U \cdot Q = U^2 C$
$\dot{Q} = I$	$\dot{\Phi} = U$	$\dot{\boldsymbol{D}} = I$

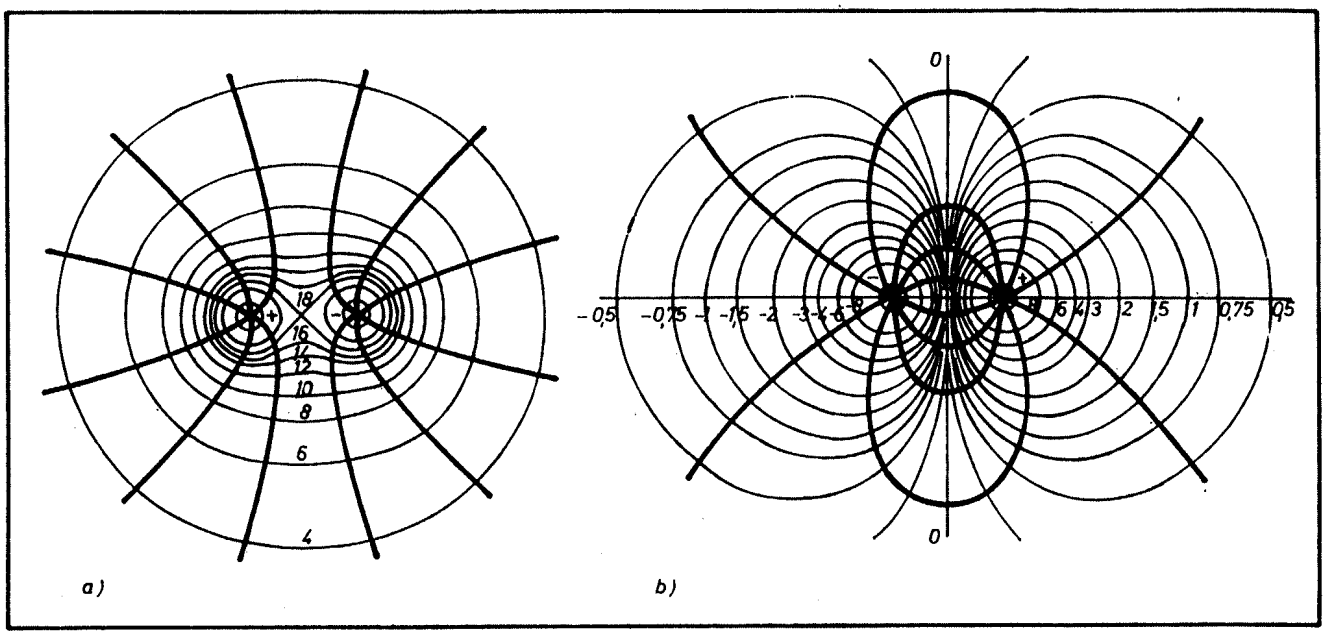


Bild 1.3.1. Verschiebungslinien (= Kraftlinien: ausgezogen) und Äquipotentiallinien von zwei a) gleichen und b) entgegengesetzten Ladungen

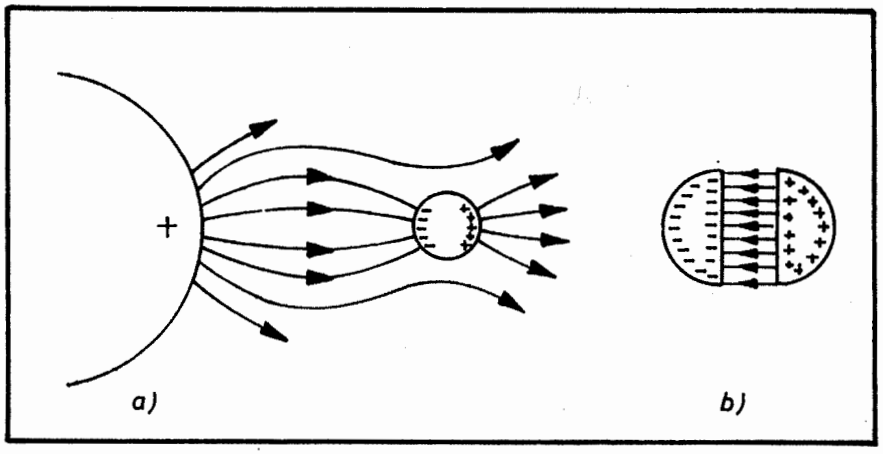


Bild 1.3.2. a) Von einer Ladung (links) ausgehende Feldlinien bewirken auf einer metallischen Kugel Ladungstrennung. b) Wird diese Kugel im Feld geteilt, so entstehen zwei Teile mit eigener Ladung

1.3.1. Isolierstoffe

Isolierstoffe werden in vielfältiger Weise eingesetzt, u. a. als Dielektrikum in Kondensatoren, zur Isolation von Leitern (z.B. Drähten), zur Vermeidung von Luftblasen und ähnlichen Löchern, an denen Durchschläge erfolgen könnten (u. a. als Tränkmittel), und zum Schutz, z.B. als Gehäuse oder Einbettungen. Diese Vielfalt der Anwendungen verlangt auch eine Vielfalt von *Materialien*. Dabei können u. a. drei unabhängige Einteilungen vorgenommen werden:

- Chemie: anorganisch — organisch
- Herkunft: natürlich — synthetisch
- Phase: flüssig — fest (kompakt — Folie — dünne Schichten).

Tabelle 1.3.2. Kennwerte einiger Isolierstoffe

Material	ε_r	Spezif. Widerstand $\Omega \cdot$ cm	Temperaturkoeffizient $10^6 K^{-1}$	Durchschlagfestigkeit kV/cm
Hartpapier	5		20 ... 40	>200
Papier	2,5		500	≈100
Papier (clophengetränkt)	5,4	$>10^{14}$		600
Glimmer	7	10^{15}	20 ... 30	>1000
Glas	3.5 ... 9	10^{13}		100 ... 400
Polystyrol (Styroflex)	2,5	10^{17} ... 10^{21}	−150	800
Polyester	3,2		+500	2000
Polycarbonat	2,9	10^{18}	+100	1500
Polyethylen	2,7	10^{16}	160	400
Polypropylen	2,2	10^{16}	−50	1400
Polytetrafluorethylen (PTFE)	2,1	10^{18}	100	1000
Polyacetat	3,7	10^{15}	100	600
Al_2O_3	7,5	10^{11}	+100	2000
Ta_2O_3	27		+100 ... +500	6000
SiO_2	4		+20 ... +40	2000
Si_3N_4	6,2	10^{15}	+4 ... +10	1000
NDK-Keramik	10 ... 80	10^{13}	−800 ... +100	≈ 200
HDK-Keramik ($BaTiO_3$)	1000—10000	10^9	−20000	20 ... 100

Besonders vielfältig sind heute die *hochpolymeren Kunststoffe*. Relativ groß ist auch das Spektrum der *keramischen Materialien*. Diese bestehen aus den Hauptkomponenten (wie z. B. TiO_2, ZrO_2, $BaCO_3$, Pb_3O_4), den Nebenkomponenten (wie z. B. $MgCO_3$, $CaCO_3$, $SrCO_3$, SnO_2, Nb_2O_5) und werden mit Oxiden der Übergangsmetalle dotiert. Dabei sind die NDK- und HDK-Stoffe, also solche mit niedriger und hoher Dielektrizitätskonstanten zu unterscheiden. Die HDK sind stark temperatur- und erheblich spannungsabhängig. Sie besitzen auch eine gewisse Hysterese.

Die wichtigsten Eigenschaften von Isoliermaterialien sind in Tab. 1.3.2 zusammengestellt. Nicht berücksichtigt sind dabei zwei spezielle Eigenschaften, nämlich die der *Piezoelektrizität* (s. 6.3.4.2.) und die der Elektrete. Die *Elektrete* sind in der Regel Folien, denen eine permanente Polarisation eingeprägt werden kann. Sie werden hierzu mit Feldstärken um 10 kV/cm behandelt und besitzen danach Oberflächenladungen mit einer Dichte um 10^7 As/cm^2.

1.4. Wechselstrom

1.4.1. Definition

Bisher wurden nur Quellen mit Gleichstrom bzw. Gleichspannung betrachtet. Dabei wurde vorausgesetzt, daß ihre Größen sich nicht änderten. Dies ist jedoch ein Ideal, was zwar für verschiedene Anwendungen angestrebt wird, jedoch nur sehr selten der Wirklichkeit entspricht. *Änderungen* sind in *drei Arten* vorhanden:

1. Infolge Erschöpfung einer Batterie ändert sich ihre Spannung bzw. ihr Strom so langsam, daß über relativ große Zeiträume statische Verhältnisse garantiert sind. Hier genügt also eine *quasistationäre* Betrachtung *mit Gleichstromcharakter*.
2. Die Spannung bzw. der Strom nehmen ständig und relativ schnell zu und ab. Die Ursache hierfür kann z. B. eine sich ändernde Belastung oder ein Wackelkontakt sein. In geringem Umfang tragen hierzu auch die thermischen Bewegungen der Elektronen bei. Es gibt dann einen *mittleren Gleichstromwert*, dem ein *zeitlich sich ändernder* positiver und negativer *Strom überlagert* ist.
3. Es liegt eine Spannung vor, die *keinen* mittleren *Gleichanteil* mehr enthält, z. B.:

$$\int_0^T U(t)\,\mathrm{d}t = 0\,. \tag{1}$$

Eine solche Spannung heißt *Wechselspannung*. Es gilt der *allgemeine Satz*:

> Jede sich verändernde Spannung bzw. jeder sich verändernde Strom kann in zwei Anteile — den Gleich- und den Wechselanteil — zerlegt werden. In Sonderfällen ist einer der beiden Anteile Null. Dann liegt reiner Gleichstrom oder reiner Wechselstrom (-spannung) vor.

1.4.2. Periodische Wechselströme

Zur Vereinfachung wird im folgenden immer von Strömen gesprochen. Gleichberechtigt gelten dieselben Betrachtungen auch für Spannungen.

Reiner Wechselstrom ist dann besonders einfach sowohl theoretisch als auch praktisch zu handhaben, wenn periodisch nach einer bestimmten Zeit T immer wieder derselbe Wert angenommen wird:

$$F(t) = F(t - T)\,. \tag{2}$$

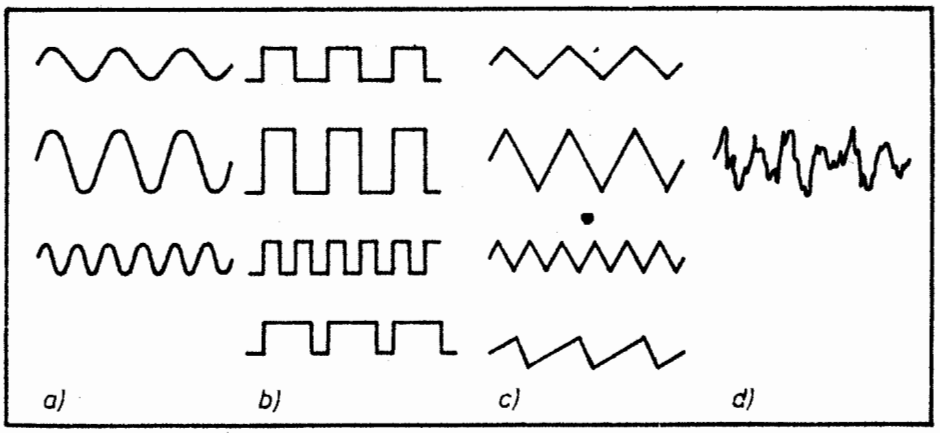


Bild 1.4.1. Schwingungen verschiedener Form: a) Sinusschwingung, b) Rechteckschwingung, c) Dreieckschwingung, d) Rauschen. In der ersten Zeile ist die Bezugsschwingung dargestellt, die zweite Zeile zeigt eine veränderte Amplitude, die dritte eine veränderte Frequenz, die vierte eine unsymmetrische Schwingung

Unter den *periodischen Schwingungen* gibt es *drei* wichtige *Sonderfälle* (Bild 1.4.1)

- Sinus-, auch Kosinus- oder *harmonische Schwingung* genannt. Sie ist vollständig durch drei Größen definiert: Amplitude, Frequenz und Anfangsphase.
- *Rechteckschwingung.* Sie ist durch zwei Spannungswerte, die Wiederholfrequenz (Periodendauer) und das Zeitverhältnis der beiden Spannungswerte (Impulsverhältnis) gegeben.
- *Dreieckschwingung.* Sie kann als zeitliche Ableitung der Rechteckschwingung aufgefaßt werden.

Die Sinusschwingung ist theoretisch besonders übersichtlich. Sie ist gut auszufiltern und relativ leicht zu erzeugen. Die Rechteckschwingung ist Grundlage der digitalen und der Impulstechnik und steht damit heute nahezu gleichberechtigt neben der Sinusschwingung. Eine Sonderform der Rechteckschwingung sind die WALSH-*Funktionen.* Die Dreieckschwingung hat verschiedene Sonderanwendungen (u. a. Ablenkung in Oszillographen und Fernsehempfängern und AD-Wandler).

Große Bedeutung besitzen weiterhin die *stochastischen Signale*, z. B. in Form von Rauschspannungen. Bei ihnen sind die momentane Amplitude und die Änderung der Spannung durch Zufallsfunktionen bestimmt. Weißes Rauschen enthält alle Frequenzen mit gleicher Amplitude. Beim GAUSSschen Rauschen gehorcht die Amplitude der Normalverteilung. Es gibt dann also einen mittleren Wert, den Erwartungswert, und eine Streuung, die statistische Abweichung von diesem Wert.

1.4.3. Harmonische Schwingung

Zur Beschreibung der harmonischen Schwingung existieren zwei Formen: die reelle Schreibweise

$$u = \hat{U}_0 \cdot \cos(2\pi f t + \varphi) = \hat{U}_0 \cos(\omega t + \varphi) \tag{2a}$$

und die komplexe

$$\underline{u} = \hat{U}_0 \cdot e^{j(2\pi f t + \varphi)} = \hat{U}_0 \, e^{j(\omega t + \varphi)} . \tag{2b}$$

Darin bedeuten:

$\hat{U}_0$ Amplitude der Schwingung
e Basis der Exponentialfunktion, $e \approx 2{,}71$

j imaginäre Einheit, $j = \sqrt{-1}$. Da i die Stromdichte bzw. Wechselstrom ist, wird in der Elektrotechnik ausschließlich j verwendet.
f Frequenz, Anzahl der Perioden pro Sekunde
ω Kreisfrequenz ($2\pi f$). Sie ergibt sich als pro Sekunde durchlaufene Strecke auf dem Einheitskreis. Da in einigen theoretischen Büchern der Unterschied zwischen normaler und Kreisfrequenz nicht eindeutig angegeben wird, ist auf den Faktor 2π besonders zu achten. In der Regelungstechnik ist es ferner üblich, die komplexe Frequenz $j\omega$ mit p zu bezeichnen.
t Zeit
φ Anfangsphasenwinkel.

Die graphische Darstellung einer Sinusschwingung kann auf vielfältige Art geschehen. In einer GAUSSschen Ebene (Bild 1.4.2) kann z.B. ein Vektor der Länge $\hat{U}_0$ mit der Frequenz f umlaufend angesetzt werden. Hier gilt dann exakt (2b). Die Projektion auf die reelle Achse ergibt den jeweiligen Momentanwert der Schwingung. Wird dieser als Funktion der Zeit aufgetragen, so entsteht die reelle Schwingung entsprechend (2a), die in Bild 1.4.2b) dargestellt ist. Anstelle des Vektors kann auch das Koordinatensystem entgegengesetzt rotieren. Dann ist es sinnvoll, nur die reelle Achse darzustellen (Bild 1.4.2c)). In vielen Fällen wird auch sie noch fortgelassen.

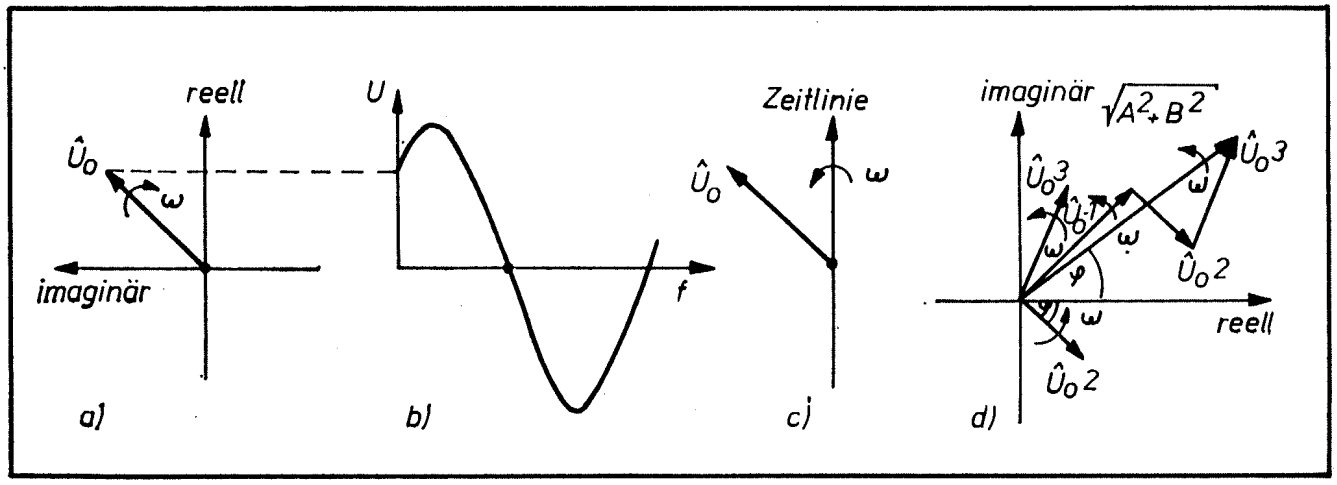


Bild 1.4.2. Darstellung der harmonischen Schwingung

a) im komplexen Koordinatensystem durch einen mit gegen den Uhrzeigersinn rotierenden Vektor der Länge $\hat{U}_0$. Das Koordinatensystem wurde (ausnahmsweise) um 90° verdreht. Dadurch läßt sich die Projektion des rotierenden Vektors auf die reelle Achse leicht in die Zeitfunktion (b) übertragen

c) Gleichwertig ist die Darstellung mit feststehendem Vektor $\hat{U}_0$ und rotierender reeller Achse, Zeitlinie

d) Addition von drei Schwingungen gleicher Frequenz, unterschiedlicher Anfangsphase und Amplitude in der Darstellungsform von a)

1.4.4. Zusammenschaltung von Wechselquellen

Bei jeder Wechselquelle ändert sich ständig der Momentanwert. Werden mehrere Quellen zusammengesetzt, so addieren sich die Momentanwerte. Bei einer Darstellung gemäß Bild 1.4.2 wären also sowohl die Amplituden, Phasen als auch Frequenzen der einzelnen Vektoren zu erfassen. Dies kann zu komplizierten Additionen führen. In der mathematischen Darstellung von (2b) gilt:

$$\begin{aligned} u &= u_1 + u_2 + u_3 + \cdots \\ &= \hat{U}_{01} \cdot \cos(\omega_1 t + \varphi_1) \\ &\quad + \hat{U}_{02} \cdot \cos(\omega_2 t + \varphi_2) \\ &\quad + \hat{U}_{03} \cdot \cos(\omega_3 t + \varphi_3) + \cdots . \end{aligned} \tag{3}$$

Sind die Frequenzen ω_1 bis ω_n unabhängig voneinander, so existiert in jedem Moment ein anderes Schwingungsbild. Es kann so kompliziert sein, daß eine graphische Darstellung praktisch nicht möglich ist. Besonders einfach wird die Addition, sobald die einzelnen Frequenzen und Anfangsphasenwinkel übereinstimmen,

$$\omega_1 = \omega_2 = \omega_3 = \cdots = \omega_0$$

und

$$\varphi_1 = \varphi_2 = \varphi_3 = \cdots = \varphi_0 \,. \tag{4}$$

Dann addiert sich die Amplitude der Vektoren wie bei einer Reihenschaltung von Gleichquellen. Die Vereinfachung ist geringer, wenn nur die Frequenzen gleich sind. Mathematisch kann dann die Zusammenfassung über das Additionstheorem

$$\cos(\alpha + \beta) = \cos\alpha \cdot \cos\beta - \sin\alpha \cdot \sin\beta \tag{5}$$

erfolgen. Dabei gilt

$$\alpha = \omega_0 t \quad \text{und} \quad \beta = \varphi \,. \tag{6}$$

Also folgt

$$\begin{aligned} u &= (\hat{U}_{01}\cos\varphi_1 + \hat{U}_{02}\cos\varphi_2 + \hat{U}_{03}\cos\varphi_3)\cos\omega_0 t \\ &\quad - (\hat{U}_{01}\sin\varphi_1 + \hat{U}_{02}\sin\varphi_2 + \hat{U}_{03}\sin\varphi_3)\sin\omega_0 t \\ &= A\cos\omega_0 t + B\sin\omega_0 t \\ &= \sqrt{A^2 + B^2}\cdot\cos(w_0 t + \varphi_g) \quad \text{mit} \quad \tan\varphi_g = \frac{B}{A} \,. \end{aligned} \tag{7}$$

Es entsteht also wiederum eine Schwingung der Frequenz ω_0 mit einem Phasenwinkel φ_g und einer Amplitude $\sqrt{A^2 + B^2}$. Eine graphische Darstellung dieser Zusammenhänge gemäß Bild 1.4.2a) zeigt Bild 1.4.2d).

1.4.5. Drehstromnetz

Es liegen drei Spannungen gleicher Amplitude und Frequenz vor. Sie sind in der Phasenlage so verkoppelt, daß zwischen jeweils zweien ein Winkel von 120° besteht. Die drei Spannungsquellen besitzen weiter einen gemeinsamen Bezugspunkt und sind daher durch das Vektorfeld Bild 1.4.3a) darstellbar. Hat jeder Vektor eine Länge von 220 V, so besteht zwischen den Endpunkten zweier Vektoren eine Spannung, die sich unter Berücksichtigung des Phasenwinkels zu 380 V ergibt:

$$2 \cdot 220 \cdot \cos 30° = 2 \cdot 220 \cdot \frac{\sqrt{3}}{2} = 380 \,. \tag{8}$$

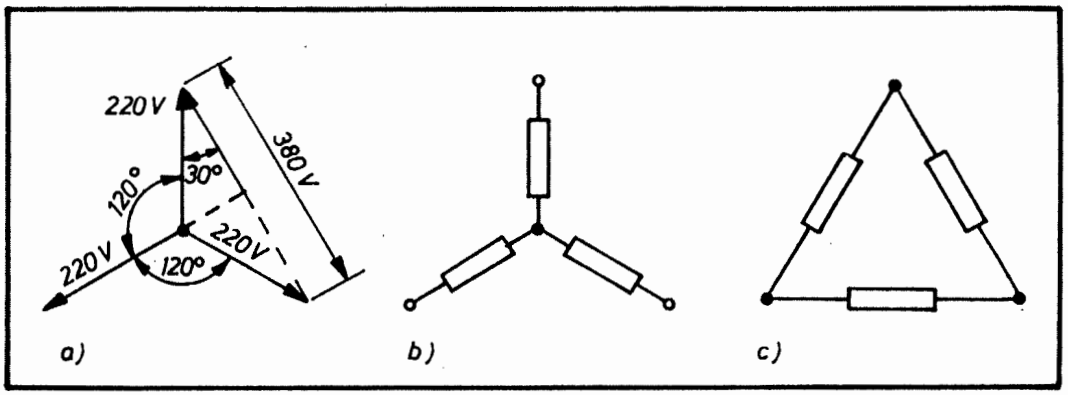


Bild 1.4.3. Darstellung von Drehstrom in Form von Bild 1.4.2c), jedoch unter Fortlassung der Zeitlinie. In b) und c) sind die drei Quellen in Stern oder Dreieck zusammengeschaltet. Dadurch werden an den drei Wicklungen für Motoren entweder 220 oder 380 V wirksam

Dieses Verhältnis 220/380 V hat unter anderem beim Betrieb von Motoren große Bedeutung. Je nachdem, ob die drei Wicklungen in Dreieck- oder Sternschaltung aus den drei Quellen betrieben werden, liegen an ihnen 380 bzw. 220 V, und es kann zwischen dem Anlauf und Betrieb des Motors unterschieden werden (Bild 1.4.3 b) und c)).

1.4.6. FOURIER-Analyse

Es mögen drei Frequenzen vorliegen, die im Verhältnis 1:2:6 stehen. Ohne Berücksichtigung des Phasenwinkels gelte also

$$u_1 = \hat{U}_{01} \sin \omega_0 t \,,$$
$$u_2 = \hat{U}_{02} \sin 2\omega_0 t \,, \tag{9}$$
$$u_6 = \hat{U}_{06} \sin 6\omega_0 t \,.$$

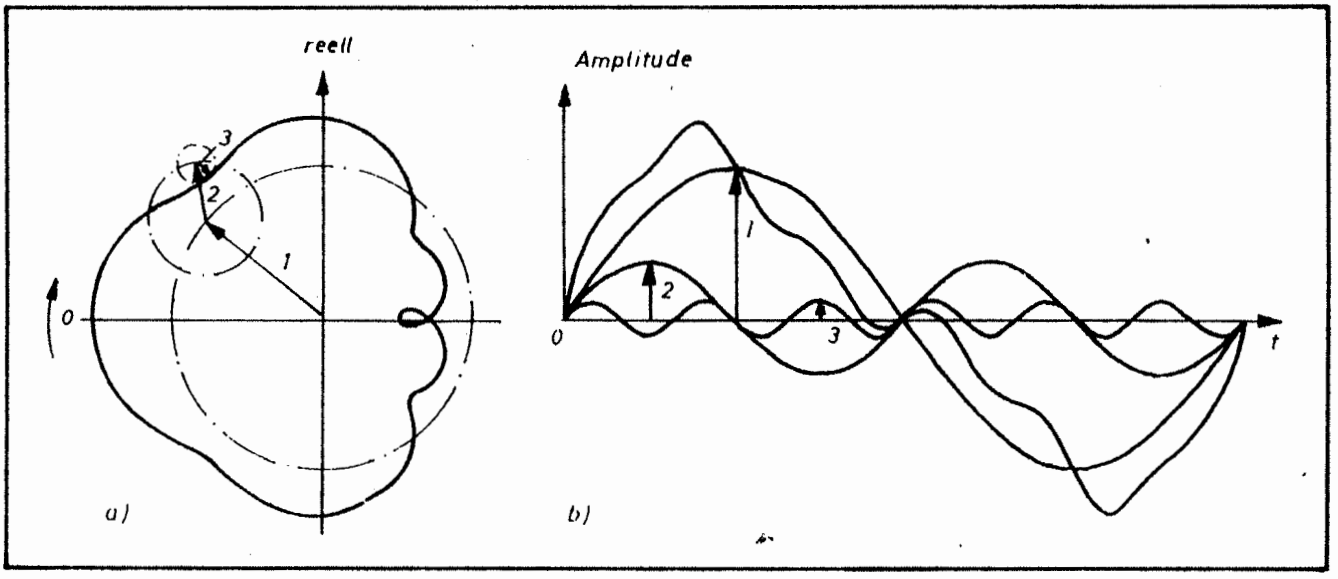


Bild 1.4.4. Zusammensetzung einer Schwingung aus drei Teilschwingungen mit einem Frequenzverhältnis 1:2:6

a) Die aus der Vektoraddition von Bild 1.4.2c) entstehende Grenzkurve weicht stark von der Kreisform ab und wird auch nicht mehr mit gleichbleibender Geschwindigkeit durchlaufen

b) Die Addition der Momentanwerte in der Zeitdarstellung zeigt den sich ergebenden Signalverlauf. Er entspricht dem Abstand zum Mittelpunkt in a)

Im Teilbild a) ist die Gausssche Ebene um −90° gedreht um die Projektion auf Teilbild b) deutlich erkennen zu lassen

Werden diese Kurven vektoriell addiert, so entsteht eine Umlaufkurve bestimmter Form (Bild 1.4.4a)). Ihre Abweichung von der Kreisform kann zugleich als Abweichung von der Sinusform interpretiert werden, die in Bild 1.4.4b) dargestellt ist.

Prinzipiell läßt sich mittels Oberwellen jeder periodische Zeitverlauf einer Schwingung darstellen, und umgekehrt sind zu jedem periodischen Zeitverlauf auch die Grundschwingung und die Oberwellen nach Amplitude und Phase zu berechnen.

Beides leistet die FOURIER-*Analyse*.

Es sei die Zeitfunktion der Form

$$f(t) = a_0 + a_1 \cos \omega t + a_2 \cos 2\omega t + a_3 \cos 3\,\omega t + \ldots + b_1 \sin \omega t + b_2 \sin 2\omega t + b_3 \sin 3\omega t + \ldots \tag{10}$$

mit der Periodendauer

$$T = \frac{2\pi}{\omega} \tag{11}$$

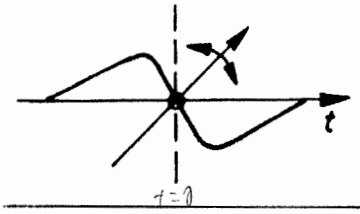
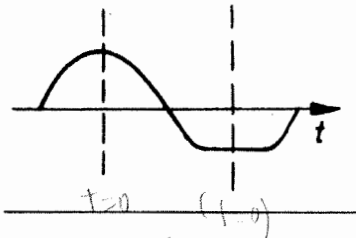
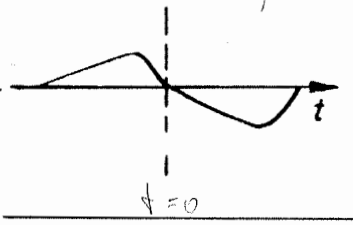
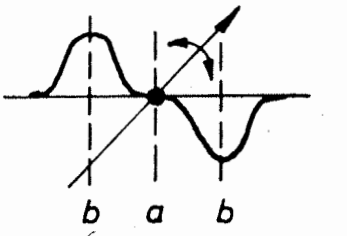
gegeben, dann gilt

$$\left.\begin{aligned} a_0 &= \frac{1}{T}\int_0^T f(t)\,\mathrm{d}t\,, \qquad a_n = \frac{2}{T}\int_0^T f(t)\cos n\omega t\,\mathrm{d}t\,, \\ b_n &= \frac{2}{T}\int_0^T f(t)\sin n\omega t\,\mathrm{d}t\,. \end{aligned}\right\} \tag{12}$$

Auch die Umkehrung von $f(t)$ nach a_n und b_n ist gültig.

Die Rechnungen werden einfacher, wenn in der Zeitfunktion *Symmetrien* existieren. Die vier typischen Fälle enthält Tab. 1.4.1. Ein Beispiel für eine Zeitfunktion und die dazugehörige Spektralfunktion zeigt Bild 1.4.5. Infolge der großen Bedeutung der FOURIER-Analyse gibt es viel-

Tabelle 1.4.1. FOURIER-Zerlegung bei Schwingungen mit Symmetrien

Schwingungsform	Zeitfunktion	FOURIER-Reihe (Spektren)
	drehsymmetrisch $f(t) = -f(-t)$	enthält nur Sinusglieder $b_1 \sin \omega t + b_2 \sin 2\omega t + b_3 \sin 3\omega t + \cdots$
	zweimal spiegelsymmetrisch $f(t) = f(-t)$	enthält nur Cosinusglieder $a_0 + a_1 \cos \omega t + a_2 \cos 2\omega t + a_3 \cos 3\omega t + \cdots$ In dieser Funktion können die ungeraden Glieder sehr klein werden
	halbwellengleich $f(t) = -f(t + \pi)$	enthält nur Glieder ungerader Ordnung $a_1 \cos \omega t + a_3 \cos 3\omega t + a_5 \cos 5\omega t + \ldots + b_1 \sin \omega t + b_3 \sin 3\omega t + b_5 \sin 5\omega t + \ldots$
	dreh- und spiegelsymmetrisch	enthält nur ungerade Glieder, darstellbar in Sinus- oder Cosinusform je nach Wahl des Zeitpunktes Null (a oder b)

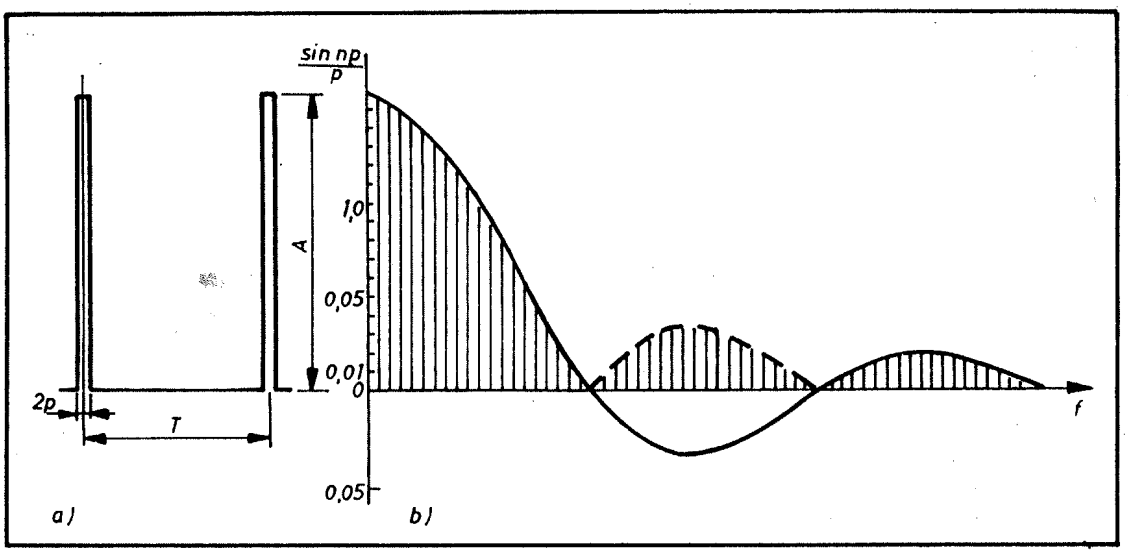


Bild 1.4.5. a) Zeitfunktion und b) Spektraldarstellung einer unsymmetrischen Rechteckschwingung mit $p = 9°$ und $\frac{2A}{\pi} = 1$. Der Ausdruck $\frac{2A}{\pi} \cdot \frac{\sin np}{n}$ bestimmt die Amplitude der jeweiligen Harmonischen. So entsteht die Hüllkurve der Oberwellenamplituden. Bis etwa $n = \pi/(8p)$ besitzen sie konstante Amplitude. In der Umgebung von $n = m\pi/(4p)$ für $m = 1, 2, 3, \ldots$ ist die Amplitude praktisch Null

fältige numerische und graphische Methoden zu ihrer Realisierung. Dabei werden oft nur eine bestimmte Anzahl von Stützpunkten aus der Funktion $f(t)$ gewählt. Die Unterschiede in den numerischen Methoden beziehen sich vor allem auf die algorithmische Bearbeitung der mathematischen Teilausdrücke. Besondere Bedeutung erlangte in der letzten Zeit die schnelle (engl.: fast) FOURIER-Transformation (FFT). Sie läßt mit Rechnern vielfach eine Spektralberechnung während des Schwingungsablaufes zu.

1.4.7. Beschreibung von Schwingungen

Zusammengefaßt sind für nichtsinusförmige (periodische) Schwingungen folgende *Darstellungsformen* möglich:

- *Zeigerdarstellung* in Analogie zu Bild 1.4.4a)
- *Zeitfunktion* gemäß Beispielen von Bild 1.4.4b) und 1.4.5a)
- *Spektraldarstellung* gemäß Beispiel in Bild 1.4.5b)
- *analytischer Ausdruck* für die Zeitfunktion gemäß (10) oder im Beispiel einer Sägezahnschwingung

$$f(t) = t \quad \text{für} \quad -T/2 \ldots T/2 \,. \tag{13}$$

  Dabei könnte allgemein geschrieben werden

$$f(t) = a + bt + ct^2 + dt^3 + \ldots \,. \tag{14}$$

  Dieser Ausdruck wird relativ selten benutzt.
- *Spektralfunktion* mittels der harmonischen Frequenzen entsprechend der FOURIER-Reihe. Hierbei genügen die FOURIER-Koeffizienten a_n und b_n aus (12) .

Die Zeitfunktion ist besonders einfach auf dem Oszillographen sichtbar zu machen. Die Spektraldarstellung erfordert dagegen spezielle Meßgeräte, z.B. Spektrographen, Echtzeitspektralanalysatoren usw. Im Bereich der Optik ist die Spektraldarstellung, z.B. mit Prismen, relativ leicht zu realisieren. Hier besitzt die Zeitfunktion fast keine Bedeutung.

Zeit- und Spektralfunktion können als zwei *Grenzfälle* einer allgemeineren Schwingungsbeschreibung betrachtet werden. Dies läßt sich besonders klar aus der Sicht der Meßtechnik verstehen. Für die Gewinnung von Meßwerten ist eine bestimmte Bandbreite ΔB und eine Meßzeit ΔT erforderlich. Zwischen beiden gilt die KÜPFMÜLLER-*Beziehung*

$$\Delta T \, \Delta B \geqq 2 \,. \tag{15}$$

Je nachdem, welche der beiden Größen gegen Unendlich bzw. Null strebt, entstehen

$$\left.\begin{aligned} &\text{— Zeitfunktionen:} && \Delta T \to 0 \,, && \Delta B \to \infty \\ &\text{— Spektralfunktionen:} && \Delta T \to \infty \,, && \Delta B \to 0 \,. \end{aligned}\right\} \tag{16}$$

Zwischen diesen beiden Extremfällen existieren vielfältige Übergänge. Dies deutet Bild 1.4.6 an. Bei diesen Zwischenfällen können Elementarsignale (Logons) definiert werden, aus denen sich wiederum die Funktionen erzeugen lassen.

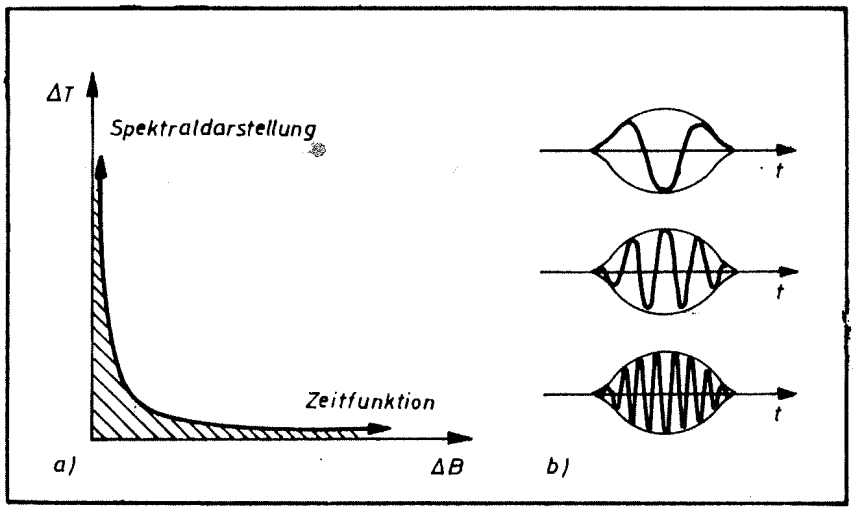


Bild 1.4.6. Aus $\Delta B \cdot \Delta T = 2$ folgt die Hyperbel der sinnvollen Kombination von Meßzeit und Bandbreite. Das schraffierte Gebiet ist prinzipiell nicht erreichbar. Die beiden Sonderfälle Spektraldarstellung und Zeitfunktion entsprechen den Asymptoten. Punkte der erreichbaren Teilebene führen zu sogenannten Elementarsignalen (Logons) begrenzter Meßzeit und Bandbreite. Drei Beispiele von Logons fester Bandbreite und Meßzeit, jedoch verschiedener Frequenz zeigt b)

Von den periodischen Signalen besteht auch die Möglichkeit, zu den *einmaligen Signalen* überzugehen. Dies kann auf zweierlei Art erfolgen:

- Die Funktion existiert nur in einem endlichen Zeitabschnitt, und es wird dann dieser Zeitabschnitt formal periodisch aneinandergereiht.
- Die Periodendauer T wird unendlich gewählt.

Im ersten Fall tritt ein gewisser Fehler auf, dafür können aber die schon geschilderten Methoden angewendet werden. Der zweite Fall führt zum FOURIER-Integral, und die Spektrallinien gehen in eine kontinuierliche Spektraldichteverteilung im Frequenzbereich von $-\infty$ bis $+\infty$ über. Bei der praktisch immer begrenzten Bandbreite und Meßzeit ist dies also nur eine theoretisch wichtige Grenzbetrachtung. Sie hat aber insbesondere Bedeutung für die immer wichtiger werdenden stochastischen Signale. Allgemein wird dieser Grenzfall in der Praxis mittels zweier typischer Signale untersucht: *Einheitssprung* und DIRAC-*Impuls*. Sie sind beide in Bild 1.4.7 dargestellt und besitzen in der Spektraldarstellung im Grenzfall eine konstante Spektraldichte.

Eine weitere Betrachtung von Signalen ist mittels der *Autokorrelation* möglich. Hierbei wird das Signal mit sich selbst bei einer veränderlichen Zeitverschiebung τ

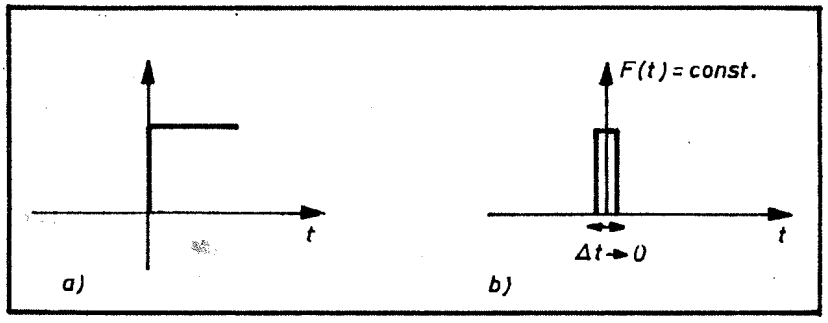


Bild 1.4.7. a) Einheitssprung und b) DIRAC-Impuls bei Beispiel für einmalige Impulse. Sie besitzen ein kontinuierliches Spektrum. Alle Frequenzen sind mit gleicher Amplitudendichte enthalten

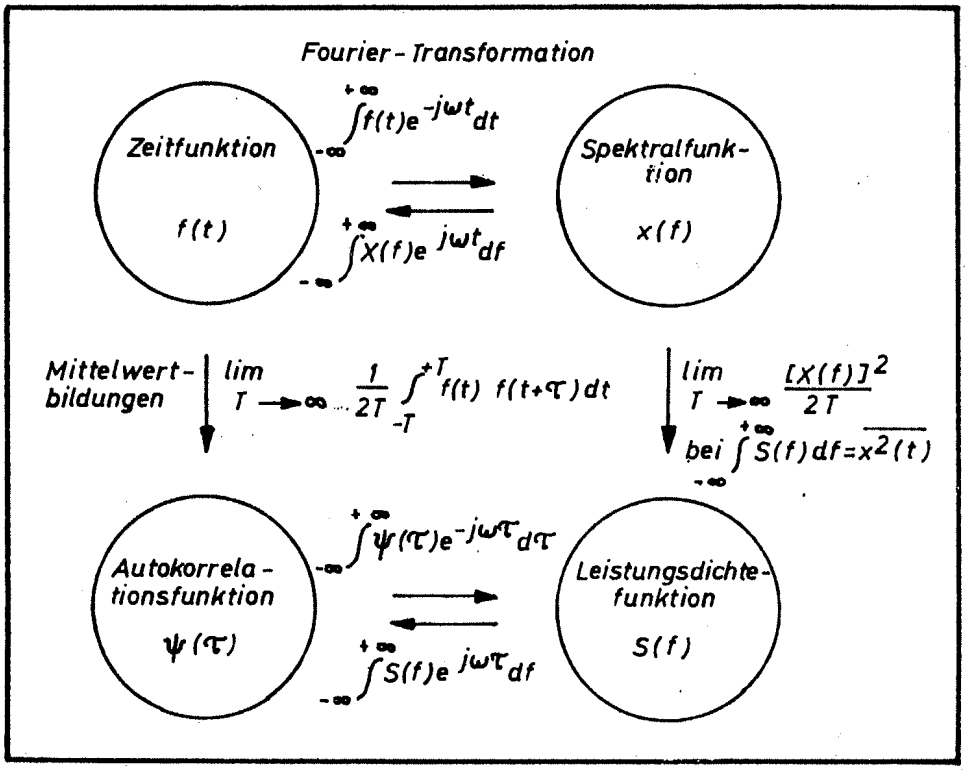


Bild 1.4.8. Zusammenhang von Zeit-, Spektral-, Autokorrelations- und Leistungsdichtefunktion

gefaltet (unter dem Integral multipliziert). So geht die Zeitfunktion $f(t)$ in eine Funktion $f(\tau)$ über. Dies entspricht einer zeitlichen Mittelwertbildung, die deutlich periodische Anteile mit der Zeit τ hervorhebt:

$$\tau = nT \quad \text{für} \quad n = 0, \pm 1; \pm 2; \ldots . \tag{17}$$

In ähnlicher Weise wird auch die *Leistungsdichte* gebildet. Diese Zusammenhänge veranschaulicht Bild 1.4.8. Daraus ist auch ersichtlich, daß bei der Autokorrelation und Leistungsdichtefunktion die Phasenbeziehungen vorlorengehen. Dadurch ist keine Rücktransformation zur Zeit- und Spektralfunktion mehr möglich.

Weiterhin sei noch auf die Zerlegung in orthogonale Rechteckfunktionen, z.B. die WALSH-*Funktionen*, hingewiesen. Sie haben bisher keine besondere praktische Bedeutung erlangt.

1.4.8. Entstehung von Oberwellen und Gleichrichtung

Das mikro-Ohmsche Gesetz ergab die Proportionalität zwischen Strom und Spannung. Zuweilen werden auch Strom und Spannung als Synonyme benutzt, z.B. Wechsel-

stromnetz, wo es eigentlich Wechselspannungsnetz heißen müßte. Dies führt für den Lernenden zu einer oft unklaren Abgrenzung der unterschiedlichen Eigenschaften. Daraus ergibt sich dann das schwierige Verständnis der Phasenverschiebung zwischen Strom und Spannung. Aus diesem Grunde wird hier zunächst der Unterschied mittels nichtlinearer Widerstände eingeführt.

Bei der Erwärmung von Leitern oder Hälbleitern existieren zwei Möglichkeiten zur Widerstandsänderung. Beim Kaltleiter wächst der Widerstand mit steigender Temperatur. Hierzu gehören alle Metalle. Beim Heißleiter dagegen fällt der Widerstand mit steigender Temperatur.

Die Temperaturänderung kann sowohl durch äußere Einflüsse als auch durch die Stromwärme selbst erfolgen. Bei der Selbsterwärmung entstehen Widerstände, deren Größe von der angelegten Spannung bzw. vom hindurchfließenden Strom abhängig sind. Solche spannungs- bzw. stromabhängigen Widerstände können auch auf andere physikalischen Effekten, also nicht nur auf Erwärmung, beruhen. Hierzu sind ausführliche Angaben bei den Halbleitern enthalten.

Eine graphische Darstellung der strom- bzw. spannungsabhängigen Widerstände zeigt Bild 1.4.9, und zwar einmal im Strom-Spannungs-Diagramm und zum anderen bezüglich des Widerstandes selbst.

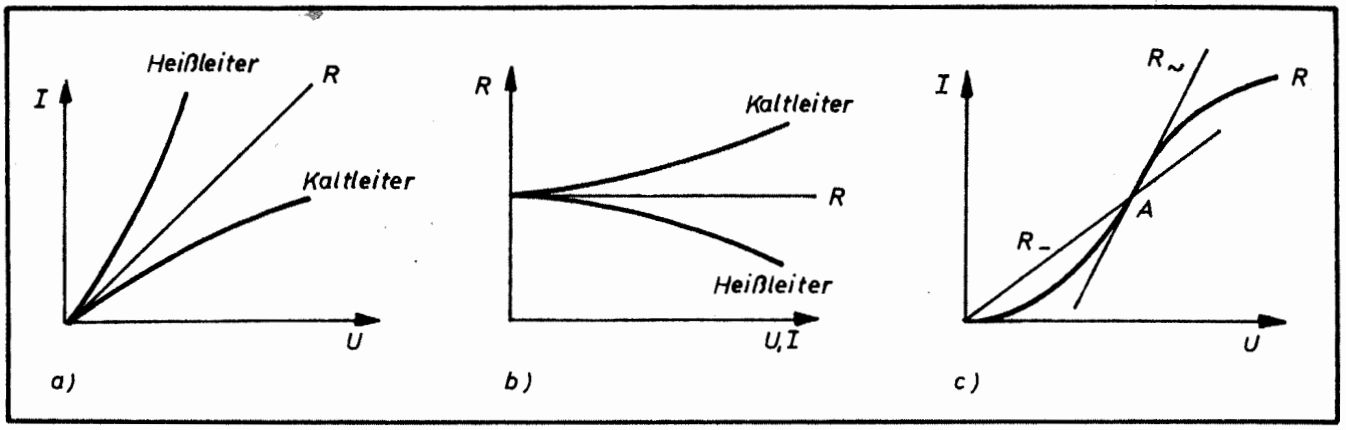


Bild 1.4.9. Heiß- und Kaltleiter als Beispiele strom- und spannungsabhängiger Widerstände a) im $I-U$-Diagramm und b) im $R-U(I)$-Diagramm. Darstellung eines nichtlinearen Widerstandes R. Im Arbeitspunkt ist der (Gleichstrom-) Widerstand durch die Gerade R_- und der differentielle (Wechselstrom-) Widerstand $R_\sim$ durch die Tangente bestimmt (c)

Strom- bzw. spannungsabhängige *Widerstände* heißen *nichtlinear*. Ein nichtlinearer Widerstand werde in einem bestimmten Arbeitspunkt A betrieben (Bild 1.4.9c)). Dann sind zwei typische Widerstandsgrößen zu unterscheiden. Das Stromspannungsverhältnis bestimmt den Gleichstromwiderstand

$$R_- = U/I\,. \tag{18}$$

Die Tangente im Arbeitspunkt ist für kleine überlagerte Wechselströme wirksam und bestimmt daher den *differentiellen Widerstand*

$$R_\sim = \frac{\mathrm{d}U}{\mathrm{d}I}. \tag{19}$$

Von jetzt an werden alle Gleichgrößen mit großen Buchstaben U, I und alle Wechselgrößen mit kleinen Buchstaben bezeichnet, so gilt

$$U + u = R_- I + R_\sim i\,. \tag{20}$$

Infolge der Kennlinienkrümmung ist es sinnvoll, um den Arbeitspunkt eine Reihenentwicklung entsprechend

$$R_{-}I + R_{\sim 1}i + R_{\sim 2}i^2 + \ldots \tag{21}$$

vorzunehmen. Wird der Wechselstrom jetzt harmonisch angenommen,

$$i = \hat{I}_0 \cos \omega t\,, \tag{22}$$

so entsteht nach Einsetzen in (21) und Umformungen mit den trigonometrischen Additionstheoremen

$$U + u = U_{00} + \hat{U}_{01} \cos \omega t + \hat{U}_{02} \cos 2\omega t + \hat{U}_{03} \cos 3\omega t + \ldots\,. \tag{23}$$

Während der Strom durch den Widerstand also rein harmonischen Verlauf zeigt, besitzt die Spannung am Widerstand Oberwellen. Die Koeffizienten $\hat{U}_{0n}$ setzen sich aus den $R_{\sim}$ und R_{-} kompliziert zusammen. Auch der Gleichanteil U_{00} enthält neben dem Gleichstrom I Anteile von $\hat{I}_0$. Bei einer angelegten harmonischen Spannung fließt umgekehrt durch einen nichtlinearen Widerstand ein oberwellenhaltiger Strom. Bezüglich des *Entstehens der Oberwellen* sind zwei *Fälle* zu unterscheiden:

- Die Oberwellen sind *erwünscht*. Dies ist u. a. der Fall bei der Frequenzverdopplung, Erzeugung spezieller Kurvenformen, z.B. Rechteck. Bei der Gleichrichtung interessiert vor allem der entstehende Gleichstromanteil U_{00} in (23).
- *Unerwünschte* Oberwellen. Dies ist überwiegend bei der nachrichtentechnischen Übertragung, z.B. Telefon, Rundfunk und Fernsehen der Fall.

Die Größe des Oberwellenanteils wird durch die *Klirrkoeffizienten*

$$k_n = \frac{\hat{U}_{0n}}{\hat{U}_{01}} \tag{24}$$

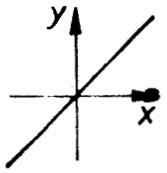
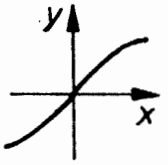
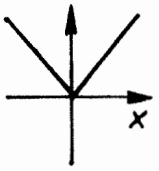
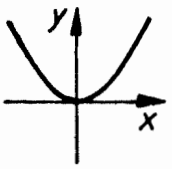
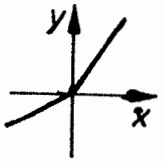
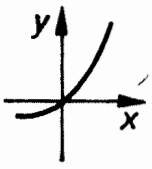
angegeben. Der *Klirrfaktor* ist die quadratische Summe der einzelnen Klirrkoeffizienten:

$$k = \frac{\sqrt{\sum_{n=2}^{\infty} U_{0n}^2}}{\sqrt{\sum_{n=1}^{\infty} U_{0n}^2}} \leqq 100\%\,. \tag{25}$$

Zuweilen wird noch die alte Klirrfaktor-Definition verwendet, bei der im Nenner nur U_{01} steht. Bei ihr kann der Klirrfaktor größer als 100% werden. Bis etwa 10% stimmen beide gut überein. Bei akustischen Übertragungen sind Klirrfaktoren von etwa 1% hörbar. Auf dem Oszillographen sind Werte von etwa 10% an sichtbar. Meßbar sind Klirrfaktoren heute bis zu 10^{-6} herab. Die Verzerrungen können auch als Kennlinien-Abweichung von der Geraden angegeben werden. *Linearitätsfehler* von 10% entsprechen einem Klirrfaktor von etwa 1%. Schließlich läßt sich die Nichtlinearität noch mittels des *Intermodulationsfaktors* bestimmen. Hierbei werden auf den nichtlinearen Widerstand zwei harmonische Frequenzen gegeben, und gemessen wird die sich bildende Differenzfrequenz.

Es gibt einige *wenige typische Kennlinien*, aus denen die wesentlichen Schwingungsänderungen zu folgern sind. Sie sind in Tab. 1.4.2 zusammengestellt. Beim Typ 1 entstehen bei einer harmonischen x- oder y-Schwingung vor allem ungeradzahlige Harmonische. Beim Typ 2 wird bei harmonischer x-Schwingung in y nur eine Gleichkomponente und geradzahlige Harmonische erzeugt. Die Grundwelle wird vollständig unterdrückt. y-Eingang ist für diese Kennlinie nicht definiert. Kennlinientyp 3 ist wieder in beiden Richtungen zu betreiben. Bei harmonischem Eingangsverlauf sind im Ausgang Grundwelle, alle Oberwellen und Gleichkomponente enthalten.

Tabelle 1.4.2. Grundtypen von nichtlinearen Kennlinien
In der Spalte „Anteile" bedeuten: daß bei Sinuserregung — Gleichstrom, ~ die Grundwelle, *g* geradzahlige Harmonische, *u* ungeradzahlige Harmonische in der Ausgangsschwingung enthalten sind

Typ	Knickkennlinie	Gekrümmte Kennlinie	Funktion ...	Anteile — ~ *g u*	Ausgang bei Sinuseingang (vgl. Tab. 1.4.1)
1			$y = k_1x + k_3x^3 + k_5x^5$	0 + 0 +	dreh- und spiegelsymmetrisch
2			$y = k_2x^2 + k_4x^4 + k_6x^6$	+ 0 + 0	Spezialfall von zweimal spiegelsymmetrisch
3			z. B. $y = e^x - 1$	++++	zweimal spiegelsymmetrisch

Typ 1 und 2 sind vorrangig zur *Frequenzvervielfachung* geeignet. Mit Typ 2 und 3 können aus Wechselgrößen Gleichgrößen erzeugt werden. Sie werden deshalb zur Grundlage von *Gleichrichtern* herangezogen. Hierauf wird u. a. im Abschnitt 5.4. noch genauer eingegangen. Werden die Kennlinien in Tab. 1.4.2 durch eine zusätzliche Eingangsgleichgröße aus dem Ursprung verschoben, ergeben sich keine wesentlich neuen Qualitäten.

1.4.9. Messung von Wechselgrößen

Die Wechselgrößen einer harmonischen Schwingung sind durch die Amplitude U_0, die Frequenz ω und die Phase φ vollständig beschrieben. Eine hervorragende Bedeutung für viele Anwendungen haben außerdem noch die Energie bzw. die Leistung des Wechselstromes. Sie stehen im Zusammenhang mit der Amplitude. Zu ihrer Bestimmung sei zunächst noch einmal die Überlagerung von Gleichstrom und Wechselstrom entsprechend Abschnitt 1.4.1. betrachtet. Ein mit Wechselstrom überlagerter Gleichstrom $U(t)$ könnte bezüglich seines *arithmetischen Mittelwertes* durch die Beziehung

$$U_M = \frac{1}{T}\int_0^T U(t)\,dt \tag{26}$$

beschrieben werden. Hierbei bedeutet T die Zeit, über die die Mittelung erfolgt.

In ähnlicher Weise kann auch bei einer reinen Wechselgröße ein *absoluter Mittelwert* eingeführt werden:

$$U_A = \frac{1}{T}\int_0^T |U(t)|\,dt\,. \tag{27}$$

Jetzt ist T die Periode des beliebigen, auch nicht harmonischen Zeitverlaufs $U(t)$ ohne Gleichanteil. Gl. (27) ist so interpretierbar, daß die negative Halbwelle als ebenfalls positive betrachtet und danach der arithmetische Mittelwert gebildet wird. Bei Wechselströmen mit gleichen positiven und negativen Halbwellen genügt es dementsprechend, Gl. (26) für eine halbe Periodenlänge zu benutzen,

$$U_A = U_M = \frac{2}{T}\int_0^{T/2} U(t)\,dt\,. \tag{28}$$

Anschaulich läßt sich diese Beziehung so interpretieren, daß der Inhalt unter dem Schwingungsverlauf in ein *flächengleiches Rechteck* mit der Höhe U_M verwandelt wird (Bild 1.4.10a)).

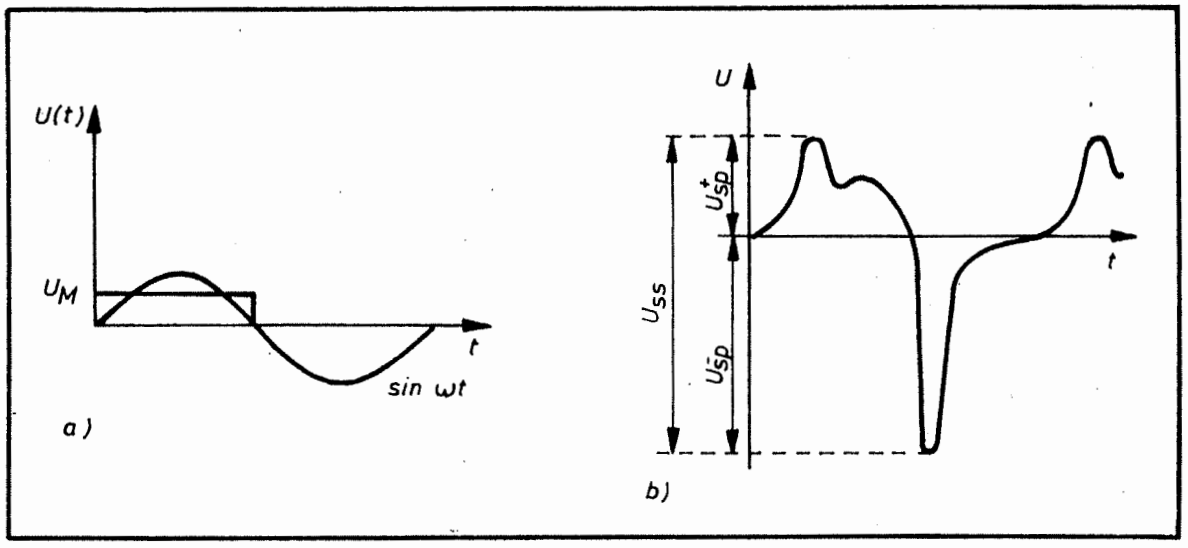


Bild 1.4.10. Zur Definition von Meßwerten bei Wechselspannungen

a) Arithmetischer Mittelwert
b) Die drei Spitzenspannungen

Für eine harmonische Schwingung errechnet sich der arithmetische Mittelwert zu

$$U_{M,\sin} = \frac{2}{\pi}\,\hat{U}_0 \approx 0{,}635\,\hat{U}_0\,. \tag{29}$$

Für den Zusammenhang zwischen der *Leistung* und der Spannung gilt bei Gleichstrom

$$P = \frac{U^2}{R}. \tag{30}$$

Für Wechselgrößen muß daher ein *quadratischer Mittelwert* gewählt werden:

$$U_{\text{eff}}^2 = \frac{1}{T}\int_0^T U^2(t)\,\mathrm{d}t\,. \tag{31}$$

An einem Widerstand erzeugt hiernach die Wechselspannung $U(t)$ eine gleichgroße Wärmewirkung wie eine Gleichspannung von der Größe U_{eff}. In diesem Sinne ist die Bezeichnung „*effektiv*" zu verstehen. Für den Sonderfall einer harmonischen Schwingung folgt

$$U_{\text{eff, sin}} = \frac{1}{\sqrt{2}}\hat{U}_0 \approx 0{,}707\hat{U}_0\,. \tag{32}$$

Für eine allgemeine Schwingung gemäß Bild 1.4.10b) können weiterhin ein *positiver*, ein *negativer* und ein *Spitzen-Spitzenwert* mit

$$U_{\text{ss}} = U_{\text{sp}}^{+} + U_{\text{sp}}^{-} \tag{33}$$

definiert werden. Für Sinusschwingungen gilt

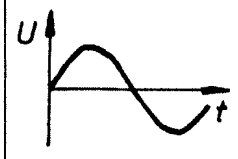
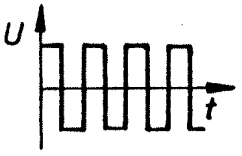
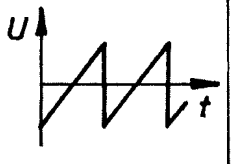
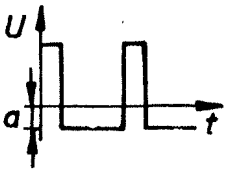
$$\hat{U} = U_{\text{sp}}^{+} = U_{\text{sp}}^{-} = U_{\text{ss}}/2\,. \tag{34}$$

Es gibt nun Schaltungen, die nur auf eine Art dieser Möglichkeiten ansprechen. Aus diesem Grunde ist es wichtig, von einer Größe zur anderen *Umrechnungsfaktoren* zu besitzen. Sie hängen z.T. erheblich von der Kurvenform des Signals ab. Als *Formfaktor* s wird das Verhältnis aus Effektivwert und absolutem arithmetischem Mittelwert bezeichnet,

$$s = U_{\text{eff}}/U_{\text{A}} \tag{35}$$

Für eine Sinusschwingung beträgt s rund 1,1107 bzw. sein reziproker Wert 0,9003. In der Regel erfolgt bei Meßinstrumenten eine Eichung mit Sinusschwingungen. Deshalb

Tabelle 1.4.3. Zum Kurveneinfluß bei Wechselspannungen mit $U_{\text{ss}} = 2$

						
					$a = 0{,}2$	$a \ll 2$
U_{eff}	$\frac{1}{\sqrt{2}} \approx 0{,}707$	1	$\sqrt{\frac{2}{3}} \approx 0{,}816$	$\sqrt{a(2-a)}$	0,6	$\approx \sqrt{2a}$
U_{M}	$\frac{2}{\pi} \approx 0{,}635$	1	$\frac{1}{2}$	$a(2-a)$	0,36	$\approx 2a$
U_{sp}^{+}	1	1	1	$2-a$	1,8	≈ 2
U_{sp}^{-}	1	1	1	a	0,2	a
$s = \frac{U_{\text{eff}}}{U_{\text{M}}}$	$\approx 1{,}11$	1	$\approx 1{,}64$	$\approx \frac{1}{\sqrt{a(2-a)}}$	$\approx 1{,}64$	$\approx \frac{1}{\sqrt{2a}}$
$q = \frac{s}{s_{\text{sinus}}}$	1	$\approx 0{,}9$	$\approx 1{,}48$	$\approx \frac{0{,}9}{\sqrt{a(2-a)}}$	$\approx 1{,}84$	$\approx \frac{0{,}64}{\sqrt{a}}$

wird oft der *relative Formfaktor*

$$q = s/s_{\text{sinus}} \tag{36}$$

verwendet. Beide Werte sind in Tab. 1.4.3 für einige ausgewählte Funktionen dargestellt.

1.5. Grundbauelemente

1.5.1. Definition

Entsprechend den drei Arten des elektromagnetischen Feldes lassen sich auch *drei Bauelemente* definieren. Ihre Größen hängen nur von den geometrischen Abmessungen und der Materialkonstante ab (vgl. Tab. 1.1.3). Die Definition der Bauelemente folgt aus dem

$$\left.\begin{aligned} &\text{stationären Strom:} && R = \frac{U}{I}, \\ &\text{elektrostatischen Feld:} && C = \frac{Q}{U}, \\ &\text{magnetostatischen Feld:} && P = \frac{\Theta}{\Phi}. \end{aligned}\right\} \tag{1}$$

Der stationäre Strom definiert also den Ohmschen Widerstand, das elektrostatische Feld die Kapazität und das magnetostatische Feld den magnetischen Widerstand.

Um die so definierten Bauelemente vergleichen zu können, sind alle in ihnen enthaltenen Größen auf Strom und Spannung zu beziehen. Dabei bleibt lediglich die Gleichung für den Ohmschen Widerstand erhalten. Die *Kapazität* wird gewonnen, indem die Ladung auf den Strom zurückgeführt wird. Bei stationärem Strom gilt

$$I = \frac{ne_0}{t} = \frac{Q}{t}. \tag{2}$$

Für einen zeitlich veränderlichen Strom entsprechend

$$\boxed{I = \frac{\mathrm{d}Q}{\mathrm{d}t}} \tag{3}$$

und (1) gilt also

$$\boxed{U = \frac{1}{C}\int I\,\mathrm{d}t} \tag{4}$$

oder

$$I = C\frac{\mathrm{d}U}{\mathrm{d}t}. \tag{5}$$

Das Symbol der Kapazität zeigt Bild 1.5.1a) und b). Ihre Einheit ist das Farad (F).

Für den *magnetischen Widerstand* besteht zunächst das Durchflutungsgesetz

$$\Theta = nI\,. \tag{6}$$

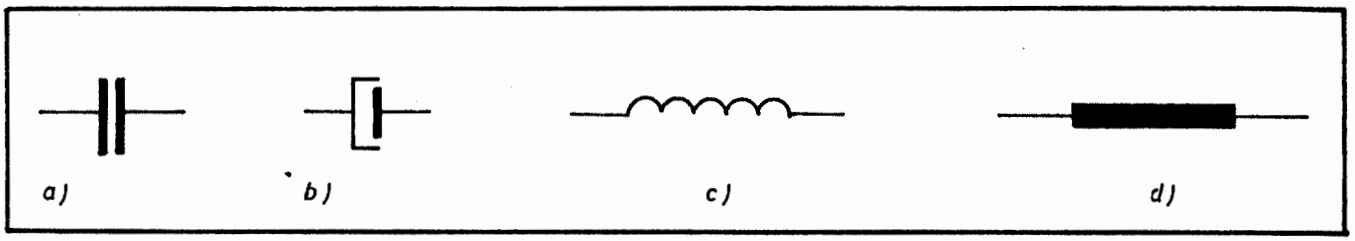


Bild 1.5.1. Schaltsymbole für kapazitive und induktive Bauelemente:
a) Kondensator, b) Elektrolytkondensator, c) Spule mit genau gegebener Induktivität, d) Spule allgemein, insbesondere mit Eisenkern

Dabei sei n die *Windungszahl* um den *magnetischen Kreis*. Außerdem gilt das sogenannte *Induktionsgesetz*, nach dem jede Flußänderung eine Spannung an den Enden der Spule hervorruft:

$$\boxed{U = -n\frac{\mathrm{d}\Phi}{\mathrm{d}t}}. \tag{7}$$

Diese *Flußänderung* kann drei verschiedene *Ursachen* haben:

- Die *Spule bewegt sich im Magnetfeld* so, daß ein $\mathrm{d}\Phi/\mathrm{d}t$ auftritt. Dies ist bei der Dynamomaschine und bei Magnetköpfen der Fall.
- Das *Feld stammt von einer zweiten Spule*, und dort wird der Strom geändert. Dann wird von *Gegeninduktivität* gesprochen. Dieser Fall wird im Abschnitt 1.6.1. behandelt.
- Die Feldänderung wird *von der Spule selbst hervorgerufen*. Dann wird von *Selbstinduktion* gesprochen. Dies wird jetzt weiterbehandelt.

Aus (1), (6) und (7) folgt dann

$$I = -\frac{P}{n^2}\int U\,\mathrm{d}t \tag{8}$$

oder die LENZ*sche Regel*

$$\boxed{U = -L\frac{\mathrm{d}I}{\mathrm{d}t}}. \tag{9}$$

Darin bedeutet L die *Induktivität*. Sie enthält die Windungszahl und die geometrischen Abmessungen sowie die Materialkonstante des magnetischen Kreises:

$$\boxed{L = \frac{n^2}{P}}. \tag{10}$$

Der magnetische Widerstand P ist also das eigentliche magnetische Bauelement und L das dazu gehörende elektrische. Die Kopplung zwischen beiden Bauelementen erfolgt durch die Wicklung. Das Symbol der Induktivität zeigt Bild 1.5.1c) und d). Ihre Einheit ist das Henry (H).

1.5.2. Einschaltvorgänge

Entsprechend Bild.1.5.2a) lade eine Spannungsquelle U_0 über einen Widerstand R einen Kondensator C auf. Am Widerstand fällt dann je nach dem Strom die Spannung

$$U_R = I_C R \tag{11}$$

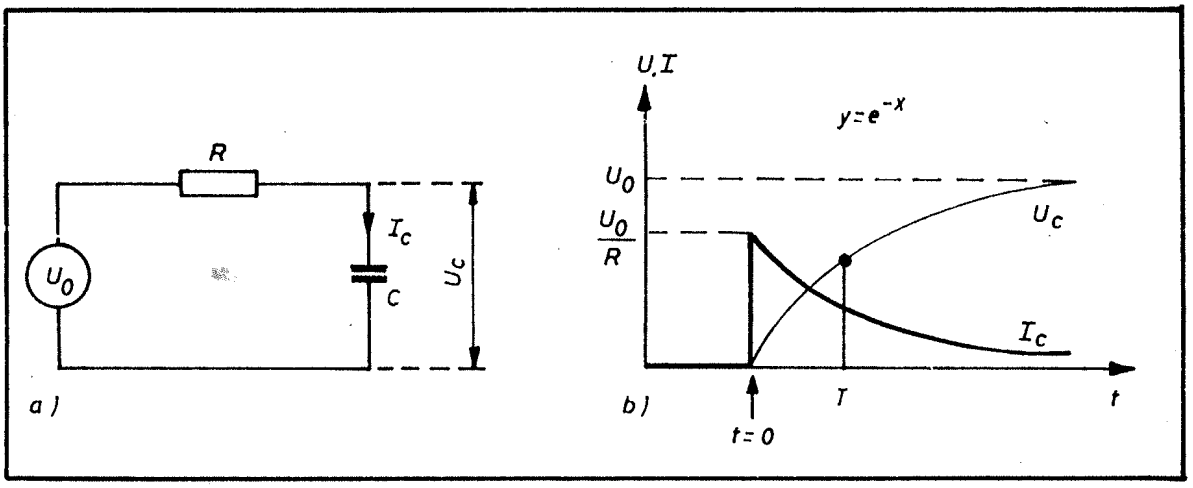


Bild 1.5.2. a) Schaltung zur Aufladung eines Kondensators über einen Widerstand, b) Spannungs- (U_C) und Stromverlauf (I_C) bei Aufladung

ab, und die Spannung am Kondensator beträgt

$$U_C = U_0 - I_C R \,. \tag{12}$$

Andererseits gilt für den Kondensator Gl. (5). Daher folgt die Differentialgleichung

$$U_C = U_0 - RC\frac{\mathrm{d}U_0}{\mathrm{d}t}. \tag{13}$$

Sie hat die Lösung

$$\boxed{U_C = U_0\left(1 - \mathrm{e}^{-\frac{t}{RC}}\right)}\,. \tag{14}$$

Es sei angenommen, daß für $t = 0$ die Spannung eingeschaltet werde. Dann gilt zu diesem Zeitpunkt am Kondensator $U_C = 0$. Nach sehr langer Zeit gilt $U_C = U_0$. Der Übergang zwischen beiden Größen erfolgt gemäß der Exponentialfunktion (14) und ist in Bild 1.5.2b) dargestellt. Die typische Größe ist die *Zeitkonstante*

$$\boxed{T = RC}\,. \tag{15}$$

Mit ihr läßt sich Gl. (14) normieren. Der Strom durch den Widerstand ist proportional der Spannung an ihm,

$$I_C = I_R\frac{U_0 - U_C}{R} = \frac{U_0}{R}\,\mathrm{e}^{-t/T}\,. \tag{16}$$

Den zugehörigen Verlauf zeigt ebenfalls Bild 1.5.2b). Er beginnt mit dem plötzlichen Anstieg auf $I_C = U_0/R$, um dann exponentiell abzuklingen. Aus dem Vergleich von Strom- und Spannungsverlauf folgt, daß die Spannung am Kondensator dann am größten ist, wenn der ihm zufließende Ladungsstrom am kleinsten ist, und umgekehrt.

Sollen die Kurven nach Bild 1.5.2b) experimentell bestimmt werden, so ist zur besseren Korrektur der Fehler eine Darstellung in halblogarithmischem Papier vorteilhaft. Die zugehörige Gleichung aus Gl. (16) lautet dann z.B.

$$\ln\frac{RI_C}{U_0} = -\frac{t}{T} \tag{17}$$

und entspricht damit einer Geraden.

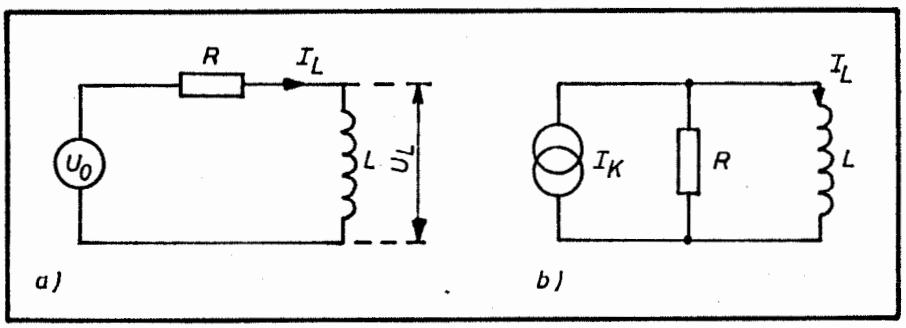


Bild 1.5.3. Einschaltvorgang bei einer Induktivität a) in Spannungsdarstellung und b) in Stromersatzschaltung

Für den Einschaltvorgang mit einer Induktivität gilt Bild 1.5.3a). In Analogie zu (12) und mit (9) folgt jetzt die Differentialgleichung

$$U_0 = I_L R - L\frac{\mathrm{d}I_L}{\mathrm{d}t}. \tag{18}$$

Mit dem Kurzschlußstrom

$$I_K = \frac{U_0}{R} \tag{19}$$

kann sie in die Form

$$I_K = I_L - \frac{L}{R}\frac{\mathrm{d}I_L}{\mathrm{d}t} \tag{20}$$

übergeführt werden und legt dann das duale Ersatzschaltbild 1.5.3b) nahe. Durch Vertauschung von Strom und Spannung entsprechend (13) und (20) gilt dann wieder Bild 1.5.2b). Als *Zeitkonstante* ergibt sich

$$\boxed{T = \frac{L}{R}}. \tag{21}$$

Die bisherige Dualitätsbetrachtung kann daher erweitert werden auf:

U	R	L	Parallel-	Dreieck-	
↕	↕	↕	↕	↕	Schaltung
I	G	C	Reihen-	Stern-	

. (22)

Zur Veranschaulichung des Aufladevorganges von einem Kondensator sei ein Fluß angenommen, der rechts und links je einen breiten Landungssteg mit einem schmalen Anmarschweg besitzt. Marschieren auf diesen Wegen Menschen zum oder vom Landungssteg, so entspricht ihre passierende Anzahl dem Strom und die Differenz der sich auf dem Landungssteg befindenden Menschen der Spannung am Kondensator. Erst wenn ein Landungssteg voll oder leer ist, können keine weiteren Menschen auf dem Anmarschweg zum Kondensator marschieren.

In ähnlicher Weise kann eine Induktivität durch einen sandigen Weg erklärt werden, der durch das Begehen festgetreten wird und damit schneller zu begehen ist. Wenn er nicht begangen wird, verschlechtert er sich wieder. Ganz zu Beginn (Stromeinschaltung) ist er nur sehr mühevoll und damit extrem langsam begehbar (geringer Strom).

1.5.3. Rechteckströme

Statt der einmaligen Einschaltung von Strömen und Spannungen sei jetzt ein periodisches Ein- und Ausschalten betrachtet. Dies führt zur Rechteckspannung bzw. zum

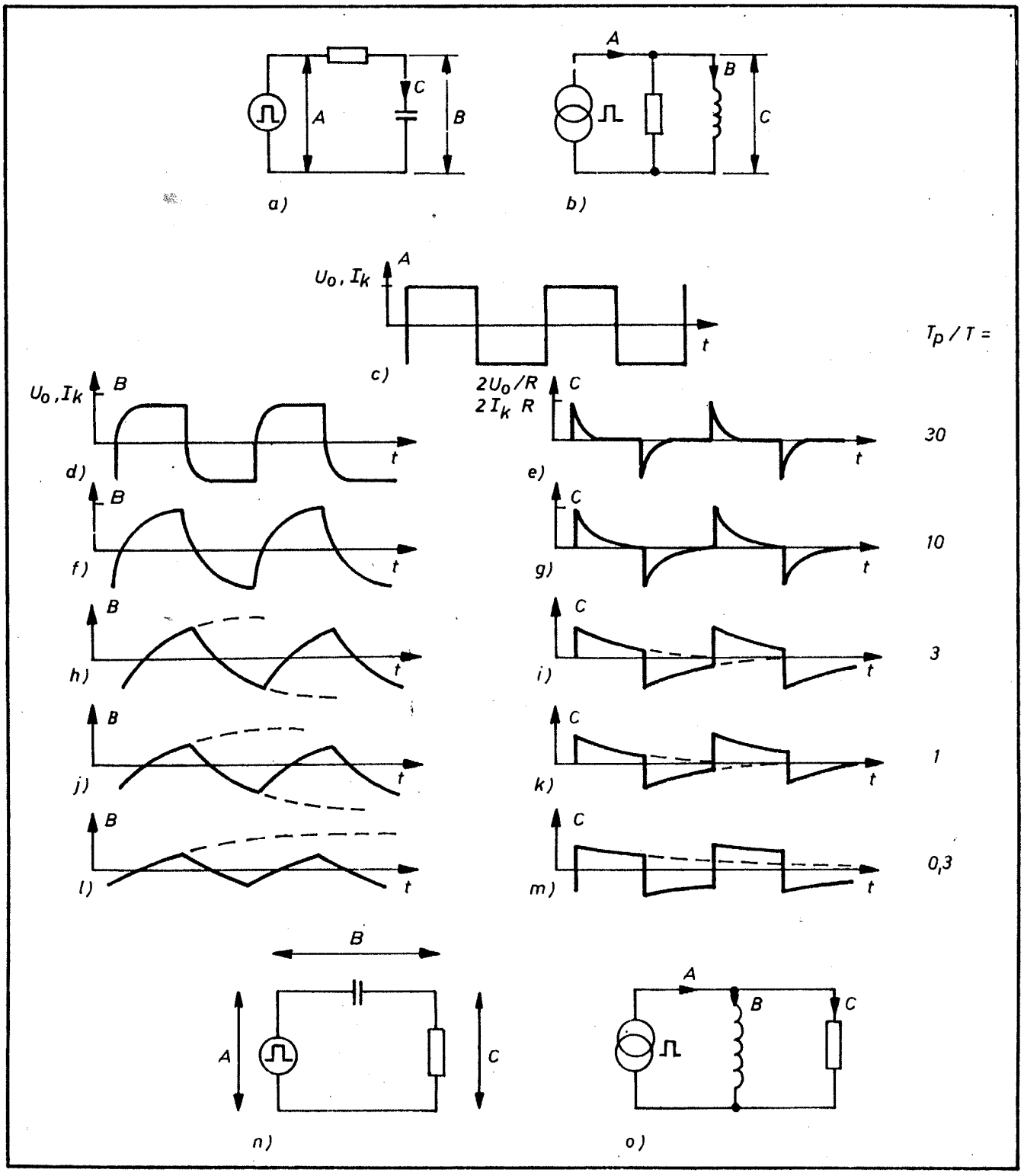


Bild 1.5.4. Eine Rechteckschwingung wirkt auf eine Kapazität (a) bzw. Induktivität (b) ein. Die Größen A, B, C sind durch die Dualität zugeordnet. Je nach dem Verhältnis von Periodendauer T_p der Rechteckschwingung (c) (sie bleibt im Bild konstant) und der Zeitkonstante $T = RC$ bzw. $T = L/R$ ergibt sich an der Kapazität bzw. Induktivität ein unterschiedlicher Schwingungsverlauf B und C. Unten im Bild sind als n) und o) noch reziproke Schaltungen zu a) und b) dargestellt.

Rechteckstrom. Entsprechend Bild 1.5.4 wird je nach dem Verhältnis aus Periodendauer T_p des Ein- und Ausschaltens und der Zeitkonstante (15) bzw. (21) ein mehr oder weniger kleiner *Ausschnitt aus der Exponentialkurve* von Bild 1.5.2b) wirksam. Für einige $T_p : T$-Verhältnisse zeigt dies Bild 1.5.4. Unter Ausnutzung der Dualitäten sind in a) und b) die beiden möglichen Schaltungen dargestellt. A bedeutet hierin die

Rechteckspannung bzw. den Rechteckstrom entsprechend der Kurve c). Zunächst werde angenommen, daß die Umschaltung relativ langsam entsprechend der Zeitkonstante erfolge, dann ist in f) auch recht deutlich der zugehörige Spannungseinschwingverlauf am Kondensator (B) bzw. für den Strom bei der Induktivität zu erkennen. g) zeigt den zugehörigen Kondensatorstrom bzw. die entsprechende Spannung an der Induktivität (C). Auffallend sind hier die plötzlichen, stark *ausgeprägten Spitzen*. Sie sind in Bild 1.5.2b) ebenfalls beim Einschalten für $t = 0$ zu sehen. Infolge der Umschaltung von $+U$ auf $-U$ bzw. $+I_k$ auf $-I_k$ besitzen sie hier die Höhe $2U_0/R$ bzw. $2I_kR$.

Für eine noch kleinere Zeitkonstante T zeigt d) die allmähliche Annäherung an die ursprüngliche Rechteckschwingung und e) den Übergang zu extrem kurzen Impulsen.

Mit größer werdender Zeitkonstante wird ein immer kleinerer Ausschnitt aus der Exponentialkurve ausgenutzt. Für die Spannung am Kondensator zeigt die Folge c), d), f), h), j), l), wie so aus der Rechteckschwingung eine *Dreieckschwingung* entsteht, wobei die Amplitude ständig abnimmt. In h) und j) ist noch angedeutet, wie die Exponentialfunktion weiterlaufen würde, wenn keine Umschaltung erfolgt.

Umgekehrt wandeln sich mit zunehmender Zeitkonstante die Spitzenimpulse von e) über g), i), k) und m) allmählich zur Rechteckschwingung mit voller Amplitude. Auch hierbei ist in i) und k) die Exponentialkurve für den Fall angedeutet, daß keine Umschaltung erfolgt. Auffallend ist weiter, daß in allen Kurven der Sprung die gleiche Höhe, nämlich $2U_0/R$ bzw. $2I_kR$ beibehält. Der geringe Abfall in Kurve m) im Vergleich zur Kurve c) wird mit *Dachschräge* bezeichnet.

Werden die zu Bild 1.5.4a) und b) reziproken Schaltungen gebildet, so entstehen Bild 1.5.4n) und o). Hierbei wird mit den Ohmschen Widerständen der Strom in Spannung bzw. umgekehrt umgewandelt. So entstehen in jeder Schaltung drei Schwingungsgrößen gleichen Bezugs — Spannung in a), Strom in b) —, für die gilt

$$A \approx B + C\,. \tag{23}$$

1.5.4. Harmonische Ströme

Die harmonische Spannung

$$u = \hat{U}_0 \sin \omega t \tag{24}$$

liege an einem Kondensator. Wegen (5) fließt dann der Strom

$$i = C\omega \hat{U}_0 \cos \omega t\,. \tag{25}$$

Außer dem Amplitudenfaktor $C\omega$ sind also Strom- und Spannung am Kondensator genau um 90° verschoben (vgl. Bild 1.5.5a)). Immer wenn der Strom durch den Kondensator maximal ist, liegt an ihm die Spannung Null. Ist dagegen die Spannung an ihm maximal, so fließt kein Lade- bzw. Entladestrom. Immer dann, wenn der Strom zu fließen beginnt, lädt sich also der Kondensator auf. Wenn der Strom dagegen durch Null geht, ändert er seine Richtung, und hierdurch fällt die Spannung am Kondensator weiter ab. Aus (25) folgt weiter, daß der Strom durch den Kondensator bei konstanter Spannung mit der Frequenz zunimmt. Werden also mehrere Frequenzen gleichzeitig an den Kondensator angelegt, so überwiegen im Strom vergleichsweise die höheren Frequenzen zu ihrem Anteil in der Spannung. *Der Klirrfaktor ist bei einem Kondensator im Strom also stets größer als der Klirrfaktor der Spannung, die anliegt.* Es muß jedoch betont werden, daß ein idealer Kondensator selbst keinen Klirrfaktor erzeugt. Er ist ein lineares Bauelement.

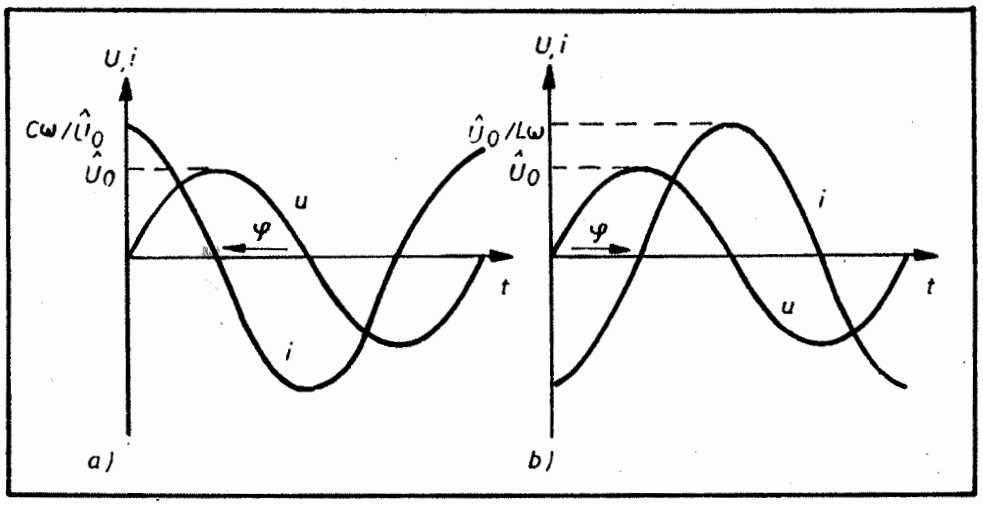


Bild 1.5.5. Phasenverschiebung zwischen Strom und Spannung bei a) einer Kapazität und b) einer Induktivität

Wird eine harmonische Spannung gemäß (24) an eine Induktivität angelegt, so gilt mit (9)

$$i = \frac{1}{L} \int u \, dt \,, \tag{26}$$

und es folgt

$$i = -\frac{\hat{U}_0}{L\omega} \cos \omega t \,. \tag{27}$$

Es sind also wiederum *Strom und Spannung um* 90° *verschoben*, jedoch liegt in diesem Fall zunächst die Spannung an der Induktivität, und der Strom folgt (Bild 1.5.5b)). Außerdem nimmt der Widerstand der Induktivität im Gegensatz zur Kapazität nach hohen Frequenzen hin zu. Sowohl bezüglich der Richtung der Phasenverschiebung als auch bezüglich des Frequenzverhaltens zeigt sich auch hier, daß Induktivität und Kapazität dual sind (vgl. (22)).

1.5.5. Technische Kapazitäten

1.5.5.1. Allgemeines

Technische Kapazitäten werden in einem sehr großen Bereich von pF bis zu einigen F und in einem großen Spannungsbereich von V bis zu mehreren kV benötigt. Dies und weitere Kennwerte haben im Laufe der Zeit eine sehr große Vielfalt von Kondensatoren entstehen lassen. Wenn von den heute unbedeutenden Anwendungen der Vakuum- und Luftkondensatoren abgesehen wird, können *drei Hauptgruppen* unterschieden werden:

- keramische Kondensatoren
- Folienkondensatoren
- Elektrolytkondensatoren.

Darüber hinaus werden später noch die *integrierten Kapazitäten* und die *gedruckten Kapazitäten* zu besprechen sein.

Generell gibt es keinen idealen Kondensator. Jeder realisierte Kondensator besitzt einen *Isolationswiderstand*, der überwiegend durch die Eigenschaften seines Dielektri-

kums bestimmt ist. Für beide gelten die Gleichungen

$$C = \varepsilon_r \frac{A}{d}, \qquad R = \varrho \frac{d}{A} \tag{28}$$

näherungsweise. Dementsprechend ist das Produkt mit Maßeinheit einer Zeitkonstanten weitgehend unabhängig von den jeweiligen konkreten Abmessungen,

$$T = CR \approx \varepsilon_r \cdot \varrho \,. \tag{29}$$

Darüber hinaus besitzt jeder Kondensator eine gewisse Induktivität ESL (= engl.: equivalent series inductance) und einen Serienwiderstand ESR (= engl.: equivalent series resistance). So entsteht das Ersatzschaltbild gemäß Bild 1.5.6a). Wie aus Kap. 2. genauer ersichtlich ist, ergibt sich daraus eine frequenzabhängige Impedanz gemäß Bild 1.5.6b).

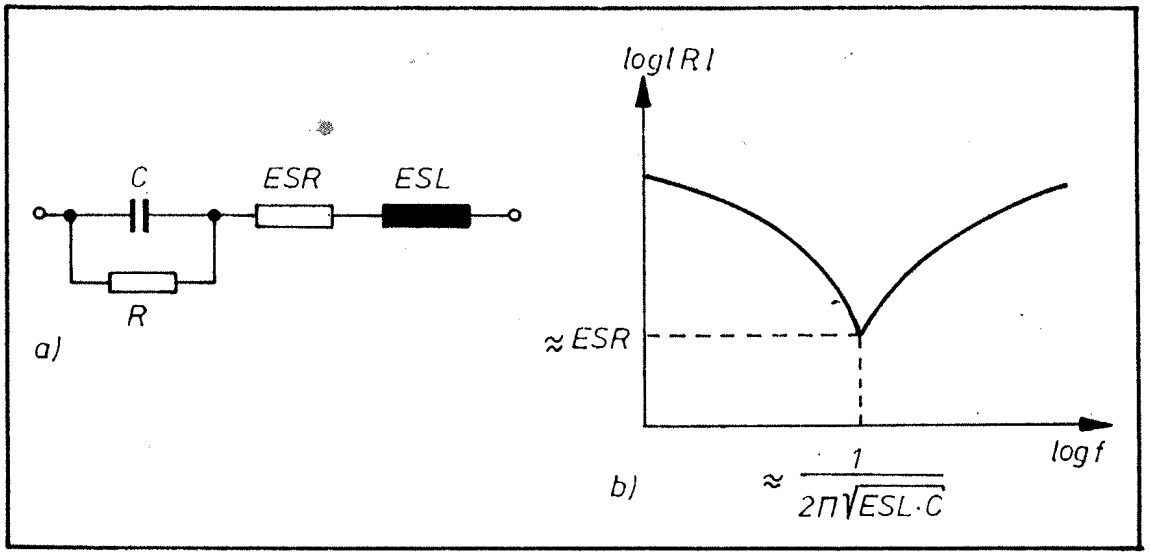


Bild 1.5.6. a) Ersatzschaltbild, b) Frequenzgang der Impedanz eines Kondensators

Für hohe Frequenzen sind die Verluste eines Kondensators besonders wichtig. Sie werden meist durch den *Verlustwinkel* δ gemäß

$$\tan \delta = \frac{1}{\omega R_v C} \approx \delta \tag{30}$$

angegeben. Dabei ist R_v ein Verlustwiderstand bei der Meßfrequenz. Er setzt sich aus mehreren Komponenten zusammen, die teilweise nicht unmittelbar aus der Ersatzschaltung von Bild 1.5.6 ersichtlich sind, wie z.B. Verluste beim ständigen Umpolen der Feldstärke im Dielektrikum.

Da ein Kondensator oft in frequenzbestimmenden Schaltungen angewendet wird, werden an die Stabilität seiner Kapazität sehr hohe Anforderungen gestellt. Deshalb besitzen unter anderem der TK (Temperaturkoeffizient), die Spannungsabhängigkeit und die *Langzeitkonstanz* der Kapazität große Bedeutung.

1.5.5.2. Keramische Kondensatoren

Keramische Kondensatoren gehen von Oxiden wie $BaCO_3$, TiO_2, ZrO_2, SnO_2, $MgCO_3$, $SrCO_3$ u. a. aus. Die ursprüngliche Masse ist knet- und gießbar. Nach der Formgebung wird sie bei ca. 1400 °C gesintert. Zuweilen ist dem Sintern noch ein Glühbrand um

900 °C vorgelagert. Infolge dieser Prozesse betragen die Dicken des kompakten Dielektrikums 0,1 bis 1 mm. Die Elektroden werden durch Spritzen, Drucken, Pinseln usw. aufgebracht und bei ca. 800 °C fest aufgebrannt. Durch Abtragen von Teilen der Belage lassen sich die Kapazitäten eng tolerieren. Die endgültigen Anschlüsse werden aufgelötet, und abschließend wird der Kondensator mit einem Lack geschützt. Die üblichen Gestalten der keramischen Kondensatoren sind die *Röhrchen-, Scheiben- oder Pillenformen.* Eine besondere und neue Form der Keramikkondensatoren sind die *Vielschicht-Keramik-Kondensatoren.* Hierbei wird die ursprüngliche Masse auf eine ebene Unterlage gegossen und vorgetrocknet. Die entstehenden Folien werden mit einem Metall, z.B. Palladium, metallisiert, dessen Schmelzpunkt oberhalb der Sintertemperatur liegt. Die einzelnen Folien werden dann übereinandergestapelt und gemeinsam gesintert. Danach werden die Teilbelage seitlich wechselseitig verlötet. Mit dieser Technologie lassen sich weitaus höhere spezifische Kapazitäten (auf das Volumen bezogen) erreichen.

Bei den *keramischen Massen* werden NDK- und HDK-Materialien (niedere bzw. hohe Dielektrizitätskonstante) unterschieden; siehe auch Tab. 1.3.2. Sie haben sehr verschiedene Eigenschaften. Dies geht beispielhaft aus den Kurvenscharen des Bildes 1.5.7 hervor. Das weitaus höhere ε der HDK wird also durch starke Temperaturabhängigkeit, Spannungsabhängigkeit und auch höhere Verluste erkauft. Bei den NDK ist der TK in Grenzen positiv und negativ einstellbar. Der *Sperrschichtkondensator* ist schließlich eine spezielle Form des Keramikkondensators. Hierbei besteht das Dielektrikum nur aus einer dünnen reoxydierten Schicht. Sie sind dadurch aber nur bis zu einer Spannung von etwa 16 V geeignet.

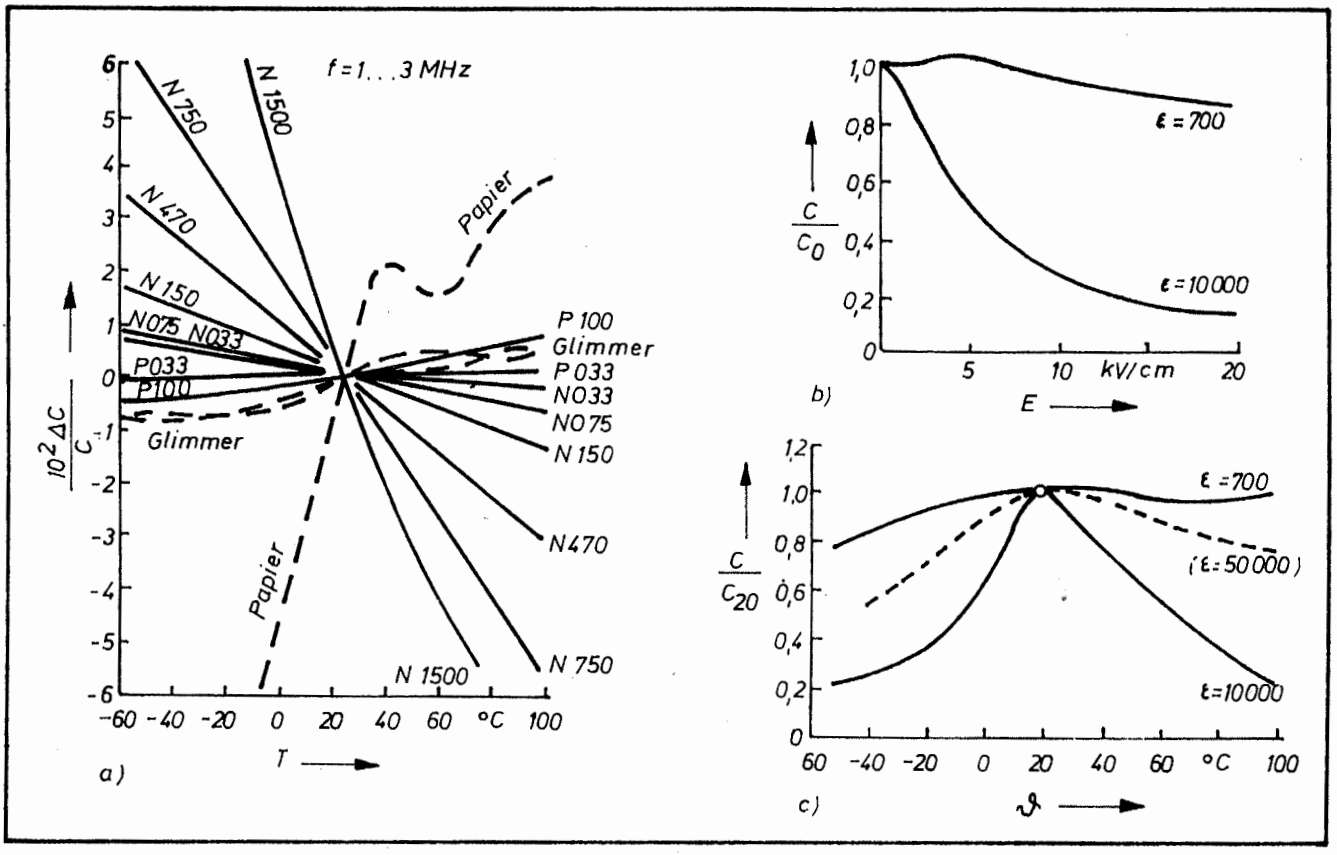


Bild 1.5.7. Eigenschaften von Keramikkondensatoren

a) Temperaturabhängigkeit verschiedener NDK-Kapazitäten im Vergleich zu Papier und Glimmer. Die Zahlenwerte geben ungefähr den TK in 10^{-6} je K an und sind zugleich die typischen Werkstoffbezeichnungen

b) Feldstärkeabhängigkeit von zwei extremen HDK-Materialien

c) Temperaturabhängigkeit bei HDK-Kondensatoren

1.5.5.3. Folienkondensatoren

Folienkondensatoren bestanden früher aus getränkten Papierwickeln (Öl bzw. Wachs) und Aluminiumfolie. Dieser Aufbau hat heute praktisch nur noch für höhere Spannungen und insbesondere für die Starkstromtechnik Bedeutung. Statt dessen haben sich folgende Kunststoffolien durchgesetzt:

K: Polystyrol (engl.: Polystyrene; auch Styroflex)
P: Polypropylen
T: Polyäthylenterephtholat (auch Polyester, Mylar, Hostafan)
C: Polycarbonat (auch Makrofol)
L bzw. U: Celluloseacetat.

Die Herstellungen erfolgen einmal in der von den Papierkondensatoren bekannten *Wickeltechnik* mit einer 5 bis 6 µm dicken Aluminiumfolie, wie sie Bild 1.5.8a) zeigt, oder durch einseitiges Bedampfen der Kunststoffolie mit Aluminium von 0,02 bis 0,1 µm Dicke, wobei die Kunststoffolie 2 bis 10 µm stark ist. Dann werden zwei derartig bedampfte Folien zu einem Wickel gemäß Bild 1.5.8b) zusammengefügt.

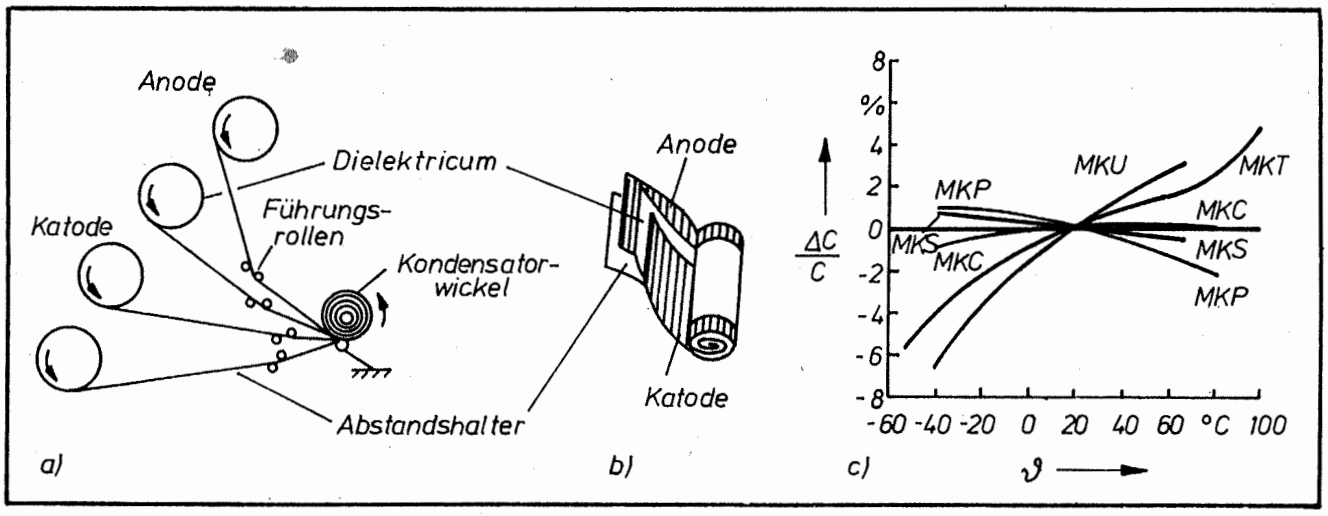


Bild 1.5.8. Aufbau und Eigenschaften von Folienkondensatoren
a) Prinzip des Wickelvorganges
b) Aufbau des Wickels
c) Temperaturgang verschiedener Materialien (vgl. Tab. 1.5.1)

Eine dritte Variante stellen die *Lackfolienkondensatoren* (MKL bzw. MKU) auf der Basis des Celluloseacetats dar. Hier wird die Folie gegossen (ca. 1 µm) und dann der Metallbelag aufgetragen. Alle dünn metallisierten Folien besitzen wie die alten *MP-Kondensatoren* (Metall-Papier) die Selbstheilung. Schlägt der Kondensator an einer Stelle durch, so verdampft das Metall und der Kondensator ist nach dem Durchschlag wieder brauchbar. Nach diesem Prinzip erfolgt auch das kurze Einbrennen der Kondensatoren nach der Fertigung. Es werden weiter Kunststoffnacktwickel und geschützte Wickel unterschieden. Die Nacktwickel werden nach dem Wickeln bis kurz vor dem Erweichungspunkt der Folie erhitzt und dabei zu einem kompakten Gebilde zusammengebacken. Die geschützten Kondensatoren werden darüber hinaus eingegossen bzw. in ein verschlossenes Gehäuse eingebaut, z.B. in Keramikrohr eingelötet. Bezüglich der Kontaktierung können eingelegte Kontakte und Spritzen (Schoopieren), kompaktes Verbinden der beiden Endflächen durch aufgetragene Metalle, unterschieden werden. Die zweite Form hat geringe Induktivitäten und ESR. Sie ist deshalb dämpfungsarm.

Tabelle 1.5.1. Richtwerte für den Reststrom verschiedener Elektrolytkondensatoren bei 25°C, Nennspannung nach mindestens 10 Minuten Betrieb

Kondensatortyp	Reststrom $\frac{\mu A}{\mu F \cdot V}$	Mindestwert μA
Aluminium		
● normal	0,02	3
● rauscharm professionell	0,002	0,2
● trocken	0,06	
Tantal		
● feucht	0,0005	
● trocken	0,005	

Die verschiedenen oben genannten Materialien führen zu Kondensatoren unterschiedlicher *Qualität*. Der Temperaturgang ist aus Bild 1.5.8c) zu ersehen. Weitere Daten enthält Tab. 1.5.1. Kurz *zusammengefaßt* gilt:

Polystyrol: geringste Verluste, hohe Kapazitätskonstanz, negativer TK, geringe spezifische Kapazität

Polyester: billiger Kondensator, guter TK, Folien bis 2 µm, starke Frequenzabhängigkeit ab 50 kHz

Polycarbonat: niedrige Verluste, großer Temperaturbereich, Folien bis 1,5 µm, hohe Kapazitätskonstanz, hoher Isolationswiderstand, professioneller Kondensator

Polypropylen: ähnlich Polystyrol, jedoch besseres Temperaturverhalten und größere spezifische Kapazität

Polyacetat: relativ hoher Verlustwiderstand, kleiner Isolationswiderstand, müssen vor Feuchtigkeit durch Umhüllen geschützt werden, größte volumenspezifische Kapazität von allen Koppelkondensatoren.

1.5.5.4. Elektrolytkondensatoren (Elko)

Elektrolytkondensatoren werden heute auf der Basis von *Aluminium* und *Tantal* sowie jeweils im *nassen* und *trockenen Aufbau* gefertigt. Versuche wurden auch bereits mit anderen Metallen, u. a. Ti, Nb und Zr, gemacht. Das Dielektrikum ist das Oxid, also Al_2O_3 bzw. Ta_2O_5. Dieses Oxid wird durch anodische Oxydation (Formierung) gerade so dick hergestellt, wie es für die Nennspannung des jeweiligen Elkos notwendig ist. Bei Aluminium sind dies 1,1 bis 1,4 und bei Tantal 1,7 bis 2,1 nm/V. Die Luftoxidschicht bei Al beträgt ca. 5 nm. Formiert wird immer auf den 1,5fachen Wert der Nennspannung. Bei diesen dünnen Schichten herrschen also in der Oxidschicht extrem hohe Feldstärken im Bereich von MV/cm. Deshalb müssen diese Schichten sehr gleichmäßig sein. Das bedeutet sehr reine Ausgangsmaterialien von mindestens 99,99%. Anderseits werden hierdurch aber gerade die hohen spezifischen Kapazitäten der Elkos erreicht.

Der eine (*positive*) *Kondensatorbelag* ist das *Metall*. Jenseits des Oxids läßt sich kein zweiter Leiter anbringen. Hier wird ein *Elektrolyt* verwendet. Er muß *Sauerstoff abgeben* können, damit die nicht mehr sehr stabile Oxidschicht im Betrieb durch anodische Oxydation *ständig regeneriert* (durch den Reststrom nachformiert) werden

kann. Dies ist auch der Grund, warum an einem Elko immer eine richtig gepolte Gleichspannung und niemals eine negativ werdende Spannung anliegen darf. Es gibt allerdings spezielle Elkos, die in beiden Richtungen formiert sind. Sie werden u. a. bei Lautsprecherweichen und als Anlaufkondensatoren für Motoren verwendet.

Bei Alu-Elkos wurden früher u. a. wäßrige Lösungen Ammoniumborat verwendet. Heute werden wasserfreie Elektrolyte auf der Basis von Glycerin oder Ethylenglykol bevorzugt. Bei den *Tantalelektrolyten* wird meist H_2SO_4 verwendet. Der Elektrolyt wird meist in eine Schicht saugfähiges Papier von 30 bis 60 µm gebunden. Jenseits des Elektrolyts befindet sich dann die zweite (*negative*) *Elektrode*. Als Festelektrolyt wird der Halbleiter MnO_2 benutzt.

Zur Erreichung einer *großen Oberfläche* werden beim Aluminium und Tantal unterschiedliche Wege gegangen. Die Aluminiumfolie von etwa 60 bis 100 µm wird speziell geätzt. Dadurch entstehen im Aluminium tiefe Poren (bei Hochvolt-Elkos bis 30 µm) von wenigen µm Durchmesser. Diese *Poren* müssen so gestaltet sein, daß sie beim Oxydieren offen bleiben und der Elektrolyt in sie eindringen kann. Auf diese Weise wird die wirksame Oberfläche bis auf das 100fache erhöht. Bei den relativ neuen Trocken-Alu-Elkos muß das MnO_2 in diese Poren mit einem pyrolytischen Prozeß hineingebracht werden. Die Kondensatoren heißen SAL (solid Aluminium).

Bei den Tantal-Elkos wird die Anode heute in *Sintertechnologie* hergestellt. Hierbei wird von einem Pulver mit einem mittleren Teilchendurchmesser von 3 bis 25 µm ausgegangen. Dieses Pulver wird in einem Preßkörper, der meist den Anodendraht enthält, auf 4 bis 9 g/cm³ verdichtet. Anschließend folgt der Sinterprozeß bei 1600 bis 2050 °C für 3 bis 40 min. Es werden so Oberflächen bis zu 6000 cm² je cm³ erreicht. Auf das Pulver bezogen, ergeben sich so Werte von 10000 µFV/g. Beim fertigen Tantal-Elko gilt etwa 5 µFV/cm³.

Alu-Elkos werden heute fast ausschließlich als Wickel aufgebaut und in einem geschlossenen Gehäuse untergebracht. Einen Ausschnitt aus den sich gegenüberstehen-

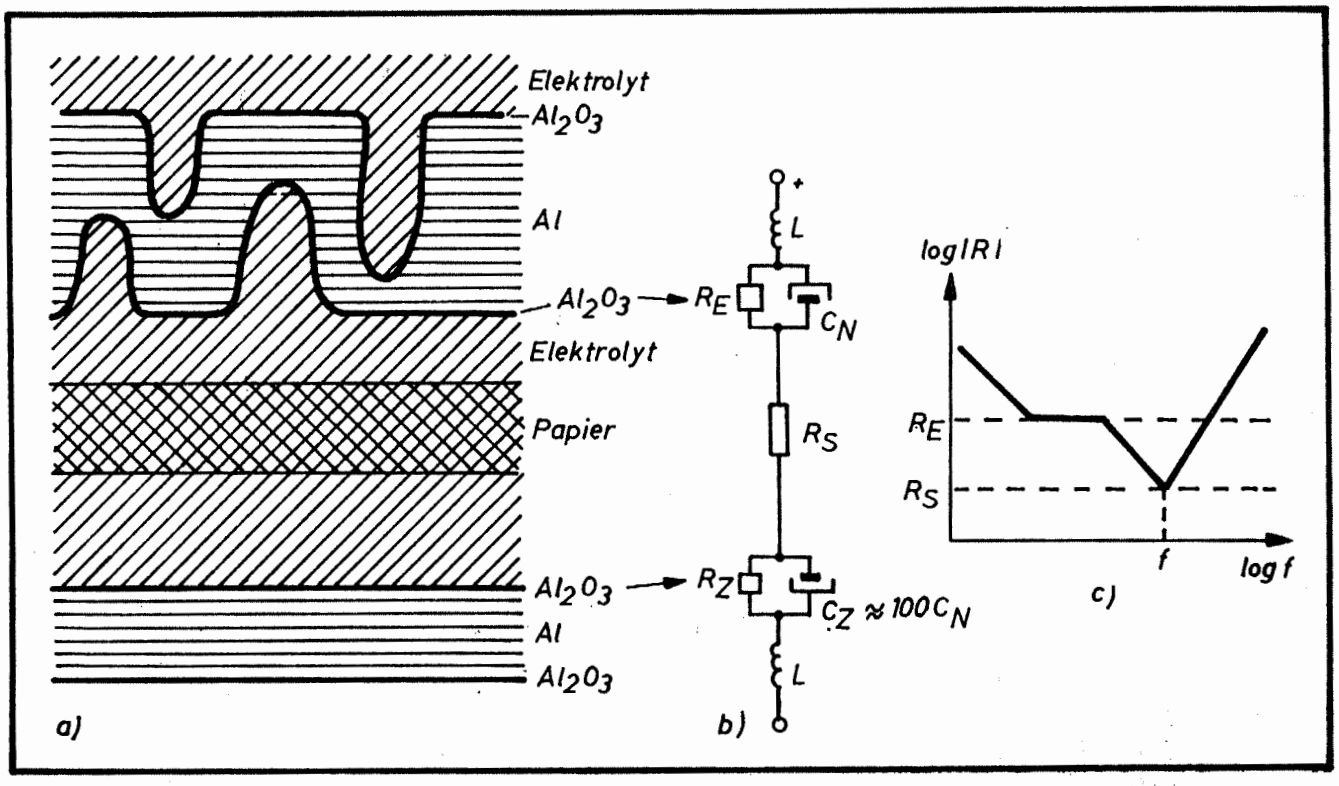


Bild 1.5.9. Zum Aluminium-Elektrolytkondensator

a) Ausschnitt aus einem Wickel

b) Ersatzschaltbild

c) Frequenzgang der Impedanz

den Elektroden zeigt Bild 1.5.9. Außer der gewollten Nennkapazität C_N der *rauhen Anode* entsteht an der *luftoxydierten Katode* eine Zusatzkapazität C_Z, die bei Niedervoltelkos etwa 100mal so groß wie die Nennkapazität und bei Hochvoltelkos sogar das 1000fache betragen kann. Zwischen beiden Kapazitäten liegt der Serienwiderstand des Elektrolyts, der partiell in den Poren sogar noch erheblich vergrößert ist. Parallel zu den Oxidschichten liegen die Isolationswiderstände R_E und R_Z. Ferner existiert die Zuleitungsinduktivität, welche durch induktive Einflüsse des Wickels und von Elektrolyten erhöht wird. So entsteht das *Ersatzschaltbild* eines Elkos gemäß Bild 1.5.9b). Es bewirkt einen typischen Frequenzgang gemäß Bild 1.5.9c). Die Impedanz nimmt zunächst gemäß C_Z ab, um dann auf den Wert von R_E zu fallen. Dann wird C_N wirksam, und es tritt die *Resonanz* bei f_0 auf mit dem ERS annähernd R_S. *Typische Werte* für f_0 liegen im Bereich von ca. 10 kHz bis einige MHz. Die bei Gleichstrom vorhandene Kapazität ist ca. 1,3mal so groß wie die Nennkapazität.

Den Aufbau von Tantal-Elkos zeigt Bild 1.5.10. Wichtige Eigenschaften von Elkos sind in Bild 1.5.11 zusammengestellt. Hierin fallen insbesondere die meist besseren Eigenschaften des Tantal-Elkos auf. Weiter ist deutlich zu erkennen, wie der Reststrom mit der Temperatur überproportional und oberhalb der *Nennspannung* extrem steil zunimmt. Dies ist der Grund, weshalb die Lebensdauer entsprechend schnell abnimmt. Es wird daher empfohlen, Elkos stets mit einer Spannung zu betreiben, die kleiner als die Nennspannung ist. Wegen der ständigen Nachformierung der Anode sollte sie anderseits aber auch nicht zu klein sein. Die Spitzenspannung

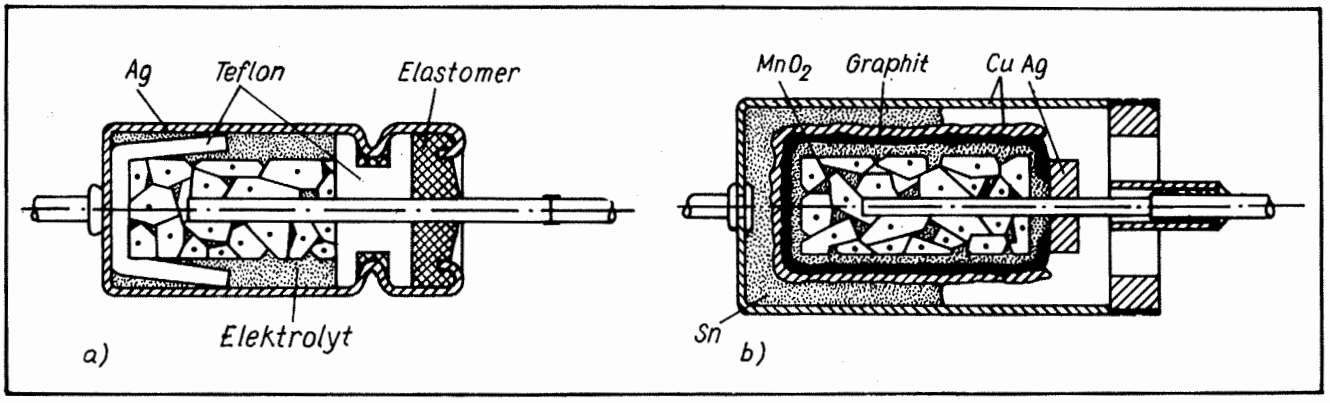


Bild 1.5.10. Schematischer Aufbau von Tantal-Elektrolytkondensatoren, a) feucht, b) trocken

(die auftretende maximale Spannung, oft Gleichspannung plus überlagerte Wechselspannung) sollte nie überschritten werden. Dann tritt nämlich ein großer Reststrom auf, der in Wärme umgewandelt wird und zur Explosion des Elkos führen kann. Beim Einschalten eines Elkos nach längerer Pause tritt in der Regel ein relativ großer Reststrom auf. Er regeneriert insbesondere stark abgebaute Stellen der Oxidschicht. Nach ca. 10 Minuten (spätestens 30) sollte er seinen Normalwert erreicht haben. Für die Richtwerte von *Reströmen* gilt Tab. 1.5.1. Tantal-Elkos sind empfindlich gegenüber Stoßbelastungen. Deshalb wird mindestens ein Innenwiderstand der Quelle von 3 Ω/V verlangt. Bereits bei Werten um 0,1 Ω/V ist die Ausfallrate der Elkos um den Faktor 10 größer.

Aluminium-Elkos sind dagegen *schaltfest* und lassen sich in dieser Hinsicht durch zusätzliche Maßnahmen erheblich stabil gestalten. Dies ist vor allem für Blitzkondensatoren und Schaltnetzteile wichtig. Die speziellen Maßnahmen betreffen u. a. Mehrfachkontaktierung des Wickels, Elektrolyten mit kleinem Widerstand, gute Wärmeableitung und spezielle Separatoren.

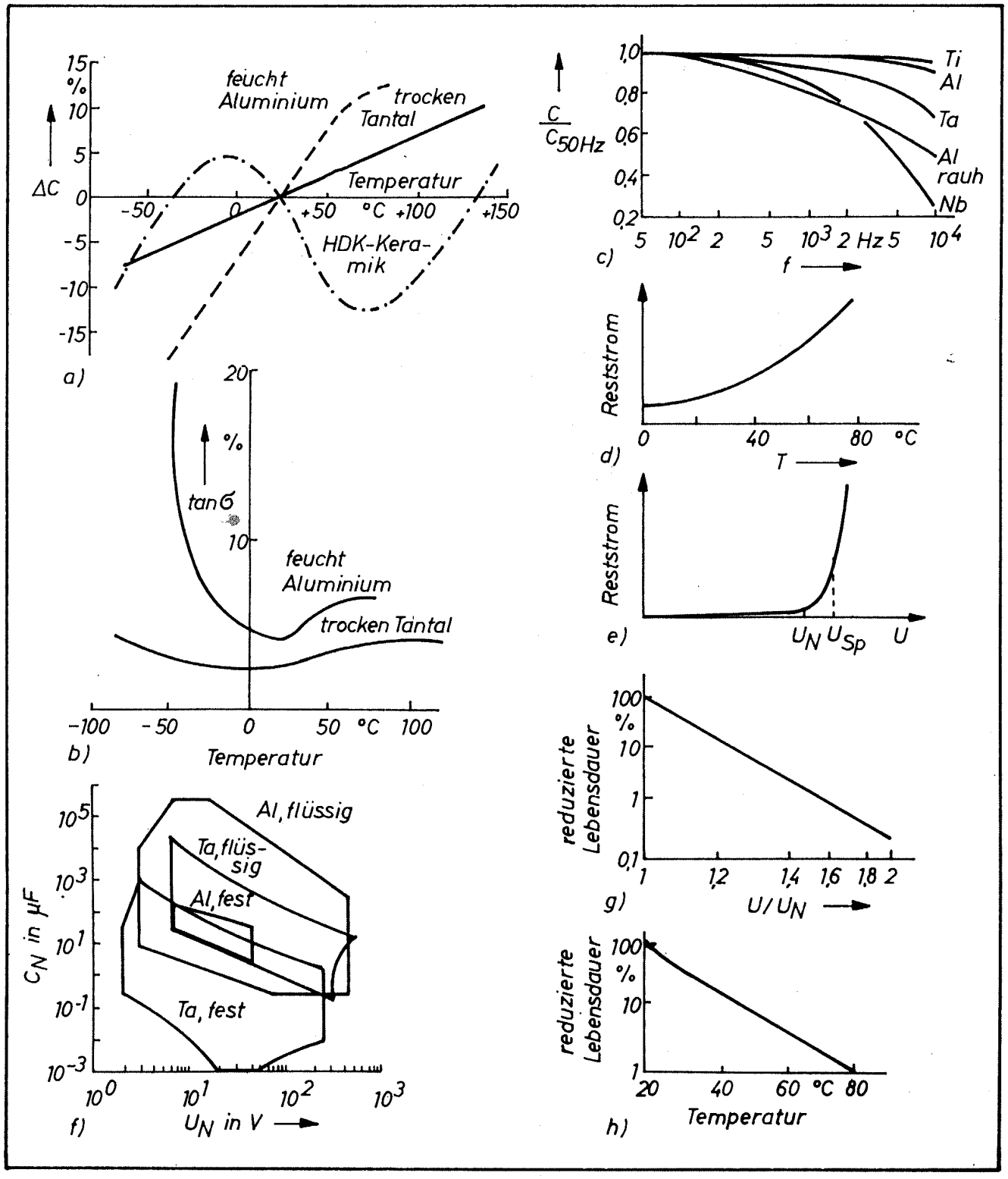


Bild 1.5.11. Eigenschaften von Elektrolytkondensatoren

a) Temperatureinfluß auf die Kapazität mit Vergleich zur HDK-Keramik
b) Verlustwinkel als Funktion der Temperatur
c) Änderung der wirksamen Kapazität mit der Frequenz
d) Reststrom als Funktion der Temperatur
e) Reststrom als Funktion der Spannung
f) Arbeitsbereiche der verschiedenen Elkos
g) Verringerung der Lebensdauer als Funktion der Überspannung
h) Verringerung der Lebensdauer als Funktion der Temperatur

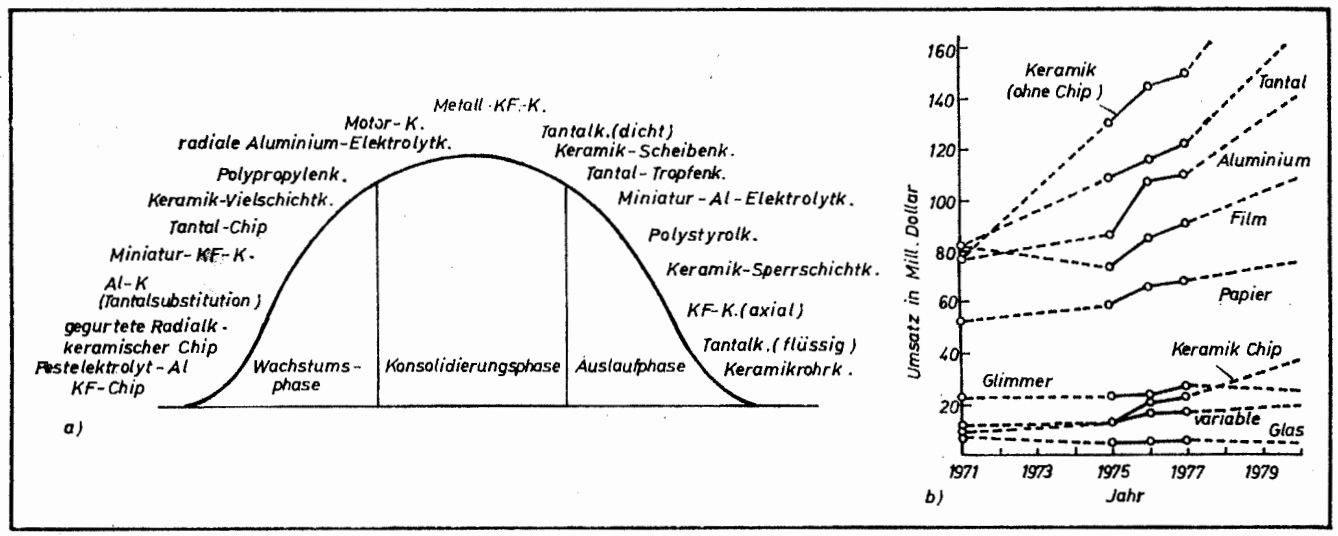


Bild 1.5.12. a) Lebenszyklus und b) Absatzgrößen für verschiedene Kondensatortypen

1.5.5.5. Vergleiche und Ergänzungen

Bild 1.5.12 zeigt Werte zur Verbreitung der vielfältigen Typen von Kondensatoren. Im Teilbild a) ist die Kurve des Lebenszyklus von Bauelementen dargestellt. Sie weist aus, welche Kondensatoren in der Produktion stark zunehmen (Wachstumsphase) und welche zurückgehen (Auslaufphase). Hieraus ist z.B. ersichtlich, daß die Alu-Elkos (insbesondere die Trocken-Elkos), trotzdem sie schon lange existieren, noch stark im Kommen sind. Hierbei dürfte auch der stark ansteigende Tantalpreis von Einfluß sein. Eine andersartige Darstellung zeigt der Umsatz von Kondensatoren in den USA.

Die Kurven in Bild 1.5.13 beziehen sich auf Kennwerte wie Verlustfaktor, Zeitkonstante, Energiespeicherung und Kapazitätsbereiche.

Neben den hier behandelten Kondensatoren gibt es noch einige weitere. Chipkondensatoren werden im Abschnitt 9.4., Dünnschichtkondensatoren in 9.4.5., Sperrschichtkondensatoren in 5.2. behandelt. Als weitere Kondensatoren sind die variablen Typen, z.B. gemäß Bild 1.5.14, und der *Durchführungskondensator* zu erwähnen. Die Durchführungskondensatoren führen Ströme und Spannungen in abgeschirmte Gehäuse. Spezielle variable Kondensatoren, die nur selten eingestellt werden (meist nur zum einmaligen Abgleich), heißen Trimmer.

Es sei erwähnt, daß in der letzten Zeit mehrfach über spezielle Kondensatoren berichtet wurde, die nach besonderen Prinzipien hohe Energiewerte speichern können. Eine Variante sind die ESD (engl.: energy storage device). Hierüber sind jedoch wenige Details bekannt. Unter anderem wird mitgeteilt, daß sie eine Silberanode, Graphitkatode und $RbAg_4J_5$ als Elektrolyten enthalten. An anderer Stelle wird mitgeteilt, daß an den Elektroden Ladungsdoppelschichten auftreten. Es werden insgesamt 10- bis 25fache Kapazitätswerte gegenüber anderen Elkos genannt.

1.5.6. Technische Induktivitäten

1.5.6.1. Grundlagen

Induktivitäten bestehen aus einer *Wicklung* und einem *magnetischen Kreis*. Der magnetische Kreis kann auch ganz oder teilweise ein Luftweg sein. In jedem Fall wird aber eine Spule zumindest auf einen Wickelkörper gewickelt. Wegen der großen

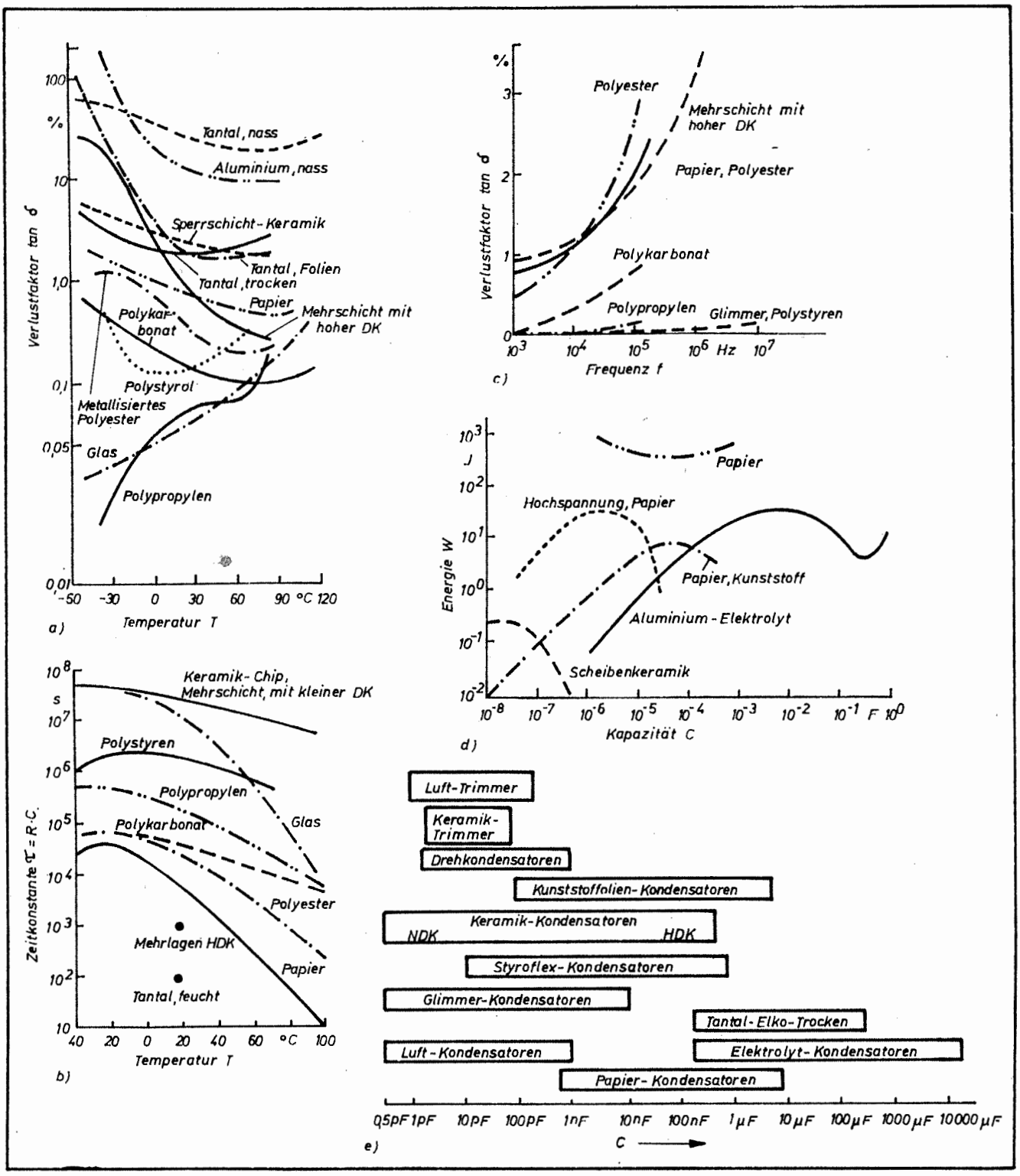


Bild 1.5.13. Vergleichende Kenndaten von Kondensatoren nach verschiedenen Gesichtspunkten

Vielfalt und der notwendigen verschiedenen Eigenschaften werden Spulen nur selten als fertige Bauelemente angeboten. Die fertigen Bauelemente sind die Wickelkörper bzw. magnetischen Kreise. Der magnetische Kreis ist durch den *magnetischen Widerstand* gekennzeichnet. Sein reziproker Wert wird häufig als A_L-Wert angegeben. Es gilt daher die Beziehung

$$\boxed{L = A_L n^2} \,. \tag{31}$$

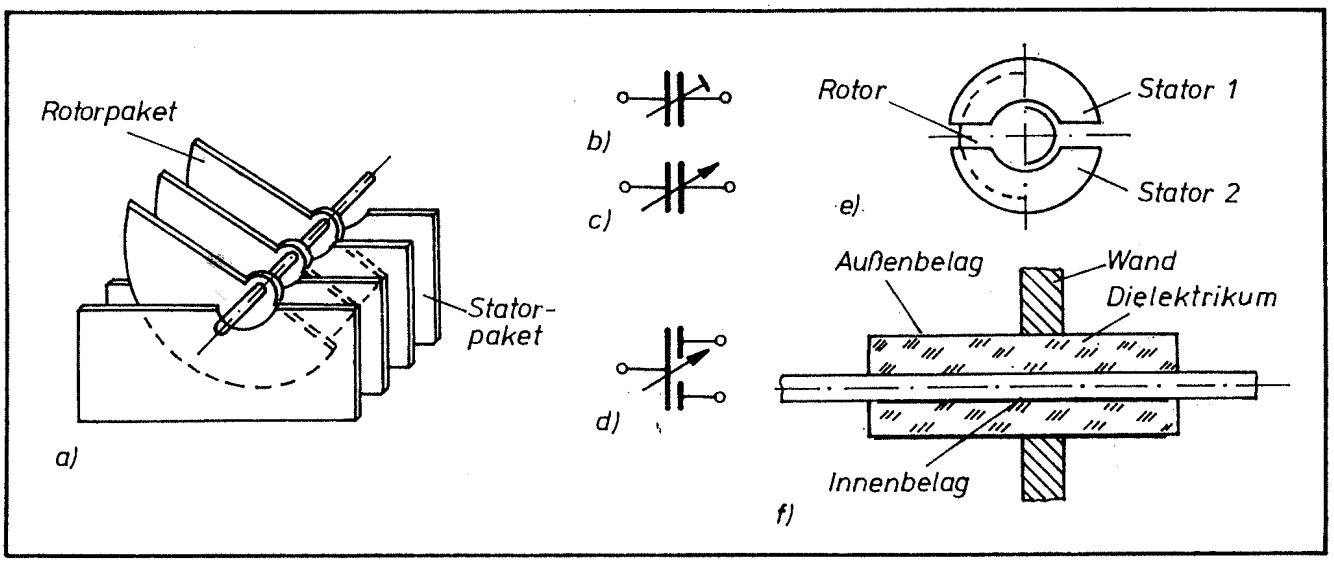


Bild 1.5.14. Sonderkondensatoren

a) Aufbauschema und c) Schaltbild eines Drehkondensators (Dreko)
b) Schaltbild eines Trimmers
d) Schaltbild und e) Aufbau eines Differentialdrekos
f) Aufbau eines Durchführungskondensators

Der A_L-Wert wird meist in nH/Wdg² angegeben. Für die üblichen Ferritkerne sind Werte zwischen 10 und 1000 vorhanden. Die *Verluste einer Spule* setzen sich aus drei Größen zusammen:

— *ohmscher Widerstand.* Der Wickeldraht einer jeden Spule besitzt einen gewissen ohmschen Widerstand. Er läßt sich nur dadurch vermindern, daß der vorgesehene Wickelraum durch eine entsprechende Drahtstärke voll ausgenutzt wird. Dieser Verlust-Widerstand wird mit R_{Cu} bezeichnet. Er steigt lediglich bei sehr hohen Frequenzen infolge des Skineffektes etwa proportional f^2 an. Dieser Einfluß ist durch die Parallelschaltung vieler dünner Drähte zu einer Litze zu senken.

— *Hystereseverluste.* Jedes ferromagnetische Material besitzt eine Hysteresekurve, die bei jedem Magnetisierungszyklus, also je Sekunde f-mal durchlaufen wird. Die Verluste sind proportional der Fläche der Hysteresekurve und der Frequenz. Sie sind elektrisch ersetzbar durch einen Widerstand

$$R_h = hn^2 f\,. \tag{32}$$

Darin bedeutet h einen Faktor, der vom Material abhängt und etwa proportional der Feldstärke $\boldsymbol{H}$ ist.

— *Wirbelstromverluste.* Durch Induktion bilden sich im ferromagnetischen Kern Wirbelströme, die proportional dem Quadrat der Frequenz sind. Sie können durch einen Ersatzwiderstand

$$R_{wi} = kn^2 f^2 \tag{33}$$

mit einer Konstanten k berücksichtigt werden. In diese Konstante kann auch der Einfluß des Skineffektes einbezogen werden.

Als *Güte* G einer Spule wird das Verhältnis von induktivem Widerstand zu den Verlusten bezeichnet. Zur Bestimmung der relativen Verlustanteile dient vielfach der reziproke Wert

$$\gamma = \tan\delta = \frac{R_{verl}}{\omega L}\,. \tag{34}$$

Nach den vorangegangenen Betrachtungen gilt

$$\gamma = \frac{R_{Cu} + a\omega + b\omega^2}{\omega L} = \frac{R_{Cu}}{\omega L} + \frac{a}{L} + \frac{b\omega}{L}. \tag{35}$$

Dieser Wert besitzt ein Minimum für

$$\omega_{opt}^2 = \frac{R_{Cu}}{b}, \tag{36}$$

also genau dann, wenn die Wirbelstrom- und Skinverluste gleich den Kupferverlusten sind. In diesem Bereich hat die Spule eine besonders hohe Güte. Sowohl bei tieferen als auch bei höheren Frequenzen ist entsprechend Bild 1.5.15a) der Verlustanteil größer. Deshalb sind *Induktivitäten meist nur für einen relativ kleinen Frequenzbereich optimal* verwendbar. Im Bild 1.5.15b) sind konstruktive Hinweise zur Verbesserung der Güte eingetragen.

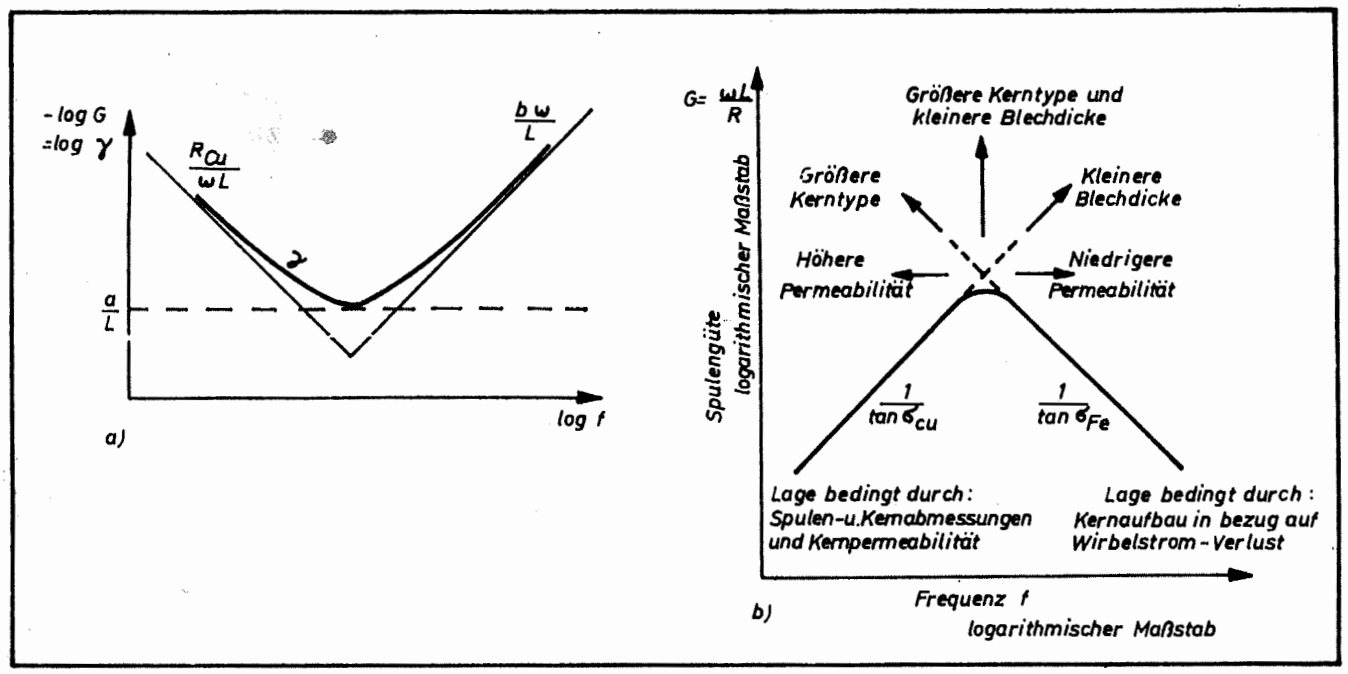


Bild 1.5.15. Zusammensetzung der Spulengüte aus den Verlustanteilen nach Gl. (36) (a) und schematisch mit konstruktiven Hinweisen (b)

In vielen Fällen werden noch zwei *weitere Verluste* berücksichtigt:

- *Nachwirkungsverluste*, die auf das Kernmaterial zurückgehen. Sie werden hier nicht weiter behandelt.
- *Resonanzverluste*, die wesentlich durch die Wickelkapazität bedingt werden.

Die Windungen einer Spule liegen so nebeneinander, daß sie eine Kapazität erzeugen. In vereinfachter Weise ist dies in Bild 1.5.16 schematisch dargestellt. Hierdurch liegt in Summe parallel zur Spule eine Ersatzkapazität, die *Wicklungskapazität*.

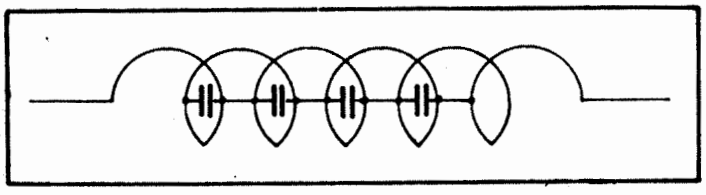


Bild 1.5.16. Zur Erklärung der Kapazität einer Spule

Sie senkt bzw. kompensiert den induktiven Anteil der Spule nach hohen Frequenzen. Bei sehr hohen Frequenzen (jenseits der Resonanzfrequenz, s. Kap. 2.) bleibt dann nur noch der kapazitive Anteil wirksam.

Insgesamt zeigt sich also, daß Induktivitäten, gleich welcher Art, aus den genannten Gründen nur in einem relativ kleinen Frequenzbereich brauchbar sind. Dies ist der Grund, weshalb für die einzelnen Frequenzbereiche sehr unterschiedliche Spulenformen verwendet werden.

Zuvor sei aber noch darauf hingewiesen, daß jeder Widerstand, jede Kapazität (vgl. 1.5.5.) ebenfalls induktive Komponenten enthält. Sie treten z. B. durch die spiralförmige Wendelung der Schichten auf, sind aber noch stärker bei den gewickelten Drahtwiderständen vorhanden. Durch spezielle *bipolare Wicklung* läßt sich der induktive Teil erheblich reduzieren. Dabei erfolgt die Wicklung mit zwei Drähten fortlaufend. Zu Beginn werden beide Drähte kurzgeschlossen, und am Ende der Wicklung stellen die beiden Drähte die Anschlüsse dar. Da so praktisch kein Magnetfeld entstehen kann, ist die Induktivität sehr klein. Andererseits ist aber die Wickelkapazität groß.

1.5.6.2. Blechkernspulen

Für Frequenzen bis zu einigen Tausend Hertz haben die Transformatorbleche als magnetischer Kreis besondere Bedeutung. Die verwendbare höchste Frequenz wird durch die *Stärke der Bleche* infolge der *Wirbelstromverluste* bestimmt. Neben der normalen Stärke von etwa 0,2 mm existieren Bleche bis herab zu 0,02 mm. Die Anfangspermeabilität der *Dynamobleche* liegt bei 500 und die Magnetisierung bei 2 Tesla. Vielfach werden auch Spezialbleche mit besonders hoher Anfangspermeabilität (ca. 10000), aber geringerer Magnetisierung verwendet. Bei hohen Leistungen hat meist Dynamoblech den Vorteil, während für hohe Frequenzen Bleche aus Speziallegierungen zu bevorzugen sind. Es gibt eine Anzahl *normierter Kernbleche*. Sie sind in Tab. 1.5.2 gestellt. Für Sonderfälle existieren auch M-Bleche mit einem Luftspalt. Der *M-Schnitt* (M = Mantel) hat den Vorteil, den Wickel gegen mechanische Beschädigungen zu schützen. Der Kern läßt sich aber nicht einfach schichten. Deshalb wird der M-Schnitt immer mehr durch den *EI-Schnitt* ersetzt, der bei gleichen Eigenschaften einen einfachen (maschinellen) Kernaufbau ermöglicht. Nachteilig ist die erforderliche Halterung für das Joch. Der *UI-Schnitt* wird bevorzugt bei Trenntransformatoren eingesetzt. Er hat den Vorteil, daß er relativ verlustfrei auszustanzen ist.

Neben diesen drei Blechformen setzen sich verstärkt *Bandkerne* durch. Sie verwenden ein Material, das durch Kaltwalzen vorzugsgerichtet ist. Dieses Material wird bandförmig hergestellt und zum Kern aufgewickelt. Daher sind derartige Kerne immer rund oder oval gestaltet. Sie sind besonders streuarm und benötigen bei gleicher Leistung nur 50% des Volumens. Sie sind jedoch schwierig zu bewickeln. So entstanden die *Schnittbandkerne*. Hier wird der Wickel geklebt und mit Trennschleifen in zwei Teile zerlegt. Diese Technik wird heute so gut beherrscht, daß beim Zusammenfügen der beiden Kernteile, die in die beiden Spulenhälften gesteckt werden, nur ein Luftspalt von ca. 5 μm entsteht.

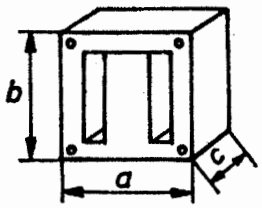
In verschiedenen Anwendungen fließt durch die *Spulen* ein *Gleichstrom*. Um dann eine möglichst gute Induktivität zu erreichen, muß ein genau definierter Luftspalt eingefügt werden. Mit der Induktion im Luftspalt B, die günstig zu 0,7 T gewählt wird, der Spaltweite d in m und dem Gleichstrom I in A gilt für die Windungszahl

$$n \approx \frac{80dB}{I}. \tag{37}$$

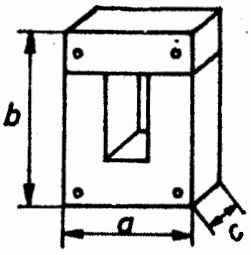
Für einen Eisenquerschnitt A in m² gilt dann für die Induktivität in Henry

$$L \approx \frac{10^{-4}An^2}{d}. \tag{38}$$

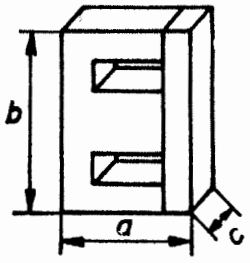
Tabelle 1.5.2. Bezeichnungen und Hauptabmessungen der wichtigsten aus Blechen aufgebauten Eisenkerne
Bei den EI- und M-Kernen bezeichnet die hinter dem Querstrich stehende Zahl die Dicke des Kernpakets. EI und UI stammt von der Form der Bleche, während M für Mantelkern steht.



Typ		M 20	M 30	M 42	M 55	M 65	M 74	M 85	M 102a 102/35	M 102b 103/53
Abmessungen in mm	*a*	20	30	42	55	65	74	85	102	102
	b	20	30	42	55	65	74	85	102	102
	c	5	7	15	20	26	22	32	35	53
Übertragbare Scheinleistung in VA				4	12	25	50	60	100	135



Typ		UI 30	UI 39	UI 48	UI 60	UI 75	UI 90	UI 102	UI 114	UI 132	UI 144	UI 168
Abmessungen in mm	*a*	30	39	48	60	75	90	102	114	132	144	168
	b	50	65	80				170	190	220	240	280
	c	15	20	25	30	40	50	55	62	70	75	90
Übertragbare Scheinleistung in VA			15	30	60	150	300	530	800	1200	1600	3000



Typ	EI 70	EI 92	EI 106	EI 130/36	EI 130/46	EI 150/40	EI 150/50	EI 170/60	EI 170/65	EI 170/75	EI 195/56
a	55	74	85	105	105	120	120	120	140	140	180
b	70	92	106	130	130	150	150	150	170	170	195
c	20	34	34	38	47	42	52	62	67	77	58
s in VA	25	50	125	250	320	370	450	550	750	850	1000

1.5.6.3. Ferritkernspulen

Den Aufbau der für Ferritkerne typischen Schalenkerne zeigt Bild 1.5.17. Die Wickelkörper gibt es in mehreren Ausführungen. Im Mittelteil besitzt der Ferritkern einen Luftspalt, mit dem er auf den A_L-Wert eingestellt wird. Im geringen Umfang kann dann die Induktivität mit einem Gewindekern abgeglichen werden. Die wichtigsten Abmessungen von *Schalen-* und *EE-Kernen* enthält Tab. 1.5.3. Darüber hinaus werden noch *X-* und *RM-Kerne* gefertigt. Sie sind insbesondere für gedruckte Schaltungen vorteilhaft, da sie unmittelbar quaderförmigen Aufbau zulassen. Sie besitzen jedoch keinen so gut geschlossenen magnetischen Kreis.

Bild 1.5.18 gibt Hinweise auf den Verlustwinkel und die nutzbare Frequenz bei Ferritkernen. Insgesamt gibt es etwa folgende Gruppen:

1. Mn-Zn-Ferrite — bis etwa 2 MHz
2. kubische Ni-Zn-Ferrite mit Co-Zusätzen — von etwa 2 bis 30 MHz
3. Ni-Zn-Ferrite mit Perminvar-Charakter — von etwa 30 bis 200 MHz
4. hexagonale Ferrite mit planarer Anisotropie — über 300 MHz.

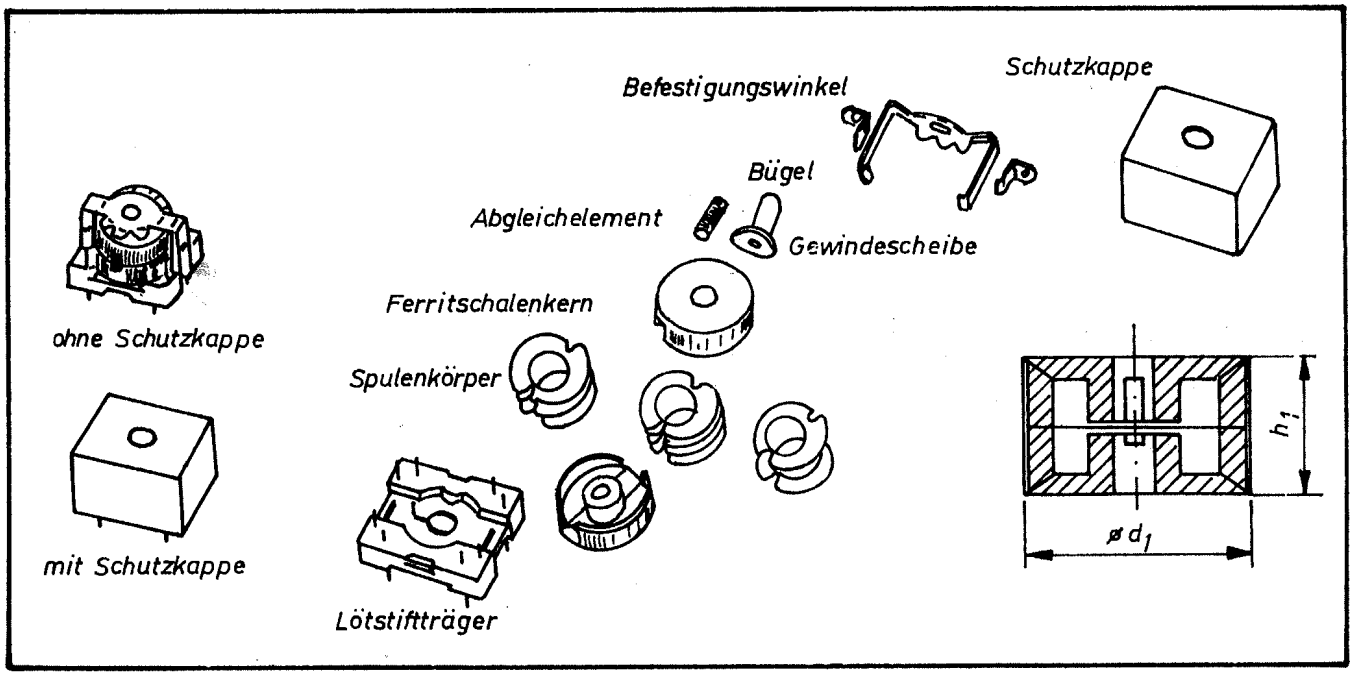


Bild 1.5.17. Aufbau von Schalenkernen: links mit und ohne Schutzkappe in der Ausführung für gedruckte Schaltungen; in der Mitte Explosionszeichnung; rechts unten Lage des Abgleichkerns zwischen den beiden Halbschalen aus Ferrit mit Andeutung des mittleren Luftspaltes

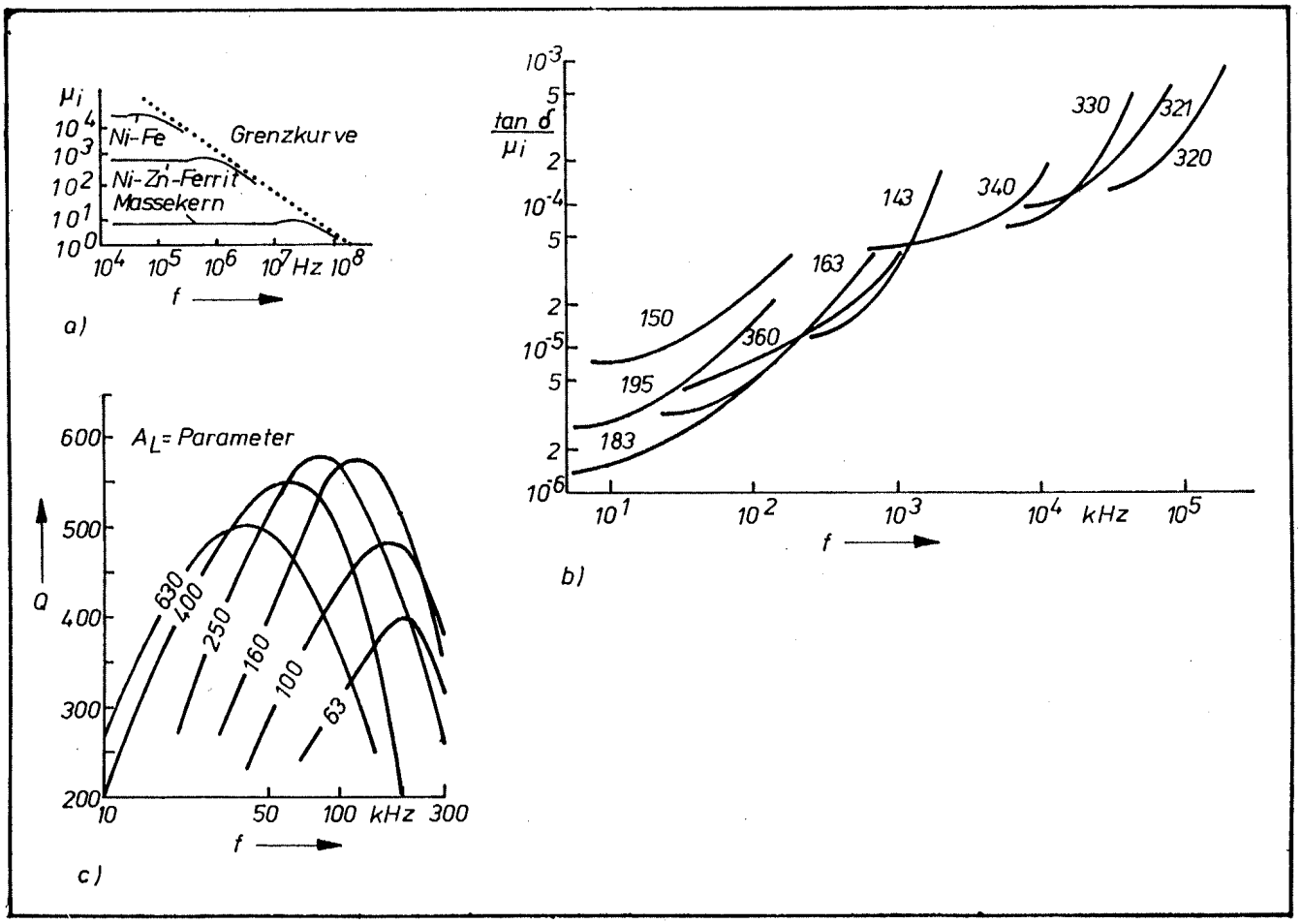


Bild 1.5.18. Zu Frequenzeigenschaften der Ferritspulen

a) Typischer Verlauf der Permeabilität; die Grenzkurve entspricht etwa Gl. (40)

b) Verlustwinkel, dividiert durch die Permeabilität, als Funktion der Frequenz mit Parametern verschiedener Manipermaterialien; die 100er Serie betrifft MnZn und die 300er NiZn mit Co-Zusatz

c) Gütewerte für Schalenkern 22 × 13 Manifer 163; 100 Windungen; $R_{Cu} = 0{,}94\ \Omega$; $L_0 = 99{,}8\ \mu H$

Tabelle 1.5.3. Abmessungen von Ferritkernen
Die Angaben des Wickelraumes gelten für den Wickelkörper mit einer Kammer. Folgende A_L-Werte werden bevorzugt benutzt: 25; 40; 63; 100; 160; 250; 400; 630; 1000; 1600; 2500; 4000; 6300. Die Werte *a*, *b*, *c* beziehen sich auf die Maßskizze des M-Kerns in Tab. 1.5.2. Der EE 42 wird auch als EE 42/20 mit $c = 20$ mm gefertigt.

Nenngröße	Durchmesser mm	mm	Höhe mm	Nutzbarer Wickelraum-querschnitt mm²	Mittlere Windungslänge cm
5 × 3	5,8		3,3		
7 × 4	7,35		4,2	1,4	1,4
9 × 5	9,3		5,4	2,1	1,9
11 × 6	11,3		6,6	2,7	2,2
14 × 8	14,3		8,5	6,9	2,8
18 × 11	18,4		10,7	11,9	3,6
22 × 13	22,0		13,6	17,6	4,4
26 × 16	26,0		16,3	27,6	5,2
30 × 19	30,5		19,0	40,3	6,1
36 × 22	36,2		22,0	56,1	7,3
42 × 29	43,1		29,9	120	8,5
	a	*b*	*c*		
EE 20	20	20,4	5,3	27,5	3,8
EE 30	30	30,4	7,3	27,5	5,0
EE 42	42	42,4	15,3	138	9,2
EE 55	55	55,6	21,0	229	12

Für die erreichbaren Grenzfrequenzen gelten etwa die folgenden Formeln für die Permeabilität

$$\mu_r - 1 \leqq 5 \cdot 10^9 \text{ Hz}/f \tag{39}$$

und für die Güte

$$G \leqq 5 \cdot 10^8 \, (f/\text{Hz})^{2/3} \, . \tag{40}$$

Ab Frequenzen von einem MHz ist es vorteilhaft, HF-Litze zu verwenden, um die Verluste des Skineffektes möglichst klein zu halten.

1.5.6.4. Offene Spulen

Mit steigender Frequenz werden die Magnetkerne immer weniger geeignet für den Spulenaufbau. Dadurch ist die Spule ohne Kern, die Luftspule, das andere Extrem. Hierfür existieren mehrere Aufbaumöglichkeiten, die Bild 1.5.19 beispielhaft zeigt. Die Spulen unterscheiden sich vor allem durch die *Wicklungsart*, also ein- und mehrlagig, Kreuzwicklung, und durch den Wickelkörper. Im weitesten Sinne gehören hierzu auch die gedruckten Spulen. Wichtig bei den Luftspulen (wegen der hohen Frequenzen) ist die Wickelkapazität. Zur *Berechnung* der *Luftspulen* gibt es viele Formeln. Sie sind alle mehr oder weniger genau. Hier sei eine Formel für einlagige Spulen mit dem Durchmesser D (cm), der Länge l (cm) und der Windungszahl n angegeben,

$$L/\text{nH} = \frac{\pi^2 D^2 n^2}{1 + 0{,}45D - 0{,}003D/l} \, , \tag{41}$$

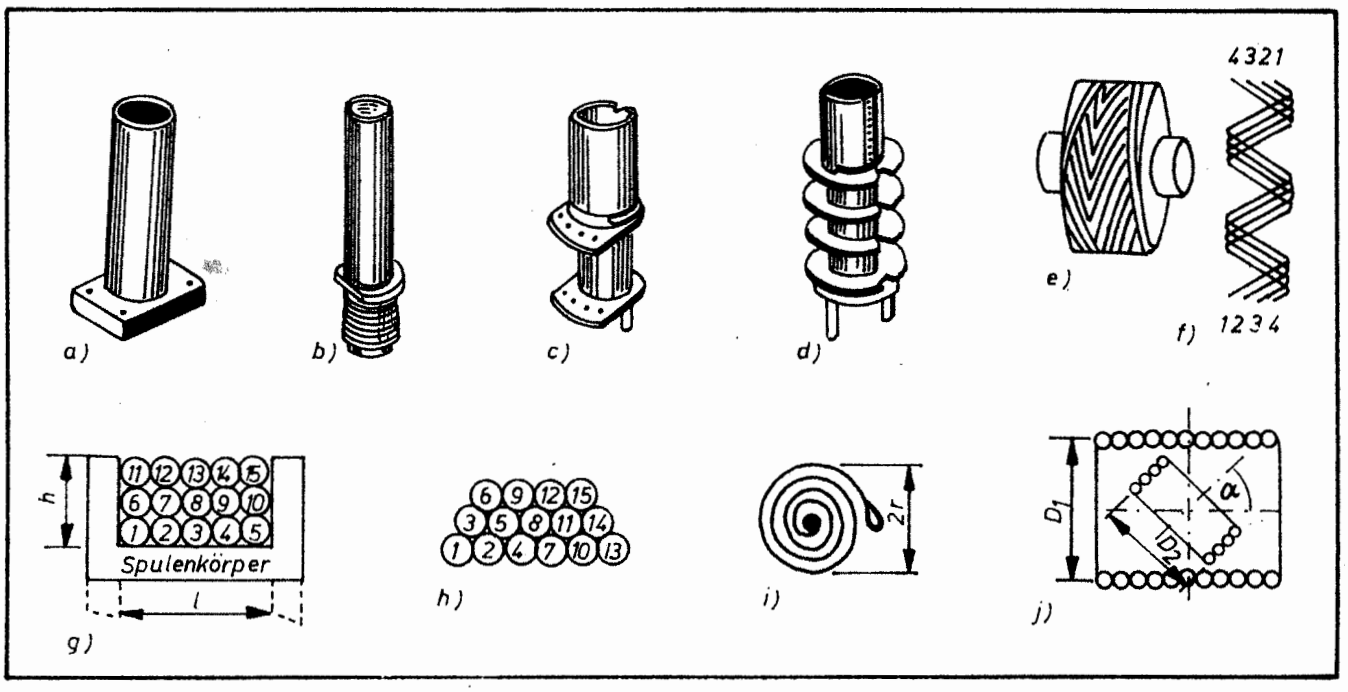


Bild 1.5.19. Beispiele von Luftspulen

a) bis d) Wickelkörper aus Polystyrol mit eventuell einfügbarem Gewindekern
e) Kreuzwickelspule
f) Drahtfolge beim Kreuzwickel
g) Lagenweise Wicklung in einer Kammer eines Wickelkörpers
h) Pyramidenwicklung
i) Beispiel einer gedruckten Spule, $L \approx 0{,}56\mu_0 n^2 r$
j) Luftvariometer (es gibt auch Ausführungen mit Kernen)

die in einem großen Bereich von D und l gilt.

Viele Luftspulen im MHz-Bereich können in ihrer Induktivität mittels eines einschraubbaren Eisenkernes (z.T. Carbonyleisenpulver) in einem gewissen Bereich eingestellt werden. Zum Teil wird dieses Prinzip auch zum ständigen Abgleich verwendet. Dann heißen diese Spulen *Variometer*. Das eigentliche Variometerprinzip (Bild 1.5.19e)) beruht dagegen auf der Verstellung von zwei Spulen gegeneinander. Es wird also die *Gegeninduktivität* benutzt, deren Prinzip erst im nächsten Abschnitt behandelt wird.

1.6. Transformatoren und Übertrager

1.6.1. Gegeninduktivität

Entsprechend Bild 1.6.1 seien zwei Spulen so angeordnet, daß der Magnetfluß einer erregten Spule L_1 die Spule L_2 zumindest teilweise durchdringt. Jede Stromänderung in der erregten Spule bewirkt dann in der zweiten Spule eine Flußänderung, und an ihren Enden entsteht eine Spannung. Die Spannungsänderung der stromerregten Spule war ein Maß für ihre Selbstinduktion (Gl. (1.5.9)). In diesem Sinne kann jetzt die Spannung an der zweiten Spule als Folge einer Gegeninduktivität bezüglich der ersten Spule interpretiert werden. Die Gegeninduktivität entspricht damit dem Kernwiderstand des Vierpols (Gl. (1.1.40)). Solange die Permeabilität im Bereich der Magnetfelder bei der Spulen nicht von der Feldstärke abhängt, ist die Gegeninduktivität von der ersten zur zweiten Spule M_{12} gleich der Gegeninduktivität von der zweiten zur ersten M_{21}. Stets ist dies bei Spulen und Leiteranordnungen in Luft erfüllt. Diese Relation wird als *Reziprozität* bezeichnet. Mit gewissen Einschränkungen ist

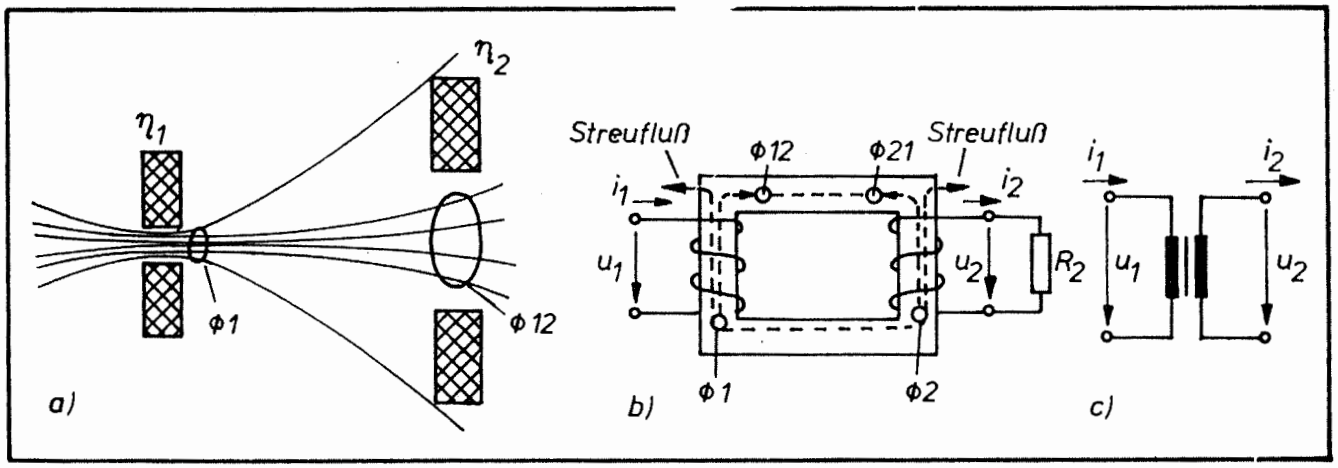


Bild 1.6.1. Zur Erklärung der Gegeninduktivität

a) Schematische Spulenanordnung
b) Flüsse und Ströme in einem Transformator
c) Ersatzschaltbild eines Transformators

sie auch noch dann gegeben, wenn bei der Erregung der beiden Spulen an entsprechenden Stellen ein gleich großes H auftritt. In allen anderen Fällen ist das Verhältnis der beiden Gegeninduktivitäten kaum zu berechnen. Für die Induktivität folgt aus Abschnitt 1.2. und Tab. 1.3.1 sowie der Windungszahl n der Spule

$$\sum I = n \cdot I\,, \tag{1}$$

$$L = \frac{\Phi n}{I}. \tag{2}$$

Deshalb wird für die Gegeninduktivität

$$M_{12} = \frac{\Phi_{12} n_2}{I_1}. \qquad M_{21} = \frac{\Phi_{21} n_1}{I_2} \tag{3}$$

definiert. Mit dem *Kopplungsfaktor* kann das Verhältnis des Flusses in der erregten Spule zum Fluß in der passiven Spule ausgedrückt werden:

$$k_{21} = \frac{\Phi_{21}}{\Phi_2} \quad \text{bzw.} \quad k_{12} = \frac{\Phi_{12}}{\Phi_1}. \tag{4}$$

Bei geltender Reziprozität sind beide Kopplungsfaktoren gleich groß. Andernfalls wird häufig ein mittlerer Kopplungsfaktor für beide Spulen angegeben,

$$k_m = \sqrt{k_{21} \cdot k_{12}}. \tag{5}$$

Ein anderer Ausdruck für die Kopplung zwischen den Spulen ergibt sich durch die *Streufaktoren* σ. Es gilt

$$k_{12} = 1 - \sigma_1\,, \qquad k_{21} = 1 - \sigma_2\,. \tag{6}$$

Sie berücksichtigen gerade jenen Anteil, der nicht in die Kopplung eingeht. So wie es den *mittleren Kopplungsfaktor* gibt, existiert auch ein totaler Streufaktor σ gemäß

$$k_m^2 = 1 - (\sigma_1 + \sigma_2 - \sigma_1 \cdot \sigma_2) = 1 - \sigma\,. \tag{7}$$

Aus (2) bis (4) folgt

$$M_{12} = k_{12} L_1\,, \qquad M_{21} = k_{21} L_2\,. \tag{8}$$

Für die *mittlere Gegeninduktivität* gilt daher

$$M_m = k_m \sqrt{L_1 L_2}\,. \tag{9}$$

1.6.2. Idealer Transformator

Bei einer *vollkommen festen Kopplung* zwischen beiden Spulen,

$$k = 1 \quad \text{oder} \quad \sigma = 0\,, \tag{10}$$

wird die gesamte magnetische Energie von der Primär- zur Sekundärspule übertragen. Mit der Primärinduktivität L_1 und der Sekundärinduktivität L_2 gilt daher

$$M = \sqrt{L_1 L_2}\,. \tag{11}$$

Es sei weiter angenommen, daß beide Induktivitäten sich auf einem gemeinsamen Kern mit dem magnetischen Widerstand P befinden. Dann gilt (Gl. (1.5.10))

$$L_1 = \frac{n_1^2}{P}, \qquad L_2 = \frac{n_2^2}{P}, \qquad M = \frac{n_1 n_2}{P}. \tag{12}$$

Infolge des Induktionsvorganges und einer Erregung des Kerns werden beide Spulen unabhängig davon, woher die Erregung stammt, vom gleichen Fluß durchsetzt. Es entstehen an ihnen die Spannungen

$$U_1 \sim n_1 \frac{\mathrm{d}\Phi}{\mathrm{d}t}, \qquad U_2 \sim n_2 \frac{\mathrm{d}\Phi}{\mathrm{d}t}, \tag{13}$$

und folglich gilt

$$\boxed{\frac{U_1}{U_2} = \frac{n_1}{n_2} = \ddot{u}}\,. \tag{14}$$

Die Spannungen an beiden Spulen verhalten sich also genau so wie ihre Windungszahlen. $\ddot{u}$ heißt Übersetzungsverhältnis. Wird der Fluß von beiden Spulen durch die Ströme i_1 und i_2 hervorgerufen, so gilt

$$\left.\begin{aligned} \boldsymbol{u}_1 &= \mathrm{j}\omega(L_1 \boldsymbol{i}_2) = \frac{\mathrm{j}\omega n_1^2}{P}\left(\boldsymbol{i}_1 \pm \frac{1}{\ddot{u}}\,\boldsymbol{i}_2\right), \\ \boldsymbol{u}_2 &= \mathrm{j}\omega(L_2 \boldsymbol{i}_2 \pm M \boldsymbol{i}_1) = \frac{\mathrm{j}\omega n_2^2}{P}\,(\ddot{u}\boldsymbol{i}_1 \pm \boldsymbol{i}_2)\,. \end{aligned}\right\} \tag{15}$$

Mit der komplexen Schreibweise erfolgt hier ein Vorgriff in das Kapitel 2.) Hieraus folgt die zugehörige Ersatzschaltung gemäß Bild 1.6.2a). Aus dem ersten Teil der

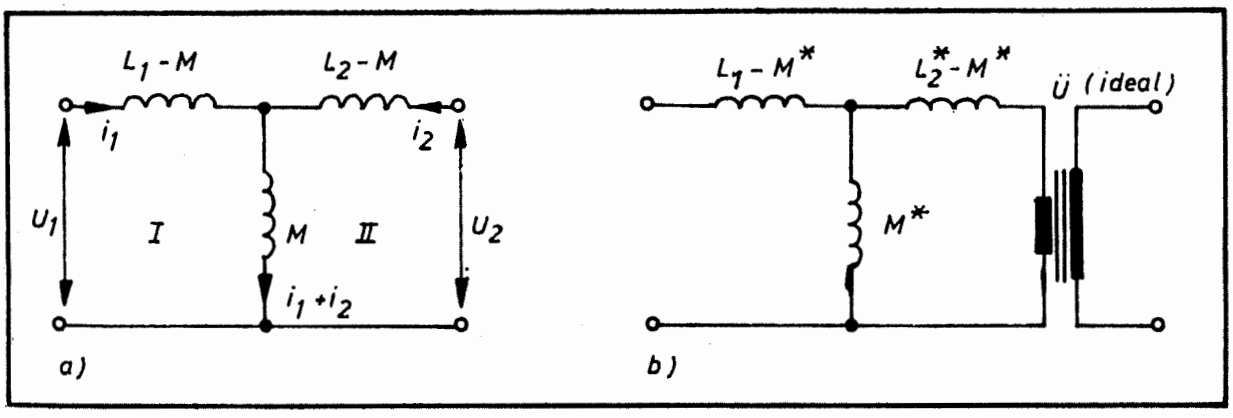


Bild 1.6.2. Ersatzschaltbilder,

a) von einem idealen Transformator,

b) von einem Transformator mit Streuverlusten in den $*$-Größen und einem nachfolgenden idealen Übertrager

Gl. (15) folgt für den Eingangsstrom

$$\underline{i}_1 = \frac{\underline{u}_1}{\mathrm{j}\omega L_1} \pm \frac{1}{\ddot{u}}\underline{i}_2 . \tag{16}$$

Wird angenommen, daß der Blindstrom infolge der Induktivität sehr gering ist,

$$\underline{i}_L = \frac{1}{\mathrm{j}\omega L_1} < \frac{1}{\ddot{u}}\underline{i}_2 , \tag{17}$$

so ergibt sich auch ein einfaches Verhältnis für den Primär- und Sekundärstrom:

$$\boxed{\frac{\underline{i}_2}{\underline{i}_1} = \ddot{u}} . \tag{18}$$

Mit guter Näherung gilt dies Stromverhältnis immer bei starker Belastung der Sekundärseite. Prinzipiell kann dieses Verhältnis auch dadurch erreicht werden, daß der Blindstrom $\underline{i}_L$ für die einzelne erregende Frequenz durch Resonanz mit einem Kondensator kompensiert wird. Aus (14) und (18) folgt

$$\boxed{\frac{\underline{U}_1 \cdot \underline{i}_1^*}{\underline{U}_2 \cdot \underline{i}_2} = \frac{P_1}{P_2} = 1} . \tag{19}$$

Der ideale Transformator überträgt also die volle Leistung. Ist die Sekundärspannung höher, so muß dafür der Sekundärstrom geringer sein und umgekehrt. In diesem Sinne folgt für die an der Sekundär- und Primärseite angeschlossenen Widerstände

$$\boxed{\frac{n_1}{n_2} = u = \sqrt{\frac{R_1}{R_2}}} . \tag{20}$$

Die Widerstände (auch komplexe) werden also entsprechend dem Quadrat des Windungszahlverhältnisses übertragen.

1.6.3. Verbessertes Ersatzschaltbild

Ist das Übersetzungsverhältnis größer oder kleiner als 1, so ist eine der Längsinduktivitäten negativ. Da negative Induktivitäten physikalisch nicht realisierbar sind, ist es üblich, alle Werte auf die Primär- oder Sekundärseite umzurechnen. Dann besteht der *ideale Transformator aber nur aus einer Gegeninduktivität und der Übersetzung.* Werden jedoch die Kopplungsfaktoren zugelassen, so entsteht eine Ersatzschaltung gemäß Bild 1.6.2b). Sie enthält *zwei Teile*: den *idealen Transformator* nur zur Übersetzung und eine *T-Schaltung* aus Induktivitäten, die einen Transformator für die Übersetzung 1:1, jedoch mit Streuverlusten, darstellt. Es gilt

$$L_2^* = \frac{L_2}{\ddot{u}^2} . \qquad M^* = \frac{k_\mathrm{m}}{\ddot{u}}\sqrt{L_1 L_2} . \tag{21}$$

Zusätzlich können jetzt auch ohmsche Verluste berücksichtigt werden. So entsteht eine Ersatzschaltung gemäß Bild 1.6.3a). Dabei wurde die Übersetzung zur Vereinfachung fortgelassen. Die Kupferverluste gehen in den Längs- und die Wirbelstrom- und Hystereseverluste in den Querzweig ein.

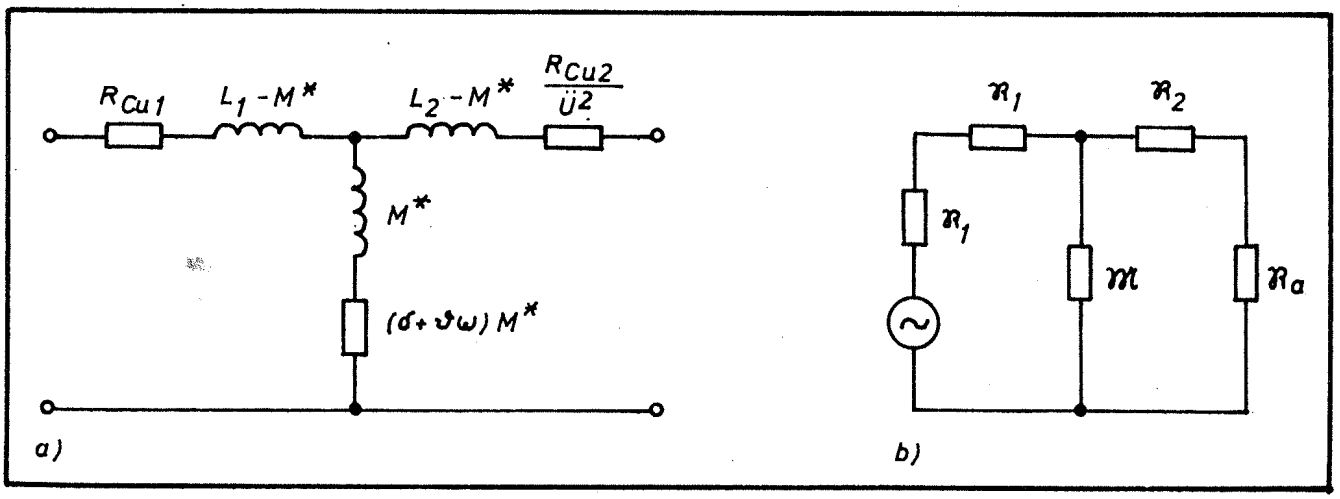


Bild 1.6.3. Ersatzschaltbilder

a) eines realen Transformators unter Berücksichtigung der ohmschen, hysteretischen und wirbelstrombedingten Verluste
b) zur Bestimmung des Frequenzganges infolge von Belastung

1.6.4. Frequenzgang

Der Frequenzgang des Transformators entsteht im wesentlichen bei seiner Belastung an der Primär- und Sekundärseite. Die Berechnung ist daher entsprechend dem Bild 1.6.3b) durchzuführen. Nach einer längeren Rechnung, auf die hier verzichtet wird und die eine ideale Anpassung des Eingangs- und Ausgangswiderstandes voraussetzt, läßt sich zeigen, daß die mittlere Übertragungsfrequenz durch

$$\omega_m = \sqrt{\frac{R_a R_e}{\sigma L_1 L_2}} \tag{22}$$

bei

$$\frac{L_2}{L_1} = \frac{R_e}{R_a}$$

gegeben ist. Für die Spannungs- bzw. Stromübersetzung folgt dann

$$\frac{u_2}{u_1} = \ddot{u}\,\frac{1}{1 + j\sqrt{\sigma}\,\dfrac{\omega}{\omega_m}}, \qquad \frac{i_1}{i_2} = \frac{1}{\ddot{u}}\,\frac{1}{1 - j\sqrt{\sigma}\,\dfrac{\omega_m}{\omega}}. \tag{23}$$

Sie entsprechen *in erster Näherung einem Schwingkreisverhalten*, dessen Güte durch die Streuung bestimmt ist. Mit der üblichen Definition der Grenzfrequenzen gilt daher

$$\boxed{\frac{\omega_g}{\omega_m} = \frac{\omega_m}{\omega_k} = \frac{1}{\sqrt{\sigma}}} \tag{24}$$

ω_g ist dabei die höchste und ω_k die niedrigste Eckfrequenz. Da σ bei guten Übertragern etwa 1% beträgt, läßt sich ein Frequenzverhältnis von 1:100 übertragen. Lediglich bei sehr guten Übertragern ist 1:1000 erreichbar. Hierbei haben dann aber auch die Kapazitäten der Wicklungen Bedeutung. Sie bewirken nämlich Resonanzfrequenzen, die zusätzlich die Bandbreite einengen.

1.6.5. Leistungstransformator

Bei allen bisherigen Übertragern wurde eine kleine Aussteuerung und damit konstante Permeabilität vorausgesetzt. Bei den verschiedenen Leistungstransformatoren für *Netz- oder Tonfrequenz* ist nun aber gerade die zu übertragende Leistung von entscheidender Bedeutung. Sie wird vor allem durch den Querschnitt Q des Eisenkerns begrenzt. Für Netztransformatoren gilt bei einem Volt-Ampere-Produkt als Leistung P

$$Q/\mathrm{cm}^2 \approx \sqrt{P/\mathrm{W}}\,. \tag{25}$$

Bei einer Induktion von 10000 G gilt dann weiter für die Abhängigkeit der mittleren Windungszahl von der Spannung U

$$n \approx \frac{45\,U/\mathrm{V}}{Q/\mathrm{cm}^2}. \tag{26}$$

Bei Ausgangstransformatoren ist die untere Grenzfrequenz f_u und die übertragene Leistung entsprechend

$$Q/\mathrm{cm}^2 \approx 10\sqrt{\frac{2P/\mathrm{W}}{f_u/\mathrm{Hz}}} \tag{27}$$

entscheidend. Zur Minderung der Streuung ist es zuweilen üblich, die Primär- und Sekundärwicklung auf Transformatoren zu verschachteln.

1.6.6. Spartransformator

Wird eine Induktivität angezapft, so kann sie ebenfalls als Transformator mit fester Kopplung aufgefaßt werden. In diesem Falle liegt zwar keine galvanische Trennung zwischen der Primär- und der Sekundärseite vor, aber dennoch kann entsprechend Bild 1.6.4 der Spartransformator genau so wie ein Transformator mit zwei Spulen behandelt werden. Eine besondere Form des Spartransformators ist der *Stelltrans-*

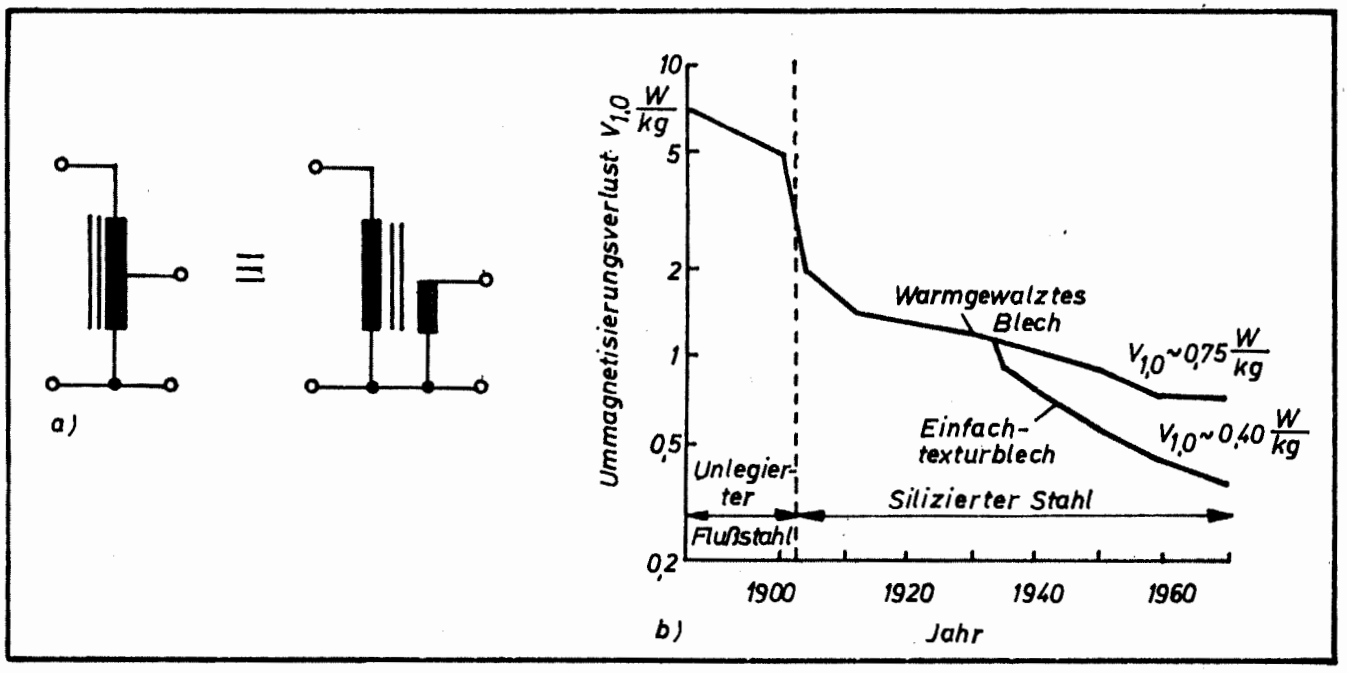


Bild 1.6.4. a) Zur Erklärung des Spartransformators
b) Verringerung der Ummagnetisierungsverluste für Transformatorenblech im Verlauf der zeitlichen Entwicklung

formator. Er ist auf einen Ringkern gewickelt und kann daher frei wählbar an jeder Stelle mit einer Spezialkohle angezapft werden. In Bild 1.6.4b) ist dargestellt, wie im Laufe der Zeit die Eisenverluste von Leistungsübertragern und Transformatoren gesenkt werden konnten.

1.6.7. Variometer

Werden zwei Spulen so angeordnet, daß ihre Kopplung zwischen +1 über Null nach −1 geändert werden kann, und werden diese beiden Spulen durch Reihenschaltung zu einer Induktivität vereinigt, so stellen sie eine stetig veränderbare Induktivität dar. In der Stellung ohne Kopplung addieren sich einfach die Induktivitäten, während in den anderen Stellungen gemäß Bild 1.5.19j) je nach Kopplungsrichtung die doppelte Gegeninduktivität addiert bzw. subtrahiert wird:

$$L_{\text{ges}} = L_1 + L_2 \pm M = L_1 + L_2 \pm k\sqrt{L_1 L_2}\,. \tag{28}$$

2. Lineare Wechselstromschaltungen

2.1. Einführung

Aus den drei passiven Bauelementen R, C, L lassen sich eine Vielzahl von Schaltungen aufbauen. Bei diesen Schaltungen interessiert vor allem der Frequenzgang verschiedener Parameter. Am übersichtlichsten sind die Verhältnisse, wenn ein Zweipol gemäß Bild 2.1.1a) ausgewählt wird. Bei ihm besteht zwischen Spannung $\boldsymbol{u}$ und Strom $\boldsymbol{i}$ ein von der Frequenz abhängiger Widerstand $\boldsymbol{R}$, wobei zwischen Strom und Spannung sowohl in Amplitude als auch in Phase eine Frequenzabhängigkeit besteht. Es wird deshalb eine *komplexe Schreibweise* eingeführt,

$$\boldsymbol{R} = \boldsymbol{u}/\boldsymbol{i} = a + \mathrm{j}b\,. \tag{1}$$

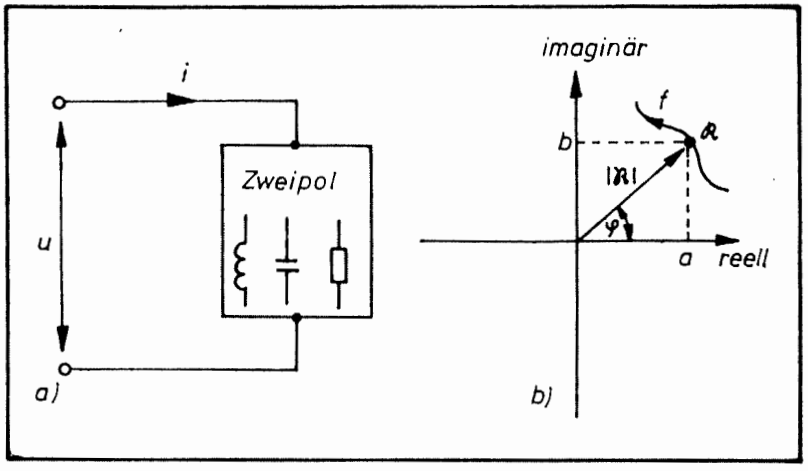


Bild 2.1.1. Zur Definition a) eines allgemeinen Zweipols und b) der Ortskurve

Ein *komplexer Widerstand* entspricht somit genau einem Punkt in der GAUSSschen Ebene. In der Regel ändert er mit der Frequenz seinen Wert und damit seinen *Ort in der* GAUSS*schen Ebene*. Aus diesem Grund besitzt also jeder Zweipol eine Ortskurve (Bild 2.1.1b)). Aus den Größen a und b lassen sich der Betrag des Widerstandes

$$|\boldsymbol{R}| = \frac{|\boldsymbol{u}|}{|\boldsymbol{i}|} = \sqrt{a^2 + b^2} \tag{2}$$

und der Phasenwinkel φ zwischen Strom und Spannung gemäß

$$\tan\varphi = \frac{b}{a} \tag{3}$$

berechnen.

Allgemein bestehen *für die Zweipole folgende Fragestellungen*:

1. Welche Ortskurve besitzt eine spezielle Schaltung?
2. Wie kann man eine gewünschte Ortskurve durch eine Schaltung aus Bauelementen möglichst genau realisieren?
3. Welche Ortskurven sind prinzipiell möglich?

Die erste Frage ist ein *meßtechnisches* oder theoretisch-*analytisches* Problem und damit in jedem Falle zumindest prinzipiell lösbar. Sie wird im folgenden an Beispielen behandelt. Die zweite Frage betrifft die *Synthese* von Schaltungen. Sie wird im Abschnitt 2.5. am Beispiel der Entwicklung spezieller Filter betrachtet. Die dritte Frage ist genereller Natur und führt zur *allgemeinen Systemtheorie* der Elektrotechnik. Auf sie wird im Abschnitt 2.4. eingegangen.

Größere Bedeutung als die Zweipole haben die Vierpole in der Elektrotechnik. In reeller Schreibweise wurden sie bereits im Abschnitt 1.1.5. eingeführt. Die dortigen Betrachtungen gelten auch hier, müssen jedoch ins Komplexe übertragen werden, und darüber hinaus sind sie frequenzabhängig. Für jede der Kenngrößen des Vierpols kann damit jeweils eine Ortskurve gemäß Bild 2.1.1 b) eingeführt werden. Dies wird im Abschnitt 2.3. behandelt, und es gelten analog wieder die drei obigen Fragen.

2.2. Einfache Zweipole

2.2.1. Grundbauelemente

Gemäß Abschnitt 1.2. und 1.3. gilt für die Kapazität

$$i = C\frac{\mathrm{d}u}{\mathrm{d}t} \tag{4}$$

bzw. für die Induktivität

$$u = L\frac{\mathrm{d}i}{\mathrm{d}t}. \tag{5}$$

Wird die Spannung nun harmonisch, gemäß

$$u = \hat{U}_0\, \mathrm{e}^{\mathrm{j}\omega t} \tag{6}$$

angesetzt, so folgen die komplexen Widerstände zu

$$\boxed{R_c = \frac{1}{\mathrm{j}\omega C}} \tag{7}$$

bzw.

$$\boxed{R_L = \mathrm{j}\omega L}\,. \tag{8}$$

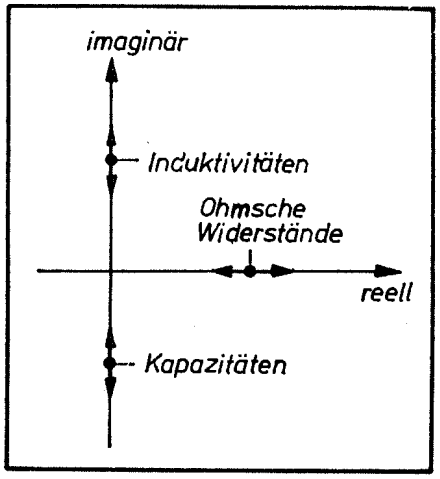


Bild 2.2.1. Zur Lage der Ortskurvenpunkte von rein ohmschem Widerstand, verlustloser Induktivität bzw. verlustloser Kapazität

Hierin sind automatisch die im Zusammenhang mit Bild 1.5.5 eingeführten *Phasenverschiebungen zwischen Strom und Spannung* enthalten. Weiter ist ersichtlich, daß alle Induktivitäten (ohne Verluste) auf der positiv imaginären Achse und alle verlustfreien Kapazitäten auf der negativ imaginären Achse liegen (Bild 2.2.1). Die ohmschen Widerstände liegen auf der positiv reellen Achse. Zur negativ reellen Achse gehören keine passiven Bauelemente. Derartige Elemente besitzen einen negativen reellen Widerstand. Seine Realisierungsmöglichkeiten werden mit der Rückkopplung im Abschnitt 7.1. behandelt.

2.2.2. Reihenschwingkreis

Entsprechend Bild 2.2.2a) liege eine Reihenschaltung aus Widerstand, Kapazität und Induktivität vor. Da auch für die komplexen Widerstände dieselben Rechenregeln wie für reelle Widerstände gelten, ergibt sich

$$\boldsymbol{R} = R + \mathrm{j}\omega L + \frac{1}{\mathrm{j}\omega C} = R + \mathrm{j}\left(\omega L - \frac{1}{\omega C}\right). \tag{9}$$

Der Realteil dieser Schaltung ist also unabhängig von der Frequenz und gleich R. Der *Imaginärteil* ist dagegen stark von der Frequenz abhängig, und es gibt eine Frequenz, für die er *Null* ist:

$$\boxed{\omega_0 = \frac{1}{\sqrt{LC}}}\,. \tag{10}$$

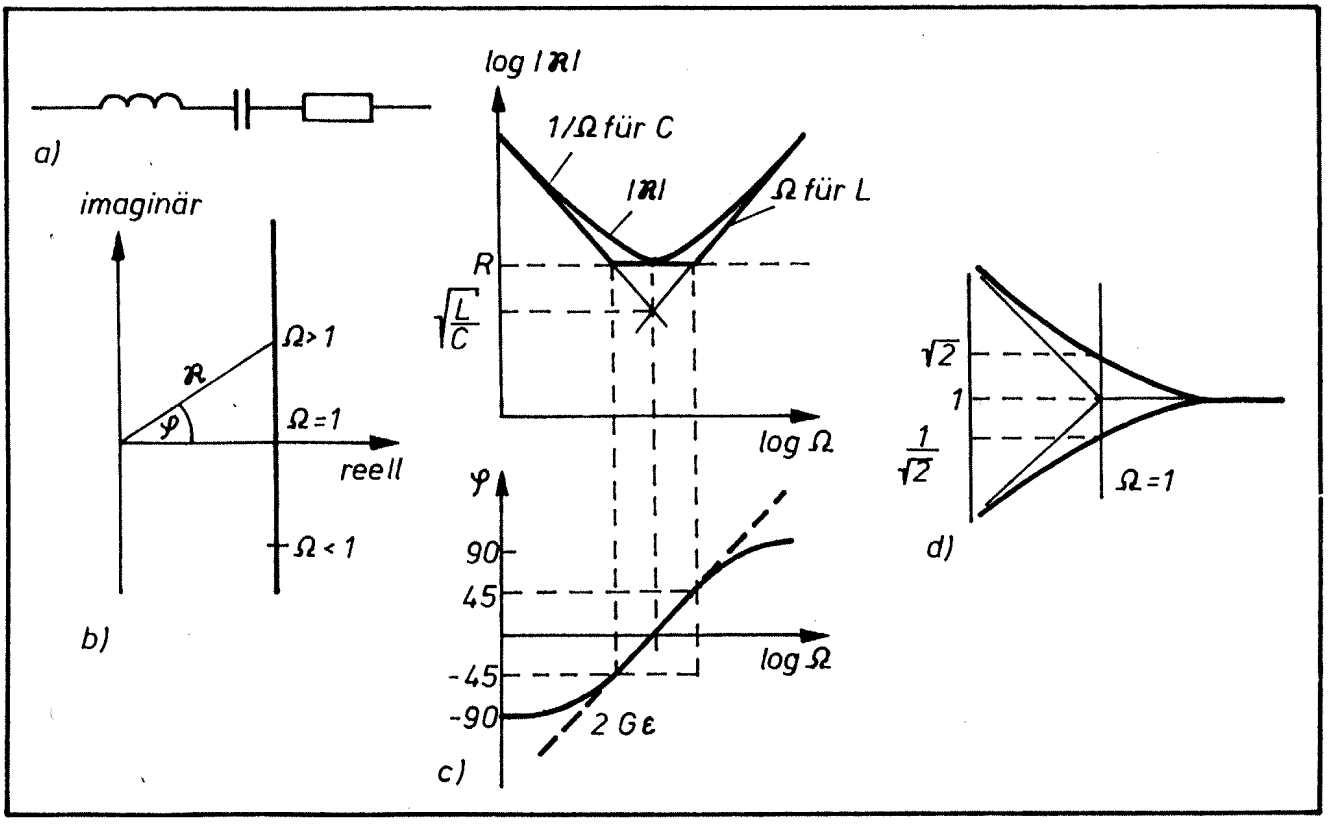


Bild 2.2.2. Zum Reihenschwingkreis

a) Schaltung
b) Ortskurve
c) Verlauf des Betrages (oben) und des Phasenwinkels (unten)
d) Zur Definition der Fehler bei Knickkennlinien

Diese Frequenz heißt *Resonanzfrequenz*. Die Ortskurve des Reihenschwingkreises ist folglich eine senkrechte Gerade durch den reellen Punkt R (Bild 2.2.2b)).

Zur Verallgemeinerung ist es sinnvoll, alle Frequenzen hierauf zu normieren. Für die *normierte Frequenz* gilt dann

$$\boxed{\Omega = \frac{\omega}{\omega_0} = \omega\sqrt{LC}} \,. \tag{11}$$

Bei $\Omega = 1$ liegt der Ortskurvenpunkt genau auf der reellen Achse.

Für $\Omega > 1$ wirkt die Reihenschaltung der drei Widerstände überwiegend induktiv und für $\Omega < 1$ kapazitiv. Hieraus ist wiederum ersichtlich, daß Induktivität und Kapazität zueinander duale Größen sind. Bei der Frequenz $\Omega = 1$ kompensieren sich beide vollständig. Als weitere *Normierungsgröße* für Gl. (11) ist der reelle Widerstand R geeignet. So entsteht

$$\boxed{\boldsymbol{r} = \frac{\boldsymbol{R}}{R} = 1 + \mathrm{j}G\left(\Omega - \frac{1}{\Omega}\right)} \,. \tag{12}$$

Hierin ist nur noch

$$G = \frac{1}{R}\sqrt{\frac{L}{C}} \tag{13}$$

von den spezifischen Werten der Schaltung abhängig. Dieser Ausdruck bestimmt auch nur die Verteilung der Frequenzen auf der Ortskurve. Je größer er ist, desto schneller durchläuft die Ortskurve die reelle Achse. Daher wird der Wert G *Phasensteilheit* bzw. *Güte* genannt.

Der *Betrag des Widerstandes* ist durch

$$|\boldsymbol{R}| = \sqrt{R^2 + \frac{L}{C}\left(\Omega - \frac{1}{\Omega}\right)^2} \tag{14}$$

gegeben und in Bild 2.2.2c) dargestellt. Er hat ein Minimum mit $|\boldsymbol{R}| = R$ bei $\Omega = 1$. Für große und kleine Frequenzen kann Gl. (14) durch einfache *Beziehungen genähert* werden:

$$\begin{aligned} |\boldsymbol{R}| &\approx \sqrt{\frac{L}{C}}\,\Omega \quad \text{für} \quad \Omega \gg 1 \,, \\ &\approx R \quad \text{für} \quad \Omega \approx 1 \,, \\ &\approx \sqrt{\frac{L}{C}}\,\frac{1}{\Omega} \quad \text{für} \quad \Omega \ll 1 \,. \end{aligned} \tag{15}$$

Die drei Näherungen ergeben die drei Geraden in Bild 2.2.2c), aus denen die stark ausgezogene Knickkennlinie zu konstruieren ist. Die wirkliche Kennlinie nähert sich stets den Knickkennlinien recht gut. Der maximale Fehler befindet sich gerade am Knickpunkt und beträgt dort $\sqrt{2}$ statt 1 (bzw. für spätere Fälle $1/\sqrt{2}$ statt 1), wie es in Bild 2.2.2d) dargestellt ist. Aus der relativen Lage von R zu $\sqrt{L/C}$ wird auch die Güte verständlich. Je tiefer $\sqrt{L/C}$ unterhalb von R liegt, desto flacher wird das Minimum. Je höher es über R liegt, desto steiler wird das Minimum.

Für den *Phasenwinkel* gilt wegen (12)

$$\tan\varphi = \left(\Omega - \frac{1}{\Omega}\right). \tag{16}$$

Die Auswertung zeigt Bild 2.2.2c) unten. In der Umgebung von $\Omega = 1$ gilt als Näherung

$$\tan \varphi \approx 2G\varepsilon \quad \text{mit} \quad \Omega = 1 + \varepsilon \,. \tag{17}$$

Bei den Knickpunkten in Bild 2.2.2c) liegt gerade ein Phasenwinkel von 45° vor. Bis hierher gilt Gl. (17) als brauchbare Näherung. Sofern $\sqrt{L/C}$ größer als R ist, bleibt nur die Phasensteilheit als Näherung. Die Darstellung von Bild 2.2.2c) wird auch als Darstellung in logarithmischen Frequenzen oder als BODE-Diagramm bezeichnet.

2.2.3. Einschwingvorgänge

Für *RC*- bzw. *RL*-Glieder wurden Einschwingvorgänge im Abschnitt 1.5.2. bzw. 1.5.3. bereits behandelt. Dies sei jetzt für den Reihenschwingkreis fortgesetzt. In Bild 2.2.3a) ist die entsprechende Schaltung dargestellt. Der Schalter wird zum Zeitpunkt $t = 0$ auf die Spannung U umgeschaltet. Entsprechend den Formeln von Kapitel 1. gilt dann bezüglich der Ladung Q die Differentialgleichung

$$L\frac{\mathrm{d}^2Q}{\mathrm{d}t^2} + R\frac{\mathrm{d}Q}{\mathrm{d}t} + \frac{Q}{C} = U \,. \tag{18}$$

Zur Lösung wird der Ansatz

$$Q = A\,\mathrm{e}^{pt} \tag{19}$$

gewählt. Für die dann in p quadratische Differentialgleichung stellt

$$p_{1,2} = -\frac{R}{2L} \pm \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} \tag{20}$$

eine Lösung dar. Die vollständige Lösung ergibt sich mit den Konstanten A_1 und A_2 sowie C_1 und C_2 aus den Anfangsbedingungen zu

$$Q = A_1\,\mathrm{e}^{p_1 t} + A_2\,\mathrm{e}^{p_2 t}\,, \qquad I = C_1\,\mathrm{e}^{p_1 t} + C_2\,\mathrm{e}^{p_2 t}\,. \tag{21}$$

Hierfür sind drei Fälle zu unterscheiden:

— *überaperiodische Dämpfung.* Wegen

$$\left.\begin{aligned} &\frac{R^2}{4L^2} > \frac{1}{LC} \qquad \text{oder} \\ &R\sqrt{\frac{C}{L}} > 2 \quad \text{oder mit (13)} \quad G < \frac{1}{2} \end{aligned}\right\} \tag{22}$$

sind beide p reell und negativ, jedoch verschieden groß, und weil für $t = 0$ kein Strom fließt, folgt

$$C_1 - C_2 = 0 \,. \tag{23}$$

Der unterschiedliche Zeiteinfluß in den Exponenten bewirkt, daß der Strom zunächst ein Maximum erreicht und danach asymptotisch auf Null abklingt (Bild 2.2.3c)).

Das langsame Einschwingen der Kondensatorspannung auf den Endwert U zeigt Bild 2.2.3g).

— *aperiodischer Grenzfall.* Hier gilt

$$\boxed{\frac{R^2}{4L^2} = \frac{1}{LC} \quad \text{bzw.} \quad G = \frac{1}{2}}\,, \tag{24}$$

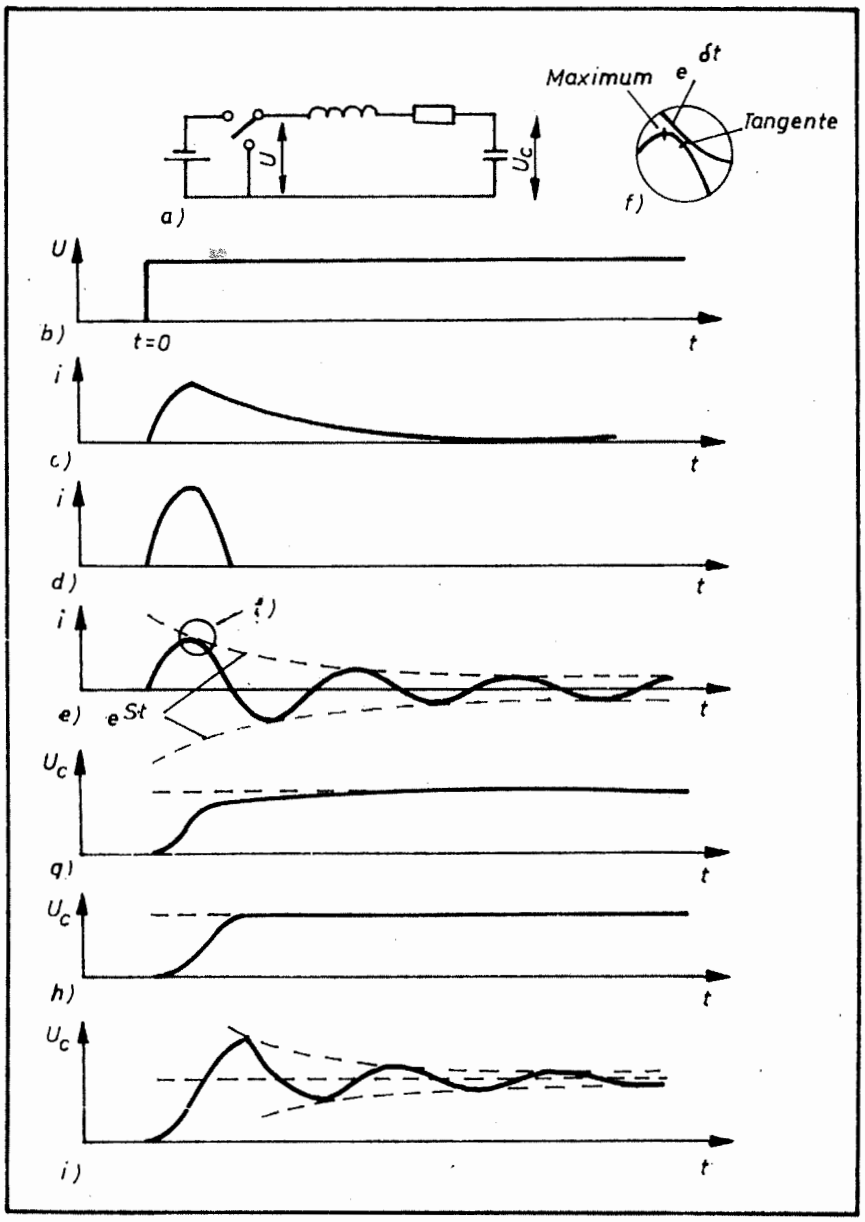


Bild 2.2.3. Einschaltvorgang beim Reihenschwingkreis

a) Schaltbild
b) Rampenfunktion bei Betätigung des Schalters
c) Stromverlauf bei überperiodischer Dämpfung
d) Stromverlauf bei aperiodischer Dämpfung
e) Stromverlauf der gedämpften Schwingung
f) Vergrößerter Ausschnitt von e) in der Umgebung eines Maximums
g) Spannungsverlauf am Kondensator bei überperiodischer Dämpfung
h) Spannungsverlauf am Kondensator bei aperiodischer Dämpfung
i) Spannungsverlauf am Kondensator bei gedämpfter Schwingung

und daher fallen beide p zusammen und sind gleich $-\frac{R}{2L}$. Das An- und Abklingen des Stromes verläuft damit gleichartig reziprok. Hierdurch wird der stromlose Zustand besonders schnell erreicht, da der Strom nach seinem Maximum in Null einmündet (Bild 2.2.3d)). Den Spannungsverlauf zeigt Bild 2.2.3h).

— *freie Schwingung.* In den meisten Fällen werden beide p wegen

$$R\sqrt{\frac{C}{L}} < 2 \quad \text{bzw.} \quad G > \frac{1}{2} \tag{25}$$

imaginär in der Form

$$p_{1,2} = \delta \pm \mathrm{j}\bar{\omega} \tag{26}$$

mit

$$\boxed{\delta = -\frac{R}{2L}, \qquad \bar{\omega} = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}}} \tag{27, 28}$$

vorliegen. Damit lautet die Lösung

$$\begin{aligned} Q &= \mathrm{e}^{\delta t}(A_1\,\mathrm{e}^{\mathrm{j}\bar{\omega}t} + A_2\,\mathrm{e}^{-\mathrm{j}\bar{\omega}t}) \\ &= \mathrm{e}^{\delta t}[(A_1 + A_2)\cos\bar{\omega}t + \mathrm{j}(A_1 - A_2)\sin\bar{\omega}t]\,. \end{aligned} \tag{29}$$

Hierbei hat in den eckigen Klammern nur der reelle Anteil einen physikalischen Sinn. Es liegt also eine Kosinusschwingung für die Ladung vor, deren Amplitude mit $\mathrm{e}^{\delta t}$ abnimmt. Aus den Anfangsbedingungen gilt für den Strom

$$I = -Q\,\frac{\bar{\omega}^2 + \delta^2}{\bar{\omega}} \cdot \mathrm{e}^{\delta t}\sin\bar{\omega}t\,. \tag{30}$$

Den Verlauf dieser Schwingung zeigt Bild 2.2.3e). Die Frequenz der freien Schwingung (28) liegt also wegen (27) gemäß

$$\bar{\omega} = \sqrt{\omega_0^2 - \delta^2} \tag{31}$$

etwa um den Wert δ tiefer als die eigentliche Resonanzfrequenz

$$\omega_0 = \frac{1}{\sqrt{LC}}. \tag{32}$$

Die Maxima der freien Schwingung liegen jedoch — wie durch Differentiation von (30) gezeigt werden kann — genau im Abstand

$$T = \frac{2\pi}{\omega_0}. \tag{33}$$

Zur Veranschaulichung dieses Zusammenhanges zeigt Bild 2.2.3f) einen Ausschnitt in der Umgebung des Maximums vergrößert. Bild 2.2.3i) zeigt schließlich noch die Spannung am Kondensator. Bezüglich der Dämpfung gilt

$$-\delta T = \ln\frac{I_1}{I_2} = \vartheta = \pi R\sqrt{\frac{C}{L}} = \frac{\pi}{G}. \tag{34}$$

Darin bedeuten I_1 und I_2 zwei aufeinanderfolgende Amplituden. ϑ wird das logarithmische Dekrement genannt, während G wiederum die Güte des Schwingkreises ist.

2.2.4. Pole und Nullstellen

In der Nachrichtentechnik ist es üblich, die *komplexe Frequenz*

$$p = \mathrm{j}\omega \tag{35}$$

zu verwenden. Hiermit geht (9) über in

$$\boldsymbol{R}(p) = R + pL + \frac{1}{pC} \tag{36a}$$

$$= \frac{p^2L + pR + 1/C}{p}. \tag{36b}$$

Das quadratische Zählerpolynom besitzt die Wurzeln

$$p_{01}; p_{02} = \delta \pm \mathrm{j}\bar{\omega} \tag{37}$$

gemäß (20) bzw. (26). Damit kann (36b) in die Form

$$\boldsymbol{R}(p) = \frac{(p - p_{01})\,(p - p_{02})}{(p - p_{\infty 1})} \tag{38}$$

gebracht werden. Im Nenner wurde die *Nullstelle* $p_{\infty 1} = 0$, die zu einem *Pol* im Frequenzgang führt (Pol), ergänzt. Damit ist der Reihenresonanzkreis durch die zwei Nullstellen und den einen Pol vollständig beschrieben.

Die im vorigen Abschnitt eingeführten drei Fälle stellen sich dabei entsprechend Bild 2.2.4 dar. Im aperiodischen Grenzfall (Teilbild b)) existiert eine doppelte reelle und negative Nullstelle bei $-R/2L$. Im aperiodischen Fall, also bei starker Dämpfung

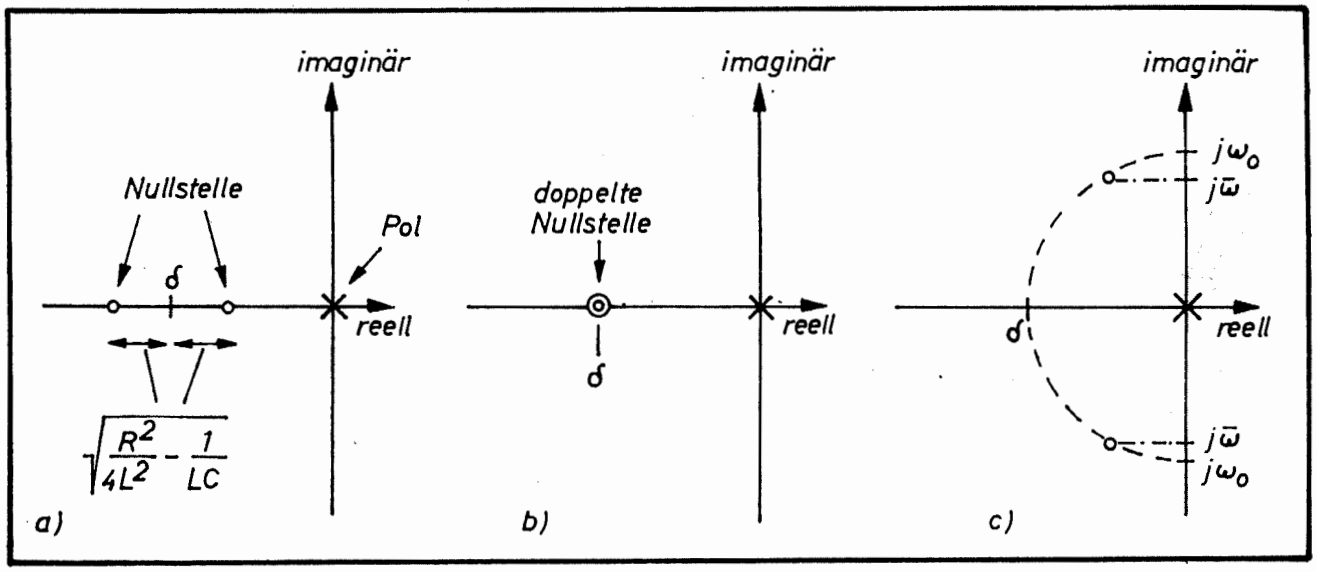


Bild 2.2.4. Lage von Pol und Nullstellen des Reihenschwingkreises
a) Überaperiodischer Fall
b) Aperiodischer Grenzfall
c) Freie Schwingung

spaltet sich diese Nullstelle in zwei auf, die jeweils um den im Teilbild a) angezeigten Betrag nach rechts bzw. links verschoben sind. Im Fall der freien Schwingung spaltet sich die doppelte Nullstelle aus Teilbild b) in zwei konjugiert komplexe Nullstellen auf, die auf dem gestrichelten Halbkreis von Teilbild c) liegen.

Es wird sich später im Abschnitt 2.4. zeigen, daß generell alle Zwei- und Vierpole vollständig durch ihre Pole und Nullstellen zu beschreiben sind. Darüber hinaus wird gezeigt werden, daß diese Darstellung auch anschaulich interpretiert werden kann.

2.2.5. Kreisdiagramm

Beim Reihenschwingkreis verläuft die Ortskurve von $-\mathrm{j}\,\infty$ nach $+\mathrm{j}\,\infty$. Bei den meisten Zusammenschaltungen wird ähnlich der unendlich ferne Punkt einbezogen. Dies ist für die Darstellung recht ungünstig. Da andererseits nur die rechte Halbebene verwendet wird, bringt eine Transformation Vorteile, die sie in das Innere eines Einheitskreises abbildet. Sie führt zum sogenannten SMITH-*Diagramm*. Wie jede analytische Transformation ist sie durch zwei Fixpunkte, z. B. 0 und 1, gekennzeichnet. Den Zusammenhang von der GAUSS-Ebene und dem SMITH-Diagramm zeigt Bild 2.2.5. Das SMITH-Diagramm bringt vielfältige Vorteile, u. a. zeigt es auch die Dualität von Widerstand und Leitwert bzw. Reihen- und Parallelschaltung. Dies geht deutlich

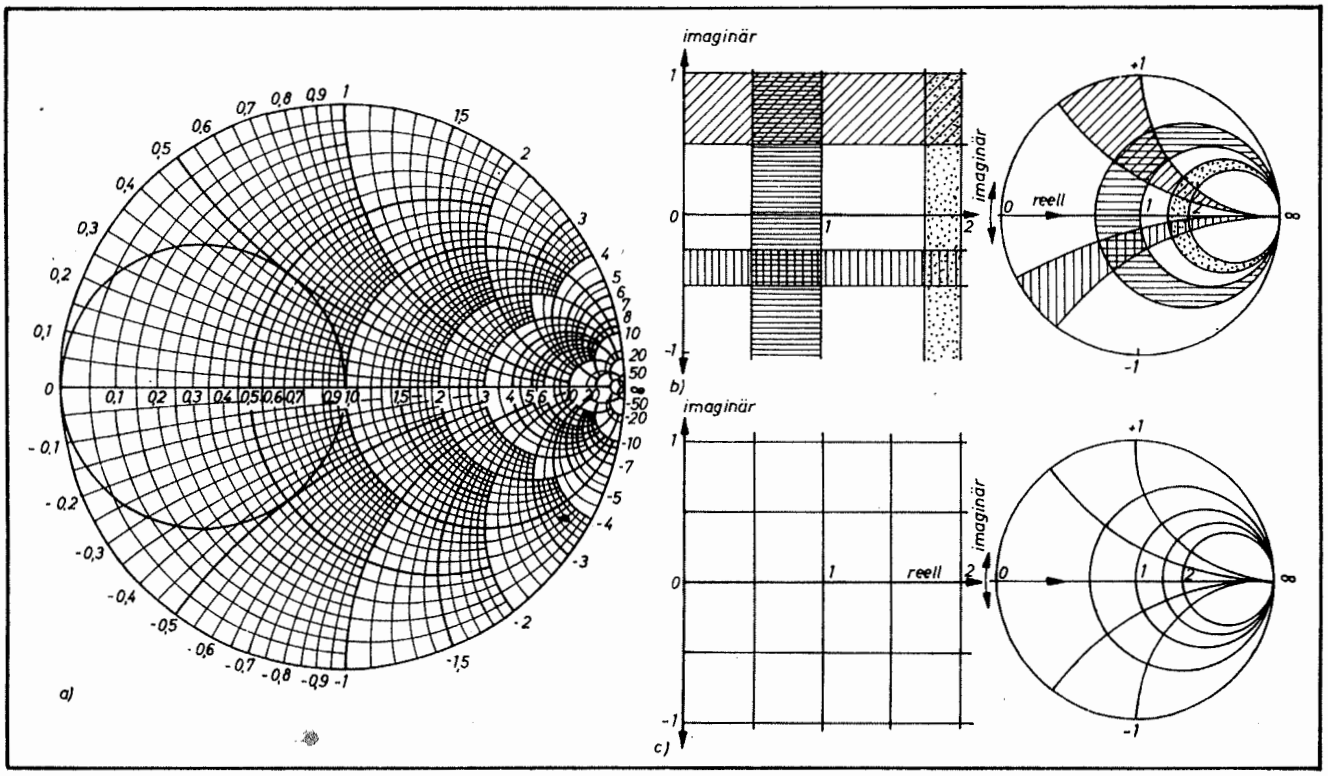


Bild 2.2.5. a) SMITH-Diagramm mit Koordinateneinteilung, Vergleich von b) gleichen Flächen und c) gleichen Linien im SMITH-Diagramm und in der GAUSS-Ebene

aus der Zusammenstellung von Bild 2.2.6 hervor. Der Übergang von der Reihen- zur Parallelschaltung und umgekehrt erfolgt danach vor allem durch das Vertauschen von Null und Unendlich.

2.2.6. Relative Bandbreite

In Bild 2.2.6 sind u. a. die normierten Widerstände des Reihen- und Parallelschwingkreises enthalten. Beiden Gleichungen gemeinsam ist der Teil

$$x = 1 + \mathrm{j}G\left(\Omega - \frac{1}{\Omega}\right). \tag{39}$$

Für den Betrag dieser Beziehung gilt

$$x = \sqrt{1 + G^2\left(\Omega - \frac{1}{\Omega}\right)^2}. \tag{40}$$

Als Bandbreite eines Schwingkreises wird der Abstand jener Frequenzen bezeichnet, für die sich der relative Widerstand auf $\sqrt{2}$ bzw. $1/\sqrt{2}$ verändert hat. Dies führt in (40) zu

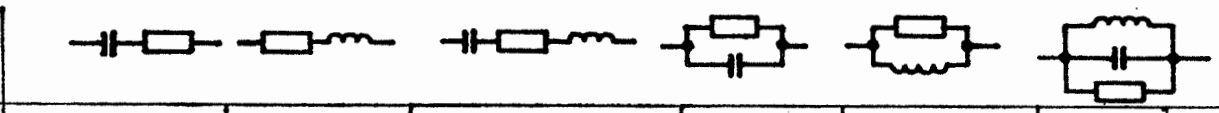
$$x = \sqrt{2}\,. \tag{41}$$

Der rechte quadratische Ausdruck unter der Wurzel in (40) muß also gleich 1 werden. Folglich gilt

$$G\left(\Omega - \frac{1}{\Omega}\right) = \pm 1\,. \tag{42}$$

Hieraus läßt sich die quadratische Gleichung

$$\Omega^2 \mp \frac{1}{G}\,\Omega - 1 = 0 \tag{43}$$



$\Re$	$R+\frac{1}{j\omega C}$	$R+j\omega L$	$R+j\left(\omega L-\frac{1}{\omega C}\right)$	$\frac{R}{1+j\omega CR}$	$\frac{j\omega LR}{R+j\omega L}$	$\frac{1}{\frac{1}{R}+j\left(\omega C-\frac{1}{\omega L}\right)}$
$\lvert\Re\rvert$	$\sqrt{R^2+\frac{1}{\omega^2C^2}}$	$\sqrt{R^2+\omega^2L^2}$	$\sqrt{R^2+\left(\omega L-\frac{1}{\omega C}\right)^2}$	$\frac{R}{\sqrt{1+\omega^2C^2R^2}}$	$\frac{\omega LR}{\sqrt{R^2+\omega^2L^2}}$	$\frac{1}{\sqrt{\frac{1}{R^2}+\left(\omega C-\frac{1}{\omega L}\right)^2}}$
$\tan\varphi$	$\frac{1}{\omega CR}$	$\frac{\omega L}{R}$	$\frac{\omega L-\frac{1}{\omega C}}{R}$	$-\omega CR$	$-\frac{R}{\omega L}$	$\frac{\omega L-\frac{1}{\omega C}}{R}$
ω_g, ω_0	$\frac{1}{RC}$	$\frac{R}{L}$	$\frac{1}{\sqrt{LC}}$	$\frac{1}{RC}$	$\frac{R}{L}$	$\frac{1}{\sqrt{LC}}$
Ω	ωRC	$\frac{\omega L}{C}$	$\omega\sqrt{LC}$	ωRC	$\frac{\omega L}{R}$	$\omega\sqrt{LC}$
τ	$1+\frac{1}{\Omega}$	$1+j\Omega$	$1+j\frac{1}{R}\sqrt{\frac{L}{C}}\left(\Omega-\frac{1}{\Omega}\right)$	$\frac{1}{1+j\Omega}$	$\frac{1}{1+\frac{1}{\Omega}}$	$\frac{1}{1+jR\sqrt{\frac{C}{L}}\left(\Omega-\frac{1}{\Omega}\right)}$
$\tan\varphi$	$\frac{1}{\Omega}$	Ω	$\frac{1}{R}\sqrt{\frac{L}{C}}\left(\Omega-\frac{1}{\Omega}\right)$	$-\Omega$	$-\frac{1}{\Omega}$	$R\sqrt{\frac{C}{L}}\left(\Omega-\frac{1}{\Omega}\right)$
Ortskurve						
SMITH-Diagramm						
Knick-Kennlinie						
Phasenwinkel						
PN-Plan						

Bild 2.2.6. Zusammenstellung der wichtigsten Schaltungen komplexen Widerständen. Sie demonstriert zugleich die 7 äquivalenten Beschreibungsmethoden des gleichen Sachverhaltens (a), nähmlich:

a) physikalisches Gebilde = Bauelement
b) Schaltbild
c) komplexer Widerstand $\boldsymbol{R}$
d) Betrag $|\boldsymbol{R}|$ und Phase φ (tan φ)
e) Ortskurve
f) SMITH-Diagramm
g) BODE-Diagramm als Knickkennliniendarstellung für Frequenz- und Phasengang
h) PN-Plan

Die vier nicht hiermit begründeten Spalten oberhalb der Ortskurve betreffen verschiedene Normierungen, die den Zweck haben, das Prinzipielle der jeweiligen Schaltung deutlich hervorzuheben. So betrifft die Grenzfrequenz ω_g den Knickpunkt im BODE-Diagramm bzw. den zugehörigen Phasenwinkel von 45° und ω_0 die Resonanzfrequenz eines Schwingkreises. Die normierte Frequenz Ω wird auf diese Frequenzen bezogen. Die beiden folgenden Spalten betreffen dann den hierauf und auf den Wert 1 normierten komplexen Widerstand bzw. den Phasenwinkel.

Beispielhaft sind die entsprechenden Bezüge für den Schwingkreis in den Abschnitten 2.2.2. bis 2.2.5. behandelt.

ableiten. Es seien Ω_1 und Ω_2 die beiden Lösungen. Ihre Differenz ergibt dann die relative Bandbreite b gemäß

$$\Omega_1 - \Omega_2 = \boxed{b = \frac{B}{f_0} = \frac{1}{G}} \,. \tag{44}$$

Sie ist also nur von der Güte des Schwingkreises abhängig.

2.2.7. Ergänzungen zu realen Bauelementen

Für eine reale Induktivität galt (vgl. 1.5.6., (1.5.35))

$$R_v = R_{Cu} + a\omega + b\omega^2 \,. \tag{45}$$

Hierbei sind die Verluste an die Anschlußklemmen transformiert. Es ist jedoch auch möglich, sie auf den magnetischen Widerstand zu beziehen. Dabei gilt (1.5.10) und (1.5.20):

$$L = \frac{n^2}{P} = \frac{n^2 A}{l}\mu \,. \tag{46}$$

In diesem Sinne sind die Verluste dann durch eine komplexe Permeabilität ausdrückbar. Für den induktiven Widerstand mit Verlusten gilt

$$\begin{aligned} \boldsymbol{R}_L &= \mathrm{j}\omega L + R_v \\ &= \mathrm{j}\omega n^2 \frac{A}{l}\mu_w + n^2(a_1 + a_2\omega + a_3\omega^2) \\ &= \mathrm{j}\omega n^2 \frac{A}{l}[\mu_w - \mathrm{j}(\alpha + \beta\omega + \gamma\omega^2)] \,. \end{aligned} \tag{47}$$

In der eckigen Klammer steht die sogenannte *komplexe Permeabilität* des Kerns bzw. des magnetischen Materials. Es ist üblich, auch sie durch eine *Ortskurve* zu beschreiben. Sie wird dann in der

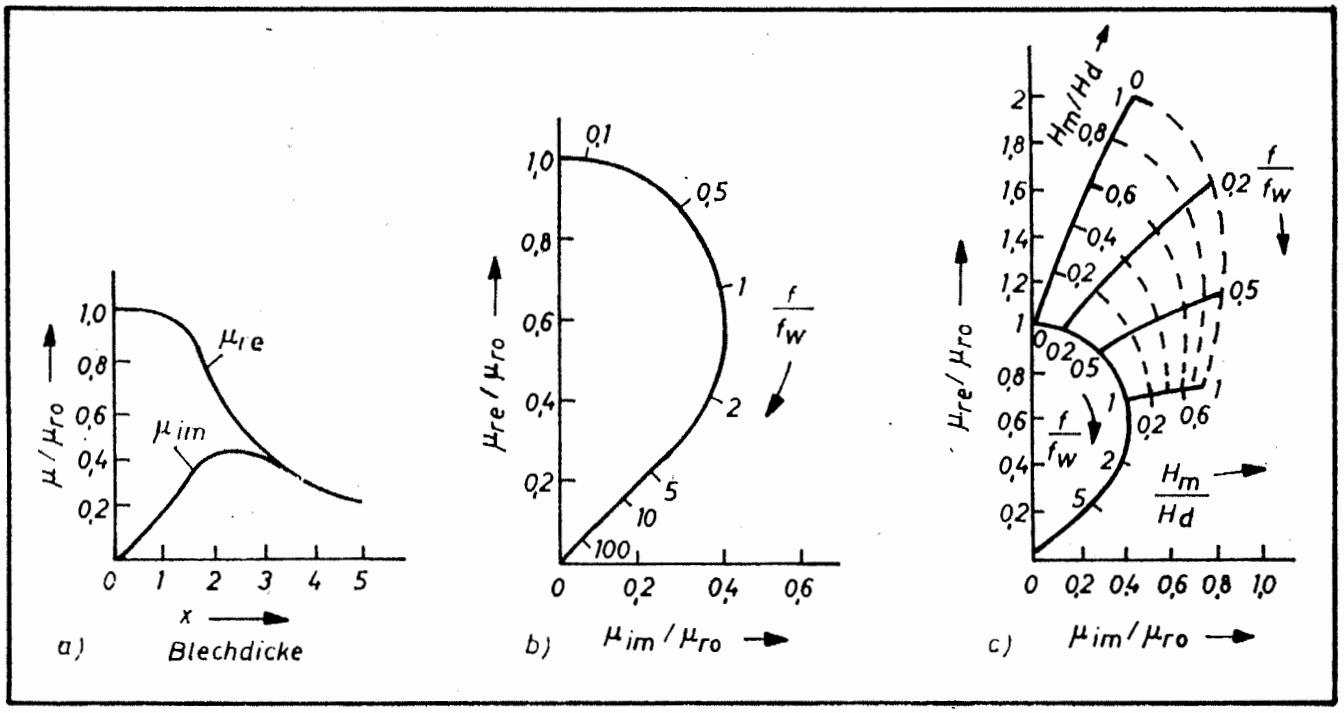


Bild 2.2.7. Zur komplexen Permeabilität

a) Abhängigkeit des reellen und imaginären Anteils von der Blechdicke bezüglich der Wirbelstromverluste
b) Beispiel einer Ortskurve mit f_w als WOLMAN-Frequenz
c) Unter Berücksichtigung der Hystereseverluste als Funktion der relativen Aussteuerung H_m/H_d

Form

$$\mu_r = \mu_{re} - j\mu_{im} \tag{48}$$

geschrieben und wird in der Regel in einer Ebene dargestellt, die um 90° gegenüber der GAUSS-Ebene gedreht ist (Bild 2.2.7). Die Frequenz, bei welcher der Betrag von μ_r auf $1/\sqrt{2}$ abgefallen ist, heißt die WOLMAN-*Grenzfrequenz*. Da die Hystereseverluste von der Aussteuerung des Materials abhängen, ist die Ortskurve der komplexen Permeabilität ebenfalls aussteuerungsabhängig.

Die *Verluste von Kondensatoren* können zwar im Prinzip ebenfalls durch eine komplexe Dielektrizitätskonstante beschrieben werden, bei ihnen ist dies jedoch nicht üblich. Hier werden die Verluste unmittelbar in der Ortskurve erfaßt.

2.3. Komplexe Spannungsteiler

2.3.1. Einfache frequenzselektive Teiler

Ein komplexer Spannungsteiler ist entsprechend Bild 2.3.1 links oben durch zwei komplexe Widerstände, den Längswiderstand r_l und den Querwiderstand r_q gegeben. Die Eigenschaften des Spannungsteilers hängen wesentlich von den Werten der komplexen Widerstände ab. Für sein Übertragungsmaß gilt die Beziehung

$$K = \frac{u_a}{u_e} = \frac{r_q}{r_l - r_q}. \tag{1}$$

Es ist wiederum eine komplexe Größe und kann damit ebenfalls als Ortskurve in der GAUSSschen Ebene dargestellt werden.

Eine Zusammenstellung für die einfachen Spannungsteiler aus Widerständen und Kondensatoren zeigt mit ihren typischen Eigenschaften und Bezug auf die gleichartigen Ortskurven bei komplexen Widerständen Bild 2.3.1. Daraus ist ersichtlich,

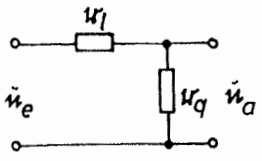
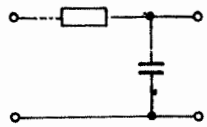
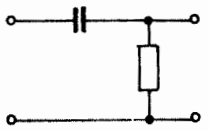
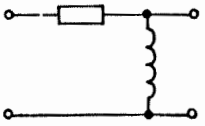
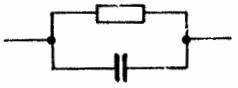
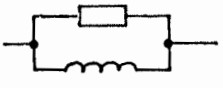
	Tiefpaß		Hochpaß	
				
K	$\frac{1}{1 + j\omega CR}$	$\frac{1}{1 + j\omega \frac{L}{R}}$	$\frac{1}{1 + \frac{1}{j\omega CR}}$	$\frac{1}{1 + \frac{R}{j\omega L}}$
Sperrbereich	$\omega > \frac{1}{RC}$	$\omega > \frac{R}{L}$	$\omega < \frac{1}{RC}$	$\omega < \frac{R}{L}$
K	$\frac{1}{1 + j\Omega}$		$\frac{1}{1 + \frac{1}{j\Omega}}$	
Sperrbereich	$\Omega > 1$		$\Omega < 1$	
entspricht Zweipol				

Bild 2.3.1. Schaltungen und Daten einfacher komplexer Spannungsteiler sowie Vergleich mit frequenzäquivalenten Zweipolen

Integrierglied

$\breve{u} \rightarrow \boxed{\int} \rightarrow \int \breve{u}\,dt$

$\breve{u}_a = \frac{1}{j\Omega}\hat{u}_0 e^{j\Omega t}$

$\boldsymbol{K} = \frac{1}{j\Omega}$

Tiefpaß für $\Omega \gg 1$

Differenzierglied

Ansatz $\breve{u} = \hat{u}_0 e^{j\Omega t}$

$\breve{u} \rightarrow \boxed{d} \rightarrow \frac{d\breve{u}}{dt}$

$\breve{u}_a = j\Omega\hat{u}_0 e^{j\Omega t}$

$\boldsymbol{K} = j\Omega$

Hochpaß für $\Omega \ll 1$

Verallgemeinerung

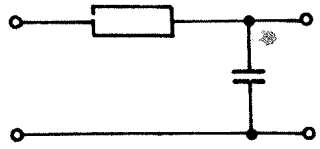
$|\boldsymbol{K}| \sim \frac{1}{\Omega}$

$\varphi = -90°$

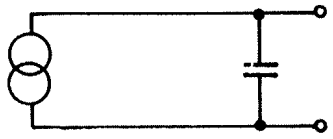
$|\boldsymbol{K}| \sim \Omega$

$\varphi = +90°$

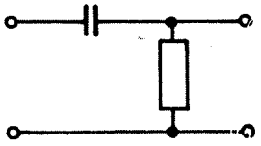
reale *RC*-Pässe



$\Omega \ll 1$ bedeutet $R \rightarrow \infty$
also Stromquelle wirkt auf C



Strom → Spannung am Kondensator



$\Omega \ll 1$ bedeutet $R \rightarrow 0$
Ausgangsspannung sehr klein, nur Ausgangsstrom wichtig

Bild 2.3.2. Zur differenzierenden und integrierenden Wirkung von *RC*-Gliedern

daß diese Spannungsteiler in *Hoch- und Tiefpässe* zu teilen sind. Die jeweilige Grenzfrequenz ist besonders deutlich aus der *Knickliniendarstellung* (Bild 2.2.6) zu ersehen. Sie trennt den *Durchlaß- vom Sperrbereich.* Die vollständige Analogie zur Parallelschaltung von Widerständen hat trotz der anderen physikalischen Eigenschaften zur Folge, daß immer häufiger in integrierten Schaltungsanordnungen mittels dieser Spannungsteiler-Schaltungen und aktiver Verstärkeranordnungen Bauelemente synthetisiert werden.

Im Sperrbereich weisen Spannungsteiler von Bild 2.3.1 ein spezielles Verhalten auf. Bezüglich des Verhältnisses von Ein- und Ausgangsspannung besitzen sie hier integrierendes bzw. differenzierendes Verhalten. Dieser Zusammenhang ist in Bild 2.3.2 dargestellt. Daraus folgt u. a., daß auch eine beliebige Frequenzgangkurve immer dort *integrierendes* bzw. *differenzierendes Verhalten* zeigt, wo die Phase um $-90°$ bzw. $+90°$ verschoben wird und ein Frequenzgang umgekehrt proportional bzw. proportional der Frequenz existiert. Aus dem Bild ist auch ersichtlich, daß dieses Verhalten wesentlich auf die Stromspannungsbeziehung bei einer Kapazität zurückgeht. Die entsprechenden Betrachtungen für die Induktivität wurden wegen ihrer geringen praktischen Bedeutung fortgelassen. Bei diesen *RC*-Gliedern ist im Sperrbereich die Ausgangsgröße stets sehr klein gegenüber der Eingangsgröße. Der Fehler

für die Integration bzw. Differentiation entspricht somit etwa der relativen Größe der Ausgangsspannung. Für ideale Integratoren bzw. Differentiatoren kann sie am Eingang kompensiert werden.

2.3.2. Schwingkreisspannungsteiler

Ein Spannungsteiler mit besonderen Eigenschaften besteht aus Induktivität, Widerstand und Kapazität (Bild 2.3.3). Die Resonanzfrequenz des Serienkreises führt bei ihm zu einem speziellen Verhalten. Mit den normierten Größen

$$\omega_0 = \frac{1}{\sqrt{LC}}; \qquad \Omega = \frac{\omega}{\omega_0}; \qquad G = \frac{1}{R}\sqrt{\frac{L}{C}} \tag{2}$$

gelten die Beziehungen

$$\left.\begin{aligned} \boldsymbol{K} &= \frac{1}{1 - \Omega^2 + \mathrm{j}\frac{1}{G}\Omega}, \\ |\boldsymbol{K}| &= \frac{1}{\sqrt{1 + \left(\frac{1}{G^2} - 2\right)\Omega^2 + \Omega^4}}, \qquad \tan\varphi = \frac{-\frac{1}{G}\Omega}{1 - \Omega^2}. \end{aligned}\right\} \tag{3}$$

Ihre Auswertung zeigt ebenfalls Bild 2.3.3. Danach ist die maximale Ausgangsspannung etwa um den Gütefaktor G größer als die Eingangsspannung. Das genaue *Maximum* liegt bei

$$\Omega_{\max} = \sqrt{1 - \frac{1}{2G^2}}, \tag{4}$$

also etwas *tiefer als* die *Resonanzfrequenz* des Schwingkreises.

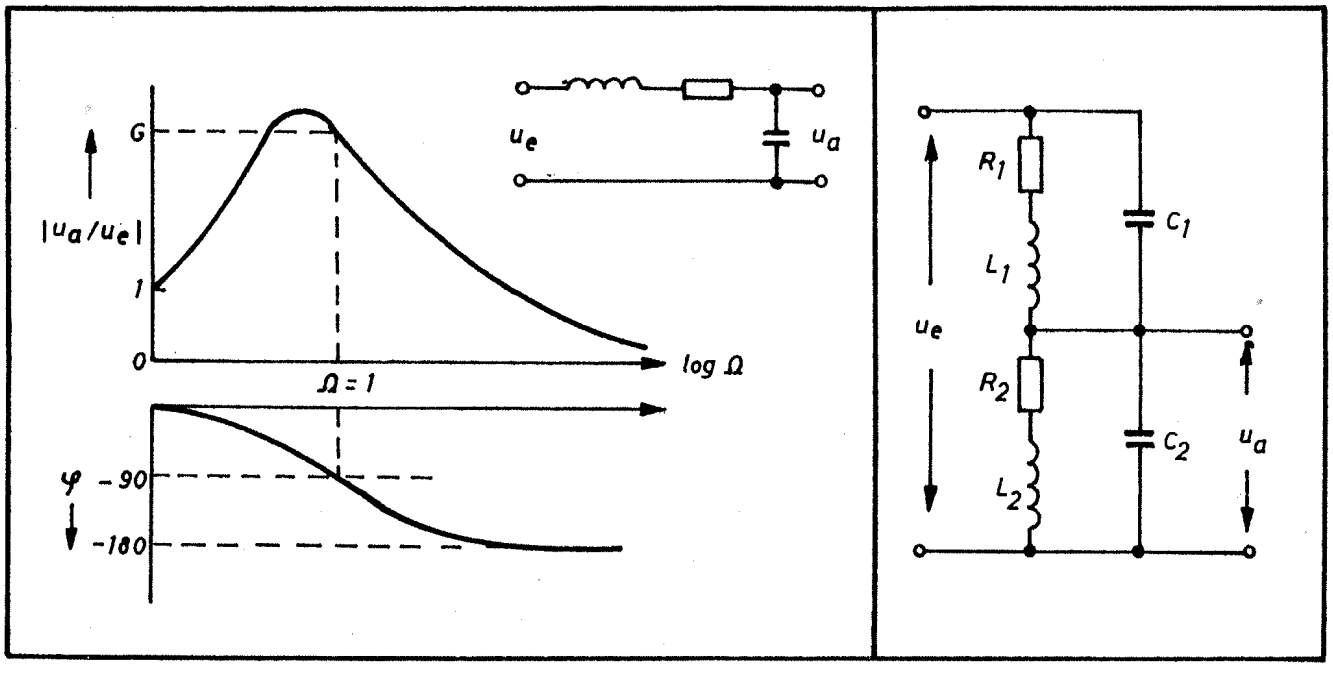


Bild 2.3.3. Schaltung, Frequenz- und Phasengang des Schwingkreisspannungsteilers

Bild 2.3.4. Schema des frequenzunabhängigen Teilers

Die Phasenverschiebung von $-90°$ liegt dagegen genau bei der Resonanzfrequenz. Die Maximalüberhöhung beträgt

$$|\boldsymbol{K}_{\max}| = \frac{G}{1 - \frac{1}{4G^2}} \approx G\left(1 - \frac{1}{8G^2}\right). \tag{5}$$

Bei der Resonanzfrequenz ist die *Überhöhung* genauso groß wie die *Güte* des Schwingkreises. Mit dem Schwingkreisspannungsteiler kann also eine Spannungsverstärkung erreicht werden. Da jedoch keine Energieverstärkung erfolgt, sind die Ausgangsströme stets entsprechend kleiner als die Eingangsströme. Die Ausgangsströme erzeugen nämlich eine zusätzliche Dämpfung des Schwingkreises und erniedrigen somit die Resonanzüberhöhung. Der Schwingkreisteiler hat in gewisser Hinsicht eine *Ähnlichkeit mit* einem *Transformator* (vgl. 1.6.). Ein Unterschied besteht darin, daß der Spannungsteiler nur für eine Frequenz wirksam ist. Für eine Frequenz wird er daher auch vielfach *zur Anpassung* benutzt. Anwendungen sind z.B. die industrielle Trocknung von Holz im HF-Feld des Kondensators und die Längsinduktivität bei Antennen. Es sei weiter an das Einschwingverhalten dieses Spannungsteilers gemäß Abschnitt 2.2.3. erinnert.

2.3.3. Frequenzunabhängiger Teiler

Da alle Bauelemente sowohl Kapazitäten als auch Induktivitäten enthalten, ist es recht schwierig, einen Spannungsteiler zu bauen, der für alle Frequenzen gleichermaßen wirksam ist. Hierzu ist eine Anordnung nach Bild 2.3.4 günstig. Sie zeigt für alle Frequenzen eine gleichbleibende Teilung, sobald

$$\frac{R_1}{R_2} = \frac{L_1}{L_2} = \frac{C_2}{C_1} \tag{6}$$

erfüllt ist. Für die meisten Anwendungen ist es möglich, die Induktivität zu vernachlässigen, dann ist nur einer der beiden Kondensatoren gemäß dem Widerstandsverhältnis abzugleichen.

2.3.4. *RC*-Siebglieder

In den vierziger Jahren wurde es üblich, frequenzselektive Spannungsteiler ohne Induktivitäten in vielfältiger Form zu benutzen. Ein typisches Bauelement dieser Art ist das WIEN-*Glied* gemäß Bild 2.3.5a). Mit den Größen

$$\omega_0^2 = \frac{1}{R_1 R_2 C_1 C_2}; \qquad \Omega = \frac{\omega}{\omega_0}; \qquad r = \frac{R_1}{R_2}; \qquad c = \frac{C_2}{C_1} \tag{7}$$

ergibt sich sein normiertes Übertragungsmaß zu

$$\begin{aligned} \frac{1}{\boldsymbol{K}} &= (1 + r + c)\left[1 + \mathrm{j}\,\frac{rc}{1 + r + c}\left(\Omega - \frac{1}{\Omega}\right)\right] \\ &= v\left[1 + \mathrm{j}G\left(\Omega - \frac{1}{\Omega}\right)\right]. \end{aligned} \tag{8}$$

In den eckigen Klammern steht der Ausdruck (2.2.12). Die analoge *Güte* des *RC*-WIEN-Gliedes beträgt

$$G = \frac{rc}{1 + r + c}. \tag{9}$$

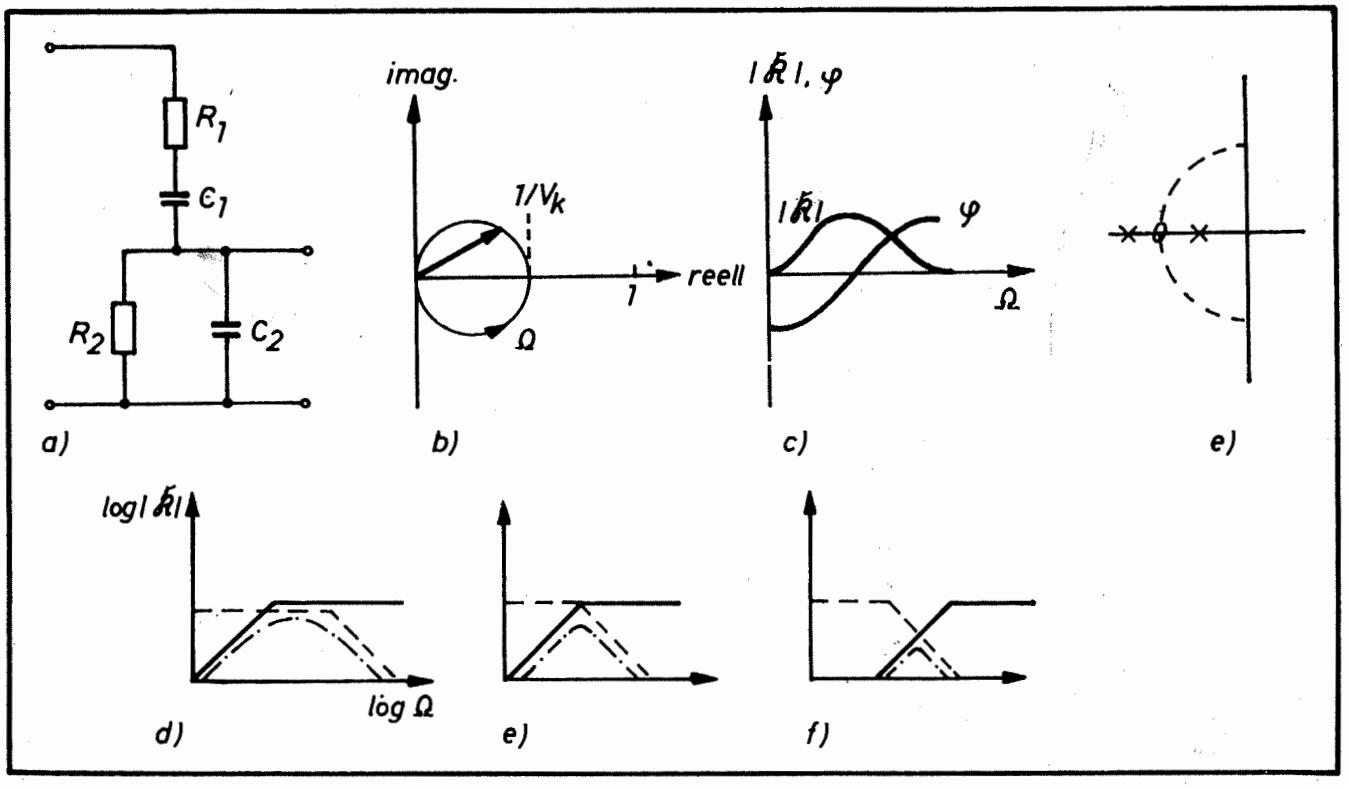


Bild 2.3.5. Zum WIEN-Spannungsteiler

a) Schaltbild
d) Ortskurve
c) Frequenz- und Phasengang
b) bis f) Knickkennliniendarstellungen zur Erklärung der unterschiedlichen Güte
e) Pol-Nullstellen-Plan

Im Vergleich zum Schwingkreis ist sie wesentlich geringer. Sie erreicht *maximal* den *Wert* 0,5 und liegt für die meisten Ausführungen bei etwa 1/3. Bezüglich des Einschwingverhaltens (Abschnitt 2.2.3.) liegt also stets die aperiodische Dämpfung vor (Gl. (2.2.22)). Auch die Ortskurve entspricht gemäß Bild 2.3.5b) qualitativ der des Schwingkreises (vgl. Bild 2.2.6). Diese Eigenschaften führten dazu, daß solche Glieder immer mehr anstelle von Schwingkreisen eingesetzt werden. Dabei ist es jedoch notwendig, daß sie zur Verbesserung der Güte mit aktiven Bauelementen gekoppelt werden. Sie haben dann aber den Vorteil, daß alle Bauelemente voll in die Schaltung integrierbar sind, da keine Induktivitäten notwendig sind.

Neben dem WIEN-Glied existieren noch *drei weitere Schaltungen*, die *gleichartiges Verhalten* zeigen. Lediglich ihr Übertragungsmaß und ihre Güte sind anders. Ihre Schaltung und Eigenschaften zeigt Bild 2.3.5. Besonders übersichtlich läßt sich beim Hoch- und Tiefpaßglied mit Hilfe der Knickkennlinien zeigen, wie die Selektion dieser Glieder entsteht (Bild 2.3.5d) bis f)). Die Steilheit in weiter Entfernung von der „Resonanzfrequenz" ω_0 ist proportional bzw. umgekehrt proportional zur Frequenz. Die Breite der Resonanz (Resonanzschärfe, Güte) ergibt sich aus dem Abstand der Knickfrequenz für den Hoch- und Tiefpaß. Sie liegt nahe dem Maximum, wenn beide Frequenzen etwa gleich sind. Wird die Hochpaßfrequenz niedriger als die Tiefpaßfrequenz gewählt, so wird keine wesentlich bessere Resonanzschärfe erreicht, sondern es tritt ein starker Amplitudenverlust ein. Weitere Amplituden- und Güteverluste treten noch durch die gegenseitige Belastung der Teilglieder auf.

Genau so wie es *Serien- und Parallelschwingkreis* gibt, lassen sich auch *RC*-Glieder finden, die eine Frequenz besonders stark dämpfen. Sie werden aus Bild 2.3.7a) verständlich. Zeigt das Verhältnis von Ausgangs- zu Eingangsspannung $u_a : u_e$ für eine Frequenz ein Maximum, so muß die Spannung u_a' ein Minimum besitzen. Ist also für

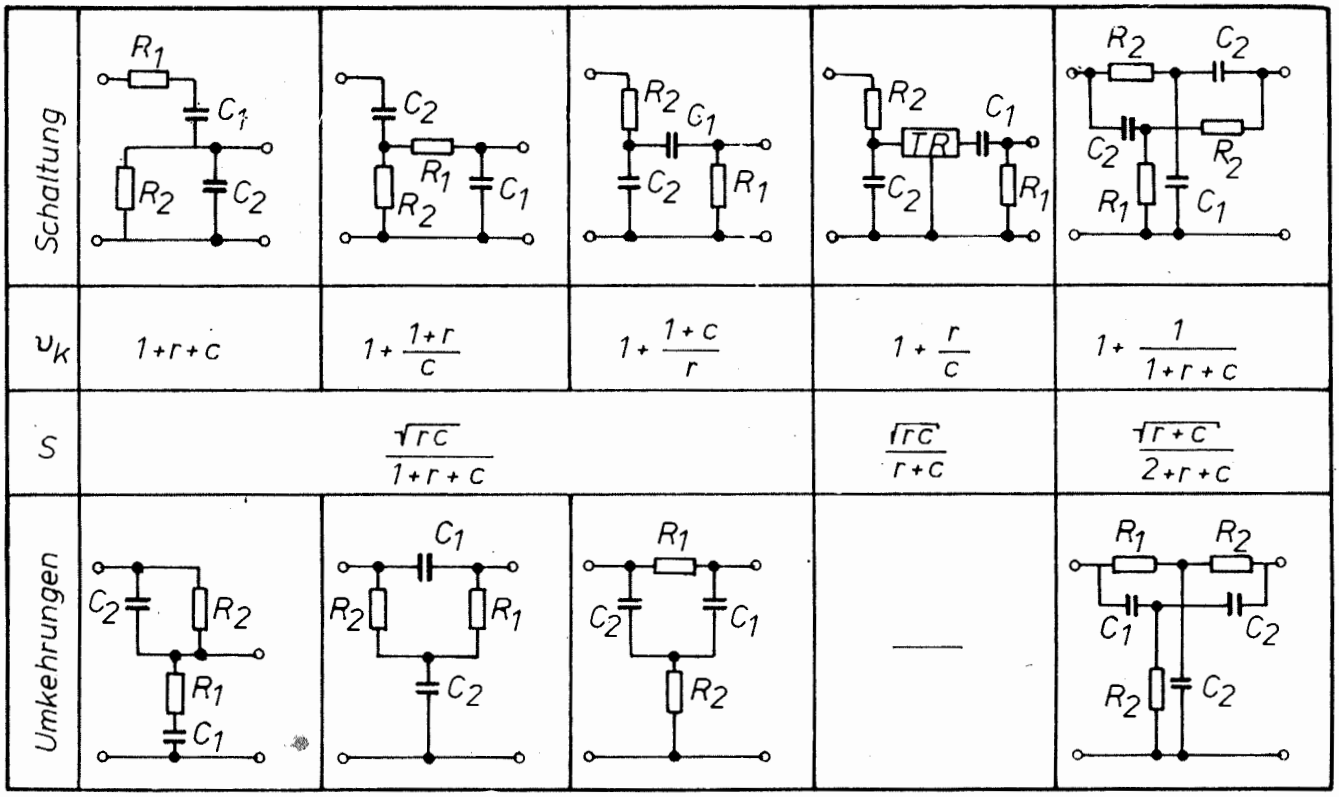


Bild 2.3.6. Aufbau der selektiven *RC*-Glieder, ihre Eigenschaften und Umkehrungen. Es bedeuten $r = R_1/R_2$; $c = C_2/C_1$ und *Tr* Trennstufe zur Entkopplung

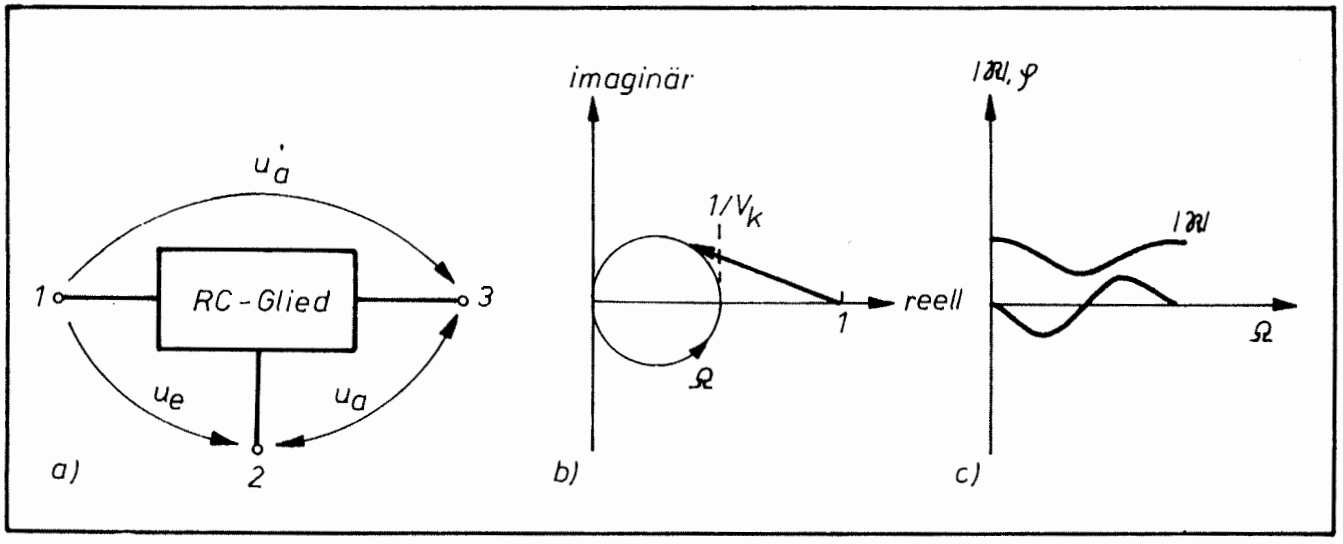


Bild 2.3.7. a) Zur Herleitung der Umkehrungen der *RC*-Glieder. Durch die Umkehrung des *RC*-Gliedes geht der Vektor der Ausgangsspannung nicht vom Ursprung in der GAUSS-Ebene, sondern vom Punkt +1 aus (b). Dadurch ergibt sich für diese Glieder ein Frequenz- und Phasengang gemäß c)

die Glieder mit *Bandpaßverhalten* der gemeinsame Anschluß 2, so muß es für die *frequenzbandsperrenden Glieder* 1 sein. Damit wird 2 zur Eingangs- und 3 zur Ausgangsklemme. In diesem Sinne ergeben sich die Umkehrungen der *RC*-Glieder. Sie sind in Bild 2.3.6 in der untersten Zeile mit aufgeführt. Die Frequenzkurven der Umkehrungen mit der zugehörigen Ortskurve zeigt Bild 2.3.7 b) und c). Der Ausgangsvektor beginnt also nicht in 0, sondern im Punkt 1.

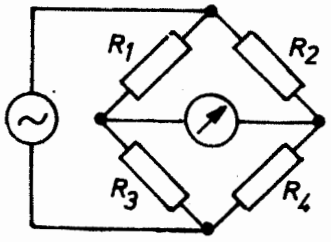
2.3.5. Brückenschaltungen

Alle bisher behandelten Vierpole hatten für den Ein- und Ausgang einen gemeinsamen Pol. Für den allgemeinen Vierpol ist dies jedoch nicht notwendig. Ein besonders einfaches Beispiel dafür ist die WHEATSTONE-*Brücke* (Bild 2.3.8). Sie wird in der Regel so abgeglichen, daß keine Spannung von der Quelle zum Galvanometer gelangt. Hierzu muß das Potential an den beiden Galvanometeranschlüssen gleich sein. Folglich gilt die im Bild stehende Gleichung.

Da alle Widerstände *komplex* sein können, gelten die Beziehungen *getrennt für den reellen und imaginären* Teil. Für die abgeglichene Brücke ist es ohne Bedeutung, *welches Klemmenpaar als Ein- bzw. Ausgang* benutzt wird. *Unterschiede* bestehen in dieser Hinsicht nur

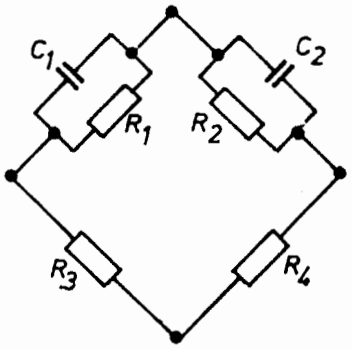
— bezüglich der Stromaufnahme
— außerhalb des Abgleichs bezüglich der Last
— bezüglich der Steilheit des Nulldurchganges.

Drei wichtige spezielle Brücken mit den zugehörigen Nullbedingungen zeigt Bild 2.3.8. Während die *Kapazitätsmeßbrücke* einen Nullabgleich unabhängig von der Frequenz zeigt, ist es für die WIEN-*Brücke* gerade typisch, daß der Nullabgleich



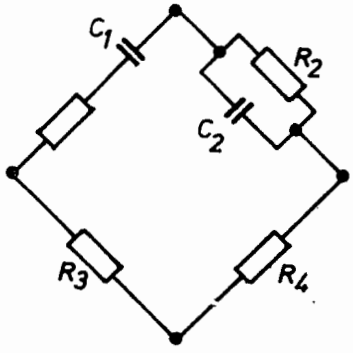
WHEATSTONE-Brücke

$$R_2 R_3 = R_1 R_4.$$



Kapazitätsmeßbrücke

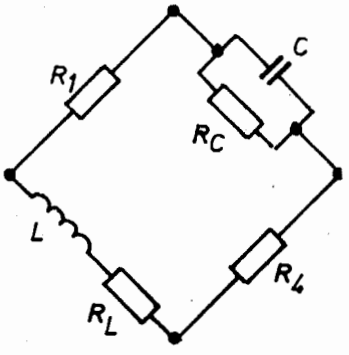
$$\frac{R_3}{R_4} = \frac{R_1}{R_2} = \frac{C_2}{C_1}$$



WIEN-Brücke

$$\omega_0^2 = \frac{1}{R_1 R_2 C_1 C_2}$$

$$\frac{R_1}{R_2} + \frac{C_2}{C_1} = \frac{R_3}{R_4}$$



MAXWELL-WIEN-Brücke

$$L = R_1 R_4 C$$

$$R_1 R_4 = R_C R_L$$

Bild 2.3.8. Bezeichnungen, Schaltungen und Abgleichbedingungen der allgemeinen WHEATSTONE-Brücke sowie dreier daraus ableitbarer Wechselstrombrücken

nur für eine Frequenz erfolgt. Sollen mit ihr also Kapazitäten gemessen werden, so muß die Spannungsquelle frei von Oberwellen sein. Andererseits kann auf diese Art die WIEN-Brücke aber auch dazu benutzt werden, die Grundwelle einer Schwingung zu sperren. Da sie die Oberwellen weiterhin durchläßt, eignet sie sich zur relativ *einfachen Messung des Klirrfaktors.*

Mit dem Abgleich der WIEN-Brücke ändert sich entsprechend Bild 2.3.9 aber auch der Frequenzgang. Das reine WIEN-Glied (a) zeigt das schon bekannte Verhalten. Durch eine geringe Verschiebung in Richtung der Nullstelle (b) wird das Maximum etwas flacher, und bei allen Frequenzen kommt eine Spannung hindurch. Der Verschiebung entspricht eine Verlagerung des Ausgangspunktes des Vektors bezüglich der Ortskurve. Beim weiteren Abgleich gelangt der Aus-

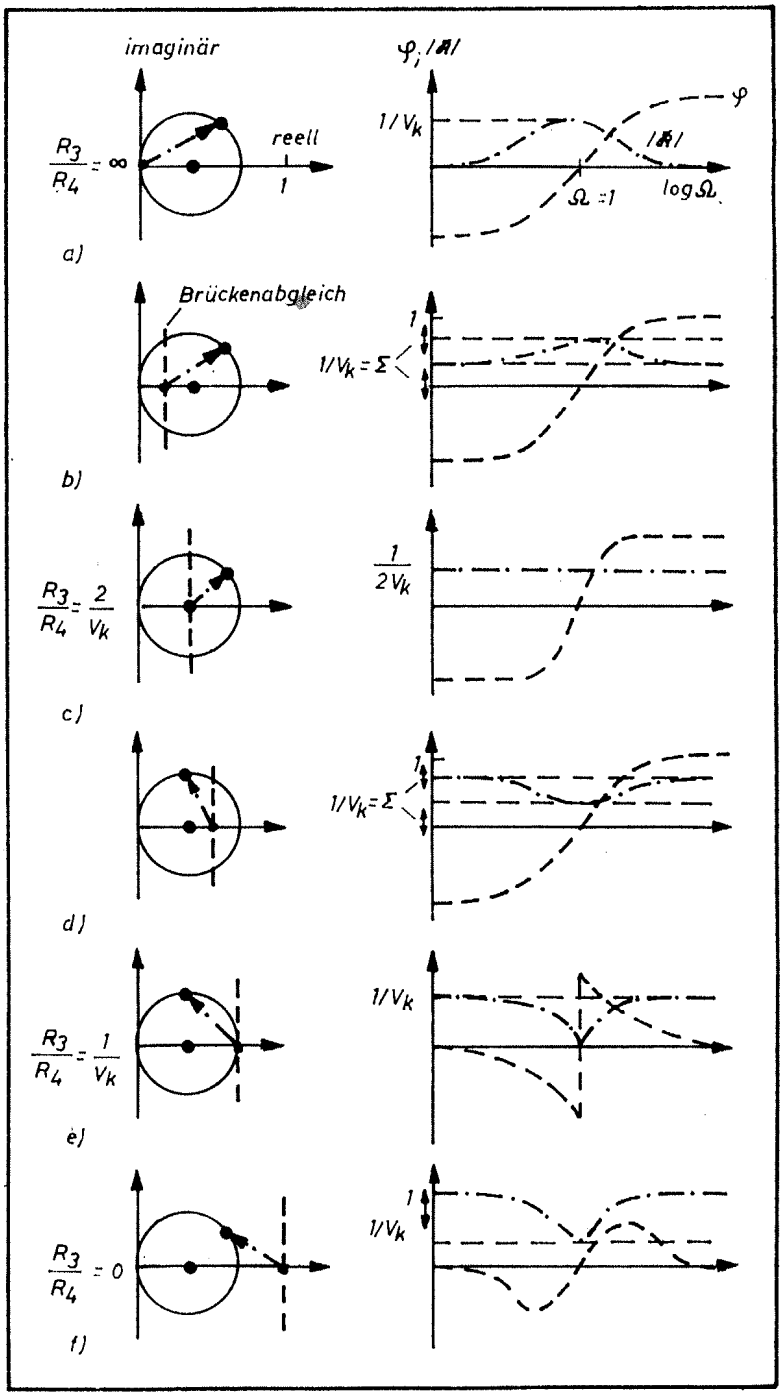


Bild 2.3.9. Zum Verhalten der WIEN-Brücke gemäß Bild 2.3.8 bei verschiedenen Abgleichbedingungen. Die Bezeichnungen im Bild nehmen Bezug auf Gl. (8). Zugleich gelten daher diese Betrachtungen für alle *RC*-Glieder von Bild 2.3.8, wenn sie an die Stelle des WIEN-Gliedes in dieser Brücke treten

gangspunkt des Vektors genau in den Kreismittelpunkt der Ortskurve (c). Dadurch werden alle Frequenzen mit gleicher Amplitude hindurchgelassen. Für jede Frequenz existiert aber eine andere Phasenverschiebung zwischen Ein- und Ausgang. Bei weiterer Näherung zum Nullpunkt wird ein Minimum sichtbar (d). Die Phasensteilheit in der Nähe des Nullpunktes erhöht sich weiter. Beim genauen Abgleich der Brücke (e) wird das Minimum zum Nulldurchgang, und die Phase zeigt einen Sprung. Bei weiterer Verschiebung über den Nullpunkt hinaus wird das Minimum wieder flacher, und die Phase zeigt Maxima und Minima, bis schließlich das umgekehrte WIEN-Glied (f) erreicht wird. Aus diesem Beispiel ist ersichtlich, wie vielfältig die Eigenschaften des WIEN-Gliedes durch die Brückenschaltung werden.

In Bild 2.3.8 ist auch die MAXWELL-WIEN-*Brücke* enthalten. Sie dient vor allem zur *Messung von Induktivitäten* und hat den großen Vorteil, daß diese Messung aus einer Kapazität als Normal abgeleitet wird. Bei dieser Schaltung ist es auch relativ leicht, die Induktivität zusätzlich mit einem Gleichstrom bei der Messung vorzumagnetisieren.

Am Beispiel der MAXWELL-WIEN-Brücke sei genauer gezeigt, *wie der Nullabgleich erfolgt.* Mit ihm müssen ja die beiden Gleichungen in Bild 2.3.8 erfüllt werden. Bei der Induktivitätsmessung sind L und R_L unbekannt und sollen mittels Abgleich von R_4 und R_C bestimmt werden. Die entsprechend abgewandelte Schaltung und die entsprechend aufgelösten Gleichungen zeigt Bild 2.3.10a). Es ist offensichtlich, daß

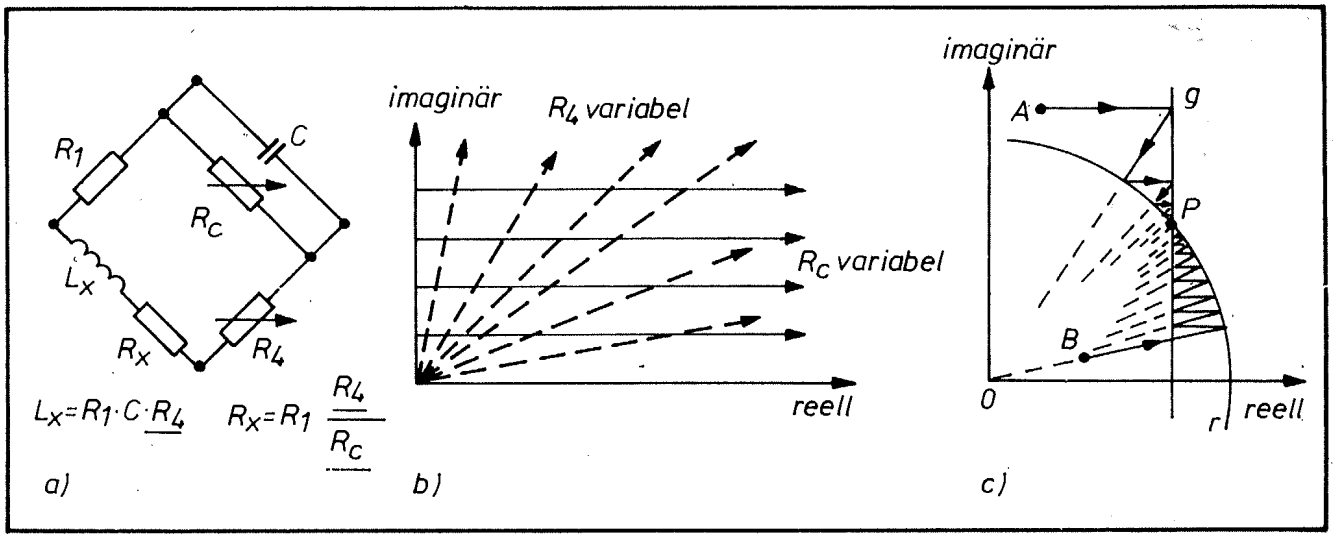


Bild 2.3.10. Zum Abgleich einer komplexen Meßbrücke

a) Schaltung und zugehörige Abgleichbedingungen
b) Durch Änderung von R_4 entstehen Strahlen, die vom Nullpunkt ausgehen; durch Änderung von R_C entstehen Parallelen zur reellen Achse
c) Wechselseitige Abgleiche von R_4 und R_C, die von den zufälligen Punkten A bzw. B ausgehen und zum Nullabgleichspunkt P führen

der Abgleich des Real- und Imaginärteils nicht unabhängig erfolgen. Eine Veränderung von R_C hat keinen Einfluß auf den Imaginärteil. Sie ist also durch Parallelen zur waagrechten Achse darstellbar. Veränderung von R_4 beeinflußt dagegen Real- und Imaginärteil gleichermaßen. Hierfür gelten also Strahlen, die vom Ursprung ausgehen (Bild 2.3.10b)). Der Abgleich der Brücke auf den unbekannten Punkt P für die Induktivität in der Ortskurve muß folglich iterativ mittels relativer Minima erfolgen. Dies zeigt Teilbild 2.3.10c), wenn z.B. vom zufällig vorhandenen Punkt A bzw. B ausgegangen wird. Der Radius r gibt dabei den minimalen Abstand von P für Veränderungen von R_4 und die Gerade g für Änderung von R_C an. Hieraus folgt:

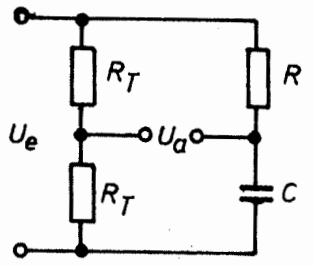
— Ein *exakter Abgleich* ist erst nach *unendlich* vielen wechselseitigen Abgleichen von R_4 und R_C möglich.

— Ein *näherungsweiser Abgleich* dauert um so *länger, je größer der Verlustwiderstand* der Induktivität ist. Dann liegt P nämlich weiter unten auf dem Kreis r.

— Der *Abgleichvorgang erreicht* dann *schneller* eine bestimmte *Näherung*, wenn *mit zu großem R_C begonnen* wird (Vergleich Punkt A und B).

Die Problematik des komplizierten und langwierigen iterativen Abgleichs ist in der letzten Zeit zu vermeiden versucht worden. Eine Methode koppelt den Abgleich der Widerstände R_4 und R_C so, daß die Bewegungsrichtungen etwa senkrecht aufeinander stehen (Orthonull). Eine zweite Methode ersetzt R_3 durch eine Diode, die mit einer phasenempfindlichen Schaltung immer automatisch auf das relative Minimum abgeglichen wird.

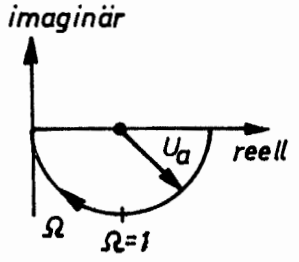
Bei der WIEN-Brücke wurde bereits gezeigt, wie sie als reiner *Phasenschieber* ohne Amplitudengang betrieben werden kann (Bild 2.3.9c)). Andere Phasenschieberschaltungen zeigt Bild 2.3.11 mit den Ortskurven und den zugehörigen Formeln. Die obere Schaltung wird vor allem dann benutzt, wenn eine bestimmte Phase für eine Frequenz durch den Widerstand oder Kondensator eingestellt werden soll. Die untere Schaltung ist ein spezieller Breitband-Phasenschieber, der für einen großen Frequenzbereich zwischen den beiden Ausgangsspannungen 90° aufweist. Hierbei ändert sich jedoch noch das Amplitudenverhältnis. Durch kompliziertere *Breitband-Phasenschieberschaltungen* läßt sich dies vermeiden. Sie sind auch für beliebige Phasenwinkel dimensionierbar.

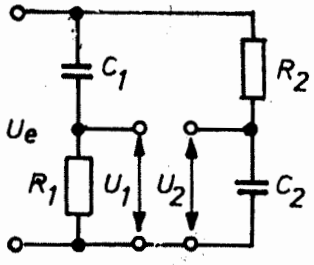


$$\underline{K} = \frac{2U_a}{U_e} = \frac{1 - j\Omega}{1 + j\Omega}$$

$$[\underline{K}] = 1 \qquad \Omega = \omega CR$$

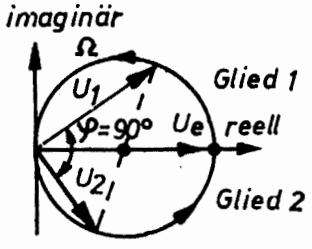
$$\tan\varphi = -\frac{2\Omega}{1 - \Omega^2}$$





$$\underline{U}_1 = \underline{U}_e \frac{1}{1 + j\Omega_1} \qquad \Omega_1 = \omega R_1 C_1$$

$$\underline{U}_2 = \underline{U}_e \frac{1}{1 + \frac{1}{j\Omega_2}} \qquad \Omega_2 = \omega R_2 C_2$$



$$\frac{\underline{U}_1}{\underline{U}_2} = \frac{1 - \frac{\Omega_1}{\Omega_2} + j\left(\frac{1}{\Omega_2} - \Omega_1\right)}{1 + \Omega_1^2}$$

für $\Omega_1 = \Omega_2 = \Omega$ folgt

$$\frac{\underline{U}_1}{\underline{U}_2} = \frac{j\left(\frac{1}{\Omega} - \Omega\right)}{1 + \Omega^2}, \text{ also } \Delta\varphi = 90° \text{ für alle Frequenzen}$$

allerdings mit $\frac{|U_1|}{|U_2|} = \sqrt{\frac{1 + \frac{1}{\Omega^2}}{1 + \Omega^2}}$

Bild 2.3.11. Schaltung, Formeln und Ortskurven für zwei spezielle Phasenschieber

2.4. Systemtheorie

2.4.1. Allgemein komplexe Frequenzen

Die Pole und Nullstellen wurden bereits im Abschnitt 2.2.4. am Beispiel des Serienkreises eingeführt. In Verallgemeinerung dessen muß zunächst die Bedeutung der einzelnen *Punkte für* die entsprechende GAUSS-*Ebene* eingeführt werden. Sie entsprechen einer *verallgemeinerten komplexen Frequenz*

$$p = \sigma + \mathrm{j}\omega \,. \tag{1}$$

Entsprechend dem komplexen Ausdruck

$$f(t) = \mathrm{e}^{-\sigma t} \cdot \mathrm{e}^{\mathrm{j}\omega t} = \mathrm{e}^{pt} \tag{2}$$

ist erkennbar, daß hiermit die harmonische Frequenz verallgemeinert wurde, indem sie mit einem *exponentiellen Dämpfungsfaktor* multipliziert wird. Daher entspricht jedem Punkt in der GAUSS-Ebene mit σ und $\mathrm{j}\omega$ eine entsprechend σ an- bzw. abklingende Schwingung der Frequenz ω. Lediglich für $\sigma = 0$ wird die normale, ewig dauernde harmonische Schwingung mit konstanter Amplitude auftreten. In Bild 2.4.1 ist dies schematisch dargestellt. Für theoretisch Interessierte sei erwähnt, daß sich mit dem Übergang von der normalen, reellen Frequenz ω zur komplexen Frequenz p auch leicht ein Bezug zwischen FOURIER- und LAPLACE-Transformation herstellen läßt.

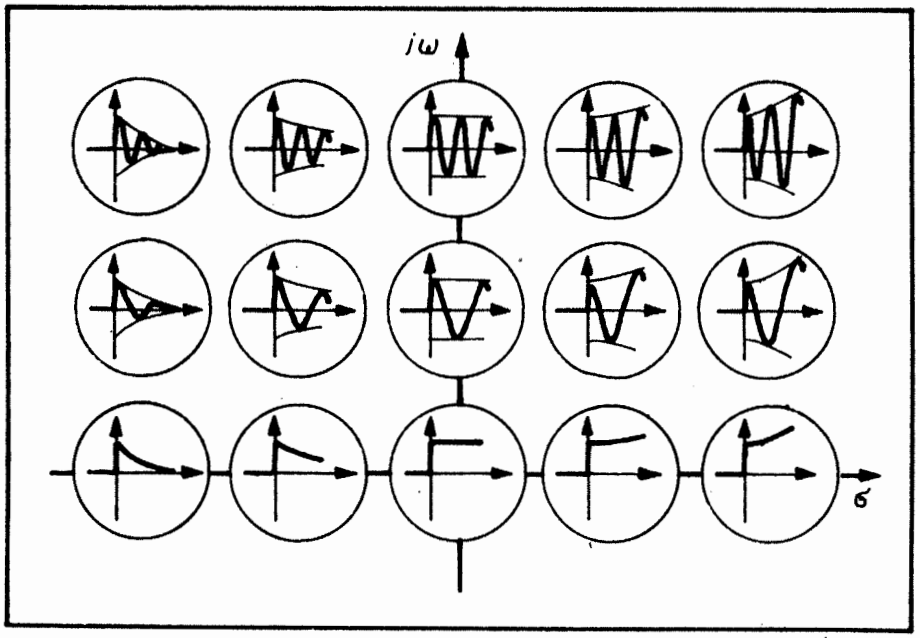


Bild 2.4.1. Darstellung einiger ausgewählter Schwingungsverläufe der Funktion $\exp(\sigma + \mathrm{j}\omega)$. Das Bild kann auch an der σ-Achse nach unten gespiegelt werden. Es ist dann auch für negative ω gültig. Dabei tritt nur eine um 180° verschobene Anfangsphase auf

2.4.2. Allgemeine Systeme

Ein *allgemeines System* habe die *Eingangsgröße* $x_\mathrm{e}(t)$ und die *Ausgangsgröße* $x_\mathrm{a}(t)$. Bei den Zweipolen können beiden Strom bzw. Spannung zugeordnet werden, wobei die Eingangsgröße danach ausgewählt wird, oder der Zweipol vom Strom oder von der Spannung erregt wird. Bei Vierpolen können die Eingangs- und Ausgangswiderstände wie Zweipole betrachtet werden. Die Übertragung vom Eingang zum Ausgang kann

für Spannungen oder Ströme oder gemischt betrachtet werden. Insgesamt zeigt sich also, daß zur vollständigen Beschreibung eines Vierpols wieder vier Kenngrößen entsprechend den Koeffizienten der Vierpolmatrizen von Abschnitt 1.1.5. notwendig sind.

Die folgende Betrachtung ist also für jeden einzelnen Koeffizienten getrennt möglich und z. T. auch notwendig. Wir setzen für das *System* die folgenden *Eigenschaften* voraus:

- *Stabilität*: Wenn die Ursache Null wird, folgt ihr die Wirkung Null in endlicher Zeit. Ohne Eingangsgröße stellt sich also alsbald ein stabiles System in einem unerregten Zustand ein und verharrt dort. Zuweilen wird in diesem Zusammenhang auch von kausalen Systemen gesprochen.
- *Linearität*: Verschiedene Teilursachen addieren sich auch in der Wirkung.
- *Zeitinvarianz*: Die feststellbaren Eigenschaften und Zusammenhänge sind unabhängig von dem Zeitpunkt, in dem sie ermittelt werden. Die Parameter des Systems ändern sich also nicht.

Diese drei *Bedingungen* sind dann *automatisch erfüllt*, wenn die Systeme nur aus *passiven* (nicht alternden) *Bauelementen*, also Widerständen, Kapazitäten und Induktivitäten aufgebaut sind. Bei anderen, insbesondere bei aktiven Bauelementen sind vor allem gesonderte Stabilitäts-, aber auch Linearitäts- und Zeitinvarianzuntersuchungen notwendig.

Für derartige Systeme gilt die lineare Differentialgleichung

$$a_0 x_e(t) + a_1 \frac{\mathrm{d}x_e(t)}{\mathrm{d}t} + a_2 \frac{\mathrm{d}^2 x_e(t)}{\mathrm{d}t^2} + \ldots + a_n \frac{\mathrm{d}^n x_e(t)}{\mathrm{d}t^n}$$

$$= A \left[(b_0 x_a(t) + b_1 \frac{\mathrm{d}x_a(t)}{\mathrm{d}t} + b_2 \frac{\mathrm{d}^2 x_a(t)}{\mathrm{d}t^2} + \ldots + b_m \frac{\mathrm{d}^m x_a(t)}{\mathrm{d}t^m} \right]. \tag{3}$$

Infolge von Stabilität und Kausalität können in der Wirkung keine höheren Ableitungen als in der Ursache auftreten. Dies bedeutet:

$$m \leqq n\,. \tag{4}$$

Da es sich um ein physikalisch reales und zeitinvariantes System handelt, sind weiter

alle Koeffizienten reell und konstant.

Der Faktor A sorgt dafür, daß je nach den gewählten Ursachen und Wirkungen (Strom bzw. Spannung) die Koeffizienten frei von Einheiten sind. Im Sinne einer Normierung kann er auch unterdrückt werden ($A = 1$).

Die Anfangsbedingungen des Systems seien dadurch bestimmt, daß es sich zum Startzeitpunkt $t = 0$ in Ruhe befindet. Dann gilt $x_e(-0) = x_a(0) = 0$. Weiter sind alle Ableitungen beider Größen zu diesem Zeitpunkt Null.

Wird auf (3) die LAPLACE-Transformation angewendet, so folgt eine gebrochen rationale Funktion, die zwischen den komplexen Ein- und Ausgangsfrequenzen vermittelt:

$$H(p) = \frac{b_0 p^m + b_1 p^{m-1} + b_2 p^{m-2} + \ldots + b_m}{a_0 p^n + a_1 p^{n-1} + a_2 p^{n-2} + \ldots + a_n}. \tag{5}$$

Dieser Ausdruck ist die Verallgemeinerung von (2.2.36). Beim Vierpol heißt dieser Zusammenhang *Übertragungsfunktion*. Beim Zweipol wird von *Imetanz-* bzw. *Zweipolfunktion* gesprochen.

2.4.3. Pol-Nullstellen-Plan

Vielfach ist es nützlich, von der Übertragungsfunktion

$$H(p) = \frac{Z(p)}{N(p)} \tag{6a}$$

das Zähler- und Nennerpolynom getrennt zu betrachten. Beide lassen sich mittels ihrer im allgemeinen komplexen Nullstellen, p_{0q} für das Zähler- und $p_{\infty q}$ für das Nennerpolynom, in Linearfaktoren zerlegen:

$$H(p) = \frac{b_0(p - p_{01})^{k_1} \cdot (p - p_{02})^{k_2} \cdot \ldots \cdot (p - p_{0u})^{k_u}}{a_0(p - p_{\infty 1})^{l_1} \cdot (p - p_{\infty 2})^{l_2} \cdot \ldots \cdot (p - p_{\infty v})^{l_v}}. \tag{6b}$$

Dabei bedeuten die k_q und l_q die Ordnungen der Nullstellen. Für sie muß gelten

$$\sum k_q = m \quad \text{und} \quad \sum l_q = n\,. \tag{6c}$$

Für die Übertragungsfunktion bewirken die *Nullstellen* des Zählerpolynoms Nullstellen; die des Nennerpolynoms dagegen *Unendlichkeitsstellen* (*Pole*). Fallen Nullstellen des Nenner- und Zählerpolynoms zusammen, so sind die entsprechenden Linearfaktoren bis auf ein Differenzglied im Zähler oder Nenner kürzbar.

Neben der polynomialen Form von Gl. (5) und der Darstellung in Linearfaktoren gemäß Gl. (6) existieren noch zwei weitere nützliche Darstellungen der Übertragungsfunktion. Eine davon ist die Zerlegung in *Partialbrüche*

$$H(p) = A_0 + \sum_{q=1}^{n} \sum_{v=1}^{t} \frac{A_{rv}}{(p - p_{\infty q})^r}. \tag{7a}$$

Darin ist r die Vielfachheit der Pole $p_{\infty q}$ und A_{rv} das zugehörige Residuum. Für den Sonderfall einfacher Pole sind die *Residuen* relativ leicht berechenbar:

$$A_{1k} = \frac{\prod_{q=1}^{m} (p_{\infty k} - p_{0q})}{\prod_{q=1}^{n} (p_{\infty k} - p_{\infty q})} \qquad \text{mit} \quad q \neq k\,. \tag{7b}$$

In der Gleichung sind also alle Strecken enthalten, die von allen Nullstellen und Polen zu dem ausgewählten k-ten Pol weisen.

Aufgrund physikalischer Gegebenheiten eines Systems unterliegt die Anordnung der Pole und Nullstellen in der komplexen σ, ω-Ebene verschiedenen Einschränkungen.

Allgemein gilt für stabile (lineare, zeitinvariante) Systeme:

- Das *Nennerpolynom* ist ein HURWITZ-*Polynom*, d. h., seine Nullstellen, also die Pole der Übertragungsfunktion, liegen in der linken Halbebene. Daher gilt $\mathrm{Re}\,(p_{\infty q}) < 0$.
- Alle nicht *reellen* Nullstellen und Pole sind *konjugiert komplex*.

Aufgrund der zweiten Aussage existieren also zwei typische Arten von Nullstellen und Polen, nämlich die reellen und die konjugiert komplexen. Die konjugiert komplexen lassen sich besonders übersichtlich in einem *biquadratischen Ausdruck*

$$p^2 + ap + b \tag{8a}$$

mit den Wurzeln

$$p_{1,2} = -\frac{a}{2} \pm \sqrt{\frac{a^2}{4} - b} \tag{8b}$$

beschreiben. Insgesamt existiert damit die vierte Form der Übertragungsfunktion

$$H(p) = K \frac{\Pi(p + \sigma_{0k}) \cdot \Pi(p^2 + a_{0l}p + b_{0l})}{\Pi(p + \sigma_{\infty k}) \cdot \Pi(p^2 + a_{\infty l}p + b_{\infty l})}. \tag{9a}$$

Dabei gilt für die Indizes

$$\sum 0k + 2 \sum 0l = m\,, \tag{9b}$$

$$\sum \infty k + 2 \sum \infty l = n\,. \tag{9c}$$

2.4.4. Einteilung der Systeme

Aus den Betrachtungen des vorigen Abschnitts geht deutlich hervor, daß die Nullstellen eines Systems nicht beliebig angeordnet sind. Es läßt sich allgemein zeigen, daß bezüglich der Lage von Polen und Nullstellen die Einteilung gemäß Bild 2.4.2 gilt. Die *stabilen Systeme* verlangen also, daß keine Pole in der rechten Halbebene liegen. Von den stabilen Systemen können dann wiederum zwei spezielle Systeme herausgetrennt werden:

Die *Allpaßsysteme* übertragen dabei alle Frequenzen, bewirken aber Phasenverschiebungen. Bei ihnen ist gemäß Bild 2.4.3a) jedem Pol in der linken Halbebene symmetrisch eine Nullstelle in der rechten Halbebene zugeordnet. Die Allpaßeigenschaft ist besonders leicht auf der Basis von Gl. (7) zu erkennen. Es gilt für sie

$$Z(-p) = N(p)\,. \tag{10}$$

Nullstellen	Pole	
	nur linke Halbebene	auch rechte Halbebene
nur links	Minimalphasensysteme	instabile Systeme
beliebig	allgemeine stabile Systeme	
nur rechts	Allpaßsystem	

Bild 2.4.2. Einteilung der Systeme nach Lage der Nullstellen und Pole

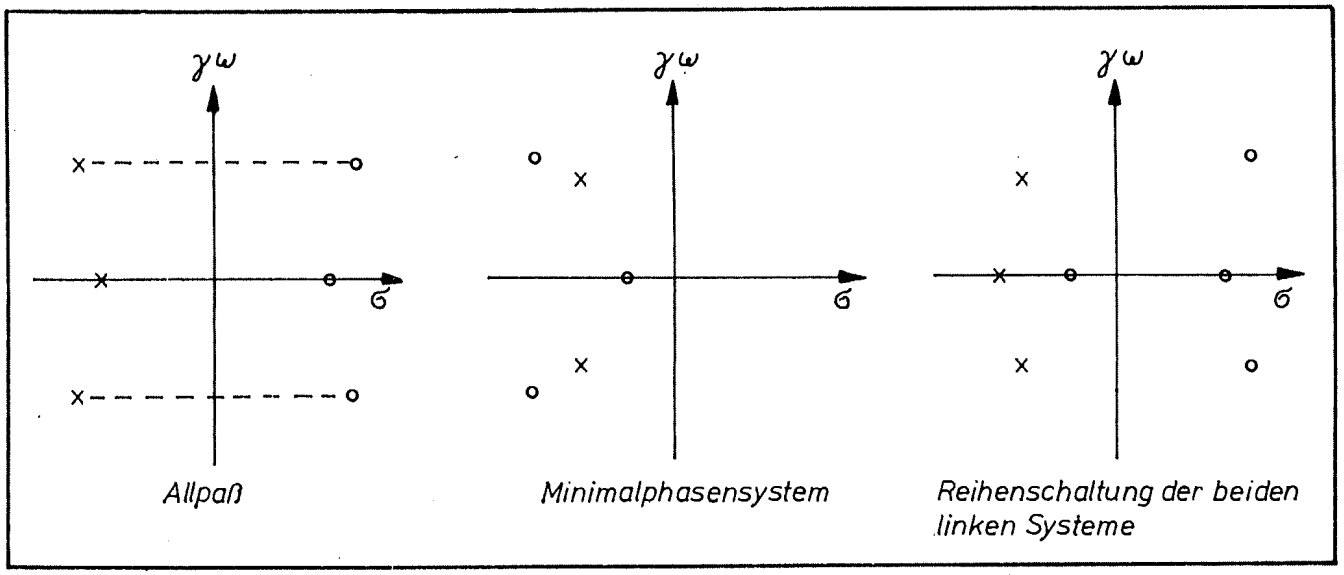


Bild 2.4.3. Lage der Pole und Nullstellen bei a) einem Allpaß, b) einem Minimalsystem und c) einem allgemein stabilen System, welches durch Reihenschaltung der beiden ersten entsteht

Die *Minimalphasensysteme* besitzen gemäß Bild 2.4.2 alle ihre Nullstellen und Pole in der linken Halbebene, wobei sie entweder auf der σ-Achse liegen oder konjugiert komplex sind (Bild 2.4.3b)). Auf dieser Grundlage besteht zwischen ihrem Amplituden- und Phasengang eine feste Kopplung. Ist einer von beiden gegeben, so folgt der andere daraus. Die Umrechnung ist allerdings nicht immer ganz einfach, da sie über die HILBERT-Transformation erfolgt. Vereinfacht (wenn keine Pole und Nullstellen in der rechten Halbebene und auf der ω-Achse liegen) gilt, wenn $A(\omega)$ der Amplituden- und $\varphi(\omega)$ der Phasenfrequenzgang ist:

$$\ln A(\omega) = -\frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{\varphi(u)}{u-\omega}\,\mathrm{d}u\,, \qquad \varphi(\omega) = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{\ln A(u)}{u-\omega}\,\mathrm{d}u\,. \tag{11}$$

Der Begriff Minimalphasensystem kann u. U. irreführend sein. Es lassen sich nämlich auch Übertragungsglieder realisieren, die zwar einen Amplitudengang, aber keinen Phasengang besitzen. Sie treten z. B. als Apperturfehler in der Optik bzw. beim Frequenzgang des Magnetbandkanals auf. Sie besitzen aber nicht den eindeutigen Zusammenhang zwischen Amplituden- und Phasengang. Das Minimalphasensystem wird auch Mindestphasen- oder Phasenminimumsystem genannt.

Es läßt sich zeigen, daß sich *jedes stabile System* aus der *Reihenschaltung* je eines *Minimalphasen- und eines Allpaßsystems* aufbauen läßt. Dies zeigt demgemäß Bild 2.4.3c). Dies ist schließlich der Grund, weshalb dem Allpaß- und dem Minimalphasensystem so große Bedeutung zukommt.

Bei Minimalphasensystemen kann für (11) eine grobe Näherung angegeben werden:

$$\varphi(\omega) \approx \frac{\mathrm{d}A(\omega)}{\mathrm{d}\omega}. \tag{11a}$$

Für viele Anwendungen ist die *Gruppenlaufzeit* von Bedeutung:

$$T_{\mathrm{gr}}(\omega) = \frac{\mathrm{d}\varphi(\omega)}{\mathrm{d}\omega}. \tag{11b}$$

In einigen Gebieten, z.B. Fernsehen, wird auch die *Phasenlaufzeit*

$$T_{\mathrm{ph}} = \varphi(\omega)/\omega \tag{11c}$$

verwendet. Mit der Näherung (11a) gilt

$$T_{\mathrm{gr}}(\omega) \approx \frac{\mathrm{d}^2\varphi(\omega)}{\mathrm{d}\omega^2}. \tag{11d}$$

2.4.5. Potentialanalogien

Die bisherigen Betrachtungen zeigten, daß neben einfachen Nullstellen und Polen besonders häufig *drei Kombinationen* je zwei benachbarter Nullstellen und/oder Pole auftreten:

- Nullstelle — Pol
- Pol — Pol
- Nullstelle — Nullstelle.

Die beiden letzten Typen ergeben sich vor allem infolge der konjugiert komplexen Pole bzw. Nullstellen. Das Gefüge *Nullstelle — Pol* kommt ausgeprägt bei den Allpässen und Zweipolen vor. Wichtige Eigenschaften dieser drei Möglichkeiten können

mittels der Linien konstanter Dämpfung (Amplitude) und Phase in der σ, ω-Ebene erhalten werden. Dabei entstehen Gebilde, wie sie bereits im Abschnitt 1.3. im Zusammenhang mit dem Bild 1.3.1 behandelt wurden. Auf dieser Basis ist relativ leicht eine *anschauliche Interpretation* der Pole und Nullstellen möglich. Hierzu diene das Beispiel zweier konjugiert komplexer Pole in der linken Halbebene und einer Nullstelle im Ursprung gemäß Bild 2.4.4. Es entspricht dem Leitwert des Reihenschwingkreises gemäß Bild 2.2.2, wobei durch den Übergang zum Leitwert in Bild 2.2.4 die Nullstellen

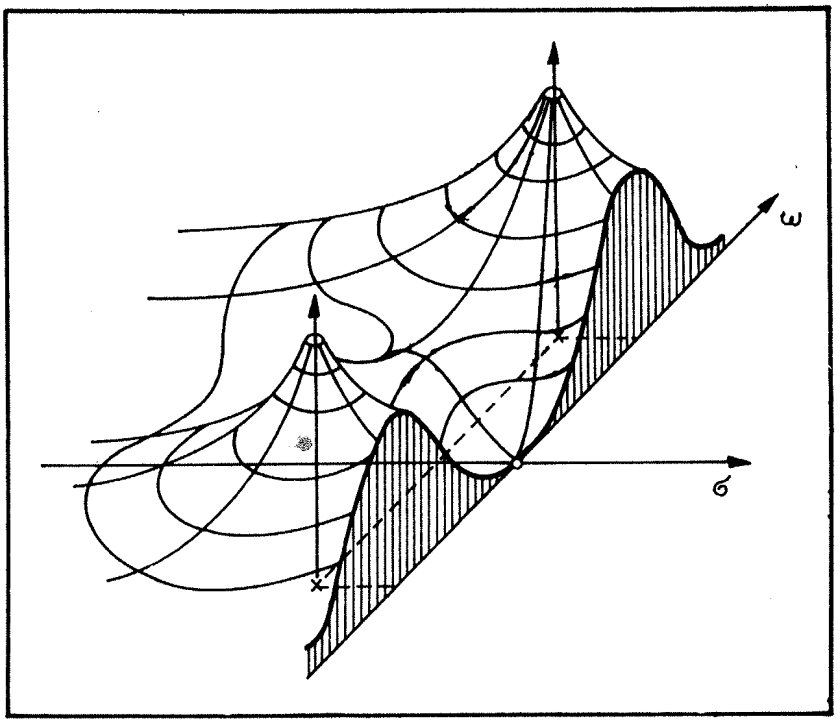


Bild 2.4.4. Anschauliche Interpretation eines Pol-Nullstellen-Planes und Vergleich zur Potentialdarstellung gemäß Bild 1.3.1

zu Polen und der Pol zur Nullstelle wird. Physikalische Bedeutung hat in der Darstellung natürlich nur die positive ω-Achse. Sie zeigt einen Frequenzgang, der notwendigerweise weitgehend dem des komplexen Spannungsteilers von Bild 2.2.3 entspricht. Es ist auch deutlich ersichtlich, wie mit Veränderung der Dämpfung (in etwa Verschieben der Schnittebene nach rechts bzw. links) die Resonanzschärfe geändert wird. Wird in das Bild 1.3.1 zurückgegangen, so entsprechen die dünnen Höhenlinien dem Amplitudengang und die dicken Stromlinien (Flußlinien) dem Phasengang des entsprechenden Systems.

Aus der Zusammenstellung gemäß Bild 2.4.3 wird so noch einmal die Notwendigkeit der Symmetrien von Allpaß- und Minimalphasensystem und des daraus entstehenden allgemeinen, stabilen Systems deutlich: Einzelne Pole oder Nullstellen können eben nur auf der σ-Achse liegen, Paare aus Polen und Nullstellen oder Polen oder Nullstellen können nur symmetrisch zur σ- oder ω-Achse liegen. Andernfalls würden Überlagerungen im Potentialbild und damit Amplituden- und Phasengang entstehen, die den realen Möglichkeiten widersprächen.

2.4.6. Zulässige Zweipolfunktionen

Bestehen die *Zweipole nur aus den drei passiven Grundbauelementen*, also R, C und L, so kommen neben den Einschränkungen von Abschnitt 2.4.4. noch folgende Eigenschaften hinzu (dabei sind die ersten drei Punkte das Kriterium für sogenannte positive Funktionen):

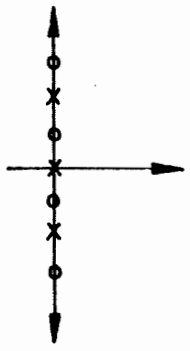
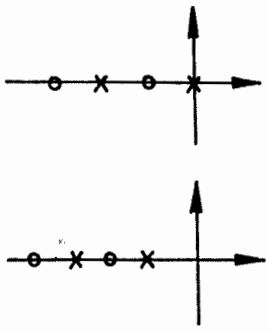
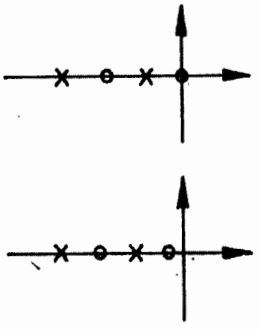
- In der positiven Halbebene existieren weder Pole noch Nullstellen.
- $H(p)$ muß für reelle p reelle Funktionswerte annehmen.
- $H(p)$ muß für alle p der rechten Halbebene einen positiven Realteil besitzen.
- Der Gradunterschied zwischen Zähler- und Nennerpolynom beträgt -1; 0 oder $+1$, d. h. $|m - n| \leqq 1$.
- Alle Pole und Nullstellen auf der ω-Achse können nur einfach sein.
- Ist $H(p)$ eine Zweipolfunktion, so ist es auch $1/H(p)$. Dies bedeutet den Übergang von Widerstand auf Leitwert.

Weitere Einschränkungen ergeben sich, wenn *nur zwei Bauelementetypen* benutzt werden. Hierzu gibt Tab. 2.4.1 einen Überblick.

Tabelle 2.4.1. Zur Lage der Nullstellen und Pole bei Zweipolschaltungen, die nur aus zwei Arten von den passiven Bauelementen aufgebaut sind
Generell gilt dann:

- Alle Nullstellen und Pole sind einfach
- Sie befinden sich auf der $j\omega$- oder σ-Achse
- Sie sind der Frequenz nach geordnet. Es wechseln sich also Pole und Nullstellen alternativ ab.

Einschränkungen für spezielle Schaltungen

L *C*	*R* *C*	*R* *L*
• $m = n + 1$ • Pole und Nullstellen auf der $j\omega$-Achse • bei $p = 0$ und bei $p = \infty$ befindet sich je ein Pol oder eine einfache Nullstelle • $Z(p) + N(p)$ ist ein HURWIRTZ-Polynom	• $n - m = 0$ oder 1 • Pole und Nullstellen auf der σ-Achse • entweder liegt im Ursprung ($p = 0$) oder ihm am nächsten ein Pol	• $m - n = 0$ oder 1 • Pole und Nullstellen auf der σ-Achse • entweder liegt im Ursprung ($p = 0$) oder ihm am nächsten eine Nullstelle
		

Besonders betont sei, daß *alle Pole nur einfach* sein können. Dadurch läßt sich nämlich jede Zweipolfunktion günstig in *Partialbrüche* zerlegen:

$$H(p) = \frac{A_0}{p} + \sum_{n=1}^{r} \frac{2A_n p}{p^2 + \omega_n^2} + A_\infty p \,. \tag{12}$$

Die Koeffizienten A_0, A_n und A_∞ lassen sich aus den a_n und b_k von Gl. (5) und umgekehrt berechnen. Hieraus leitet sich eine *systematische*, sogenannte *kanosche Realisierung* von Zweipolfunktionen ab. Für den Fall, daß nur L und C vorhanden sind, entstehen so die beiden Möglichkeiten von Bild 2.4.5. Es besteht aber auch die Möglichkeit, den Partialbruch schrittweise abzubauen. Dadurch geht (12) in einen *Ketten-*

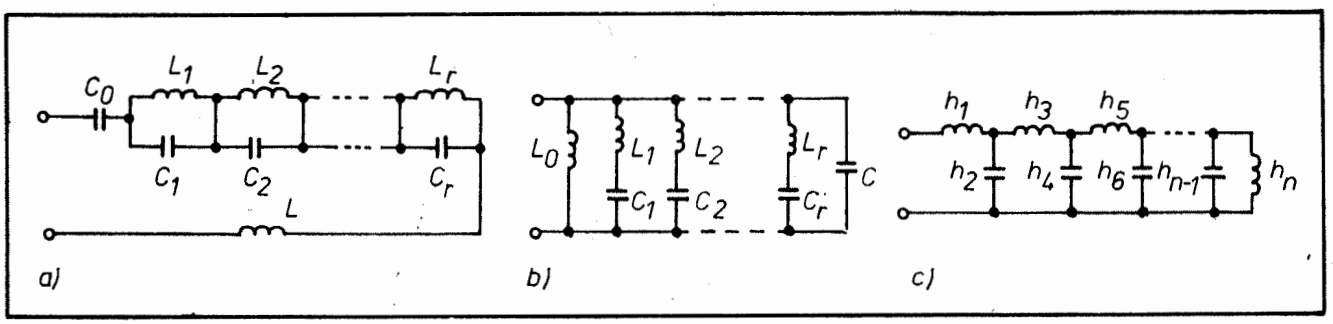


Bild 2.4.5. Darstellung einiger kanonischer Formen von Zweipolnetzwerken. a) und b) beruhen auf der Partialbruchzerlegung und c) auf einem unendlichen Kettenbruch für die Darstellung der Zweipolfunktion

bruch über:

$$H(p) = h_1(p) + \cfrac{1}{h_2(p) + \cfrac{1}{h_3(p) + \cfrac{1}{\quad + h_{n-1}(p) + \cfrac{1}{h_n(p)}}}}. \tag{13}$$

Damit entsteht eine Kettenschaltung gemäß Bild 2.4.5c). Gut lassen sich Kettenschaltungen aus Gl. (6b) herleiten. Dabei entsteht der Vorteil, daß die einzelnen Pole und Nullstellen weitgehend unabhängig voneinander realisiert werden können. Daraus, daß jedes reale System auf diese Form gebracht werden kann, folgt, daß *jedes reale System* in seinen Eigenschaften *als Kettenschaltung realisiert* werden kann. Allein diese hier gegebene kleine Auswahl von Schaltungen zeigt, daß es zu jeder Zweipolfunktion *mehrere technische Realisierungen* gibt. Das Gebiet der *äquivalenten Schaltungen* ist dementsprechend in der Literatur sehr breit untersucht, z.B. [F 19, F 21, R 25, U 2].

2.5. Theorie der Filter

2.5.1. Allgemeine Einführung

Die kompliziertesten Strukturen aus passiven Bauelementen sind die Filter. Sie werden in der Elektronik in vielfältigen Varianten und Anwendungsfällen benötigt. Neben den hier zu behandelnden Filtern aus passiven Bauelementen wird auf sie bezüglich anderer Realisierungsformen noch in den Abschnitten 6.3.4., 6.3.5., 7.5.8. und 7.9., 9. 3.4. eingegangen.

Die Theorie der Filter hat heute einen sehr hohen Stand erreicht. Dadurch ist es möglich, relativ schnell für jeden Anwendungsfall ein optimales Filter zu finden. Diese Theorie kann hier nur angedeutet werden. Dabei existieren *zwei* getrennte *Lösungswege*:

1. die Wellenparametertheorie
2. die Betriebsparametertheorie.

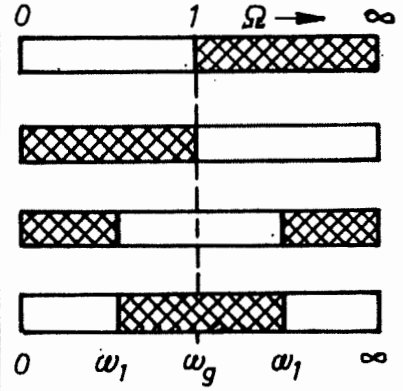
Die erste ist älter und gestattet nur iterativ brauchbare Lösungen. Sie schien deshalb zwischenzeitlich völlig bedeutungslos zu werden. Sie erlangt jedoch wieder eine ge-

wisse Bedeutung, insbesondere bezüglich des Einflusses der Spulenverluste. Die zweite Theorie geht von den Forderungen aus und führt auf geschlossenem Weg bis zu den endgültigen Bauelementen. Eine besondere Bedeutung bei der Filtertheorie hat die *Normierung*. Es wird primär nicht mit Frequenzen, Kapazitäten, Induktivitäten und Widerständen selbst gerechnet, sondern mit auf Eins normierten Größen. Weiter werden nicht die vier möglichen *Filterfälle* (TP, HB, BP und BS) sofort erfaßt, sondern primär nur der *Tiefpaß* (TP). Die anderen Fälle ergeben sich daraus durch eine *Transformation*. So gilt Tab. 2.5.1. Dabei ist zu beachten, daß auch Bauelemente entsprechend ausgetauscht werden, z.B. TP: L; HP: C; BP: Serienschwingkreis; BS: Parallelschwingkreis.

Tabelle 2.5.1. Transformation von den normierten Tiefpaßfrequenzen Ω zu den wirklichen Frequenzen

Es bedeuten ω_g die Grenzfrequenz bei TP und HP; $\omega_g = \sqrt{\omega_{-1}\omega_1}$ die Mittenfrequenz; $\gamma = (\omega_1 - \omega_{-1})/2$ die halbe Band- bzw. Sperrbreite bei Bandpaß bzw. Bandsperre. ▩ bedeutet den gesperrten Frequenzbereich. $\Omega = 1$ entspricht dem Übergang vom Sperr- zum Durchlaßbereich.

TP	Tiefpaß	$\omega = \Omega\omega_g$
HP	Hochpaß	$\omega = -\omega_g/\Omega$
BP	Bandpaß	$\omega = \sqrt{\omega_0^2 + (\gamma\Omega)^2} \pm \gamma\Omega$
BS	Bandsperre	$\omega = \sqrt{\omega_0^2 + (\gamma/\Omega)^2} \pm \gamma/\Omega$



2.5.2. Wellenparametertheorie

Die Basis der Wellenparametertheorie ist das *Halbglied* aus einem Längs- und einem Querwiderstand, r_l bzw. r_q. Beide können bereits zusammengesetzt sein, deshalb die komplexe Schreibweise. Zwei Halbglieder können zu einem Π- bzw. T-Glied zusammengesetzt werden. Dementsprechend werden beim Halbglied die Π- und T-Seite unterschieden (Bild 2.5.1). Bei einem Halbglied sind unmittelbar der *Leerlauf-* und der *Kurzschlußwiderstand* $\boldsymbol{W}_L$ bzw. $\boldsymbol{W}_K$ und daraus der Widerstand $\boldsymbol{Z}$ zu bestimmen.

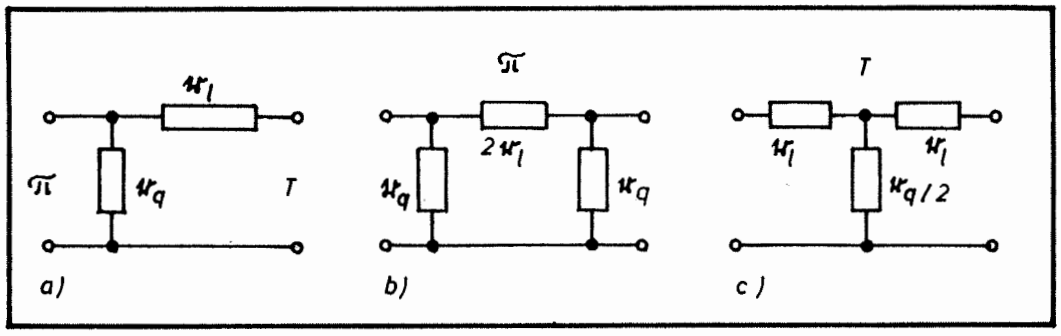


Bild 2.5.1. a) Halbglied mit Π- und T-Seite

b) Π-Vollglied

c) T-Vollglied

Tabelle 2.5.2. Leerlauf-, Kurzschluß- und Wellenwiderstand des Halbgliedes nach Bild 2.5.1.

	T-Seite	Π-Seite
W_L	$r_l + r_q$	r_q
W_K	r_l	$\frac{r_l \cdot r_q}{r_l + r_q}$
$Z = \sqrt{W_L \cdot W_K}$	$\sqrt{r_l \cdot r_q}\sqrt{1 + \frac{r_l}{r_q}}$	$\frac{\sqrt{r_l \cdot r_q}}{\sqrt{1 + \frac{r_l}{r_q}}}$

Hierfür gilt Tab. 2.5.2. Die *Wellenwiderstände* sind dann reell, wenn

$$r_l \cdot r_q = Z_L^2 \tag{1}$$

gilt. Z_L ist der Nennwert des Wellenwiderstandes. Ist z.B. der Längswiderstand eine Induktivität und der Querwiderstand eine Kapazität, so gilt

$$Z_L = \sqrt{\frac{L}{C}}. \tag{2}$$

Für die *normierte Frequenz* Ω gilt damit

$$\frac{r_l}{Z_L} = \frac{Z_L}{r_q} = j\Omega\,. \tag{3}$$

Der Wellenwiderstand aller möglichen Filterschaltungen ist nur im *Durchlaßbereich* reell. Für ihn gilt, je für die T- bzw. Π-Seite:

$$Z_T = Z_L\sqrt{1 - \Omega^2}\,, \qquad Z_\Pi = Z_L \frac{1}{\sqrt{1 - \Omega^2}}. \tag{4}$$

Die Auswertung dieser Gleichungen zeigt Bild 2.5.2. Daraus ist ersichtlich, daß ein Halbglied auf beiden Seiten mit einem reellen Widerstand abgeschlossen werden müßte, dessen Größe von der Frequenz abhängig ist. Da solche Widerstände nicht

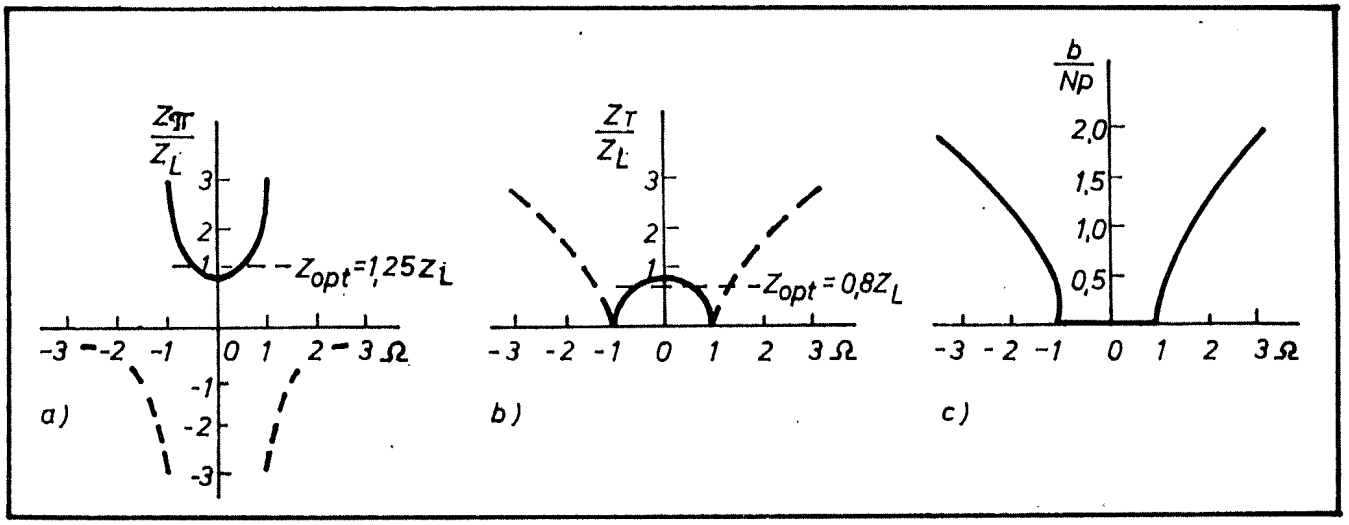


Bild 2.5.2. Verlauf des Wellenwiderstandes für die a) Π-Seite und b) T-Seite von Halbgliedern. Im Durchlaßbereich $(-1 < \Omega < +1)$ ist der Verlauf reell (ausgezogen). Im Sperrbereich ist er komplex (gestrichelt). Da es einen frequenzabhängigen reellen Widerstand nicht gibt, wird ein Mittelwert für den Durchlaßbereich als Z_{opt} gewählt. c) zeigt die Wellendämpfung des ideal abgeschlossenen Halbgliedes

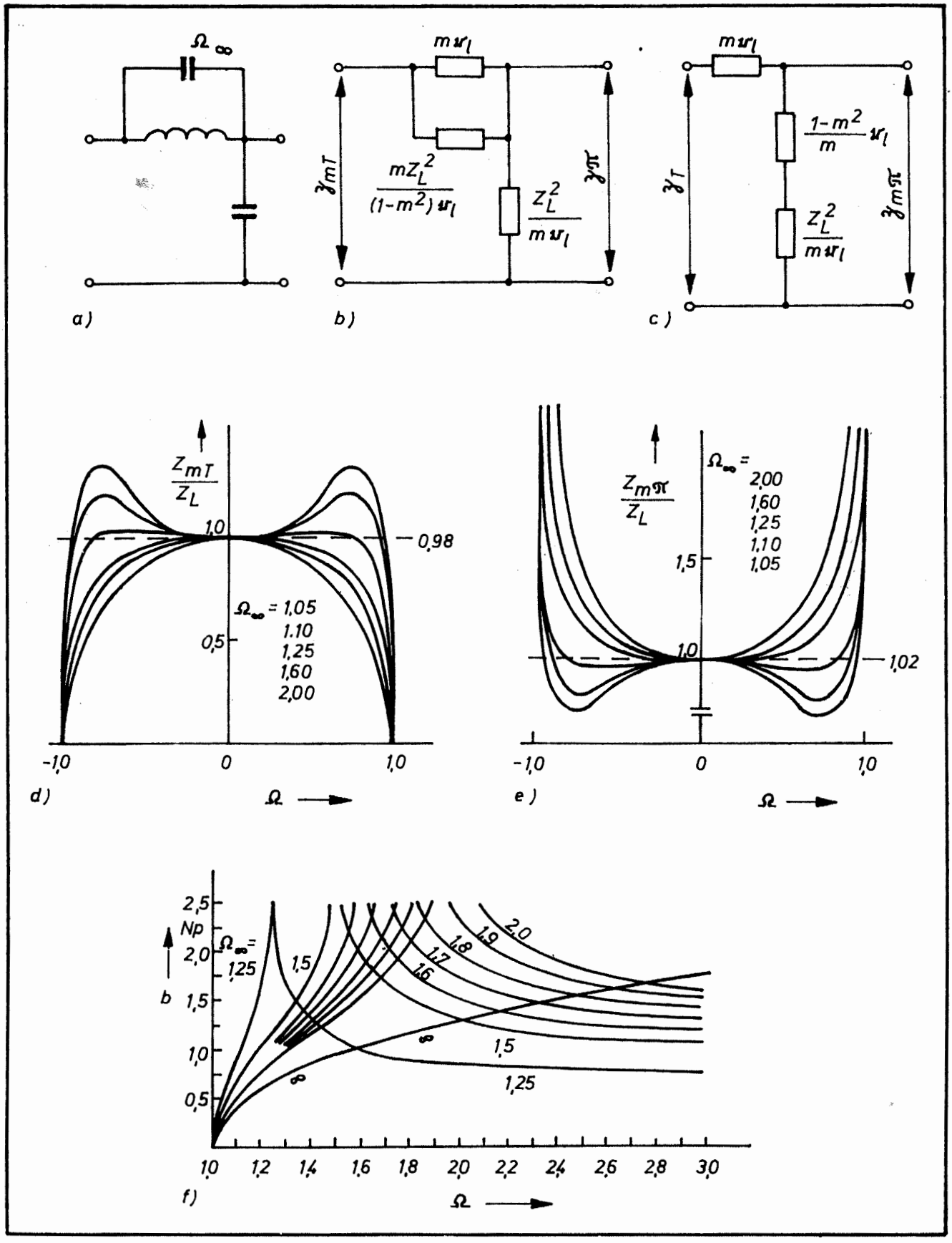


Bild 2.5.3. Aufbau und Eigenschaften von ZOBEL-Halbgliedern

a) Schaltung des mT-Gliedes für einen Tiefpaß
b) Dimensionierung des allgemeinen mT-Gliedes
c) Dimensionierung und Schaltung des mΠ-Gliedes
d) Wellenwiderstand eines mT-Gliedes
e) Wellenwiderstand eines mΠ-Gliedes
f) Wellendämpfung von ZOBEL-Halbgliedern

realisierbar sind, wird in der *Praxis eine Näherung* verwendet:

$$Z_{\mathrm{T,\,opt}} = 0{,}8Z_{\mathrm{L}}\,,$$
$$Z_{\Pi,\,\mathrm{opt}} = 1{,}25Z_{\mathrm{L}}\,. \tag{5}$$

Hierdurch treten im Durchlaß- und im Sperrbereich *Abweichungen gegenüber der Theorie* auf. Sie verflachen den Übergang vom Durchlaß- zum Sperrbereich und rufen zusätzlich Phasenfehler hervor. Durch die direkte Zusammenschaltung mehrerer Halbglieder mit gleichartigem Wellenwiderstandsverhalten läßt sich an den Stellen dieser Zusammenschaltung eine ideale Anpassung erreichen. *Fehler* treten dann *nur* noch je einmal *an den Enden* des zusammengesetzten Filters auf.

Für den Tiefpaß besteht das Halbglied aus einer Längsinduktivität und einer Querkapazität. Ein Aufbau der Halbglieder aus drei Bauelementen führt zu den ZOBEL- bzw. *m-Halbgliedern*. Sie besitzen *zwei Vorteile*:

- Der Frequenzgang des Wellenwiderstandes (Bild 2.5.2a) und b)) läßt sich verbessern.
- Die Dämpfung gemäß Bild 2.5.2c) läßt sich versteilern.

Dabei wird ein Parameter m frei wählbar. Er ist unmittelbar mit dem Parameter Ω_∞ gekoppelt:

$$\Omega_\infty = \frac{1}{1-m^2} > 1 \quad \text{bzw.} \quad m = \sqrt{1 - 1/\Omega_\infty}\,. \tag{6}$$

Diese normierte Frequenz ist ein Dämpfungspol. Dies ist neben dem Aufbau der ZOBEL-Glieder und anderen Eigenschaften aus Bild 2.5.3 deutlich zu erkennen. Weiter ist aus Bild 2.5.3 zu erkennen, daß es *zwei Typen* von ZOBEL-Gliedern gibt, die sich dadurch unterscheiden, daß im Längs- oder Querzweig zwei Blindelemente enthalten sind. Die Wellenwiderstände der ZOBEL-Glieder am $\boldsymbol{Z}_{\mathrm{T}}$- bzw. $\boldsymbol{Z}_{\Pi}$-Anschluß entsprechen genau denen der einfachen Halbglieder. Auf diese Weise ist immer eine Zusammenschaltung beliebiger Vollglieder möglich. Den Verlauf des Wellenwiderstandes am anderen Anschluß zeigt Bild 2.5.3d) bzw. e). Er folgt aus den Gleichungen

$$Z_{m\Pi} = Z_\Pi\left(1 - \frac{\Omega^2}{\Omega_\infty^2}\right), \qquad Z_{m\mathrm{T}} = Z_{\mathrm{T}}\,\frac{1}{1 - \dfrac{\Omega^2}{\Omega_\infty^2}}\,. \tag{7}$$

Aus Bild 2.5.3d) und e) ist ersichtlich, daß ein besonders geradliniger Verlauf für $m = 0{,}6$ bzw. $\Omega_\infty = 1{,}25$ vorhanden ist. Als *optimale Anpassung* folgt daraus

$$Z_{m\Pi,\,\mathrm{opt}} = 1{,}02Z_{\mathrm{L}}\,, \qquad Z_{m\mathrm{T},\,\mathrm{opt}} = 0{,}98Z_{\mathrm{L}}\,. \tag{8}$$

Aufgrund dieser Eigenschaft werden bei den meisten Siebschaltungen diese ZOBEL-Halbglieder als *Abschlußglieder* verwendet. ZOBEL-Halbglieder mit anderen Unendlichkeitsfrequenzen werden vor allem wegen ihrer Selektionserhöhung im Sperrbereich verwendet. Gebräuchlich sind dabei Werte für $\Omega_\infty = 1{,}05;\ 1{,}1;\ 1{,}25;\ 1{,}6$ und $1{,}7$. Der Aufbau eines Filters unter Zusammenziehung der verschiedenen Halb- und Vollglieder ist in Bild 2.5.4 dargestellt. Hieraus ist auch ersichtlich, wie durch Vertauschen von Π und T-Gliedern eine minimale Zahl von Induktivitäten zum Aufbau erreicht werden kann. Bild 2.5.4 zeigt weiter einige übliche Schaltungen mit ihren charakteristischen Eigenschaften. Ohne Probleme ist der Durchlaßbereich nur bis etwa 0,8 nutzbar. Die Berücksichtigung der Dämpfung der Spule führt zu komplizierten Betrachtungen.

Im Zentrum der Wellenparametertheorie steht der Wellenwiderstand von Halbgliedern. Die einzelnen Halbglieder werden mit Ausnahme der Endglieder zu Voll-

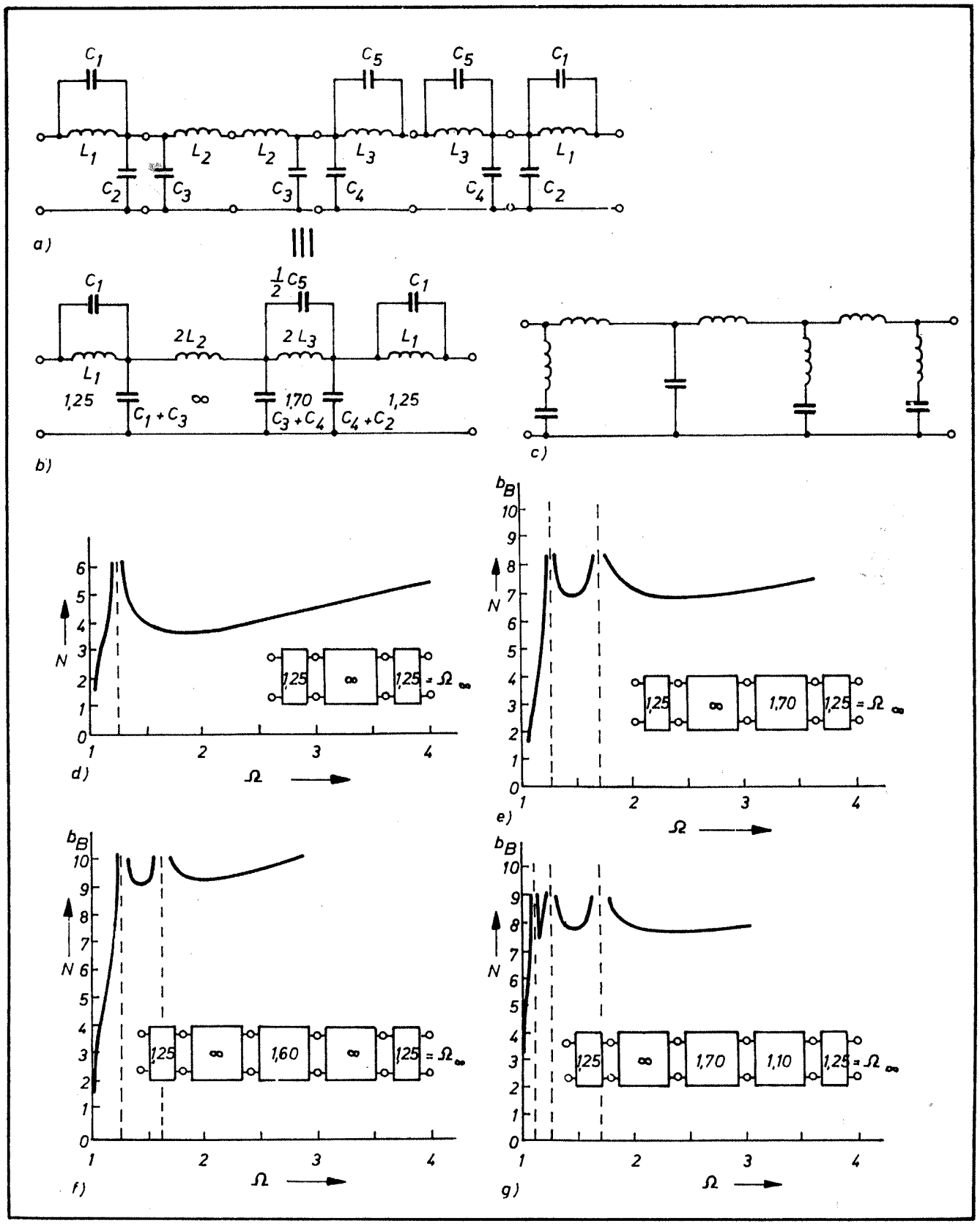


Bild 2.5.4. a) Filteraufbau mit T-Glied-Abschluß mit den Ω_∞-Größen 1,25; ∞; 1,70; 1,25

b) Zusammenfassung von Induktivitäten und Kapazitäten für den realen Aufbau

c) Gleiches Filter, jedoch mit Π-Abschluß, benötigt weniger Kapazitäten, aber mehr Induktivitäten

d) bis g) Vier Beispiele von Standardfiltern mit ihren Dämpfungsverläufen

gliedern zusammengefaßt und dann so gekoppelt, daß keine Reflexionsverluste auftreten. Bei den Endgliedern werden gewisse Verluste mit Dämpfungs- und Phasenfehlern als Folge in Kauf genommen. Für eine gewünschte Filterfunktion wird aus einer großen Tabelle ein ähnliches Filter, also m- bzw. Ω_∞-Werte, entnommen und dann auf technische Werte umgerechnet und so realisiert. Ein Hauptproblem besteht darin, daß auf diese Weise keine geschlossene Berechnung für das gewünschte Filter möglich ist.

2.5.3. Betriebsparametertheorie

In der Betriebsparametertheorie geht man generell von den *Forderungen des Filters* aus. Sie sind durch *drei Größen* gegeben:

A_{max} = maximale Dämpfung in dB im Bereich der normierten Frequenzen 0 bis 1

A_{min} = minimale Dämpfung in dB im Sperrbereich von Ω_s bis ∞

Ω_s = Beginn des Sperrbereiches mit $\Omega_s > 1$.

Zur einfacheren Schreibweise wird hier eingeführt

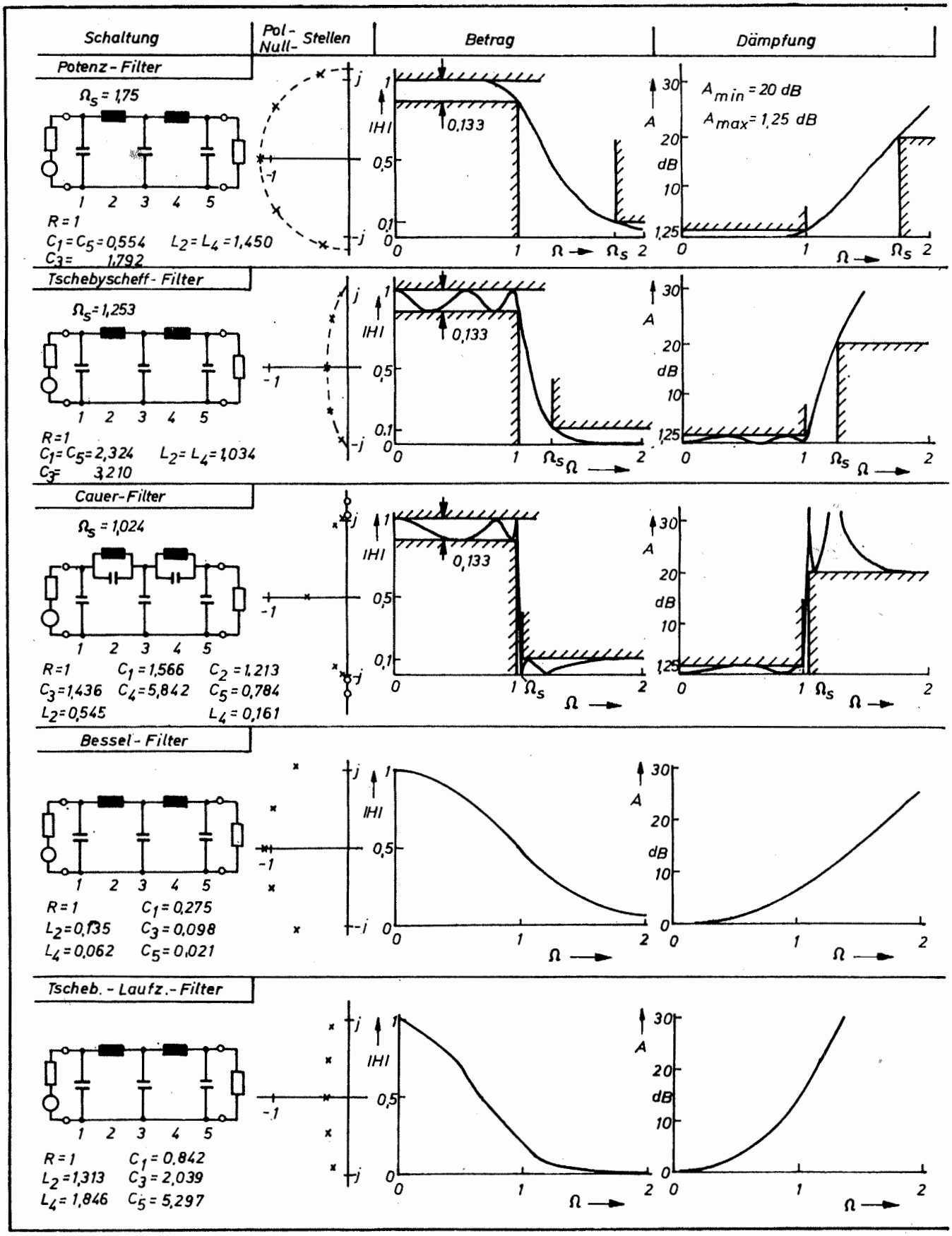
$$A = \sqrt{\frac{10^{0,1\,A_{max}} - 1}{10^{0,1\,A_{min}} - 1}}. \tag{9}$$

Ist dieses Toleranzfeld festgestellt, so müssen Entscheidungen darüber getroffen werden, ob eine *Welligkeit* im Durchlaß und/oder im Sperrbereich zulässig ist. Hiernach ergeben sich die *drei* ersten, *typischen Fälle* des Bildes 2.5.5:

- *Potenz-*, auch BUTTERWORTH- oder GAUSS-*Filter* genannt. Bei ihm verläuft die Dämpfungskurve glatt.
- TSCHEBYSCHEFF-*Filter*. Bei ihm existiert im Durchlaßbereich eine Dämpfungskurve, welche voll und gleichmäßig den zulässigen Bereich 0 bis A_{max} mit Welligkeit ausfüllt. Die Welligkeit wird um so dichter, je größer die Ordnung n des Filters ist. Neben diesem normalen TSCHEBYSCHEFF-Filtertyp gibt es ein zweites, das die Welligkeit im Sperrbereich besitzt (Bild 2.5.6a)).
- CAUER-, auch *elliptisches Filter* genannt. Es besitzt im Durchlaß- und Sperrbereich die gleichmäßige Welligkeit. Auch bei diesem Filter gibt es mehrere Varianten. Sie hängen vom Grad n ab und davon, ob bei $\Omega = 0$ die Dämpfung A_{max} bzw. A_{min} beträgt. Den typischen Fall für ungerades n zeigt Bild 2.5.5, die beiden Fälle für gerades n Bild 2.5.6b), c).

Sind auch diese Entscheidungen gefallen, so liefert die dann folgende, *geschlossene Rechnung* unmittelbar die normierten Pole und Nullstellen. Wenn nun weiter — wie allgemein üblich — die Kettenschaltung benutzt wird, verläuft die Rechnung über die Entnormierung bis zur Festlegung aller Bauelemente.

Dieser geschlossene *Rechnungsgang* kann heute ohne Probleme *interaktiv* bereits auf Tischrechnern realisiert werden und ist für den Entwickler daher von großem Vorteil. Auf diesem Prozeß bauen u. a. auch die später zu behandelnden Aktivfilter (7.5.8.) und die Digitalfilter (7.9.) auf. Selbst für den im Abschnitt 9.3.4. zu behandelnden Analogprozessor ist er von grundlegender Bedeutung. Deshalb seien hier einige weitere Details angeführt.


Schaltung
Pol- Null- Stellen
Betrag
Dämpfung
Potenz-Filter
Ωs = 1,75
R = 1
C1 = C5 = 0,554
L2 = L4 = 1,450
C3 = 1,792
Amin = 20 dB
Amax = 1,25 dB
Tschebyscheff-Filter
Ωs = 1,253
R = 1
C1 = C5 = 2,324
L2 = L4 = 1,034
C3 = 3,210
Cauer-Filter
Ωs = 1,024
R = 1
C1 = 1,566
C2 = 1,213
C3 = 1,436
C4 = 5,842
C5 = 0,784
L2 = 0,545
L4 = 0,161
Bessel-Filter
R = 1
C1 = 0,275
L2 = 0,735
C3 = 0,098
L4 = 0,062
C5 = 0,021
Tscheb. - Laufz. - Filter
R = 1
C1 = 0,842
L2 = 1,313
C3 = 2,039
L4 = 1,846
C5 = 5,297

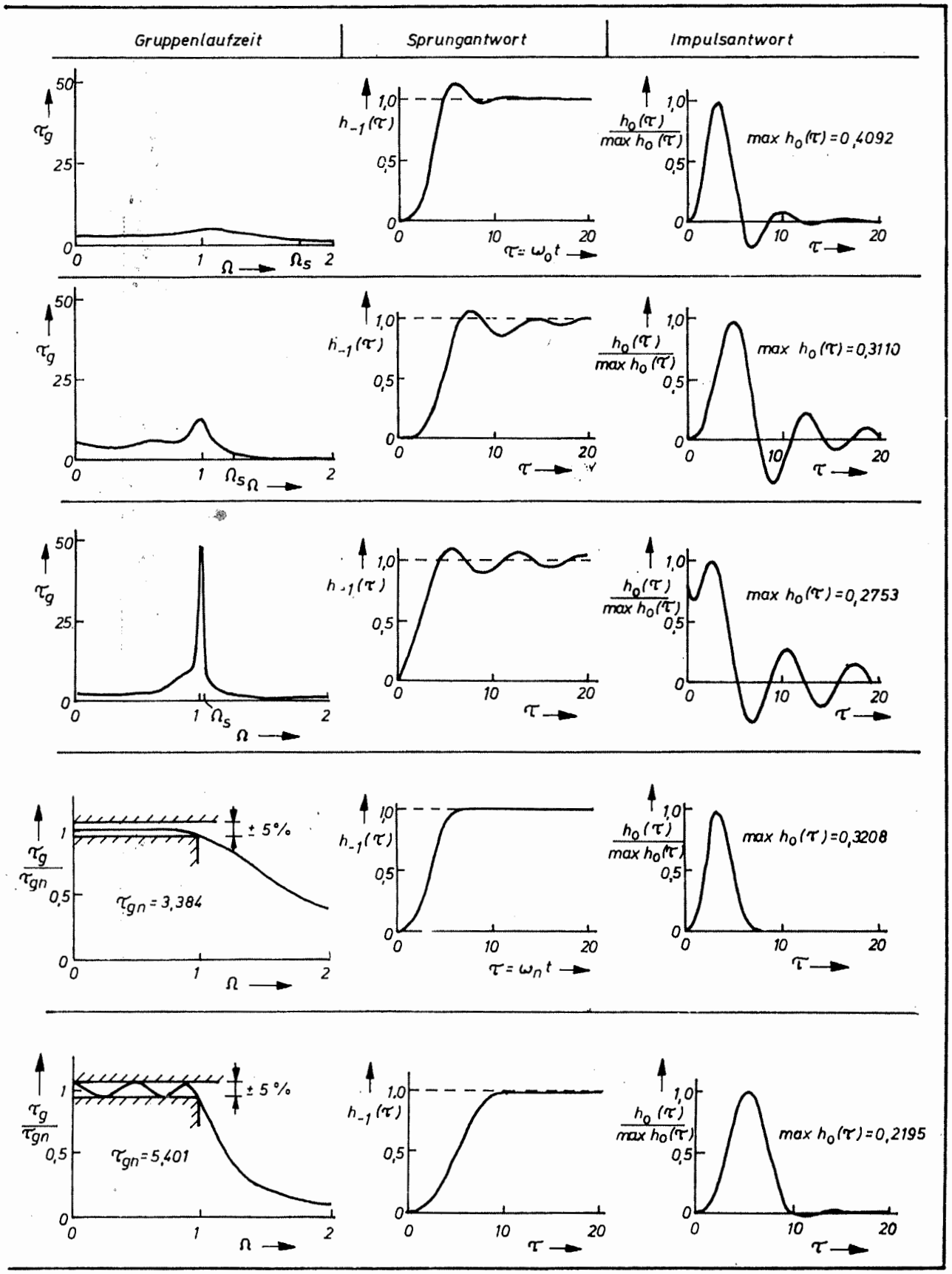


Bild 2.5.5. Aufbau der Pole und Nullstellen sowie Eigenschaften der wichtigsten Filtertypen. Die ersten drei Filtertypen dienen der Frequenzselektion, bis $\Omega = 1$ geht der Durchlaßbereich, bei Ω_s beginnt der Sperrbereich. Die beiden unteren Filtertypen sind bezüglich einer konstanten Laufzeit approximiert [S 66]

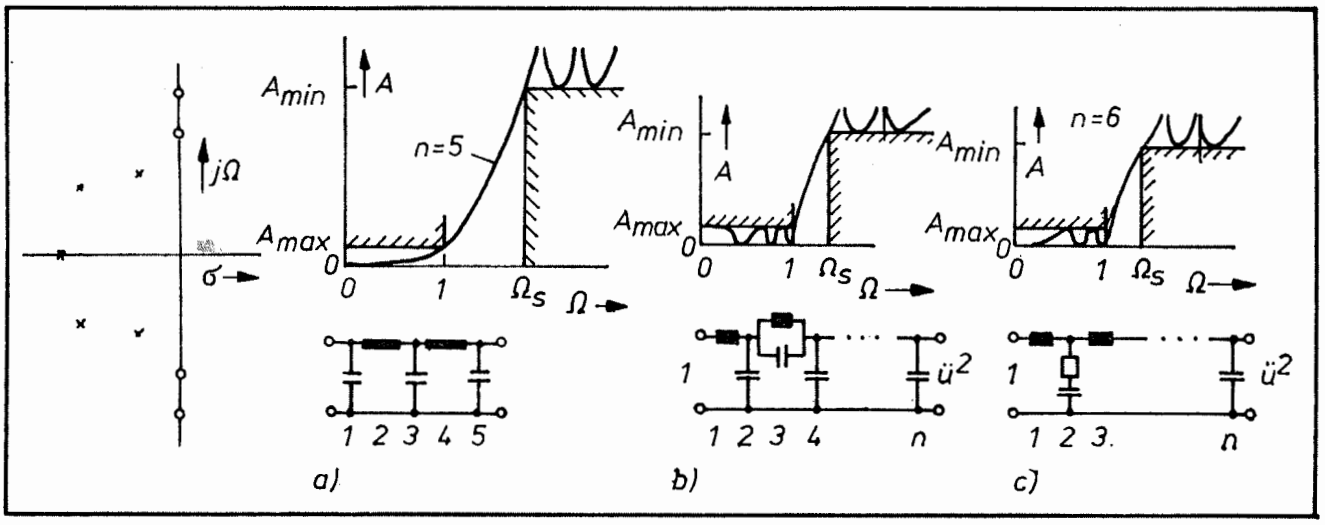


Bild 2.5.6. Ergänzungen zum Bild 2.5.5. a) zeigt Pole und Nullstellen, Aufbau und Dämpfungsverlauf des TSCHEBYSCHEFF-Filters II. Art. b) und c) zeigen Aufbau und Dämpfungsverlauf der beiden CAUER-Filter bei gerader Ordnung [F 22]

2.5.3.1. Potenzfilter

Beim Potenzfilter wird keine Welligkeit im Durchlaß- und Sperrbereich zugelassen. Dann leistet eine Funktion

$$D(\Omega) = \delta \Omega^n \tag{10}$$

das Geforderte. Die Funktion wächst monoton. Die natürliche Zahl n bestimmt, wie steil die Funktion für $\Omega > 1$ ansteigt, und δ ist eine Konstante, welche bewirkt, daß der Verlauf der Funktion ins Toleranzfeld paßt. Für den Betrag der Übertragungsfunktion gilt weiter

$$H(\mathrm{j}\Omega) \cdot H(-\mathrm{j}\Omega) = \frac{1}{1 + D^2(\Omega)} = 10^{-0{,}1\,A(\Omega)}\,. \tag{11}$$

Infolge der obigen Bedingungen gilt

$$\left.\begin{aligned} \Omega &= 1 \colon 1 + \delta^2 = 10^{0{,}1\,A_{\max}}\,, \\ \Omega &= \Omega_s \colon 1 + \delta^2 \Omega^{2n} = 10^{0{,}1\,A_{\min}}\,. \end{aligned}\right\} \tag{12}$$

Aus diesen beiden Gleichungen läßt sich sofort der Grad n berechnen:

$$n \geqq \frac{\log A}{\log \Omega_s}\,. \tag{13}$$

Für n wird also die nächstgrößere ganze Zahl benutzt. Die Auflösung der obigen Gleichung nach den Polen liefert

$$p_k = \frac{1}{\sqrt[n]{\delta}}\left[-\sin\frac{2k-1}{2n}\pi + \mathrm{j}\cos\frac{2k-1}{2n}\pi\right] \quad \text{für} \quad k = 1, \ldots, n\,. \tag{14}$$

Sie liegen alle auf einem Kreis mit dem Radius $1/\sqrt[n]{\delta}$. Es ist nur der links von der Ω-Achse liegende Teil als Schaltung zu realisieren.

2.5.3.2. TSCHEBYSCHEFF-Filter

Die gleichmäßige Approximation, d. h. Welligkeit vorgegebener Höhe im Durchlaß- bzw. Sperrbereich, leisten die TSCHEBYSCHEFF-*Funktionen* $T_n(\omega)$ gemäß dem Ansatz

$$D(\Omega) = \delta T_n(\Omega) \tag{15}$$

Sie lassen sich auf verschiedene Weise, z.B. iterativ oder gemäß

$$T_n(\Omega) = \begin{cases} \cos\,(n \arccos \Omega) \\ \cosh\,(n \operatorname{arccosh} \Omega) \end{cases} \text{ für } \Omega \begin{cases} \leqq \\ > \end{cases} 1\,, \tag{16}$$

berechnen. Für die Ordnung n (Anzahl der Pole, Nullstellen) folgt dann

$$n \geqq \frac{\operatorname{arccosh} A}{\operatorname{arccosh} \Omega_s}\,. \tag{17}$$

Die Pole liegen auf einer Ellipse gemäß

$$p_k = -s_1 \sin \frac{2k-1}{n}\frac{\pi}{2} + \mathrm{j} s_2 \cos \frac{2k-1}{n}\frac{\pi}{2} \qquad \text{für} \quad k = 1, \ldots, n \tag{18}$$

mit der reellen Halbachse

$$s_1 = \sinh\left(\frac{1}{n} \operatorname{arcsinh} \frac{1}{\delta}\right) \tag{19}$$

und der imaginären Halbachse

$$s_2 = \cosh\left(\frac{1}{n} \operatorname{arcsinh} \frac{1}{\delta}\right). \tag{20}$$

2.5.3.3. CAUER-Filter und Vergleiche

Der Rechengang bei den CAUER-Filtern verläuft analog zu den beiden anderen. Er wird jedoch dadurch komplizierter, daß bei ihm die entscheidenden Funktionen die *vollständigen elliptischen Integrale* sind. Sie werden meist durch rationale Polynome approximiert oder rekursiv berechnet.

Aus dem *Vergleich der drei Filtertypen* (Potenz-, TSCHEBYSCHEFF- und CAUER-Filter) in Bild 2.5.5 ist deutlich folgendes zu erkennen: Bei festgehaltener Ordnung (im Bild: 5) nimmt der Übergang vom Durchlaßbereich zum Sperrbereich stark zu: $\Omega_s = 1{,}75;\ 1{,}253;\ 1{,}024$. Diese erhöhte Selektivität bedingt dagegen zwei andere Mängel: einmal die Welligkeit und zum anderen unausgeglichenere Laufzeit sowie stärkeres Einschwingen. Hieraus wird die Forderung nach speziellen Filtern deutlich, die gerade hier in Toleranzschemata passen. Die beiden typischen *Laufzeit-approximierten Filter* enthält ebenfalls Bild 2.5.5. Sie zeigen zugleich gutes Einschwingverhalten.

Drei bezüglich des *Zeitverhaltens approximierte Filter* enthält Bild 2.5.7. Die beiden obersten zeigen eine gewisse Dämpfung im Sperrbereich (im Bild: 40 dB), und bezüglich des Einschwingens bei der Anregung mit der Sprungfunktion bzw. dem Deltastoß ist eine Toleranz (im Bild :1%) festgelegt.

Das dritte Filter von Bild 2.5.7 hat die Aufgabe, das Eingangssignal bezüglich eines *gleitenden Rechtecks* der Länge 2π *zu mitteln*. Tab. 2.5.3 faßt die Daten der Pole und Nullstellen zu den Bildern 2.5.5 und 2.5.7 zusammen. Die Nullstellen entsprechen den Resonanzfrequenzen der Parallelschwingkreise in den Filtern.

Aus den letzten Beispielen dürfte deutlich geworden sein, daß es heute für die Filtertechnik eine *Vielzahl von Approximationsaufgaben* gibt. Sie betreffen einzeln oder gemeinsam folgende *drei Ziele*:

- Übertragung bzw. Dämpfung bez. der Frequenz
- Laufzeitverhalten
- Zeitverhalten.

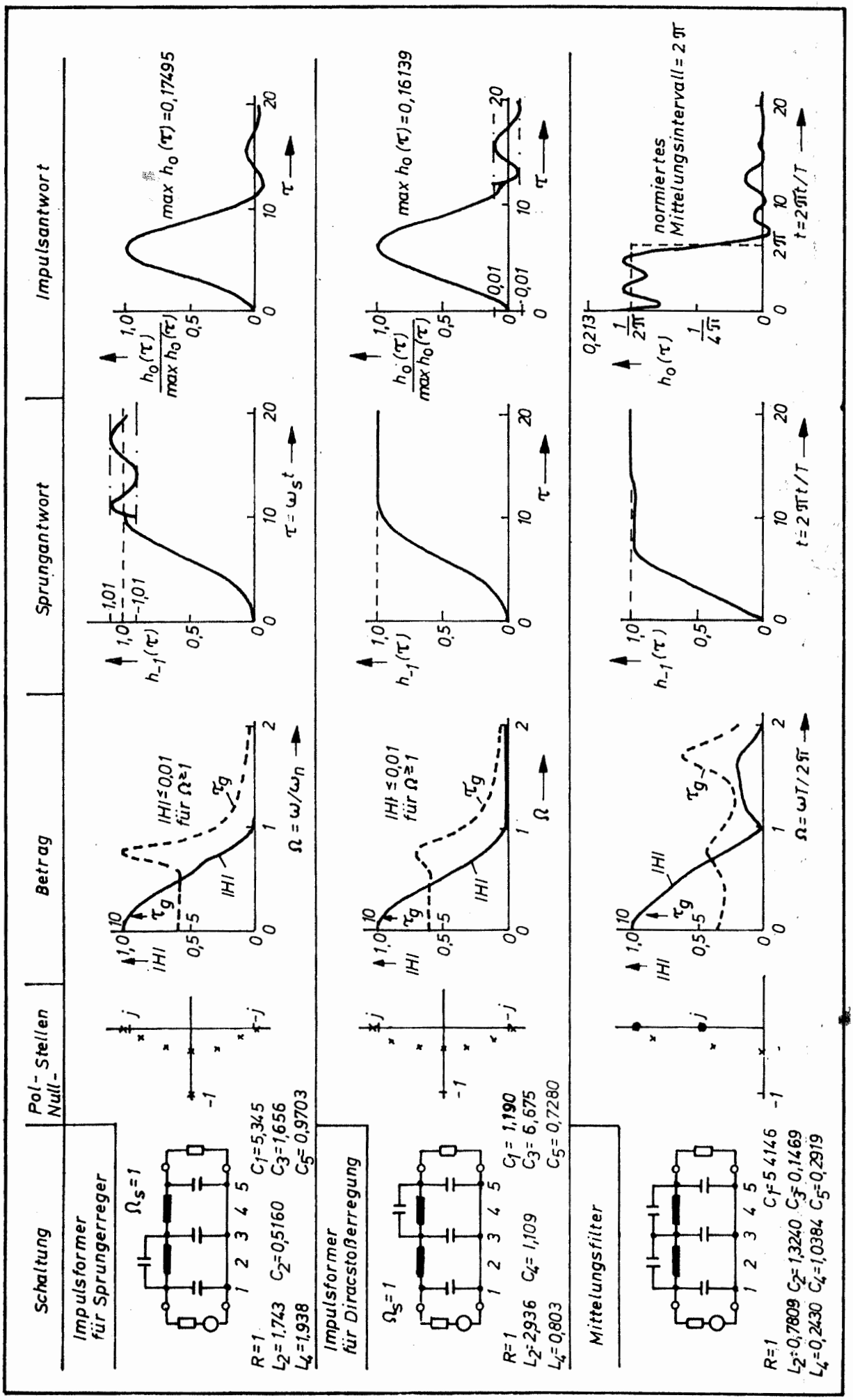


Bild 2.5.7. Filter mit spezieller Approximation für das Zeitverhalten [S 66]

Tabelle 2.5.3. Daten der Filter aus Bild 2.5.5 und 2.5.7

Filtertyp	$s_{\infty 1}$	$s_{\infty 2,4}$	$s_{\infty 3,5}$	$j\omega_{01,3}$	$j\omega_{03,4}$	b
Potenz	−1,116121 ±0,0 · j	−0,344901 ±1,061496 · j	−0,902963 ±0,656041 · j	—	—	1,732051
TSCHEBYSCHEFF	−0,266448 ±0,0 · j	−0,082336 ±0,984238 · j	−0,215560 ±0,608292 · j	—	—	0,108253
CAUER	−0,536828 ±0,0 · j	−0,014127 ±1,000059 · j	−0,145480 ±0,896079 · j	±1,033049	±1,229953	0,274122
BESSEL	−1,077641 ±0,0 · j	−0,990531 ±0,514970 · j	−0,68690 ±1,055267 · j	—	—	2,129472
TSCHEBYSCHEFF-Laufz.	−0,300282 ±0,0 · j	−0,292381 ±0,485877 · j	−0,245340 ±0,937654 · j	—	—	0,090707
Impulsformer 1	−0,35686 ±0,0 · j	−0,30094 ±0,37416 · j	−0,12306 ±0,76144 · j	1,05435	—	0,044034
Impulsformer □	−0,33101 ±0,0 · j	−0,30169 ±0,38178 · j	−0,19657 ±0,77908 · j	—	1,05834	0,045173
Mittelungsfilter	−0,40515 ±0,0 · j	−0,29454 ±0,77984 · j	−0,17811 ±1,69462 · j	0,98345	1,99094	0,213226

Mit einem entsprechenden Filteraufwand sind die meisten Probleme (sofern sie physikalisch zulässig sind) gut approximierbar. Es gibt aber auch noch Probleme, die rechentechnisch äußerst schwierig und damit nicht geschlossen lösbar sind. Für die meisten Anwendungsfälle existieren jedoch komplexe, iterativ abzuarbeitende Algorithmen.

Abschließend sei noch erwähnt, daß es für die Pole und Nullstellen sowie weitere Daten von Filtern auch umfangreiche *Tabellenwerke* gibt, z.B. [F 17, F 20, S 1].

2.5.3.4. HF-Bandfilter

Bandfilter sind aus den vorangegangenen Filtern formal mittels der Transformation Bandpaß in Tab. 2.5.1 zu gewinnen. Dadurch werden die mit den Tiefpaßgleichungen berechneten Pole und Nullstellen an zwei Stellen des eigentlichen ω-Frequenzbereichs gelegt. Hochfrequenzbandfilter vereinfachen diese Operation etwas dadurch, daß die *Mittenfrequenz ω_g groß gegenüber der Bandbreite* $\omega_1 - \omega_{-1}$ ist. Es sei jetzt weiter angenommen, daß zwei gleiche Schwingkreise mit dem Faktor k verkoppelt sind. Dann liegen für kleine ϱ und k je zwei Pole auf einem Halbkreis um die Punkte $\pm j$ mit dem Radius $r = \frac{1}{2}\sqrt{\varrho^2 + k^2}$, und der Ursprung ist eine Nullstelle dritter Ordnung. Exakt gilt

$$H(s_n) = \frac{k\varrho}{1-k^2} \frac{s_n^3}{\left(s_n^2 + \frac{\varrho}{1-k} s_n + \frac{1}{1-k}\right)\left(s_n^2 + \frac{\varrho}{1+k} s_n + \frac{1}{1+k}\right)}, \tag{21}$$

für $\varrho \ll 1$ und $k \ll 1$ folgt

$$s_{n\infty 1 \ldots 4} \approx -\frac{\varrho}{2} \pm \mathrm{j}\left(1 \mp \frac{k}{2}\right). \tag{22}$$

Die entsprechende Darstellung für die positive ω-Achse zeigt Bild 2.5.8. In dem Bild ist weiterhin der Frequenzgang für die normierte Kopplung ψ als Parameter dargestellt.

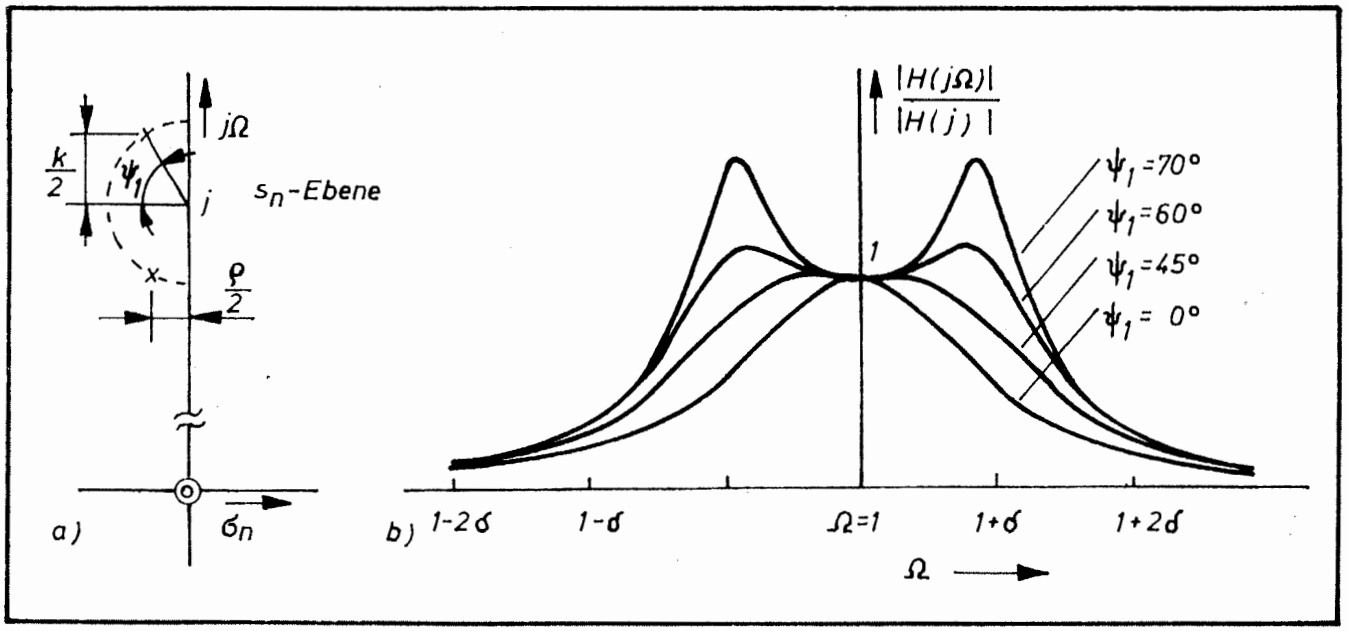


Bild 2.5.8. a) Lage der Pole und Nullstellen der Übertragungsfunktion $H(s_n)$ für gekoppelte Schwingkreise

b) Betragsfrequenz $|H(\mathrm{j}\Omega)|$ bei gekoppelten Schwingkreisen

3. Grundlagen der Halbleiterbauelemente

In diesem Kapitel werden die Eigenschaften von Halbleitermaterialien, Halbleiterwiderständen, Dioden, Feldeffekt- und Bipolartransistoren behandelt. Ergänzend hierzu werden behandelt

- Technologie der Halbleitertechnik in Kapitel 4.
- Schaltungen mit Standardhalbleitern in Kapitel 5.
- Sonderbauelemente, wie Optoelektronik, Thyristoren, Tunneldioden usw. in Kapitel 6.

3.1. Eigenschaften von Halbleitermaterialien

3.1.1. Bändermodell

Ein einzelnes Atom besteht aus einem positiven Kern, der in mehreren Schalen von den zugehörigen Elektronen umkreist wird. Für die weiteren Betrachtungen interessieren nur die Elektronen der äußersten Schale. Zur Vereinfachung sei zunächst ein einzelnes äußeres Elektron angenommen. Im stationären Zustand des Atoms läuft es auf einer stabilen Bahn um. Hierbei wird die Anziehungskraft zwischen dem positiven Kern und dem negativen Elektron gerade durch die Zentrifugalkraft kompensiert. Damit dieses Elektron auf einer weiter entfernten Bahn umläuft, muß ihm eine Energie zugeführt werden. So entsteht die Kurve der Potentialenergie in Bild 3.1.1a). Dieser Verlauf wird auch als *Potentialtopf* bezeichnet. Die Anziehung der Elektronen zum Kern hin kann damit auch so gedeutet werden, daß sie danach streben, möglichst tief in den Potentialtopf zu fallen.

Gemäß der *Quantentheorie* sind nicht alle Energien für ein Elektron zulässig. Seine *Materiewelle* wird, anschaulich interpretiert, an den Wänden des Potentialtopfes teilweise reflektiert und durchtunnelt teilweise die Potentialberge. Hierdurch entsteht *für gewisse Energiewerte eine Interferenz,* die Nichtexistenz der Materiewelle bedeutet. Elektronen mit derartigen Energien sind nicht möglich. So entstehen die *möglichen Energieterme* in Bild 3.1.1a).

Im *Kristall* (Festkörper) überlappen sich die Potentialtöpfe der einzelnen Atome, und es entsteht ein periodisches Potentialgebilde gemäß Bild 3.1.1b). Hierdurch lassen sich die äußeren Elektronen (oberen Terme) nicht mehr dem einzelnen Atom zuordnen. Diese Elektronen sind als Elektronengas frei beweglich. Dies bildete die Grundlage für die einfachen Betrachtungen auf der Basis der DRUDE-Theorie im Abschnitt 1.1.

Infolge der dicht beieinanderliegenden Atome tritt auch eine *Kopplung zwischen den Energietermen* ein. Sie werden in Linien mit leicht unterschiedlichem Niveau aufgespalten. Das PAULI-Verbot der Atomtheorie besagt nämlich stark vereinfacht, daß in einem Verband keine Elektronen mit der gleichen Energie (genauer mit gleichem

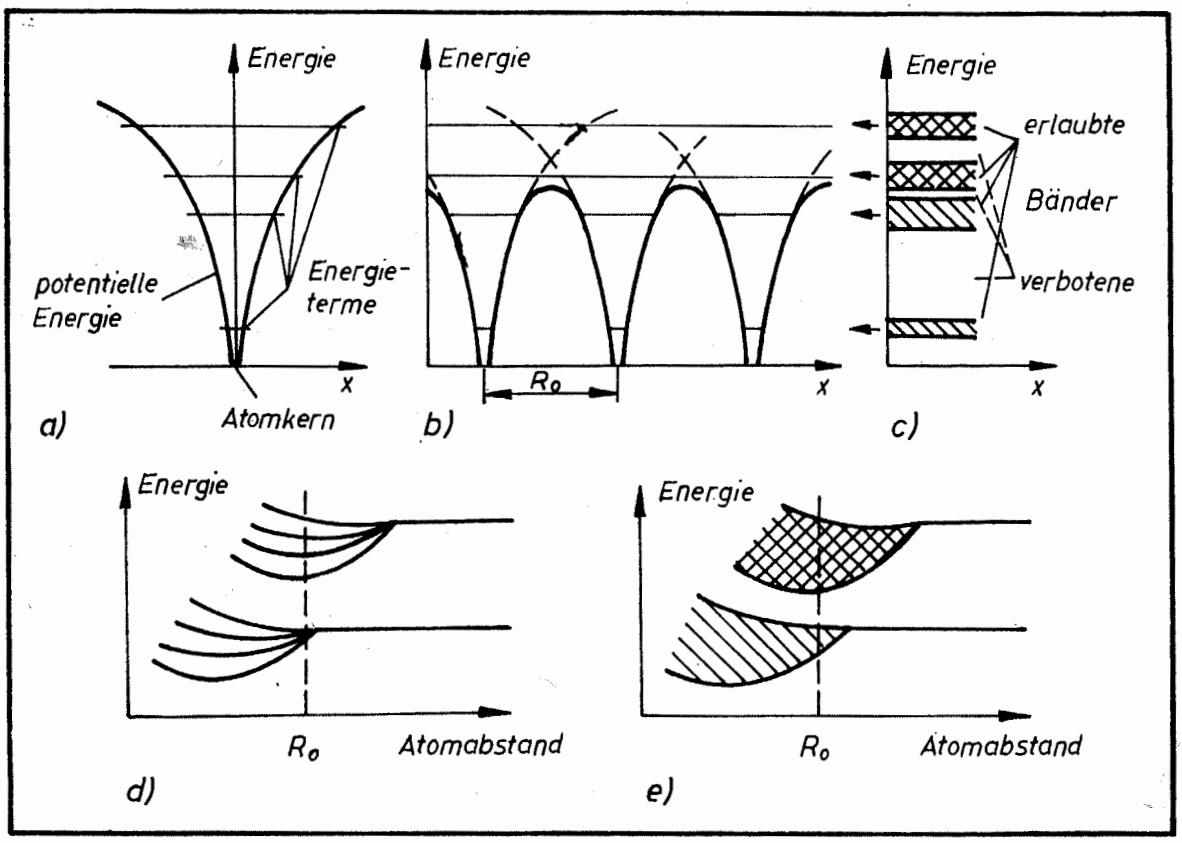


Bild 3.1.1. Zum Bändermodell

a) Potentialtopf und Energieterme eines einzelnen freien Atoms
b) Kopplung der Potentialtöpfe bei Atomen im Abstand R_0
c) Ausbildung der Bänder durch die Verkopplung der Atome infolge des PAULI-Verbotes
d) Aufspaltung von zwei Energietermen durch vier gekoppelte Atome
e) Entstehung von Bändern aus den Termen durch Kopplung vieler Atome

Zustand) vorkommen dürfen. Der Grad der Aufspaltung hängt sowohl vom Abstand der Atomterme als auch von der Anzahl der wechselwirkenden Atome ab. Für vier Atome beispielsweise gilt qualitativ der Verlauf des Bildes 3.1.1d). Für mehr Atome kann die Aufspaltung kaum breiter werden. Es entstehen lediglich zusätzliche Zwischenterme. Bei sehr vielen Atomen liegen die Zwischenterme schließlich so dicht, daß scheinbar ein Kontinuum entsteht (Bild 3.1.1e)). Die Liniendichte wird noch durch einen weiteren Effekt zum Kontinuum übergeleitet. Die thermische Energie des Kristalls läßt die Atome sich statistisch um ihre Ruhelage bewegen. Hierdurch wird jede einzelne *Interferenzlinie* (Energieterm) *unscharf*. Insgesamt bilden sich so aus den diskreten Energietermen erlaubte Energiebänder aus. Die dazwischenliegenden Lücken heißen *verbotene Bänder* (Bild 3.1.1c)). Diese anschaulich leicht verständliche Betrachtung heißt Bändermodell des Festkörpers. Das von unten gesehen letzte, ganz oder teilweise mit Elektronen besetzte Band heißt *Valenzband*. Es ist für die chemischen Bindungen verantwortlich. Das nächsthöhere Band heißt *Leitungsband*, da es wesentlich für die elektrische Leitfähigkeit verantwortlich ist.

Genau genommen sind die Energieterme und damit die Lagen der *Bänder auch ortsabhängig*. Es ist ein Unterschied, ob sich z.B. ein Elektron dichter oder weiter entfernt von einem Atomkern befindet. Die periodische Anordnung der Atome im Kristall läßt so auch ein *periodisches Bandbild* entstehen. Sein Verlauf ist in verschiedenen Richtungen unterschiedlich. So entsteht Bild 3.1.2 für die drei wichtigsten Halbleiter.

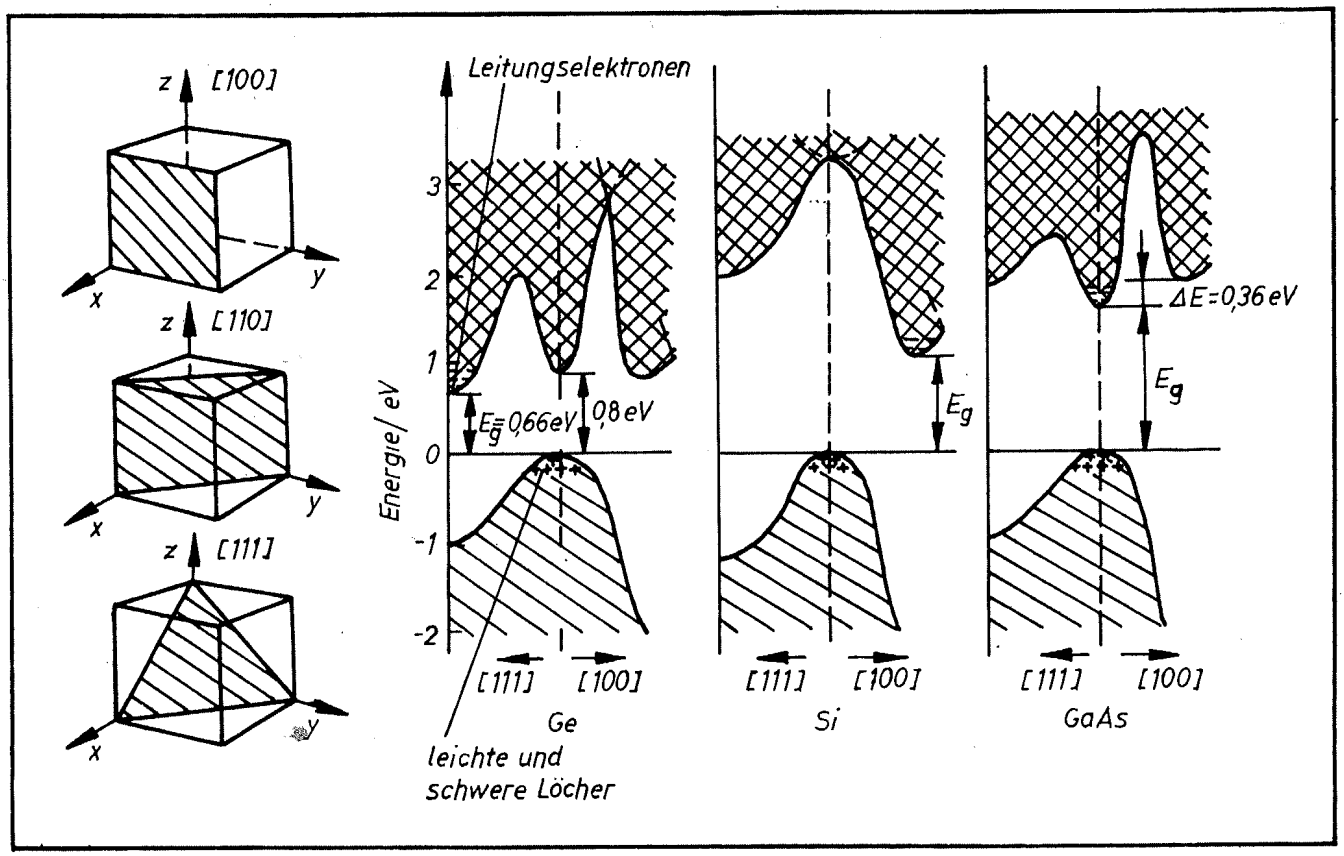


Bild 3.1.2. Schematische Darstellung der ortsabhängigen Bänder für Ge, Si und GaAs. Die Bezeichnungen [1 1 1] und [1 0 0] sind die MILLERschen Indizes und beziehen sich auf eine entsprechende Ebene in der Kristallzelle

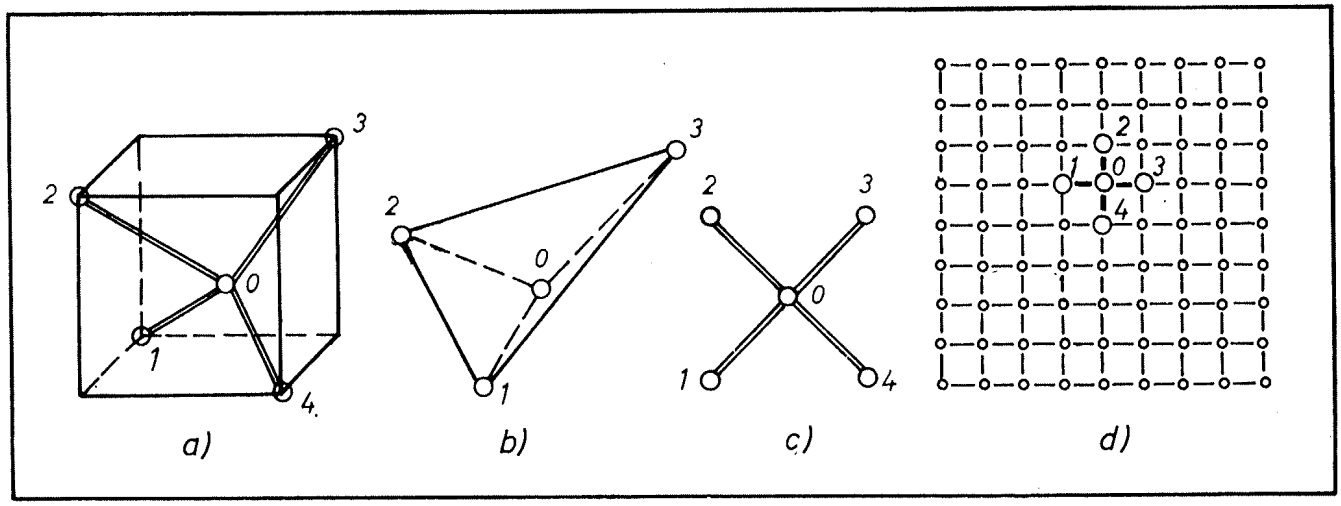


Bild 3.1.3. Zum Kristallaufbau der Halbleiter als Diamantgitter

a) Anordnung der Atome in der Elementarzelle
b) Tetraeder, Ausschnitt aus der Elementarzelle
c) Ebene Projektion der Elementarzelle
d) Ebene Zusammenfassung eines Kristallgitters

3.1.2. Leitungsmechanismen

Der Kristallaufbau vieler Halbleitersubstanzen entspricht dem *Diamanttyp*. Jedes Atom hat dabei vier Kovalenzbindungen zu vier Nachbarn. Bild 3.1.3a) zeigt diese Verhältnisse im Einheitswürfel. Je eines der Valenzelektronen des Zentralatoms wechselwirkt mit je einem Elektron der vier Nachbaratome. Von den vier Atomen

an den Würfelecken gehen also jeweils Bindungen zum Zentralatom. Ein derartiger Kristallaufbau ist auch noch anders zu veranschaulichen. Jeweils 4 benachbarte Atome spannen ein regelmäßiges Tetraeder auf. Im Bild 3.1.3b) sind zur Veranschaulichung die Atome 0, 1, 2, 3 des Teilbildes a) herausgezogen. Der Kristall kann deshalb auch aus *dicht gepackten Tetraedern* bestehend gedacht werden.

Zur Veranschaulichung des Kristallaufbaus ist es üblich, den Einheitswürfel gemäß Bild 3.1.3c) auf eine Ebene projiziert darzustellen. Viele Atome eines Kristalls ergeben dann ein regelmäßiges quadratisches Raster, wie es Bild 3.1.3d) zeigt. Hierbei ist immer daran zu denken, daß dieses Raster einen dreidimensionalen Kristall lediglich vereinfacht darstellt.

In dem ebenen Raster von Bild 3.1.3d) läßt sich die Eigenleitung des Halbleiters wie folgt veranschaulichen: Die *thermische Energie* kann dazu führen, daß eine *Bindung aufgebrochen* wird. Das freigewordene *Elektron* kann sich dann im Kristall *frei bewegen*. Unter äußerem Feldeinfluß verhält es sich etwa wie ein Leitungselektron gemäß Abschnitt 1.1.1. Durch die Aufspaltung einer Bindung bleibt eine ungesättigte Ladung des zugehörigen Atomkerns zurück. Sie stellt ein *Loch* dar, das durch ein beliebiges Elektron aufgefüllt werden kann. Die thermische Aufspaltung, auch *Generation* genannt, bildet also immer ein Trägerpaar, das aus einem Elektron und dem zurückbleibenden Loch besteht. Dieses Loch verhält sich in einem Feld wie ein Elektron mit positiver Ladung. Dadurch, daß es von einem Nachbarelektron aufgefüllt werden kann und dort wiederum ein Loch verbleibt, läßt sich der Vorgang auch so interpretieren, als ob das Loch entgegengesetzt zu den Elektronen wandert. Deshalb ist es oft einfacher und nützlicher, ein *Loch als positiven Ladungsträger* aufzufassen.

Bei der Bindung eines freien Elektrons an ein Loch wird immer ein Ladungsträgerpaar vernichtet. Dieser Vorgang heißt *Rekombination*. Im thermodynamischen Gleichgewicht sind Generation und Rekombination gleich groß. Die eben geschilderten Zusammenhänge sind in der Mitte von Bild 3.1.4 dargestellt.

Ein exakter Kristallaufbau nach Bild 3.1.3d) setzt extrem reines Material und völlig gleichmäßig angeordnete Atome (Einkristall) voraus (vgl. Kap. 4.). In dieses reine Material können bewußt *Fremdatome* eingebaut werden. Der entsprechende Vorgang wird *Dotieren* genannt. Die Wirkung dieser Elemente im Kristallgitter zeigt Bild 3.1.4.

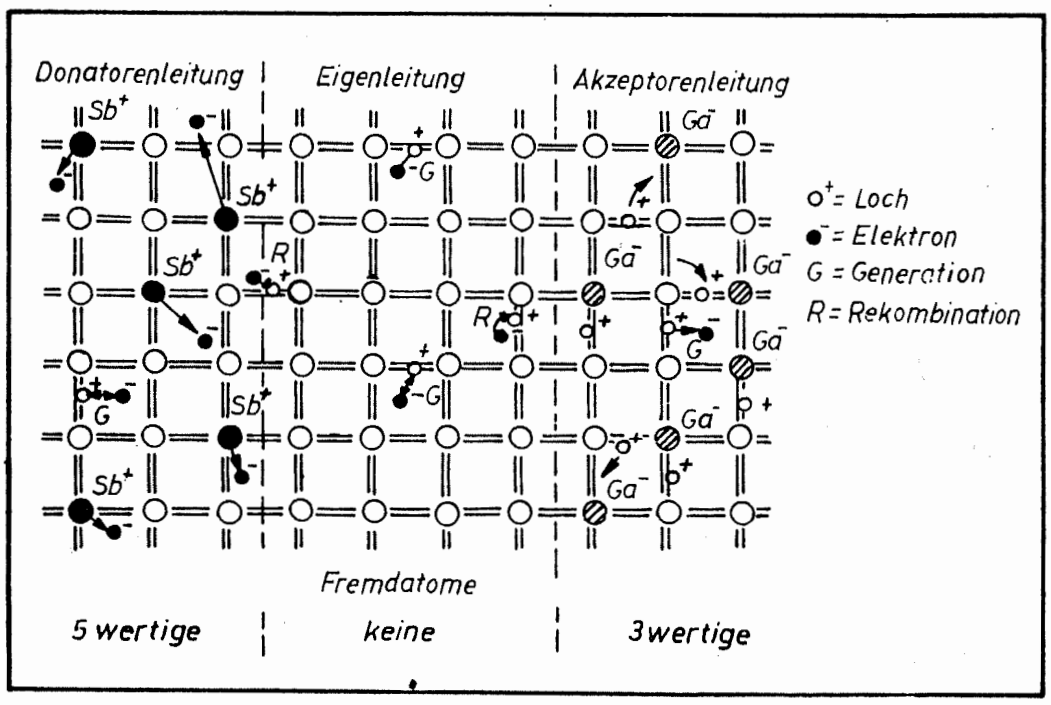


Bild 3.1.4. Zur Erklärung von Eigenleitung und Störstellenleitung

Fünfwertige Elemente erzeugen ein freies, nicht gebundenes Elektron. Entsprechend der Anzahl der *fünfwertigen Atome* im Kristall wird also die Anzahl der beweglichen Elektronen erhöht. Da so die negative Leitung im Kristall überwiegt, wird derartiges Halbleitermaterial *n-leitend* genannt. Die fünfwertigen Elemente heißen *Donatoren*. Sie bewirken eine Art der Störstellenleitung. Als in der Überzahl vorhandene *Majoritätsträger* existieren die *n*-leitenden Elektronen. Die *p*-leitenden Löcher bilden die Minoritätsträger.

Beim Einbau von *dreiwertigen Atomen* bleibt jeweils eine Bindung des Grundgitters ungesättigt. Es entstehen also zusätzliche *Löcher*. Das so dotierte Material ist also überwiegend *p-leitend*. Die Löcher bilden die Majoritäts- und die Elektronen die Minoritäts-Ladungsträger. Dreiwertige Materialien heißen *Akzeptoren*.

3.1.3. Leiter, Halbleiter, Isolatoren

Die *drei* typischen *Leitermechanismen* ergeben sich unmittelbar aus dem Bandabstand und deren Besetzung (Bild 3.1.5)

- Ein *Leiter* liegt dann vor, wenn das Leitungsband mit hinreichend vielen Elektronen besetzt ist. Dies kann auch so interpretiert werden, daß sich Leitungs- und Valenzband überlappen.
- *Halbleiter* liegen vor, wenn der Bandabstand um 1 eV liegt. Dieser Fall wird noch unten weiter erklärt.
- *Isolatoren* liegen vor, wenn der Bandabstand größer als 5 eV ist.

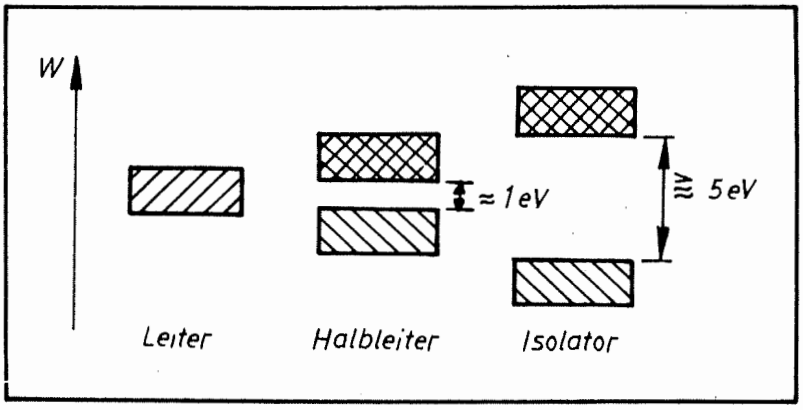


Bild 3.1.5. Relative Lage der Energiebänder und des Bandabstandes beim Leiter, Halbleiter und Isolator

Dieser Zusammenhang erklärt sich daraus, daß die Elektronen im vollbesetzten Valenzband im Mittel die Energie kT (mit der BOLTZMANN-Konstanten $k = 1{,}38 \times 10^{-23}$ J/K und der absoluten Temperatur T) annehmen können. Bei *Zimmertemperatur* besitzt sie den Wert 26 meV. Die *Wahrscheinlichkeit* dafür, daß mit dieser Energie Elektronen *ins Leitungsband kommen*, beträgt grob

$$p(\Delta W, T) \approx \mathrm{e}^{-\Delta W/kT} . \tag{1}$$

Für $\Delta W = 1$ eV folgt $p \approx 2 \cdot 10^{-17}$ und für $\Delta W = 5$ eV ca. $3 \cdot 10^{-84}$.

3.1.4. FERMI-Verteilung

Jedes einzelne Energieniveau in einem erlaubten Band darf auch höchstens von einem Elektron belegt werden. Andernfalls ist das Niveau leer. Dann existiert in dem

Verband kein Elektron mit der zugehörigen Energie. Die *Wahrscheinlichkeit* dafür, daß *zu einer Energie W ein Elektron gehört,* wird durch die FERMI-Verteilung angegeben:

$$f(W) = \frac{1}{e^{\frac{W - W_F}{kT}} + 1} \,. \tag{2}$$

Die Auswertung dieser Funktion für verschiedene Temperaturen zeigt Bild 3.1.6a). Für $T = 0$ sind alle niedrigen Niveaus mit $W \leqq W_F$ besetzt. Die Elektronen ordnen sich also so an, daß sie die geringstmögliche Energie annehmen. So entsteht die Rechteckkurve in Bild 3.1.6a). Die FERMI-Energie W_F folgt also aus der Termdichte und der Zahl der Elektronen. Bis zur Energie W_F existieren genau so viele Terme, wie es Elektronen im Verband gibt.

Am *absoluten Nullpunkt* sinken alle Elektronen auf das niedrigste Niveau. Es sind daher alle Terme bis W_F besetzt. Die zugehörige Energieverteilung zeigt die dünne *rechteckige Linie* im Bild 3.1.6a).

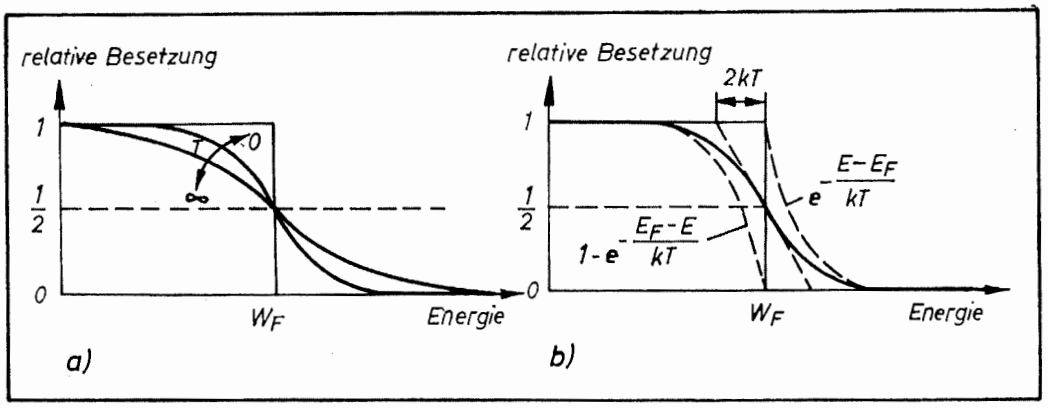


Bild 3.1.6. Zur FERMI-Verteilung

a) FERMI-Funktion nach Gl. (2) bei konstantem FERMI-Niveau W_F und der Temperatur T als Parameter

b) Näherung der FERMI-Verteilung durch die BOLTZMANN-Verteilung. Hieraus ist auch die Breite des Überganges $n_i = 0$ nach 1 ersichtlich. Sie beträgt etwa 4 kT (statt E_F und E lies W_F und W)

Für eine *unendlich hohe Temperatur* nimmt die Exponentialfunktion im Nenner von Gl. (2) den Wert Eins an. Folglich ist die Besetzungswahrscheinlichkeit für alle Energien *genau* 1/2. Es liegt also eine Gleichverteilung bezüglich der Besetzungsdichte vor. Dies wird durch die gestrichelte waagerechte Linie in Bild 3.1.6a) angedeutet.

Zwischen den beiden Extremwerten der Temperatur Null und Unendlich existieren Übergangsfunktionen. Sie sind ebenfalls in Bild 3.1.6a) dargestellt und besitzen alle für $W = W_F$ die Besetzungswahrscheinlichkeit 1/2. Das FERMI-Niveau ist also für jede Temperatur deutlich ausgezeichnet.

In Bild 3.1.6b) ist die FERMI-Verteilung für eine bestimmte endliche Temperatur dargestellt. Der Übergang von nahezu vollständig besetzten zu nahezu leeren Niveaus erfolgt danach innerhalb eines Energieintervalls von etwa 4 kT. Dabei wurde vorausgesetzt, daß die thermische Energie kT klein gegenüber der FERMI-Energie ist.

Werden nun Energien W betrachtet, die groß gegenüber der FERMI-Energie W_F sind, so kann im Nenner von Gl. (2) die Eins vernachlässigt werden, und es folgt die BOLTZMANN-*Verteilung* als *Näherung*:

$$f(W) = e^{-\frac{W - W_F}{kT}} \,. \tag{2a}$$

Sie kann ebenfalls als Näherung für Energien angewendet werden, die sehr klein gegenüber der FERMI-Energie sind. Ihr Wert gibt dann an, wieviel der Niveaus relativ leer, d. h. unbesetzt sind. Solange die Anzahl der Leerstellen im Vergleich zu den vorhandenen Termen sehr klein ist, können die unbesetzten Energieterme auch als Löcher, d. h. als positiv geladene Elektronen interpretiert werden. Besetzt nämlich ein anderes Elektron diesen freien Energieterm, so wird dadurch das vorausgehende Energieniveau leer.

3.1.5. Zustandsdichte

Im Valenz- und Leitungsband existieren, auf das Volumen und die Energie bezogen, eine bestimmte *Anzahl von Zuständen*, welche die Elektronen bzw. Löcher einnehmen können. Sie wird durch die *Zustandsdichte* gemäß

$$N(W)\,\mathrm{d}W = \frac{4\pi}{h^3}\,(2m^*)^{3/2}\,\sqrt{W}\,\mathrm{d}W \tag{3}$$

mit der PLANCKschen Konstanten $h = 6{,}626 \cdot 10^{-34}$ J und der effektiven Masse m^* der Ladungsträger bestimmt, je nach Material und Ladungsträgerart liegt sie bei 0,02 bis 1,8 Elektronenmassen.

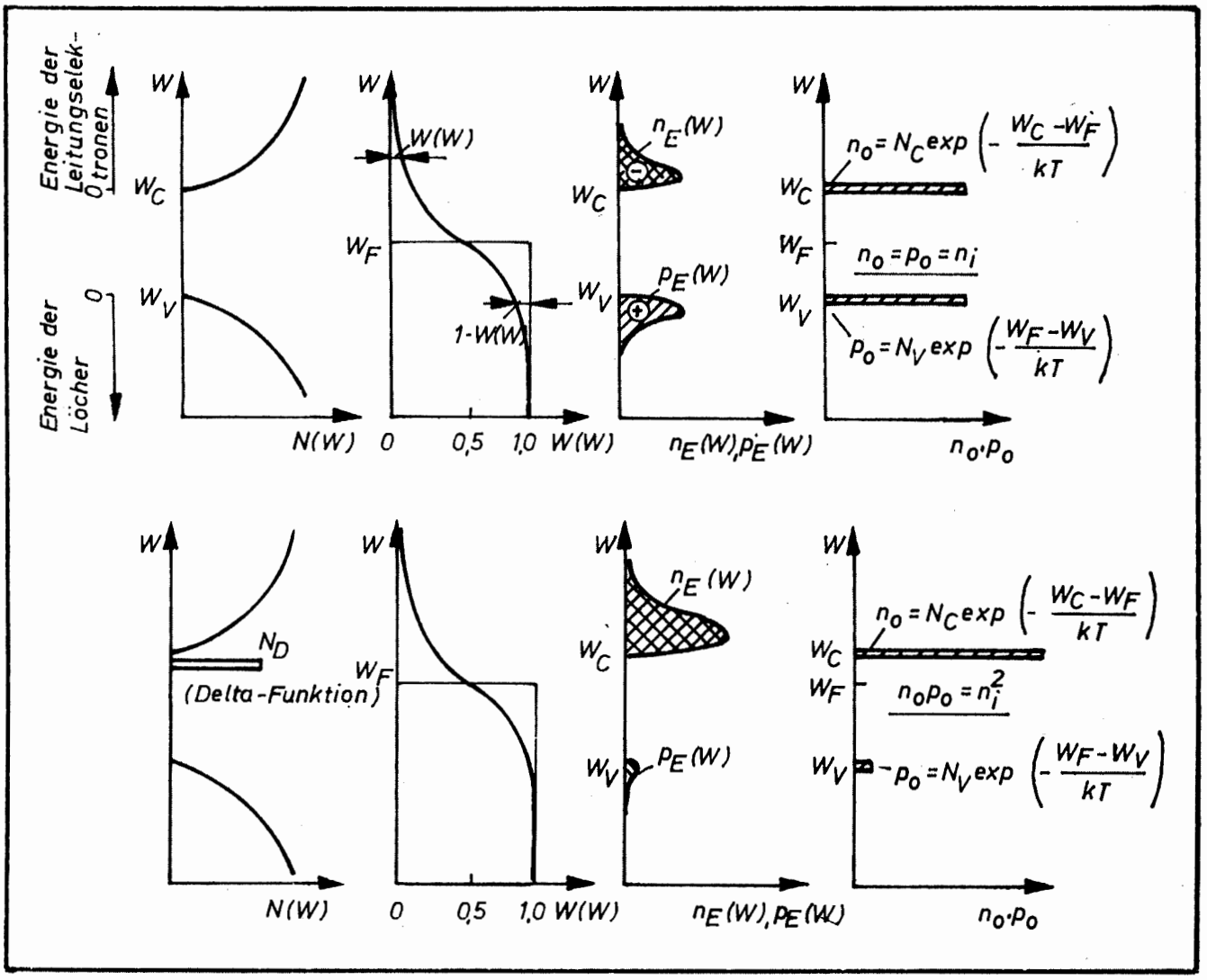


Bild 3.1.7. Verteilung der Elektronen und Löcher bezüglich der Energie beim eigenleitenden Halbleiter (oben) und beim n-Halbleiter (unten)

Dieser Verlauf ist in Bild 3.1.7a) dargestellt. Die Wahrscheinlichkeit für die Besetzung der Zustände ergibt sich aus der Faltung mit der FERMI-Verteilung. Deshalb ist sie um 90° gedreht passend zu Bild 3.1.7a) in Teilbild b) gezeichnet. Das Ergebnis enthält dann gemäß

$$n = \int_{W_c}^{+\infty} N(W) f(W)\, \mathrm{d}W \tag{4}$$

Teilbild c). Für die Rechnung ist einmal genau genommen die Obergrenze des Integrals gleich der Obergrenze des Valenzbandes. Doch von da an ist in der Regel der FERMI-Term infolge der Exponentialfunktion schon sehr klein. Bei nicht zu großer Dotierung gibt die BOLTZMANN-Verteilung eine ausreichende Näherung:

$$n = N_c \exp\left(\frac{W_F - W_c}{kT}\right) \tag{5}$$

mit der äquivalenten Zustandsdichte

$$N_c = 2\left(\frac{2\pi m^* kT}{h^2}\right)^{3/2} = c_1 T^{3/2}\,. \tag{6}$$

Die Konstante c_1 liegt bei ca. $10^{21}\ \mathrm{K^{2/3}\ m^{-3}}$.

Das Ergebnis von Gl. (5) ist in Bild 3.1.7d) dargestellt. Im Bild 3.1.7 sind auch die entsprechenden Verläufe für die Löcher enthalten.

Mit einer *Dotierung* entstehen *im verbotenen Band* in der Nähe des Valenz- bzw. Leitungsbandes *zusätzliche Energieterme*. Dadurch verändert sich die Situation entsprechend dem unteren Teil von Bild 3.1.7. Es läßt sich zeigen, daß generell gilt

$$n_i = \sqrt{np} = \sqrt{N_c N_v}\, \exp\left(-\frac{\Delta W}{2kT}\right), \tag{7}$$

wobei ΔW der Bandabstand, N_c, N_v die äquivalenten Zustandsdichten analog Gl. (6), n die im Volumen vorhandenen Elektronen und p die Löcherzahl sind. n_i ist die nur vom Material und der Temperatur (und nicht von der Dotierung) abhängige Ladungsträgerdichte, auch Inversionsdichte, Eigenkonzentration oder Intrinsiczahl genannt. Werte für sie enthält Tab. 3.1.1, den Temperaturverlauf zeigt Bild 3.1.8a). Die

Tabelle 3.1.1. Inversionsdichten einiger Halbleiter bei 300 K in $\mathrm{m^{-3}}$

Ge	Si	GaAs	GaP	InP	InAs	GaSb	InSb	Diamant
$2{,}3 \cdot 10^{19}$	$1{,}6 \cdot 10^{16}$	$9{,}2 \cdot 10^{12}$	$2{,}7 \cdot 10^{6}$	$6{,}9 \cdot 10^{12}$	$8{,}3 \cdot 10^{20}$	$9{,}6 \cdot 10^{17}$	$1{,}5 \cdot 10^{22}$	$6.7 \cdot 10^{-22}$

Änderung der Trägerzahl ist von der Temperatur und dem Bandabstand abhängig. Näherungsweise gilt

$$\frac{\mathrm{d}n_i^2}{n_i^2} \approx \left(3 + \frac{\Delta W}{kT}\right)\frac{\mathrm{d}T}{T}. \tag{8}$$

Für Germanium und Zimmertemperatur gilt $\Delta W \approx 27\, kT$. Deshalb ändert sich die Trägerzahl und damit auch die Leitfähigkeit um annähernd 10% pro Grad.

Der Zusammenhang

$$n_i^2 = np \tag{9}$$

aus Gl. (7) heißt auch *Massenwirkungsgesetz des Halbleiters*.

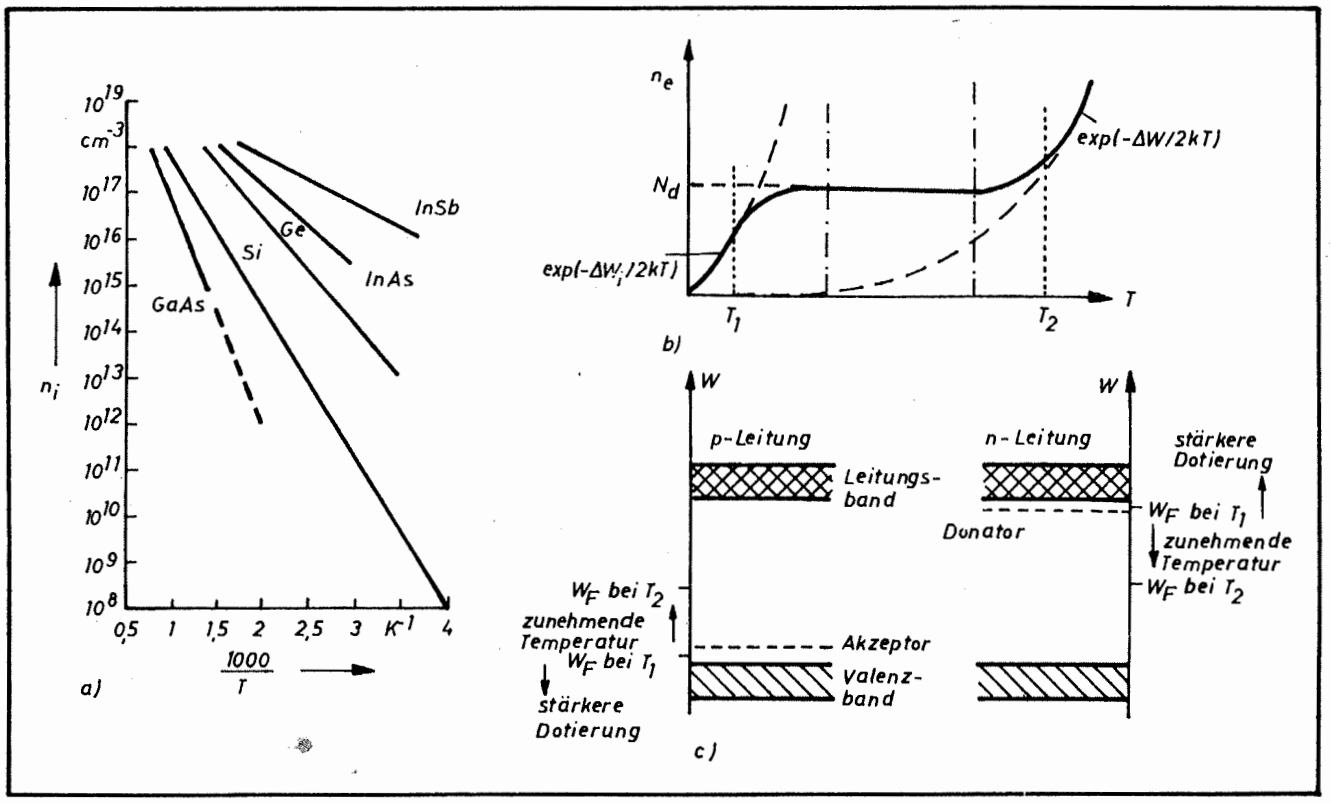


Bild 3.1.8. a) und b) Abhängigkeit der Inversionsdichte von der Temperatur bei einigen Halbleitermaterialien. Schematische Darstellung der Temperaturabhängigkeit der Trägerdichte. Halbleiter weisen in den Bereichen (a) und (b) Störstellenleitung auf, im Gebiet (c) dagegen Eigenleitung. Die gestrichelte Kurve gibt den Verlauf der Trägerdichte bei Eigenleitung an
c) Verlagerung des FERMI-Niveaus bei Störstellenhalbleitern in Abhängigkeit von der Temperatur und von der Störstellenkonzentration

Für die *Temperaturabhängigkeit* der Ladungsträger der Donatoren bzw. Akzeptoren gilt ebenfalls Gl. (4). Für einen dotierten Halbleiter (Störstellenhalbleiter) zeigt den Verlauf der Trägerzahl in Abhängigkeit von der Temperatur Bild 3.1.8b). Dabei sind *drei Gebiete unterscheidbar*. Im Gebiet (a) wird mit Erhöhung der Temperatur die Störstellenleitung infolge der Donatoren bzw. Akzeptoren wirksam. Die Trägerzahl nimmt mit der Temperatur so lange zu, bis alle Donatoren bzw. Akzeptoren ihre Wirkung zeigen. Von da ab, d. h. im Bereich (b), bleibt die Trägerzahl zunächst konstant. Sie wächst erst wieder im Gebiet (c), wenn die Trägerbildung des reinen (störstellenfreien) Halbleitermaterials merklich wird (gestrichelte Kurve). Von da ab verhält sich ein Störstellenhalbleiter so, als ob er nicht dotiert wäre. Die *Eigenleitung* bestimmt sein Verhalten.

Aus dem Bild 3.1.8b) kann auch einiges zum FERMI-Niveau abgeleitet werden. Bei der Temperatur T_1 und n-leitendem Material muß die FERMI-Kante zwischen dem Donatoren-Niveau und der unteren Grenze des Leitungsbandes liegen. Nur so ist in beiden Bändern eine gleich große Ladungsträgerzahl erreichbar. Bei der Temperatur T_2 liegt Eigenleitung vor. Folglich muß die FERMI-Kante in der Mitte zwischen dem Leitungs- und Valenzband liegen. Bei n-leitendem Material wandert folglich die FERMI-Kante mit steigender Temperatur abwärts, bei p-leitendem Material dagegen aufwärts (vgl. hierzu Bild 3.1.8c)).

In Bild 3.1.8c) ist auch gezeigt, wie sich das FERMI-Niveau mit der Dotierung verlagert. Bei extrem starker Dotierung kann es bis ins Valenz- bzw. Leitungsband gelangen.

Ladungsträger einer Art, die durch Dotierung gewonnen werden, reduzieren die andere Art von Ladungsträgern. Auf dieser Basis ist der *„vergiftete" Halbleiter* verständlich. Er enthält Donatoren und Akzeptoren, und die wirksame Ladungsträgerdichte wird zweifach reduziert [M 20].

Die Niveaus der Dotierungselemente für die hauptsächlichen Halbleiter enthält Bild 3.1.9.

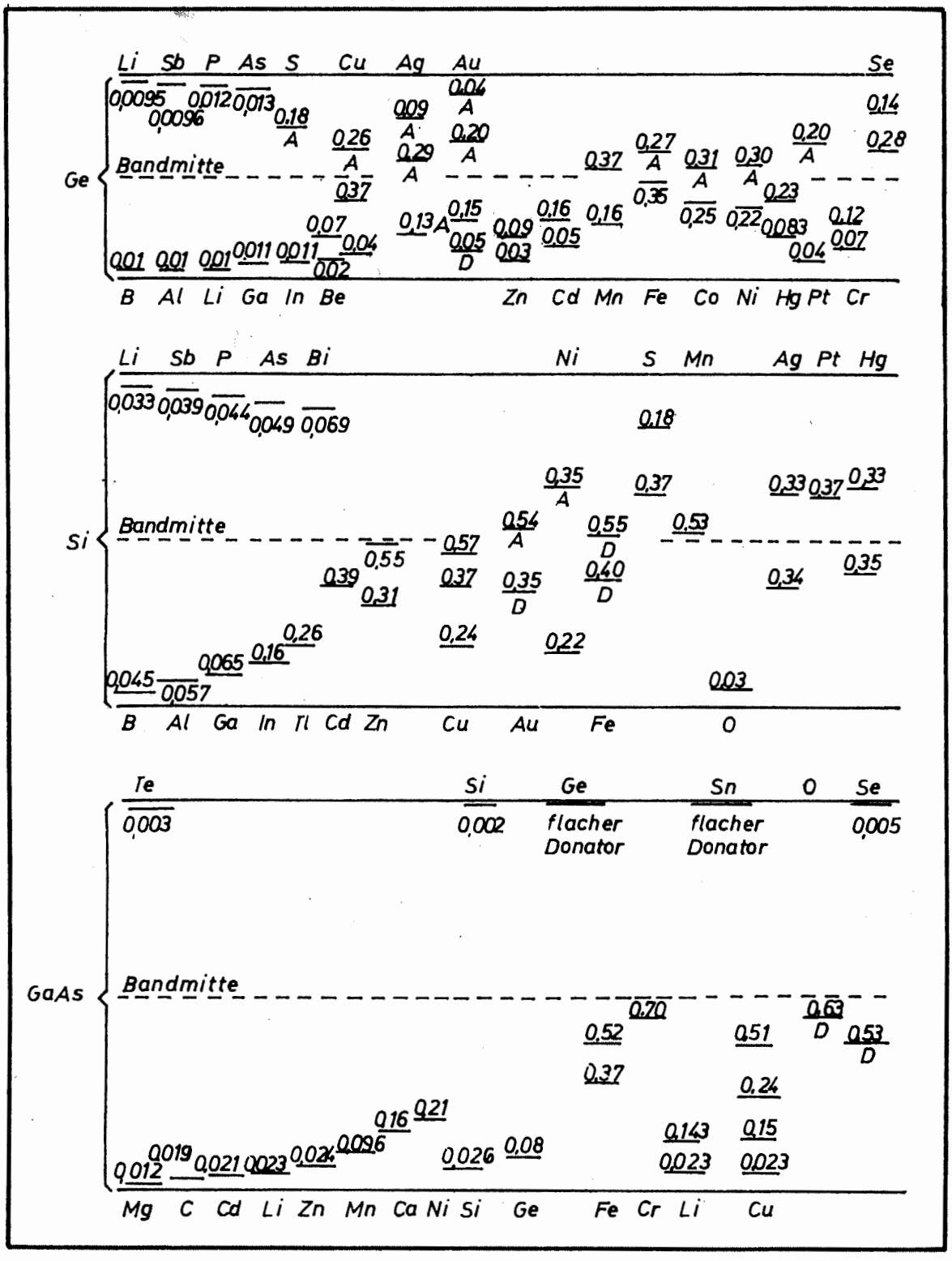


Bild 3.1.9. Ionisationsenergien einiger Dotierungselemente in Ge, Si und GaAs. Die Niveaus unter der Bandmitte sind von der oberen Bandkante des Valenzbandes gemessen und stellen Akzeptorniveaus dar, wenn nicht durch *D* (Donator) gekennzeichnet; die Niveaus über der Bandmitte sind von der unteren Bandkante des Leitungsbandes gemessen und stellen Donatorniveaus dar, wenn nicht durch *A* (Akzeptor) gekennzeichnet

3.1.6. Leitungsmechanismen

Ein mit der Geschwindigkeit v bewegtes Elektron besitzt gegenüber einem ruhenden mit der Energie W die Energie

$$W^+ = W + \tfrac{1}{2} mv^2 . \tag{10}$$

Aus diesem Grunde ist ein Elektron in einem voll besetzten Band nicht bewegbar, denn auch die Energie W^+ ist ja besetzt.

Entsprechend Gl. (1.1.5) besteht bei Leitern Proportionalität zwischen der Driftgeschwindigkeit der Elektronen und der Feldstärke $\boldsymbol{E}$. Für die Halbleiter wird entsprechend mit der *Beweglichkeit* μ geschrieben

$$v_D = \mu \boldsymbol{E} . \tag{11}$$

Die Beweglichkeit ist unterschiedlich für *Löcher* μ^+ und für *Elektronen* μ^-. Beide Größen sind untereinander und für die einzelnen Materialien verschieden (vgl. Tab. 3.1.2 und 3.1.3). Sie hängen aber auch von der Feldstärke ab. Dies zeigt Bild 3.1.10a).

Tabelle 3.1.2. Stellung der Halbleiter im Periodensystem der chemischen Elemente
Links oben am Elementzeichen steht die Ordnungszahl, rechts davon der Bandabstand in eV. Für Kohlenstoff gilt die Diamantkonfiguration. Unterhalb der Elemente steht links die Elektronen-, rechts die Löcherbeweglichkeit in cm^2/Vs.

III		IV		V		VI		VII	
^{5}B	1,28	^{6}C	5,2	^{7}N	—	^{8}O	—	^{9}F	—
7;	>7	1800;	200	—		—		—	
^{13}Al	—	^{14}Si	1,12	^{15}P	1,5	^{16}S	2,5	^{17}Cl	—
20;	—	1350;	480	—		—		—	
^{31}Ga	2,25	^{32}Ge	0,7	^{33}As	1,2	^{34}Se	2,1	^{35}Br	—
450;	20	3900;	1900	—		6,05;	0,13	—	
^{49}In	—	^{50}Sn	0,1	^{51}Sb	0,12	^{52}Te	0,3	^{53}J	1,25
—		3000;	—	—		1750;	1250	—	

Die *Leitfähigkeit* σ ist reziprok zum spezifischen Widerstand ϱ. Sie ist proportional der Ladungsträgerzahl und zu deren Beweglichkeiten:

$$\frac{1}{\varrho_i} = \sigma_i \sim N_c n_i (\mu^+ + \mu^-) . \tag{12}$$

Da sich die Beweglichkeit bei den meisten Halbleitern mit $T^{-3/2}$ ändert, folgt die Leitfähigkeit etwa der Ladungsträgerzahl (Gl. (5)). Auf die Abhängigkeit des spezifischen Widerstandes kann daher mit Hilfe von Bild 3.1.8b) und (12) geschlossen werden. Die entsprechenden Verhältnisse zeigt Bild 3.1.10b). Im Gebiet (a) und (c) überwiegt jeweils die Zunahme der Ladungsträgerzahl mit der Temperatur. Der Widerstand nimmt also mit steigender Temperatur ab. Im Gebiet (b) ist dagegen die Trägerzahl etwa konstant, und es macht sich vor allem die Änderung der Beweglichkeit der Ladungsträger bemerkbar. In diesem Gebiet verhält sich also ein Halbleiter bezüglich der Temperatur fast genau so wie ein Leiter. Sein Widerstand steigt geringfügig mit der Temperatur an. Oberhalb dieser Temperatur (im Gebiet (c)) ist der eigenleitende Anteil überwiegend.

Tabelle 3.1.3. Bandabstände in eV und Beweglichkeiten in cm^2/Vs einiger Verbindungshalbleiter der Gruppe III—V (2. Zeile: Beweglichkeit der Elektronen, darunter der Löcher), rechts oben der Gruppe II—IV—V, in der Reihenfolge von oben nach unten Si, Ge, Sn für IV; ganz unten weitere Halbleiter mit Bandabständen, darüber zusätzliche Beweglichkeiten
Es muß bemerkt werden, daß z. T. die Daten in verschiedenen Büchern abweichen.

III	V			
	N	P	As	Sb
B	6,4	2,1	1,46	—
Al	3,8	3,0 80	2,4 180	1,8 200 400
Ga	3,84 150	2,34 2100 1000	1,38 8500 400	0,81 4000 650
In	2	1,42 4600 650	0,43 30000 240	0,2 70000 1000

V	II		
	Zn	Cd	Mg
P	2,1 2,2 1,4	2,1 1,7 1,2	2,3 >2
As	2,1 0,9 0,66	1,6 0,5 0,23	

Au	Ag	Cu	Bi	SiC	PBS	Cu_2O	BiTe
32	66	43	5000	60	600	80	750
—	—	—	—	8	500	100	400
PbSe	HgTe	CdS	CdSe	ZnO		ZnS	PbTe
1200	30000	300	800	200		165	6000
900		50	40	—		5	4000

II—VI		IV—VI		II—VI	
ZnS	3,7	PbS	0,41	ZnO	3,2
ZnSe	2,6	PbSe	0,29	Cu_2O	1,06
CdS	2	PbTe	0,32		
ZnTe	2,2	HgSe	0,1		
CdSe	1,74	SiC	≈ 3		
CdTe	1,51	SnS	1,26		
		BiTe	0,16		

Aus dem *spezifischen Widerstand* eines Halbleitermaterials bei Zimmertemperatur kann auf die *Störstellenzahl* geschlossen werden. Bild 3.1.10c), d) zeigt die Zusammenhänge für *n*- bzw. *p*-leitendes Silicium und Germanium.

Schließlich sei erwähnt, daß *Aufbaufehler im Kristall*, wie z.B. Versetzungen, Zwillingsbildung, genau wie Störstellen wirken. Deshalb muß sowohl für reine Eigenleitung als auch für einwandfreie dotierte Materialien stets vom fehlerfreien Kristall ausgegangen werden. (Vgl. hierzu Kap. 4.)

Bezüglich der *Reinheitsstufen* läßt sich etwa folgende Einteilung angeben:

chemisch reine Stoffe	99,9%
analysenreine Stoffe (p. a.)	99,99%
spektroskopisch reine Stoffe	99,999%
lumineszenzreine Stoffe	99,9999%
halbleiterreine Stoffe	99,99999%.

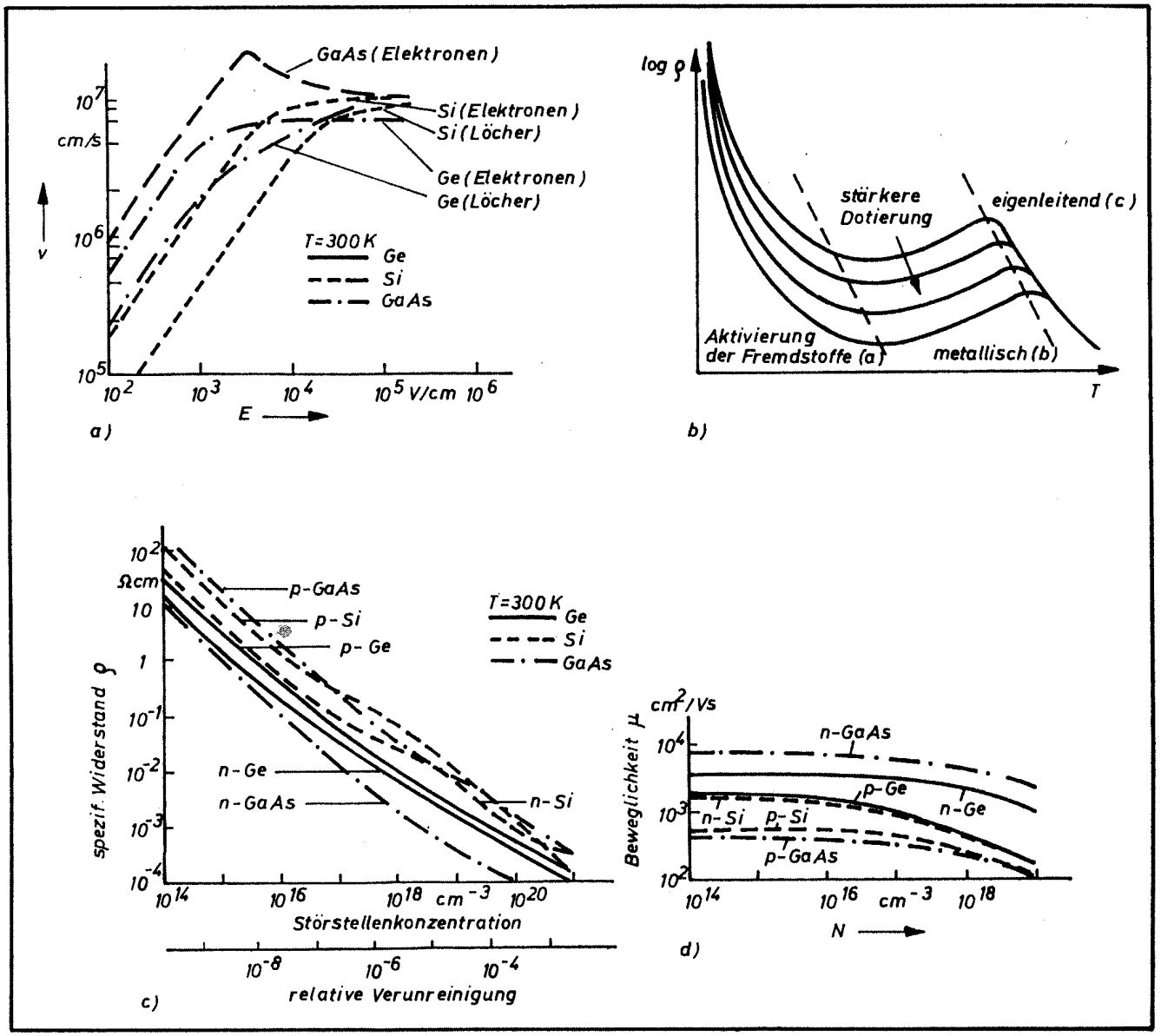


Bild 3.1.10. Beweglichkeit und spezifischer Widerstand von Halbleitermaterialien

a) Beweglichkeit als Funktion der Feldstärke
b) Typischer Verlauf des spezifischen Widerstandes als Funktion der Temperatur
c) Zusammenhang von spezifischem Widerstand und Störstellenkonzentration
d) Zusammenhang von Beweglichkeit und Störstellenkonzentration

3.2. Der pn-Übergang

Bisher wurden nur homogene Halbleiter besprochen. Wenngleich sich mit diesen Materialien auch spezielle Bauelemente herstellen lassen, so stellen sie jedoch nur eine Voraussetzung für die Vielzahl der Halbleiterbauelemente dar. Diese Vielfalt beruht grundlegend auf der Kombination von p- und n-Material. Stoßen zwei derartige Materialien zusammen, so entsteht ein pn-Übergang. Aus seinen Eigenschaften lassen sich die Eigenschaften fast aller Halbleiterbauelemente ableiten.

3.2.1. Stromloser pn-Übergang

Die folgenden Betrachtungen werden stark vereinfacht durchgeführt. In Bild 3.2.1 sind jeweils nur die Donatoren und Akzeptoren und die durch sie erzeugten Elektronen

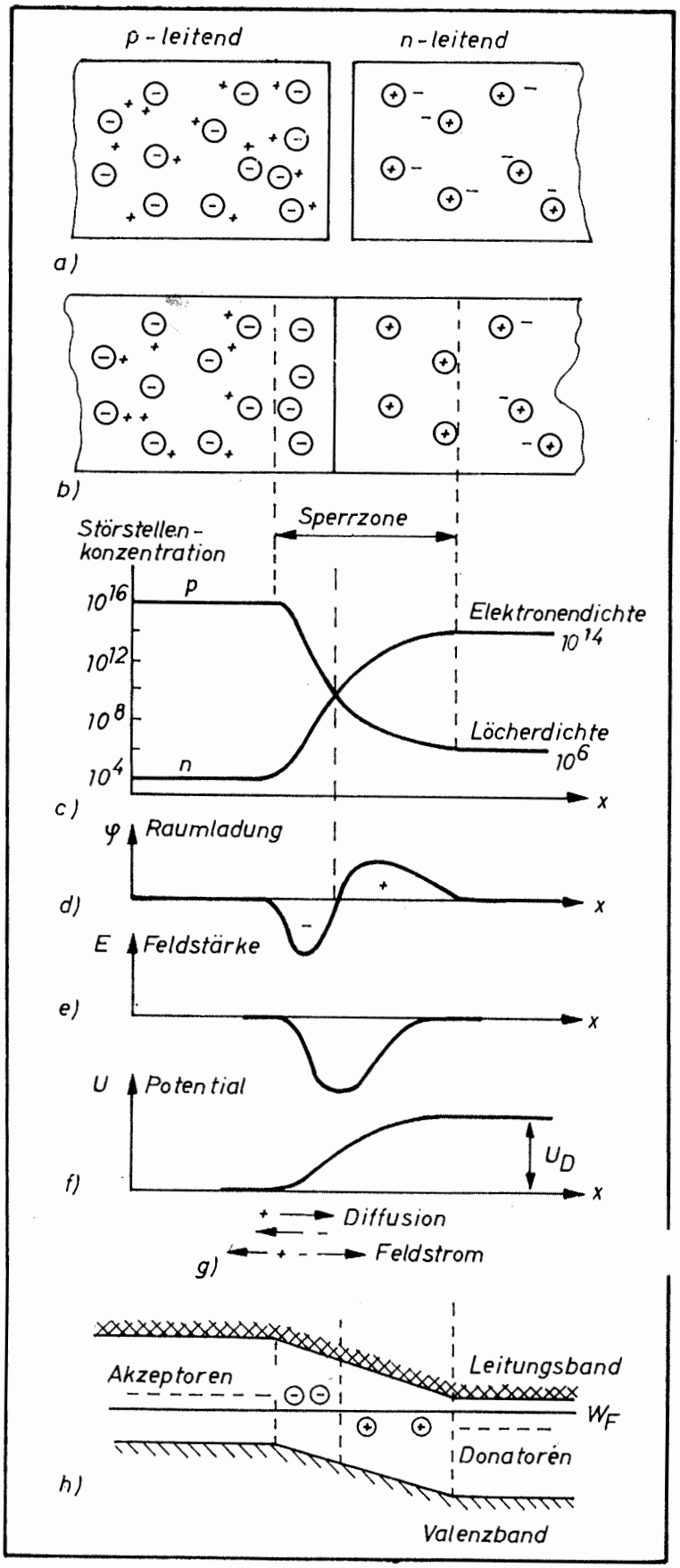


Bild 3.2.1. Zum Aufbau und zur Wirkung eines *pn*-Überganges. Dargestellt sind nur die eingebauten Donatoren ⊕ und Akzeptoren ⊖ sowie die dadurch bedingten Elektronen — und Löcher -|-

a) Gegenüberstehender, b) verbundener *pn*-Übergang. In der Umgebung des *pn*-Überganges diffundieren die Ladungsträger so, daß sie weitgehend rekombinieren. Dadurch ergibt sich in der Sperrzone eine Ladungsträgerverteilung, wie sie c) zeigt. Die eindiffundierten Löcher bilden eine positive, die Elektronen eine negative Raumladung (d). Als Folge entsteht eine ortsabhängige Feldstärke im Kristall (e), die integriert zur Diffusionsspannung U_D führt (f). Die Richtung des Diffusionsstromes und des Feldstromes deutet g) an. Die Interpretation im Bändermodell zeigt h)

und Löcher dargestellt. Die anderen Atome sind fortgelassen. Während die Donatoren und Akzeptoren fest in Kristallgittern eingefügt sind, können die Elektronen und Löcher durch Energie, z.B. Wärme oder elektrisches Feld, nahezu frei bewegt werden.

Bild 3.2.1a) zeigt die Verhältnisse, wenn ein *p*- und ein *n*-Material sich ohne Berührung gegenüberstehen. Dabei wird für das *n*-Material eine Störstellenkonzentration von 10^{11} und für das *p*-Material von 10^{14} je cm³ gewählt (siehe Grenzen in Teilbild c)). Die Dichte n_i beträgt 10^{10}.

In Bild b) ist nun angenommen, daß diese *beiden Materialien fehlerstellenfrei aneinandergefügt* sind. Dies ist technologisch durch Sondermaßnahmen über einen Si-Einkristall zu sichern (s. Kap. 4.). An der Grenze zwischen *p*- und *n*-Material liegt damit ein *abrupter pn-Übergang* vor. Hier gehen die Störstellenkonzentrationen sprunghaft von einem *p*- in den anderen *n*-Störstellenwert über. Die Folge davon ist, daß sich die beweglichen Ladungsträger (also die freien Elektronen und Löcher) unter diesem Einfluß zu kompensieren versuchen. Die *Elektronen diffundieren nach links*, die *Löcher nach rechts*. So entsteht der *Diffusionsstrom* (Bild 3.2.1g)). In ihm sind die Diffusionskoeffizienten D_n und D_p sowie die Konzentrationsgradienten $\mathrm{d}p/\mathrm{d}x$ und $\mathrm{d}n/\mathrm{d}x$ enthalten:

$$I_D = e\left(D_n \frac{\mathrm{d}n}{\mathrm{d}x} - D_p \frac{\mathrm{d}p}{\mathrm{d}x}\right). \tag{1}$$

Dadurch bleiben die Akzeptoren- und Donatorenrümpfe als in den Kristall eingefrorene *Ladungen zurück*, und so entsteht eine *Raumladungsverteilung*, wie sie Bild 3.2.1d) zeigt. Sie läßt sich in eine Feldstärke E und in ein Potential U umrechnen. Das so entstehende Feld wirkt seinerseits wieder auf freie Ladungsträger zurück und erzeugt so einen *Feldstrom*

$$I_F = e(\mu^- n + \mu^+ p)\, E\,. \tag{2}$$

Er ist also dem Diffusionsstrom entgegengesetzt (Bild 3.2.1g)). Insgesamt bildet sich ein dynamisches Gleichgewicht zwischen Feld- und Diffusionsstrom. Dabei entstehen in etwa die Verläufe, wie sie Bild 3.2.1 zeigt. Die Berechnungen dieser Zusammenhänge sind nicht ganz einfach (z.B. [M 24, P 23, S 51]). Lediglich für die *Diffusionsspannung* entsteht eine übersichtliche Formel:

$$\boxed{U_D = \frac{kT}{e_0} \ln \frac{{}^{p}n_i}{{}^{n}n_i}}\,. \tag{3}$$

Die hochgestellten Indizes stellen den Bezug zur Ladungsträgerdichte im *p*- bzw. *n*-Gebiet her. Vielfach liegt das Diffusionspotential bei etwa 0,2 V. Im thermischen Gleichgewicht betragen sowohl der Diffusion- als auch der Feldstrom etwa 10000 A je cm².

Es sei weiter erwähnt, daß die Diffusionsspannung unmittelbar nicht meßbar ist. Das Halbleitermaterial muß dazu nämlich am *p*- und *n*-Ende kontaktiert werden. Dabei entstehen infolge der unterschiedlichen Leitungseigenschaften Kontaktpotentiale, die genau entgegengesetzt so groß wie das Diffusionspotential sind. Weiter sei erwähnt, daß in Metallen nur der Feldstrom existiert.

Die soeben geschilderten Verhältnisse lassen sich auch mittels des Bändermodells erklären. Bei üblicher Dotierung und Zimmertemperatur gilt jeweils für *n*- und *p*-Material getrennt die FERMI-Grenze in Bild 3.1.8c). Werden beide Materialien zu einem Kristall zusammengefügt, so muß als Bezugslinie die dann gemeinsame FERMI-Kante gewählt werden. Im *p*-Material liegen also Leitungs- und Valenzband vergleichsweise höher. Beim Kontakt beider Materialien rekombinieren Elektronen aus dem Leitungsband des *n*-Materials mit den Löchern der Akzeptoren, so daß die Akzeptoren als negativ geladene Ionen am Rand der *pn*-Schicht erscheinen. Sie stellen die negative

Raumladung links vom *pn*-Übergang in Bild 3.2.1 dar. Analog rekombinieren die Löcher des Valenzbandes mit den Elektronen der Donatoren, und die positiv geladenen Ionen der Donatoren bilden die rechte positive Raumladung in Bild 3.2.1. Die beiden Raumladungen führen zur Verformung der Bandkanten gemäß Bild 3.2.1 h). Sie führt die Leitungs- und Valenzbänder in beiden Materialien zusammen. Die bisher besprochenen Effekte entsprechen wieder den Diffusionsströmen. Durch die Verbiegung der Bänder können unter Energiegewinn Elektronen im Leitungsband von links nach rechts und Löcher im Valenzband von rechts nach links verschoben werden. Dies entspricht dem Feldstrom. Das Gleichgewicht zwischen beiden Strömen führt zu einer Verbiegung der Bänder, die genau der Diffusionsspannung entspricht und immer kleiner als der Bandabstand sein muß.

Die Leitfähigkeit eines Halbleitermaterials ist unmittelbar von der vorhandenen Trägerzahl abhängig. Durch Rekombination von Ladungsträgern entsteht in der Umgebung des *pn*-Überganges ein stark *an Trägern verarmtes Gebiet*. Es besitzt infolgedessen einen sehr großen Widerstand und sperrt den Stromfluß. Deshalb heißt dieses Gebiet auch *Sperrzone*. Es ist meist sehr schmal (Größenordnung von 0,1 μm). Daher rührt sein anderer Name *Grenzgebiet*. Wegen der Raumladung gemäß Bild 3.2.1 d) wird es am genauesten *Raumladungszone* genannt.

3.2.2. Stromdurchflossener *pn*-Übergang

Wird eine *äußere Spannung* an den *pn*-Übergang gelegt, so *wirkt* sie fast ausschließlich innerhalb der *Raumladungszone*. Dort besteht nämlich der größte Widerstand innerhalb des Kristalls. Als Folge wird die Diffusionsspannung um die äußere Spannung je nach deren Polung vergrößert bzw. verkleinert. In Bild 3.2.1 ändert sich infolgedessen die Feldstärke bzw. die Raumladung des *pn*-Überganges. Je nach der Polung ist die Verbiegung der Bänder mehr oder weniger stark ausgeprägt. Dabei bleibt der feldunabhängige *Diffusionsstrom* praktisch *unverändert*. Seine Summe aus Löchern und Elektronen werde als *Sättigungssperrstrom* I_{sp} bezeichnet. Die Feldströme dagegen werden von einem mehr oder weniger höheren Potential beschleunigt. Unter etwas vereinfachten Annahmen kann für den *Feldstrom* von dem Feldstrom im stromlosen Zustand und der BOLTZMANN-Verteilung ausgegangen werden:

$$I^+ = I_F^+ \, e^{e_0 U/kT} \,. \tag{4}$$

Dieser Strom muß wegen des dynamischen Gleichgewichts bei $U = 0$ entgegengesetzt gleich dem Sättigungssperrstrom I_{sp} sein. Die Differenz zu ihm ist der außen feststellbare Strom. Für ihn gilt folglich:

$$\boxed{I = I_{sp} \, (e^{e_0 U/kT} - 1)} \,. \tag{5}$$

Diese Beziehung ist die *grundlegende Ausgangsgleichung der Halbleitertechnik*. Ihr prinzipieller Verlauf ist in Bild 3.2.2a) dargestellt. Je nach der Polung der angelegten Spannung werden Durchlaß- und Sperrichtung unterschieden. Die Größe kT/e_0 beträgt bei Zimmertemperatur rund 26 mV. Der Sättigungssperrstrom I_{sp} ist abhängig von verschiedenen Eigenschaften des *pn*-Überganges. Bei normalen kleinen Dioden und Zimmertemperatur beträgt er für Ge etwa 10 μA, und für Si liegt er bei etwa 10^{-3} μA.

Gl. (5) gilt exakt nur unter Beachtung der folgenden fünf *Voraussetzungen*:

— An der Grenze des *pn*-Überganges treten keine Störungen des Kristallgitters auf.
— Die thermisch erzeugten Trägerpaare rekombinieren nicht in der Raumladungszone. Ihr Diffusionsweg ist also groß gegenüber der Länge des *pn*-Überganges.
— Die Feldstärken in der Raumladungszone bleiben so gering, daß sie zu keiner Trägererzeugung führen.

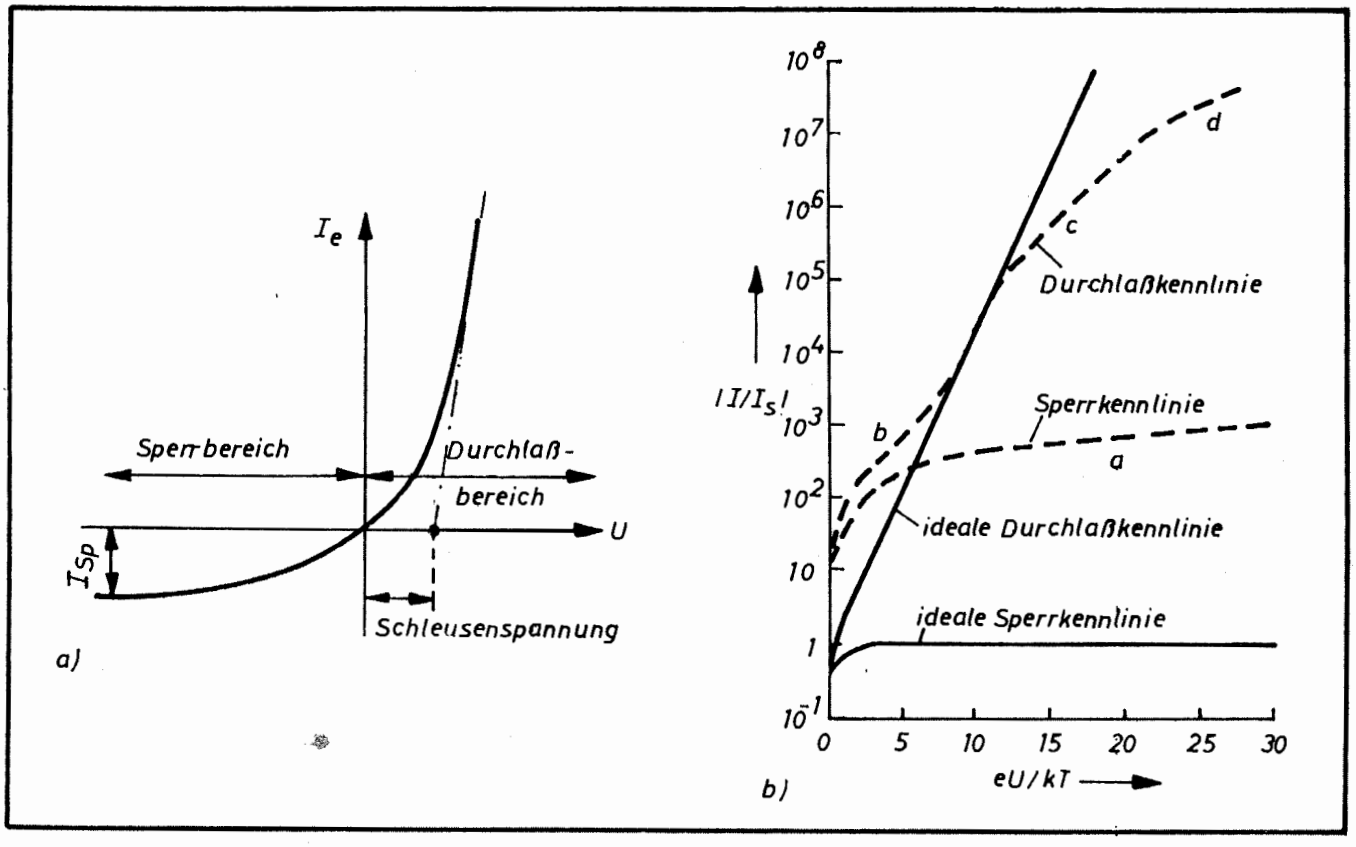


Bild 3.2.2. a) Strom-Spannungs-Kennlinie eines idealen *pn*-Überganges in linearer Darstellung gemäß Gl. (5)
b) Strom-Spannungs-Kennlinie in logarithmischer Darstellung mit realen Abweichungen infolge spezieller Einflüsse: (a) Trägererzeugung in der Raumladezone, (b) Rekombination in der Raumladezone, (c) Einfluß starker Injektion, (d) Einfluß von Bahnwiderständen

— Die Ströme im Kristall bleiben so klein, daß der ohmsche Widerstand des Kristalls außerhalb der Sperrzone, der sogenannte Bahnwiderstand, vernachlässigbar ist.
— Es gibt keine Oberflächeneffekte am *pn*-Übergang, d. h., senkrecht zum *pn*-Übergang wird der Kristall als unendlich ausgedehnt angenommen.

Wie sich diese Einflüsse bei einem realen *pn*-Übergang auswirken, geht aus Bild 3.2.2b) hervor.

3.2.3. Durchlaßbetrieb

Beim Betrieb in *Durchlaßrichtung* liegt am *p-Material positive* und am *n*-Material negative Spannung. Das äußere Feld treibt damit die Majoritätsträger in Richtung zum *pn*-Übergang. Anders ausgedrückt, werden Elektronen in die *p*-Zone und Löcher in die *n*-Zone injiziert. Majoritätsträger der beiden Gebiete gelangen also als zusätzliche Minoritätsträger in die anderen Gebiete. Dadurch wird die Leitfähigkeit des *pn*-Überganges verbessert, und der Strom kann stark ansteigen. Dieses Geschehen bewirkt, daß statt (9) jetzt $pn > n_i^2$ in der Umgebung des *pn*-Überganges gilt. Bild 3.2.3 zeigt die dabei entstehenden Ladungsträgerkonzentrationen im Vergleich zu Bild 3.2.1c). Insbesondere fällt der exponentielle Verlauf der zusätzlichen Minoritätsträger links und rechts von der ursprünglichen Sperrzone auf. Wird ein *pn*-Übergang mit relativ starkem Strom in Durchlaßrichtung betrieben, so wächst auch die Trägerinjektion sehr stark an. Dabei wird die Ladungsträgerverteilung im Valenz- und Leitungsband erheblich verändert. Sie kann sich dann auch durch strahlenden Über-

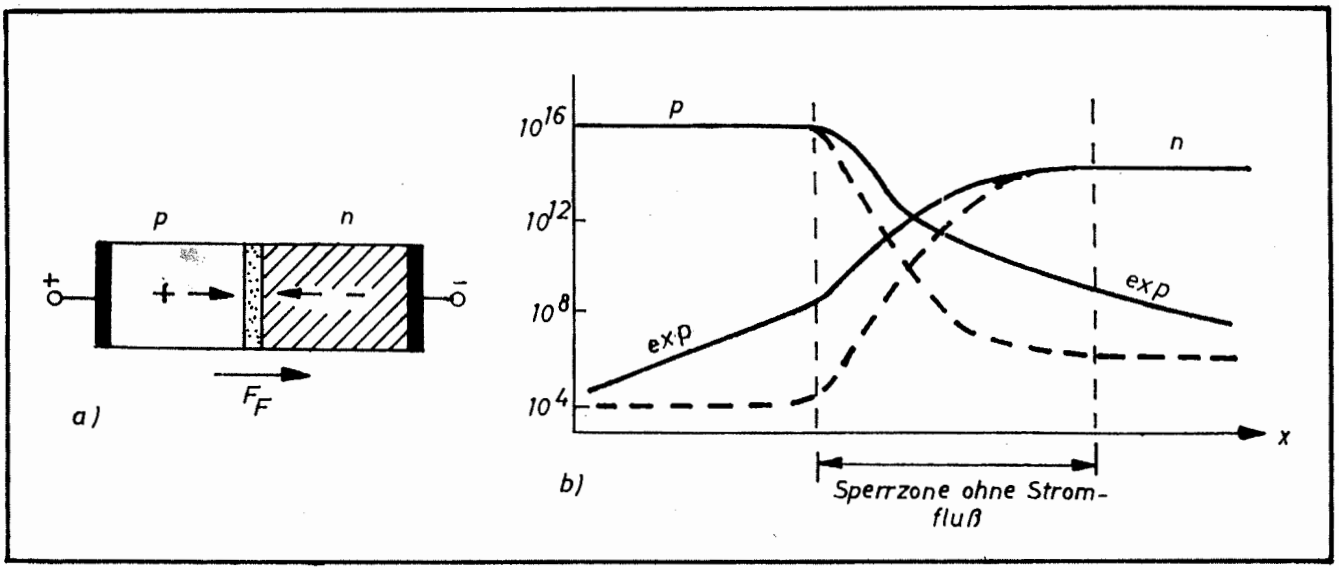


Bild 3.2.3. Betrieb eines *pn*-Überganges in Durchlaßrichtung

a) Polung und Bewegung der Träger

b) Änderung der Trägerkonzentration:
ausgezogen: bei Durchlaßbetrieb;
gestrichelt: stromlos im Vergleich zu Bild 3.2.9c). exp deutet auf den exponentiellen Abfall der zusätzlichen Minoritätsträger jenseits des ursprünglichen Sperrbereiches hin

gang wieder ausgleichen. Dies wird bei Luminizenz- und Laserdioden genutzt (Kap. 6.).

3.2.4. Sperrbetrieb

Beim Betrieb des *pn*-Überganges in *Sperrichtung* liegt am *p-Material* negative und am *n*-Material positive Spannung. Hierdurch werden beide Majoritätsträger aus dem *pn*-Übergang herausgezogen. Dadurch verbreitert sich die Sperrzone. Der Widerstand nimmt dabei in etwa mit der angelegten Spannung zu, so daß der Sperrstrom im wesentlichen von der angelegten Spannung unabhängig ist. Bild 3.2.4 zeigt die entsprechenden Zusammenhänge für die geänderte Trägerkonzentration und für die Größe der Raumladungen bei unterschiedlicher symmetrischer und unsymmetrischer Dotierung.

Während bei Polung des *pn*-Überganges in Durchlaßrichtung die Trägerkonzentration nahezu beliebig anwachsen kann, ist es im Sperrbetrieb nur möglich, sie bis nahezu Null abzusenken. Unter anderem deshalb tritt eine *Änderung der Breite d* für die *Sperrzone* auf:

$$d = \sqrt{\frac{2\varepsilon}{e}\,(U_D - U)\,\frac{1}{\frac{1}{{}^{p}n_i} + \frac{1}{{}^{n}n_i}}}\,. \tag{6}$$

Die Breite der Sperrzone ist also etwa der Wurzel der angelegten Sperrspannung U proportional. Sie wird formal Null, wenn die Sperrrpannung in Flußrichtung umgepolt wird und dabei gleich der Sperrspannung wird. Dieser Punkt steht in einem gewissen Zusammenhang mit der schon in Bild 3.2.2 angedeuteten und noch im Abschnitt 5.4. genauer zu behandelnden Schleusenspannung.

Der Sperrstrom ist in einem relativ weiten Bereich der Ladungsträgerdichte n_i^2 proportional. Diese hängt, wie schon im Abschnitt 3.1.5. behandelt, von der Tem-

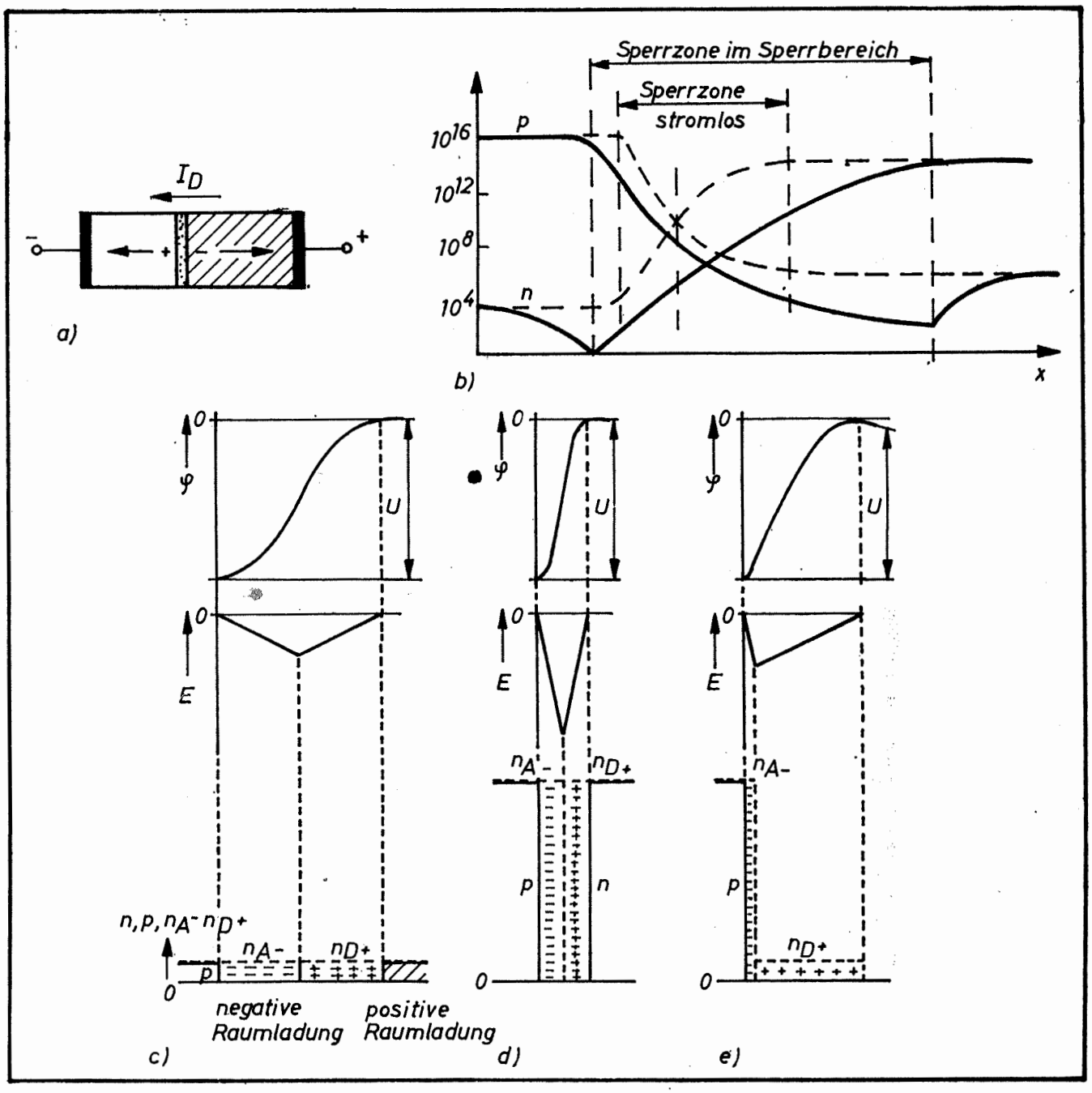


Bild 3.2.4. Betrieb eines *pn*-Überganges in Sperrichtung

a) Bewegung der Majoritätsträger
b) Änderung der Ladungsträgerkonzentrationen mit der Wirkung: Verbreiterung der Sperrzone (stromloser Fall gemäß Bild 3.2.1c), gestrichelt)
c) bis e) Feldstärke- und Raumladungsverteilung bei unterschiedlicher Dotierung: c) symmetrisch schwach, d) symmetrisch stark, e) unsymmetrisch dotiert

peratur ab. Aus diesen Zusammenhängen ist die Temperaturabhängigkeit des Sperrstromes ableitbar. Für Zimmertemperatur gilt etwa:

$$\left.\begin{aligned} \text{Ge:}\ \frac{\mathrm{d}I_s}{\mathrm{d}T} &= I_s\,(0{,}01 + 0{,}1\ T/K)\,, \\ \text{Si:}\ \frac{\mathrm{d}I_s}{\mathrm{d}T} &= I_s\,(0{,}01 + 0{,}16\ T/\mathrm{K})\,. \end{aligned}\right\} \tag{7}$$

Der Sättigungsstrom verdoppelt sich danach bei Ge mit 10 K und Si bereits mit 6 K.

3.2.5. Raumladekapazität

Für kleine Wechselspannungen, die der Sperrspannung überlagert sind, kann die *Sperrzone als Dielektrikum eines Kondensators* aufgefaßt werden (Raumladungs- oder Sperrschichtkapazität genannt). Dann ist mit der Sperrgleichspannung die Breite der Sperrzone gemäß (6) und damit entsprechend auch die *Kapazität steuerbar*. Dieser Effekt wird bei *Kapazitätsdioden* genutzt. Um bestimmte Kennlinien und z.T. auch

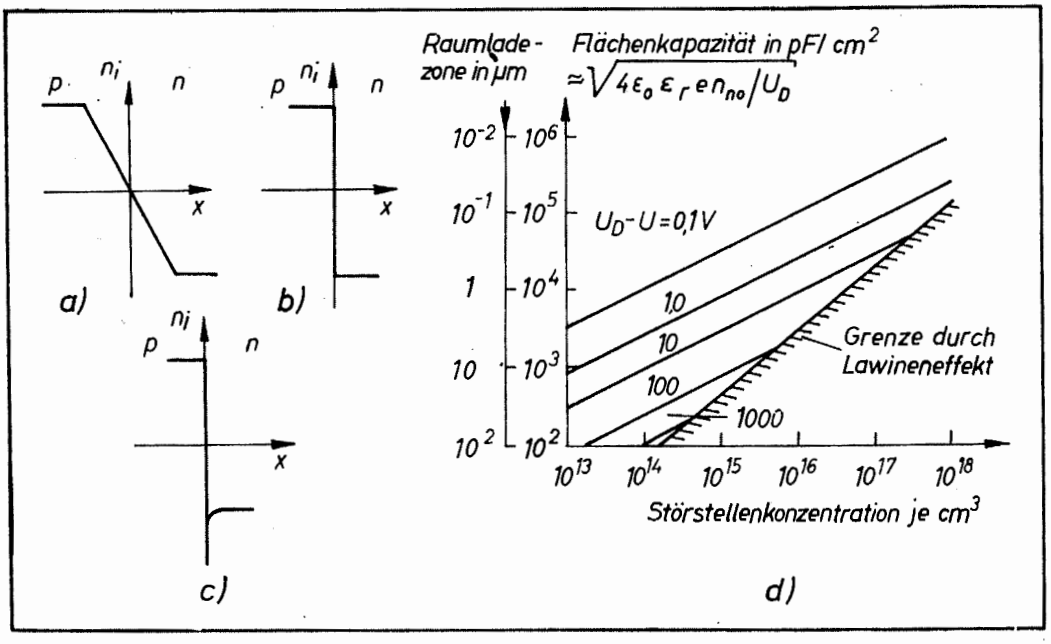


Bild 3.2.5. Dotierung von *pn*-Übergängen und ihre Wirkung auf den Exponenten q bei Kapazitätsdioden in Gl. (8)

a) Linear, $q \approx 0{,}33$ b) Abrupt, $q \approx 0{,}5$ c) Hyperabrupt, $q = f(U) > 0{,}5$

d) Sperrschichtkapazität pro Flächeneinheit C_s/A und Weite der Raumladungszone l als Funktion der Dotierung für „einseitig abrupte" *pn*-Übergänge in Si; eine Seite ist sehr stark dotiert (z.B. p^+), die andere Seite (z.B. *n*-Zone) sei mit der Konzentration $|N|$dotiert; Parameter ist die Spannung $U_D - U$ am Übergang; die gestrichelte Gerade gilt für $U = 0$ ([4], S. 89)

große Kapazitätsbereiche zu erreichen, wird der *pn*-Übergang unterschiedlich (also nicht nur abrupt) dotiert (Bild 3.2.5). Dann gilt näherungsweise die Formel

$$C = \frac{C_0}{\left(1 + \dfrac{U}{U_D}\right)^q}, \tag{8}$$

wobei C_0 die Kapazität ohne Gleichspannung und U_D wieder die Diffusionsspannung bedeutet. Der Exponent q hängt von dem Dotierungsverlauf und z. T. auch wieder von der Sperrspannung U ab. Für einseitig abrupten Übergang zeigt Bild 3.2.5d) die erreichbaren, auf die Fläche A bezogenen Kapazitätswerte.

3.2.6. Zener- und Avalanche-Effekt

Werden *pn*-Übergänge sehr stark in Sperrichtung vorgespannt, so tritt irgendwann ein *Durchbruch* auf. Hierzu sind *Feldstärken* um $2 \cdot 10^5$ V cm^{-1} notwendig. Dabei nimmt der Sperrstrom plötzlich und sehr steil zu. Es ergeben sich Kennlinien, wie sie Bild 3.2.6a) zeigt. Vor allem die Kennlinie bei Si hat wegen ihres scharfen Knicks

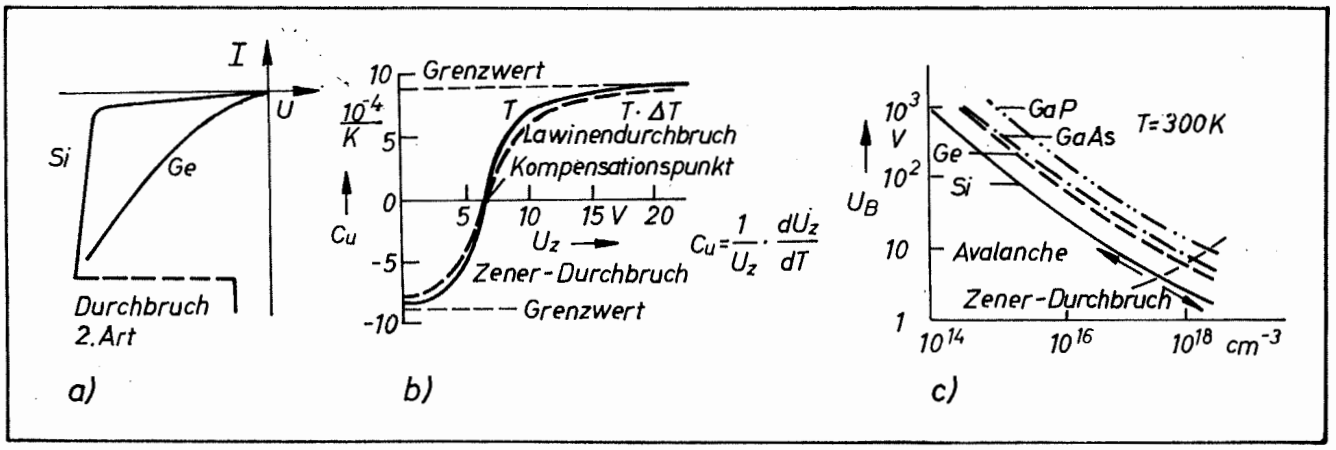


Bild 3.2.6. Zum ZENER- und Avalanche-Effekt bei *pn*-Übergängen

a) Durchbruchverhalten bei Si und Ge

b) Temperaturabhängigkeit der Durchbruchspannung (ZENER-Spannung) als Funktion der Spannung bei Si-Dioden

c) Durchbruchspannung bei verschiedenen Materialien als Funktion der Ladungsträgerdifferenz des abrupten *pn*-Überganges

große praktische Bedeutung. Dieser Durchbruch geht auf zwei *unterschiedliche Effekte* zurück, die im folgenden beschrieben werden.

Der sogenannte ZENER-*Durchbruch* erfolgt für Si bei Spannungen unterhalb etwa 4,5 V. Infolge der hohen Feldstärke werden dabei im *pn*-Übergang einzelne *kovalente Bindungen aufgerissen*. Die so frei werdenden Ladungsträger erhöhen den Sperrstrom. Je höher die Temperatur des Kristalls ist, um so weniger Spannung wird hierfür benötigt. Mit steigender Temperatur beginnt der ZENER-Durchbruch also früher. Der *Temperaturkoeffizient* der ZENER-Spannung ist daher *negativ* und zugleich stark stromabhängig. Er liegt zwischen -10^{-3} und -10^{-4} je Grad Temperaturänderung.

Oberhalb von 6 V und vor allem in dickeren Sperrschichten können Ladungsträger eine so hohe Energie annehmen, daß sie durch Stoß weitere Ladungsträger freisetzen. Dieses kann sich eventuell auch für die neuen Ladungsträger wiederholen. Insgesamt steigt der Strom so lawinenartig an. Da diesem Effekt (Avalanche-, Lawinen-, Stoßionisation) die thermische Bewegung (Stöße) entgegenwirkt, ist sein Temperaturkoeffizient positiv und wenig stromabhängig.

Infolge des lawinenartigen und damit unregelmäßigen Stromanstiegs rauscht eine Diode mit *Avalanche-Durchbruch* im Gegensatz zu einer mit ZENER-Durchbruch relativ stark. Da beide Effekte nach außen gleichartige Strom-Spannungs-Charakteristika zeigen, werden Diodenarten nach beiden Prinzipien unter dem Namen Z-Dioden zusammengefaßt. Im Übergangsbereich vom ZENER- zum Avalanche-Durchbruch gibt es ein Gebiet um etwa 5,7 V, wo der Temperaturkoeffizient der Z-Spannung praktisch Null ist. Hier kompensieren sich gerade die Temperatureinflüsse beider Effekte (Bild 3.2.6b)).

In Bild 3.2.6 ist bei der Si-Kennlinie noch ein *zweiter Durchbruch* eingezeichnet. Er tritt bei sehr hohen Strömen auf. Dann bilden sich einzelne Strompfade extrem hoher Stromdichte aus. Sie erzeugen eine kraterförmige Schmelzzone und führen zu Kurzschlüssen des *pn*-Überganges. Genauer wird hierauf noch bei den Leistungstransistoren eingegangen (5.7.5.).

Schließlich sei noch der *thermische Durchbruch* erwähnt. Die zugeführte Energie erwärmt den *pn*-Übergang. Dadurch steigt der Strom, und die Erwärmung nimmt weiter zu. Diese Rückkopplung führt schließlich zur thermischen Zerstörung des *pn*-Überganges (5.7.6.).

Bild 3.2.6c) zeigt die *erreichbare Sperrspannung* bei verschiedenen Materialien als Funktion der Differenz der Trägerkonzentrationen. Eine anschaulichere Inter-

pretation dieser Zusammenhänge kann aus Bild 3.2.4 gewonnen werden. Bei niedriger Konzentration in c) im Vergleich zu d) ist die Sperrzone breiter und damit die Feldstärke im Kristall kleiner.

3.2.7. Trägheit von *pn*-Übergängen

Im statischen Zustand gelangen Majoritätsträger jedes der beiden Gebiete als Minoritätsträger durch Diffusion in das andere Gebiet. Für eine Änderung dieses Zustandes, z. B. durch Änderung der äußeren Spannung, wird eine bestimmte Zeit benötigt. Die richtige Verteilung der Minoritätsträger nach einem Spannungswechsel erfolgt also verzögert. Dieser Effekt wirkt sich, von außen gesehen, wie eine zusätzliche Kapazität, die Diffusionskapazität, aus. Dieser Effekt wird besonders groß, wenn ein *pn*-Übergang plötzlich von Durchlaßrichtung in Sperrichtung umgeschaltet wird. Derartige dynamische Effekte seien jetzt in beiden Richtungen betrachtet.

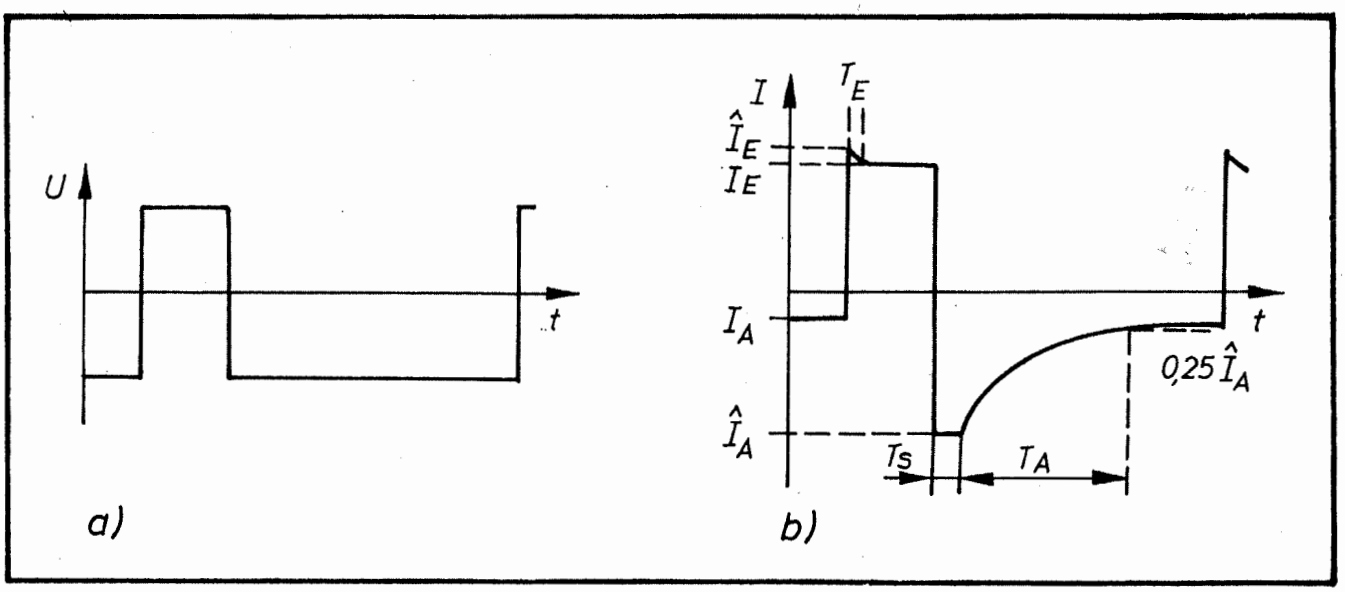


Bild 3.2.7. Stromverlauf (b) in einer Diode bei sprunghaftem Umschalten der Spannung (a) Es bedeuten: T_E Durchlaßerholzeit; T_S Speicherzeit; T_A Sperrerholzeit

Zunächst sei angenommen, daß ein *pn*-Übergang in Sperrichtung mit dem Strom I_A betrieben werde. Dann werde der Strom plötzlich in Durchlaßrichtung umgeschaltet und soll dabei als zeitlichen Endwert die Größe I_E erreichen (Bild 3.2.7). Zuvor muß jedoch noch die bei Sperrichtung vorhandene Raumladung abgebaut werden. Sie bewirkt ein kurzzeitiges Überschwingen des Durchlaßstromes im Bereich von einigen ns. Der Durchlaßstrom erzeugt nun viele Ladungsträger. Sie müssen nach einem erneuten Umschalten in Sperrichtung abtransportiert und durch Rekombination vernichtet werden. Dies erfolgt in zwei Stufen. Während der Speicherzeit T_S erfolgt keine Änderung der Trägerzahl. Es wird also die volle Stromstärke, allerdings in der entgegengesetzten Richtung, erreicht. Danach klingt der Strom exponentiell ab. Daraus ergibt sich die Sperrerholzeit T_A. Sie ist meist relativ lang und kann bei Spezialdioden auf 10 bis 500 ns gesenkt werden. Wesentlich sind Minoritätsträger mit typischen Lebensdauerwerten von

Ge: $10^{-6} \dots 10^{-3}$ s,

Si: $10^{-10} \dots 10^{-3}$ s,

GaAs: $10^{-10} \dots 10^{-8}$ s.

Eine vollständige Unterdrückung der Minoritätsträger ist u. a. durch SCHOTTKY- *bzw. Heterodioden* möglich, dann sind Zeiten bis zu 1 ns möglich.

3.2.8. *psn*-Struktur

Wie aus Bild 3.2.4c) bis e) und 3.2.6c) zu erkennen ist, muß für eine *hohe Sperrspannung* mindestens eines der beiden Materialien *schwach dotiert* sein. Dies bedeutet aber, daß der pn-Übergang im Durchlaßbereich einen *relativ großen Bahnwiderstand* besitzt, denn die Wege zum pn-Übergang sind infolge des niedrig dotierten Materials relativ hochohmig. Einen Ausweg aus dieser Widersprüchlichkeit ermöglicht die 1952 von HALL vorgeschlagene psn-Struktur. Sie fügt in die Mitte zwischen die beiden hochdotierten p- und n-Materialien ein schlecht leitendes (s-) Material ein. Dieses Material kann dabei sowohl n- als p-leitend sein.

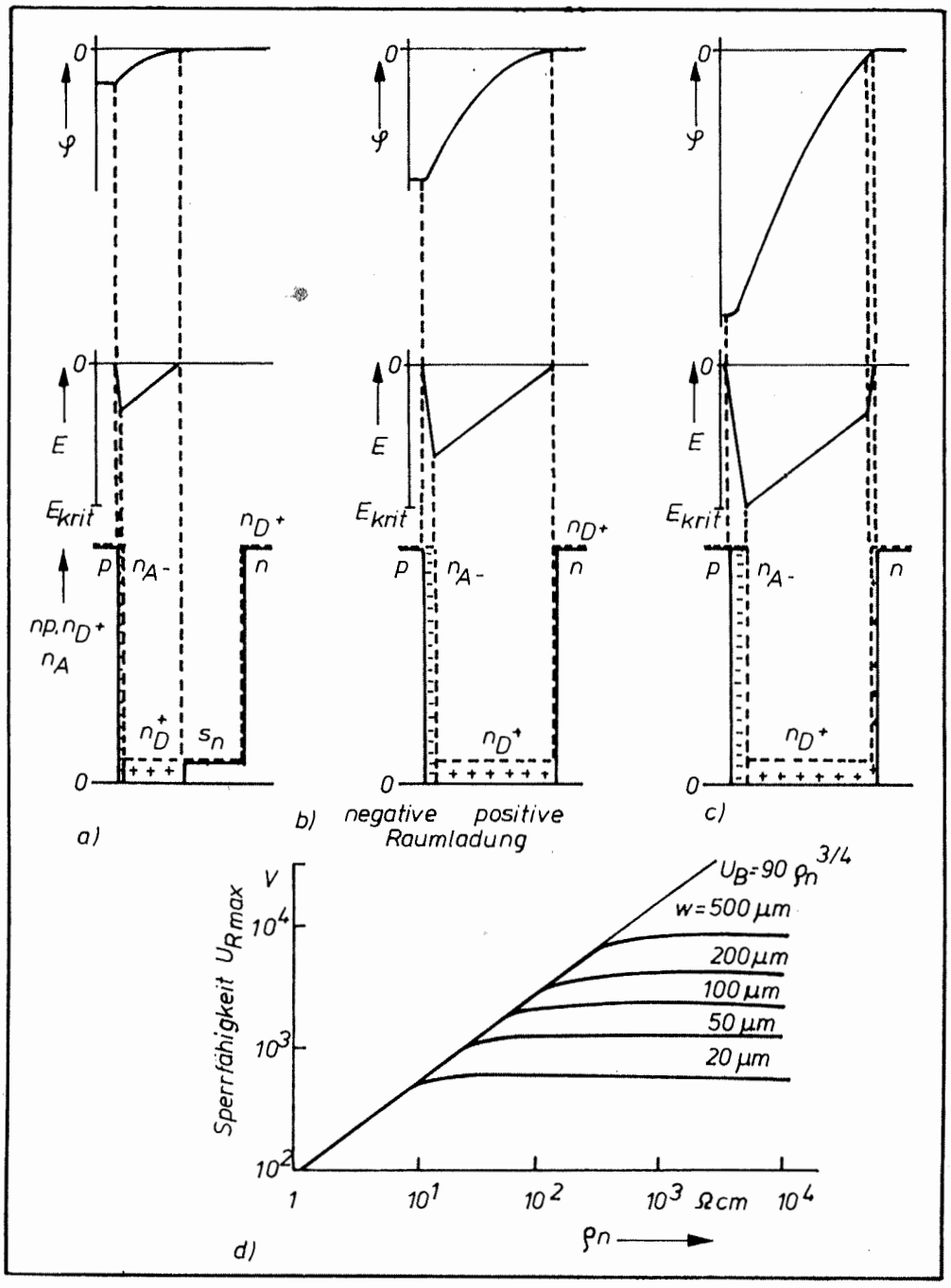


Bild 3.2.8. Eigenschaften einer ps_nn-Struktur

a) bei geringer Sperrspannung,
b) bei mittlerer Sperrspannung,
c) bei höchster Sperrspannung
d) Erreichbare Sperrspannung als Funktion der Leitfähigkeit des s_n-Gebietes mit der Weite w des Gebietes zwischen dem hochdotierten p- und n-Gebiet

Im Durchlaßbereich wird dieses Material von Trägern überschwemmt und leitet damit gut. Im Sperrbereich kann sich hier eine Raumladung ausbilden. Das gesamte Gebiet kann bis zur kritischen Feldstärke betrieben werden und so infolge seiner relativ großen Länge viel Spannung aufnehmen. Wie sich die Raumladung mit zunehmender Sperrspannung zunächst ausbreitet, bis das ganze s-Gebiet erfüllt ist, und danach die Feldstärke bis zur kritischen Feldstärke erhöht werden kann, zeigt Bild 3.2.8a) bis c). Bild 3.2.8d) zeigt, wie mit zunehmender Weite w der s-Struktur die mögliche Sperrspannung erhöht werden kann [P 23, S 51].

3.2.9. Metall-Halbleiter-Kontakte

Werden Metall und Halbleiter verbunden, so entstehen in vielen Fällen pn-Übergänge. Dies ist schon lange vom alten *Kristalldetektor* bekannt und wird bei *Spitzendioden* genutzt. Es sind aber auch *sperrfreie Kontakte* möglich. Beides läßt sich relativ gut mit dem Bändermodell erklären. Hierzu dient Bild 3.2.9. Dabei wird ein n-Halbleiter vorausgesetzt. Im Bild bedeuten W_V die obere Kante des Valenzbandes, W_L das Leitungsband, W_F FERMI-Niveau, W_D Donatorenniveau, W_A die Elektronenaffinität des Halbleiters, $\oplus$ örtliche feste Donatoren und $-$ Elektronen im Leitungsband.

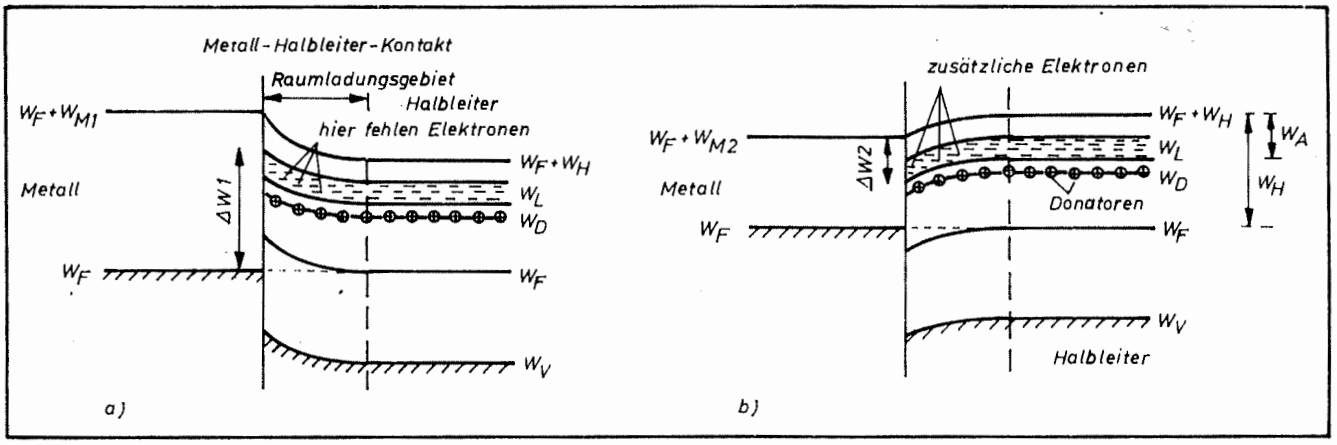


Bild 3.2.9. Bändermodell von Metall-Halbleiter-Kontakten
a) als pn-Übergang, b) als ohmscher Kontakt

W_H entspricht der Austrittsarbeit beim Metall. Jeweils links befinden sich Metalle, links ein Metall mit einer Austrittsarbeit $W_{M1} > W_H$ und rechts mit $W_{M2} < W_H$. An der Kontaktgrenze müssen das oberste Niveau des Halbleiters und die Lage der Austrittsarbeit des Metalls auf gleichem Niveau liegen. Sie gehen an der Grenze also kontinuierlich ineinander über. Im großen Abstand vom Kontaktgebiet besitzt das Bändermodell des Halbleiters wieder seine ursprüngliche Lage mit gleichem FERMI-Niveau von Metall und Halbleiter. Da im Metall wegen seiner sehr guten Leitfähigkeit keine Verformungen des FERMI-Niveaus stattfinden können, geschieht dies ausschließlich im Halbleiter und führt zum Raumladegebiet. Ist die Austrittsarbeit des Metalls größer als der Abstand W_H des Halbleiters, so erfolgt eine Krümmung nach oben (linkes Bild), und im Raumladegebiet fehlen Ladungsträger im Leitungsband, ähnlich wie im Raumladegebiet eines Halbleiters. Im Vergleich zu ihm gemäß Bild 3.2.1 erfolgt die Krümmung jedoch nur einseitig von der Grenzfläche beider Materialien. Dennoch verhält sich dieser Übergang fast genauso wie ein üblicher pn-Übergang. Es *fehlen* jedoch die *Löcher als Ladungsträger*, deshalb besitzen Metall-Halbleiter-Übergänge stets *sehr kurze Schaltzeiten*. Derartige Dioden heißen SCHOTTKY-*Dioden*.

Ist nun $W_{M2} < W_H$ (rechtes Schema), so wird das Halbleiterband nach unten gekrümmt, dadurch wandern zusätzliche Elektronen ein, und das Raumladegebiet ist besonders gut leitend. Es liegt ein sperrfreier Metall-Halbleiter-Kontakt vor.

In Wirklichkeit sind die Verhältnisse nicht ganz so einfach. An der Grenzfläche Halbleiter—Metall sind meist Störterme aus verschiedenen Ursachen, z.B. Kristallbaufehler, Entstehen zusätzlicher Niveaus im verbotenen Band usw., vorhanden. Sie beeinflussen sowohl W_M als auch W_H, und dadurch hängt es auch von technologischen Randbedingungen ab, ob ein *pn*-Übergang oder ein sperrfreier Kontakt entsteht.

Wird statt des *n*-Halbleiters — wie in den bisherigen Betrachtungen — ein *p*-Halbleiter verwendet, so kehren sich die Verhältnisse bezüglich des *pn*-Übergangs und des sperrfreien Kontakts genau um. Im Halbleiter ist dann ja nicht das Donatoren-, sondern das Akzeptorenniveau wirksam (vgl. Bild 3.1.8c)).

Einen Überblick zu den *Barrierenhöhen* bei verschiedenen Metall-Halbleiter-Kontakten zeigt Bild 3.2.10.

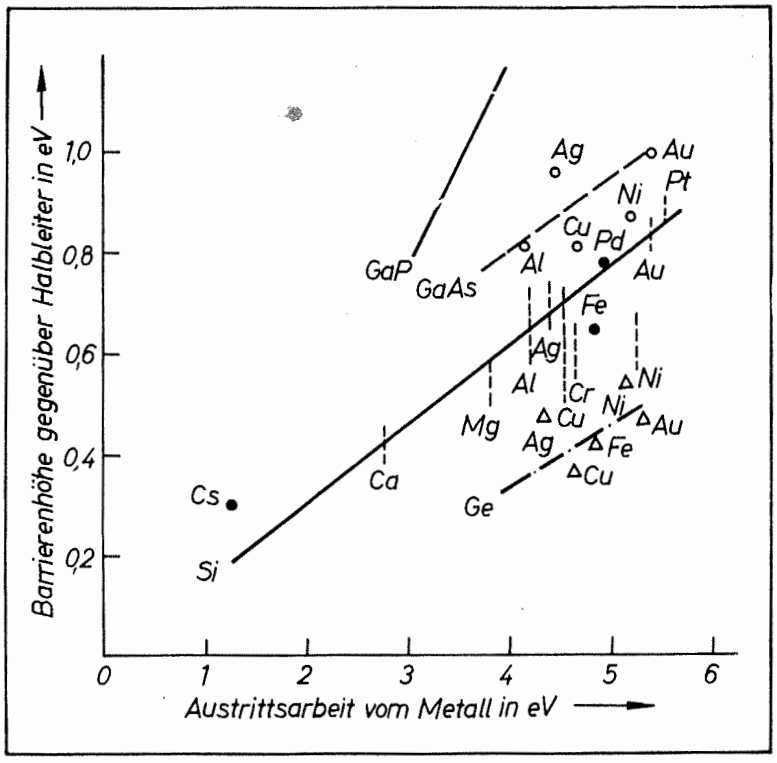


Bild 3.2.10. Barrierenhöhe von Metall-Halbleiter-Übergängen als Funktion der Austrittsarbeit von Metallen für die drei Halbleiter Ge, Si und GaAs

3.3. Feldeffekttransistoren

Entscheidend für den Aufbau elektronischer Schaltungen sind aktive Bauelemente. Soll mit ihnen eine *Verstärkung* realisiert werden, so gibt es nur *zwei Möglichkeiten*:

— Ausnutzung einer Kennlinie mit einem *negativen Kennlinienausschnitt*.
— *Steuerung eines Bauelementes* mit einer geringeren Eingangsleistung, als sie infolge der Steuerung am Ausgang entnommen werden kann.

Der erste Fall wird noch später im Abschnitt 4.1.1.3. allgemein zu behandeln sein. Der zweite Fall sind die üblichen Verstärker. In beiden Fällen wird die *Energie einer äußeren Stromquelle* (meist Gleichstrom) von einer Steuerenergie so beeinflußt, daß

FET

Feldeffekttransistoren

Funktion beruht auf der Steuerung des Leitwertes eines leitfähigen Kanals. Anfang heißt Quelle oder Source (*S*), Ende heißt Senke oder Drain (*D*).

SFET

Sperrschichtfeldeffekttransistoren

auch JFET = Junction-field-effect-Transistor
Leitwertsteuerung wird durch elektronische Steuerung des Kanalquerschnitts mit Hilfe einer Entblößungszone (Sperrschicht) realisiert.

IGFET

Feldeffekttransistoren mit isolierter Steuerelektrode (Gate G)

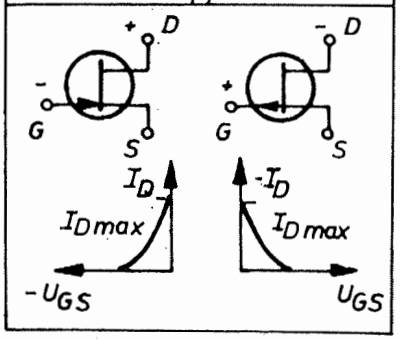
Leitwertsteuerung wird durch Änderung der Ladungsträgerdichte im Kanal mit Hilfe von Influenz realisiert.

SFET

***pn*-Sperrschichtfeldeffekttransistor**

Als Entblößungszone zur Querschnittsmodulation wird die Sperrschicht eines *pn*-Überganges verwendet.

n-Kanal | *p*-Kanal



MESFET

SCHOTTKY-Barrieren-Feldeffekttransistor

Als Entblößungszone zur Querschnittsmodulation wird die Verarmungsrandschicht einer Metall-Halbleiter-SCHOTTKY-Barriere verwendet.

MOSFET ≈ MIS

MOSFeldeffekttransistor

Die isolierte Steuerelektrode (Gate G) wird als Metall- oder Silicium-(polykristallin)-elektrode isoliert durch eine SiO_2-, Si_3N_4- oder Al_2O_3-Schicht auf die in Silicium-Planartechnik hergestellte Transistorstruktur aufgebracht. Das Grundmaterial heißt Bulk (B).

TFT

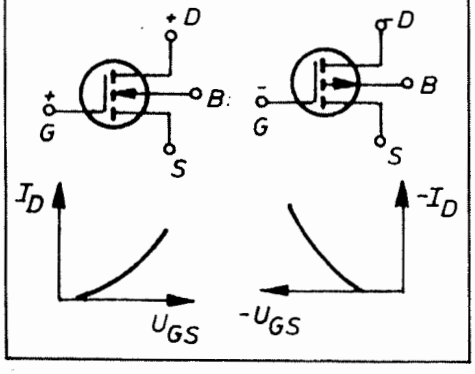
Dünnschicht-Transistor

Alle Bestandteile des Transistors (Gateelektrode, Gateisolator, Halbleiter, Drain- und Sourceelektrode) werden in Form dünner Schichten auf ein Glas- oder Keramiksubstrat gebracht.

Anreicherungstyp

Enhancement-Typ
Selbstsperrender Typ
Ohne Gatespannung ist kein leitender Kanal vorhanden. Dieser wird erst durch eine positive (*n*-Kanal-) oder negative (*p*-Kanal-) Gatespannung angereichert.

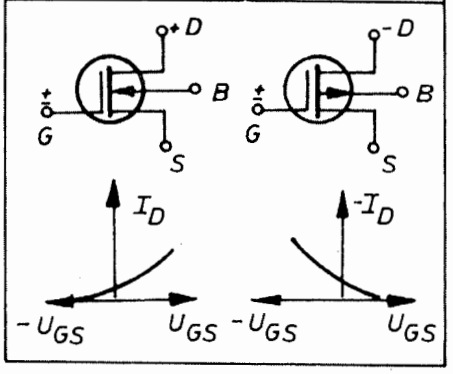
n-Kanal | *p*-Kanal



Verarmungstyp
Entblößungstyp

Depletion-Typ
Selbstleitender Typ
Bei der Gatespannung Null ist bereits ein leitender Kanal vorhanden. Durch Anlegen einer Gatespannung wird in diesem Kanal die Ladungsträgerzahl verändert.

n-Kanal | *p*-Kanal



Speicher-FET
(FEFET, MNOST)

Durch Verwendung von ferroelektrischen bzw. geschichteten Dielektrika zwischen Gate und Kanal können Speicherfunktionen realisiert werden.

Bild 3.3.1. Überblick über die feldeffektgesteuerten Bauelemente und ihre Funktionsprinzipien

sie mit entsprechendem Wirkungsgrad umgewandelt wird und dann in der gewünschten Form am Ausgang zur Verfügung steht. Den einfachsten Aufbau und zugleich den besten Zugang zu derartigen Mechanismen bieten die Feldeffekttransistoren (FET). Vom Prinzip her sind sie bereits seit 1930 bekannt. Technologisch konnten sie jedoch erst um 1960 realisiert werden, also ca. 10 Jahre später als der bipolare Transistor. Es gibt heute eine größere Anzahl von Typen. Ihre Benennung ist z. T. unterschiedlich. Von der Steuerelektrode her sind zwei prinzipielle Arten unterscheidbar:

Beim ersten Typ ist die räumliche Ausdehnung einer Sperrschicht entscheidend für die Wirkung. Hierbei handelt es sich um den *Sperrschicht-FET* (SFET) oder, wenn SCHOTTKY-Kontakte verwendet werden, um den MESFET (Metallsperrschicht).

Der zweite Typ nutzt die Influenzwirkung an einem Isolator aus. Die Steuerelektrode ist hierbei gut isoliert. Gleichwertige Bezeichnungen für diese Transistortypen sind: IGFET (isolated gate), MISFET (Metall-Isolation-Silicium) und die gebräuchlichste MOSFET (Metall-Oxid-Silicium). In Sonderfällen wird der IGFET mittels polykristalliner dünner Schichten hergestellt. Hierfür sind die Namen *TFT* (Thin-Film-Transistor) oder Dünnfilmtransistor gebräuchlich. Zu diesen Transistortypen gibt es noch weitere Einteilungen. Sie unterscheiden, ob der leitende Kanal aus *n*- oder *p*-Material besteht und, bei den IGFET, ob der Kanal ohne Steuerspannung gesperrt oder leitend ist. Einen schematischen Überblick hierzu gibt Bild 3.3.1. Weitere Details hierzu werden noch behandelt.

3.3.1. Sperrschicht-FET

Den prinzipiellen Aufbau eines Sperrschicht-FET zeigt Bild 3.3.2a). Er besteht überwiegend aus *n*-dotiertem Halbleitermaterial. An seinen Enden befinden sich die sperrschichtfreien Anschlüsse *S* und *D* (Source und Drain.) Der Abstand zwischen

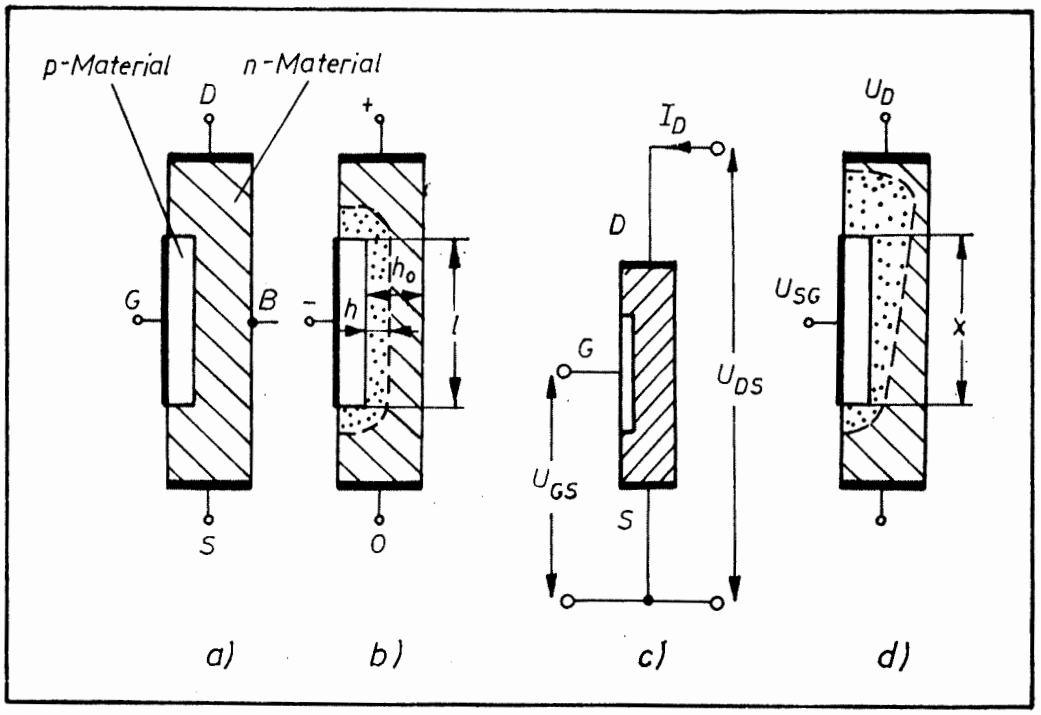


Bild 3.3.2. Zum Sperrschicht-FET

a) Prinzipieller Aufbau

b) Mit Abmessungen und eingetragener Sperrzone (gepunktet)

c) Zur Definition der Spannungen und Ströme

d) Unsymmetrische Sperrzone infolge der Überlagerung von Gate- und Drainspannung

den beiden Anschlüssen beträgt etwa 10 ... 500 μm. In das Grundmaterial ist der Länge nach ein dünner p-leitender Abschnitt eingelassen. Er dient zur Steuerung des Widerstandes zwischen S und D und trägt den Namen *Tor* oder *Gate* (G).

Prinzipiell kann ein Sperrschicht-FET auch mit einem p-dotierten Kanal (Grundmaterial) und einem n-leitenden Gate aufgebaut werden. Ohne Einschränkung der Allgemeinheit soll im folgenden fast ausschließlich der n-Kanal-Sperrschicht-FET behandelt werden. Die Majoritätsträger für die Stromleitung von S nach D sind dann Elektronen. Für den p-Kanal-Typ gelten alle Betrachtungen analog, wenn gleichzeitig die Vorzeichen der Betriebsspannungen sowie Elektronen und Löcher vertauscht werden.

Zur Vereinfachung der Betrachtungen werde zeitweilig ein Punkt B gegenüber dem Gate als Bezugspunkt für die Spannung am Gate gewählt. Beim realen Sperrschicht-FET ist ein derartiger Anschluß nie herausgeführt. Wird an das Gate eine negative Spannung gelegt, so ist der pn-Übergang vom Gate zum n-Material gesperrt. Der Strom, der hierbei durch das Gate fließt, ist bei Silicium sehr klein, nämlich weniger als 10^{-9} A, das bedeutet einen Eingangswiderstand um 10^9 Ω. Änderungen der Gatespannung erfolgen daher praktisch leistungslos. Dies gilt jedoch nur so lange, wie die Gatespannung negativ bleibt. Beim komplementären FET mit p-Kanal muß dagegen die Gatespannung stets positiv bleiben. Bei GaAs sind die Sperrströme noch kleiner.

In der Umgebung eines gesperrten pn-Überganges bildet sich stets eine trägerarme Raumladungszone aus. Gemäß Bild 3.2.4 hängt ihre Ausdehnung von der Größe der Gatespannung ab. Für die Breite h der trägerarmen Zone in Bild 3.3.2b) gilt die Näherung

$$\sqrt{\frac{h}{h_0}} \approx \frac{U_G}{U_0}. \tag{1}$$

Hierin bedeutet h_0 die nutzbare Breite des n-Materials, wenn die Gatespannung Null ist. U_0 ist jene Spannung, bei der die Raumladungszone die volle Höhe h erreicht.

Der Widerstand des Kanals von S nach D wird sowohl von der nutzbaren Höhe $h - h_0$ als auch von der Länge l des Kanals und seiner Breite b sowie von seinem spezifischen Widerstand ϱ bestimmt. Mit guter Näherung gilt

$$R_{DS} = \varrho \frac{l}{(h_0 - h)\, b} \tag{2}$$

oder

$$\frac{1}{R_{DS}} \approx \frac{b}{\varrho l}\left[1 - \left(\frac{U_G}{U_0}\right)^2\right]. \tag{3}$$

Dieser Widerstand wird also von U_G gesteuert. Es werde deshalb in Bild 3.3.2 das untere Ende als gemeinsamer Punkt betrachtet, so daß eine Schaltung gemäß Bild 3.3.2 entsteht. Das untere Ende wird dann als Quelle oder Source (S) bezeichnet. Wird hierauf die Gatespannung bezogen, heißt sie U_{GS}. Um den Widerstand zu messen, kann eine Spannung an den oberen Anschluß gelegt werden. Er ist dann die Senke oder Drain (D). Der Strom durch den Transistor entspricht dann Gl. (3). Dies gilt aber nur so lange, wie durch den Spannungsabfall an der Strecke S—D keine Veränderung der Feldverhältnisse erfolgt, also für sehr kleine Spannungen U_{DS}. Wird die Drainspannung größer, so beeinflußt der Spannungsabfall im Halbleitermaterial die örtliche Feldstärke gegenüber dem Gate. In Bild 3.3.2d) gehört zur Koordinate x die Spannung U_x. Die Breite der Raumladungszone an dieser Stelle ist dann durch (1) und folgende Spannung gegeben:

$$U_G = U_{GS} - U_x\,. \tag{4}$$

Insgesamt tritt also infolge der Drainspannung eine unsymmetrische Raumladungszone auf. Sie bewirkt ihrerseits, daß auch der wirksame Widerstand und damit die Feldstärke im Material in der Richtung x zunimmt. Der größere Teil der Drainspannung fällt infolgedessen in der Nähe des Drain ab.

Entsprechend der eingezeichneten Polarität von U_{DS} erhöht sich auch die wirksame negative Gatespannung. Mit wachsender Drainspannung nimmt deshalb die leitfähige Fläche des Kanals stark ab. Als Folge wächst der Widerstand R_{DS} erheblich, und insgesamt wird so der Strom I_D nahezu unabhängig von der Drainspannung. Dieser Effekt heißt *pinch-off*. Er zeigt sich deutlich in den Ausgangskennlinien von Bild 3.3.3a). Da hierbei der Kanal zumindest in der Nähe des Drain sehr stark abgeschnürt ist, wird der Pinch-off-Bereich auch vielfach *Abschnürbereich* genannt. Die Kennlinien weisen hierdurch einen *Sättigungscharakter* auf. Es wird deshalb zuweilen auch vom Sättigungsbereich gesprochen.

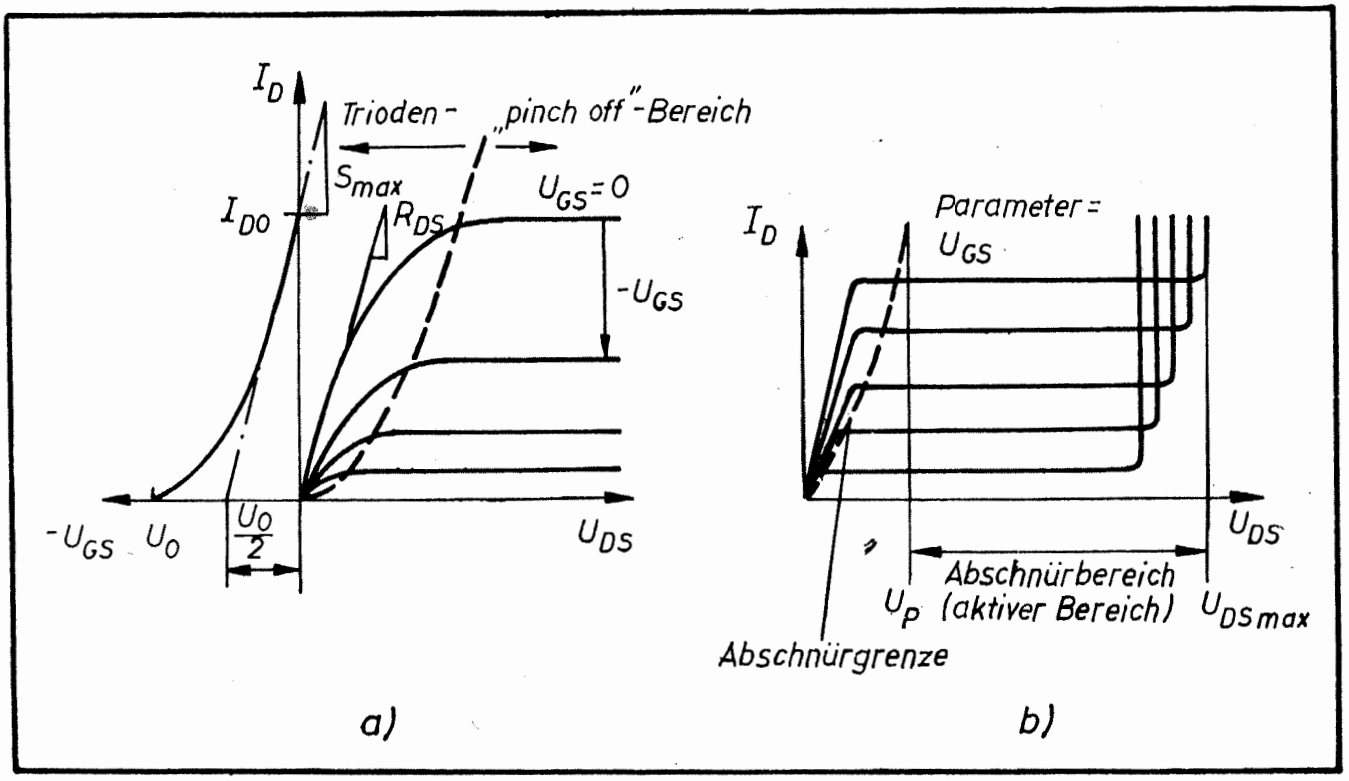


Bild 3.3.3. Kennlinienfelder eines Sperrschicht-FET

a) Steuer- (links) und Ausgangskennlinie (rechts). Die Bezeichnung Triodenbereich ist geschichtlich bedingt und geht auf analoge Kennlinien bei der Röhren-Triode zurück. Im ähnlichen Sinn wird der Pinch-off-Bereich auch Pentodenbereich genannt

b) Andere Bezeichnung für die Grenzen und zusätzlich eingezeichnete Kennlinien bei Durchbruch, bei zu hoher Drainspannung

Der Pinch-off-Effekt tritt auf, wenn die Raumladungszone an ihrer breitesten Stelle, das heißt am oberen Ende in Bild 3.3.2d) die Höhe h_0 erreicht. Diese Grenze ist in Bild 3.3.3a) gestrichelt gezeichnet.

Wird die Drainspannung weiter erhöht, so erfolgt ein Durchbruch, wie er im Abschnitt 3.2.6. diskutiert wurde. Hierdurch entstehen dann Kennlinien, wie sie Bild 3.3.3b) zeigt.

3.3.2. MOSFET

Den prinzipiellen Aufbau eines MOSFET zeigt Bild 3.3.4. Auch hier sei wiederum von einem speziellen, in diesem Fall einem p-Substrat ausgegangen. Sein Anschluß

wird vielfach getrennt herausgeführt und mit B (Bulk) bezeichnet. In das Substrat sind zwei n-leitende Bereiche eingefügt. Sie werden wieder als Source und Drain benutzt. Unabhängig von der Polung kann beim MOSFET zunächst also kein Strom von Source nach Drain fließen. In jedem Falle ist nämlich einer der beiden pn-Übergänge zum p-Substrat gesperrt. Über dem Zwischengebiet befindet sich eine meist mit SiO_2 isolierte metallische Elektrode. Sie besteht gewöhnlich aus Aluminium und stellt das Gate des MOSFET dar. Die Dicke der Isolierschicht hängt u. a. von der geforderten Durchschlagsfestigkeit ab. Bei der üblichen Dicke von etwa 0,1 μm sind etwa 90 V zulässig. Der Isolationswiderstand des Gate beträgt ca. 10^{15} Ω. Wird das Gate nun positiv gegenüber Bulk aufgeladen, so sammeln sich an der Oberfläche des Substrats gemäß Abschnitt 1.3. durch Influenzwirkung Elektronen. Dieser Elektronenüberschuß sättigt zunächst die freien Löcher der Akzeptoren ab. Bei weiterer Erhöhung der positiven Gatespannung bildet sich im Substrat unmittelbar

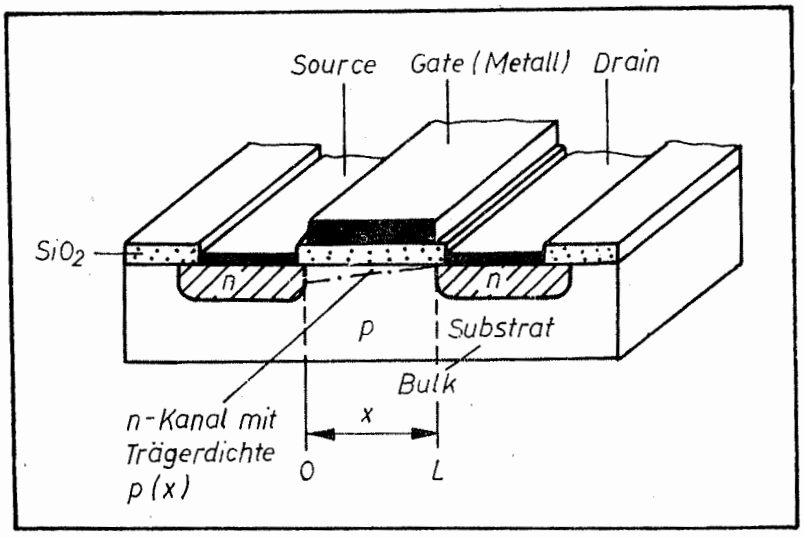


Bild 3.3.4. Prinzipieller Aufbau eines MOSFET

gegenüber dem Gate eine sehr dünne Schicht (10^{-2} μm) mit freien Elektronen aus. Dadurch wird das ursprüngliche p-Material an dieser Stelle n-leitend, so daß die pn-Übergänge zu Source und Drain überbrückt werden. In dem dünnen n-leitenden Kanal kann ein Strom von S nach D fließen.

Da durch die Gatespannung die Anzahl der influenzierten Elektronen beeinflußbar ist, steuert es den Widerstand zwischen Source und Drain. Jene Gatespannung, für die sich erstmalig freie Elektronen durch Influenz ausbilden und damit ein Drainstrom fließen kann, heißt Schwellenspannung. Sie entspricht der Spannung U_0 in Bild 3.3.3a). Durch verschiedene Bedingungen innerhalb des MOSFET, auf die im Abschnitt 3.3.2.3. eingegangen wird, kann U_0 sowohl positive als auch negative Werte annehmen.

Solange die Drainspannung U_{DS} hinreichend klein ist, können wie beim Sperrschicht-FET Source und Drain ohne wesentliche Folgen vertauscht werden. Für größere Drainspannungen wird dagegen wieder durch den Spannungsabfall längs des Kanals die Kanalbreite beeinflußt (Bild 3.3.2d)). Infolgedessen gelten für den MOSFET bezüglich seiner Ausgangskennlinien ähnliche Beziehungen wie beim Sperrschicht-FET. Wichtig hierfür ist wiederum die richtige Polarität der Drainspannung. Beim bisher behandelten n-Kanal-MOSFET muß sie positiv und beim komplementären p-Kanal-MOSFET negativ sein. Dies bedeutet, daß jeweils die Draindiode gesperrt ist. In ihrer Nähe wird auch wieder bei MOSFET der stromführende Kanal abgeschnürt.

Der Unterschied zwischen dem Sperrschicht-FET und MOSFET besteht hauptsächlich hinsichtlich der Gate-Isolation. Im Gegensatz zum Sperrschicht-FET kann daher jeder MOSFET je nach Bedarf mit positiver und negativer Gatespannung betrieben werden. Zusätzlich entfällt beim MOSFET der temperaturabhängige Gatesperrstrom des Sperrschicht-FET. Durch den extrem hohen Isolationswiderstand besteht beim MOSFET der große Nachteil einer ungewollten hohen Aufladung des Gate. Bei der zusätzlich kleinen Kapazität kann leicht die Spannung so groß werden, daß ein Durchschlag erfolgt. Er führt in der Regel zur Zerstörung des MOSFET. Während des Versandes und beim Einbau des MOSFET in die Schaltung wird daher stets das Gate zu den beiden anderen Elektroden kurzgeschlossen. Beim Betrieb des MOSFET sind mitunter zum Schutz des Gate besondere Schaltungsmaßnahmen notwendig.

3.3.3. Bändermodell beim MOSFET

Die Eigenschaften des MOSFET lassen sich aus dem Bändermodell ableiten. Hierzu ist von den Bändermodellen für Metalle, Isolatoren und p-Halbleiter entsprechend Bild 3.1.5 bzw. 3.2.9 auszugehen. Für den entsprechenden Übergang folgt dann Bild 3.3.5a). Dabei ist vorausgesetzt, daß Gate und Substrat auf dem gleichen Potential liegen. Bei positiver Gatespannung werden die beweglichen Löcher gegenüber dem Gate infolge der Influenz von der Oberfläche des Substrats vertrieben. Dadurch entsteht eine Raumladung, die von den negativ geladenen Akzeptoren herrührt. Sie bewirken eine Bandkrümmung in der Nähe des Isolators (Bild 3.3.5b)). Wird die positive Gatespannung noch mehr vergrößert, so steigt die Raumladung weiter an, und die Biegung der Bänder nimmt zu, bis das Leitungsband das FERMI-Niveau erreicht hat. Dann können freie Elektronen im Leitungsband influenziert werden (Bild 3.3.5c)). Erst dann ist die Schwellenspannung für das Gate erreicht. Das p-Material hat in unmittelbarer Nähe des Gate die Eigenschaften eines n-Leiters angenommen. Dieser Vorgang heißt Ladungsträgerinversion.

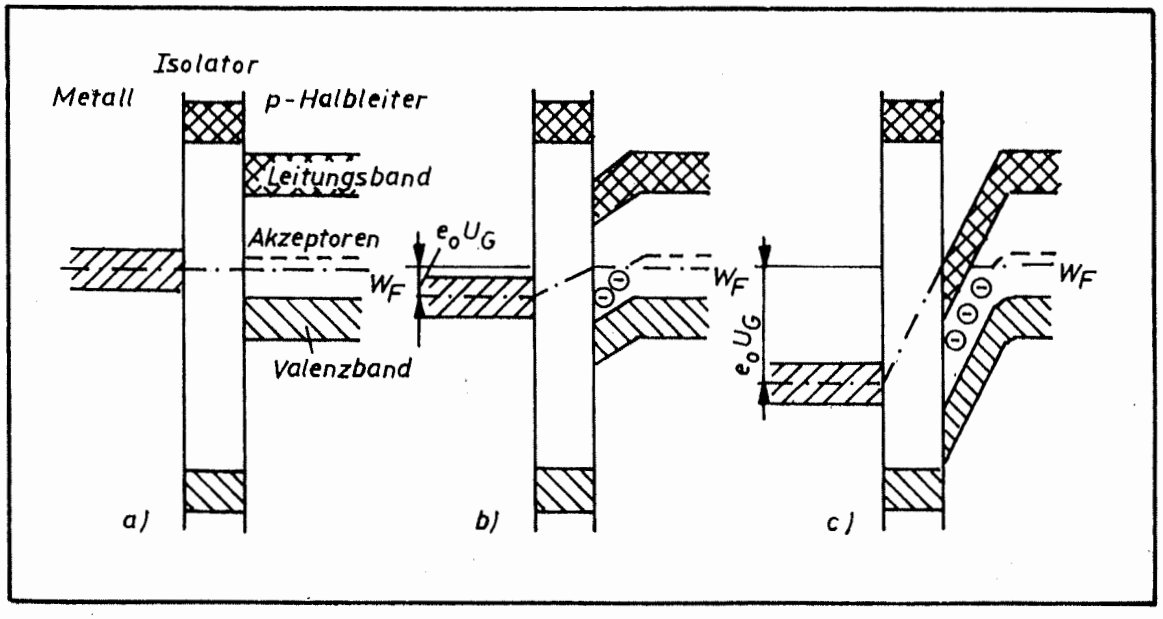


Bild 3.3.5. Bändermodell des Gate-Isolator-Substratbereiches beim MOSFET. Von (a) nach (c) steigt die positive Gatespannung. Bei (b) werden zunächst nur Löcher gebunden, so daß Akzeptoren eine negative Raumladung bilden. Bei (c) ist die Schwellenspannung überschritten. Es werden freie Ladungsträger ins Leitungsband influenziert

3.3.4. Anreicherungs- und Verarmungstyp

Sowohl Verschiebungen des FERMI-Niveaus zwischen dem Leiter und dem Halbleiter als auch *Krümmungen des Bandes* des Halbleiters können durch Effekte hervorgerufen werden, die *nicht von einer äußeren Gatespannung* herrühren. *Drei Arten* sind hierbei zu unterscheiden:

- In der Praxis wird die SiO_2-Schicht durch Oxydation des Siliciums hergestellt. Hierbei wird meist eine *positive Ladung im Isolator dauerhaft gebunden.* Sie bewirkt die gleiche Bandkrümmung wie eine positive Gatespannung.
- Metall und Halbleiter besitzen in der Regel unterschiedliche *Austrittsarbeit.* Ihre Differenz bewirkt die Verlagerung des FERMI-Niveaus von Halbleiter und Metall ohne Bandkrümmung. Sie entspricht also einer additiven Verschiebung der äußeren Gatespannung. Durch entsprechende Materialzusammensetzung kann so die Schwellenspannung zu negativen oder positiven Werten verschoben werden. (Vgl. 3.2.9.)
- An der Oberfläche des Halbleitermaterials können gewisse Unregelmäßigkeiten im Kristallaufbau existieren. Diese *Oberflächenzustände* entsprechen eingebauten Störtermen. Sie können zusätzlich freie Ladungsträger binden. Die Influenzladung muß zunächst die Störterme befriedigen, bevor die Löcher vertrieben und freie Elektronen gebildet werden können. Bei den heutigen Technologien sind Oberflächenzustände weitgehend zu beherrschen.

Die Summe der drei genannten Einflüsse führt zu einer *Verschiebung der Eingangskennlinie,* wie sie in Bild 3.3.6 angedeutet ist. Dabei sind die beiden möglichen Typen des *n*-Kanal-MOSFET mit einem *p*-Substrat und des *p*-Kanal-MOSFET mit einem *n*-Substrat getrennt betrachtet.

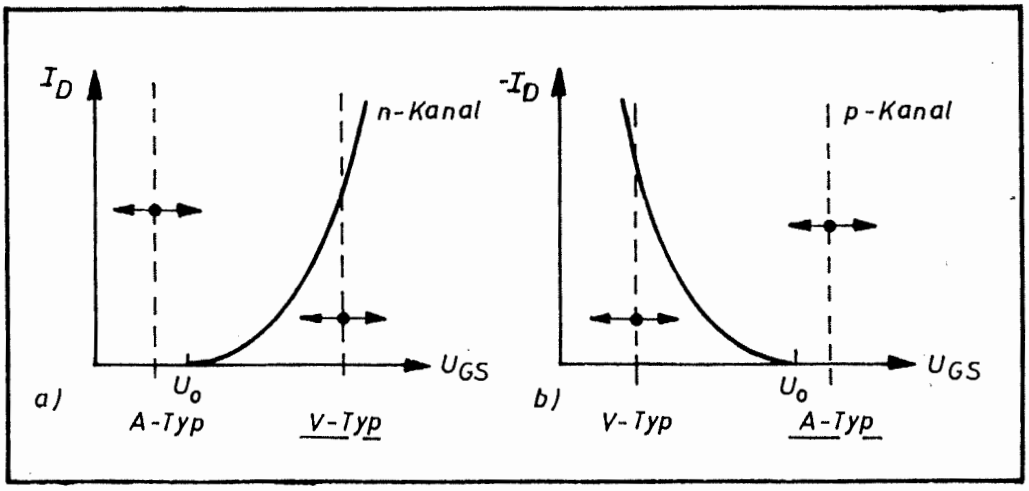


Bild 3.3.6. Zur Einteilung von MOSFET nach der Eingangskennlinie. A-Typ bedeutet Anreicherungs- und V-Typ Verarmungstyp. In der üblichen Technologie überwiegen die jeweils unterstrichenen Typen

Vielfach werden MOSFET für digitale Schaltungen verwendet. Dies erlaubt, zwei typische Fälle für die Lage der Eingangskennlinie (Bild 3.3.6) zu unterscheiden. Bei der Gatespannung Null kann einmal bereits ein leitender Kanal vorhanden sein, zum anderen kann er aber auch noch fehlen. Ohne äußere Gatespannung kann also ein MOSFET leitend oder nichtleitend sein. Durch eine zweckmäßig gepolte Gatespannung wird er gesperrt bzw. geöffnet.

Falls der MOSFET *durch eine äußere Gatespannung gesperrt wird,* erfolgt, eine Ladungsträgerverarmung in der Kanalzone. Dieser Typ heißt daher *Verarmungs-, Entblößungs-, selbstleitender* bzw. *Depletion-Typ* (V-Typ). Im anderen Fall werden durch die Gatespannung Ladungsträger erzeugt. Er heißt daher *Anreicherungs-, selbstsperrender-* bzw. *Enhancement-Typ* (A-Typ).

Der Sperrschicht-FET gehört also zum Verarmungstyp. Zusätzlich darf bei ihm die Gatespannung jedoch nicht den Wert Null überschreiten.

In der Regel überwiegt von den drei Einflüssen für die Verlagerung der Schwellenspannung meist der Einbau von Ladungen. Aus diesem Grunde sind MOSFET mit

p-Kanal überwiegend vom Anreicherungstyp und jene mit n-Kanal meist vom Verarmungstyp. Diese beiden Fälle sind deshalb in Bild 3.3.6 unterstrichen.

3.3.5. Einfluß der Substratspannungen und punch-through

Der Kanal beim MOSFET stellt eine Inversionsschicht dar. Mit ihrer Ausbildung entsteht also ein pn-Übergang innerhalb des Substrats. Beträgt die *Dicke der Inversionsschicht* etwa 0,01 µm, so hat die folgende Verarmungsschicht Abmessungen um 1 µm. In der Verarmungsschicht sind die Löcher durch die influenzierten Elektronen abgesättigt, und die notwendige Raumladung entsteht durch die negativ geladenen Akzeptoren. Die Lage der Verarmungs- und Inversionsschicht zeigt schematisch Bild 3.3.7a). Der pn-Übergang zur Inversionsschicht kann nun auch wie beim MOSFET durch die *Substrat- oder Bulkspannung* beeinflußt werden. Diese Spannung wird ebenfalls auf Source bezogen und trägt die Bezeichnung U_{BS}. Mit ihr lassen sich Änderungen der Eingangskennlinien erreichen. Die Polarität der Spannung kann nur so gewählt werden, daß sich die Verarmungsschicht verbreitert, d. h. in Sperrichtung. Beispiele für den Einfluß der Bulkspannung an einem n-Kanal-Verarmungs-MOSFET zeigt Bild 3.3.7b).

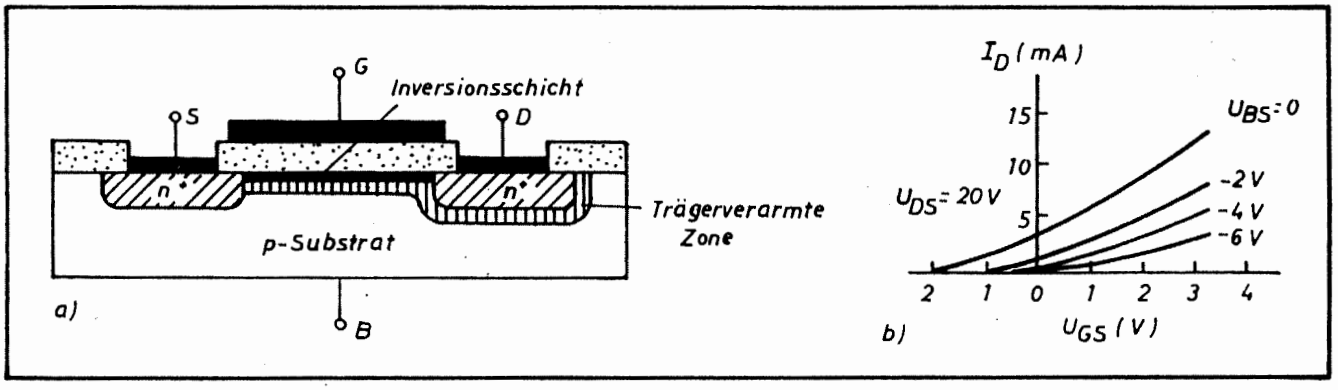


Bild 3.3.7. a) Kanal und trägerverarmte Zone in einem normal betriebenen p-Kanal-MOSFET
b) Einfluß der Bulkspannung auf die Kennlinie eines MOSFET

Die Steuersteilheit über das Substrat ist häufig näherungsweise genau so groß wie die Steilheit von Gate her. Durch eine etwas positivere Gatespannung kann leicht für beide Spannungen ein optimaler Arbeitspunkt erreicht werden.

Bei den meisten MOSFET sind innerhalb des Gehäuses Substrat und Source kurzgeschlossen. Hierbei wird lediglich der ohnehin durch die Drainspannung leitende pn-Übergang Source—Substrat zusätzlich kurzgeschlossen.

Bei einer *hohen Drainspannung* wird der pn-Übergang in der Umgebung des Drain stark beansprucht (trägerverarmte Zone in Bild 3.3.7a)). Es besteht also dann die *Gefahr eines Durchschlages*. Sie ist besonders groß bei schwacher Dotierung des Substrats.

Mit steigender Drainspannung dehnt sich die Breite der gesamten trägerarmen Zone aus. Bei einem MOSFET mit kurzem Kanal kann diese Zone unmittelbar das Source erreichen. Die dort vorhandenen Träger können dann, ohne den Kanal zu benutzen, direkt zum Drain gelangen. Dieser Effekt heißt punch-through. Er wirkt sich als steiler Stromanstieg in Bild 3.3.3b) aus und begrenzt vor allem die Frequenzeigenschaften der Hochfrequenztransistoren entsprechend. Ferner begrenzt er die Aussteuerbarkeit der Drainspannung.

3.3.6. Theorie der Ladungssteuerung

Die Ausführungsformen und die *Wirkungsprinzipien* der *MOSFET* und *SFET* sind *unterschiedlich*. Dennoch zeigen beide die praktisch *gleichen Kennlinien*. Darüber hinaus führen auch die theoretischen Ableitungen zu gleichen Kennlinien. Hier soll eine möglichst einfache Ableitung am Beispiel des MOSFET unter Berücksichtigung von Bild 3.3.4 erfolgen. Das Gate soll dabei von Source bis Drain reichen und die Länge L besitzen. Seine Kapazität gegenüber dem Substrat sei

$$C_G = \frac{\varepsilon_0 \varepsilon_r A}{d}, \tag{5}$$

wobei d die Dicke der Oxidschicht, A die Fläche des Gate, ε_0 die absolute und ε_r die relative Oxid-Dielektrizitätskonstante ist. Mit dem Kanalpotential $V(x)$ am Ort x und dem Potential V_G des Gate gilt dann für den Ort x die Ladungsdichte in x-Richtung

$$dQ = \frac{C_G}{L} [V(x) - V_G] \, dx . \tag{6}$$

Die Geschwindigkeit der Ladungsträger folgt aus dem Gradienten des Potentials und der Beweglichkeit der Majoritätsträger zu

$$v = -\mu \frac{dV}{dx} . \tag{7}$$

Bei allen FET gibt es nur Majoritätsträger. Deshalb sind infolge der etwa 3fach höheren Beweglichkeit der Elektronen gegenüber den Löchern (s. Bild 3.1.10) die n-Kanal-FET stets schneller bzw. für höhere Frequenzen geeignet. Mit dem spezifischen Widerstand in x-Richtung ϱ_x folgt für den Strom

$$I = \varrho_x \cdot v = \frac{dQ}{dx} \cdot v = \frac{C_G}{L} \cdot \mu [V_G - V(x)] \frac{dV}{dx} . \tag{8}$$

Durch Integration mit den Randbedingungen

$$V(0) = 0; \qquad V(L) = U_{DS} \quad \text{und} \quad V_G = U_G \tag{9}$$

im Bereich $U_{DS} < U_G$ folgt dann

$$\boxed{I = \frac{\mu C_G}{L^2} \left(U_G U_{DS} - \frac{1}{2} U_{DS}^2 \right)} . \tag{10}$$

Der Faktor vor der Klammer ist eine Konstante mit der Einsicht A/V^2 und hängt nur vom Material und geometrischen Abmessungen ab.

Die *Kennlinien sind Parabeln*, wie sie in Bild 3.3.3a) bis zur Pinch-off-Grenze zu erkennen sind. Für die Pinch-off-Grenze gilt genau $\partial I / \partial U_{DS} = 0$ und damit $U_G = U_{DS}$. Dies bedeutet, am drainseitigen Ende ist die Kondensatorspannung gleich Null und damit auch die Ladung dQ. Der Kanal ist also gerade eingeschnürt. Der Strom im Scheitelpunkt beträgt

$$\boxed{I_{max} = \frac{\mu C_G}{2L^2} U_G^2} . \tag{11}$$

Es kann nun angenommen werden, daß für Spannungen $U_{DS} > U_G$ Sättigung vorliegt und der maximale Strom erhalten bleibt (Bild 3.3.3a)). Die Erklärung hierfür ist, daß $\mu = \text{const}$, also je Volumeneinheit das Ohmsche Gesetz gilt. Dies ist aber nicht mehr bei der *Kanaleinschnürung* oberhalb des Pinch-off-Bereiches gegeben.

Der Zusammenhang zwischen Strom und der Spannung U_{DS} ist gemäß (10) nicht linear. Deshalb kann nicht mehr von einem durch das Gate gesteuerten Widerstand

gesprochen werden. Dies gilt nur für kleine U_{DS}, wenn das quadratische Glied vernachlässigt werden kann.

In vorangegangenen Betrachtungen wurde gezeigt, daß durch die Ladungsträger an den Grenzflächen ein Kontaktpotential bzw. eine zusätzliche *Potentialverschiebung* auftritt. Deshalb ist die bisher wirksame Gatespannung U_G um diese innere Spannung (Schwellenspannung) U_S bezüglich der außen angelegten Spannung U_{GS} zu korrigieren,

$$\boxed{U_G = U_{GS} - U_S}\,, \tag{12}$$

und es gilt

$$I = \begin{cases} \text{(Triode)} & \dfrac{\mu C_G}{L^2}\left[(U_{GS} - U_S)\,U_{DS} - \dfrac{1}{2}\,U_{DS}\right] \\ \text{(Sättigung)} & \dfrac{\mu C_G}{L^2}\,(U_{GS} - U_S)^2 \end{cases}$$

$$\text{für} \quad |U_{GS} - U_S| \begin{Bmatrix} < \\ \geqq \end{Bmatrix} |U_{DS}|\,. \tag{13}$$

Diese Gleichung besitzt *zwei Parameter* zur Anpassung an reale Kennlinien, den Faktor vor der Klammer und die Schwellenspannung U_S. Die größten Abweichungen im Kennlinienverlauf treten vor allem bei der Sättigung, also der unteren Gleichung, auf. Die verschiedenen Ansätze zur besseren Anpassung verwenden dann aber meist nicht physikalische Zusatzterme.

3.3.7. Kleinsignalverhalten und Ersatzschaltbild

In Analogie zur alten Röhrentechnik werden für den FET drei Kenngrößen verwendet (bezüglich der Röhrentechnik sei auf die 1. Auflage dieses Buches bzw. [B 3] verwiesen):

der ausgangsseitige *Innenwiderstand*

$$R_i = \frac{1}{g_D} = \left.\frac{\partial U_{DS}}{\partial I}\right|_{U_{GS}=\text{ const}}\,, \tag{14}$$

die *Steilheit*

$$S = g_m = \left.\frac{\partial I}{\partial U_{GS}}\right|_{U_{DS}=\text{const}}\,, \tag{15}$$

der *Durchgriff*

$$D = \left.\frac{\partial U_{GS}}{\partial U_{DS}}\right|_{I\,=\,\text{const}}\,. \tag{16}$$

Mit g_m und g_D sind entsprechende Leitwerte gegeben, die oft nicht weiter verwendet werden. Der eigentliche Steuereinfluß ist durch g_m bzw. S gegeben.

Es läßt sich zeigen, daß für den *Durchgriff* (wie bei den Röhren) ein *Kapazitätsverhältnis* wirksam ist:

$$D = \frac{\text{Drain-Kanal-Kapazität}}{\text{Gate-Kanal-Kapazität}}\,. \tag{17}$$

Aus der Ableitung der drei typischen FET-Kenngrößen folgt wiederum die aus der Röhrentechnik stammende BARKHAUSEN-Beziehung

$$\boxed{R_i S D = 1}\,. \tag{18}$$

Werden die Gl. (14) und (15) auf den Triodenbereich bzw. Sättigungsbereich angewendet, so folgt aus (13):

Triodenbereich:	Sättigungsbereich:	
$\frac{1}{R_i} = \frac{\mu C_G}{L^2}(U_{GS} - U_V - U_{DS})$,	$R_i \to \infty$,	(19)
$S = \frac{\mu C_G}{L^2} U_{DS}$,	$S = \frac{\mu C_G}{L^2}(U_{GS} - U_V) = S_{max}$.	(20)

Typisch für die FET ist eine Gleichung, die kein Analogon in der Röhrentechnik hat. Im Triodenbereich gilt infolge von (19) und (20)

$$\boxed{\frac{1}{R_i} + S = S_{max}} \quad . \tag{21}$$

Es ist bedeutsam, daß im allgemeinen experimentelle Werte recht gut mit diesen theoretischen Ergebnissen übereinstimmen (Bild 3.3.8).

Aus den abgeleiteten Daten läßt sich leicht ein Ersatzschaltbild ableiten. Dies zeigt Bild 3.3.9. Der Eingang ist durch die Gate-Source-Kapazität C_{GS} und den meist vernachlässigbaren Sperr- bzw. Isolationswiderstand R_{GS} gegeben. Ferner existieren Rückwirkungen vom Drain, die Kapazität C_{GD} und der Widerstand R_{GD}. Die entscheidende aktive Quelle ist die gesteuerte Stromquelle SU_{GS}. Parallel dazu liegt der Verlustwiderstand (meist vernachlässigbare Serienschaltungen der beiden Dioden zum Substrat) R_{DS} und die Kapazität C_{DS}. Dieses Schaltbild und Abwandlungen daraus werden im Abschnitt 5.5.1. für die Schaltungsberechnungen verwendet. Auf dieser

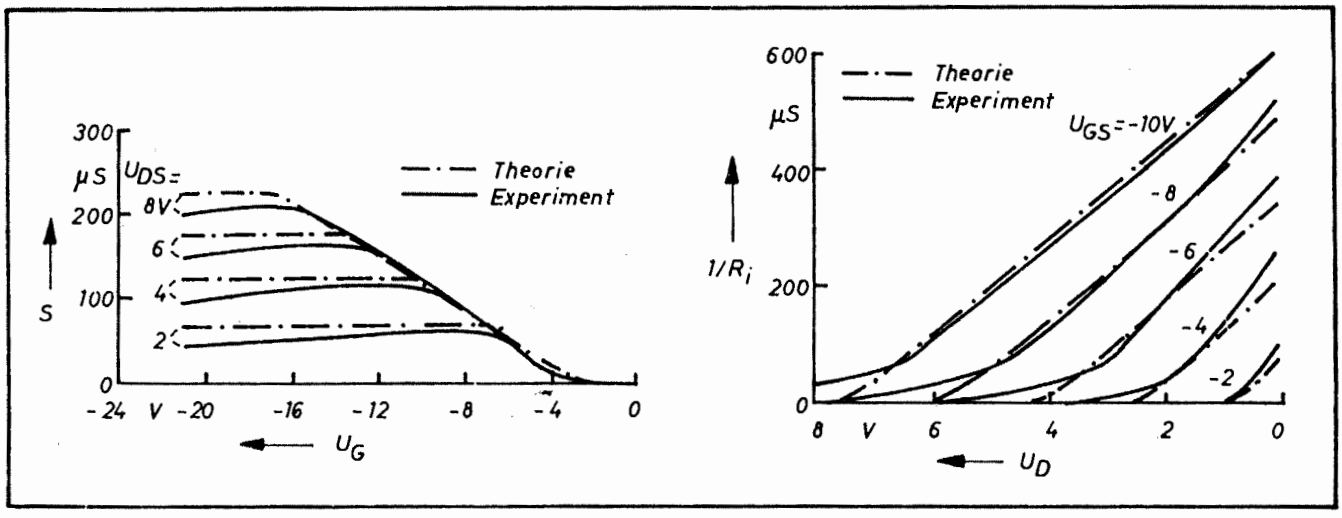


Bild 3.3.8. Vergleich von theoretischen und experimentellen Kennwerten bei einem FET [M 25]

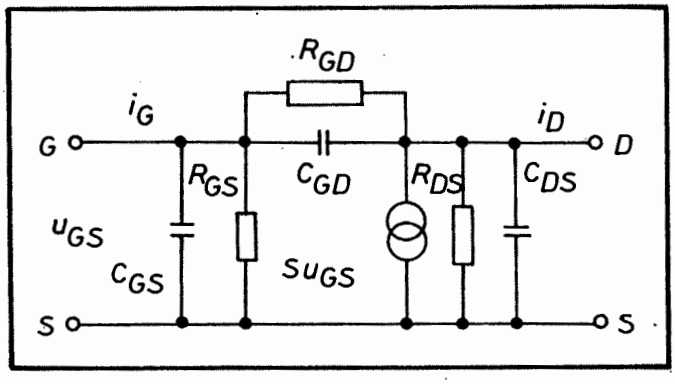


Bild 3.3.9. Ersatzschaltbild für Kleinsignalaussteuerung von FET

Basis läßt sich (wie bei Röhren) ein *Gütefaktor G* ableiten. Er ist das maximal mögliche Produkt aus Bandbreite und Verstärkung der mit dem jeweiligen FET möglichen Schaltungen,

$$G \approx \frac{3\mu\,|U_{GS} - U_S|}{2L} \approx \frac{3v}{2L}, \tag{22}$$

wobei v die im Kanal erreichte Driftgeschwindigkeit ist. Hieraus läßt sich auch jene Frequenz ableiten, bei der die Verstärkung gerade auf Eins abgesunken ist:

$$f_G = \frac{k \cdot G \text{ Hz}}{L/\mu\text{m}}. \tag{23}$$

Der Faktor k hat für SFET die Werte: für GaAs ≈ 50; Si ≈ 25; Ge ≈ 8 und für GaAs-MOSFET $\approx 7{,}5$.

3.3.8. Potentialtopfmodell

Bei den bisherigen Betrachtungen diente das Gate dazu, an der Oberfläche unter dem Isolator einen Kanal für die zu steuernden Ladungsträger zu erzeugen. Der prinzipiell gleiche Aufbau wie bei MOSFET wird aber auch für die später (6.1.5.) zu behandelnden Ladungsverschiebungselemente verwendet. Hierbei wird nun aber gerade ausgenutzt, daß mittels Potentialänderungen am Gate unterschiedlich tief reichende Zonen erzeugt werden können, die von Ladungsträgern entblößt sind. Diese Zonen heißen Potentialtöpfe. In Analogie zum Bild 3.2.9a) und 3.3.5 erklärt sie Bild 3.3.10. Das Potential wird also so gewählt, daß die Bandkante sich im Gegensatz zum MOSFET nach oben verbiegt. Diese Verbiegung und damit die *Tiefe des Potentialtopfes* ist dann etwa *proportional der angelegten Spannung*. In diesen Potentialtopf

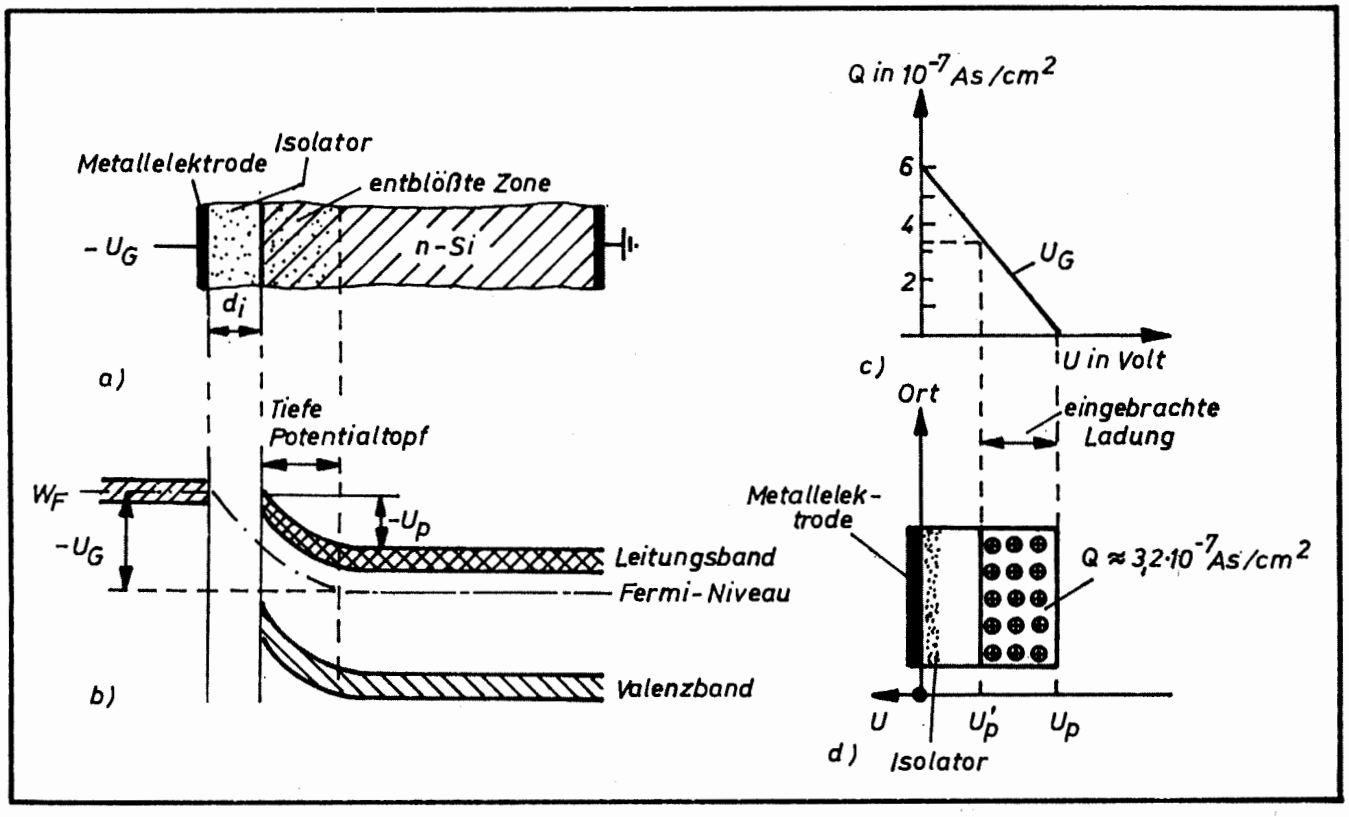


Bild 3.3.10. Zum Prinzip von Ladungstransportbauelementen

a) Schema des Aufbaus gemäß MOSFET

b) Durch entsprechende Polung der Gatespannung wird eine von Ladungsträgern entblößte Zone geschaffen. Ihr entspricht eine Verbiegung der Bänder am Rande des Isolators. Die entblößte Zone stellt einen Potentialtopf dar, der mit Minoritätsträgern von unten beginnend gefüllt wird (d). Die Ladungsmenge hängt u. a. von der Füllhöhe ab (Gerade in c))

können *Minoritätsträger* (in diesem Fall Löcher) eindiffundieren bzw. eingebracht werden. Sie füllen den Topf *ähnlich* wie eine *Flüssigkeit* von unten beginnend auf. Die vom Potentialtopf zu speichernde Ladungsmenge hängt näherungsweise von der Dicke d_i und der Dielektrizitätskonstanten ε_i des Isolators sowie von der wirksamen Potentialtopftiefe U_p und der Fülltiefe U_p' ab:

$$Q \sim \frac{\varepsilon_i}{d_i}(U_p - U_p') . \tag{24}$$

Wenn sich nebeneinander zwei unterschiedlich tiefe Töpfe befinden, wird folglich zuerst der unterste Topf aufgefüllt. Wenn die Topftiefe mit der Gatespannung gesteuert wird, können diese Ladungsträger auch definiert bewegt werden. Dieser ganze Effekt ist nur möglich, bis sich wieder ein thermodynamisches Gleichgewicht einstellt. Diese Zeit hängt u. a. von der Breite der verbotenen Zone, von Oberflächentermen (Oxid—Halbleiter) und von der Temperatur ab. Deshalb liegt die Zeit bis zum thermodynamischen Gleichgewicht im Bereich von etwa 100 µs bis zu Minuten. In dieser Zeit sind die gesteuerten *Transportprozesse der Ladungsträger* zu realisieren.

3.4. Bipolartransistoren

Der Bipolartransistor wird auch zuweilen *Injektionstransistor* genannt. Damit steht er im Gegensatz zu den bisher behandelten Feldeffekttransistoren FET, die auch Unipolartransistoren heißen. Der Grund dieser Bezeichnung liegt darin, daß die FET nur eine Ladungsträgerart verwenden, die Bipolartransistoren dagegen beide Arten benötigen. Der Bipolartransistor ist der historisch ältere. Er geht auf die am 23. 12. 1947 von BARDEEN und BRATTAIN entdeckte Injektion von Minoritätsträgern an *pn*-Übergängen zurück. Auf dieser Basis sagte SHOCKLEY 1949 den Transistor voraus, und zwei Jahre später konnte er realisiert werden. 1956 erhielten die drei Forscher dafür den Nobelpreis für Physik.

Über das Wort Transistor gibt es eine Notiz, die anläßlich des 20. Geburtstages des Transistors in der Zeitschrift "Electronic Design" erschien. Danach stammt es von JOHN R. PIERCE, der leitender technischer Mitarbeiter in BRATTAINS Labor bei Bell war. Er spürte, daß der Spitzenkontakt schaltungsanalog zur Röhre ist. Bei ihr ist die "transconductance" wesentlich. Hier ist es die "transresistance". So bildete er das Wort „Transistor", und BRATTAIN stimmte ihm zu. Der Transistor verwendet einen „übertragenen Widerstand".

Der theoretisch seit 1930 bekannte Unipolartransistor erlangte aus technologischen Gründen erst ab etwa 1960 praktische Bedeutung.

3.4.1. Grundmodell

Das heutige Grundmodell aller Bipolartransistoren sind zwei eng benachbarte *pn*-Übergänge. Sie können also in der Folge *npn* bzw. *pnp* auftreten und sind damit bezüglich der Spannungen und Ströme zueinander komplementär. Das Mittelgebiet muß immer relativ schmal sein, ca. 0,2 µm (HF-Transistor) bis 30 µm (NF-Leistungstransistor), und ist meist recht niedrig dotiert. Die anderen beiden Gebiete sind meist erheblich höher dotiert. Welches hiervon Emitter bzw. Kollektor ist, hängt von geometrischen und technologischen Parametern, aber auch von den angelegten Spannungen ab. Im Normalfall, dem sogenannten aktiven Betrieb, sind die Spannungen so gepolt, daß die *Emitter-Basis-Strecke in Durchlaßrichtung* und die *Kollektor-Basis-Strecke gesperrt* betrieben werden (Bild 3.4.1). Bei dieser Betriebsart bildet sich eine Verteilung der Minoritätsträger aus, wie sie Bild 3.4.2a) zeigt. Gegenüber den Gleichgewichtsverteilungen (gestrichelt) der Ladungsträger werden zusätzliche La-

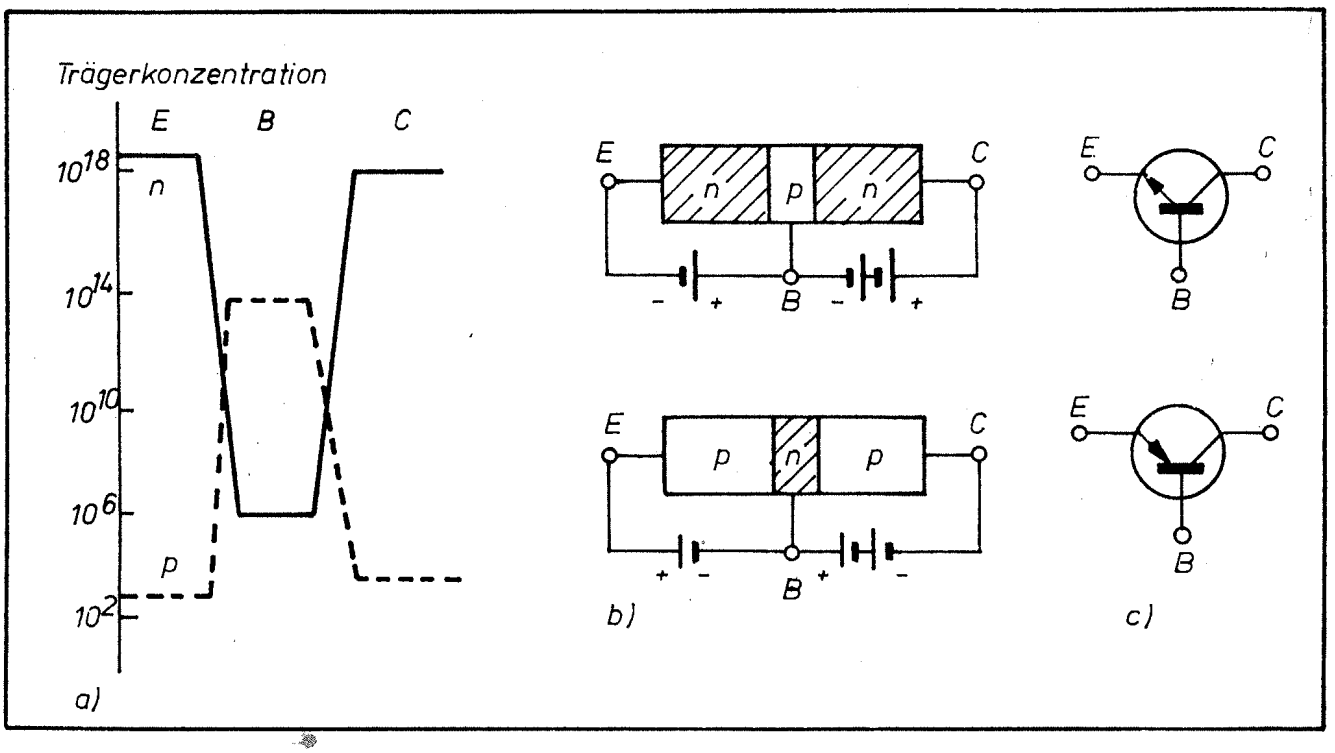


Bild 3.4.1. Grundprinzip der Bipolartransistoren

a) Typischer Dotierungsverlauf für einen *npn*-Transistor
b) Materialfolge und Spannungen an je einem *npn*- und *pnp*-Transistor
c) Schaltbilder für die beiden Typen. Es bedeuten E — Emitter, B — Basis und C — Kollektor

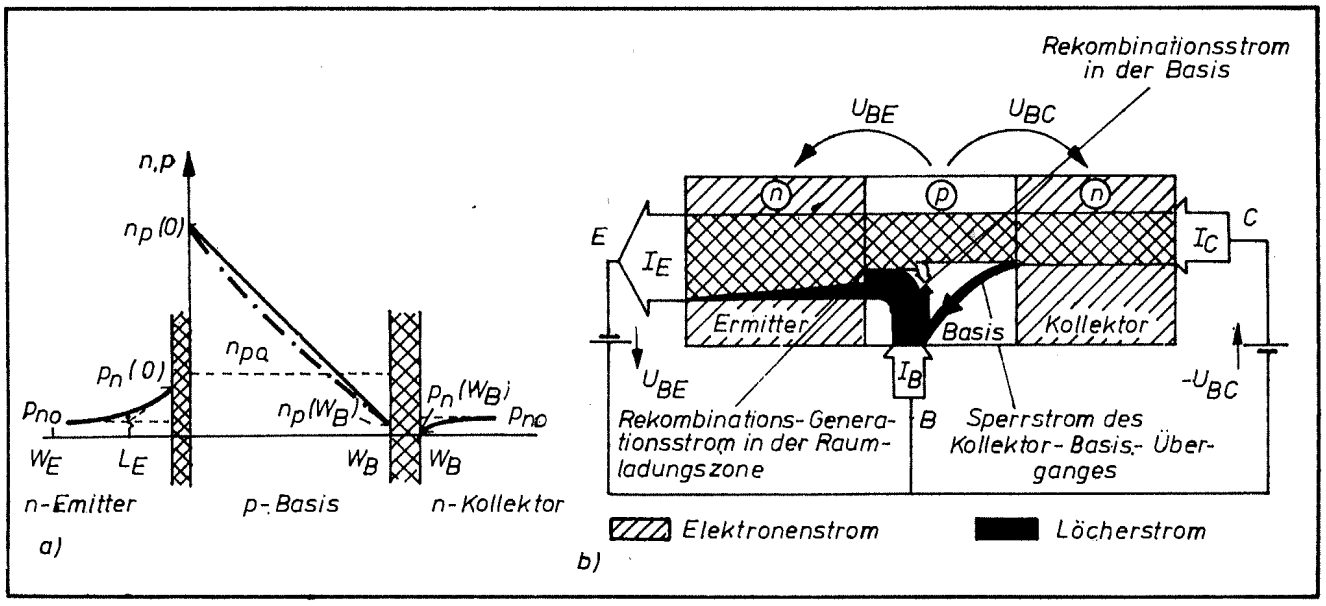


Bild 3.4.2. a) Minoritätsträgerverteilung in der Basis eines *npn*-Diffusionstransistors im aktiven Arbeitsbereich ausgezogene Kurve: bei Vernachlässigung der Rekombination; strichpunktierte Kurve: bei Berücksichtigung der Basisrekombination [$n \sim \sinh(W - d_{RC})/L$]
b) Formale Aufteilung der Löcher- und Elektronenströme in einem *npn*-Transistor

dungsträger im Emitter-Basis-Bereich generiert. Im Kollektor-Basis-Bereich erfolgt dagegen eine starke Ladungsträgerentblößung. Beide Effekte bewirken Raumladungen, die im Bild schraffiert sind.

Wesentlich ist die etwa lineare Ladungsträgerabnahme im Basisgebiet. Ladungsträger, die in der Emitter-Basis-Strecke generiert werden, können auf diese Weise in das Kollektor-Basis-Gebiet diffundieren. Dort werden sie vom vorhandenen Feld beschleunigt und gewinnen so erheblich an Energie. Beim idealen Transistor gehen

alle in der Emitter-Basis-Strecke erzeugten Ladungsträger zum Kollektor über. Dies würde bedeuten, daß kein Strom zur Basis gelangt. Er wird aber in der Tat nur durch Verlustströme (Rekombination) bestimmt und ist in der Regel daher viel kleiner.

Die *Verstärkung* des Bipolartransistors *beruht* darauf, daß eine kleine Spannungsänderung an der Emitter-Basis-Strecke eine große Spannung an der Kollektor-Basis-Strecke beeinflußt. Eine formale Zusammenstellung der Ströme zeigt Bild 3.4.2b).

3.4.2. Mathematisches Modell

Eine einfache Ableitung, das EBERS-MOLL-*Modell*, geht davon aus, daß beide *pn*-Übergänge isoliert seien und durch zusätzliche Stromquellen gemäß Bild 3.4.3 mit den Faktoren A_1 bzw. A_2 über aktive Stromquellen verkoppelt. R_{BB} stellt den Widerstand der Basiszuleitung zu den *pn*-Übergängen dar.

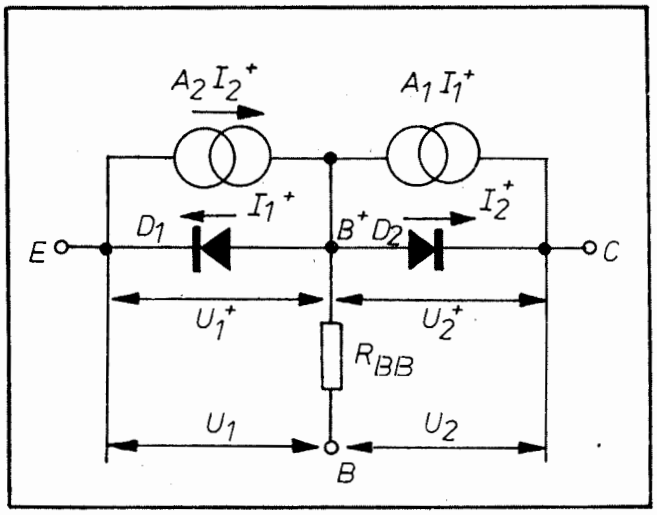


Bild 3.4.3. Ersatzschaltbild eines *npn*-Transistors zur Berechnung der Kennlinien

Für jeden *pn*-Übergang gilt dann Gl. (3.2.5):

$$I_1^+ = I_{sp1}\,(e^{e_0 U_1^+/kT} - 1)\,, \tag{1a}$$

$$I_2^+ = I_{sp2}\,(e^{e_0 U_2^+/kT} - 1)\,. \tag{1b}$$

U_1^+ und U_2^+ sind die Spannungen direkt an den *pn*-Übergängen, also jeweils auf den Punkt B^+ bezogen.

Mit Bild 3.4.3 ergibt sich weiter

$$I_1 = I_1^+ - A_2 I_2^+\,, \tag{2a}$$

$$I_2 = I_2^+ - A_1 I_1^+ \tag{2b}$$

sowie

$$U_1 = U_1^+ + R_{BB}(I_1 + I_2)\,, \tag{3a}$$

$$U_2 = U_2^+ + R_{BB}(I_1 + I_2)\,. \tag{3b}$$

Insgesamt existieren also jetzt sechs Gleichungen für die acht Größen I_1, I_2, I_1^+, I_2^+, U_1, U_2, U_1^+ und U_2^+.

Es ist sinnvoll, anstelle von I_{sp1} und I_{sp2} zwei neue Größen, die sogenannten Restströme einzuführen:

$$I_{1r} = (1 - A_1 A_2)\, I_{sp1}\,, \tag{4a}$$

$$I_{2r} = (1 - A_1 A_2)\, I_{sp2}\,. \tag{4b}$$

Der Reststrom I_{1r} ergibt sich z.B., wenn $I_2 = 0$ und $U^+ = +\infty$ ist. Dies bedeutet, daß der Diode 2 weder Strom zugeführt noch entnommen wird und die Diode 1 stark gesperrt ist. Dann ruft nämlich gemäß Bild 3.4.3 der Strom durch Diode 1 den Strom $A_1 I_1^+$ durch die Diode 2 hervor, und dieser Strom wirkt wieder dem ursprünglichen Sperrstrom gemäß dem Faktor $A_1 A_2$ entgegen.

Aus (1), (2) und (4) folgen die Transistorgleichungen

$$\boxed{I_1 = \frac{I_{1r}}{1 - A_1 A_2} (A_1 - 1 + e^{e_0 U_1^+/kT} - A_1 e^{e_0 U_2^+/kT}),} \tag{5a}$$

$$\boxed{I_2 = \frac{I_{2r}}{1 - A_1 A_2} (A_2 - 1 + e^{e_0 U_2^+/kT} - A_2 e^{e_0 U_1^+/kT}).} \tag{5b}$$

Für einen idealen Transistor gilt noch

$$A_2 I_{2r} = A_1 I_{1r}. \tag{6}$$

Hierdurch ist es möglich, Gl. (5a) bzw. (5b) auf einen auszuwählenden Reststrom zu beziehen.

Das *Gleichungspaar* (5) *gilt für alle möglichen Betriebsbedingungen* des Transistors. Allerdings ist die zuvor eingeführte Betriebsweise Emitter—Basis leitend und Kollektor—Basis gesperrt noch nicht darin enthalten. Deshalb wird diese Gleichung nur in speziellen Fällen angewendet. Es werden daher folgende Vereinfachungen gemacht: Statt 1 und 2 werden Emitter und Kollektor unterschieden. Die Kollektorspannungen werden als unendlich angenommen, so folgt

$$I_E = \frac{A_C}{A_E} \cdot \frac{I_{Cr}}{1 - A_E A_C} [e^{e_0 U_E^+/kT} - (1 - A_E)], \tag{7a}$$

$$I_C = -\frac{A_C I_{Cr}}{1 - A_E A_C} \left[e^{e_0 U_E^+/kT} + \frac{1 - A_C}{A_C} \right]. \tag{7b}$$

Wird hier durch gegenseitiges Einsetzen U_E^+ eliminiert, so folgt

$$\boxed{I_C = A_E I_E + I_{Cr}}. \tag{8}$$

Der Kollektorstrom ändert sich also proportional mit dem Emitterstrom. Mittels des Emitterstroms läßt sich daher der Kollektorstrom steuern. Für den Knoten B^+ muß der erste KIRCHHOFFsche Satz in Form der Knotenregel gelten:

$$I_B + I_C + I_E = 0. \tag{9}$$

Wird in Verbindung mit (8) die Abhängigkeit des Kollektorstroms vom Basisstrom bestimmt, so ist es sinnvoll, eine Größe

$$B_E = \frac{A_E}{1 - A_E} \tag{10}$$

einzuführen. Dann folgt

$$\boxed{I_C = B_E I_B + (1 + B_E) I_{Cr}}. \tag{11}$$

Auch zwischen Basis- und Kollektorstrom besteht ein linearer Zusammenhang. Da A_E stets wenig kleiner als Eins ist, folgt für B_E ein Wert, der groß gegen Eins ist. Von der Basis zum Kollektor besteht also eine erhebliche Stromverstärkung.

Aus (7a) und (7b) läßt sich auch die Spannung berechnen, für die der Basisstrom Null ist. Dann muß nämlich $I_E = -I_C$ gelten, und daraus folgt

$$U_E = \frac{kT}{e_0} \ln\left(1 + \frac{B_E}{B_C}\right) \quad \text{für} \quad I_B = 0\,. \tag{12}$$

Darin ist B_C in Analogie zu B_E gemäß (10) gebildet, und wegen (3) bzw. Bild 3.4.3 gilt in diesem Spezialfall auch $U_E = U_E^+$. Diese Spannung muß je nach dem Verhältnis B_E/B_C einen Bezug zur *Temperaturspannung* $kT/e_0 \approx 26$ mV besitzen. Meist liegt sie um 100 mV.

3.4.3. Die vier Betriebszustände

Das Gleichungspaar (5) gilt ganz allgemein für beliebig gepolte Spannungen U_1^+ und U_2^+ bzw., entsprechend eingesetzt, U_1 und U_2 bzw. U_{BE} und U_{CB}. *Mit den entsprechenden gepolten Spannungen* werden *vier Zustände* unterschieden:

1) *normal aktiv*: Hier ist der Emitter-*pn*-Übergang leitend und der Kollektor-*pn*-Übergang gesperrt. Dieser Fall ist mit Gl. (7) usw. behandelt, und auf ihn wird noch weiter eingegangen.

2) *invers aktiv*: Hier sind Emitter und Kollektor vertauscht, also Emitter-*pn*-Übergang gesperrt, Kollektor-*pn*-Übergang leitend. Bei voll symmetrischem Transistoraufbau unterscheidet dieser Betrieb sich nicht von 1). Bei den realen Transistoren liegt aber eine erhebliche Unsymmetrie vor. Die Faktoren A sind folglich sehr unterschiedlich, und so ergeben sich in diesem Betriebszustand auch stark abweichende Daten und Kennlinien.

3) *gesättigt*: Beide *pn*-Übergänge sind leitend. Der Basisraum wird dadurch von beiden *pn*-Übergängen her mit Ladungsträgern überschwemmt. Eine Steuermöglichkeit besteht hier so gut wie nicht mehr. Dieser Fall tritt nur als Ausnahme in Grenzfällen auf.

4) *gesperrt*: Beide *pn*-Übergänge sind gesperrt. Von keiner Seite werden Ladungsträger injiziert. Ladungsträger, die in der Basisschicht z.B. durch thermische Generation entstehen, werden sofort nach beiden Seiten abgesaugt. Sie bilden die Restströme aller drei Anschlüsse.

3.4.4. Physikalisches Modell

Die zuvor gegebene Ableitung ist zwar allgemein gültig, sie läßt aber kaum eine Beziehung zum eigentlichen physikalischen Geschehen erkennen. Deshalb besitzt ein zweites Modell erhebliche Bedeutung, das u. a. in [M 20, P 8, P 23, S 27] beschrieben ist. Es gilt allerdings *nur für den aktiven Betrieb*. Mit den Daten

D_n Diffusionskonstante der Elektronen, für Si etwa 30 cm²/s,
e_0 Elektronenladung $\approx 1{,}6 \cdot 10^{-19}$ As,
L_n Diffusionslänge der Elektronen, für Si Größenordnung $^1/_{10}$ mm,
n_{Bo} dotierte Elektronendichte im Basisgebiet,
U_T Temperaturspannung $kT/e_0 \approx 25$ mV bei Zimmertemperatur,
W Basisweite

gilt dann

$$I_C = e_0 \frac{D_n}{L_n} n_{Bo} \left[-\left(\exp\left(\frac{U_{EB}}{U_T}\right) - 1 \right) \frac{1}{\sinh\,(W/L_n)} + \left(\exp\left(\frac{U_{CB}}{U_T}\right) - 1 \right) \cdot \coth\,(W/L_n) \right]. \tag{13}$$

Die hyperbolischen Funktionen entsprechen den Faktoren A im vorangegangenen Modell. Das Verhältnis von Basisweite zur Diffusionslänge bestimmt damit wesentlich den vom Emitter zum Kollektor gelangenden Stromanteil. Weiter ist deutlich zu erkennen, in welchem Umfang die Kollektorspannung den Kollektorstrom beeinflußt. Dieser Einfluß ist bei großen Kollektorspannungen sehr gering. So wird Gl. (7b) wieder erhalten. Ihr Analogon in Form von Gl. (13) lautet

$$I_C = e_0 \frac{D_n}{L_n} n_{Bo} \left(\exp\left(\frac{U_{EB}}{U_T}\right) - 1 \right) \frac{1}{\sinh\,(W/L_n)}. \tag{14}$$

3.4.5. Effektive Basisweite und punch-through

Durch die Kollektorspannung entsteht gemäß Bild 3.4.2 eine trägerentleerte Raumladungszone im Kollektor-Basis-Gebiet. Da der Emitter im allgemeinen niedriger als der Kollektor dotiert ist, muß im wesentlichen die Basis die Kollektorspannung aufnehmen (vgl. Bild 3.4.2e)). Hierdurch wird aber der wirksame Bereich der *Basis* W_{eff} geringer. Er ist also eine *Funktion der Kollektorspannung*:

$$W_{eff} = W - d_{RC}\,, \tag{15}$$

wobei etwa gilt

$$d_{RC} = \sqrt{U_C \varrho_n}\,. \tag{16}$$

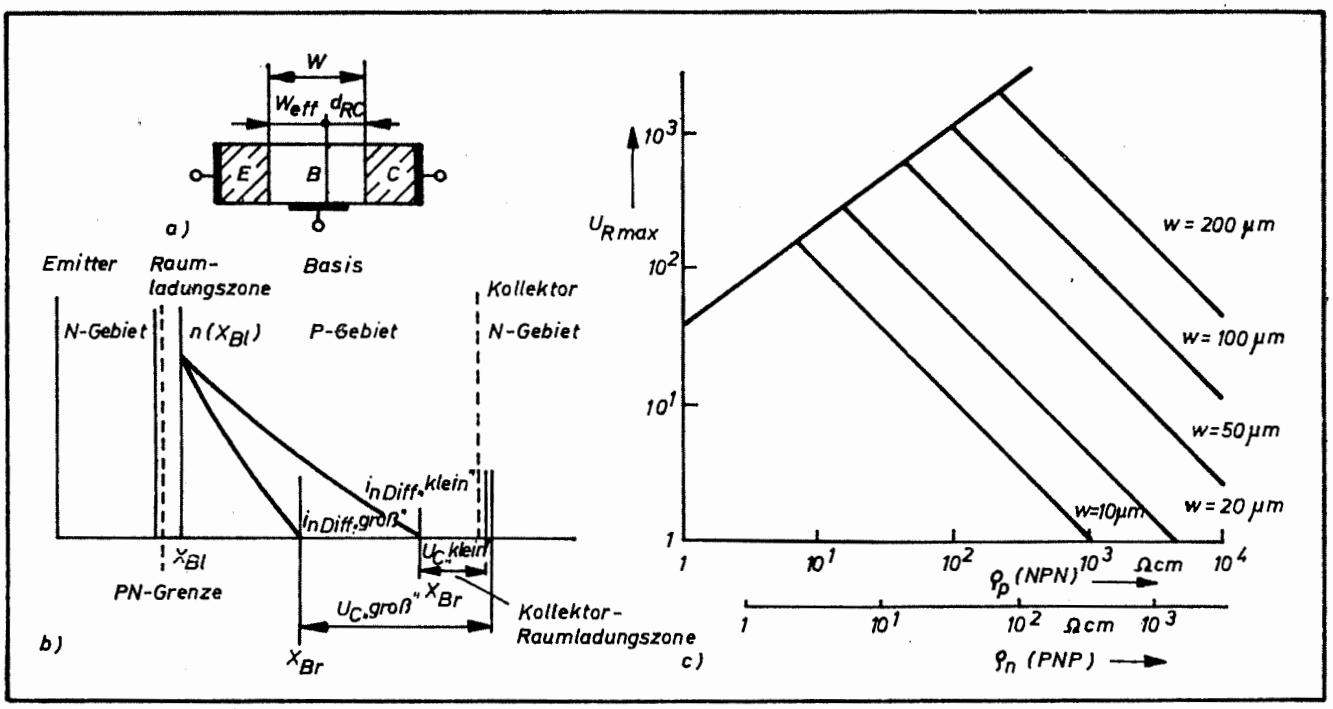


Bild 3.4.4. Zum Einfluß der Kollektorspannung auf die wirksame Basisweite

a) Schematische Darstellung der Ausbreitung der Raumladezone d_{RC}

b) Änderung des Konzentrationsverlaufs im Basisgebiet und damit Änderung des Diffusionsstromes in der Basis bei veränderter Kollektorspannung

c) Begrenzung der Sperrfähigkeit eines Transistors

Dies ist schematisch im Bild 3.4.4 angedeutet. Mit Steigerung der Kollektorspannung tritt also der Fall ein, wo die *Raumladungszone* d_{RC} *vollständig die Basisweite* W ausfüllt. Dann schlägt der Transistor durch (*punch through*), und der Kollektorstrom steigt steil an. Hierdurch ist mit der Basisweite und weiteren Größen die maximale Betriebsspannung definiert:

$$U_p = \frac{W^2}{2\varepsilon_0 \varepsilon_i \mu_n \varrho_n}. \tag{17}$$

Eine Auswertung dieser Formel zeigt Bild 3.4.4c). Bezüglich Basisweite und Dotierung des Basisgebietes müssen folglich Kompromisse hinsichtlich A (möglichst nahe bei Eins) und der zulässigen Kollektorspannung (möglichst groß) gefunden werden.

Genau genommen gilt dies nur für einen Transistor mit *homogen dotierten Gebieten* und Relationen der Dotierung, wie sie Bild 3.4.1a) zeigt. Zwei Methoden lassen *bessere Werte* erreichen:

- Die Basis wird zum Emitter hin höher dotiert. Dann ist es gemäß (16) schwieriger, daß die Raumladezone bis zum Emitter reicht.
- Der Kollektor wird niedriger dotiert, dann entsteht auch hier ein Spannungsabfall (vgl. 3.2.8.).

3.4.6. Kennlinien

Für einen normalen *pn*-Übergang gilt Gl. (3.2.5). Der Verlauf ist in Bild 3.2.2a) dargestellt. Etwas vereinfacht ergibt sich die stark ausgezogene Kurve in Bild 3.4.5a). Die Auswertung von Gl. (8) führt dann zu den weiteren Kurvenscharen im dritten Quadranten. Infolge des Spannungsabfalls an dem Basiswiderstand gemäß (3) entsteht eine Verlagerung der Kurven gemäß Bild 3.4.5b). Sie zeigt den Zusammenhang zwischen Kollektorspannung U_C (nicht mehr U_C^+) und dem Kollektorstrom I_C. Diese Kennlinie ist nur noch bezüglich Strom und Spannung in den ersten Quadranten zu drehen, und schon existiert das Ausgangskennlinienfeld gemäß Bild 3.4.5c). Bild 3.4.5 zeigt schließlich noch die Eingangskennlinie der steuernden Emitterdiode, ebenfalls

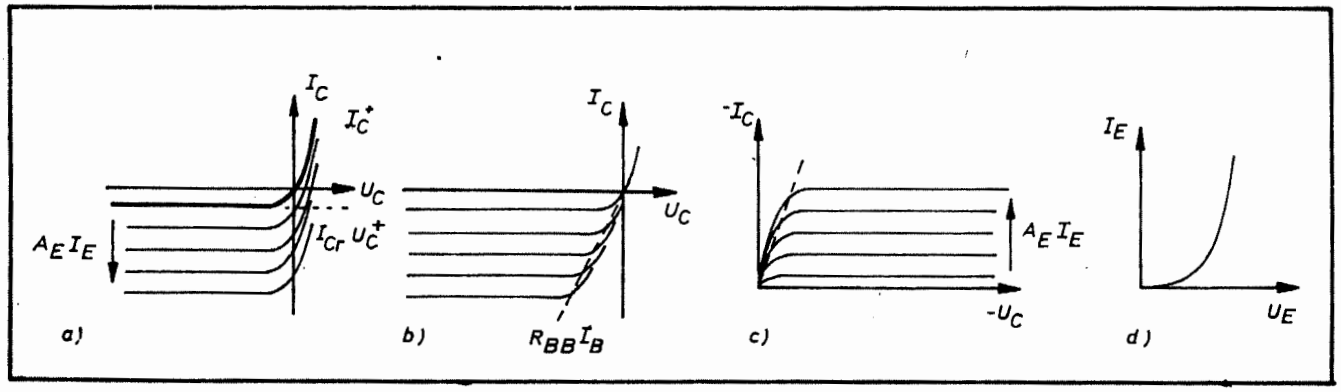


Bild 3.4.5. Zur Entstehung des Kennlinienfeldes eines Transistors

a) Unter Auswertung von Gl. (8) entsteht aus dem Strom einer Halbleiterdiode (dick ausgezogen) die Kennlinienschar im 3. Quadranten

b) Durch den Spannungsabfall verschieben sich die Kennlinien so, daß sie praktisch nur noch im 3. Quadranten liegen. In diesem Teilbild ist U_C statt U_C^+ im Teilbild a) als Abszisse gewählt

c) Wird der Kollektor als Ausgangsgröße betrachtet, so ist es sinnvoll, negative Spannung und negativen Strom zu wählen. Dabei entsteht eine Ausgangskennlinie, die stark jener der Pentode gemäß z.B. Bild 2.3.4 und jener des FET, z.B. Bild 3.3.4, ähnelt

d) Die zugehörige Eingangskennlinie entspricht der Durchlaßkennlinie der Emitterdiode

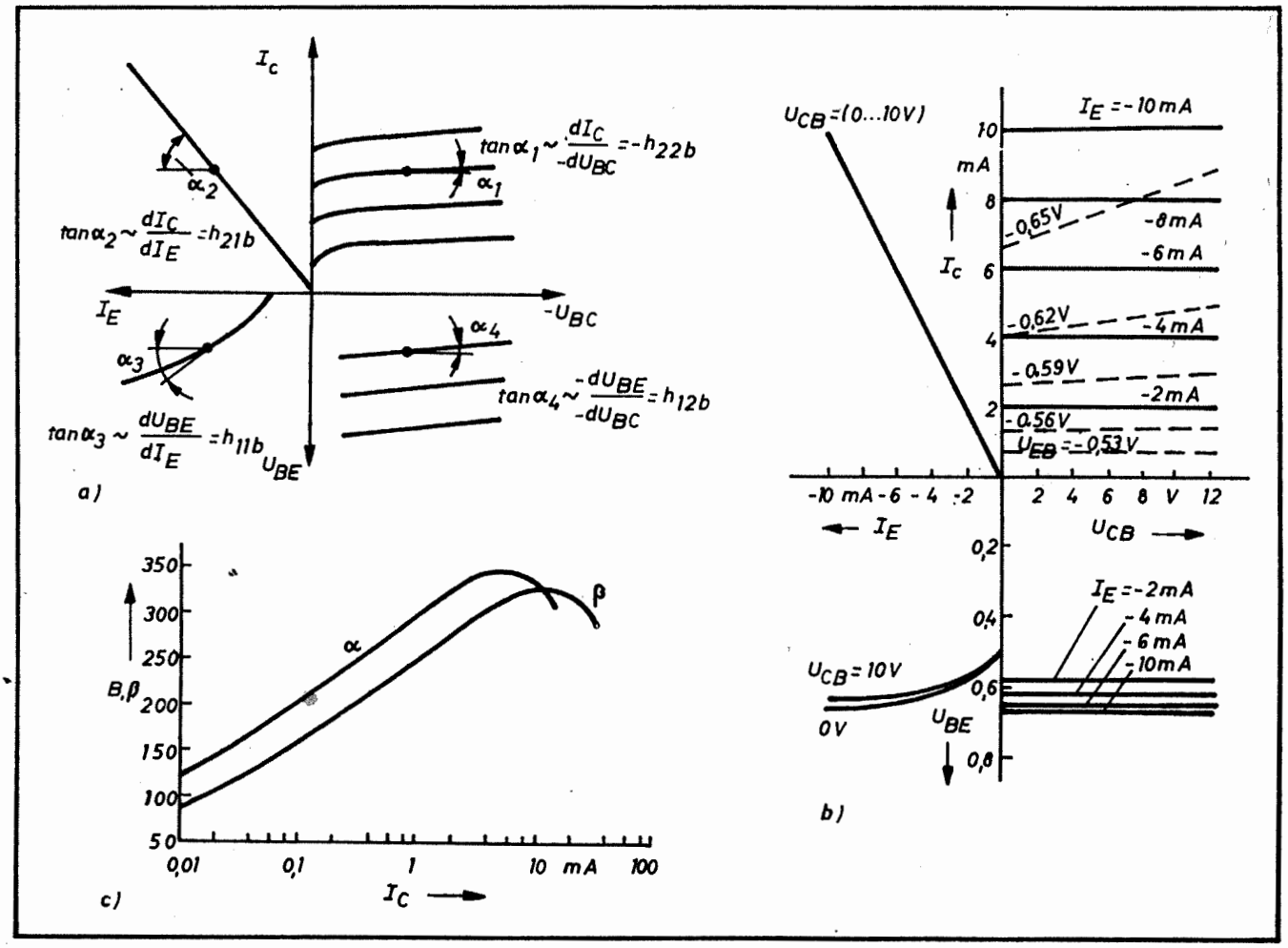


Bild 3.4.6. Vollständiges Kennlinienfeld eines Bipolartransistors in Basisschaltung

a) Schematisch mit Übergang zu den Hybridparametern
b) Für einen SC 206
c) Stromverstärkungsfaktoren für Gleichstrom B und Wechselstrom β

als *pn*-Übergang. Der Bipolartransistor benötigt im Gegensatz zu einem FET Eingangsleistung und dazu noch an einer nichtlinearen Eingangskennlinie. Hierdurch sind die Zusammenhänge zwischen Eingangs- und Ausgangsgrößen beim Bipolartransistor wesentlich komplexer als beim FET. Deshalb ist es üblich, das *vollständige Kennlinienfeld in vier unterschiedlich belegten Quadranten* darzustellen. Dies zeigt schematisch und für einen konkreten Transistor Bild 3.4.6a) und b). In Bild 3.4.6c) ist ein typischer Verlauf für die Gleichstromverstärkung B und die Wechselstromverstärkung β dargestellt.

3.4.7. Kleinsignalparameter

Dem üblichen Vierquadranten-Kennlinienfeld besonders angepaßt ist die *Hybridmatrix* (vgl. 1.1.5.). Da sie nur linear ist, gilt sie auch nur für kleine Änderungen von Wechselspannungen und Wechselströmen um einen gleichstrommäßig festgelegten Arbeitspunkt. Dies ist bereits in Bild 3.4.6a) angedeutet. Dementsprechend lautet das Gleichungssystem hier

$$\boxed{\begin{aligned} u_E &= h_{11} i_E + h_{12} u_C \,, \\ i_C &= h_{21} i_E + h_{22} u_C \,. \end{aligned}} \qquad (18)$$

Es bedeuten

h_{11}	Eingangswiderstand	Steigung im 3. Quadranten,
h_{12}	Rückwirkung der Ausgangsspannung	Steigung im 4. Quadranten,
h_{21}	Stromverstärkung	Steigung im 2. Quadranten,
h_{22}	Ausgangsleitwert	Steigung im 1. Quadranten.

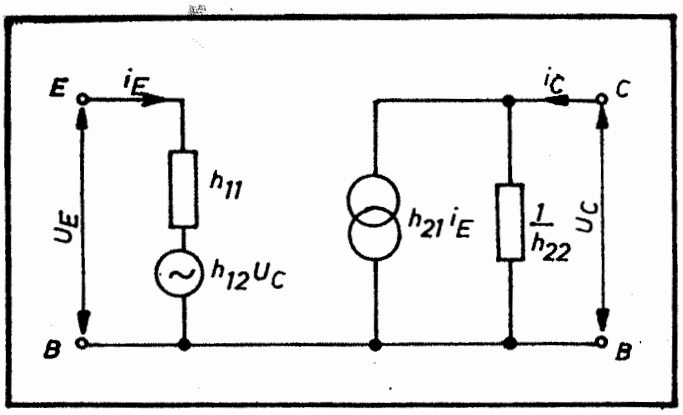


Bild 3.4.7. Ersatzschaltung eines Transistors gemäß der Hybridmatrix von Gl. (18)

Die *Stromverstärkung* h_{21} wird häufig mit $-\alpha$ bezeichnet. Sie folgt unmittelbar aus dem Bild 3.4.3, aus den A_1 und A_2. Gl. (18) führt zur Ersatzschaltung von Bild 3.4.7. Der Zusammenhang mit der Ersatzschaltung von Bild 3.4.3 wird erst später erkennbar. Hierzu wird im Abschnitt 5.6.1. auf die Emitter- und Kollektorschaltung eines Transistors einzugehen sein.

4. Technologiegrundlagen

Silicium ist das Hauptmaterial der Halbleiterbauelemente. Deshalb werden im folgenden vorrangig hierauf alle Betrachtungen bezogen. Das Rohmaterial für Si ist kein Problem. Si ist das zweithäufigste Element der Erdrinde, die bis etwa 16 km Tiefe zu 90% aus Siliciumdioxid und Silikaten besteht. 28% der Erdkruste ist Si.

Der Name Silicium wurde durch BERZELIUS 1822 bei der ersten Synthese geprägt. Er stammt vom lateinischen Wort silex, das dem deutschen Wort Kiesel entspricht. Natürliche kristalline Formen von SiO_2 sind u. a. Rauchquarz (Rauchtopas), Goldtopas, Marion (schwarz), Cristin (gelb), Rosenquarz (rosa), Chrysopas (grün), Amethyst (violett) und der Bergkristall (durchsichtig). Sauberer Dünensand ist ebenfalls reines SiO_2. Der Weltverbrauch von Si für Halbleiter lag 1970 bei 100 t, 1973 bei 1000 und 1980 bei 2000 t. Das weitere Wachstum könnte noch einmal steil durch Sonnenzellen zunehmen.

Neben Silicium hat eine gewisse Bedeutung das *Germanium*. Es war zunächst das Grundmaterial aller Halbleiter, verlor dann aber an Bedeutung, als die Siliciumtechnologie, insbesondere als ab 1960 die Epitaxie-Planartechnik gut beherrschbar wurde und so bessere Bauelemente und vor allem die integrierte Schaltkreistechnik ermöglichte. Ab 1969 ging die Germaniumproduktion sogar absolut im Weltmaßstab zurück. Es wurden nur noch NF-Transistoren für höchste Ströme und spezielle HF-Transistoren produziert. Seit etwa 1976 steigt wieder die Produktion. Die Ursache ist die breitere Anwendung von Infrarotbauelementen. 1985 wird wieder mit einem Bedarf von ca. 120 t (der bereits 1965 existierte) gerechnet, davon allein 70 t für Infrarotbauelemente. Dieser und der weiter steigende Bedarf kann abgedeckt werden, da die Weltreserve auf ca. 2500 t geschätzt wird. Ge kommt in vielen Erzen und Gesteinen in kleinen Beimengungen vor. Hauptsächlich wird es heute aus den Schmelzrückständen der Zinkproduktion gewonnen.

Das drittwichtigste Halbleitermaterial ist heute und noch stärker in der Zukunft *GaAs*. Es hat vor allem für Höchstfrequenz- und Kurzzeitbauelemente Bedeutung. Es besitzt einen stark wachsenden Produktionsanteil. Alle weiteren Halbleiter werden in geringem Umfang benötigt.

Die *Halbleitertechnologie* kann in *drei Hauptgruppen* geteilt werden

- Die Herstellung der einkristallinen Scheiben. Hierzu gehören Herstellung des Reinstsiliciums, der Einkristalle und deren Zerteilung in die Scheiben (Wafer). Dieser Teil gehört zum *Zyklus 0*.
- Die Scheibenprozesse mit den Teilprozessen der Lithografie, Epitaxie, Diffusion, Oxydation, Bedampfung und Ionenimplantation (*Zyklus I*).
- Chipprozesse bis zum fertigen Bauelement, beginnend mit der Zerteilung der Wafer in die Chips, deren Kontaktierung, Kapselung und Prüfung (*Zyklus II*).

Diesen Prozessen *vor- und nebengeordnet* ist

- der *Schaltkreisentwurf*, der von den Zielen des Bauelementes bzw. der Schaltung ausgeht und über intelligenzintensive Prozesse, die heute von Rechnern gestützt werden, zu den Masken für die Lithografieprozesse führt (s. 4.3.). Er wird zum Zyklus 0 gerechnet.

Einen anschaulichen Überblick zu den Prozessen bis zum fertigen Bauelement gibt Bild 4.0.1.

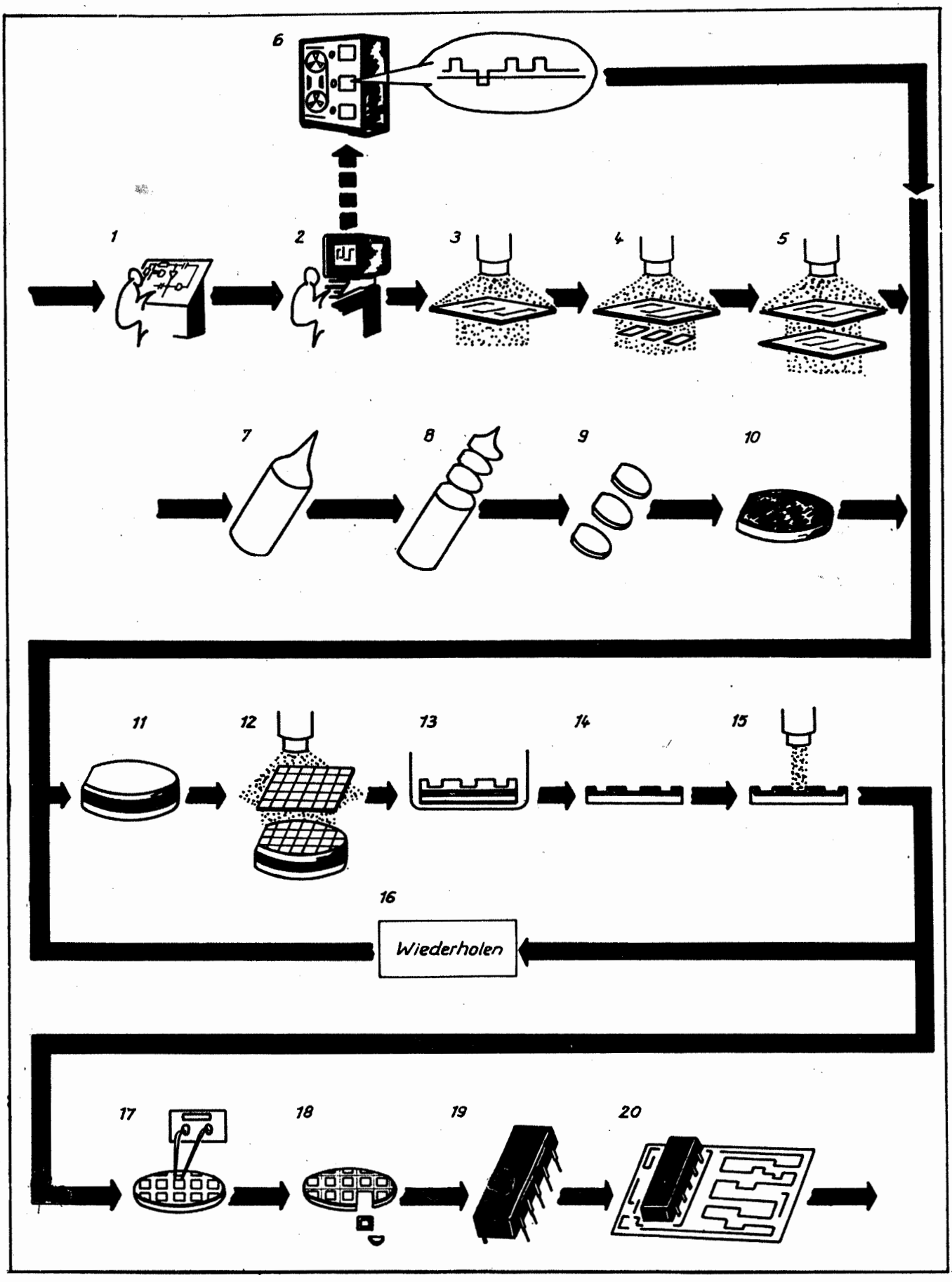


Bild 4.0.1

4.1. Reinstsilicium und Einkristalle

4.1.1. Herstellung des Rohsiliciums

Bei der Siliciumherstellung wird meist von reinem Quarz ausgegangen. Er wird in Gegenwart von Koks im *Lichtbogenofen* reduziert:

$$SiO_2 + 2\,C \rightarrow Si + 2\,CO\,. \tag{1}$$

Dann wird das meist zu 98 bis 99% reine Silicium durch trockene, aktivierte Salzsäure umgesetzt:

$$Si + 3\,HCl \xrightarrow{200\ldots400\,^\circ C} SiHCl_3 + H_2\,. \tag{2}$$

Hierbei fällt in gewissem Umfang auch $SiCl_4$ an (Bild 4.1.1a)). Das bei Zimmertemperatur flüssige *Trichlorsilan* wird einer mehrfachen fraktionierten Destillation unterzogen, dabei werden vor allem Metallchloride abgetrennt. Schließlich wird es 1:10 bis 1:100 mit Wasserstoff verdünnt und bei 1000 bis 1100 °C thermisch zersetzt und reduziert. Hierzu durchströmt das Gemisch ein langes Quarzrohr, in dessen Achse sich dünne, lange, glühende Siliciumseelen befinden (Bild 4.1.1b)):

$$4\,SiHCl_3 \rightarrow Si + 3\,SiCl_4 + H_2\,, \tag{3}$$

$$SiHCl_3 + H_2 \rightarrow Si + 3\,HCl\,. \tag{4}$$

Die Aufwachsgeschwindigkeit auf der Siliciumseele ist gering. In 24 Stunden werden ca. 20 mm erreicht. Insgesamt sind polikristalline Stäbe bis zu etwa 80 cm Länge und 8 cm Durchmesser erreichbar. Bei diesem Prozeß wird das stark störende Bor entfernt. Beim späteren Zonenschmelzen dampft der ebenfalls stark störende Phosphor ab.

◀

Bild 4.0.1. Schematische Darstellung der Technologien zum integrierten Schaltkreis. Es bedeuten

1 Schaltkreisentwurf, theoretisch und experimentell
2 Rechnergestützte Umsetzung des Entwurfs (interaktiv)
3 Herstellung der Ausgangsmasken und reticle
4 Wiederholung der Masken, verkleinert (step and repeat)
5 Erzeugung der Arbeitsmasken
6 Weg der direkten elektronischen (maskenfreien) Strukturierung der Wafer
7 Herstellung der Einkristalle
8 Trennung des Einkristalls in Wafer
9 Bearbeitung der Wafer bez. Oberflächenqualität
10 Oxydation des Wafers
11 Beschichtung des Wafers mit Fotolack
12 Belichtung des Fotolackes mit UV
13 Entwicklung des belichteten Fotolackes
14 Ätzung des Wafers
15 Diffusion, Metallisierung, Epitaxie, Oxydation oder Ionenimplantation
16 Wiederholung der Prozesse 11 bis 15 mit unterschiedlichen Masken
17 Prüfung des strukturierten Wafers
18 Zertrennung des Wafers in die Chips
19 Fertiges Bauelement
20 Prüfung des Bauelementes

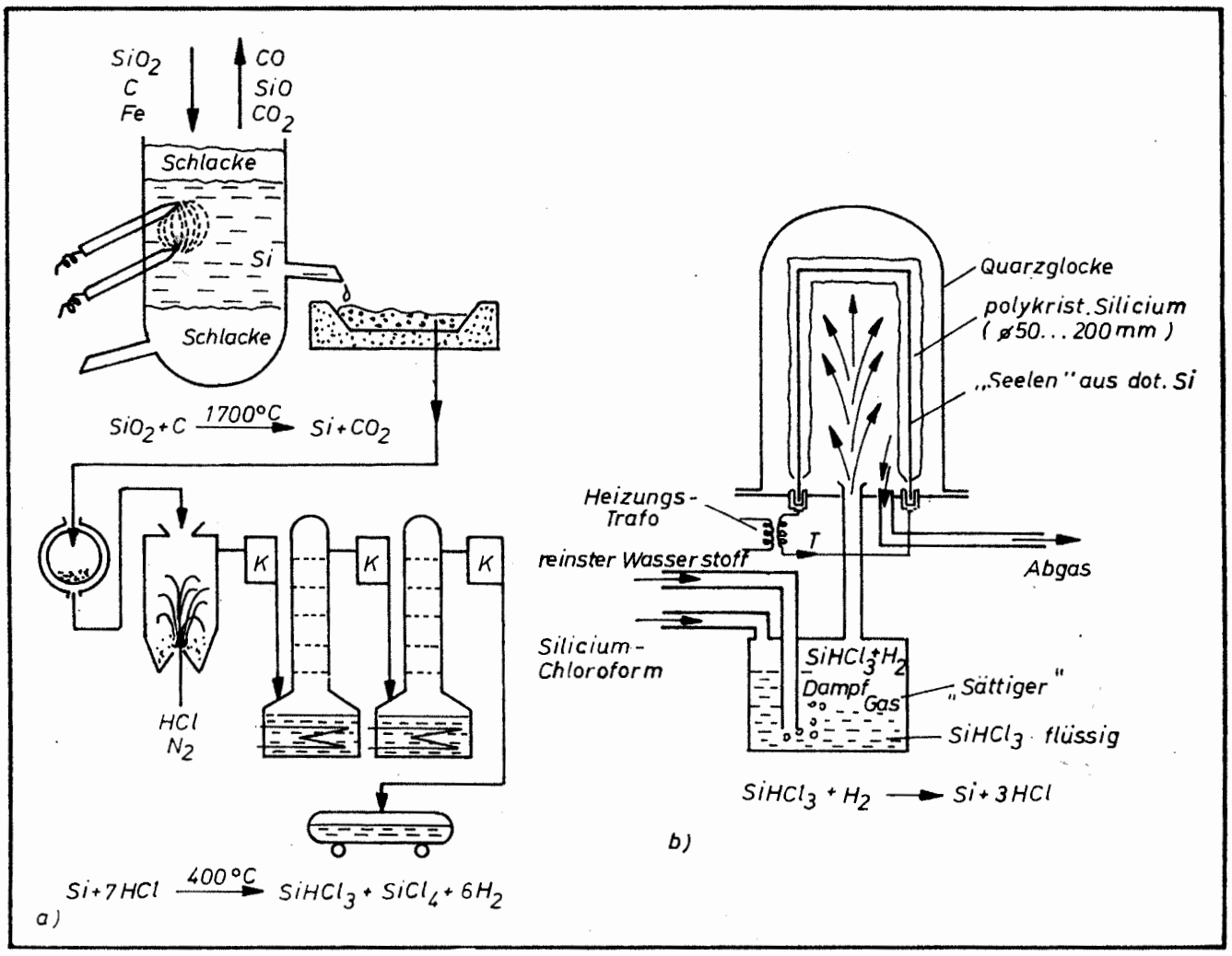


Bild 4.1.1. a) Erzeugung des $SiHCl_3$
b) Herstellung polykristalliner Siliciumstäbe

4.1.2. Physikalische Reinigung

Eine Verunreinigung geringer Konzentration hat meist in der flüssigen und festen Phase des Lösungsmittels (hier Si) unterschiedliche Löslichkeit. Das Verhältnis beider Löslichkeiten heißt *Verteilungskoeffizient k*. Einige Werte gibt Tab. 4.1.1 an. Ist k kleiner als Eins, so ist die Löslichkeit in der flüssigen Phase größer. Beim langsamen Erstarren der Flüssigkeit ist daher die Reinheit des zuerst erstarrten Materials höher als die des zuletzt erstarrten. Ist k größer als Eins, so liegen die Verhältnisse umgekehrt. Außer für P und B ist k in Silicium sogar wesentlich kleiner als Eins. Bild Bild 4.1.2a) gibt das Prinzip der vier wichtigsten hierauf beruhenden Verfahren wieder. Beim BRIDGEMAN- und CZOCHRALSKI-*Verfahren* (beide sind eigentlich Kristallzüchtungsverfahren) wird das gereinigte Material aus der Schmelze gewonnen. Sie

Tabelle 4.1.1. Verteilungskonstanten k beim Schmelzpunkt des Siliciums in Silicium und Diffusionskonstanten D in cm^2/s bei 1300 °C für einige Elemente

Material	B	Al	P	Ga	As	Sb	Bi	In
k	0,8	0,002	0,35	0,008	0,3	0,023	0,0007	0,0004
$D \cdot 10^{11}$	1,6	6,1	1,1	2	0,13	0,125	0,14	0,51

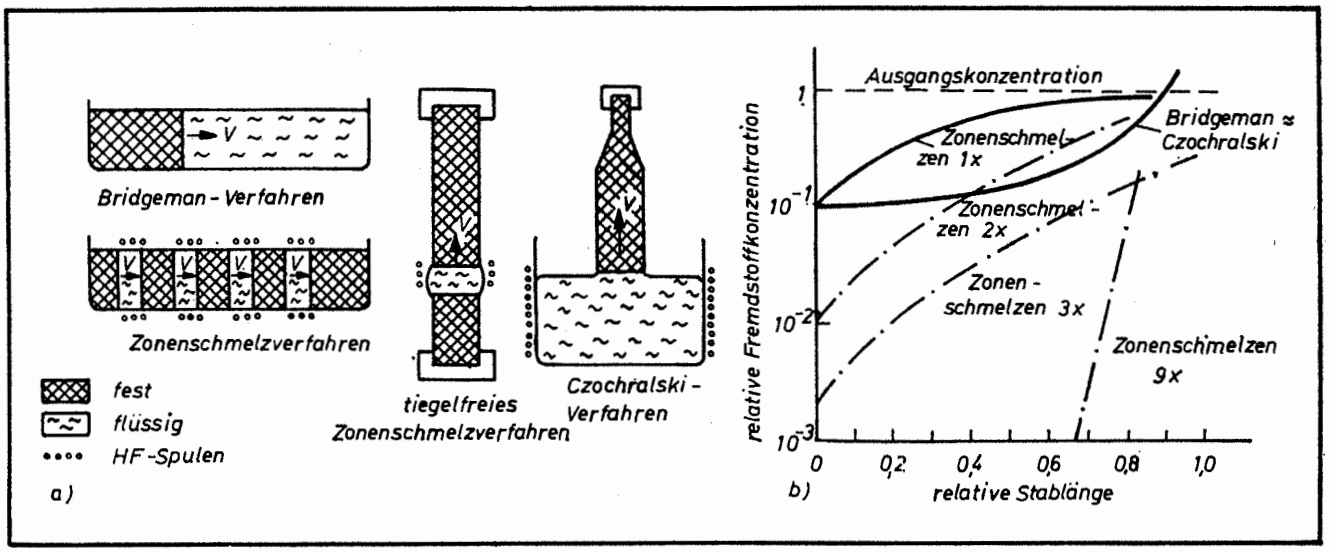


Bild 4.1.2. Physikalische Reinigung

a) Vier mögliche Verfahren

b) Wirkungsgrad physikalischer Reinigungsverfahren für einen Verteilungskoeffizienten von etwa $k = 0{,}1$

besitzen beide einen sehr guten Reinigungseffekt, da in der Schmelze ein weitreichender Konzentrationsausgleich erfolgen kann. Eine Wiederholung dieser Verfahren zur weiteren Verbesserung kann aber erst dann erfolgen, wenn das unreine Stabende vor dem erneuten Schmelzen abgetrennt wurde. Dieser Nachteil wird beim *Zonenschmelzverfahren* vermieden. Sein Reinigungseffekt ist zwar für einen einmaligen Durchgang entsprechend der geringeren Flüssigkeitsmenge kleiner, aber die unmittelbare Wiederholbarkeit wiegt dies auf. Mit jeder Wiederholung sammelt dann die Schmelzzone Fremdstoffe auf und transportiert sie zum Stabende. Einen Vergleich beider Prinzipien gibt Bild 4.1.2b). Die günstige Geschwindigkeit der Kristallisationsfront liegt bei allen Verfahren bei einigen mm pro Minute. Das *tiegelfreie Zonenschmelzverfahren* hat schließlich noch den Vorteil, daß keine Verunreinigungen durch das Gefäß entstehen können. Bei ihm kann jedoch immer jeweils nur eine Schmelzzone von unten nach oben durch den Stab bewegt werden.

Alle Verfahren werden entweder unter einer Schutzgasatmosphäre oder im Vakuum durchgeführt. Vakuum hat den zusätzlichen Vorteil, daß gewisse Störkomponenten verdampfen. Mit den genannten Verfahren werden *Fremdstoffkonzentrationen* bis unter 10^{-12} erreicht.

4.1.3. Herstellung von Einkristallen

Für die Halbleitertechnik werden möglichst ideale Kristalle benötigt. Sie müssen deshalb nicht nur frei von Fremdstoffen sein, es muß auch eine *vollkommen regelmäßige Anordnung der Atome* vorliegen. Dies bedeutet die Forderung nach einem idealen Einkristall. In der Wirklichkeit liegen jedoch Realkristalle vor. Sie sind durch Kristallfehler gekennzeichnet. Gewöhnlich werden *drei Typen von Fehlstellen* unterschieden:

- chemische, die durch *Fremdatome* hervorgerufen werden,
- strukturelle, dazu gehören alle *Abweichungen* der Atome *von der periodischen Lage*. Hierzu gibt es mehrere Untertypen, die im Prinzip Bild 4.1.3 zeigt. Sie gehen auf die Anzahl der Dimensionen zurück, die atomar verändert sind:
  - 0-Dimensionalität liegt z.B. vor, wenn Gitterplätze des Kristalls unbesetzt sind (Leerstellen) oder wenn Zwischengitterplätze von Atomen besetzt sind.

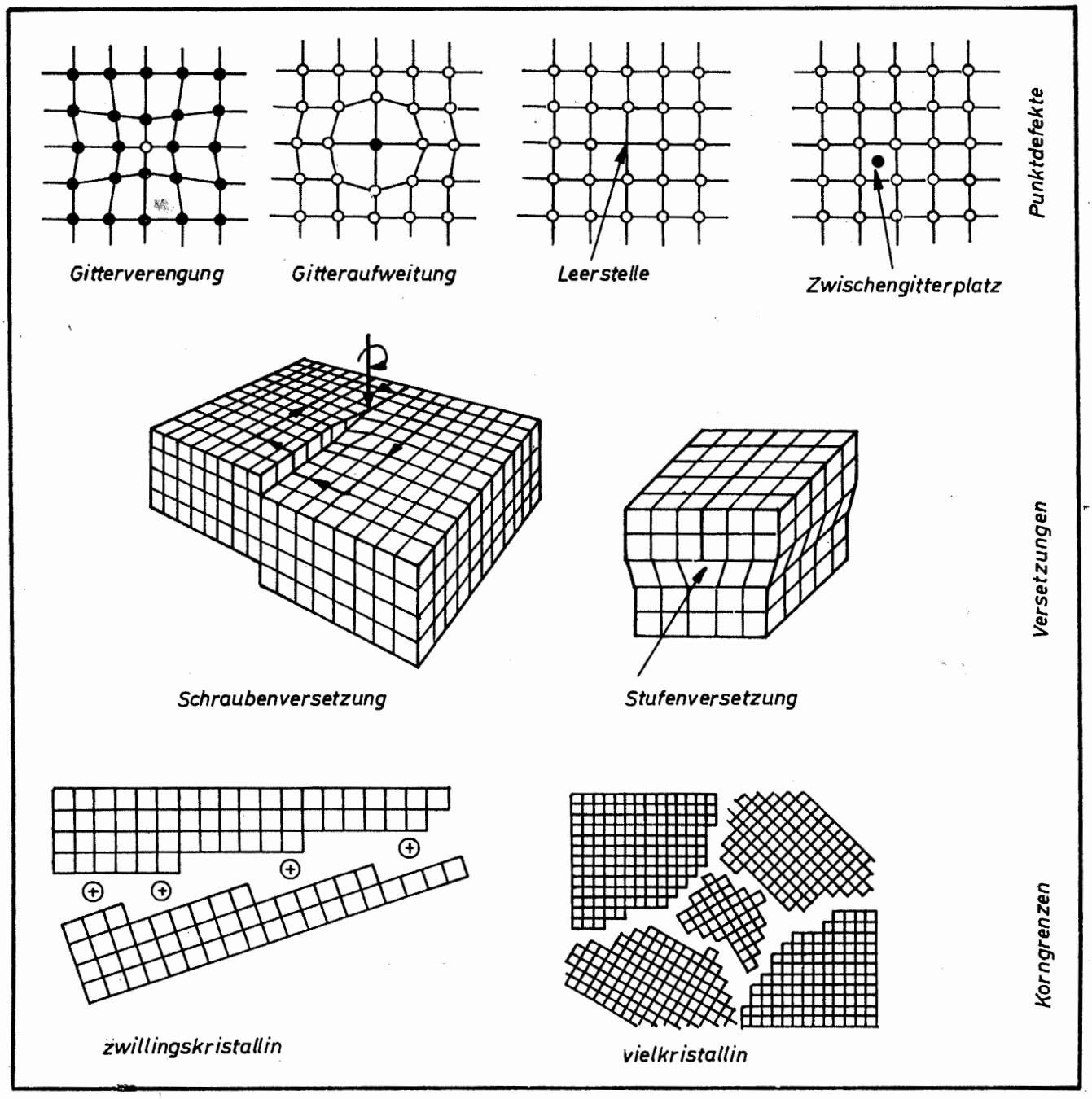


Bild 4.1.3. Zusammenstellung wichtiger Kristallfehler

Obere Zeile: Punktdefekte: Gitterverzerrungen durch Fremdstoffe, links Verengung, Mitte links Aufweitung und Wanderung von Atomen, Mitte rechts Leerstellen, indem Atome den Gitterverband verlassen haben, rechts Atome auf Zwischengitterplätzen. Für Leerstellen sind etwa Energien um 1 bis 2 eV notwendig; Zwischengitterplätze fordern die zwei- bis vierfache Energie

Mittlere Zeile: Versetzungen: die links dargestellte Schraubenversetzung trägt ihren Namen daher, daß man um die Störlinie herum die Gitterebenen wie eine Wendeltreppe schraubenförmig verbinden kann. Bei der Stufenversetzung befindet sich zu beiden Seiten der Gleitebene eine unterschiedliche Zahl von Gitterebenen

Untere Zeile: Korngrenzen: zwei zweidimensionale Gitterfehler, links die Zwillingsbildung, bei der zwei Kristalle mit unterschiedlicher Orientierung (hier 28° Kornwinkelgrenze) zusammengewachsen sind. Rechts mehrere Kristallite unterschiedlicher Orientierung, durch Korngrenzen getrennt

- 1-dimensionale Fehlstellen sind im wesentlichen die Schrauben- und Stufenversetzungen.
- 2-dimensionale gehen u. a. auf Korngrenzen, Zwillingsbildung und Stapelfehler zurück.

- *elektrische oder elektronische,* die durch geänderte Ladungsverteilung im Kristall bedingt sind. Hierzu gehören Anhäufungen von Fremdstoffen.

Alle diese *Störstellen müssen* bei der Herstellung von Halbleiterkristallen weitgehend, ja möglichst vollständig *vermieden werden*, denn sie erhöhen die Störstellendichte, d. h. Trägerdichte im Material. Damit eine definierte, auch geringe Dotierung möglich ist, muß das Grundmaterial eine erheblich geringere Störstellendichte aufweisen. Besonders störend sind deshalb Fehlstellen höherer Dimensionalität. Hierzu dienen besondere Kristallziehverfahren. Ihr Prinzip beruht im wesentlichen darauf, daß von einem störstellenfreien Kristall, dem *Impfkeim* ausgegangen wird. An ihn wächst in definierter Weise das weitere Material an. Er muß dazu genau orientiert werden. Deshalb ist er meist sehr klein. Zuweilen wird bis auf Bruchteile von Millimetern abgeätzt. Bei den Ziehverfahren wird aber auch zunächst der Durchmesser soweit vermindert (≈ 3 mm), bis alle Störstellen herausgewachsen sind. Erst dann wird wieder zum Dickenwachstum übergegangen. Beim *Wachstumsprozeß* zum eigentlichen Einkristall sind sehr *genau spezifische Bedingungen* wie Temperaturgefälle, -konstanz, Wachstumsgeschwindigkeit usw. einzuhalten. Im Prinzip sind hierzu die Reinigungsverfahren von Bild 4.1.2a) in abgewandelter Form geeignet. Silicium erschwert die Ziehprozesse infolge seiner hohen Aktivität gegenüber allen Tiegeln. Unter anderem wird bei Quarz- in hohem Maße Sauerstoff und bei Graphittiegeln Kohlenstoff aufgenommen. Deshalb werden vor allem die vier tiegelfreien Verfahren von Bild 4.1.4

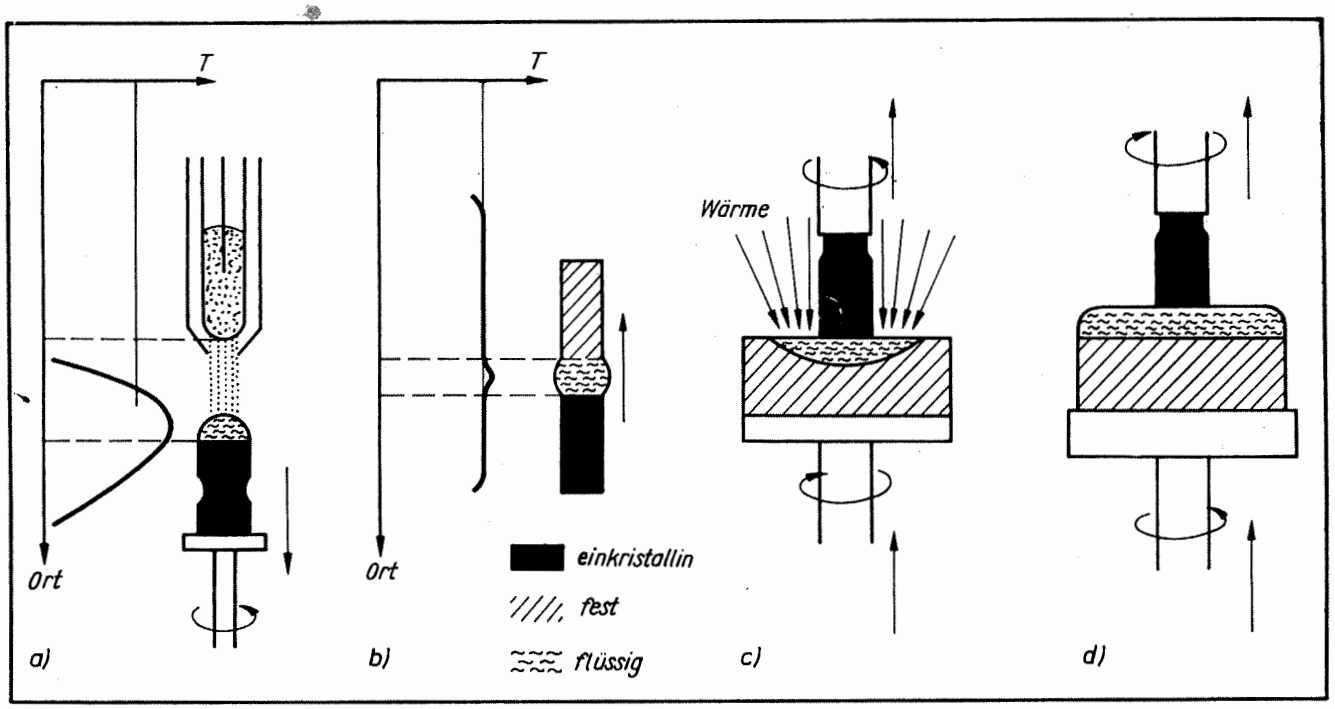


Bild 4.1.4. Schematische Darstellung von tiegelfreien Varianten zur Si-Einkristallzüchtung. Sie erfolgen im Vakuum oder in sauberer Schutzgasatmosphäre. Heute werden ca. 80% der Si-Einkristalle mit dem CZOCHRALSKI-Verfahren hergestellt, während der Rest für hochreines Si mit Zonenfloaten hergestellt wird

a) VENEUIL-Verfahren
b) Zonenfloaten mit konstanter Grundtemperatur über die Länge des Kristalls
c) Züchtung aus Si-Rohling, der zugleich als Tiegel dient
d) Pedestal-Verfahren, bei dem der Kristall von einem Rohling mit einem großen Durchmesser gezogen wird

benutzt. Beim Einkristallziehen wird meist zugleich eine zugehörige *Grunddotierung* mit erzeugt. Sie soll dabei über die Länge des Einkristalls möglichst konstant sein. Hierfür ist das *Zonenfloaten* besonders günstig. Dabei kann der Fremdstoff über die Schutzgasatmosphäre eingebracht werden. Andererseits ist das tiefgelfreie Zonenschmelzverfahren für spannungsfreie große Kristalle besonders schwierig. In der Schmelz- und Erstarrungszone existieren bei ihm nämlich ziemlich kompliziert ver-

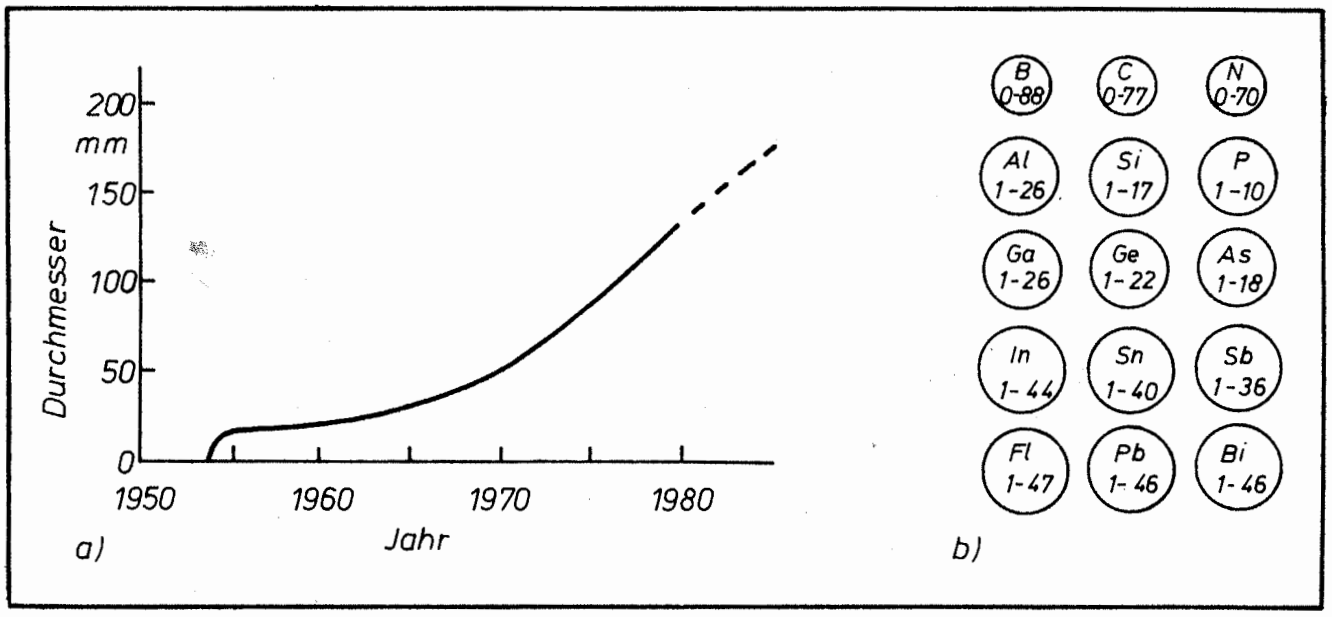


Bild 4.1.5. a) Zunahme des Durchmessers von Si-Einkristallen im Laufe der Jahre
b) PAULINGsche tetraedrische Radien in 10^{-10} m

laufende Temperaturgradienten. In dieser Hinsicht ist das CZOCHRALSKI-Verfahren günstiger. Es hat jedoch Nachteile bezüglich der konstanten Dotierung über die Länge und bezüglich des Tiegels. Die Entwicklung der erreichbaren Kristalldurchmesser über die Zeit weist Bild 4.1.5a) aus. Die Länge der Kristalle liegt heute bei mehreren dm, die erreichbare Ziehgeschwindigkeit bei 0,3 bis 10 mm je Minute, die Umdrehungszahl des Einkristalls bei etwa 5 bis 20 je Minute. Beim Einbau der Fremdstoffe stören entsprechend Bild 4.1.3 oben die unterschiedlichen tetraedrischen Radien (Bild 4.1.5 b)). Ihr Einfluß läßt sich in gewissem Umfang beheben, indem zugleich mit zwei Stoffen entgegengesetzter Abweichung gegenüber dem Grundkristall dotiert wird.

4.1.4. Herstellung der Scheiben

Am Ende dieser mehrstufigen Prozesse stehen *Scheiben mit einer Dicke* von ca. 70 bis 200 µm. Diese Scheiben müssen sehr hohen Anforderungen bezüglich der *Ebenheit, Fehlerfreiheit, Parallelität, Oberflächengüte, Kristallorientierung* usw. genügen. Der *mehrstufige Prozeß* besteht aus: Durchmesserschleifen, Trennen, Läppen, Ätzen, Polieren, Säubern und Prüfen.

Zunächst wird der Einkristall auf *gleichmäßigen Durchmesser geschliffen*. Entsprechend der kristallinen Orientierung — Wachstumsrichtung und Hauptphase — wird eine *Phase* angeschliffen. Sie dient zur Orientierung der späteren Scheibenprozesse. Teilweise werden noch Nebenphasen angebracht, die zur weiteren Kennzeichnung des Materials, z.B. *p*- oder *n*-, verwendet werden.

Das *Trennen* erfolgt heute überwiegend mittels Innenboardschneiden (Bild 4.1.6a)) (Trennschleifen). Für 80 mm dicke Einkristalle hat die aus rostfreiem Stahl bestehende Scheibe 400 mm Außen- und 150 mm Innendurchmesser bei ca. 200 µm Dicke. Allein hierdurch geht beim Schneiden über die Hälfte des Einkristallmaterials — wenn auch wiederaufbereitbar — verloren. Die Diamantschleifkörner haben Durchmesser zwischen 20 und 60 µm. Vor dem Schneiden wird der geschliffene und orientierte Einkristallstab auf Holz oder Kunststoff aufgeklebt (Bild 4.1.6b)). Die getrennten Scheiben sollen möglichst spannungsfrei und gut parallel sein sowie eine geringe Dickentoleranz bei geringer Rauhtiefe besitzen. Die schematische Darstellung der

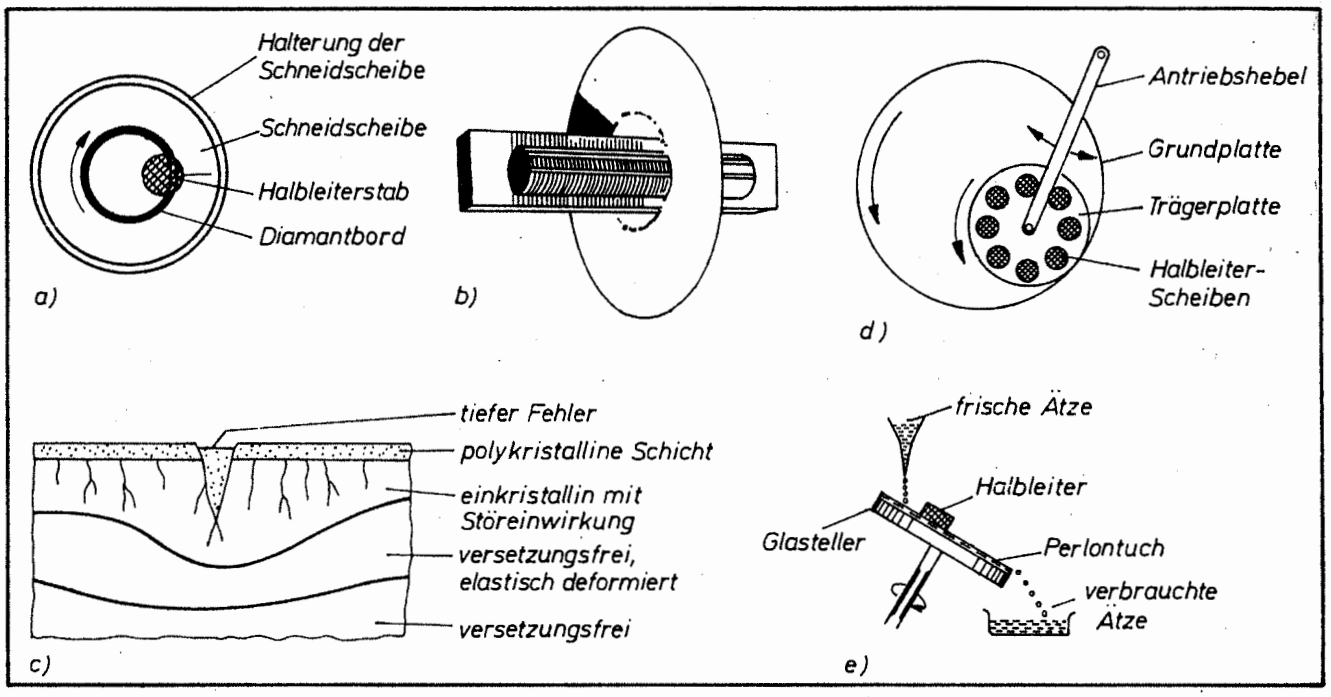


Bild 4.1.6. Prozesse zum Herstellen der Scheiben

a) Prinzip und b) Schema des Innenboardschneidens
c) Strukturfehler an der Oberfläche eines Wafers
d) Prinzip des Läppens
e) Prinzip des Wischpolierens

gestörten Oberfläche zeigt Bild 4.1.6c). Die Tiefe der Schädigung liegt bei ca. 15 bis 50 μm.

Das *Läppen* (Bild 4.1.6d)) wird meist nur angewendet, wenn beim Trennen nicht die notwendige Dickentoleranz, Parallelität und Ebenheit erreicht werden kann. Gleichzeitig wird mit dem Läppen die gestörte Oberfläche abgetragen. Das Läppen erzeugt aber andererseits selbst eine gestörte Oberfläche, jedoch nur von der Tiefe 3 bis 10 μm. Beim Läppen werden meist 60 bis 100 μm Material mit einer Geschwindigkeit von 2 bis 20 μm/min abgetragen. Ein Nachteil des Läppens ist, daß leicht der Scheibenrand geschädigt wird.

Das *Ätzen* wird für drei Anwendungen benutzt:

- Herstellung einer kristallin ungestörten Oberfläche für die Scheiben;
- bei den Scheibenprozessen zur Strukturierung;
- für die Feststellung der Orientierung und von Fehlern.

Hier interessiert nur der erste Fall. Auf die anderen wird später eingegangen. Es werden vier Ätzverfahren unterschieden:

chemisches, elektrochemisches, ionisches und thermisches Ätzen.

Besondere Bedeutung haben die speziellen Verfahren des naßchemischen, plasmachemischen, Gasphasen- und Ionenätzens erreicht. Das naßchemische Ätzen erfolgt z.B. mit HNO_3:HF. Durch das Verhältnis beider kann die Ätzgeschwindigkeit gesteuert werden. Die Ätzgeschwindigkeit wird generell im Bereich 10 bis 500 μm/min gewählt. Insgesamt werden dabei um 50 μm abgetragen. Nachteilig beim Ätzen ist, daß es oft die Dickentoleranz und die Parallelität verschlechtert.

Das *Polieren* erfolgt oft nur auf einer Scheibenseite. Es erzeugt eine ebene, spiegelglatte, schleierfreie Scheibenoberfläche ohne Störungen des Kristallgitters. Oft geschieht es in zwei Stufen: Grob- und Feinpolieren. Dabei werden insgesamt ca. 30 bis 50 μm Material mit einer Geschwindigkeit von ca. 50 μm/h abgetragen (Bild 4.1.6e)).

Das Polieren kann nach mehreren Prinzipien erfolgen: chemisch, elektrochemisch, mechanisch oder chemomechanisch. Beim mechanischen Verfahren kommen Al_2O_3, Diamant, Zirconium- oder sehr selten Ceriumoxid zur Anwendung. Die Korndurchmesser liegen zwischen 0,01 und 1 µm. Die Polierzeiten liegen bei einigen 10 Minuten bis einigen Stunden. Al_2O_3 hinterläßt ziemlich starke Deformationsschichten (einige µm), Diamantpaste erzeugt häufig Mikrorisse. Das chemomechanische Polierverfahren verwendet SiO_2 in kolloidalen Abmessungen, gelöst in Natronlauge mit dem pH-Wert 10 ... 11. Dabei entstehen wasserlösliche Natriumsilikate.

Die Reinigung der Oberfläche hat vielerlei Aufgaben, zu denen vor allem das Entfernen von Ölen, Wachsen, Fingerabdrücken, Ablagerungen aus der Luft, von Kunststoffen oder von absorbierten Ionen zählen. Hierzu dienen u. a. fluorierte Kohlenstoffe oder H_2O_2-Lösungen. Daneben hat das *Glimmreinigen* Bedeutung. Weiter gehört zum Reinigen Trocknen und Ausheizen.

Die *Fehler auf der Waferoberfläche* können besonders gut durch *richtungsabhängiges Ätzen* festgestellt werden. Hierzu werden spezielle Ätzen verwendet, die in der Regel aus drei Komponenten bestehen:

— ein basisches Oxydationsmittel, wie Hydrazin oder Kaliumhydroxid
— ein Komplexbildner, wie Isopropylalkohol
— ein Katalysator für den Oxydationsvorgang.

Die Ergebnisse je nach Defekt und Kristallorientierung zeigt Bild 4.1.7.

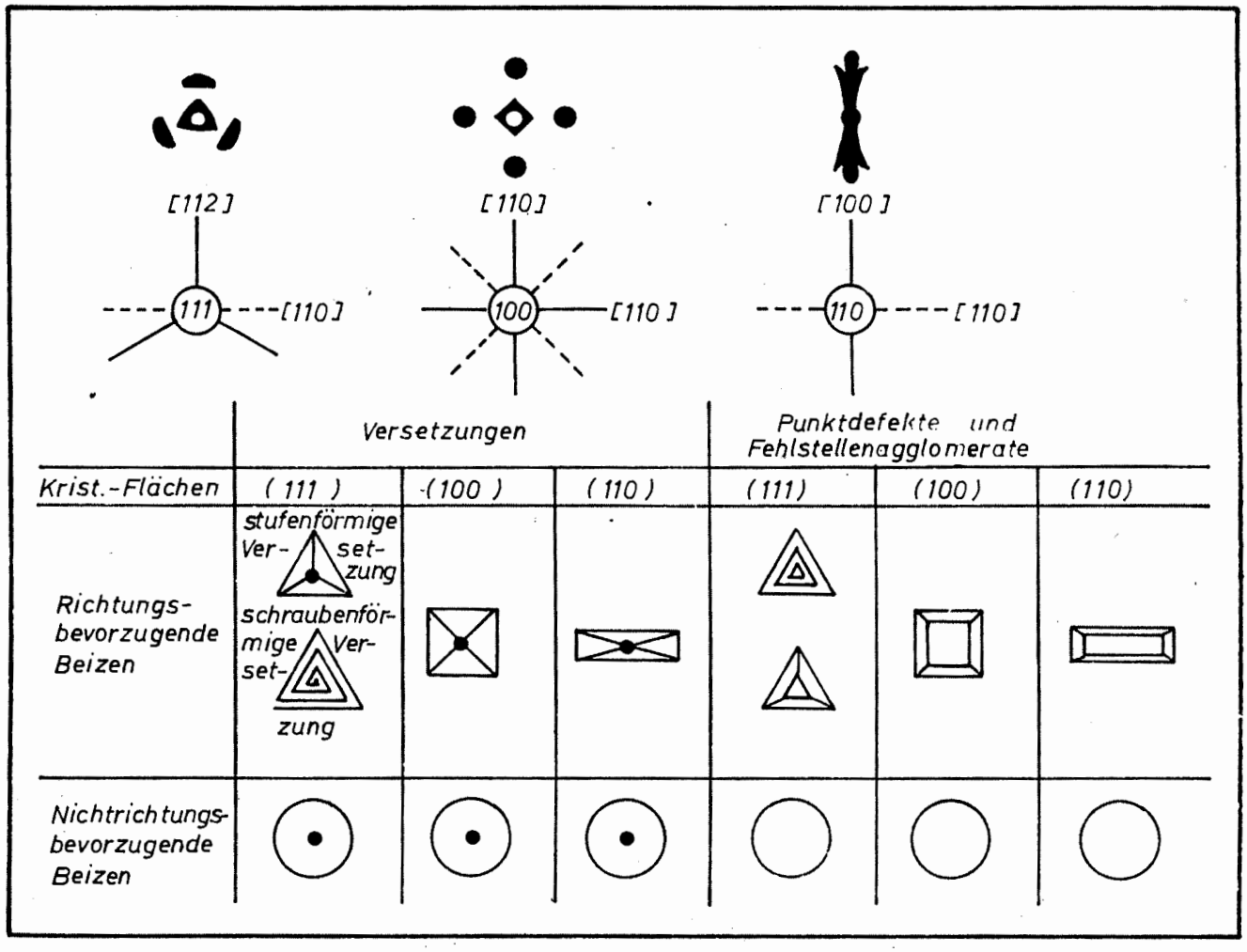


Bild 4.1.7. Oben: Lichtreflexionsbilder von verschieden orientierten Kristalloberflächen bei Ätzung der Oberfläche
Unten: Schematische Darstellung der Ätzgruben in Abhängigkeit von der Kristallorientierung

4.2. Oberflächenschichten

Die Eigenschaften der Halbleiterbauelemente werden fast ausschließlich durch Strukturänderungen von der Oberfläche der Wafer her realisiert. Dabei sind zwei Prinzipien zu unterscheiden, nämlich jene, welche die Gesamtoberfläche betreffen, und jene, die an definierten Orten vorgenommen werden. Häufig wird beides nach den gleichen Prinzipien realisiert, bei den örtlich differenzierten Strukturen werden dann nur zusätzliche Maskentechniken verwendet, die im nächsten Abschnitt beschrieben werden. Im folgenden sind vier verschiedene Oberflächenveränderungen zu behandeln: Epitaxie, Diffusion von Fremdstoffen, Erzeugung von Isolationsschichten und Aufbringen von Metallschichten.

4.2.1. Epitaxie

Bei der Epitaxie wird von einer einkristallinen Unterlage, dem Wirt oder dem Substrat ausgegangen. Hierauf kann, unter bestimmten Bedingungen unter Fortsetzung des Kristallgitters eine Schicht, der Gast oder das Deposit aufwachsen. Bei gleichem Material wird von Homo-, sonst von Heteroepitaxie gesprochen. Es gibt im Prinzip eine Vielzahl von Epitaxieverfahren. Für die wichtigsten Halbleitermaterialien Ge,

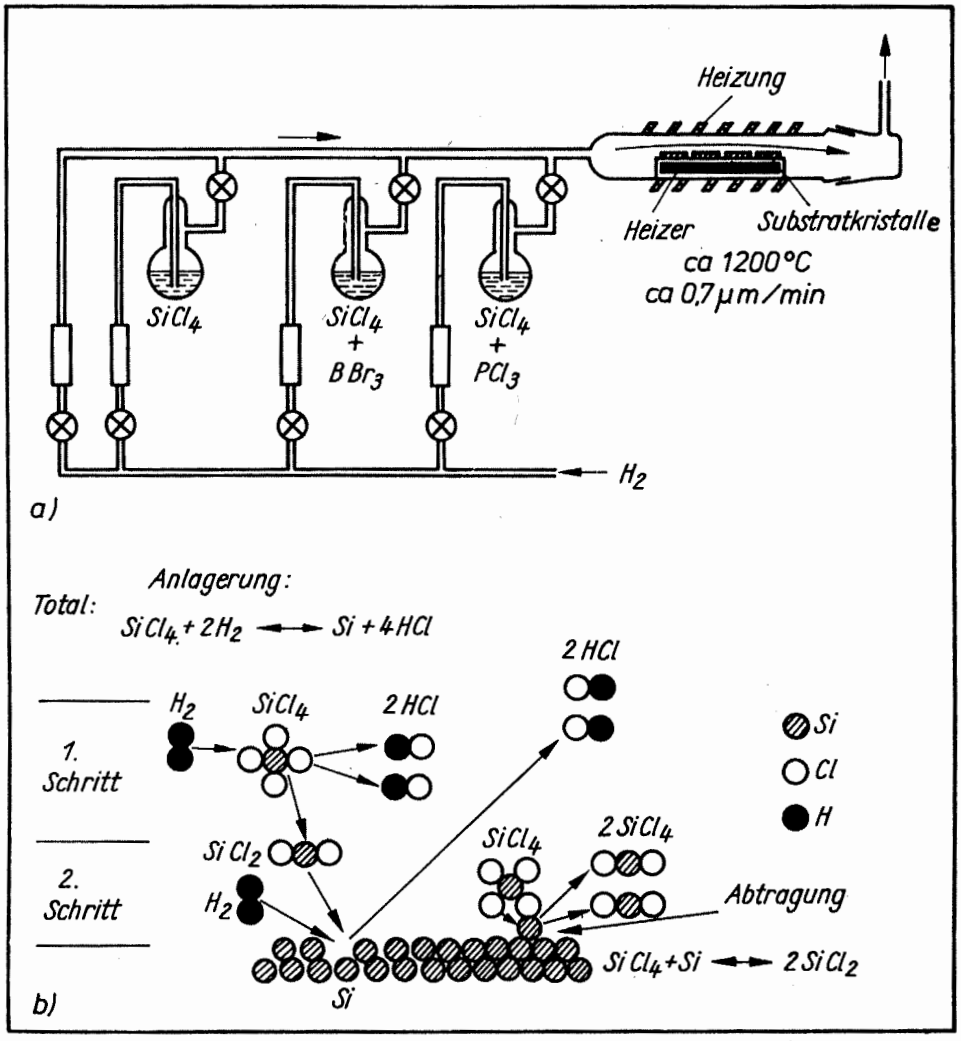


Bild 4.2.1. a) Schematische Anordnung zur Erzeugung von Epitaxieschichten auf Si-Substraten mit $SiCl_4$ (die möglichen Zusätze aus BBr_3 bzw. PCl_3 dienen zur Einstellung der gewünschten Dotierung)

b) Zusammenfassung der wesentlichen chemisch-physikalischen Vorgänge

Si, GaAs besitzen zwei größere Bedeutung, und zwar die Züchtung aus der

— Schmelze für kompakte Materialien (bereits behandelt),
— Dampfphase für die eigentlich epitaktischen Schichten.

Den prinzipiellen Aufbau einer entsprechenden Apparatur für Si zeigt Bild 4.2.1. Die darin gezeigten Vorgänge stellen eine Vereinfachung der Reaktionsabläufe dar. Es sind bisher weit über zehn Teilreaktionen nachgewiesen, die in dem doch relativ komplexen Prozeß mitwirken. Besondere Aufmerksamkeit ist bei allen Epitaxieprozessen den möglichen Kristallfehlern zu schenken.

4.2.2. Diffusion

Bei der Diffusion müssen an der Oberfläche des Einkristalls zwei Stoffströme auftreten. Einmal müssen Atome das Kristallgitter verlassen, und zum anderen müssen von der Oberfläche her Fremdatome auf diese Plätze gelangen. Aus diesem Grunde bestimmen vier Parameter maßgeblich den Diffusionsvorgang:

— Diffusionstemperatur; je höher sie ist, desto mehr Energie steht für den Vorgang zur Verfügung, desto schneller verläuft er.
— Gitterkonstante; je größer sie ist, um so leichter erfolgt die Diffusion.
— Atomradius der diffundierenden Fremdstoffe; je kleiner er ist, desto leichter die Diffusion.
— Wechselwirkungen zwischen Wirtsgitter und Fremdatomen infolge von COULOMBschen oder VAN-DER-WAALSschen Anziehungskräften.

Für die statische Strömungsgeschwindigkeit S von Atomen, bezogen auf die Zeit- und Oberflächeneinheit, gilt das erste FICKsche Gesetz

$$S = -D\frac{\mathrm{d}c}{\mathrm{d}x}, \tag{5}$$

wobei D der Diffusionskoeffizient und $\mathrm{d}c/\mathrm{d}x$ das Konzentrationsgefälle in Diffusionsrichtung ist. Werte für einige Materialien sind in Tab. 4.1.1 zusammengestellt. Für den zeitlichen Verlauf der Diffusion gilt das zweite FICKsche Gesetz

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2}. \tag{6}$$

Hieraus kann die Konzentrationsänderung an einem Ort x unterhalb der Oberfläche in Abhängigkeit von der Zeit berechnet werden. Für die Lösung der Differentialgleichungen sind drei praktisch realisierbare Grenzfälle zu unterscheiden:

— Diffusion aus der Gasphase,
— Vorablagerung an der Oberfläche,
— Herausdiffundieren von Dotierungsmaterial.

Sie alle verlaufen bei wesentlich erhöhten Temperaturen um ca. 1000 °C.

Mit jeder Methode werden unterschiedliche Konzentrationsverläufe (Diffusionsprofile) erreicht. Außerdem hat jedes Verfahren spezifische Vor- und Nachteile. Die Diffusion aus der Gasphase kann in abgeschlossenen oder offenen Systemen erfolgen, d. h. mit fest vorgegebener Fremdstoffmenge oder mit fester Konzentration des vorbeiströmenden Gases. Zum ersten Fall gehört auch die Ablagerung von zunächst festen Verbindungen des Diffusionsstoffes auf der Oberfläche der Scheiben. Dieses Verfahren wird vor allem bei der Eindiffusion von Gold verwendet. Bei der Diffusion

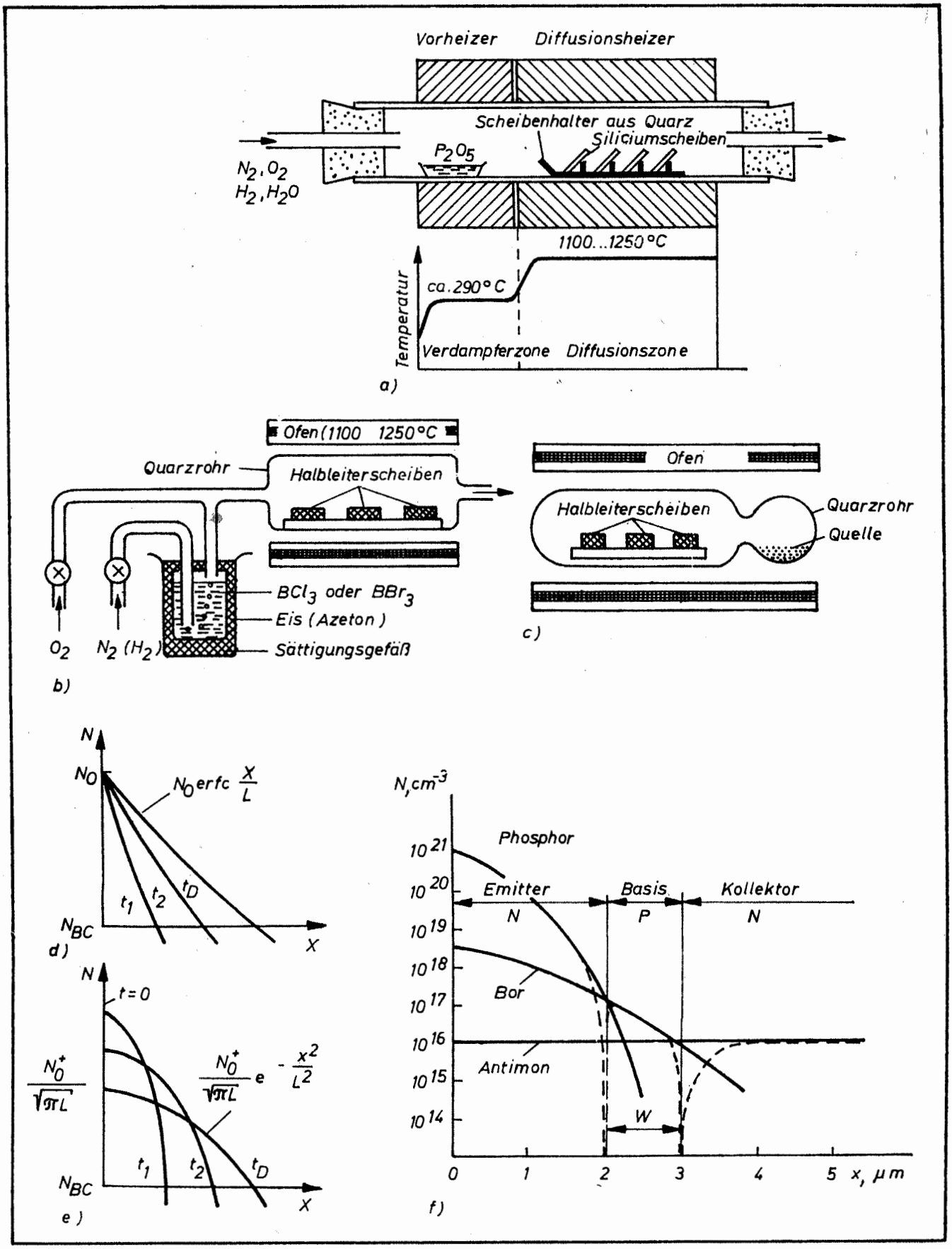


Bild 4.2.2. Beispiele von Diffusionsverfahren

a) Im Gasstrom mit vorgeheiztem Dotierungsschiffchen für P_2O_5
b) Offenes Verfahren mit gasförmiger Quelle BCl_3 oder BBr_3
c) Geschlossenes Verfahren
d) Diffusionsprofil bei konstanter Oberflächenkonzentration
e) Diffusionsprofil bei einer vorgegebenen Menge von Fremdatomen an der Substratoberfläche
f) Konzentrationsverlauf der Störstellen in einem Si-Kristall vom n-Typ nach einer Doppeldiffusion mit B und P

aus der Gasphase in geschlossenen Systemen wird meist von Stoffen ausgegangen, die bei Zimmertemperatur fest, z. B. P_2O_5, oder flüssig sind, z. B. $POCl_3$ (vgl. Bild 4.2.2). Sie werden in Gefäßen entfernt von den Wafern aufgestellt und auf eine Temperatur gebracht, die den gewünschten Dampfdruck des Dotierungsstoffes bewirkt. Für die Verfahren im Gasstrom kommen gasförmige Verbindungen, z. B. PH_3, in Betracht. Bild 4.2.2f) zeigt, wie bei gleichzeitiger Diffusion von Phosphor und Bor bei einer Grunddotierung mit Antimon ein Transistorprofil erhalten werden kann. Solche Dotierungspaare sind bei Germanium Sb und Ga oder B und bei Silicium B, Al oder Ga und P, Sb oder As. Hierbei gilt die Regel, daß Donatoren für Germanium und Akzeptoren für Silicium die höheren Diffusionskoeffizienten besitzen. Auch dies ist ein Grund, weshalb beim Germanium vorzugsweise vom *p*-Material ausgegangen wird und als Folge *pnp*-Transistoren vorrangig entstehen. Beim Si hingegen wird meist von *n*-Material mit der dann entstehenden Folge *npn* ausgegangen. Die Zeiten für die Diffusion liegen etwa zwischen 10 Minuten und mehreren Stunden.

4.2.3. Isolierschichten

Isolierschichten auf der Oberfläche von Halbleitermaterialien werden sehr vielfältig benutzt. Vorwiegend haben *vier Prinzipien* Bedeutung:

- Isolierschichten wie bei den MOSFET, die bei der Feldsteuerung mitwirken.
- Isolierschichten, die den strukturierten Halbleiterkristallen eine definierte und unveränderliche Oberfläche verleihen sollen. Sie haben es möglich gemacht, die Halbleiterbauelemente in Kunststoff einzupressen, reduzieren erheblich die ökonomischen Aufwendungen und erhöhen die Zuverlässigkeit der Bauelemente. Sie werden als passivierend bezeichnet.
- Schichten, die als Haftmasken bei der Strukturierung zu örtlich selektiven Beeinflussung herangezogen werden.
- Spezielle Schichten, die zu FET mit steuerbarer Schwellspannung, z. B. Speicher-FET (auf beide wird im Abschnitt 8.3.5. eingegangen), führen.

Im Mittel erfüllt, wenn vom letzten Punkt abgesehen wird, die SiO_2-Schicht am besten die *Forderungen*. Sie können etwa wie folgt zusammengefaßt werden:

- Undurchlässigkeit der Schicht für Stoffe aus der Umgebung. Bei den typischen Dicken von 0,15 bis 0,5 µm besteht für SiO_2 aber eine deutliche Reaktionsfähigkeit mit Wasser.
- Stabilität der notwendigerweise relativ geringen Ladungen im Isolator. Sie liegen beim SiO_2 bei etwa 10^{-8} ... 10^{-7} As cm^{-2} (10^{11} ... 10^{12} Elementarladungen cm^{-2}), also beim optimalen Wert von $-5 \cdot 10^{-8}$ As cm^{-2}. Leider existieren im SiO_2 aber noch störende bewegliche positive Ionen, z. B. Na^+ und H^+. Sie rühren unter anderem von Glasgeräten, Chemikalien und Handschweiß her und lassen sich nur mit technisch kaum vertretbarem Aufwand beseitigen. Bereits ein Na-Atom auf 10^6 Si-Atome führt zu 10^{12} positiven Ladungen je cm^2.
- Hohe chemische und physikalische Stabilität.
- Hohe Stabilität des Systems Halbleiter—Isolator gegen thermische Wechselbeanspruchung (Bedeutung der angepaßten Ausdehnungskoeffizienten).
- Möglichst großer Wert des Produktes aus Durchbruchfeldstärke und Dielektrizitätskonstante ($\geqq 5 \cdot 10^{-7}$ As cm^{-2}). Beim SiO_2 sind Betriebsfeldstärken im Bereich von 10^6 V/cm typisch.

Tab. 4.2.1 gibt einen Überblick zu wesentlichen Eigenschaften der vier wichtigsten Isolierschichten. Daraus gehen deutlich die allgemeinen Vorteile von SiO_2 hervor. Für

Tabelle 4.2.1. Eigenschaften ausgewählter Isolierschichten auf Silicium

Material	Ausdehnungskoeffizient $10^{-6}\ K^{-1}$	Dielektrizitätskonstante ε_r	Durchbruchfeldstärke $10^6\ V\ cm^{-1}$	Max. Influenzladung $10^{-6}\ As\ cm^{-2}$	Haftstellendichte cm^{-3}	Breite des verbotenen Bandes eV
SiO_2	0,56	4	6	3,55	$6 \cdot 10^{15}$	8,9
Si_3N_4	2,5 ... 4,2	6,5 ... 7,5	8	5,7	10^{16} ... 10^{17}	5,1
Al_2O_3	6	7,8 ... 8,1	5 ... 10	3,54	10^{17}	8,7
AlN	—	8 ... 10	4 ... 15	2,8 ... 13,2	—	
Si	2,3	12	0,3	—	—	1,1

spezielle Anwendungen sind eventuell andere Schichten günstiger, bei passivierenden Deckschichten sind es u. a. Si_3N_4 und Al_2O_3; als Diffusionsmaske Si_3N_4. Während die SiO_2-Schichten vorrangig durch *thermische Oxydation* hergestellt werden, kommen für die anderen Schichten überwiegend *Zerstäubungsmethoden* in Betracht, z. B.

— reaktiv mittels Gleichspannung (Gasentladung),
— HF-Zerstäubung,
— reaktiv mit Hochfrequenz.

Sie ähneln in gewissem Umfang jenen Verfahren, die im nächsten Unterabschnitt bei Metallen behandelt werden.

Für die thermische Oxydation von Si zu SiO_2 sind vier Verfahren üblich:

— trockener Sauerstoff,
— mit H_2O versetzter Sauerstoff (feuchtes O_2) (Bild 4.2.3a)),
— mit H_2O versetzter Stickstoff,
— durch PbO beschleunigte Oxydation.

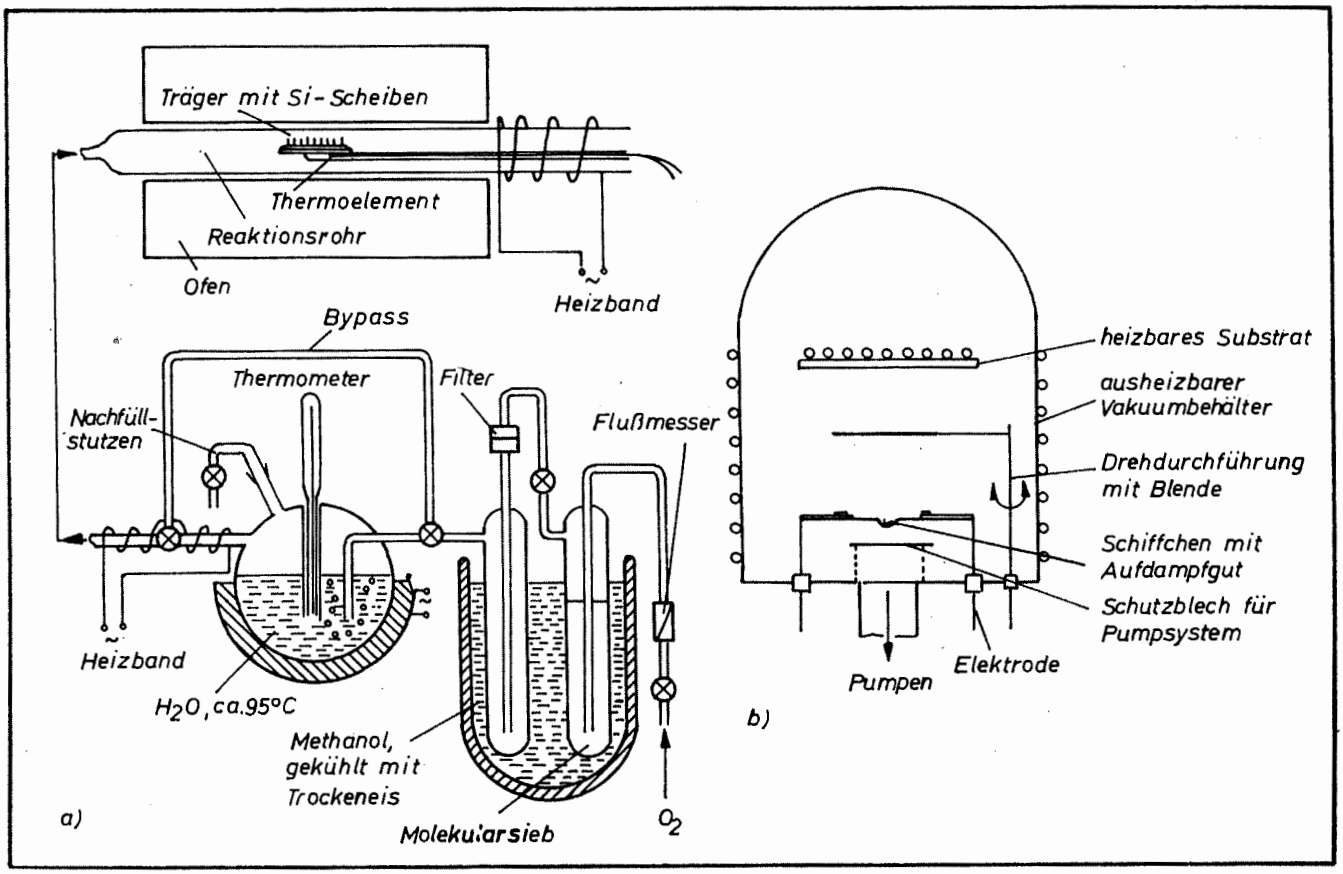


Bild 4.2.3. Schematische Darstellung a) einer Oxydationsanlage und b) einer Aufdampfanlage

Daneben besitzt noch eine geringe Bedeutung die

— anodische Oxydation.

Durch seine große Affinität zu Sauerstoff bildet Si an der Luft eine etwa 3 nm dicke SiO_2-Schicht. Sie kann elektrolytisch, ähnlich wie bei den Elektrolytkondensatoren, verstärkt werden. Details enthält Tab. 4.2.2.

Tabelle 4.2.2. Verfahren zur Herstellung von SiO_2-Schichten

Verfahren	Reaktions- bzw. Substrattemperatur °C	Abscheidungsrate µm/h
Thermische Oxydation		
● trocken O_2	1000 ... 1200	0,2
● feucht O_2	1000 ... 1200	1
● feucht N_2	1000 ... 1200	0,4
● mit PbO	750	0,4
Anodische Oxydation	20 ... 30	—
Pyrolytische Schichten (Zersetzung von SiH_4)	300 ... 800	0,8
Hydrolyse: $SiCl_4 + 2\,CO_2 + 2\,H_2 \rightarrow SiO_2 + 2\,CO + 4\,HCl$	800 ... 1200	4 .. 8
Bedampfung	400	0,5
Reaktive Zerstäubung	200	2
CO_2-Oxydation	1000	0,4
Transportreaktion	300 ... 600	0,5

4.2.4. Metallschichten

Auf beliebigen Substraten können metallische Schichten nach *folgenden Verfahren* aufgebracht werden:

— Verdampfen des Metalls durch Erhitzen im Vakuum (Bild 4.2.3b)),
— Zerstäuben des Metalls durch Gasentladungen (10^{-3} ... 10^{-1} Torr Edelgas) mit Gleichstrom oder Hochfrequenz (Sputtern),
— elektrolytisches Abscheiden,
— stromloses Abscheiden aus einer Lösung.

Für die Halbleitersubstrate haben *Zerstäuben* und ganz besonders *Vakuumbedampfung* Bedeutung. Die notwendige Temperaturerhöhung des Aufdampfmaterials erfolgt meist in Tiegeln durch Stromwärme, seltener durch Beschuß mit energiereichen Elektronen. Im letzteren Fall kann vorteilhaft tiegelfrei gearbeitet werden. Die übliche Schichtdicke für die Metallkontakte liegt bei etwa 1 µm. Überwiegend wird als Material *Aluminium* verwendet. Bei einer Schichtdicke von 1 µm ist sein spezifischer Widerstand etwa 10 ... 20% größer als der des kompakten Materials. Kontaktprobleme können durch hohe Dotierung des Si vermieden werden. Außer den Kontakt-

problemen und der guten Leitfähigkeit sind folgende Forderungen zu erfüllen:

— gute Haftfestigkeit auch auf den Isolierschichten, z. B. SiO_2,
— gute Ätzbarkeit, da oft eine nachträgliche Strukturierung für Leiterbahnen vorgenommen wird.

Beim Bedampfen werden meist vorher die Siliciumscheiben aufgeheizt. Hierdurch sollen vor allem störende Wasser-und Gasabsorptionsschichten beseitigt werden. Auch beim Bedampfen werden die Siliciumscheiben gewöhnlich noch auf höheren Temperaturen (ca. 200 °C) gehalten. Anschließend wird in geringem Umfang versucht, das Aluminium einzulegieren. Hierbei darf nicht die Temperatur für die flüssige Legierungsphase von 577 °C erreicht werden. Ferner ist zu beachten, daß Aluminium ein Akzeptor, wenn auch mit sehr geringer Löslichkeit ist. *Bei hohen Stromdichten* — ca. 10^6 A cm^{-2} — neigt Aluminium zur *Elektromigration*. Es wandert unter Stromeinfluß. Bei den Leitbahnen von nur wenigen μm^2 sind dafür nur Ströme im mA-Bereich notwendig. Für höhere Zuverlässigkeit und für besondere Anwendungen bringen Mehrkomponentenmetallierungen Vorteile. Dabei haben die Folgen Cr—Ag—Au bzw. Pt—Ti—Pt—Au insbesondere in der Beamlead-Technik (s. 9.4.3.) Bedeutung. Bei Al werden bei ca. 1300 °C Niederschlagsraten von ca. 2 nm/s erreicht.

4.2.5. Ionenimplantation

Die Ionenimplantation wurde bereits um 1950 von SHOCKLEY vorgeschlagen (1954 Patent). 1961 erschien die erste Publikation, und 1966 erfolgte die erste praktische Anwendung. Heute ist sie ein recht breit angewendetes Dotierungsverfahren.

Das Schema der Ionenimplantation geht aus Bild 4.2.4a) hervor. Ionen aus einer Ionenquelle werden beschleunigt und mittels eines Magneten so abgelenkt, daß nur Ionen mit einer Masse ausgeblendet werden. Hierdurch wird extrem hohe Reinheit der Dotierungssubstanz erreicht. Mit einem Ablenkungssystem können die Ionen gegebenenfalls an genau definierte Punkte auf der Scheibe gelenkt werden. Wie Bild 4.2.4b) zeigt, sind derartige Anlagen sehr groß und damit auch technisch aufwendig. Sie konnten sich nur durch die Vorteile der Ionenimplantation durchsetzen.

Für die Ionenimplantation gibt es vor allem *zwei Theorien*: eine für hohe Energie von BOHR, BETHE und BLOCH und eine für niedrige Energie von LINDHARD, SCHARFF und SCHIOTT (LSS-Theorie) [R 20].

Die LSS-Theorie wird heute allgemein angewendet, obwohl sie nur gut für amorphe Halbleiter gilt. Sie ergibt dann GAUSSsche Dotierungsprofile, wie sie Bild 4.2.4c) zeigt, deren Maximum in einer Tiefe R liegt. Die Streuung liegt bei etwa $0{,}5R$. Die Tiefe R hängt von der Beschleunigungsspannung (Energie) und dem Masseverhältnis der Ionen zum Substrat ab. Deshalb dringen Ionen mit höherer Masse weniger tief ein (Tab. 4.2.3). Bild 4.2.4d) zeigt einen Effekt, der sich bei fast exakter Ausrichtung der Ionenstrahlen und Kristallachse ergibt. Die Ionen dringen *auf einem Kanal*

Tabelle 4.2.3. Maxima der Eindringtiefen von Ionen mit 100 kV in nm

Substrat	Ionen							
	^{11}B	^{14}N	^{27}Al	^{31}P	^{70}Ga	^{75}As	^{115}In	^{204}Tl
Silicium	400	280	140	120	60	58	46	40
Germanium	255	180	95	82	42	40	30	24
und GaAs								

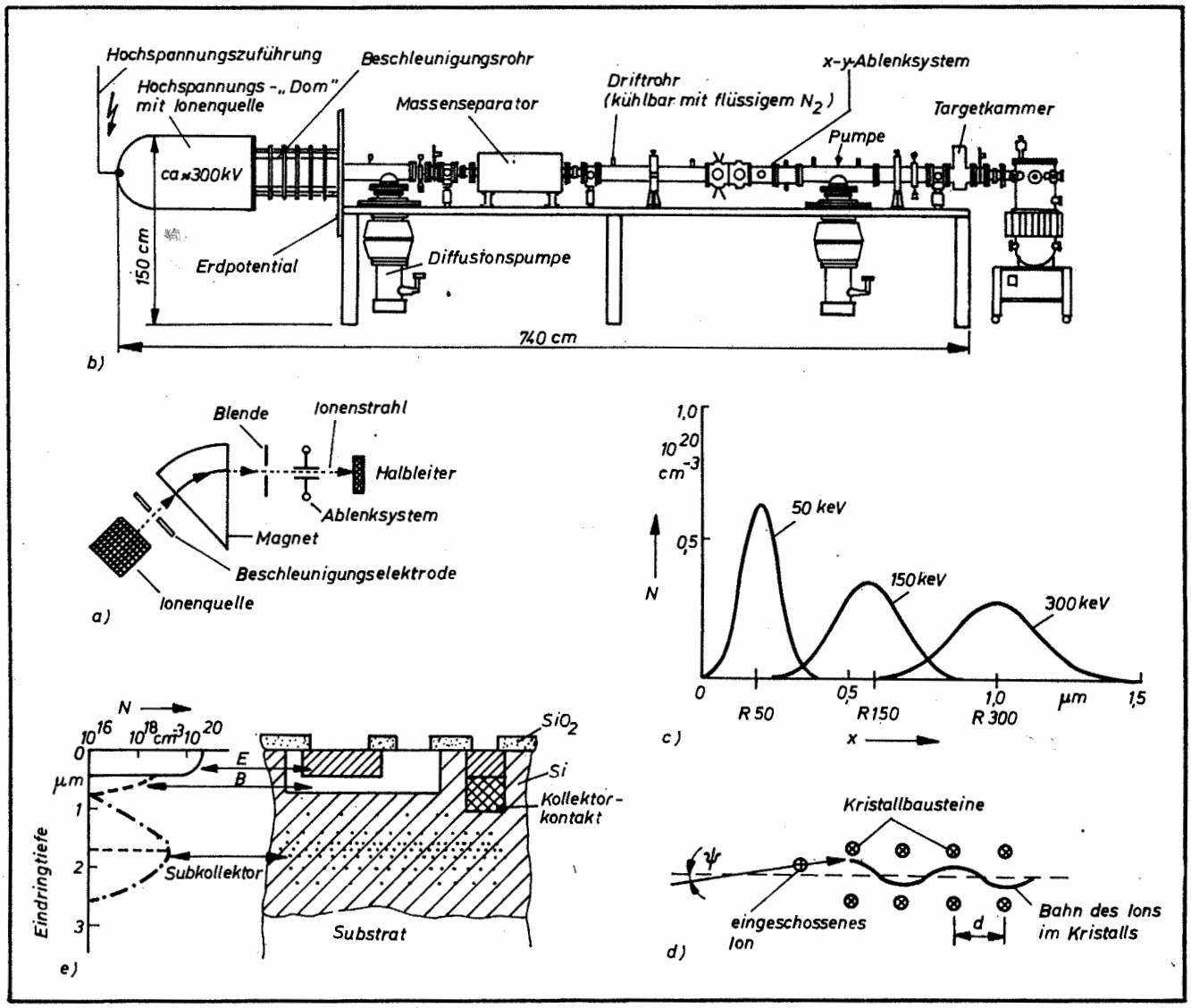


Bild 4.2.4. Zur Ionenimplantation

a) Prinzip des Mechanismus
b) Beispiel einer Apparatur
c) Ionenprofile nach der LSS-Theorie von implantiertem Bor in Silicium bei verschiedenen Einschußenergien (konstante Ionendosis 10^{15} cm^{-2})
d) Eindringen von Ionen in Einkristallkanäle (Channeling; d ist der Abstand der Atome im Kristallgitter)
e) Aufbau eines epitaxiefreien, implantierten Bipolartransistors

relativ tief in den Kristall (Channeling). Dieser Effekt wäre für extreme Dotierungen durchaus nützlich. Durch den Einfluß der Oberfläche des Kristalls und einer sich verstärkenden Auswirkung von Kristallfehlern ist er jedoch praktisch nicht anwendbar. Deshalb werden immer Methoden zu seiner Unterdrückung benutzt, z.B. Neigungen um Winkel ψ über 5 bis 15°, definierte Oberflächenschichten zur Streuung der Ionenstrahlen. Dies ist auch der Grund, weshalb die LSS-Theorie für die Einkristalle in brauchbarer Näherung gilt.

Die wesentlichen *Vorteile der Ionenimplantation* sind:

- Eine genaue Steuerung der Dotierungsmenge (von sehr klein bis sehr groß) ist über eine Zeit-Strom-Messung möglich. Dadurch ist sehr genau die Menge über den gesamten Wafer einhaltbar (Toleranzen $< 1\%$), z.B. Herstellung eng tolerierter Widerstände. Die kleinen Dotierungsmengen ermöglichen darüber hinaus, relativ große Widerstände herzustellen.
- Das Dotierungsprofil ist über Ionenart und Beschleunigungsspannung definiert beeinflußbar.

- Extreme Reinheit der Dotanten infolge Prinzip des Massenspektrometers.
- Infolge der genauen Richtung der Strahlen, deren definierter Ablenkbarkeit und der Absorption an Oxid-, Metall- und Photolackschichten ist auch eine genaue (laterale) örtliche Steuerung der Dotierung möglich.
- Der Einbau von Dotierungsatomen ist ein Nichtgleichgewichtsprozeß. Dadurch sind bei niedrigen Temperaturen hohe Dotierungen möglich. Weiter sind dadurch steilere Dotierungsprofile als bei den thermischen Prozessen der Diffusion möglich.
- Die Ionenimplantation erfolgt gegenüber der Diffusion und Epitaxie sehr schnell (Sekunden und Minuten).
- Durch mehrfache Anwendung von verschiedenen Ionenimplantationen können unterschiedliche Profile kombiniert werden. Das Beispiel eines Bipolartransistors, der ohne Epitaxie, ausschließlich mit mehrfacher Ionenimplantation hergestellt wurde, zeigt Bild 4.2.4e).

Einer der *Nachteile* der Ionenimplantation (neben ihrem hohen Aufwand) ist, daß durch sie *Störungen im Kristallaufbau* des Halbleiterkristalls erzeugt werden. Sie können aber weitgehend thermisch ausgeheilt werden.

4.2.6. Vergleich der Dotierungstechniken

Die Dotierung ist ein entscheidender Prozeß der Halbleitertechnologie, mit ihr werden u. a. die verschiedenen *pn*-Übergänge erzeugt. Wenn wir von der definierten Grunddotierung des Wafermaterials und den veralteten Techniken des *dotierten Ziehens* von *pn*-Übergängen (alte Flächentransistoren) sowie des Legierens (Legierungstransistor) absehen, stehen heute *drei* unterschiedliche *Prozesse* zur Verfügung:

- Diffusion
- Epitaxie
- Ionenimplantation.

Sie haben alle drei spezifische Eigenschaften, die kurz zusammengestellt werden sollen.

Mittels *Diffusion* können nur Profile hergestellt werden, die der Lösung der Diffusionsgleichung gehorchen. Das sind vor allem die GAUSS- und komplementäre Fehlerfunktion (erfc), wobei zusätzlich stets an der Oberfläche die höchste Dotierung liegt. Durch die Diffusionsprozesse ist der Variationsbereich bezüglich Tiefe und Gesamtkonzentration zusätzlich eingeengt. Diffusion ist ein heißer Prozeß von relativ langer Dauer. Die *Epitaxie* ermöglicht grundsätzlich, beliebige Profile herzustellen. Dabei ist jedoch zu beachten, daß immer gleichzeitig eine Diffusion stattfindet, die keine sehr steilen Gradienten zuläßt. Die *Ionenimplantation* erlaubt die exakteste Prozeßführung, läßt aber nur sehr geringe Dotierungstiefen, etwa unter 1 μm, zu.

Schließlich existiert noch ein weiteres Dotierungsverfahren:

- *Neutronenbestrahlung.*

Silicium besteht aus den Isotopen ^{28}Si, ^{29}Si, ^{30}Si. Mittels Neutronenbeschuß wird bei Treffern die Massezahl um jeweils Eins erhöht. So entsteht auch das instabile Isotop ^{31}Si. Es ist ein β-Strahler mit einer Halbwertzeit von 2,6 Stunden und geht dabei in Phosphor über. Durch Neutronenbestrahlung wird das Silicium so *n*-leitend. Die Dotierung ist *sehr homogen*. Deshalb ist das Material besonders für *Hochleistungsbauelemente* geeignet.

4.3. Strukturierung

Der große Fortschritt der Halbleitertechnik trat ein, als es ab 1960 gelang, auf einem Halbleitersubstrat nicht nur einzelne Transistoren, sondern ganze Schaltungen zu erzeugen. Heute werden auf einer Teilfläche des Wafers von nur wenigen mm^2 Schaltungen aus Tausenden von Transistorfunktionen strukturiert, und dies gleichzeitig zig- bis hundertmal nebeneinander auf einem Wafer.

Dieser Prozeß verlangt, daß die Prozesse des vorigen Abschnittes selektiv nur an genau bestimmten Orten auf dem Wafer wirksam werden, wobei die einzelnen Prozesse mehrfach nacheinander an verschiedenen Orten wirken müssen. Die so aufeinander abgestimmten Prozesse verlangen höchste Präzision. Es müssen Linienbreiten im µm-Bereich genau reproduzierbar mit Deckungstoleranzen im 0,1-µm-Bereich auf der Fläche des gesamten Wafers von ca. 100 mm^2 erreicht werden. Das ist eine Genauigkeit von rund 10^{-6}. Ins besser Anschauliche übertragen, müssen auf einer Kantenlänge von 100 m (Fußballplatz) Striche von 1 mm Breite auf $^1/_{10}$ mm genau angeordnet werden und bearbeitet werden können. Dabei muß dieser Prozeß in mehreren Stufen nacheinander vollzogen werden. So, wie etwa beim Mehrfarbendruck die Bildpunkte der einzelnen Farbauszüge nacheinander mit guter Konturdeckung aufgebracht werden müssen. Die Anforderungen an diese Prozesse sind so hoch, daß sie nur in äußerst staubarmen Räumen durchgeführt werden können.

Die Grundlage für die hier angedeutete Technologie ist die *Epitaxie-Planartechnik*. Das entscheidende Wort ist dabei heute „planar". Es bedeutet, daß *alle Prozeßschritte von der Oberfläche* des Wafers aus gesteuert werden. Hierzu werden zunächst hochauflösende Fotolacke über Masken (oder mittels analoger Prozesse) belichtet; dann wird der Lack entwickelt und an den so freien Stellen können von der Oberfläche her die einzelnen Prozesse des vorigen Abschnittes vollzogen werden. Es sind also zwei Komponenten, die diese Prozesse erheblich beeinflussen: die Fotolacke und die Masken.

Besonders die *Maskentechnik* ist heute vielfältig gestaltet. In der Regel sind *drei Maskentypen* zu unterscheiden: Original-, Zwischen- (reticle) und Arbeitsmasken. Sie unterscheiden sich u. a. in Größe, Verwendungszweck und Stellung im Prozeßablauf. Nur die Arbeitsmasken werden direkt zur Waferstrukturierung verwendet. Über diese Masken hinaus gibt es Methoden, die mittels gesteuerter Elektronenstrahlen teilweise oder ganz auf Masken verzichten.

4.3.1. Fotolithografie

Die Fotolithografie benötigt *spezielle Fotolacke*, auch *Resiste* oder *Fotoresiste* genannt. Sie sind im Prinzip schon lange in der *Drucktechnik* bekannt, wurden aber für die Zwecke der Halbleitertechnik weiterentwickelt. Es sind dabei Positiv- und Negativlacke zu unterscheiden. Ihre prinzipielle Funktion zeigt Bild 4.3.1.

Positivlacke sind Mischungen aus Diazochinon und einem Harz. Die lichtempfindliche Substanz geht also auf Azidogruppen, wie N_3, SO_2N_3 oder CNO_3, zurück. Sie sind über ungesättigte Bindungen (z.B. Isopren- oder Butadienderivate) bzw. mit CH-Gruppen von Polymerprodukten (Formaldehyd, Formaldehydharz) zu Makromolekülen vernetzt. Die Belichtung führt zur Depolymerisation. Dieses Material ist in einem Entwickler (alkalisch oder wäßrig) löslich. An den unbelichteten Stellen entsteht dagegen bei der Entwicklung eine zusätzliche Vernetzung mit dem Harz, also eine unlösliche Substanz.

Negativlacke gehen von der Venylgruppe $-CH_2 = CH-$ aus. Mittels Doppelbindung polymerisieren sie bei der Belichtung und werden dadurch unlöslich. Hier sind also nach der Entwicklung die unbelichteten Stellen lösbar.

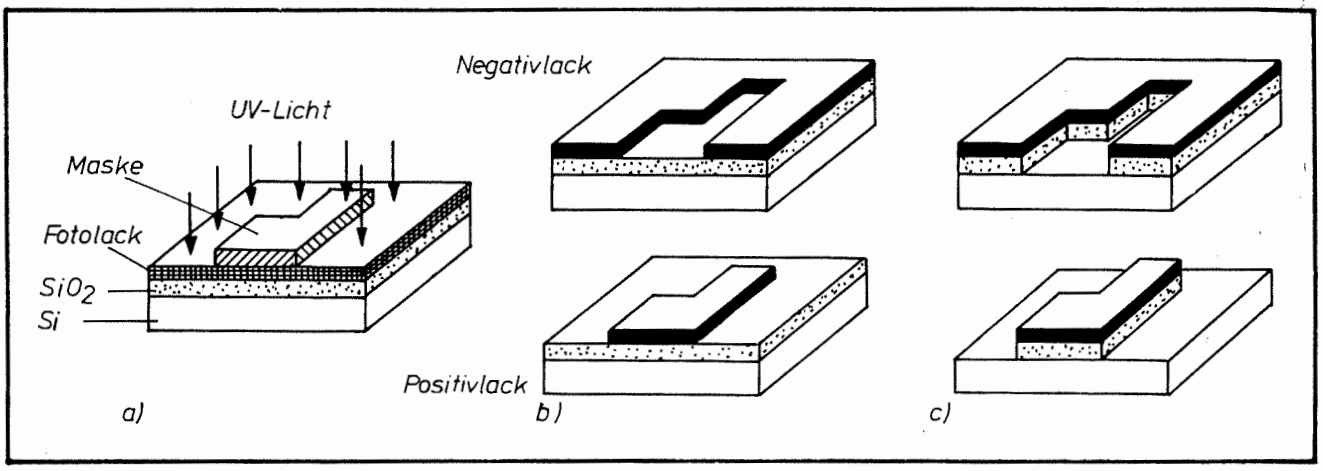


Bild 4.3.1. Zur Wirkungsweise von Fotolacken

a) Der Wafer ist mit einer Oxidschicht passiviert und darauf mit dem Fotolack beschichtet. Eine Maske deckt einen Teil des Fotolackes bei der Ultraviolettbelichtung ab

b) Die Entwicklung des Fotolackes läßt den Unterschied zwischen Negativ- (oben) und Positivlack (unten) erkennen. Es bleiben die belichteten bzw. unbelichteten Stellen des Fotolackes auf dem Wafer

c) Durch Ätzung der Oxidschicht können entsprechende Stellen der Siliciumoberfläche freigelegt werden. Diese Stellen heißen auch Fenster

Positivlacke sind in der Regel mit einem Maximum bei ca. 350 µm lichtempfindlich, während Negativlacke einen relativ breiten Empfindlichkeitsbereich von 250 bis 450 µm besitzen. Die *Empfindlichkeit der Lacke beginnt also jenseits von Grün.* Dies erlaubt ein Arbeiten bei relativ hellem gelben Licht und bringt so Vorteile für die optische Justierung der Masken. Es gibt auch spezielle Lacke für Elektronen- bzw. Röntgenstrahlen. Da Fotolacke im Gegensatz zu Silberhalogenidfilmen keine lichtempfindlichen Körner besitzen, sondern auf direkten molekularen Prozessen beruhen, ermöglichen sie ein *Auflösungsvermögen bis erheblich unter den* µm-*Bereich.* Ihre Empfindlichkeit kann so groß gemacht werden, daß mit Kohlelichtbogen oder Quecksilberdampflampen Belichtungszeiten unter 30 Minuten möglich sind.

Bei der Verwendung des Lackes muß eine hohe Reinheit, bestimmte Viskosität und gute, gleichmäßige Bindung zum Substrat vorhanden sein. Der Lack muß nämlich in einer homogenen und geschlossenen Schicht um oder unter 1 µm aufgetragen werden. Nur so ist eine fehlerfreie Strukturierung mit hinreichend hoher Auflösung möglich. Hierzu wird die Scheibe auf ca. 900 °C erhitzt, um u. a. letzte Wasserreste zu entfernen. *Ein Tropfen des Lackes* wird auf die *Mitte des Wafers* gebracht. Durch *Rotation* mit 1000 bis 10000 Umdrehungen je Minute und Vakuumabsaugung für ca. eine Minute wird die dünne, gleichmäßige Lackschicht auf dem Wafer erzeugt und anschließend bei 100 bis 200 °C für einige Minuten getrocknet. Dann erfolgt die *Ultraviolettbelichtung* über eine Maske und die Entwicklung. Nach der Entwicklung erfolgt oft noch eine Härtung durch Wärme. Dann sind alle Fotolacke gegenüber Säuren, Ätzmitteln wie Eisen-III-Chlorid, Kupferchlorid, Chromsäure, Ammoniumphosphat usw., in Sonderfällen sogar gegenüber Fluorwasserstoff beständig. Dies ist bedeutsam für das Ätzen von Glas und SiO_2-Schichten.

Mit der *Entwicklung* ist das SiO_2 der *Waferoberfläche freigelegt* und kann weggeätzt werden. So entsteht das Fenster zum Silicium für die Schichtbehandlung z. B. durch Diffusion. Danach wird der festhaftende *Fotolack entfernt.* Hierzu gibt es z. T. spezielle Lösungsmittel (u. a. Aceton, Nitroverdünnung). Auch Kochen in Natriumlauge bzw. konzentrierter Schwefelsäure wird angewendet. In der Regel bedeutet dieser Prozeß eine Belastung des Wafers. Schließlich kann der Wafer wieder mit einer SiO_2-Oberfläche versehen und neu mit Fotolack beschichtet werden. So können *in mehrfacher Abfolge viele fotolithografischen Prozesse* erfolgen. Dabei wird der Wafer immer

stärker uneben. Es bilden sich die Prozesse der Strukturierung an der Oberfläche ab. Dies hat Nachteile für eine exakte Lithografie, ist aber auch notwendig, um die jeweils nächste Maske an speziell dafür vorgesehenen Kennungen mit den vorangegangenen Masken zur guten Konturendeckung zu bringen.

Durch kleine Bläschen oder ähnliche Störungen im Fotolack treten zuweilen sehr kleine zusätzliche Fenster auf. Diese *Pin-holes* sind nur mit Aufwand zu vermeiden. Die Pin-hole-Dichte hängt von vielen Einflüssen ab.

4.3.2. Standardmethode für Masken

Die Vorlagen werden mit einem elektronischen, digital gesteuerten *Koordinatentisch* (Koordinatograf), einer Arbeitsfläche von etwa 1 m², erzeugt. Als Arbeitsmaterial wird meist Polyester- oder Polycarbonatfolie mit einer Dicke von 30 bis 200 µm verwendet. Bei Zimmertemperatur werden hierbei Stabilitäten von 10^{-5} je K bzw. 1% Luftfeuchteänderung erreicht. Die Folien sind mit einem speziellen rötlichen Lack begossen. Ein Handelsname für derartige beschichtete Folien ist Rubylith. Der

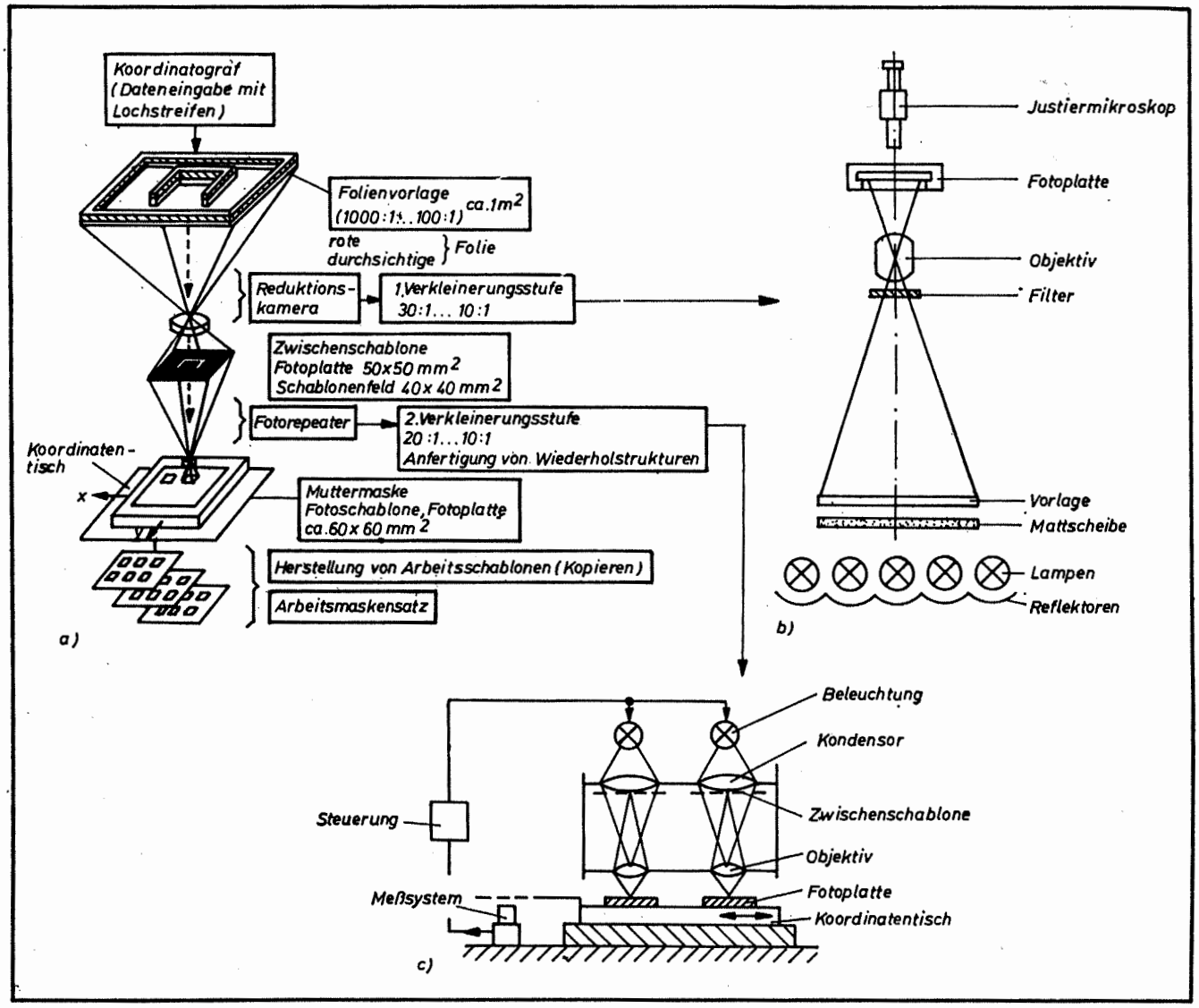


Bild 4.3.2. Schema der Standardmaskenherstellung

a) Schematische Darstellung
b) Prinzip der Reduktionskamera für den Übergang von der Vorlage zum reticle
c) Prinzip des Fotorepeaters mit zwei angedeuteten Barrels

Koordinatentisch trägt ein Messer, das den Lack einritzt. So sind Bereiche des Lackes abziehbar (*Stripping-Technik*), und die Folien werden dort durchsichtig. Die rote Lackbeschichtung soll einerseits für das Auge etwas durchsichtig sein, um bei einem Maskensatz von Vorlagen die Konturendeckung durch Übereinanderlegen kontrollieren zu können. Sie soll anderseits im grünen bis UV-Bereich stark absorbieren, damit ein hoher Kontrast für die weiteren Prozesse existiert. Die Genauigkeit guter Koordinatentische beträgt etwa $\pm 10\,\mu m$. Die Schnittgeschwindigkeit liegt um 100 mm/s. Die *Zwischenschablone* (*Reticle*) wird mittels *Reduktionskameras* (z. B. Bild 4.3.2b)) auf $^1/_{10}$ bis $^1/_{100}$ verkleinert. Für die Zwischenschablonen werden meist *gute Glasplatten* mit Silberhalogenbeschichtung und Kantenabmessungen bis etwa 100 mm verwendet. Die Reproduzierbarkeit muß hier bereits um 1 µm liegen. Es werden Spezialoptiken und monochromatisches Licht verwendet.

Die Herstellung der *Originalschablonen* erfordert die höchste Präzision. Hierbei müssen die Zwischenschablonen wiederholt mehrfach und verkleinert abgebildet werden. Dabei müssen zugleich auch die kleinen Linienbreiten im µm-Bereich und die entsprechenden noch kleineren Lagetoleranzen realisiert werden. Hierzu wurden extrem genaue *Fotorepeater* entwickelt. Sie enthalten gleichzeitig und festverkoppelt etwa sechs bis neun gleichartige optische Systeme (Barrels). Mit jedem Barrel wird je ein Reticle des Maskensatzes verarbeitet. Nur mit dieser Methode kann die notwendige Konturendeckung zwischen den endgültigen Masken realisiert werden. Mit genau definierten Verschiebungen wird dann nach jeder Belichtung der Koordinatentisch, der alle endgültigen Fotoplatten des Maskensatzes trägt in x- und y-Richtung verschoben. So entstehen die verkleinerten Wiederholungen des Reticles. Daher auch der Name dieses Prozesses „*step and repeat*" (schrittweise bewegen und wiederholen) und der Name dieser Geräte „Fotorepeater".

Diese endgültigen Masken werden aber noch nicht zum Belichten der Wafer verwendet. Hierzu dienen von ihnen im Kontaktprozeß hergestellte *Duplikate*. Diese *Arbeitsmasken* unterliegen nämlich einem sehr hohen *Verschleiß*. Durch den notwendig engen Kontakt mit dem Wafer und seinen Unebenheiten infolge vorangegangener Strukturierung werden sie so stark beansprucht, daß sie z. T. nur zehnmal verwendbar sind.

Zur relativ einfachen Positionierung der Masken werden Sonderkennzeichen verwendet, die keine Bauelementefunktionen beinhalten und auf allen Masken eines Maskensatzes genau zueinander passen und an günstigen Stellen liegen müssen. (Vgl. Bild 7.5.2.)

4.3.3. Leicht abgewandelte Techniken

Der hohe Maskenverbrauch bei der Belichtung der Wafer hat nach anderen Ausweichmethoden suchen lassen. Eine davon ist die Belichtung im genau definierten kleinen Abstand zwischen Wafer und Maske. Dieses Verfahren hat eine gewisse Bedeutung unter dem Namen *Proximity-* bzw. *Abstandsbelichtung* erhalten.

Eine andere Methode ist die 1:1-*Projektion* der Masken auf den Wafer. Hierbei werden bezüglich Auflösung und Bildfläche an die Projektionsoptik höchste Forderungen gestellt. Sie arbeiten, wie Bild 4.3.3a) zeigt, alle nahe an der theoretischen Grenze. Die entsprechenden Beziehungen sind im Bild eingetragen. Diese Methode reicht bei den üblichen Wafergrößen (ca. 100 mm) bis zu Linienbreiten um 2 bis 3 µm. Für Bilddurchmesser um 20 mm lassen sich Strichbreiten um 1,5 µm erreichen. Bei höheren Auflösungen, also größeren Wafern, und/oder feineren Linien bewährt sich ein *Step-repeat-Prozeß direkt auf dem Wafer*. Dies deutet schematisch Bild 4.3.3b) an. Dieser

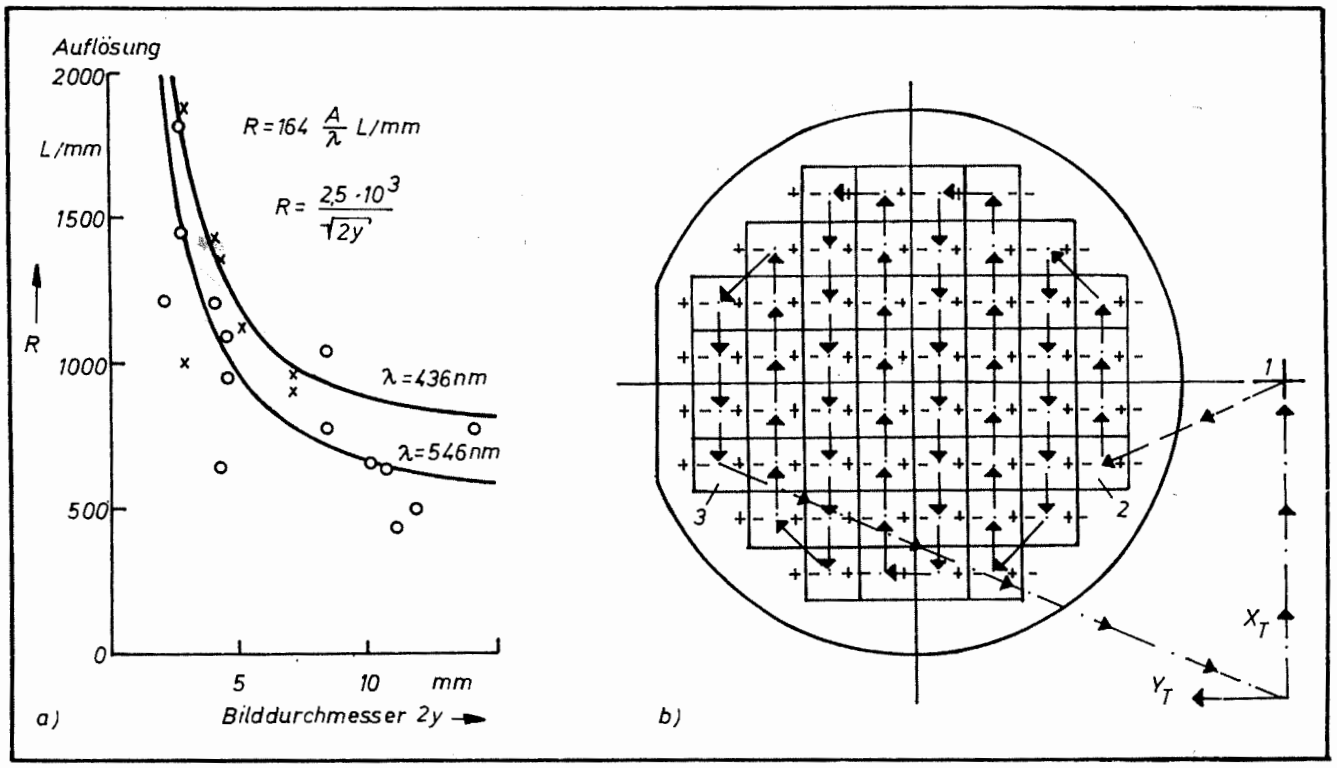


Bild 4.3.3. a) Grenzen der Auflösung bei Projektionsoptiken
b) Ausweg für große Wafer durch step-repeat direkt auf dem Wafer in schematischer Darstellung. Oft werden nur 5 bis 9 Teilfelder verwendet

Prozeß ist also entsprechend der Anzahl der Wiederholungen und der dazu notwendigen Justagen um vieles langsamer als ein Einschrittprozeß auf dem Wafer.

Verbesserte Masken ergeben sich, wenn nicht mehr Fotoemulsionen, sondern geätzte metallische oder Oxidfolien verwendet werden. Dieser Prozeß hat sowohl bei den Reticles als auch bei den Muttermasken große Bedeutung gewonnen. Insbesondere haben sich *Chromfolien* mit Dicken von 0,1 bis 0,2 μm bewährt, die mittels der Fotolithografie geätzt werden. Sie besitzen jedoch Nachteile bei der Justierung (nicht durchsichtig) bez. der Pin-hole-Dichte und infolge von Störungen durch stehende Lichtwellen mit Interferenzmustern.

4.3.4. Elektronenstrahllithografie

Die Entwicklung der Lithografie führte zu immer höheren Auflösungen, d. h. kleineren Strichbreiten. Dies ist in Bild 4.3.4 dargestellt. Daraus wird deutlich, daß bereits die *Beugungsgrenze des sichtbaren Lichtes* (etwa 1 μm) zumindest nahezu erreicht ist. Sie kann nur mit Elektronenstrahlen bzw. Röntgenstrahlen unterschritten werden. Bei der Elektronenstrahllithografie wird der im Abschnitt 6.2.1. beschriebene Ablenkmechanismus für Elektronenstrahlen verwendet. Er ermöglicht prinzipiell Auflösungen bis um oder unter 1 nm. Die Elektronenstrahllithografie wird heute auf *zwei Wegen* angewendet. Entweder werden mit ihr die *Reticles hergestellt*, oder aber sie wirkt *direkt auf den Resist* des Wafers ein.

Im ersten Fall wird wieder an die direkte Chipbelichtung in einem Prozeß oder im Chipstep angeknüpft. Hierbei entstehen vor allem die Vorteile des rechnerunterstützten Entwurfs, ohne über Koordinatograf usw. zu gehen. Die Methoden des rechnergestützten Entwurfs (CAD = computer aided design) können dann optimal genutzt werden.

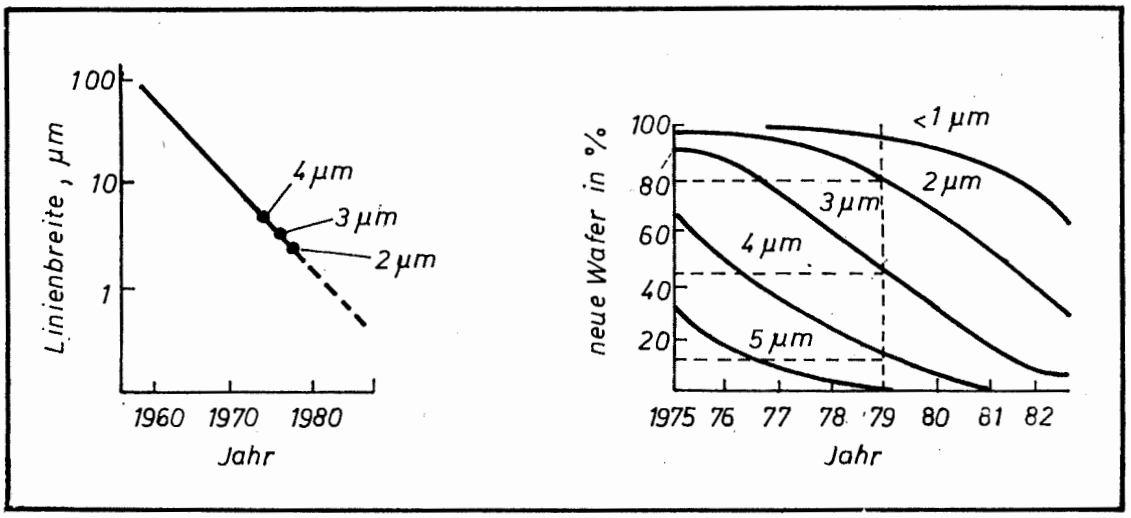


Bild 4.3.4. Zur Auflösung der Fotolithographie im Laufe der Zeit
a) Jeweils erreichte kleinste Linienbreite
b) Relativer Anteil der jeweils neu erzeugten Schaltungen bestimmter Linienbreite

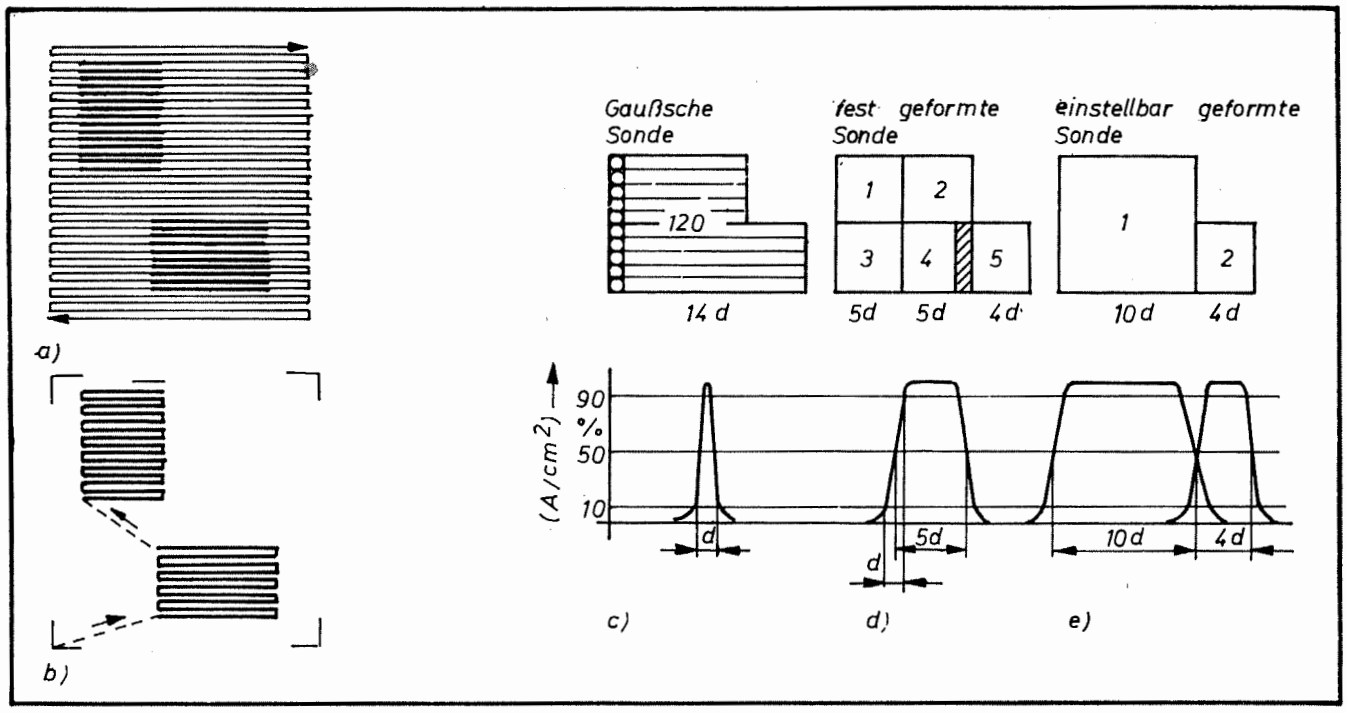


Bild 4.3.5. Zur Effektivitätssteigerung bei der Elektronenlithografie
a) Rasterscan b) Vektorscan
c) GAUSS-Strahl in Vektorscan
d) Formstrahl konstanter Fläche
c) Steuerbarer Formstrahl

Im zweiten Fall bleibt der Vorteil des CAD voll erhalten, aber es entfallen zusätzlich alle Maskenprozesse. Dieser gewaltige Vorteil wird natürlich mit anderen Nachteilen erkauft. Dies und Auswege dazu seien am Beispiel des Bildes 4.3.5 erklärt.

Der Elektronenstrahl muß für die Lithografie so fein fokussiert werden, daß er die feinsten Linien schreiben kann. Er besitzt dann ein entsprechendes GAUSS-Profil (Bild c), unterer Teil). Am einfachsten ist es nun, wenn der Elektronenstrahl über den Wafer, *ähnlich* wie beim *Fernsehbild*, geführt wird. Er wird an den Stellen hell getastet, wo belichtet werden soll. Dieses Verfahren besitzt viel *zeitliche Redundanz*, die auch dann noch groß ist, wenn Hin- und Rücklauf wie etwa in Bild a) genutzt werden. Eine effektivere Methode ist das *Vektorscanverfahren* von Bild b), wobei auf dem jeweils

kürzesten Wege nur noch die zu belichtenden Flächen angesteuert werden. Eine weitere Effektivitätssteigerung besteht darin, den *Strahl auf ein Rechteck zu formen.* Der dadurch erreichbare Zeitgewinn wird im Vergleich von Bild c) zu d) deutlich. Ein noch höherer Zeitgewinn wird erreicht, wenn auch noch die Größe des *quadratischen Feldes steuerbar* ist (Bild e)). Nach diesem Prinzip der variablen Strahlformung arbeiten die moderneren Elektronenstrahlbelichtungsverfahren.

4.3.5. Röntgenstrahllithografie

Die Röntgenstrahllithografie verlangt zwar stets eine Maske auf dem Chip. Hier ist sie also gegenüber der schon vorhandenen Elektronenstrahllithografie im Nachteil. Sie besitzt aber *Vorteile,* die sie für die Zukunft — denn noch gibt es keine effektiven Anlagen — interessant machen:

— Es entfällt das Vakuum. Röntgenstrahlen können, nachdem sie das Berylliumfenster verlassen haben, unter Atmosphärendruck, z.B. unter Helium, auf die Scheibe einwirken.
— Röntgenstrahlen werden nicht durch magnetische oder elektrische Felder beeinflußt. Es erfolgt exakt eine Belichtung als Schattenwurf von der Maske.
— Röntgenstrahlen durchdringen Staubpartikeln und dünne Schutzschichten.

Insbesondere die beiden letzten Fakten lassen geringe Fehler erwarten. Zwei *Hauptprobleme* begrenzen zur Zeit jedoch die Anwendung erheblich:

- die geringe Empfindlichkeit der Röntgenresiste
- der große Strahlenfleck von ca. 5 mm ⌀ auf der Drehanode von Röntgenröhren entsprechender Leistung.

Es wird diskutiert, ob bessere Möglichkeiten sich mittels kerntechnischer Speicherringe realisieren lassen.

4.3.6. Zu physikalischen Grenzen

Heute besitzen Halbleiterstrukturen laterale Auflösungen im Bereich von etwa 1 bis 3 μm. Die Auflösungen in der Tiefe betragen dagegen schon, insbesondere bei der Ionenimplantation, Werte ab 0,1 μm. Übliche Dotierungskonzentrationen liegen bei 10^{15} je cm^3. Der mittlere Abstand von Dotanden beträgt etwa 0,1 μm. Die Lithografieverfahren sind also aus dieser Sicht nicht mehr weit von sinnvollen Grenzen entfernt. Die höchstauflösende Lithografie muß folglich zu höheren Dotierungen übergehen. Damit sind aber auch kleine Durchbruchspannungen (Betriebsspannungen) bedingt. Da sie bei 10^{15} Dotanden je cm^3 bei ca. 100 V liegen, sind also in dieser Hinsicht noch gewisse Reserven vorhanden.

4.4. Epitaxie-Planartechnik

In diesem Abschnitt sollen beispielhaft die Prozesse der Strukturierung und Schichttechniken so zusammengefaßt werden, wie sie bei *Erzeugung von Transistoren, Widerständen, Schaltkreisen* usw. auftreten. Dabei werden die *Tiefenabmessungen* zur Erlangung einer besseren Übersicht *etwa 5- bis 10fach überhöht.* Weiter werden verschiedene Details, wie Böschungen bei den Ätzungen, nicht dargestellt. Die Prozesse werden

vereinfacht. In diesem Sinne werden u. a. *nicht erwähnt Kontrollschritte*, Reinigungsprozesse, Temperungen, Rückbeschichtungen usw.

4.4.1. Strukturierung eines bipolaren Transistors

Den schematischen *Ablauf für einen bipolaren Transistor* zeigt Bild 4.4.1. Nachdem die p-leitende Siliciumscheibe gereinigt und anschließend an der Oberfläche oxydiert ist, wird die Fotolackschicht aufgetragen und mit der Maske für das begrabene Gebiet (buried layer) belichtet. Den Zustand nach der Öffnung des Fensters zeigt Bild a), während b) das Ergebnis einer Arsen- oder Antimondiffusion darstellt. Es ist ein gut leitendes, also n^+-dotiertes Gebiet entstanden. Nach Entfernen des Fotolackes und der Oxidschichten läßt man eine n-leitende (As- oder P-dotierte) 5 bis 10 μm dicke Epitaxieschicht aufwachsen. Das Ergebnis zeigt Bild c). Es erfolgen Oxydation, Auftragen des Fotolackes und der zweite Maskenprozeß mit Öffnung der Fenster für den Isolierrahmen (Bild d)). Der Isolierrahmen dient dazu, den Transistor gegenüber anderen, auf dem gleichen Chip befindlichen Bauelementen durch einen gesperrten pn-Übergang zu isolieren. Das entsprechend eindiffundierte niederohmige p^+ wird durch die ganze Epitaxieschicht bis zum Grundmaterial eindiffundiert. Den Zustand nach der Entfernung der Fotolackschicht und erneuter Oxydation zeigt Bild e). In ganz analoger Weise werden nun mit weiteren Masken zunächst das Basisgebiet (p-dotiert; Bild f)) und dann das Emittergebiet (n^+) über Fenster eindiffundiert (Bild g)). Damit ist im Prinzip der Transistor fertig. Für seinen Betrieb müssen jetzt noch die

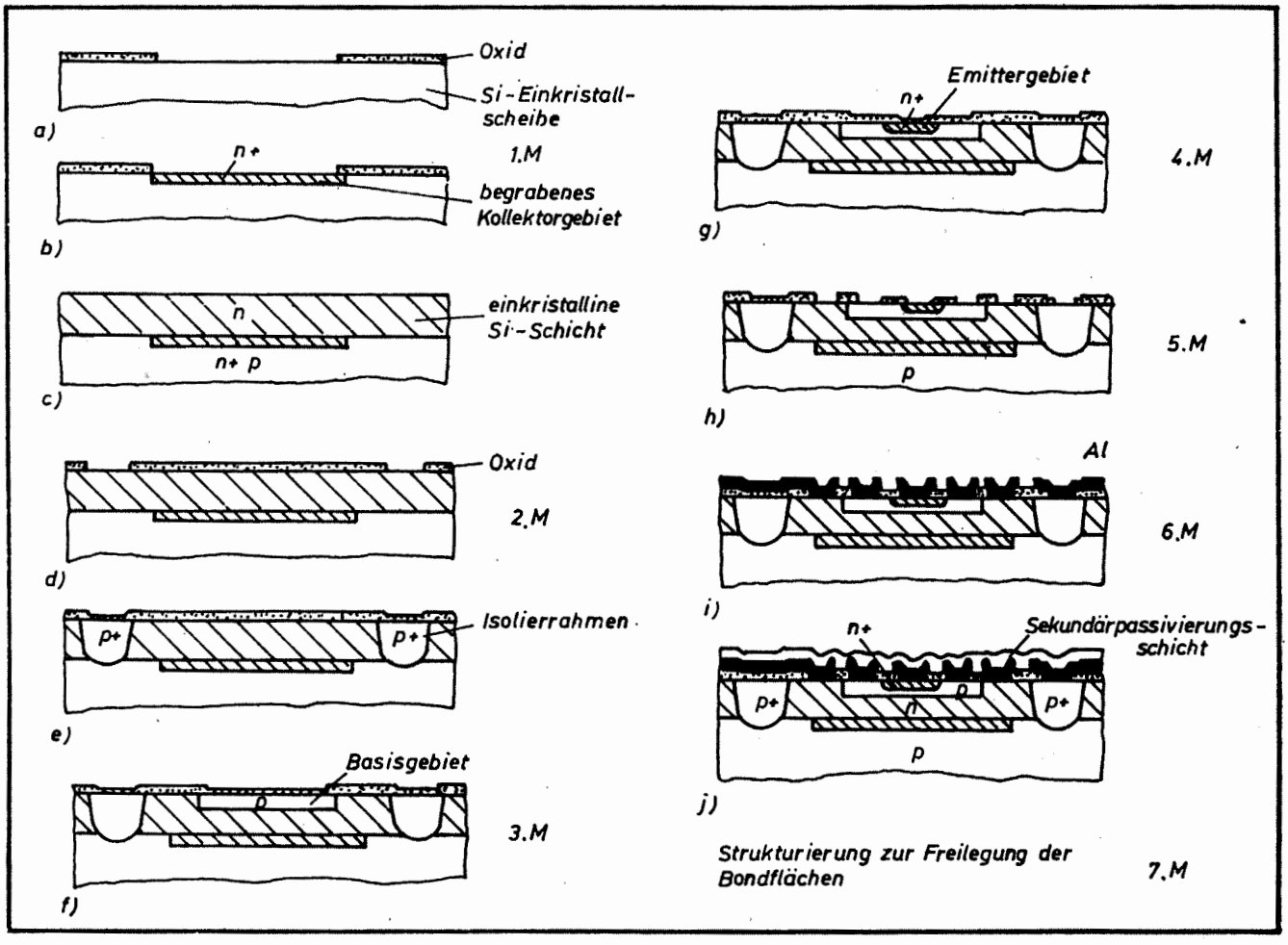


Bild 4.4.1. Vereinfachte Prozeßabfolge bei einem bipolaren Epitaxie-Planartransistor

Anschlüsse vorgesehen werden. Hierzu werden wiederum mittels Maske die Kontaktfenster geöffnet (Bild h)), der Lack entfernt und der Wafer vollständig mit Al bedampft. Fotolithografisch werden nun mit der sechsten Maske, der Leitmaske, alle nicht benötigten Aluminiumflächen weggeätzt (Bild i)). Dann wird der gesamte Wafer mittels Gasphasenabscheidung mit einer Silikatglas- (meist p-dotiertes SiO_2) oder Si_3N_4-Schicht abgedeckt (Bild j)). In dieser Schicht werden schließlich jene Metallflächen freigelegt, die für das später zu behandelnde Bonden besonders benötigt werden.

In Bild 4.4.2a) ist schließlich noch eine etwa *maßstabgerechte Darstellung* eines derartigen Transistors gegeben, während Teil b) die Oberfläche und c) das Störstellenprofil unter dem Emitter zeigt.

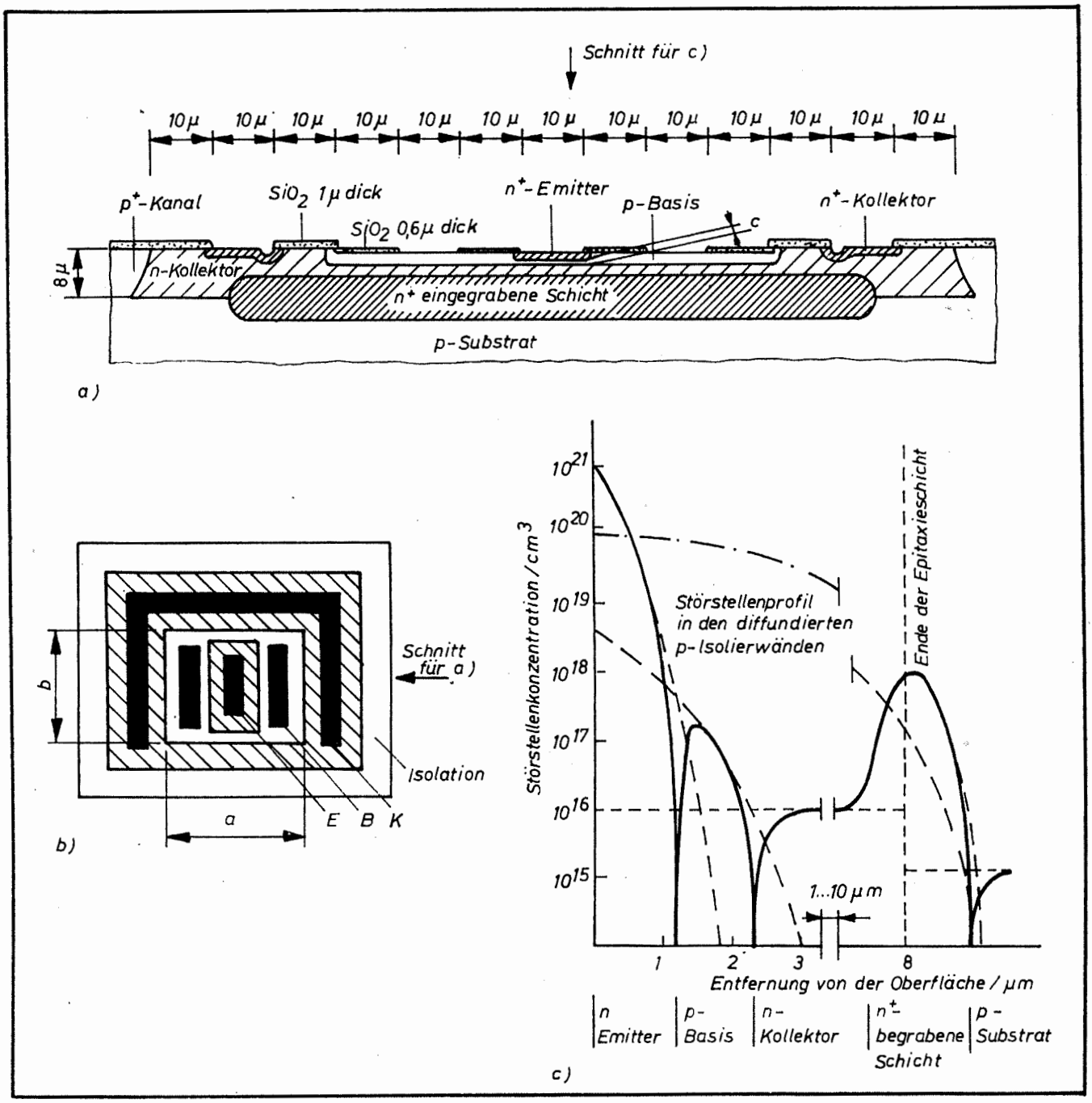


Bild 4.4.2. Reale Abmessungen eines Epitaxie-Planartransistors

a) Schnitt zur Darstellung der Tiefenabmessungen

b) Oberflächenaussehen. Für die Abmessungen gilt: Die Fläche ab bestimmt die zulässige Leistung. Die Basisweite c bestimmt die obere Grenzfrequenz etwa gemäß $f \approx \dfrac{8\,\mathrm{GHz}}{c/\mu\mathrm{m}}$

c) Störstellenprofil unterhalb des Emitters

Etwa 1980 waren folgende *typische Werte* erreichbar: Für den Emitter bzw. die Basis werden Eindringtiefen von 0,5 bis 2 bzw. 3 μm realisiert. Die Basisweite liegt bei 0,5 bis 1 μm.

4.4.2. Strukturierung eines MOSFET

Die Strukturierung eines p-MOSFET vom Anreicherungstyp zeigt Bild 4.4.3. Nach gründlicher Reinigung des Wafers wird das sogenannte Feld- oder Dickoxid erzeugt (a). Mit der ersten Maske werden die Fenster für Source und Drain (b) zur p^+-Diffusion geöffnet (c). Nach erneuter Oxydation wird das Gatefenster geöffnet (d). Hierauf

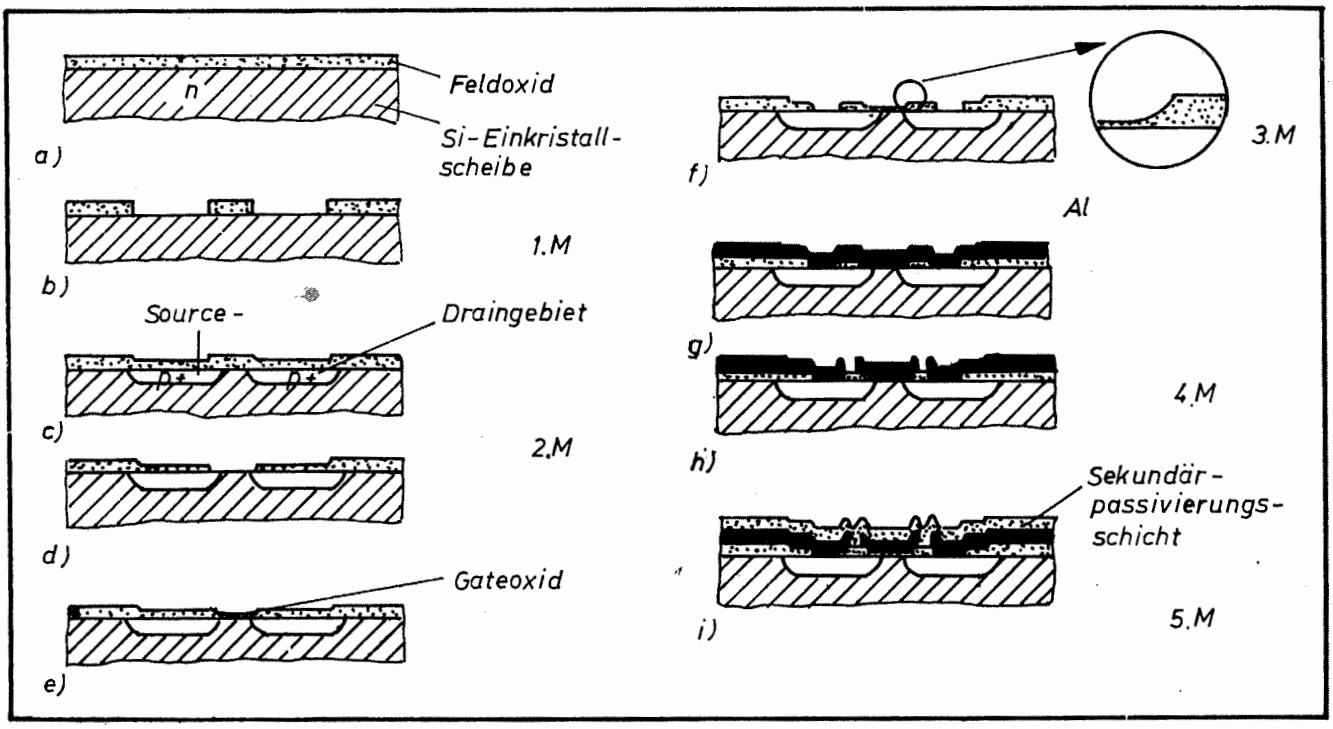


Bild 4.4.3. Vereinfachte Prozeßfolge für einen p-Kanal-MOSFET vom Anreicherungstyp

wird eine sehr dünne Oxidschicht (ca. 40 mm) aufgebracht (e). Anschließend werden die Fenster für die Source- und Drainkontakte geöffnet (f) und die gesamte Oberfläche metallisiert (g). Diese Oberfläche wird mit der vierten Maske strukturiert (h). Es folgt wieder die Sekundärpassivierungsschicht und das Fensteröffnen für die Bondinseln. Wie bereits geschildert, ist für die Grenzfrequenz die Kanallänge entscheidend. 1980 waren nach dieser Technologie Kanallängen um 1 μm realisierbar.

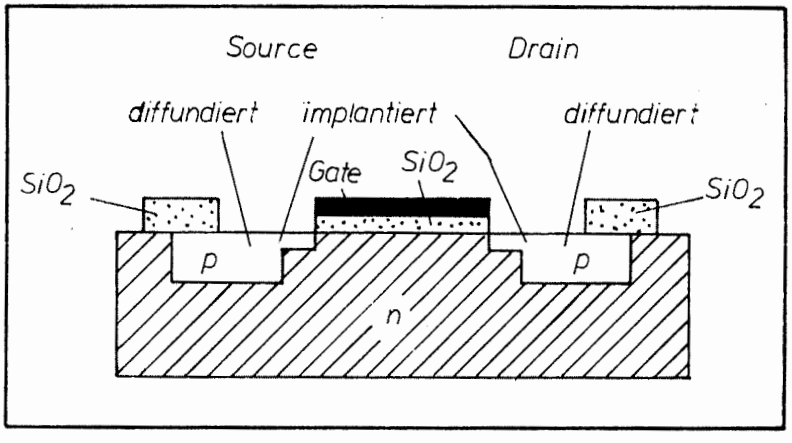


Bild 4.4.4. Schematische Darstellung eines mit Ionenimplantation hergestellten MOSFET

Das Prinzip der selbstjustierenden Maske über das Gate bei Ionenimplantation zeigt Bild 4.4.4. Hierdurch sind bessere Eigenschaften und kürzere Kanallängen infolge idealer Konturdeckungen zu erreichen.

4.5. Herstellen der Bauelemente

4.5.1. Übergang zu den Chips

Nach den Strukturierungsprozessen liegen auf jedem Chip eine Vielzahl von Bauelementen bzw. integrierten Schaltungen vor. Dies ist in Bild 4.5.1a) angedeutet. Hier kann die vollständige *Funktionstüchtigkeit mit speziellen Sonden*, die auf die Bondinseln gesetzt werden, erstmals getestet werden (Bild 4.5.1b)). Die nicht brauchbaren Strukturen werden dabei meist mit einem Lacktropfen gekennzeichnet. Im nächsten Schritt müssen die kleinen Elemente *vereinzelt* werden. Dazu wird der Wafer zwischen den Strukturen *geritzt* (Bild 4.5.1c)). Die Ritzbreite beträgt etwa 0,1 mm. Optimale Ritzrichtungen, die den geringsten Bruch ergeben, sind gemäß Bild 4.5.1d) durch die

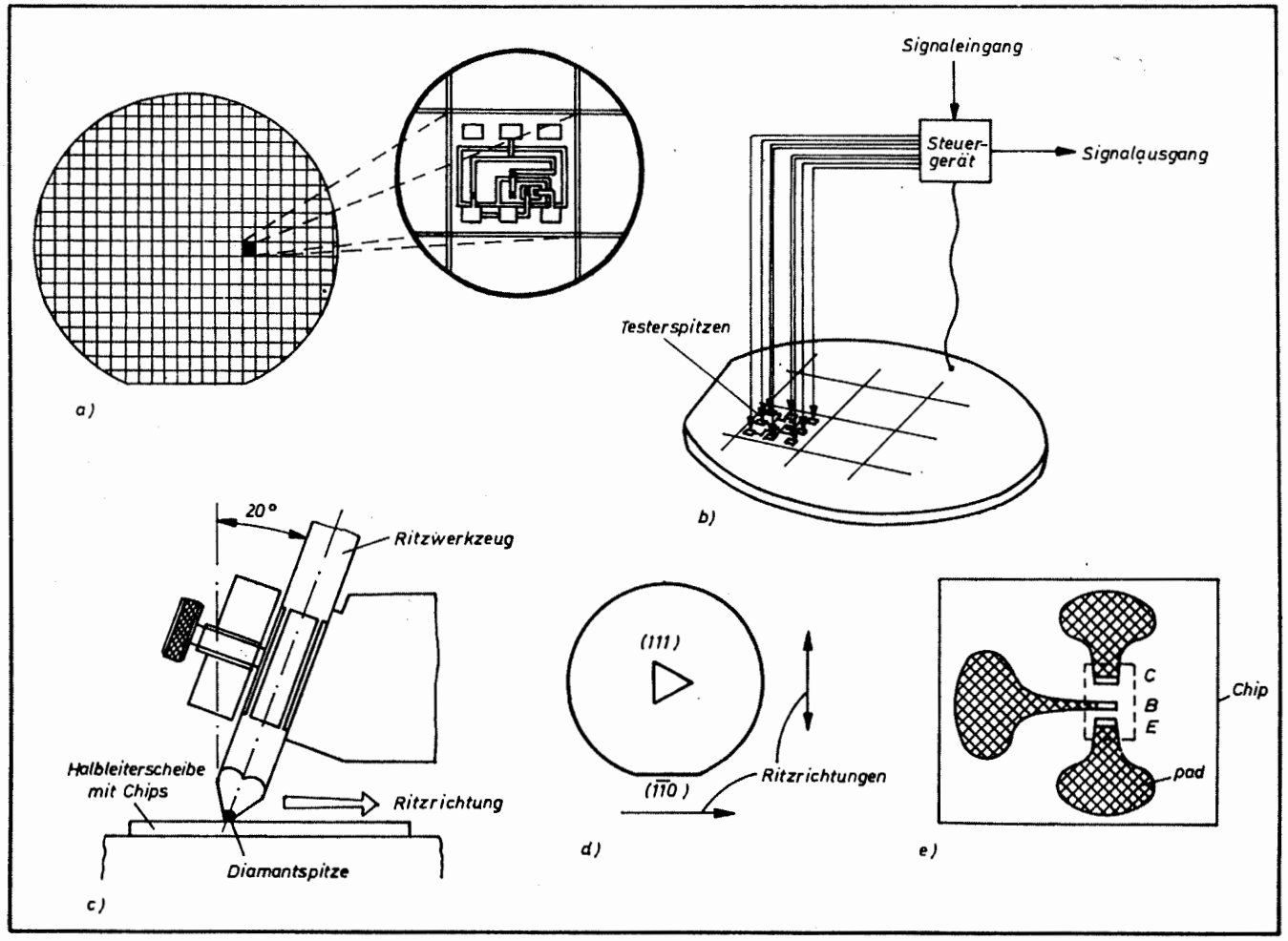


Bild 4.5.1. Übergang zu den Chips

a) Anordnung der Chips auf dem Wafer
b) Testung eines Chips auf dem Wafer
c) Prinzip des Diamantritzens beim Wafer
d) Vorzugsrichtung für das Ritzen und Brechen des Wafers in die Chips
e) Beispiel eines Miniaturtransistors, der im Vergleich zur minimalen Chipfläche klein ist. Die Bondinseln sind hier mit pad bezeichnet

Phase am Wafer bereits vorgezeichnet. Die Größe der Chips liegt je nach Komplexität der Schaltung und Leistung der Bauelemente im Bereich von 0,5 × 0,5 bis 5 × 5 mm². Bei den kleinsten Transistoren ist ihre Nutzfläche so klein, daß der Hauptteil des Chips für die *Bondinseln* verwendet wird (Bild 4.5.1e)).

Nach dem Ritzen wird der Wafer auf eine flexible Unterlage geklebt und durch deren Verbiegen gebrochen. Hierdurch verbleiben die einzelnen Chips noch in den richtigen Nachbarschaftsbeziehungen. Dies ist besonders günstig für die Hybridtechniken (9.4.). Es hat aber auch Vorteile für das Aussortieren der fehlerhaften Chips.

4.5.2. Gehäuseeinbau

Es gibt heute eine solche Vielzahl von Gehäusen für Halbleiterbauelemente, daß hier nur drei Beispiele herausgegriffen werden sollen. Sie werden insbesondere bei integrierten Schaltkreisen verwendet, zeigen aber die wichtigsten Prinzipien (Bild 4.5.2).

Beim Einbau in ein *Rundgehäuse* wird der Chip zunächst auf die Bodenfläche gebondet. Hierzu wird eine dünne Goldschicht verwendet. Bei 370 °C bildet Gold mit Silicium ein Eutektikum. Dann werden von den Bondinseln des Chips Drähte zu den

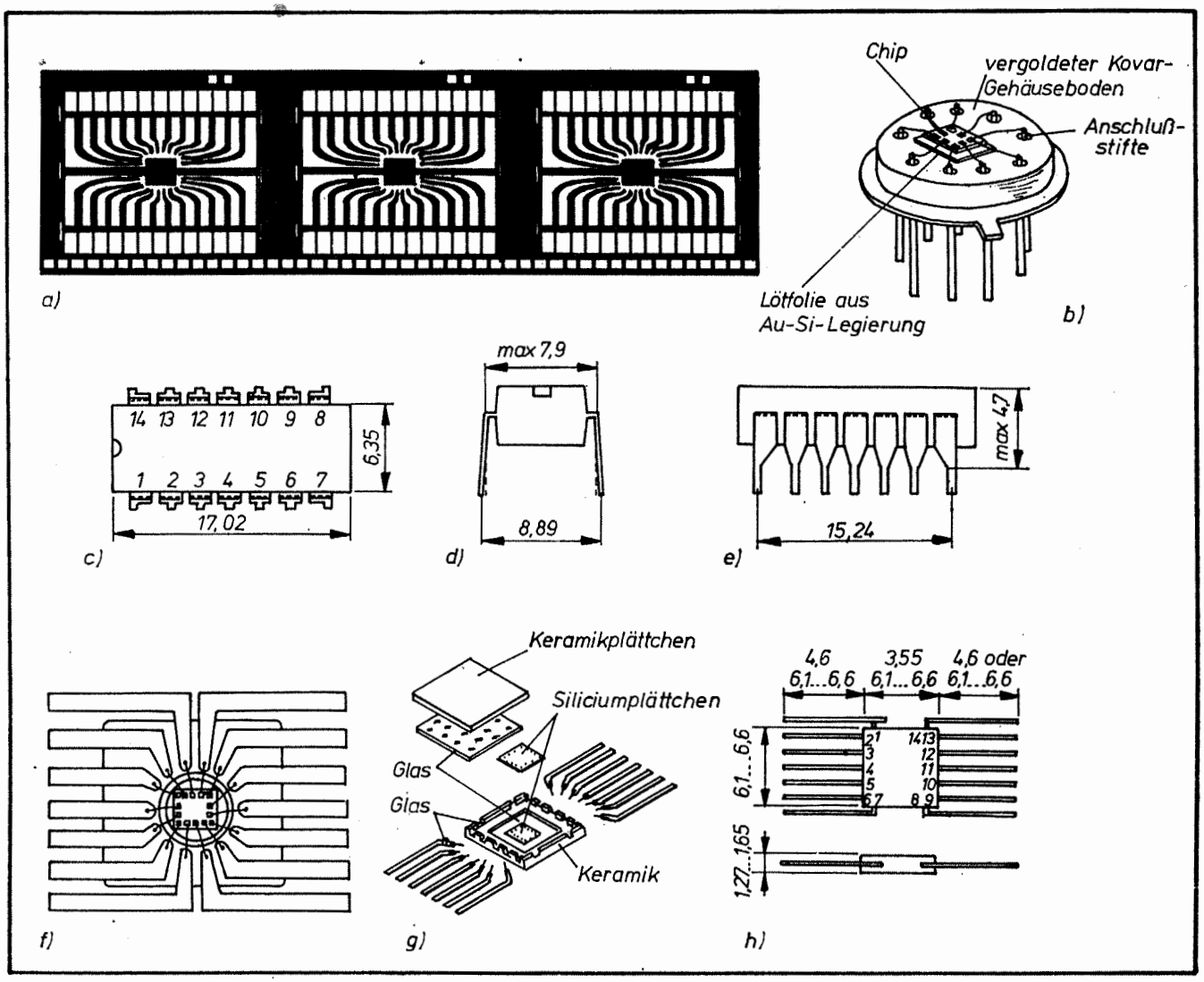


Bild 4.5.2. Beispiele vom Gehäuseeinbau der Chips

a) Der für DIL typische Kamm, hier mit 24 Anschlüssen
b) Chip auf dem Boden eines Rundgehäuses
c) bis e) Abmessungen und Aussehen des typischen DIL-14-Gehäuses
f) bis h) Bondung, Explosionszeichnung und Abmessungen von Flachgehäusen (Flat-pack)

Gehäuseanschlüssen geführt (s. nächster Abschnitt). Nachdem dies erfolgt ist (Bild b)), wird über den Boden die Kappe gelötet.

Bei den *DIL-Gehäusen* (dual-in-line — zwei Anschlüsse in je einer Linie) wird als Ausgangsträger sogenanntes *Kammaterial* verwendet (a). In der Mitte wird der Chip wiederum eutektisch mit Goldlot befestigt, und von den Bondinseln werden zu den inneren Kammenden, ähnlich wie beim Rundgehäuse, die dünnen Bonddrähte gezogen. Dann wird die Mitte des Kammes mit Kunststoff umhüllt. Durch Ausstanzen des äußeren Teils des Kammes und Umbiegen der Anschlüsse gemäß Bild c) bis e) entsteht der fertige Schaltkreis.

Die dritte hier zu besprechende Variante ist das *Flat-pack* oder *Flachgehäuse*, wie es Bild h) zeigt. Es besteht aus Keramikteilen, die mit Glaslot verbunden werden. Der Chip wird in das teilweise vormontierte Gehäuse mit Glaslot bei etwa 450 °C aufgeschmolzen. Das gleiche erfolgt für die Drahtenden (g). Danach werden mit Draht wieder die Anschlüsse gebondet, und anschließend wird das Gehäuse zugelötet.

4.5.3. Drahtbonden

Das Drahtbonden ist heute noch immer einer der aufwendigsten, fehleranfälligsten und meist nur manuell durchführbaren Prozesse. Jede Verbindung erfordert dabei zweimaliges Bonden, einmal am Chip und einmal an den Anschlüssen des Gehäuses. Von den verschiedenen Möglichkeiten des Bondens hat sich heute im wesentlichen das *Ball-* (*Kugel-*) *Bonden* durchgesetzt. Es ist daher auch in Bild 4.5.3a) ausführlicher

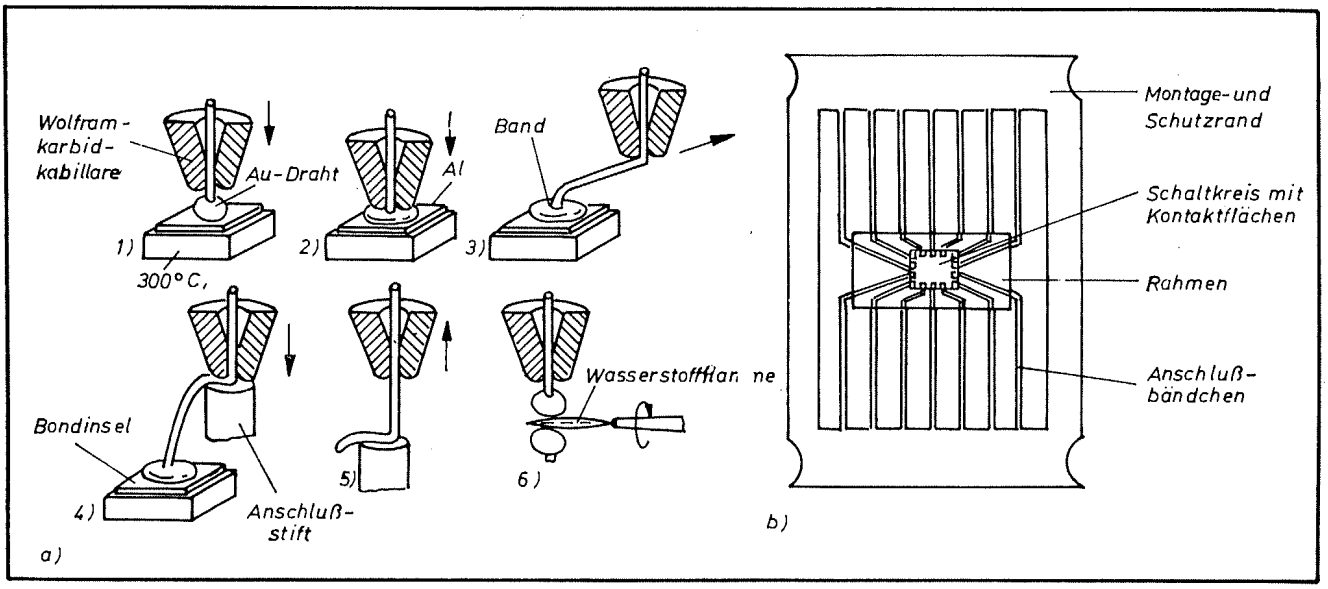


Bild 4.5.3. a) Die sechs Abfolgen für das typische Ball-Drahtbonden
b) Vereinfachung des Bondens mit einem zusätzlichen Kontaktrahmen am Beispiel des Flat-pack

dargestellt. Die Drähte werden mittels *Thermokompression* bei ca. 300 °C und einem Druck um 1000 N/mm² aufgepreßt. Die Verbindung der Drähte mit den Bondinseln hält danach einem Zug von etwa 0,1 ... 0,2 N stand. Die Drähte sind meist *Golddrähte* (selten Aluminium) mit einem Durchmesser um 10 bis 20 µm. Gold hat neben vielen Vorteilen den Nachteil, daß es in Verbindung mit dem Aluminium der Bondfläche zur Zerstörung des Kontaktes durch die sogenannte *Purpurpes* führen kann. Die Bondstellen stellen generell eine häufige Störungsursache beim Ausfall von Bauelementen dar. Infolge der Belastung beim Bonden können z. B. in unmittelbarer Nähe der Bondstelle *Mikrorisse* entstehen. Ferner können Wärme, mechanische Erschütte-

rungen und Ausdehnung des Plastwerkstoffes, in den die Chips vielfach eingegossen werden, zum Lösen der Bondstelle führen.

Neben dem hier betrachteten Ballbonden haben drei weitere Verfahren gewisse Bedeutung: das *Schneiden-, Scheren- und Ultraschallbonden.*

Ursprünglich und z. T. auch noch heute erfolgt das Bonden manuell unter dem Mikroskop. Dabei ist etwa eine Bondung je Sekunde möglich. Bei einem 14-DIL-Gehäuse (14 Anschlüsse) können dann nur ca. 50 Bauelemente je Stunde und Person gefertigt werden. Es war daher notwendig, nach *Automatisierungsmöglichkeiten* zu suchen. Hauptsächlich hat sich hier das Filmbonden bewährt. Gemäß Bild 4.5.3b) wird zwischen dem Chip und der Spinne ein weiterer kleiner Rahmen (lead-frame) statt der Bonddrähte verwendet. Diese kleinen Rahmen befinden sich auf einem Kinofilm und können so magaziniert herangeführt werden. Der Chip muß für diese Methode an den Kontaktstellen besonders vorbereitet sein. Sie sind größer und mit kleinen Metallhöckern versehen. Auf dem Leitaluminium folgt eine dünne Chromschicht und dann 25 µm hoher Kupferkontakt. Auf die orientierten Chips wird zunächst der Rahmen mit einem Mal kontaktiert. Dann wird dieser Rahmen in die Spinne gefügt. So werden ca. 2000 Bauelemente je Stunde erreicht.

4.5.4. Plastverkapselung

Werden die üblichen *Kosten* für die Herstellung einer integrierten Schaltung z. B. in einem runden Metallgehäuse zusammengestellt, so gilt etwa:

Chipherstellung	1,9%	mit Ausbeute 0,7
Funktionsprüfungen	7,1%	mit Ausbeute 0,5
Montage und Gehäuse	59%	mit Ausbeute 0,9
Endmessung	32%	mit Ausbeute 0,7 .

Dies zeigt, daß die Montage und das Einbauen in Gehäuse über die Hälfte der Kosten ausmachen (vgl. auch hierzu Bild 4.5.4). Daher bringt ein Eingießen in Plaste erhebliche Kostensenkung. Sie setzt aber voraus, daß eine Passivierung des Chips z. B. durch SiO_2 erfolgt ist. Die Verkaufspreise derartiger Bauelemente liegen daher um etwa 60 bis 70% unter denen, die mit echten Gehäusen, z. B. Metall oder

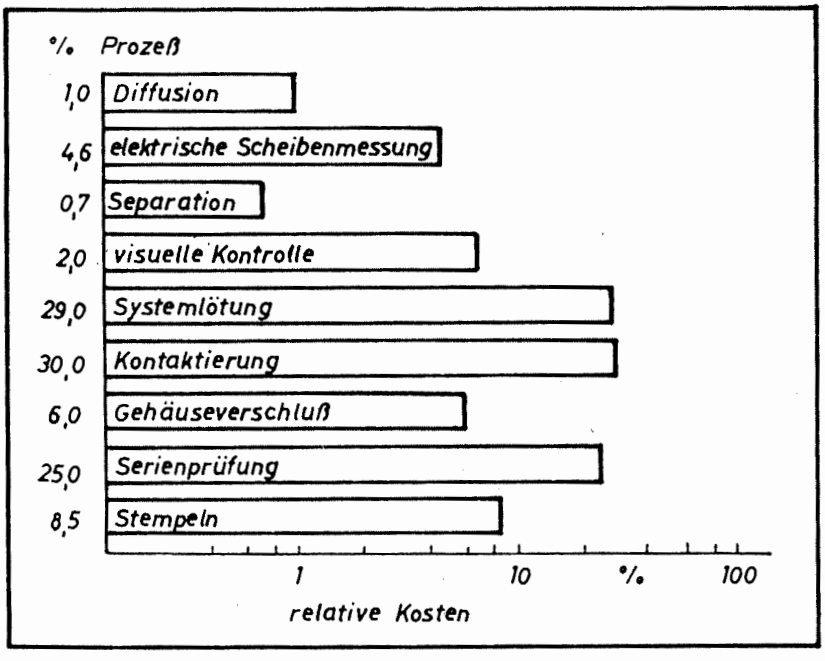


Bild 4.5.4. Relative direkte Lohnkosten je Halbleiterbauelement, aufgeteilt nach Teilfertigungsprozessen (lies 0,5 statt 8,5)

Keramik versehen sind. Für die *Kunststoffverkapselung* können folgende *Vorteile* angegeben werden:

— niedrige Kosten des Kunststoffes,
— teilautomatische Montage,
— Einbetten bei niedriger Temperatur,
— für optoelektronische Bauelemente günstiger Brechungsindex um 1,5.

Der *Hauptnachteil* ist die schlechtere *Wärmeleitung*. Auf die Zuverlässigkeit haben sich, sofern die Bondung einwandfrei erfolgt, keine Nachteile ergeben. Als Kunststoffe werden heute vorwiegend *Epoxydharze* verwendet. Ihre *guten Eigenschaften* sind

— Leichtflüssigkeit bei der Verarbeitung,
— geringere Aushärtetemperaturen und -zeiten,
— gute Verträglichkeit mit Halbleitermaterialien,
— gute elektrische und optische Eigenschaften,
— hohe Langzeitstabilität.

Das Einbetten erfolgt nach dem Gießharz- oder Fließpreßverfahren. Das Fließpreßverfahren erfolgt bei 150 bis 300 °C und Drücken von 0,2 bis 1,5 · 10^5 Pa. Eine Nachhärtung von einigen Stunden bei 150 °C führt zu stabilen chemischen Eigenschaften. Heute werden etwa die Hälfte aller Halbleiterbauelemente in Plast vergossen.

5. Bauelemente – Eigenschaften und Grundschaltungen

Die Epitaxie-Planartechnologie läßt sehr vielfältige und damit auch komplexe Strukturen realisieren. Um sie jedoch zu entwerfen, realisieren und anwenden zu können, müssen Grundprinzipien verstanden werden. Diesem Zwischenziel dient dieses Kapitel. Es behandelt den Aufbau, die wesentlichen Eigenschaften und die einfachsten Schaltungen von

- Halbleiterwiderständen,
- Kapazitäten,
- Dioden,
- Feldeffekttransistoren (FET),
- Bipolartransistoren.

Optoelektronische Bauelemente und Sonderbauelemente werden im Kapitel 6. behandelt.
Ziel dieses Abschnittes ist es also,

- Voraussetzungen für die elektronische Schaltungstechnik und
- Grundlagen der integrierten Schaltungen

zu schaffen.

5.1. Halbleiterwiderstände

Widerstände in Halbleitermaterialien können *nach folgenden Prinzipien realisiert werden*:

- durch Strukturierung und Dotierung genau definierter Gebiete im Halbleitermaterial;
- durch Ausnutzung des Source-Drain-Widerstandes vom FET (führt zu relativ großen Widerständen bei kleinen Flächen auf dem Chip).
- Die Widerstände von Dioden und bipolaren Transistoren sind infolge ihrer hohen Nichtlinearität meist nicht anwendbar. Durch eine spezielle Technologie lassen sich so jedoch spannungsabhängige Widerstände (Varistoren) realisieren.
- Für sehr kleine Widerstände könnte auch die Leitschicht genutzt werden. Dies hat jedoch praktische Bedeutung nur bei der Hybridtechnik (Kapitel 9.4.).
- Licht-, magnetfeldempfindliche und zugempfindliche Widerstände siehe Abschnitt 6.1.2.4., 6.5.2. bzw. 6.5.3.

5.1.1. Halbleitermaterial

Das Prinzip der Strukturierung von Widerständen im Halbleitermaterial zeigt Bild 5.1.1. In den Teilbildern a) und b) ist davon ausgegangen, daß der Widerstand auf p-Material basiert. Im Fall a) ist er allseitig isoliert, im Fall b) befindet er sich

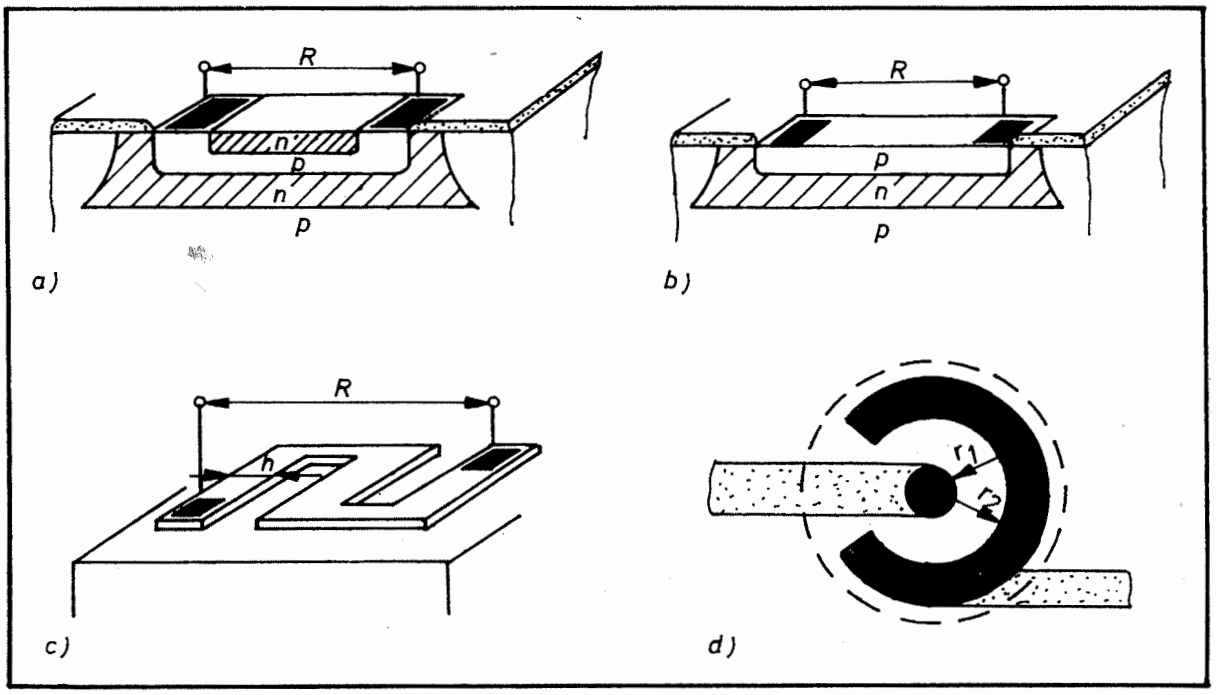


Bild 5.1.1. Prinzipieller Aufbau von Halbleiterwiderständen

a) mit durch *pn*-Übergang isolierter *p*-Leitung,
b) mit an der Oberfläche befindlicher *p*-Leitung,
c) in Mäanderform,
d) für kleine Widerstände in koaxialer Form

an der Oberfläche. Die Potentialverhältnisse müssen so gewählt werden, daß die *pn-Übergänge gesperrt bleiben*. Der Widerstand hängt von der *Geometrie des strukturierten* Widerstandsgebietes ab. Ist b die Breite, d die Dicke und l die Länge des Widerstandsgebietes mit dem spezifischen Widerstand ϱ, so gilt

$$R = \varrho \frac{l}{b \cdot d}. \tag{1}$$

Vielfach wird der *Flächenwiderstand*

$$R_\square = \varrho/d \tag{2}$$

eingeführt, dann gilt

$$R = R_\square \frac{l}{b}. \tag{3}$$

Mit dem geometrischen Verhältnis aus Länge und Breite kann also die Widerstandsgröße geändert werden und $R_\square$ ist der Widerstand einer quadratischen Struktur, unabhängig von der Größe. Sie bestimmt die Belastbarkeit. Große Widerstände lassen sich folglich in Mäanderform (Bild 5.1.1c)) und kleine in koaxialer Form (d)) herstellen. Für die *koaxiale Form* gilt

$$R \approx R_\square \ln (r_2/r_1). \tag{4}$$

Der spezifische Widerstand eines Halbleitermaterials kann aus Bild 3.1.10 entnommen werden. Er hängt von der Dotierung und der Temperatur ab. Durch eine ausgewählte Dotierung kann in einem kleinen Bereich ein fast konstantes ϱ erreicht werden. Kenndaten typischer diffundierter Widerstände enthält Tab. 5.1.1.

Nicht in diese Betrachtungen einbezogen sind die *parasitären Bauelemente*, wie Bahnwiderstände und Eigenschaften der zusätzlichen notwendigen *pn*-Übergänge. Hierzu muß die Spezialliteratur gelesen werden.

Tabelle 5.1.1. Kenndaten typischer diffundierter Widerstände. Die Flächenbelastbarkeit liegt generell bei etwa 4 ... 5 W/mm² und die Toleranz bei ± 20 ... 50%, relativ auf einem Chip ± 5%, ionenimplantiert 0,1 bis 2%.

	Basis-diffusion p	Emitter-diffusion n^+	Verdeckt Basis p	Kollektor p	Epitaxie-schicht	Ionen-implantiert
Flächenwiderstand, Ω/□	80 ... 300	1 ... 10	10^3 ... 10^4	50 ... 500	10^3 ... 10^4	10^2 ... 10^3
Diffusionsdicke, µm	3	2	1	5 ... 50		
Erreichbarer Wertebereich, Ω	10^3 ... 10^5	10 ... 10^3	10^4 ... 10^6	10^3 ... 10^5	10^3 ... 10^6	10^2 ... 10^6
Temperaturkoeffizient, 10^{-4} K^{-1}	+10 ... +30	+0,4 ... +1	+10 ... +40	ca. +20	+20 ... +35	$\leqq \|2\|$
Max. Betriebsspannung, V	20 ... 80	5 ... 7	5 ... 7	100 ... 500	50 ... 100	20 ... 80
Streukapazität, pF/mm²	200 ... 500	500 ... 1000		30 ... 150		

5.1.2. FET-Widerstände

Halbleiterwiderstände im Substrat benötigen besonders bei großen Widerständen *viel Chipfläche* und sind daher recht teuer. Hier bietet der *FET-Widerstand den Vorteil*, daß er große und dazu noch steuerbare Widerstände auf kleinster Fläche zu realisieren gestattet. Dies geht aus Bild 5.1.2a) hervor. Bei einem in etwa typischen Vorstufen-FET wird also der Bereich von ca. 300 Ω bis 1 MΩ möglich. Allgemein sind

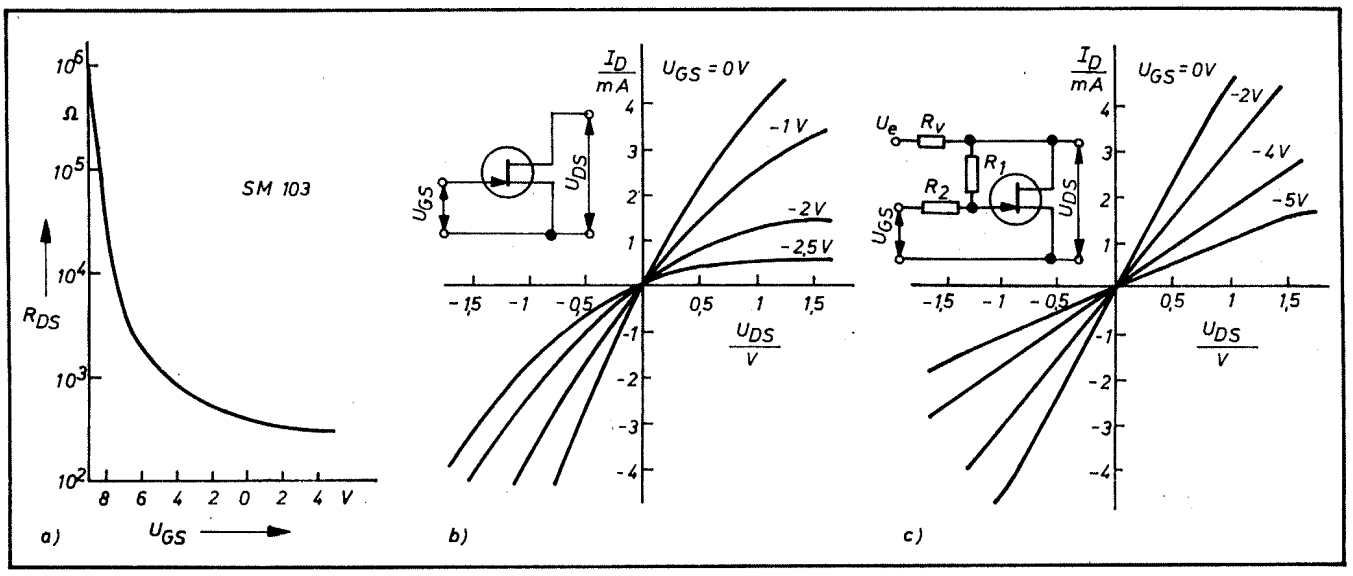


Bild 5.1.2. Zum FET als Widerstand

a) Widerstand eines SM 103 als Funktion der Gatespannung

b) Nichtlinearität der FET-Kennlinien für Drain-Source-Spannungen um Null

c) Linearisierung der Kennlinien von Bild b) durch eine Rückführung der Source-Drain-Spannung über R_1 und R_2 sowie Anwendung dieses Prinzips für einen steuerbaren Spannungsteiler aus R_V und dem Source-Drain-Widerstand

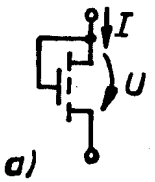
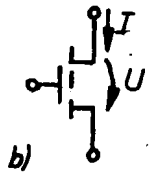
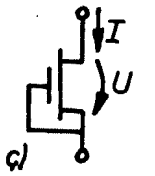
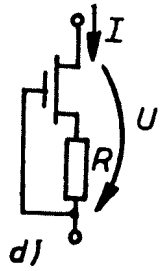
Bereiche von wenigen Ohm bis 10^9 Ohm zu erreichen. Nachteilig ist jedoch die erhebliche *Nichtlinearität* gegenüber größeren Spannungsabfällen am Source-Drain-Anschluß. Dies zeigt beispielhaft für kleine Wechselspannungen Bild 5.1.2b). Es stellt einen stark vergrößerten Ausschnitt vom Kennlinienverlauf im Bild 3.3.3 um den Nullpunkt herum dar. Im Vergleich mit Bild 3.3.3 wird aber auch deutlich, daß diese Nichtlinearität abnimmt, wenn der Wechselspannung eine relativ große Gleichspannung überlagert ist, so daß der Pinch-off-Bereich verwendet wird. Hier ist dann aber wiederum zwischen dem Gleichstrom- und Wechselstromwiderstand zu unterscheiden (vgl. Bild 1.4.9). Für spezielle Anwendungen lassen sich die Kennlinien um den Nullpunkt durch eine Rückführung über die Widerstände R_1 und R_2 (Bild 5.1.2c)) recht gut linearisieren. Dies besitzt z.B. praktische Vorteile bei *steuerbaren Spannungsteilern*. Dies ist mit dem Vorwiderstand R_V und der Eingangsspannung u_e in Bild 5.1.2c) ebenfalls angedeutet.

In integrierten Schaltungen haben sich etwa vier Schaltungen des FET als Widerstand bewährt. In Tab. 5.1.2 sind sie mit Vor- und Nachteilen zusammengestellt.

5.1.3. Heißleiter

Für viele Anwendungen sind Widerstände mit einem *bestimmten Temperaturkoeffizienten* bedeutsam. Große Temperaturkoeffizienten, wenn auch mit negativem Vorzeichen, lassen sich gut mit Halbleitermaterialien realisieren. Die reinen Substanzen Silicium und Germanium sind jedoch für die meisten Anwendungen zu teuer. Ähnliche

Tabelle 5.1.2. FET als Widerstand

Schaltung	Betriebsart	Vor-, Nachteile
 a)	*Anreicherung* ohne Steuerspannung	V: — einfach herstellbar N: — großer Spannungsabfall — schlechtes dynamisches Verhalten
 b)	*Anreicherung* mit Steuerspannung	V: — steuerbarer Widerstand (auch tastbar) — dynamisch gutes Verhalten N: — Zusatzspannung erforderlich
 c)	*Verarmung* ohne Steuerspannung	V: — siehe b) — keine Zusatzspannung erforderlich
 d)	*Verarmung* mit Steuerspannung (automatisch)	V: — großer Variationsbereich N: — Zusatzwiderstand flächenaufwendig

Werte können durch preiswerte Materialien realisiert werden. Vielfach genügt es, Donatoren und Akzeptoren als Verunreinigungen so zu wählen, daß ihre Wirkung sich gegenseitig aufhebt. In guter Näherung gilt für derartige *Materialien mit überwiegender Eigenleitung*:

$$\boxed{R(T) = R_0 \, \mathrm{e}^{-b/T}} \tag{5}$$

Darin bedeuten R_0 den Widerstand bei Zimmertemperatur und b die sogenannte Reglerkonstante. Für den zugehörigen Temperaturkoeffizienten in Abhängigkeit von der Temperatur gilt

$$\mathrm{TK} = \frac{\mathrm{d}\left(\ln R(T)\right)}{\mathrm{d}T} \approx \frac{-b}{T^2}. \tag{6}$$

Bei Zimmertemperatur wird er mit Temperaturbeiwert α bezeichnet. Er beträgt

$$\alpha = -\frac{b}{9 \cdot 10^4}. \tag{7}$$

Die wichtigsten Heißleitermaterialien enthält Tab. 5.1.3. Um 1930 wurde Urandioxid als sogenannter Urdox verwendet. Seine Leitfähigkeit läßt sich wahlweise mittels Metall- oder Oxidüberschuß einstellen. Ebenso lassen sich durch die Technologie und die Mischungsverhältnisse bei den anderen Materialien die Werte der Tabelle erhalten.

Tabelle 5.1.3. Spezifischer Widerstand ϱ und Temperaturbeiwert α wichtiger technischer Heißleiterwerkstoffe
(Werte in Klammer gelten bei Schutzatmosphäre oder im Vakuum)

Heißleitertyp	ϱ/Ωcm	$-\alpha/\%/$grd	$\varrho_{20°C}:\varrho_{min}$
Uranoxid (UO_2)	10^2 ... 10^4	1,5 ... 2,5	50 ... 2000
Kupferoxid (CuO)	10 ... 10^5	2 ... 3	10 ... 20
Magnesium-Titan-Spinell	10 ... 10^4	1,5 ... 4	7 ... 80 (30 ... 1000)
Eisenoxid-Mischspinelle (z. B. $ZnCr_2O_4 \cdot Fe_3O_4$)	10 ... 10^4	2,5 ... 4,5	50 ... 200 (100 ... 500)
Gesinterte Oxidgemische NiO; Li_2O; CrO_2; CoO	5 ... 10^6	2,8 ... 5,5	30 ... 1000 (50 ... 2000)

Der angegebene Aussteuerbereich ist durch die zulässige thermische Belastung des Materials begrenzt.

Heute werden Heißleiter hauptsächlich auf *vier Gebieten angewendet*:

— Temperaturkompensation anderer Effekte
— Temperaturmessung
— gesteuerte Widerstandsänderung
— Spannungsstabilisierung.

Näherer Erklärung bedürfen hier nur die letzten beiden Anwendungen. Eine gesteuerte Widerstandsänderung wird meist mit einem indirekt geheizten Heißleiter

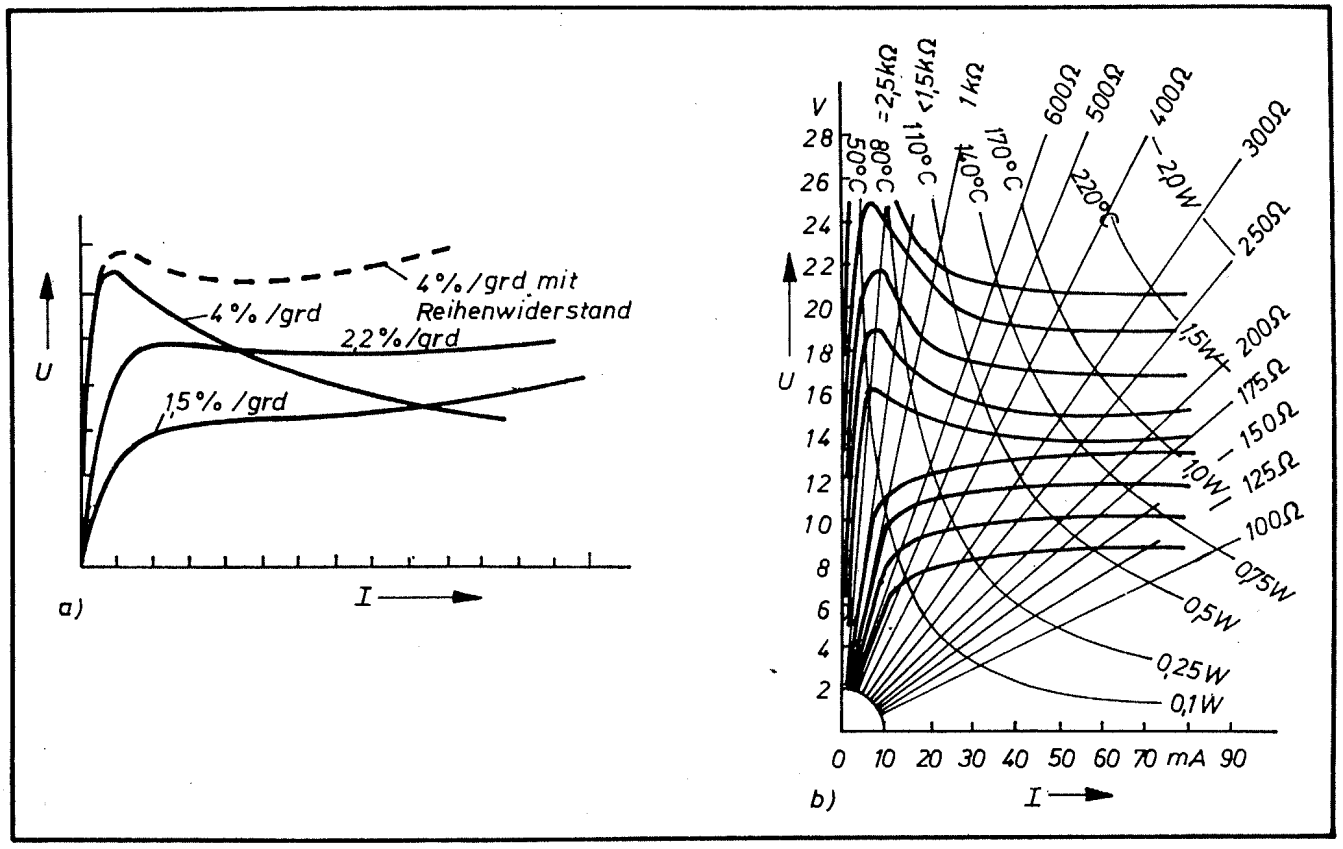


Bild 5.1.3. Kennlinien von Heißleitern

a) Prinzipieller Verlauf mit verschiedenen α-Werten

b) Reale Kennlinien technischer Heißleiter des VEB Keramische Werke Hermsdorf

vorgenommen. Sein Widerstandswert wird dann dadurch bestimmt, welche Leistung der zugehörigen, meist von ihm isoliert angebrachten Heizwicklung zugeführt wird.

Die *Spannungsstabilisierung* wird aus den Strom-Spannungs-Kennlinien eines Heißleiters verständlich. Hierbei bewirkt der Strom durch den Heißleiter selbst die Erwärmung. Die Kennlinien gemäß Bild 5.1.3a) gelten für den stationären Zustand. Dabei ist der Temperaturbeiwert α Parameter. Bei etwa 2,2% pro Grad bleibt die Spannung am Heißleiter über einen Strombereich von etwa 1:10 konstant. Bei größeren Temperaturkoeffizienten läßt sich derselbe Effekt durch einen Reihenwiderstand erreichen.

Kurven einiger technischer Heißleiter des VEB Keramische Werke Hermsdorf zeigt Bild 5.1.3b). Darin sind zugleich die Belastung des Widerstandes, die sich einstellende stationäre Temperatur und der Widerstand des Heißleiters eingetragen.

5.1.4. Kaltleiter

Heißleiter besitzen einen *negativen Temperaturkoeffizienten* (NTC-Widerstand: engl., negative temperature-coefficient). Sie entsprechen den Gebieten a und b in Bild 3.1.10b). Dazwischen liegt ein Gebiet mit positiven Temperaturkoeffizienten. Dieses Gebiet entspricht in etwa der metallischen Leitung. Leiter besitzen einen niedrigen Kalt- und einen hohen Heißwiderstand. Dieser Effekt wird u. a. bei *Glühlampen* ausgenutzt. Er ist jedoch nicht sehr ausgeprägt. Ein stärkerer positiver Temperaturkoeffizient (etwa 250fach) läßt sich mit $BaTiO_3$ erreichen. Hierbei wirken *Sperrschichten zwischen den Körnern des keramischen Materials* mit. Dieser Effekt wurde 1951 entdeckt, und es ist bisher nicht gelungen, das Verhalten mathematisch zu erfassen. Beispiele von Kennlinien technischer keramischer Kaltleiter (PTC-Widerstände) zeigt Bild 5.1.4. Es gibt also eine *kritische*, technologisch beeinflußbare *Temperatur*, von der ab der Widerstand sich steil, auf etwa das Doppelte je zwei Grad, erhöht. Je nach der Dotierung mit Fremdmaterialien, z.B. Pb oder Sr, kann die kritische Temperatur etwa zwischen −100 und +350 °C eingestellt werden. Das Strom-Spannungs-Diagramm zeigt dementsprechend zunächst einen proportionalen Strom-

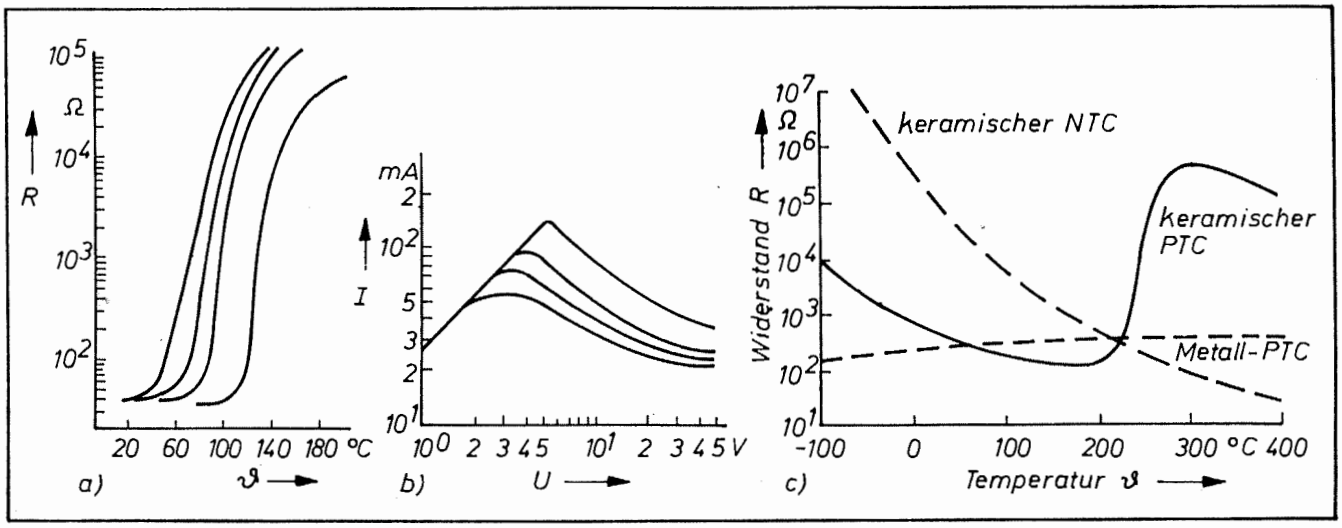


Bild 5.1.4. Beispiele von Kennlinien technischer Kaltleiter des VEB Keramische Werke Hermsdorf

a) Widerstand als Funktion der Temperatur
b) Strom-Spannungs-Kennlinie für den stationären Zustand
c) Vergleich von Heißleitern, metallischen und keramischen Kaltleitern

anstieg. Sobald die kritische Temperatur überschritten wird, fällt der Strom mit steigender Spannung ab. Eine spezielle Anwendung des keramischen Kaltleiters ist die *Temperaturbegrenzung* auf die Temperaturschwelle. Dies wird z. B. bei kleinen *Thermostaten* genutzt.

5.1.5. Varistoren

Heiß- und Kaltleiter besitzen einen Widerstand, der erheblich vom Strom bzw. von der Spannung abhängt. Dieser Effekt erfolgt jedoch stark verzögert über die Stromwärme. Anders liegen die Verhältnisse bei *pn*-Übergängen. Sie haben jedoch den Nachteil, daß der Widerstand bei ihnen auch von der Stromrichtung abhängt. Dieser Einfluß kann ausgeschaltet werden, wenn *viele pn-Übergänge* in unterschiedlicher Anordnung *zusammenwirken*. Dieser Fall liegt vor, wenn *pn*-Übergänge zwischen verschiedenen *Sinterkörnern* (von etwa 50 µm Durchmesser) gemäß Bild 5.1.5a)

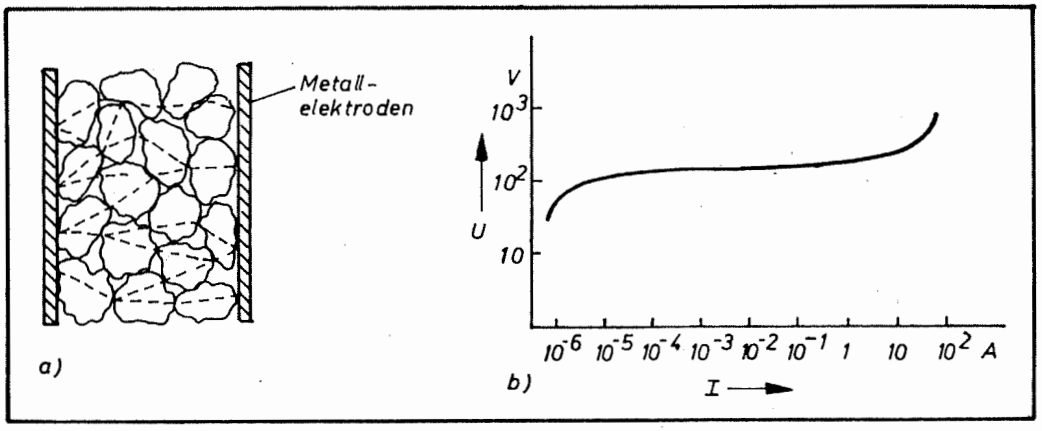


Bild 5.1.5. a) Aufbau und Stromlinien in einem Varistor
b) Typische Kennlinie eines Varistors

vorliegen. (Dieser Effekt tritt z.T. auch bei den keramischen Heiß- und Kaltleitern auf.) Die richtungsunabhängige Kennlinie eines solchen aus SiC bestehenden VDR (voltage dependent resistor) zeigt Bild 5.1.5b). Sie kann mit recht guter Näherung durch die Gleichung

$$U = K \cdot |I|^{\beta} \tag{8}$$

beschrieben werden. Der *Nichtlinearitätskoeffizient* β hängt von technologischen Bedingungen ab und liegt meist bei 0,15 bis 0,3, in Sonderfällen, z.B. bei TiO_2 oder ZnO, auch zwischen 0,02 und 0,4. Die Konstante K hängt u. a. vom *Längen-Durchmesser-Verhältnis* ab und liegt bei etwa 100 bis 1000 V. Die Stromdichte im Varistor beträgt etwa 0,1 mA/cm² bis 1 A/cm².

Infolge der vielen Korngrenzen besitzen die *pn*-Übergänge eine beachtliche *Kapazität*. Sie überbrückt *bei höheren Frequenzen* den Varistor. Dadurch sind für die meisten Varistoren bis ca. 200 Hz die Kapazitäten vernachlässigbar.

Die wichtigsten *Anwendungen* des Varistors liegen bei

- Spannungsbegrenzung, u. a. beim Abschalten von Induktivitäten,
- Spannungsstabilisierung (s. auch 5.4.6., 7.6.).

5.2. Kapazitäten

In *integrierten Schaltungen* lassen sich Kapazitäten auf *drei Arten* realisieren:

- Kapazitätsdioden (Bild 5.2.1 d))
- Elektroden und Isolator (Bild 5.2.1 a) — c))
- analog zur Gate-Kapazität bei MOSFET (Bild 5.2.1 e)).

Die Kapazitätsdioden wurden bereits im Abschnitt 3.2. behandelt. Typische Daten gehen aus Bild 3.2.5 hervor. Die wesentliche Eigenschaft der *Kapazitätsdiode* ist ihre Steuerbarkeit mittels der angelegten Sperrspannung. Je nach Anwendungsgebiet werden Kapazitäten von wenigen pF bis zu einigen hundert pF realisiert. Der Kennlinienverlauf hängt u. a. vom Dotierungsprofil (Bild 3.2.5a) bis c)) ab. Infolge der Sperrströme und anderer Verluste besitzt in Wirklichkeit eine Kapazitätsdiode einen Verlustwiderstand (und Zuleitungsinduktivitäten). Dies begrenzt den Frequenzbereich in der Anwendung. Die Güte der Kapazität wird meist über den effektiven Serienwiderstand R_s bestimmt:

$$Q = \frac{1}{\omega R_s C}. \tag{1}$$

Bei den optimalen Frequenzen werden Werte von 200 bis 1000 erreicht.

Das *Hauptanwendungsgebiet* der Kapazitätsdioden ist der Einsatz als *Schwingkreiskapazität*, u. a. zur Frequenzabstimmung oder zur automatischen Scharfabstimmung von Sendern (AFC = automatic frequency control). Während normalerweise Kapazitätsvariationen von etwa 1:5 erreicht werden, werden bei den hochkapazitiven Varianten Werte bis ca. 1:25 realisiert. Zwei typische Verläufe zeigt Bild 5.2.2a) und b). Auffallend ist, daß bei den großen Kapazitätswerten bereits sehr kleine Spannungsänderungen genügen, um den Kapazitätswert zu ändern. Daher kann

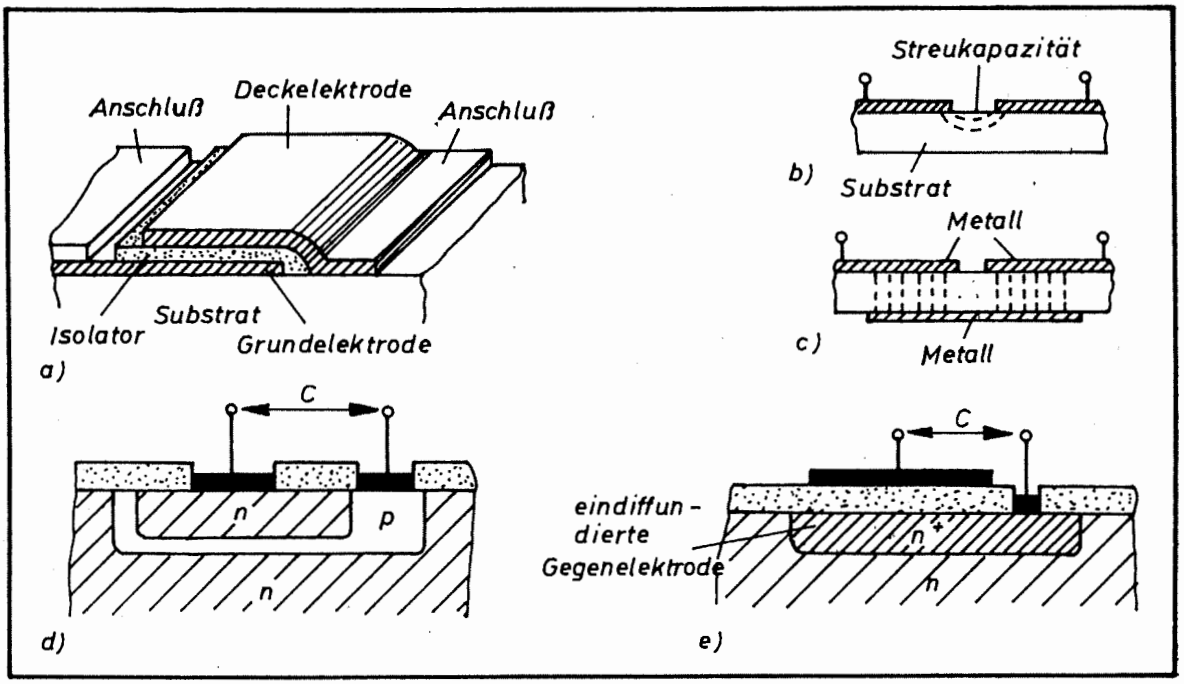


Bild 5.2.1. Aufbau typischer Schichtkapazitäten

a) Auf dem Substrat in der Reihenfolge Metall — Isolator — Metall erzeugte Kapazität
b) Streukapazität zwischen zwei nahen Leiterzügen
c) Vergrößerung der Streukapazität durch eine Gegenelektrode
d) Prinzipielle Struktur einer Kapazitätsdiode
e) Aufbau einer integrierten Kapazität mit Isolator

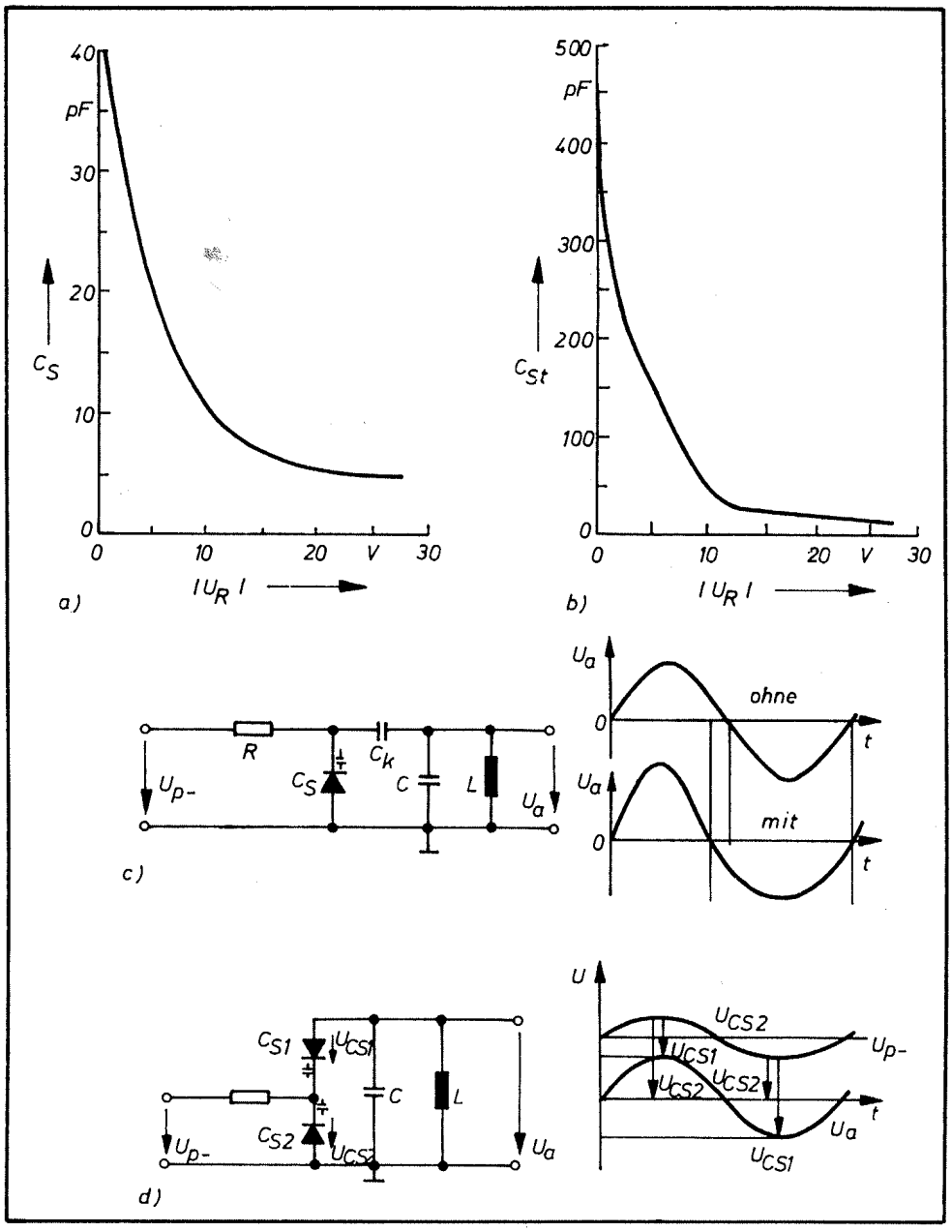


Bild 5.2.2. Kennlinien und Anwendung von Kapazitätsdioden

a) Kennlinie einer VHF-Abstimmdiode
b) Kennlinie einer AM-Abstimmdiode
c) Entstehung nichtlinearer Verzerrungen
d) Schaltung zur Kompensation der nichtlinearen Verzerrungen durch unterschiedliche Spannungen an den beiden Dioden

hier die *Wechselspannung* des Schwingkreises selbst zur Änderung der Kapazität und damit zur *periodischen Verstimmung*, d. h. zu *nichtlinearen Verzerrungen*, beitragen. Dies ist in Bild 5.2.2c) angedeutet. Eine erhebliche Kompensation dieses Effektes ist durch *Doppeldioden* gemäß Bild 5.2.2d) möglich.

Andererseits wird dieser nichtlineare Effekt auch in der Höchstfrequenztechnik bewußt ausgenutzt. Es werden auf dieser Basis einmal energiereiche Oberwellen erzeugt, und zum anderen lassen sich hiermit die *parametrischen Verstärker* realisieren. Die entsprechenden Kapazitätsdioden heißen dann Varaktordioden oder Varaktoren.

Den typischen Aufbau eines Kondensators über zwei Leiter und ein dazwischen liegendes Dielektrikum zeigt Bild 5.2.1a); Varianten davon Teilbild b) und c). Diese Typen besitzen vor allem in der *Dick-* und *Dünnschichttechnik* (9.4.5.) größere Bedeutung. Sie zeigen aber auch zugleich die Analogie zur Gatekapazität beim FET. Das

Prinzip zeigt Bild 5.2.1 d), während e) den Aufbau einer Kapazitätsdiode zeigt. Typische Daten von integrierten Kapazitäten faßt Tab. 5.2.1 zusammen. Insgesamt sind nur relativ kleine Kapazitäten um grob 100 pF zu erreichen.

Tabelle 5.2.1. Typische Werte von Kapazitäten in monolithischer bzw. Schichttechnologie

Art	Spezifische Kapazität pF/mm²	Maximale Spannung V
● Dioden		
Kollektor—Basis	100 ... 500	55 ... 100
Basis—Emitter	300 ... 1500	5 ... 15
Kollektorsubstrat	50 ... 200	50 ... 100
● MOS-Kapazität, 0,1 µm SiO_2	500	40 ... 60
● Dünnschichttechnologie:		
SiO_2 0,2 µm	50 ... 300	40
SiO 0,2 µm	20 ... 100	20
Ta_2O_5 0,3 µm	600 ... 10000	20
Al_2O_3 0,5 µm	3000	50
TiO_2 0,5 µm	1000 ... 6000	50
Si_3N_4 0,2 µm	300 ... 600	50
● Dickschichttechnologie:		
Glas 25 µm	1 ... 2	500
TiO_2 25 µm	30	500
$BaTiO_3$ 25 µm	100	500

5.3. Induktivitäten

Sind schon Kapazitäten bei den integrierten Techniken schwer zu realisieren, so gilt dies noch mehr für Induktivitäten. Hier kommt praktisch nur die mit der Leitmaske oder in der Schichttechnik zu realisierende Induktivität gemäß Bild 5.3.1 in Betracht. Leider sind auch *keine brauchbaren Effekte* bekannt, die induktiven Charakter besitzen. Dies hat zur Folge, daß fast ausschließlich Induktivitäten außerhalb der inte-

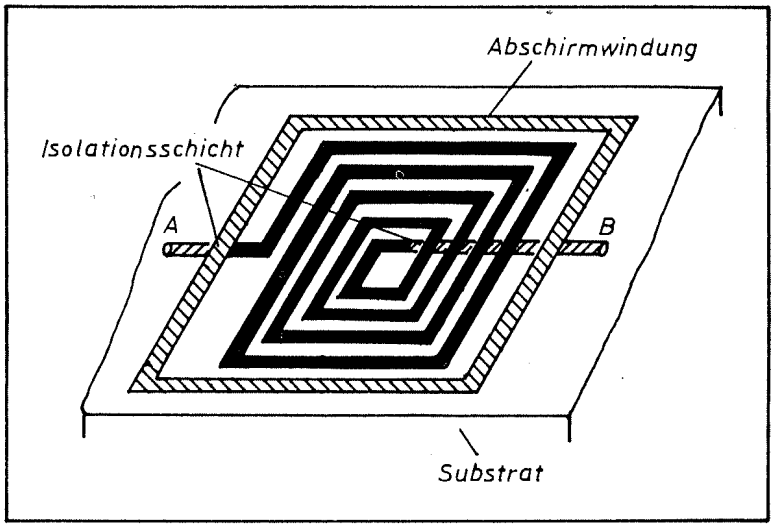


Bild 5.3.1. Aufbau einer Induktivität durch aufgedampfte oder gedruckte Leiterzüge. Eventuell kann auch eine Abschirmung erzeugt werden.

grierten Schaltung anzuschließen sind. Wie indirekt Kapazitäten und Induktivitäten mittels spezieller aktiver Schaltungen zu realisieren sind, wird in Abschnitt 7.1.4. gezeigt.

5.4. Dioden

Begriffe wie Dioden, Trioden, Pentoden usw. stammen aus der Röhrentechnik und leiten sich von lateinischen bzw. griechischen Zahlwörtern ab. In der Röhrentechnik wurde so die Anzahl der Elektroden gezählt. In der Halbleitertechnik wäre es naheliegend, die Anzahl der Dotierungsgebiete oder Anschlüsse zu zählen. Beides ist jedoch nicht konsequent erfolgt. So ist heute der Begriff Diode unbestimmt. Vorwiegend wird er auf Bauelemente angewendet, die primär die Röhrendiode ersetzen, also Zweipole mit einem *pn*-Übergang sind. Die GUNN-Diode enthält aber z. B. keinen *pn*-Übergang und die Vierschichtdiode, wie bereits ihr Name sagt, mehrere. Anderer-

Tabelle 5.4.1. Anwendungen von Dioden mit Beispielen und Verweisen

Anwendung	Beispiele	Abschnitt bzw. Literatur
Gleichrichter	Stromversorgung Backwarddiode Meßtechnik Demodulation	5.4.3. 5.4.4. 5.4.7.
Spannungsstabilisation	Z-Diode Schleusenspannung	5.4.5.
Spannungsbegrenzung	antiparallele Dioden Funkenlöschung	5.4.6.
Schalter	Leistungsschalter Logikschaltungen Meßtechnik Frequenzumsetzung	5.4.8. 5.4.9. 6.4.2.
Nichtlineare Kennlinie	logarithmische Verstärker expon. Kennlinie	7.5.7.
Steuerbarer Widerstand	*pin*-Diode bei hohen Frequenzen	[T 3]
Frequenzerzeugung Verstärkung Frequenzvervielfachung	Tunneldiode Varaktordiode GUNN-Diode Impattdiode	 [M 25] [T 3]
Kapazität	Kapazitätsdiode Varaktor	5.2.
Lichtempfindlichkeit	Fotodiode	6.1.2.
Lichterzeugung	Lumineszenzdiode Laserdiode	6.1.3.
Rauschen	Rauschdiode	[P 9]

seits werden auch nicht alle Halbleiterzweipole, z. B. Widerstände, als Dioden bezeichnet. Dies macht die Definition von Halbleiterdioden schwierig. Sinnvoller erscheint es, nach dem Stand der Technik eine Klassifikation zu versuchen. Sie soll hier unter drei Gesichtspunkten vorgenommen werden:

- Material
- Bauart
- Anwendung.

Die Reihenfolge ist bereits nach der Vielfalt der Aspekte ausgewählt. Einen Überblick zur Anwendungsbreite mit ergänzenden Hinweisen gibt Tab. 5.4.1. Dabei wurde auch versucht, die Anwendungsbreite der betonten Wirkungsprinzipien herauszustellen, indem z. B. Schwachstrom- und Leistungsschaltungen nicht unterschieden werden.

Vom Material her sind heute vor allem als Substanzen bedeutsam: Si, Ge, GaAs (in verschiedenen Abwandlungen) und z. T. noch Se und Cu_2O. Für Sonderanwendungen, auf die hier nicht eingegangen wird, werden auch weitere Halbleitermaterialien verwendet. Einen Überblick zur geschichtlichen Entwicklung der Diodentechnik zeigt Tab. 5.4.2.

Einen gewissen Übergang zur Bauart stellen die besonderen *pn*-Übergänge wie der SCHOTTKY-Kontakt, also die SCHOTTKY-Diode, der Hetero-*pn*-Übergang, also eine mehrfache Materialabfolge in einem *pn*-Übergang und *pin*- bzw. *psn*-Dioden dar. Bei den letzten beiden ist ein eigenleitendes (i = instrinsic) bzw. schwach leitendes (s = soft) Gebiet zwischen dem p- und n-Gebiet eingelagert (s. auch 3.2.8.).

Tabelle 5.4.2. Geschichtliche Daten bezüglich Dioden

1874	Publikation des Leipziger Gymnasiallehrers FERDINAND BRAUN bez. Gleichrichtung von Kupfersulfid
1877	Dito bez. Selen
1882	Gleichrichterwirkung des Hg-Lichtbogens bekannt
1885	C. E. FRITTS beobachtet Gleichrichtereffekt an Eisen-Selen-Kontakten (entspricht heutigen SCHOTTKY-Dioden)
1899	Kristalldetektor von BRAUN als HF-Gleichrichter
1899	GRAETZ-Gleichrichter
1902	Erster technischer Hg-Gleichrichter
1904	Vakuumröhre
1920	Detektor bei Einführung des Rundfunks angewendet
1926	Kupferoxidulgleichrichter im Einsatz (GRONDAHL)
1930	Großtechnischer Einsatz von Selengleichrichtern
1939	Theorie der Verarmungsschicht von SCHOTTKY abgeschlossen (u. a. SCHOTTKY-Dioden)
1940	Halbleiterspitzendioden
1947	Spitzentransistor
1951	Germanium-Leistungsgleichrichter
1952	Z-Diode
1952	Si-Technologie beginnt
1955	Thyristor (General Electric)
1958	Tunneldiode von ESAKI
1960	SCHOTTKY-Diode
1961	GaAs-LED
1962	GaAs-Laserdiode
1963	Lawinenverstärkung von GUNN entdeckt (GUNN-Diode)
1964	Impattdiode (impact avalanche transit time)

5.4.1. Bauarten

Hier wird *nicht eingegangen* auf

- Tunnel- und Backwarddioden (6.6.),
- Foto-, Lumineszenz- und Laserdioden (6.1.2., 6.1.3.),
- Höchstfrequenzdioden, wie GUNN-, Impatt-, *pin*-, Step-recovery-, Varaktordioden usw. (z. B. [M 25, P 9, P 11, T 3], Spezialliteratur).

Die *Bauarten* betreffen zwei Punkte:
Gehäuse und Schichtfolge

Die *Gehäuse* lassen sich grob in drei Gruppen einteilen:

- einfache Glasgehäuse oder Kunststoffeinpressungen mit axialen Anschlußdrähten;
- Gehäuse der integrierten Techniken, in denen meist mehrere Dioden enthalten sind;
- Leistungsgehäuse, meist in Metallausführung, um gegebenenfalls mit Kühlanordnungen in Wärmekontakt gebracht zu werden (Kühlprobleme s. 5.7.6.).

Den Prinzipaufbau einiger Dioden zeigt Bild 5.4.1.

Die *Spitzendiode* ist historisch die älteste Diode. Sie hat heute vor allem wegen ihrer sehr kleinen Sperrschichtkapazität Bedeutung. Eine s-förmig gebogene Drahtspitze aus Wolfram oder Molybdän wird federnd auf *n*-Germanium gesetzt. Durch

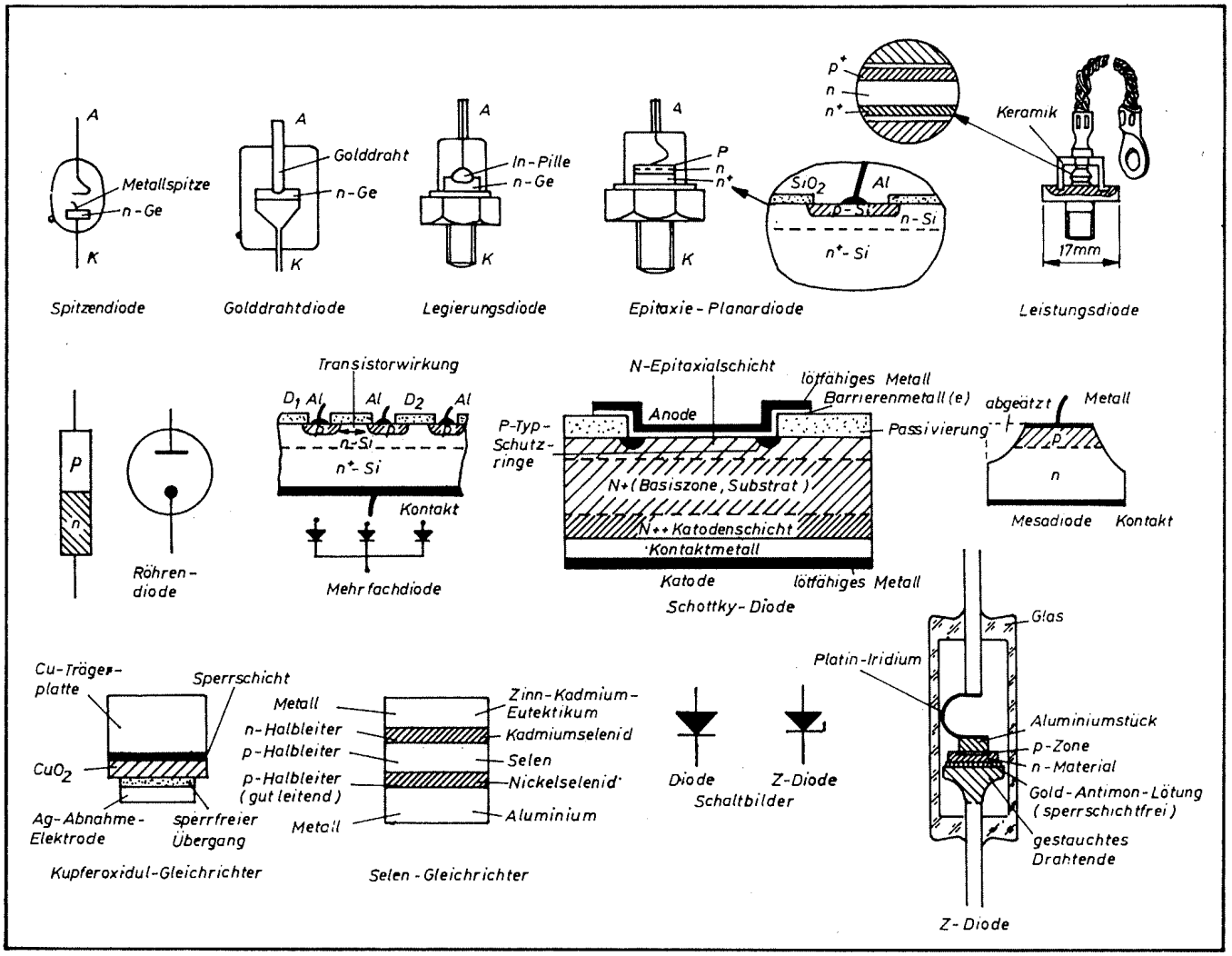


Bild 5.4.1. Beispiele von Halbleiterdioden

einen Stromstoß wird der Spitzenkontakt mit dem Germanium verschweißt. Dabei diffundieren Metallionen in das Germanium und bilden eine kleine p-leitende Zone von ca. 10 µm Durchmesser.

Die *Golddrahtdiode* wird in ähnlicher Weise hergestellt. Sie ist jedoch darauf ausgelegt, daß keine langsamen Träger entstehen können. Der Golddraht ist dazu mit Akzeptoren, z.B. Gallium oder Indium, dotiert. Infolge ihrer Eigenschaften zählt sie wie die SCHOTTKY-Diode zu den Hot-carrier-Dioden mit schnellen Schalteigenschaften.

Bei der *Legierungsdiode* wird auf ein n-Ge-Substrat eine In-Pille einlegiert. So entsteht ein flächenhafter pn-Übergang.

Die *Mesadiode* kann auf verschiedene Weise hergestellt werden. Entscheidend ist bei ihr, daß die oft störungsbeladenen Randzonen weggeätzt werden. So entsteht die Form eines Tafelberges (span.: Mesa).

Die *Epitaxie-Planardiode* wird mit den Technologien der integrierten Schaltkreise hergestellt. Sie stellt den Hauptteil der Leistungsdioden dar, die mit großer Fläche für große Ströme (bis zu mehreren hundert Ampere) geeignet sind. Weiter ist es hierbei möglich, gleichzeitig eine Vielzahl von Dioden zu kombinieren und auch in einem Gehäuse unterzubringen.

Die *Z-Diode* entspricht meist der Epitaxie-Planardiode, verwendet jedoch spezielle Dotierungen und wird in Sperrichtung betrieben.

Die SCHOTTKY-*Diode* besitzt einen Metall-Halbleiter-Kontakt. Für Sonderfälle wird sie als Schutzring-SCHOTTKY-Diode ausgeführt.

In die Reihe der Halbleiterdioden sind noch einzuordnen: Beim *Selengleichrichter* wird auf ein Al- oder Fe-Blech eine halogenhaltige Selenschicht aufgebracht. Darüber wird Cd geschichtet, welches Zusätze aus Zn oder Bi enthält. Dabei entsteht eine dünne n-leitende Cadmiumselenidschicht, die gegenüber dem p-leitenden Se eine Sperrschicht bildet.

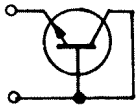
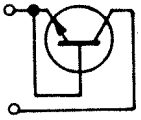
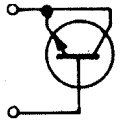
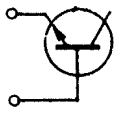
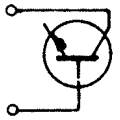
Schaltung	Durchbruchspannung	Relative Freiwerdezeitkonstante (Speicherzeitkonstante)	Sperrstrom
	5...7 V	sehr klein Bezug	sehr klein
	30...50 V	mittel ca. 6×	hoch
	5...7 V	sehr groß ca. 11×	klein
	5...7 V	mittel ca. 6×	mittel
	50 V	groß ca. 9×	mittel

Bild 5.4.2. Verschiedene Formen von integrierten Dioden, die aus dem Standardtransistor abgeleitet sind

Der CuO_2-*Gleichrichter* hat bestenfalls nur noch für Meßzwecke Bedeutung. Die Sperrschicht entsteht zwischen Cu und CuO_2.

Des öftern werden *Dioden aus Transistoren* abgeleitet. Hierzu gibt es mehrere Schaltungen, die mit ihren Eigenschaften in Bild 5.4.2 zusammengestellt sind.

5.4.2. Kennlinien und Daten

Entsprechend der Vielfalt der Anwendungen existieren auch z. T. sehr unterschiedliche Kennlinien und Daten. Hier wird vor allem das Hauptanwendungsgebiet der Gleichrichtung als Bezug gewählt. Eine Zusammenstellung der dann wichtigsten Daten für verschiedene Materialien geht aus Tab. 5.4.3 und Bild 5.4.3 hervor. Für

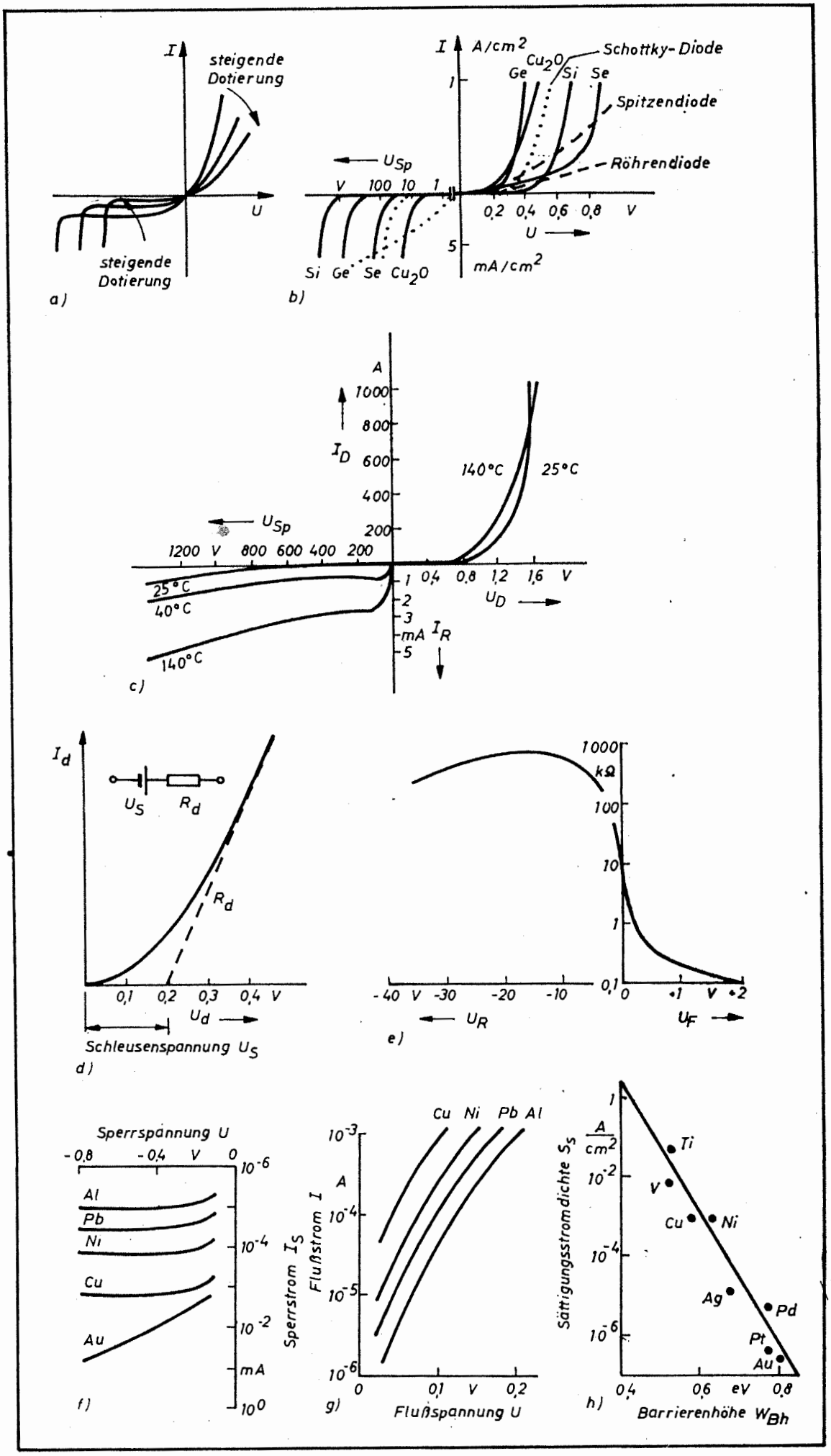
Tabelle 5.4.3. Eigenschaften der wichtigsten Gleichrichtermaterialien

	Cu_2O	Se	Ge	Si
Max. Strombelastung bei Kühlung, A/cm²	0,1	0,2	100	1000
Sperrspannung, V	15	60	300	5000
Maximale Betriebstemperatur, °C	50	90	80	180
Element-Wirkungsgrad, %	78	93	98,5	99,6
Relativer Raumbedarf bei gleicher Leistung	30	15	5	1
Schleusenspannung, V	0,25	0,5	0,35	0,7
Differentieller Widerstand, Ω/cm²	2	1	$5 \cdot 10^{-3}$	10^{-3}
Verhältnis Sperr-/Durchlaßwiderstand	10^3	10^3	$5 \cdot 10^5$	10^6
Sperrströme (ca.-Werte), A/cm²	10^{-3}	10^{-3}	10^{-3}	10^{-5}

große Spannungen und Ströme ist eindeutig die *Siliciumdiode* im Vorteil, insbesondere dann, wenn ein *i*-Gebiet eingefügt wird, wie es bereits in Abschnitt 3.2.8. behandelt wurde. Nachteilig sind ihre Eigenschaften aber, wenn nur kleine Spannungen gleichgerichtet werden sollen. Hier ist nämlich die in Bild 5.4.3d) erklärte *Schleusenspannung* bedeutsam. Sie ergibt sich, wenn eine Tangente an die Flußkennlinie gelegt wird. Die Neigung dieser Tangente kann als Durchlaßwiderstand R_d interpretiert werden. So entsteht für die Flußrichtung eine Ersatzschaltung der Diode, wie sie in Bild 5.4.3d) skizziert wurde. Diese Schleusenspannung ist gemäß Tab. 5.4.3 besonders klein bei Cu_2O und Ge. Für kleine Spannungen sei auf die Rückwärtsdiode (6.6.2.) verwiesen. Bei dem Ersatzschaltbild mit *Schleusenspannung* ist zu beachten, daß es eigentlich durch die *lineare Darstellung* und die Einschränkung auf einen *Arbeitsstrombereich* hervorgerufen wird, denn in guter Näherung besitzen alle Durchlaßkennlinien den theoretischen exponentiellen Verlauf (Gl. (3.2.5)). Hieraus folgt, daß die *Durchlaßspannung* sich jeweils um *60* bis *120* mV erhöht, wenn der Strom ver-

Bild 5.4.3. Kennlinien von Gleichrichtern ▶

a) Prinzipieller Einfluß der Dotierung
b) Vergleich verschiedener Dioden
c) Einfluß der Temperatur bei einer Leistungs-Si-Diode
d) Zur Erklärung der Schleusenspannung
e) Differentieller Widerstand einer Diode
f) Sperrkennlinien von SCHOTTKY-Dioden
g) Flußkennlinien von SCHOTTKY-Dioden
h) Zusammenhang von Sättigungsstromdichte und Barrierenhöhe gegenüber *n*-Si bei SCHOTTKY-Dioden


I
steigende Dotierung
U
steigende Dotierung
a)
I A/cm²
Schottky-Diode
Ge
Cu₂O
Si
Se
Spitzendiode
Röhrendiode
USp
V 100 10 1
0,2 0,4 0,6 0,8 V
U
mA/cm²
Si Ge Se Cu₂O
b)
A
1000
800
ID
600
400
200
140°C
25°C
USp
1200 V 800 600 400 200
1 0,4 0,8 1,2 1,6 V
UD
25°C
40°C
140°C
1
2
3
mA IR
5
c)
Id
US
Rd
Rd
0,1 0,2 0,3 0,4 V
Ud
Schleusenspannung US
d)
1000
kΩ
100
10
1
0,1
-40 V -30 -20 -10
0 +1 V +2
UR
UF
e)
Sperrspannung U
-0,8 -0,4 V 0
Al
Pb
Ni
Cu
Au
10⁻⁶
10⁻⁴
10⁻²
mA
10⁰
f)
Sperrstrom IS
Flußstrom I
10⁻³
A
10⁻⁴
10⁻⁵
10⁻⁶
Cu Ni Pb Al
0 0,1 V 0,2
Flußspannung U
g)
Sättigungsstromdichte SS
A/cm²
1
10⁻²
10⁻⁴
10⁻⁶
Ti
V
Cu
Ni
Ag
Pd
Pt
Au
0,4 0,6 eV 0,8
Barrierenhöhe WBh
h)

zehnfacht wird. Aus der Gleichung kann aber auch der *Temperatureinfluß* berechnet werden. Er beträgt etwa -2 mV *je Grad*. Bei 10 Grad Erhöhung verdoppeln sich also Sättigungs- und Sperrstrom. Prinzipiell bestätigt dies Bild 5.4.3c). Da bei dieser Diode jedoch ein *i*- bzw. *s*-Gebiet eingebaut wurde, geht dessen Widerstandsänderung ebenfalls in Flußrichtung ein und führt zur Überschneidung der Kennlinien bei ca. 1,5 V. Ähnliches Verhalten zeigen auch Z-Dioden.

In Bild 5.4.3e) ist die typische Kennlinie des *differentiellen Widerstandes* einer Diode dargestellt. Für verschiedene praktische Anwendungen (5.4.5.) ist es bedeutsam, daß er in Durchlaß- und Sperrichtung schnell auf etwa konstante Werte, hier 0,1 Ω und 500 kΩ kommt. Bei den HF-*pin*-Dioden wird der dann verlängerte lineare Mittelwert genutzt.

Kennlinien der SCHOTTKY-*Diode* zeigt schließlich Bild 5.4.3f) bis h), dabei wird auch ein Bezug zu Bild 3.2.10 hergestellt. SCHOTTKY-Dioden zeichnen sich vor allem durch ihre kurze Schaltzeit aus. Sie haben auch als Leistungsgleichrichter Bedeutung. Ein Problem ist dabei die *Durchbruchspannung*. Sie beginnt an den Grenzen des SCHOTTKY-Kontaktes. Dadurch gelang es, diesen Wert mittels *Schutzringdioden* (vgl. Bild 5.4.1k)) wesentlich, auf ca. 50 V, zu verbessern. SCHOTTKY-Dioden besitzen weiter eine relativ kleine Schleusenspannung (bei Chrom z.B. ca. 0,2 V, im Betrieb z.B. bei 60-A-Gleichrichtern bei 50 A ca. 0,45 V) und zeichnen sich durch einen exakt exponentiellen Verlauf der Durchlaßkennlinie bis zu acht Zehnerpotenzen aus.

5.4.3. Grundlagen der Gleichrichtung

Entsprechend Bild 5.4.3e) ist eine Halbleiterdiode am einfachsten durch einen *Schalter* zu veranschaulichen, der bei der einen Stromrichtung *geschlossen* und bei der anderen *geöffnet* ist. Hieraus ergibt sich die Wirkung, wie sie Bild 5.4.4a) zeigt. An den Widerstand gelangt nur die positive Halbwelle. Wenn dieser einfache Fall gegeben ist, wird oft vom *idealen Gleichrichter* gesprochen. Er ist in etwa gegeben, wenn bezüglich des Innenwiderstandes der Stromquelle R_i, des Durchlaß- und Sperrwiderstandes R_d bzw. R_{sp} der Diode und des Lastwiderstandes folgende Beziehungen bestehen (Bild 5.4.4b)):

$$R_i + R_d \ll R_L\,, \tag{1}$$

$$R_{sp} \gg R_L\,. \tag{2}$$

Bei Abweichungen treten Verluste auf. Bezüglich R_d und R_{sp} betreffen sie den Gleichrichter (hier Synonym für Diode), bezüglich R_i die Stromquelle.

Nur in seltenen Fällen wird eine Schaltung gemäß 5.4.4a) verwendet, unmittelbar nach dem Gleichrichter folgt meist ein Kondensator (*Ladekondensator* C_L). Dadurch werden die Verhältnisse sofort ganz anders (Teilbild c)). Während des *Stromflußwinkels* Θ wird der Kondensator aufgeladen (der Schalter ist nur während dieser Zeit geschlossen), und während der Zeit $2\pi - \Theta$ (Entladung) liefert der Kondensator den benötigten Strom. Daher fällt während dieser Zeit die mittlere Gleichspannung U_m (am Meßpunkt A) exponentiell gegen Null um den Betrag ΔU ab. Der Übergang zum Stromfluß (Einschalten des Gleichrichters) erfolgt, sobald abklingende Kondensatorspannung und Halbwelle der Wechselspannung gleich sind. Die Berechnung dieser Zusammenhänge ist nicht einfach. Unter der Bedingung eines sehr großen Kondensators

$$\frac{1}{\omega C} \ll R_i + R_d \tag{3}$$

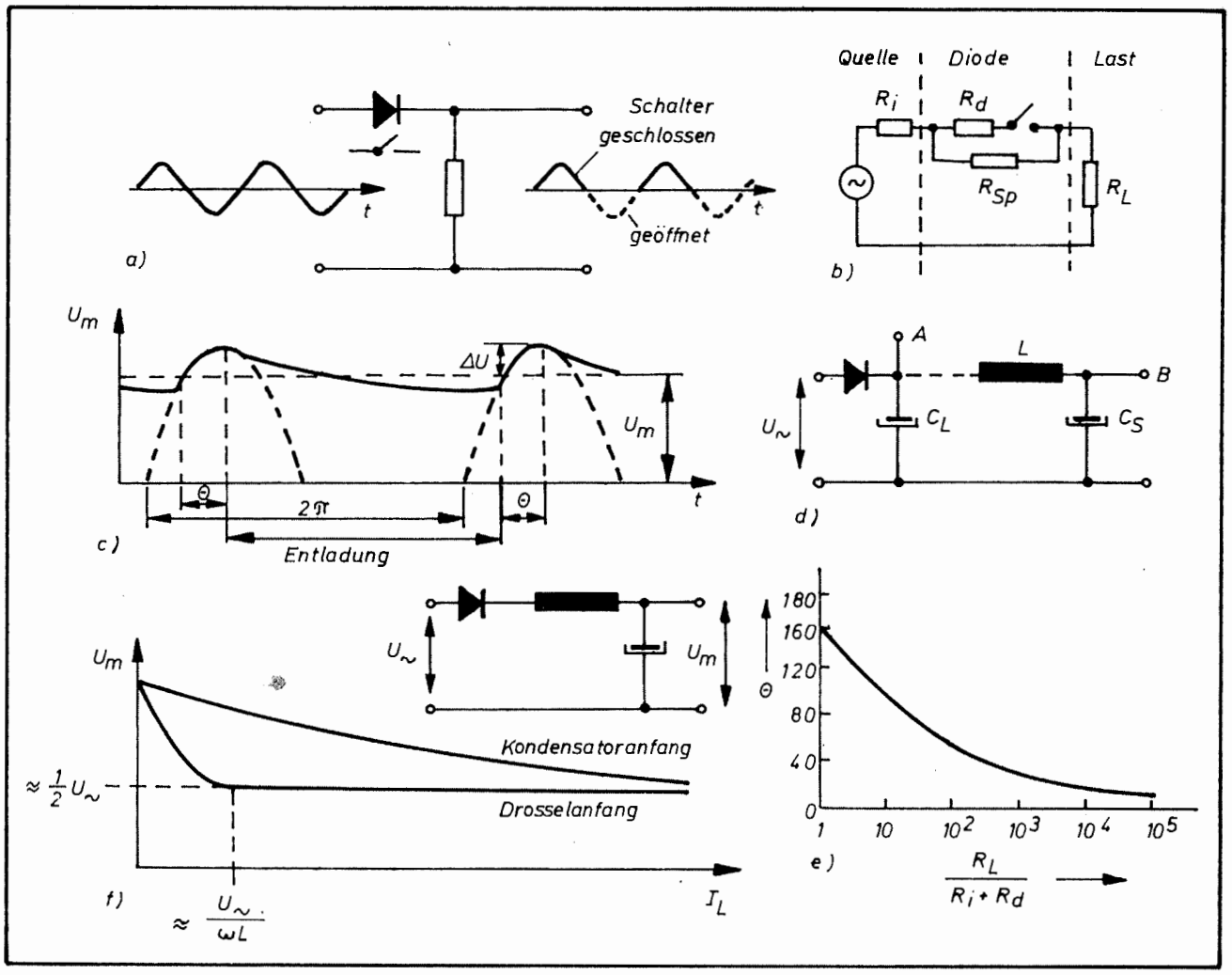


Bild 5.4.4. Einfache Gleichrichterschaltungen

a) Wirkung eines Gleichrichters als Schalter
b) Ersatzschaltung für a) mit Verlusten
c) Verhältnisse bei einem Ladekondensator
d) Schaltung mit Ladekondensator und nachgeschalteter Siebkette
e) Stromflußwinkel als Funktion der Last- und Verlustwiderstände
f) Gleichspannungsabhängigkeit bei Schaltungen mit Kondensator- und Drosselanfang sowie Schaltung mit Drosselanfang

und (2) gilt in etwa Bild 5.4.4e). Mit kleinerem Stromflußwinkel fließt ein *erheblich vergrößerter Strom* durch den Gleichrichter. Wird ein Strom I_L entnommen, so gilt für den *Spitzenwert* des Stromes etwa

$$I_{sp} \approx I \cdot 2\pi/\Theta \,. \tag{4}$$

Wird der Laststrom geändert, so ändert sich auch die mittlere Gleichspannung U_m. Der Zusammenhang von U_m zu I_L geht aus Bild 5.4.4f) hervor. Es kann ein *Innenwiderstand der Gleichquelle* definiert werden,

$$R_w = \mathrm{d}U_m/\mathrm{d}I_L \,. \tag{5}$$

In erster *Näherung* gilt hierfür

$$R_w \approx (R_d + R_i)\, 2\pi/\Theta \,. \tag{6}$$

Er hängt also von der Last selbst ab und ist besonders groß bei kleinen Stromflußwinkeln.

Die *Spannungsänderung* ΔU wird durch den Effektivwert der Grundwelle U_{br} (Brumm) gegeben. Dann gilt

$$\boxed{U_{br}/\mathrm{V} \approx A \frac{I_L/\mathrm{mA}}{C_L/\mu\mathrm{F}}} \tag{7}$$

Der Faktor A hängt von der Schaltung ab (bez. der Schaltungen s. S. 238):

Einweggleichrichtung und GREINACHER-Verdopplung	4
DELON-Schaltung	3
Zweiweg- und GRAETZ-Schaltungen	1,5.

Die Brummspannung U_{br} ist offensichtlich um so größer, je größer der Stromflußwinkel ist. Ihr Wert zeigt in etwa an, wie weit die mittlere Gleichspannung U_m vom Spitzenwert des Wechselstromes $U_{sp\sim} = \hat{U} = \sqrt{2} \cdot U_\sim$ abweicht. Der Spitzenwert des Wechselstromes kann also mittels Gleichrichtung gemessen werden, indem die Brummspannung sehr klein, d. h. auch der Stromflußwinkel sehr klein ist. Dies ist erfüllt, wenn

$$R_L \gg 1/\omega C \tag{8}$$

sowie (1) bis (3) erfüllt sind.

Für die weitere Verwendung der Gleichspannung wird die Brummspannung oft noch mit einer *Siebkette* gemäß Bild 5.4.4.d) verringert. Dies entspricht der *Wirkung von Tiefpässen*, und der *Siebfaktor*

$$S \approx \omega^2 LC \tag{9}$$

gibt an, um wieviel die Wechselspannung gesenkt wird. Oft wird auch ein Widerstand statt einer Drossel angewendet, dann gilt

$$S \approx \omega RC\,. \tag{10}$$

Dabei geht aber an dem Widerstand Gleichspannung verloren, und außerdem wird der Innenwiderstand der Gleichspannungsquelle um seinen Wert erhöht. Deshalb ist es vorteilhaft, *statt der Drossel* einen nichtlinearen Widerstand (z. B. einen *Transistor*) einzusetzen, der einen geringen Gleichstromwiderstand und einen großen Wechselstromwiderstand besitzt. Auf weitere dazugehörige Probleme wird in Abschnitt 7.3.1. eingegangen.

In verschiedenen Fällen stört der kurze Stromflußwinkel der Kondensatoranfangsschaltung. Der Grund hierfür kann sowohl der hohe Spitzenstrom gemäß (4) als auch der hohe Innenwiderstand gemäß (6) oder, als Folge davon, die sich mit der Belastung stark ändernde mittlere Gleichspannung (Bild 5.4.4f)) sein. Einen Ausweg bietet hier die *Drosselanfangsschaltung* gemäß Bild 5.4.4f) (oben). Sofern hier

$$I_L \geqq U_\sim/\omega L \tag{11}$$

gewählt wird, treten am Ausgang nur noch sehr geringe Spannungsschwankungen mit der Stromänderung auf. Sie entsprechen etwa dem Widerstand $2(R_i + R_d)$, und der Stromflußwinkel erreicht fast den maximalen Wert von 180°. Die Energiespeicherung übernimmt hier vorwiegend die Drossel. Für die Brummspannung gilt etwa

$$U_{br} \approx \omega L I_L\,. \tag{12}$$

5.4.4. Gleichrichterschaltungen

Bild 5.4.5 gibt einen Überblick über Anwendungsbereiche von Gleichrichtern. Dieses große Spektrum bezüglich Strom und Spannung hat auch eine Vielzahl von Gleich-

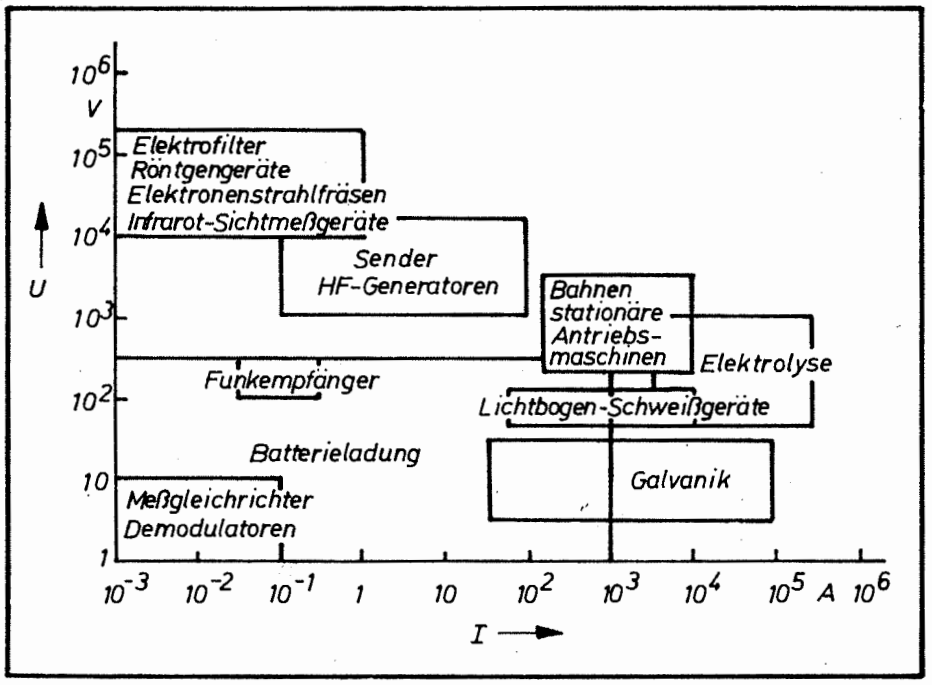


Bild 5.4.5. Anwendungsbereiche von Gleichrichtern

richterschaltungen entstehen lassen, die aber fast nur Abwandlungen des zuvor behandelten Grundprinzips darstellen. Die für Einphasenwechselstrom bedeutsamsten (ohne Einweg) sind in Bild 5.4.6 dargestellt.

Die *Zweiweggleichrichterschaltung* (auch Doppelweggleichrichtung genannt) zeigt Bild 5.4.6a). Sie benötigt einen Trafo mit Mittelanzapfung, und jede Hälfte wird nur in einer Halbwelle belastet. Auf diese Weise entsteht am Ausgang bei ohmscher Last ein Verlauf von Teilbild b). Die Zahlen deuten an, welcher Zweig zur jeweiligen Halbwelle beiträgt. Gegenüber Bild 5.4.4a) liegt also eine doppelt dichte Halbwellenfolge vor. Dies hat zur Folge, daß der Gleichspannungs- und Gleichstrommittelwert ebenfalls doppelt so groß ist. Für die übliche Kondensatoranfangsschaltung bleibt trotzdem das Bild 5.4.4c) im Prinzip gültig. Es treten jedoch *drei Änderungen* auf:

- Die Brummfrequenz ist doppelt so groß (also bei Netz 100 statt 50 Hz),
- der Stromflußwinkel ist nicht auf 2π, sondern nur auf π zu beziehen,
- der Kondensator braucht nur die Hälfte der Zeit Energie zu speichern.

Deshalb ist diese Schaltung in ihren wesentlichen Daten mehr als doppelt so gut. Der Faktor A in (7) sinkt z.B. von 4 auf 1,5, in (4) und (6) entfällt die 2, in (9) bis (12) und (3) ist ω mit $4\pi f$ einzusetzen. Der Hauptnachteil der Zweiweggleichrichtung ist die spezielle Sekundärwicklung des Trafos, so entstand bereits 1899 die GRAETZ-*Schaltung*, auch *Brückengleichrichter* genannt. Seine Schaltung zeigt Bild 5.4.6c). Er wirkt wie ein *doppelter Umschalter* (Teilbild d)), dies ist für die beiden Halbwellen in Teilbild e) und f) dargestellt. Die jeweils weggelassenen Gleichrichter sind gesperrt. Der Nachteil dieser Schaltung ist, daß die Trafospannung gegenüber der Gleichspannung keinen Gleichbezug hat, sondern auf zusätzlichem Wechselpotential liegt. Wie aus Bild 5.4.5 hervorgeht, werden oft Gleichspannungen benötigt, die über der Sperrspannung der meisten Dioden liegen. In diesen Fällen ist es z.T. möglich, Gleichrichter in Reihe zu schalten. Sie müssen dann aber gleichartige Sperrkennlinien besitzen. Andernfalls muß durch zusätzliche Parallelwiderstände hierfür gesorgt werden. Dies ist aufwendig und bedeutet zusätzliche Verluste. So hat die Reihenschaltung sich nur bei Selen- und Kupferoxidulgleichrichtern bewährt. Darüber hinaus ist auch ein Hochspannungstrafo oft problematisch. So entstanden verschiedene *Verdoppler*- und

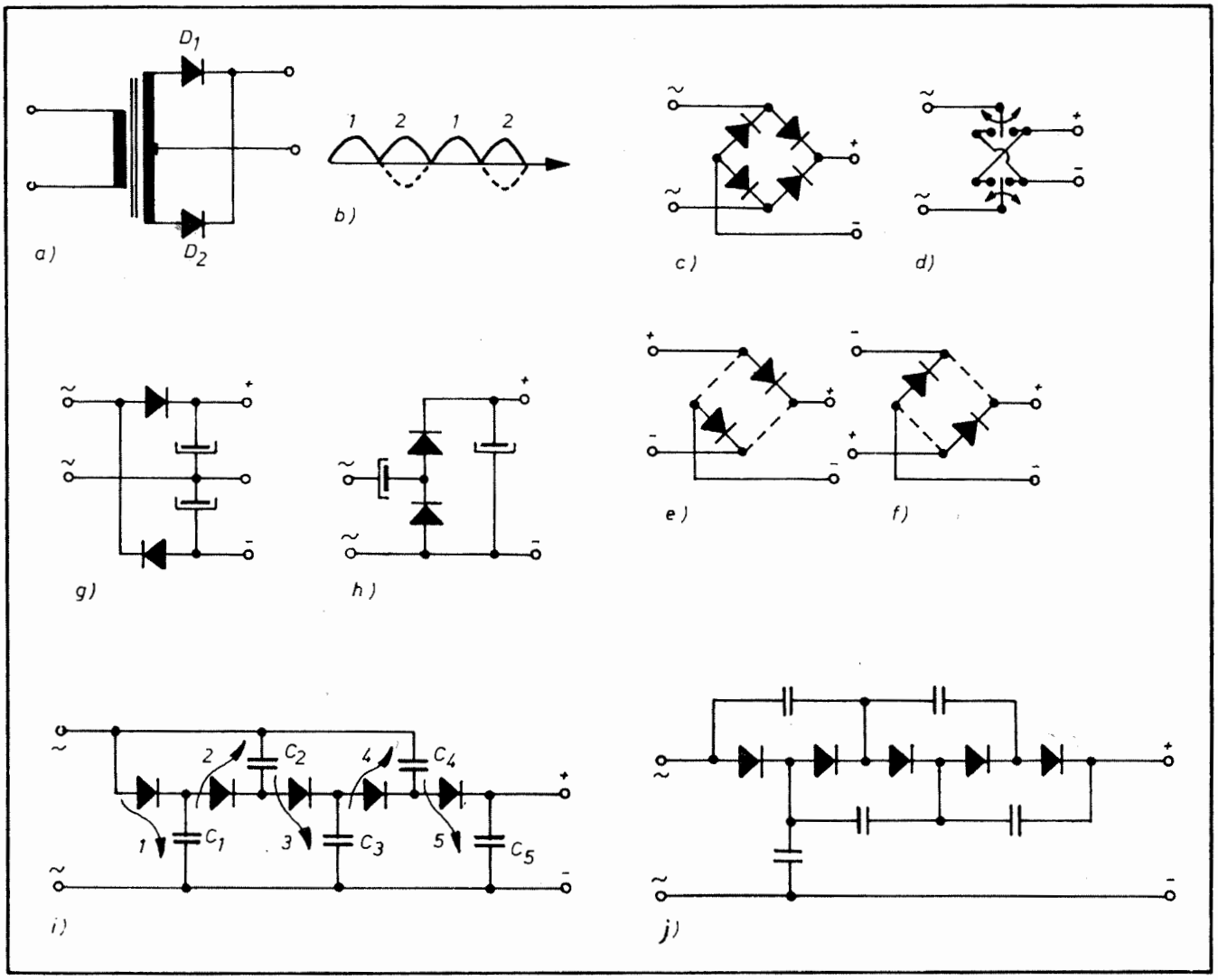


Bild 5.4.6. Gleichrichterschaltungen

a) Zweiweggleichrichter
b) Ausgangsstrom des Zweiweggleichrichters bei ohmscher Last
c) GRAETZ-Gleichrichter
d) Schalterersatzschaltbild des GRAETZ-Gleichrichters
e) Wirkung bei positiver Halbwelle
f) Wirkung bei negativer Halbwelle
g) DELON-Verdopplerschaltung
h) GREINACHER-Verdopplung
i) Mehrfache GREINACHER-Schaltung und ihre Funktion
j) Effektiver Aufbau der GREINACHER-Schaltung mit gleichen Kondensatoren

Vervielfacherschaltungen, bei denen die Ausgangsgleichspannung mehrfach so groß wie die Eingangswechselspannung ist.

Die einfachste ist wohl die DELON-*Schaltung*, die aus zwei Einweggleichrichtern für die beiden Halbwellen besteht (Bild 5.4.6g)). Es entsteht dabei eine positive und negative Spannung gegenüber dem Nullpotential. Wird nicht dieses Nullpotential weiterverwendet, so liegen die Phasen der beiden Brummspannungen gegeneinander verschoben. Dadurch ist auch der Faktor A in (7) nur 3 statt 4, wie bei der Einweggleichrichtung.

Eine andere Art der Verdopplung ist mit der GREINACHER-*Schaltung* (Bild 5.4.6h)) gegeben. Ihre Funktion ist so, daß sie beliebig fortgesetzt werden kann (Teilbild i)). Die Pfeile im Teilbild i) zeigen, wie die einzelnen Halbwellen die Kondensatoren aufladen und die Ladung des jeweils vorangegangenen Kondensators zusätzlich übernehmen. Bei diesem Prozeß treten natürlich Verluste auf, so daß es meist nicht sinn-

voll ist, über eine Verzehnfachung hinauszugehen. Der Hauptnachteil der GREINACHER-Schaltung nach Teilbild i) liegt darin, daß die einzelnen Kondensatoren für unterschiedliche Spannungen ausgelegt sein müssen, so hat sich die abgewandelte GREINACHER-Schaltung nach Bild 5.4.6j) allgemein eingeführt.

Alle bisher besprochenen Schaltungen lassen sich auch auf *Dreiphasenstrom* (*Drehstromnetz*) übertragen. Die entsprechenden Grundprinzipien zeigt dann Bild 5.4.7. Teilbild a) entspricht der Doppelweggleichrichtung und heißt Dreiphasenmittelpunktschaltung. Hier überlagern sich die Halbwellen bereits so, daß der Nullpunkt der Gleichspannung auch bei Widerstandslast nicht mehr erreicht wird. Die Brummfrequenz liegt bei $3f$ (Teilbild b)). Teilbild c) kann als systematische Fortsetzung der GRAETZ-Gleichrichtung und der DELON-Schaltung betrachtet werden. So entsteht bei ihr dann auch die doppelte Gleichspannung, und die Brummfrequenz beträgt sogar $6f$.

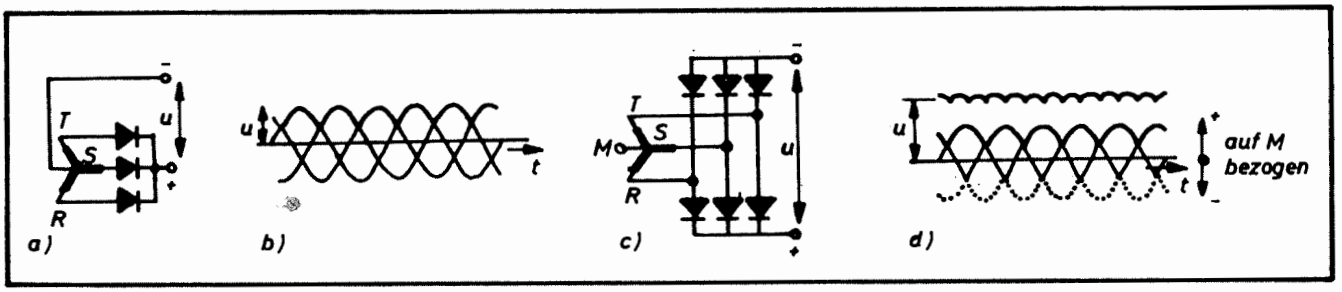


Bild 5.4.7. Gleichrichterschaltungen für Dreiphasennetze

a) Mittelpunktschaltung
b) Wechsel- und Gleichwerte bei der Mittelpunktschaltung
c) Drehstrombrückenschaltung
d) Spannungsverläufe für Drehstrombrückenschaltung; dünn: Wechselströme der drei Phasen; dick: Spannungen des unteren Zweiges gegenüber dem Mittelpunkt; dick unterbrochen: dito für den oberen Zweig; oben: Verlauf der Spannung U_m

5.4.5. Stabilisation von Spannungen

Die Stabilisation nutzt *nichtlineare Kennlinien* aus (vgl. Bild 1.4.8c)). Gemäß Bild 5.4.8a) und b) gilt für kleine Änderungen und für Gleichwerte

$$\Delta U = R_{\sim}\,\Delta I\,, \qquad U_0 = R_{-} I_0\,. \tag{13}$$

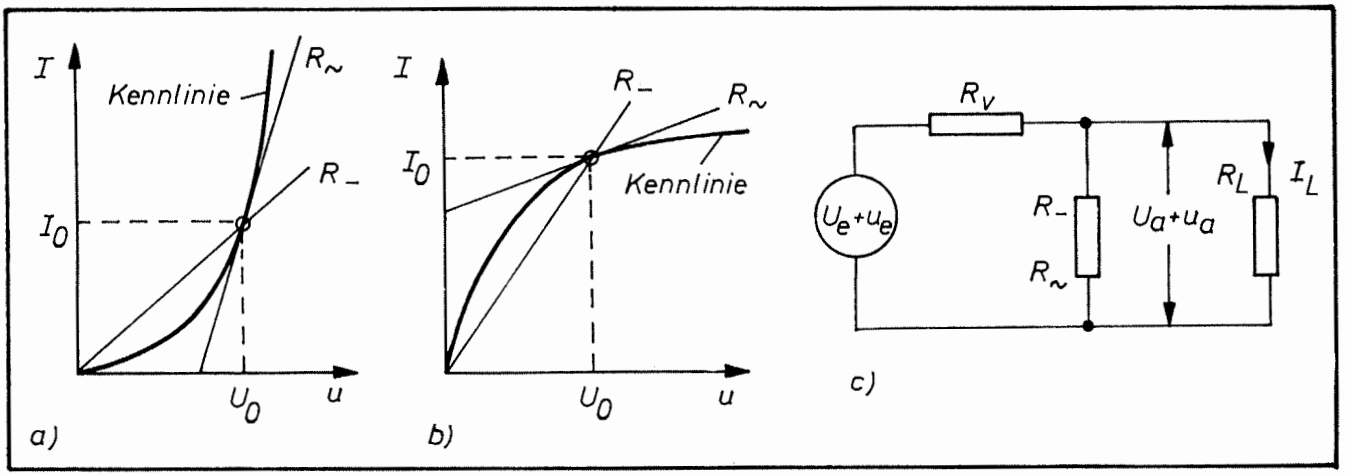


Bild 5.4.8. Zur Stabilisierung von Spannungen durch nichtlineare Widerstände

a) Kennlinie mit positiver Krümmung, wie bei normalen Dioden in Flußrichtung und ZENER-Dioden
b) Kennlinie mit negativer Krümmung, wie bei Transistoren
c) Prinzip der Spannungsstabilisierung mit Dioden (nichtlinearer Widerstand Typ a))

Für die *relativen Änderungen* folgt daraus

$$\frac{\Delta U}{U_0} = \frac{R_\sim}{R_-} \cdot \frac{\Delta I}{I_0}. \tag{14}$$

Dies bedeutet, daß sich relative Spannungsänderungen entsprechend dem Faktor $R_\sim/R_-$ als relative Stromänderungen auswirken und umgekehrt. In diesem Sinne kann die Kennlinie in Bild 5.4.8a) zur Spannungs- und Teilbild b) zur Stromstabilisation genutzt werden. Der Kennlinie a) entspricht die Diodenkennlinie im Durchlaßbereich und die Sperrkennlinie der Z-Diode, wenn die Lage im dritten Quadranten beachtet wird. Die Prinzipschaltung der zugehörigen Stabilisierung zeigt Bild 5.4.8c). Die Gleichspannungen sind darin mit großen und die Änderungen um sie mit kleinen Buchstaben gekennzeichnet, der nichtlineare Widerstand durch R_- und $R_\sim$. Mit der Parallelschaltung aus $R_\sim$ und R_L

$$R_{\sim||} = \frac{R_\sim R_L}{R_\sim + R_L} \tag{15a}$$

gilt dann für das *Verhältnis der Spannungsänderungen*, also den *Stabilisierungsfaktor*

$$\alpha = \frac{u_e}{u_a} = 1 + R_V/R_{\sim||}. \tag{15b}$$

Außerdem ist noch der *Einfluß von Lastwiderstandsänderungen* auf die Ausgangsspannung wichtig. Hierfür gilt

$$\frac{\mathrm{d}u_a}{\mathrm{d}R_L} = \frac{U_e R^2 R_V}{(R_\sim R_L + R_V R_L)^2}. \tag{16}$$

Mit den Relationen

$$R_V \gg R_\sim \quad \text{und} \quad R_L \gg R_\sim \tag{17a}$$

folgt

$$\frac{\mathrm{d}u_a}{\mathrm{d}R_L} \approx \frac{U_e}{R_V}\left(\frac{R_\sim}{R_L}\right)^2. \tag{17b}$$

Hieraus geht deutlich hervor, daß ein kleiner Wechselstromwiderstand $R_\sim$ des nichtlinearen Bauelementes (z.B. Diode) vorteilhaft beide möglichen Störungen senkt.

Nach dem Prinzip von Bild 5.4.8c) können Spannungen mittels Z-Dioden im Bereich von ca. 2 bis 20 V stabilisiert werden. Mittels normaler Dioden in Flußrichtung sind Werte im Bereich von 0,5 bis 1 V und durch Reihenschaltung darüber hinaus erreichbar.

5.4.6. Spannungsbegrenzung

Die Stabilisierung ist nur für Gleichstrom möglich. Mittels antiparalleler Dioden oder Reihenschaltung in entgegengesetzter Richtung von Z-Dioden ist eine Begrenzung von Wechselspannung realisierbar. Für normale Dioden zeigt dies Bild 5.4.9. Durch Vorspannung der Dioden ist diese Spannung darüber hinaus einstellbar. Hiermit ist noch besser die Funktion des Varistors (Bild 5.1.5) ersichtlich.

Auf ähnliche Weise ist es schließlich möglich, die negativen Überschwingimpulse beim Abschalten von Induktivitäten, z.B. Relais, zu unterdrücken (Bild 5.4.10).

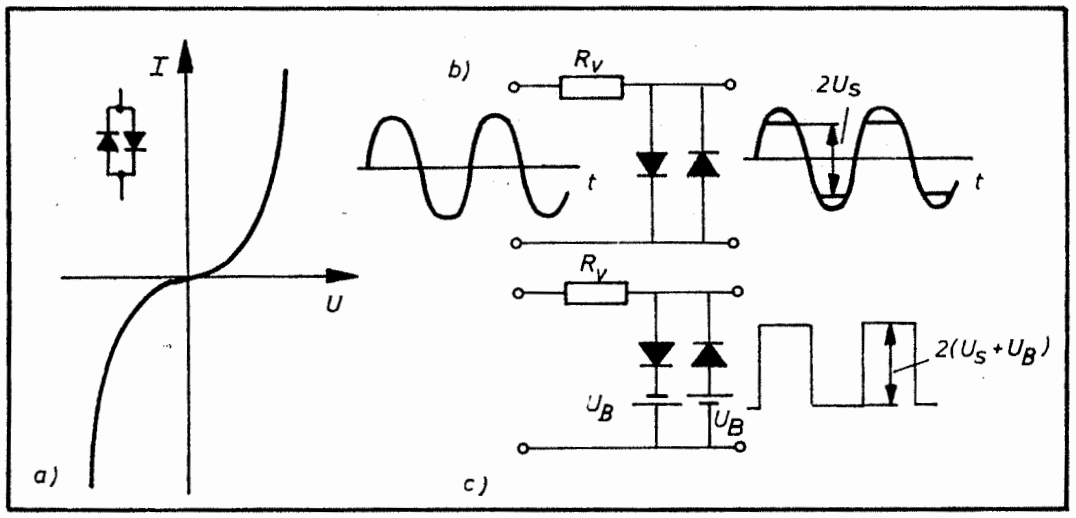


Bild 5.4.9. Spannungsbegrenzung mittels Dioden
a) Kennlinie für zwei antiparallele Dioden
b) Begrenzung mit einer Schaltung, die der Spannungsstabilisation bei Gleichstrom entspricht
c) Einstellung der Begrenzungsspannung durch zusätzliche Vorspannungen

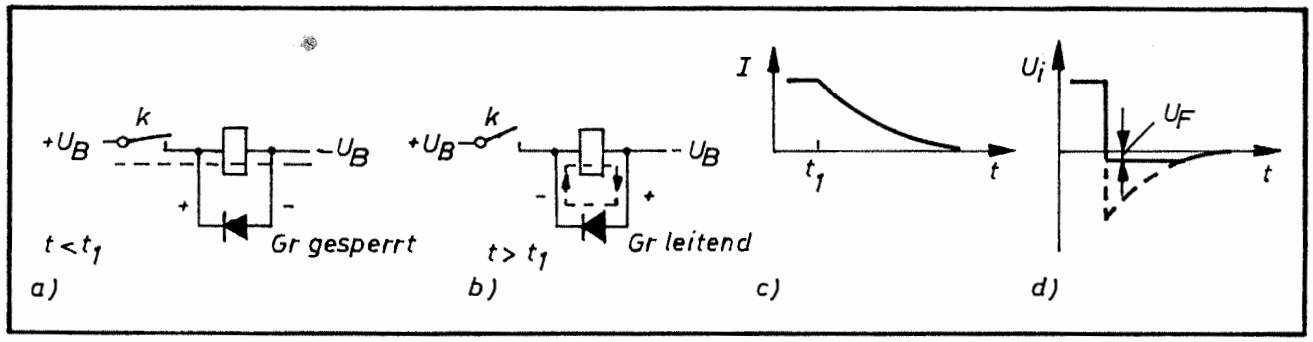


Bild 5.4.10. Wirkungsweise der Funkenlöschung mit Dioden
a) Kontakt geschlossen, b) geöffnet
c) Strom-, d) Spannungsverlauf

5.4.7. Demodulation von Amplitudenmodulation

Den Signalverlauf einer *amplitudenmodulierten HF-Schwingung* der Frequenz ω_{HF} zeigt Bild 5.4.11c). Die Modulation ist als Hüllkurve der Frequenz ω_H zu erkennen. Mittels Gleichrichtung kann nun die *Halbwelle* gemäß Teilbild d) *abgeschnitten* werden. Hierzu kommt eine der beiden Schaltungen von Teilbild a) oder b) in Betracht. Die Zeitkonstante aus C_2 und R_L muß nun so gewählt werden, daß die HF unterdrückt wird, der *Kondensator also in den Pausen noch Spannung speichert*, aber andererseits der Hüllkurve folgt. Dies ist möglich, wenn gilt

$$\omega_H < \frac{1}{R_L C_2} < \omega_{HF}/10 \,. \tag{18}$$

Günstige Werte sind erreichbar, wenn $\omega_{HF} \geqq 100\,\omega_H$ gilt.

5.4.8. Zwei Frequenzen an nichtlinearer Kennlinie

Entsprechend Bild 5.4.12a) sollen jetzt zwei in Reihe geschalteten Quellen auf einen nichtlinearen Widerstand, z.B. eine Diode, einwirken. Es gelte

$$U_\Omega(t) = \hat{U}_1 \cos \Omega t \,, \qquad U_\omega(t) = \hat{U}_2 \cos \omega t \,. \tag{19a}$$

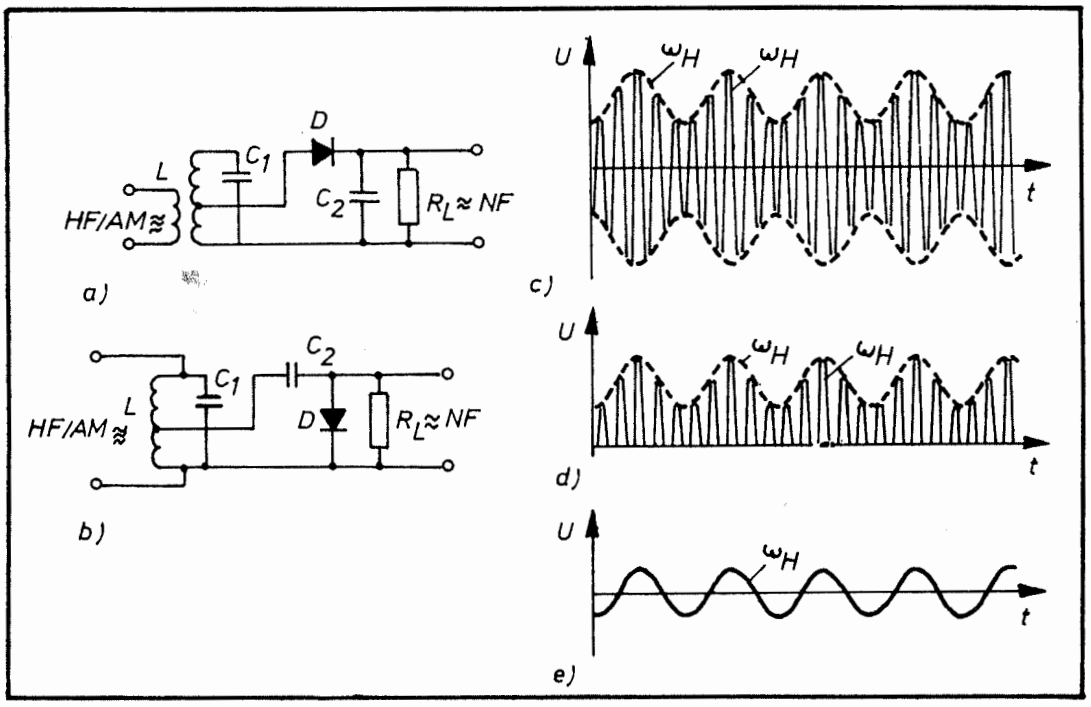


Bild 5.4.11. Zur Demodulation amplitudenmodulierter Signale

a) D und R_L in Reihenschaltung, b) in Parallelschaltung

c) HF-Signal mit Hüllkurve

d) HF-Signal nach Gleichrichtung. Die Hüllkurve wird durch passend gewählte Mittelwertbildung der Zeitkonstanten $R_L C_2$ erreicht

e) Vom Gleichwert befreite NF-Schwingung

In c) und d) ist das jeweils rechts stehende ω_H durch ω_{HF} zu ersetzen

Ganz ähnlich wie im Abschnitt 1.4.8. läßt sich zeigen, daß der Strom folgendermaßen geschrieben werden kann:

$$I(t) = \sum_{n,m=0}^{\infty} \hat{I}_{n,m} \cos(n\Omega \pm m\omega)\, t\,. \tag{19b}$$

Es treten *Kombinationsfrequenzen* auf. Die zugehörigen *Stromamplituden* $I_{n,m}$ hängen hauptsächlich *von drei Größen* ab:

— der Kennlinie des Gleichrichters,
— der Aussteuerung der Kennlinie,
— dem Amplitudenverhältnis $\hat{U}_1$ zu $\hat{U}_2$.

Relativ anschaulich werden die Zusammenhänge, wenn eine der beiden Spannungen groß angenommen wird. Ohne Beschränkung der Allgemeinheit gelte

$$\hat{U}_1 \approx \hat{U}_2\,. \tag{20}$$

Dann kann die Schaltung von Bild 5.4.12a) durch jene von Bild 5.4.12b) ersetzt werden. Die Diode wirkt als *Schalter, der im Rhythmus der Frequenz Ω betätigt* wird. Unter dieser Voraussetzung zeigt das Bild für zwei Frequenzverhältnisse $\Omega:\omega$ gleich 3:2 und 7:1 die Zeitverläufe des Stromes. Im Teilbild c) ist leicht die Differenzfrequenz zu erkennen. Sie ist zugleich sehr oberwellenreich. Im Teilbild d) ist infolge des Phasensprungs bei A gut die Summenfrequenz zu erkennen. Sie erscheint noch deutlicher, wenn die halbe Amplitude der Sinuswelle als (Gleichanteil) entsprechend der dünnen Linie subtrahiert wird. Dann entsprechen den sieben gestrichelten Rechteckschwingungen deutlich acht Schwingungen der ausgezogenen Kurve. Auch diese Schwingung ist stark oberwellenhaltig. Damit ist qualitativ der Zusammenhang von Gl. (19b) verständlich.

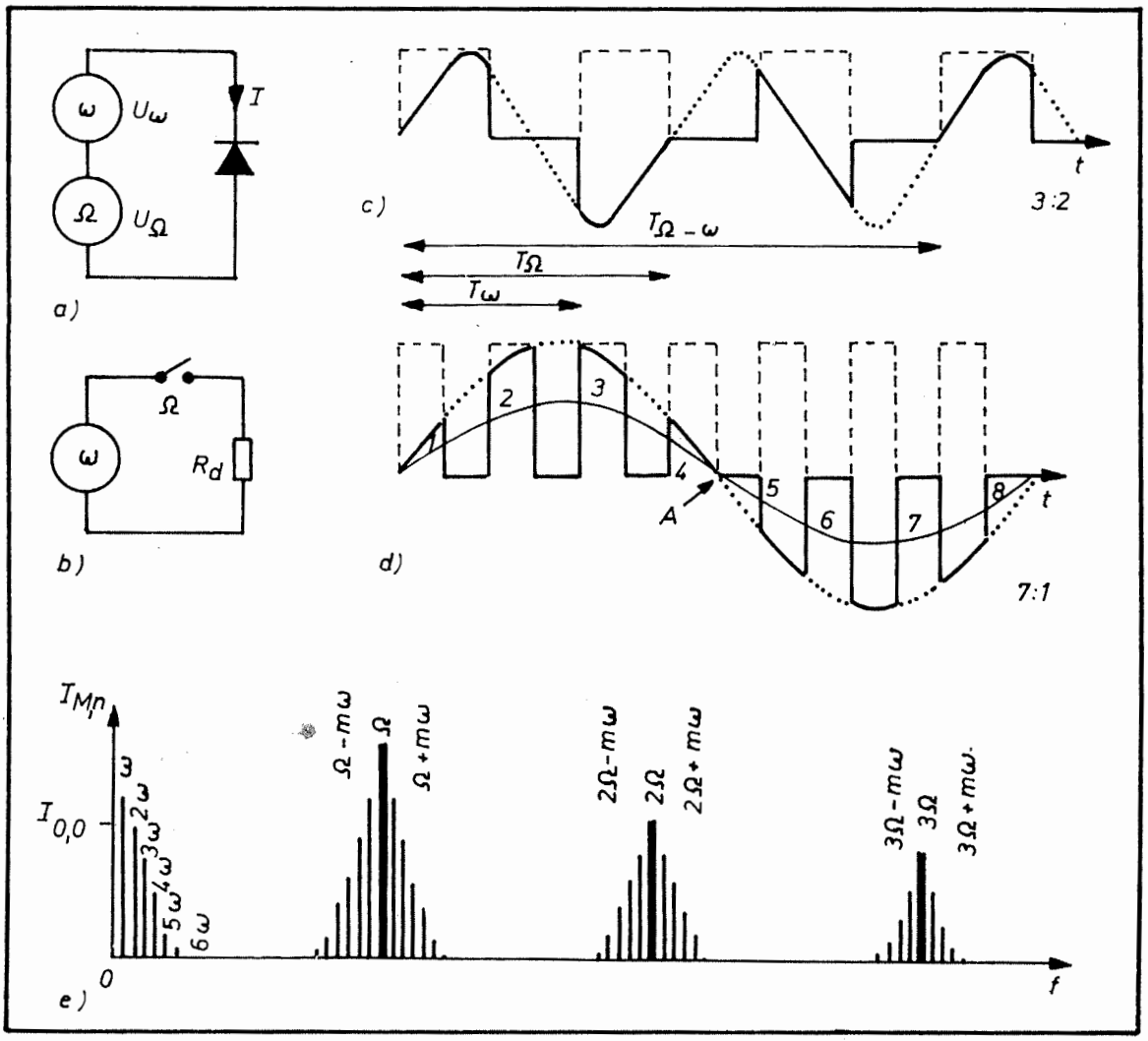


Bild 5.4.12. Einwirken von zwei Frequenzen auf eine nichtlineare Kennlinie

a) Prinzipschaltbild
b) Ersatzschaltung, wenn die Amplitude von Ω groß gegenüber der von ω ist
c) Signalverlauf bei einem Frequenzverhältnis von 3:2,
d) bei 7:1
e) Prinzipielles Aussehen des Spektrums des Stromes bei $\Omega \gg \omega$

Eine Veranschaulichung der Verhältnisse ist auch in einer Spektraldarstellung möglich. Für $\Omega \gg \omega$ zeigt Bild 5.4.12e) die Stromamplituden. Je mehr sich Ω und ω nähern, desto geringer werden die Lücken zwischen den Spektralgruppen um die Vielfachen von Ω. Bei weiterer Näherung von Ω an ω durchdringen sich die einzelnen Gruppen, und schließlich entstehen sogar negative Frequenzen. Sie sind genau so wie positive, jedoch mit einem Phasenwinkel von 180° zu behandeln. Je näher Ω und ω liegen, desto schwieriger ist eine bestimmte *Kombinationsfrequenz*, z.B. $\Omega \pm \omega$ aus den sich gegenseitig durchdringenden *Frequenzgruppen auszufiltern.*

5.4.9. Ringmodulator

Bereits im Zusammenhang mit Tab. 1.4.1. zeigte sich für eine Frequenz, daß je nach der typischen Form der Kennlinie bestimmte Oberwellen unterdrückt werden. Dies gilt ähnlich für zwei Frequenzen. Bedeuten u die ungeraden Harmonischen mit 1 als Grundwelle und g die geradzahligen mit 0 als Gleichteil, so sind neben dem allgemeinen Fall aller Kombinationsfrequenzen $n\Omega \pm m\omega$ solche wie $u\Omega \pm m\omega$; $g\Omega \pm u\omega$;

$n\Omega \pm g\omega$ usw. zu unterscheiden. Dies wird ausführlich in [H 19] untersucht. Für antiparallele Dioden gilt z. B. die Bedingung, daß $m + n = u$ sein muß; beim GRAETZ- und Zweiweggleichrichter dagegen $m + n = g$. Mit komplizierteren Schaltungen lassen sich andere Spezialfälle realisieren. Für viele Anwendungen sind die Fälle $u\Omega \pm m\omega$ und $u\Omega \pm u\omega$ günstig. Dann ist es nämlich besonders leicht möglich, die meist interessierenden Frequenzen $\Omega \pm \omega$ herauszufiltern. Durch günstige Wahl der Kennlinie und des Arbeitspunktes auf ihr ist es weiter möglich, ihre Amplitude im Vergleich zu den anderen Frequenzen recht groß zu erhalten (günstiger Störabstand).

Derartige Bedingungen erfüllen der *Gegentaktmodulator* und der *Ringmodulator*. Die entsprechenden Schaltungen zeigt Bild 5.4.13. Damit wirklich die unerwünschten Kombinationsfrequenzen unterdrückt werden, sind *hohe Anforderungen* an die *Gleichheit der Diodenkennlinien* und an die *Symmetrie der Transformatoren* zu stellen. Deshalb erfolgt in der Praxis hier ein aufwendiger Abgleich.

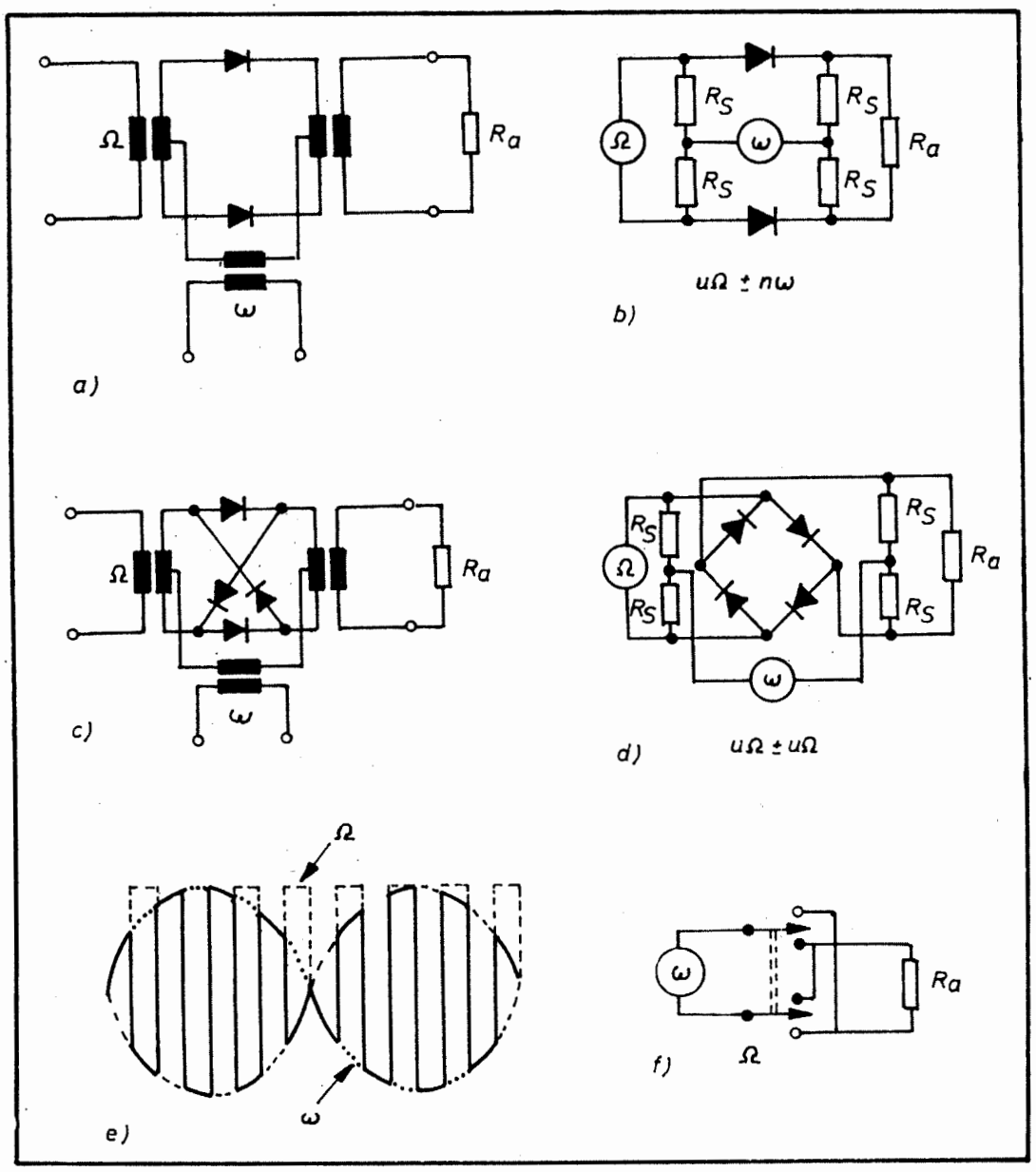


Bild 5.4.13. Gegentakt- und Ringmodulator

a) Gegentaktmodulator mit Trafos, b) mit Symmetrierwiderständen
c) Ringmodulator mit Trafos, d) mit Symmetrierwiderständen
Aus d) geht deutlich die Ringschaltung der Gleichrichter hervor, die in der Polung vom GRAETZ-Gleichrichter abweicht
e) Spannungsverläufe beim Ersatzschaltbild des Ringmodulators durch einen Umpolschalter (f) im Takt der Frequenz Ω
In e) sind deutlich die beiden Frequenzen $\Omega \pm \omega$, einmal als Hüllkurve der Spannung mit dem Phasensprung die Differenzfrequenz und zum anderen als modulierte Rechteckspannung die Summenfrequenz

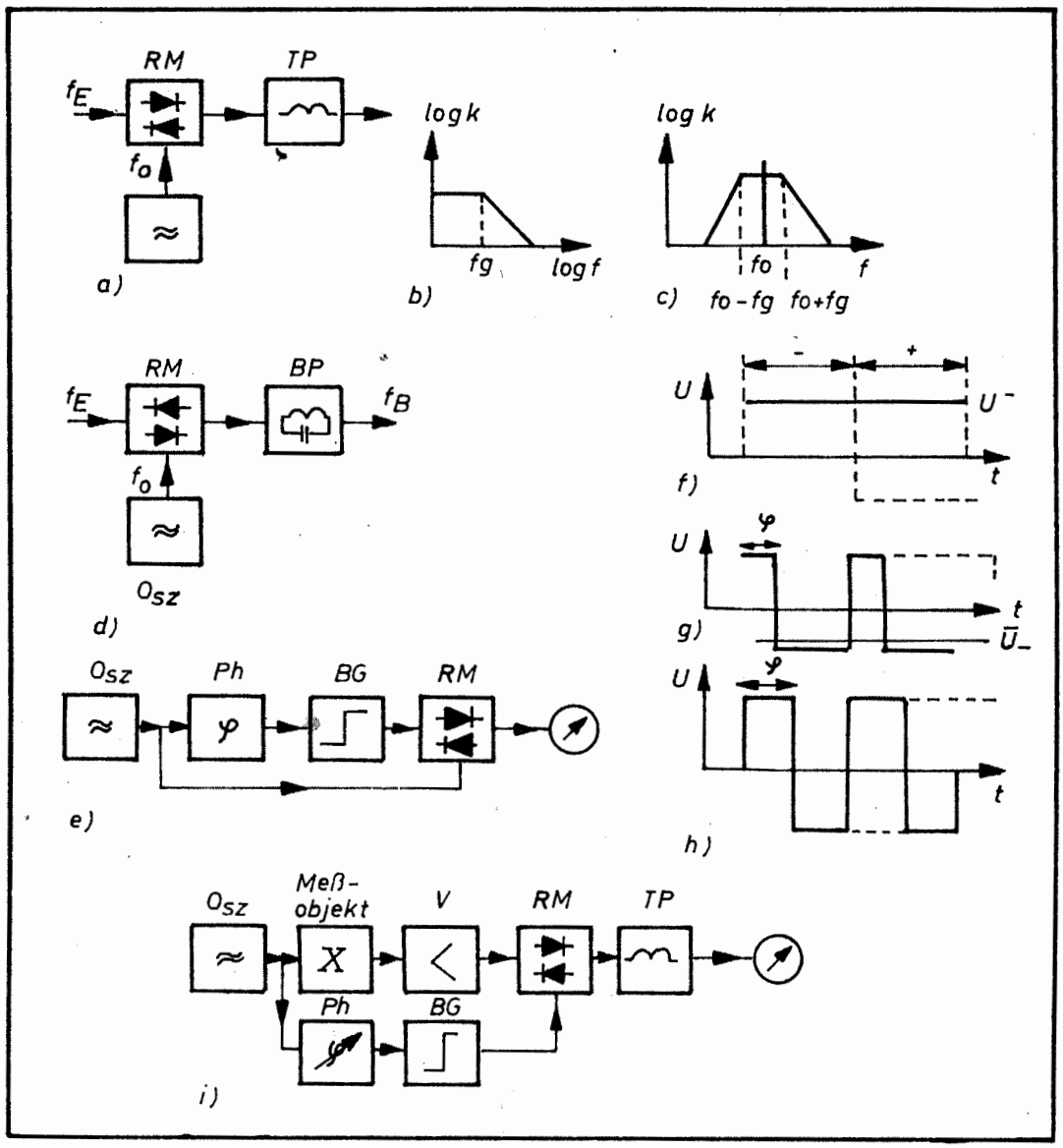


Bild 5.4.14. Anwendungsbeispiele des Ringmodulators

a) Als schmalbandiger Bandpaß um die Frequenz des Oszillators mit f_0

b) Tiefpaßkurve, c) ihre Wirkung bezüglich der Eingangsfrequenzen

d) Prinzip der Frequenzumsetzung (Mischen)

e) Phasenmessung mit ihrer Wirkung (in f) bis h)). Dabei werden in f) die beiden Halbphasen mit + und − bezeichnet (Umschalter von Bild 5.4.13 f)). Für die Phasenlagen gilt:

f) − gegenphasig (180°). Deshalb wird genau in ihrem Umschlagspunkt umgeschaltet. Als mittlere Gleichspannung tritt die volle positive Amplitude auf

g) − verschoben um $\varphi \approx 50°$. Es tritt eine mäanderförmige Schwingung mit einer mittleren negativen Gleichamplitude $\bar{U}_-$ auf

h) − verschoben um $\varphi \approx 90°$. Daher ist die mittlere Ausgangsgleichspannung Null

i) Phasenempfindlicher Gleichrichter (Lock-in-Schaltung). Dem Meßobjekt wird eine Meßschwingung zugeführt.

Bei Messung von Lichtströmen erfolgt dies z. B. durch eine periodische Unterbrechung des Lichtweges. Die Meßspannung aus dem Meßobjekt wird in V verstärkt und dem Eingang des Ringmodulators zugeführt. Parallel zu diesem Signalweg wird die Oszillatorschwingung nach Durchlaufen eines einstellbaren Phasenschiebers Ph und Begrenzers BG dem Ringmodulator als Schaltspannung zugeführt. Mit dem Phasenschieber wird die Ausgangsspannung des Ringmodulators (Anzeige) auf Maximum eingestellt. Dann ist die Anzeige nur noch von der Meßamplitude am Ausgang des Meßobjektes abhängig. Das verstärkte Signal und wesentlich der Störabstand wird von der Integrationszeit am Ausgang des Ringmodulators bestimmt.

Hauptanwendungen des Ringmodulators sind:

— Amplitudenmodulation,
— Einseitenbandmodulation,
— Frequenzumsetzung,
— Aufbau extrem schmalbandiger Filter,
— Phasenmessung,
— phasenempfindlicher Gleichrichter.

Einige seien im folgenden kurz erklärt. Der *Ringmodulator* RM wird gemäß Bild 5.4.14d) vom Oszillator mit der Frequenz $\Omega = f_0$ gesteuert. Mit der Eingangsfrequenz f_E stehen dann an seinem Ausgang vor allem die Frequenzen $f_0 \pm f_E$. Durch den Bandpaß BP mit der Mittenfrequenz f_B kann eine von beiden ausgefiltert werden. Die vollständige Schaltung realisiert damit die *Frequenzaddition bzw. -subtraktion.* Wird bei einer derartigen Schaltung die Oszillatorfrequenz geändert, so passiert die komplette Schaltung nur eine Eingangsfrequenz. Nach diesem Prinzip arbeitet u. a. das *Frequenzsuchtonverfahren* und die Abstimmung bei unseren Rundfunkempfängern. f_B wird dann als *Zwischenfrequenz* bezeichnet. Damit nicht beide Frequenzen wirksam werden, muß eventuell f_E auch gut ausgefiltert werden. Das dabei erreichte Verhältnis zwischen gewünschter und unerwünschter Eingangsfrequenz heißt *Spiegelselektion.*

An die Stelle des Bandpasses in Bild 5.4.14d) kann auch ein Tiefpaß geschaltet werden. So entsteht die Schaltung von Bild 5.4.14a). Der Tiefpaß möge ein einfaches *RC*-Glied mit der Grenzfrequenz f_g sein (Bild 5.4.14b). Entsprechend werden zum Ausgang hin nur Frequenzen aus dem Bereich

$$f_E = f_0 \pm f_g \tag{21}$$

übertragen. Es liegt also eine *Selektionskurve* gemäß Bild 5.4.14c) vor. Sie ist dabei wieder vollkommen von der *Mittenfrequenz* f_E, die ja *nur durch die Oszillatorfrequenz* f_0 *bestimmt* wird, unabhängig. Außerdem berücksichtigt sie auch bereits die Spiegelfrequenz mit. Da Tiefpässe mit nahezu beliebig niedrigen Grenzfrequenzen (z.B. 1 Hz und weit darunter) leicht aufbaubar sind, können Eingangsfrequenzen mit derselben Genauigkeit wie die Oszillatorfrequenz bestimmt werden. Frequenzänderungen können dann nach diesem Prinzip mit sehr großer Genauigkeit gemessen werden.

Weiter sei noch die Anwendung des Ringmodulators zur *Phasenmessung* beschrieben. Hierzu diene Bild 5.4.14e). Eine Meßspannung schaltet den Ringmodulator im Takt ihrer Frequenz. In einem zweiten Signalweg liege das bekannte oder zu untersuchende Glied Ph mit der Phasenverschiebung φ. Damit seine Dämpfung für verschiedene Frequenzen oder der Einfluß verschiedener zu untersuchender Schaltungen verschwinden, folgt ein Begrenzer BG z.B. in Form eines SCHMITT-Triggers. Am Ausgang entsteht dann eine Gleichspannung, die nur vom Phasenwinkel abhängig ist und direkt angezeigt werden kann:

$$U = k\left(\varphi - \frac{\pi}{2}\right) \quad \text{für} \quad 0 \leqq \varphi \leqq \pi\,. \tag{22}$$

Die Wirkung sei mittels der konstruierten Kurvenverläufe von Bild 5.4.14f) bis h) für drei Phasenwinkel, die gemäß Ersatzschaltbild 5.4.13f) entstanden, erklärt. Daraus ist auch ersichtlich, daß die Frequenz selbst keinen Einfluß auf die Anzeige hat.

Wird die Schaltung von Bild 5.4.14e) etwas geändert, so entsteht Bild 5.4.14i). Diese Schaltung wird zur optimalen Rauschunterdrückung angewendet und heißt *phasenempfindlicher Gleichrichter* bzw. *Lock-in-Schaltung* oder *Synchrodyndetektor.* Das Rauschen eines Meßobjektes ist ja etwa proportional der Bandbreite B. Sie hängt bei

dieser Schaltung nur noch von der Meßzeit T_m, d. h. der Zeitkonstante des Meßinstrumentes ab. Sie kann aber sehr groß, z.B. bis zu mehreren Sekunden, gewählt werden.

5.5. Feldeffekttransistoren

Feldeffekttransistoren (FET) zeichnen sich durch folgende *Eigenschaften* aus:

- Die Steuerung erfolgt leistungslos; es wird nur Blindleistung für das Umladen der Gate- und MILLER-Kapazität benötigt. Dadurch liegen besonders übersichtliche Vierpolparameter vor ($I_e \approx 0$).
- hohe Betriebsfrequenzen, kurze Schaltzeiten;
- positiver Temperaturkoeffizient des Drainwiderstandes, dadurch besteht ein gewisser Selbstschutz bez. Überlastungen und die Möglichkeit zur einfachen Parallelschaltung;
- relativ gut lineare Steuerkennlinie;
- kleiner Durchschaltwiderstand (kleinste Restspannung), also geringe Verluste;
- geringes Rauschen.

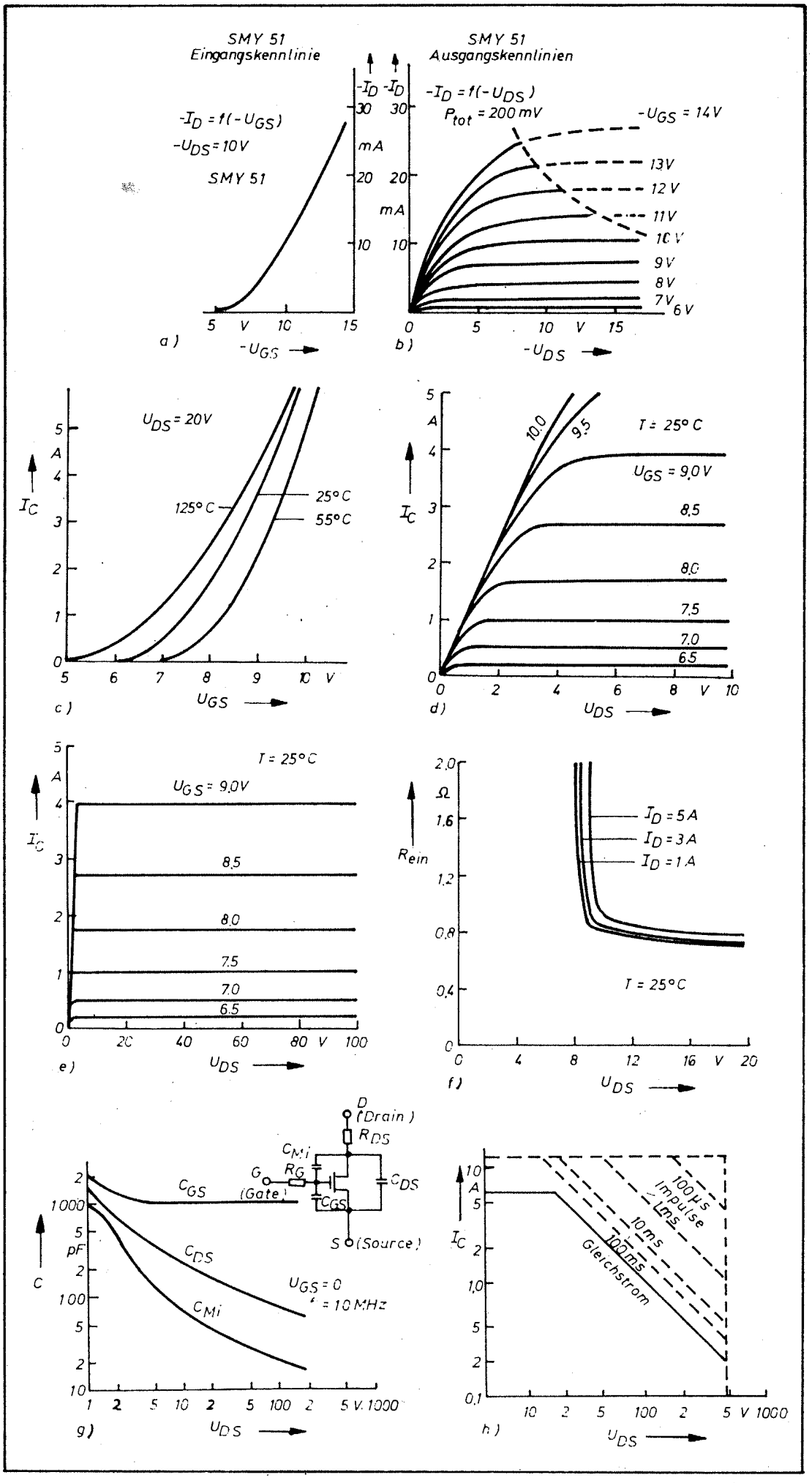
Nachteilig bei den MOSFET ist die *Empfindlichkeit der Gateisolation* gegenüber elektrostatischen Ladungen. Weiterhin waren zur Realisierung technisch brauchbarer FET *technologische Schwierigkeiten* zu überwinden. So standen die ersten brauchbaren Vorstufen-FET erst 1964 und die ersten Leistungs-FET erst 1974 serienmäßig zur Verfügung.

Tabelle 5.5.1. Daten von zwei etwa typischen Vorstufen- und vier ausgewählten Leistungs-FET

	MOSFET	SGFET	Leistungs-VMOSFET			
U_{SD}, V	20	20	100	100	100	500
I_D, A	0,01	0,02	0,5	2	30	5
R_d, Ω	100	50	0,5	3	0,05	2
S, A/V	0,001	0,005	0,1	0,5	15	2
C_G, pF	5	12	20	50	3000	1000
C_{GD}, pF	1,5	2				30
C_D, pF	5	6				100
$t_{ein/aus}$, ns	(300 MHz)	(300 MHz)	5	5	100	500

Tab. 5.5.1 stellt Daten einiger typischer FET zusammen. Bild 5.5.1 ergänzt hierzu einige Kennlinien. Die typischen Daten der großen Leistungs-FET liegen bei etwa 100 W und 100 ns. Aber auch in das Gebiet der Höchstfrequenztechnik ist der FET vorgedrungen. Hier liefert die Spitzenwerte GaAs mit gegen 30 GHz bei Vorstufen und gegen 10 GHz mit mehr als 100 W. Bei den Vorstufen-FET werden dabei Rauschwerte um wenige dB erreicht, die sonst nur bei Tunneldioden und parametrischen Verstärkern möglich waren.

Das Hauptanwendungsgebiet der FET liegt bei den integrierten Schaltungen der Digitaltechnik (8.2.7.). Hier werden nur allgemeine Grundlagen der Verstärkerschaltungen des FET behandelt. Die meisten Schaltungen werden erst nach Behandlung des bipolaren Transistors im Kapitel 7. besprochen.


SMY 51
Eingangskennlinie
$-I_D = f(-U_{GS})$
$-U_{DS} = 10V$
SMY 51
a)
$-U_{GS}$
SMY 51
Ausgangskennlinien
$-I_D = f(-U_{DS})$
$P_{tot} = 200\,mV$
$-U_{GS} = 14V$
13V
12V
11V
10V
9V
8V
7V
6V
b)
$-U_{DS}$
$U_{DS} = 20V$
125°C
25°C
55°C
c)
U_{GS}
T = 25°C
$U_{GS} = 9{,}0V$
8,5
8,0
7,5
7,0
6,5
d)
U_{DS}
T = 25°C
$U_{GS} = 9{,}0V$
8,5
8,0
7,5
7,0
6,5
e)
U_{DS}
$I_D = 5A$
$I_D = 3A$
$I_D = 1A$
T = 25°C
f)
U_{DS}
D (Drain)
R_{DS}
C_{Mi}
G (Gate)
R_G
C_{DS}
C_{GS}
S (Source)
C_{GS}
C_{DS}
C_{Mi}
$U_{GS} = 0$
f = 1 MHz
g)
U_{DS}
100 µs Impulse
1 ms
10 ms
100 ms
Gleichstrom
h)
U_{DS}

5.5.1. Verstärkung und Ersatzschaltbilder

Da beim FET praktisch kein Eingangsstrom fließt, ist es zumindest bei niedrigen Frequenzen sinnvoll, nicht von den Vierpolgleichungen, sondern von einer direkten Schreibweise auszugehen, die schon bei der Röhrentechnik im Gebrauch war. Bild 5.5.2a) zeigt eine typische Verstärkerschaltung, wobei R_S und C_S vor allem zur Arbeitspunkteinstellung dienen (s. S. 252).

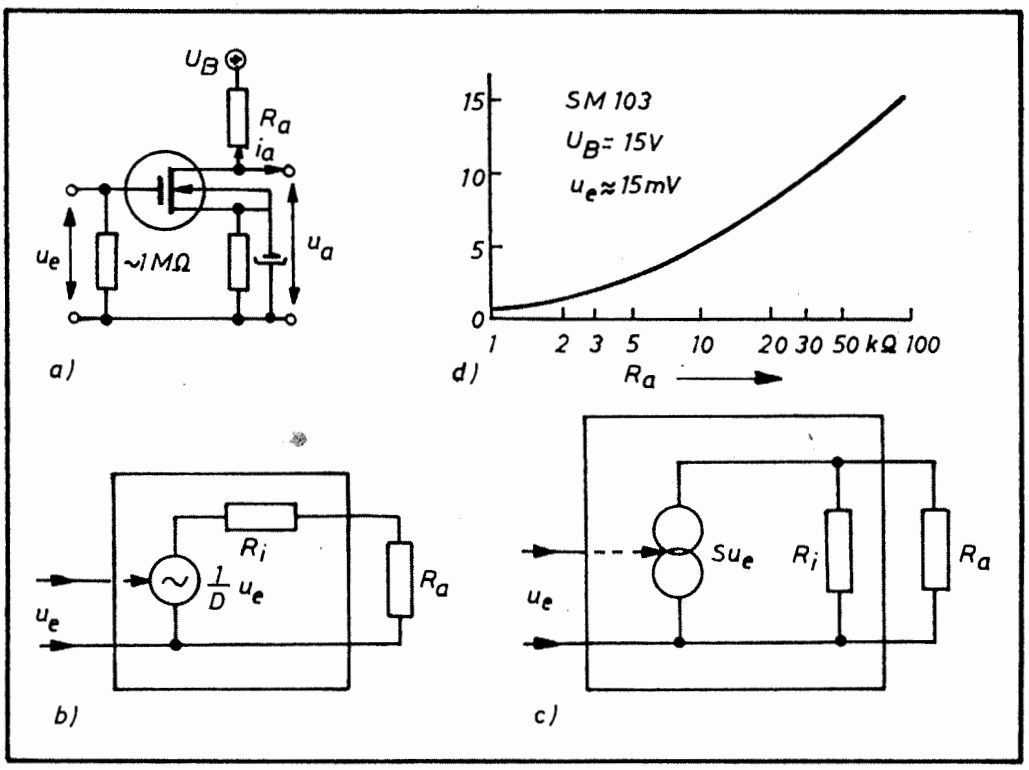


Bild 5.5.2. a) Schaltung eines n-Kanal-MOSFET vom Verarmungstyp mit Arbeitswiderstand
b) Spannungs-, c) Stromersatzschaltbild von FET-Verstärkerstufen
d) Beispiel der Abhängigkeit der Verstärkung eines FET bei konstanter Betriebsspannung und verändertem ohmschen Arbeitswiderstand

Entsprechend Abschnitt 3.3.7. gilt einmal für die Parameter des FET die BARKHAUSEN-*Beziehung*

$$\boxed{SDR_i = 1} \tag{1}$$

und zum anderen die *Steuergleichung*

$$\boxed{i_a = S(u_e + Du_a)} \,. \tag{2}$$

◀

Bild 5.5.1. Kennlinien zweier ausgewählter FET. Die Kennlinien eines analog zu verwendenden Vorstufen-FET SM 103 werden bei der Arbeitskennlinie in Bild 5.5.3 verwendet.

a) Eingangs-, b) Ausgangskennlinie eines Vorstufen-FET SMY 51 für vorwiegend digitale Anwendungen
c) bis h) Leistungs-MOSFET HPWP
c) Eingangskennlinien als Funktion der Temperatur
d) und e) Ausgangskennlinien für kleine und große Drain-Source-Spannungen
f) Ein-Widerstand für verschiedene Drainströme
g) Kapazitäten in Abhängigkeit von U_{DS}
h) Grenzkennlinien

Die kleinen Buchstaben deuten hierbei auf Wechselspannungen hinreichend kleiner Amplitude um den Arbeitspunkt hin, e bedeutet Eingang und a Ausgang. Anderseits gilt mit der Betriebsspannung U_B für die Drainspannung

$$U_D = U_B - I_D \cdot R_a \,. \tag{3}$$

Für die Wechselgrößen folgt daraus

$$u_a = -i_a R_a \,. \tag{4}$$

Hiermit kann der Ausgangsstrom in (2) ersetzt werden, und es folgt für die Verstärkung

$$\boxed{\begin{aligned} V &= -u_a/u_e && (5a)\\ &= \frac{1}{D}\,\frac{R_a}{R_a + R_i} && (5b)\\ &= S\,\frac{R_a \cdot R_i}{R_a + R_i} && (5c) \end{aligned}} \,.$$

Aus den letzten beiden Gleichungen folgen *je ein Ersatzschaltbild* (Bild 5.5.2b) und c)). Wird der FET im Pinch-off-Bereich betrieben, so ist wegen des großen R_i Ersatzschaltbild c) und Gl. (5c) geeigneter, im Triodenbereich dagegen b). Aus (5b) läßt sich der Faktor $1/D$ als *Grenzverstärkung*

$$\boxed{V_\infty = 1/D} \tag{6}$$

interpretieren. Sie wird für einen unendlich großen Arbeitswiderstand erreicht. Dies ist jedoch nur für Wechselstrom und näherungsweise möglich, da bei ohmschen Widerständen mit Änderung des Wertes auch der Arbeitspunkt und damit D sowie R_i und S geändert werden. Wird nur ein ohmscher Widerstand und konstante Betriebsspannung verwendet, so zeigt sich ein Verlauf, wie ihn beispielhaft Bild 5.5.2d) zeigt. Die Grenzverstärkung von etwa 30 wird hier nicht, auch nicht asymptotisch erreicht.

Im *Pinch-off-Bereich* und für $R_i > R_a$ geht (5c) über in

$$\boxed{V \approx S R_a} \,. \tag{7}$$

Auch hier muß (wie u. a. der Vergleich mit Bild 5.5.2d) zeigt) wieder ein fester Arbeitspunkt vorausgesetzt werden, da sich sonst S ändert.

5.5.2. Arbeitspunkt und Arbeitskennlinie

Der Arbeitswiderstand eines FET ist zur Erzeugung der Ausgangsspannung u_a notwendig. Dadurch entsteht der Unterschied von Betriebsspannung U_B und Drain-Source-Spannung U_{DS}. Außerdem ändert sich U_{DS} infolge von u_a. Dies bedeutet, daß ein FET mit Arbeitswiderstand eine andere Kennlinie, die Arbeitskennlinie, besitzt. Ihre Konstruktion im Kennlinienfeld des FET ist in Bild 5.5.3 gezeigt. Ausgegangen wird dabei von der Betriebsspannung U_B, und gemäß (3) wird die *Widerstandsgerade* in das Ausgangskennlinienfeld eingetragen. Sie beginnt bei $U_B = 15$ V und endet bei $U_{GS} = 0$ V und $I_D = 15$ V/R_a, in diesem Fall bei 10 mA ($R_a = 1500\ \Omega$). Der Arbeitspunkt wird so festgelegt, daß eine möglichst gut lineare Aussteuerung nach beiden Seiten möglich ist. Dies ist hier z. B. für $U_{GS} = -2$ V gegeben. Für den Spitzenwert der Eingangsspannung von 2 V ($u_{e,\,eff} \approx 1{,}41$ V) zeigt das Bild dann die Wechsel-

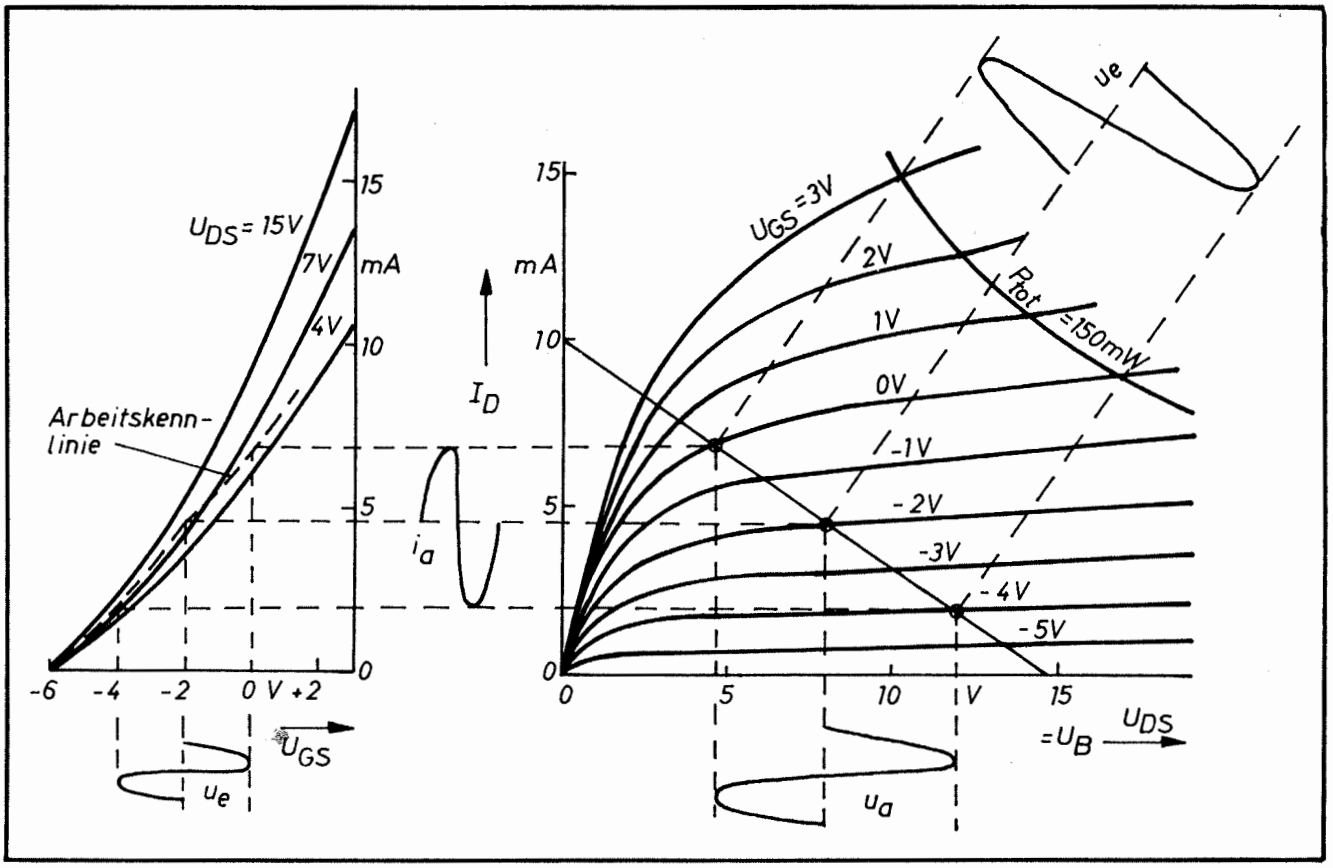


Bild 5.5.3. Konstruktion der Arbeitskennlinien mit eingetragener Aussteuerung

größen i_a links und u_a unten. Die Verstärkung beträgt ca. $V = (12 - 4{,}8)/4 \approx 1{,}8$. Sie ist infolge des klein gewählten R_a so klein. (Ein größerer Wert hätte in dem Kennlinienfeld so flach gelegen, daß die Konstruktion unübersichtlich geworden wäre. Wie später noch folgt, ist so aber eine sehr breitbandige Verstärkung möglich.) Der Wechselstrom beträgt $i_{a,\,eff} = (6{,}9 - 2)/\sqrt{8} \approx 1{,}73$ mA.

Mit Hilfe der Widerstandsgeraden kann auch die *Arbeits-Eingangskennlinie* bestimmt werden. Für sie werden die Schnittpunkte der Widerstandsgeraden mit den Kennlinien in das Eingangskennlinienfeld für die verschiedenen U_{DS} übernommen. Sie ist im Bild (links) gestrichelt dargestellt. Sie verläuft also weniger steil, dafür aber linearer. In etwa gilt für die *Arbeitssteilheit*

$$S_a = \frac{i_a}{u_e} = \frac{S}{1 + R_a/R_i}. \tag{8}$$

Aus der Konstruktion der Arbeitskennlinie ist ersichtlich, daß die *Wahl des Arbeitspunktes* stets durch eine entsprechende Gatespannung zu erfolgen hat. Sie wird in der Regel in der Schaltung selbst erzeugt. Beispiele hierfür zeigt Bild 5.5.4. Dabei ist zu beachten, daß im Teilbild d) zugleich eine Gegenkopplung (vgl. 6.1.) erfolgt. Im Teilbild b) und e) (Variante 2) ist die Gegenkopplung infolge des Elkos C nur für tiefe Frequenzen, insbesondere für Gleichstrom wirksam. Hierdurch werden (vgl. 7.1.) Stabilisierungen des Arbeitspunktes gegenüber Temperatur- und Betriebsspannungsschwankungen sowie eine Reduzierung der wirksamen Parameterstreuungen aus der Serie von FET erreicht.

5.5.3. *RC*-Verstärker

Nur in Sonderfällen ist es möglich, *mehrstufige Verstärker* direkt zu koppeln, so daß also auch Gleichpotentiale weiter verstärkt werden. Dieses Problem wird noch bei den

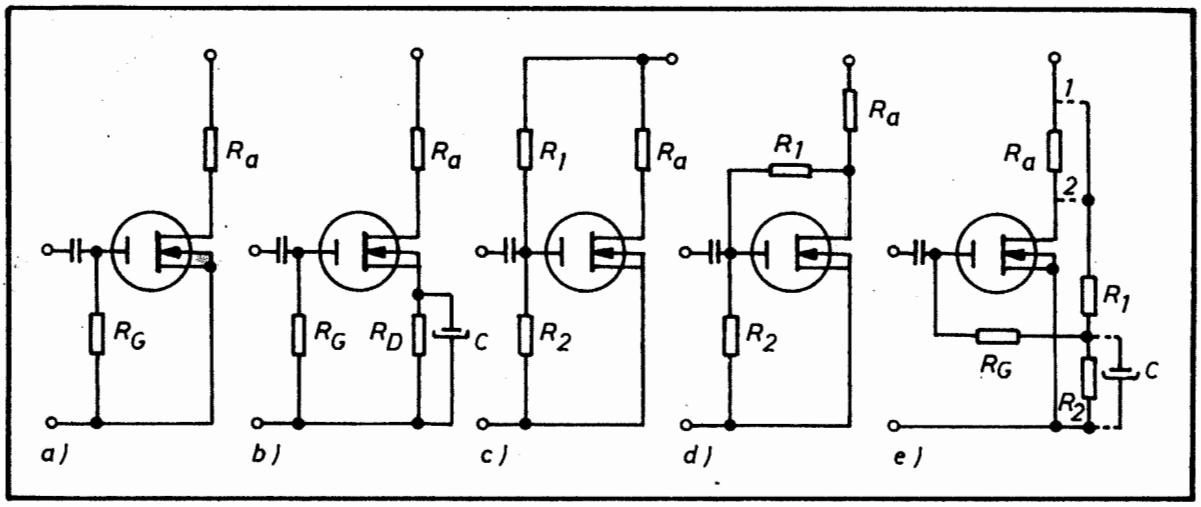


Bild 5.5.4. Zusammenstellung von Möglichkeiten zur Festlegung des Arbeitspunktes bei FET. Sie sind alle für n-Kanal-FET gezeichnet, gelten aber auch für die anderen Typen, bei p-Kanal-FET sind lediglich die Vorzeichen der Spannungen entgegengesetzt zu wählen.

a) Betrieb bei $U_G = 0$
b) U_G negativ, also Verarmungstypen
c) bis e) Anreicherungstypen
c) und d) Erzeugung der Gatespannung durch Spannungsteiler R_2 und R_1. Bei d) tritt zusätzlich Gegenkopplung auf, welche die Verstärkung verringert, aber die Stabilität der Schaltung erhöht
e) Zur Vermeidung sehr großer Widerstände R_G wird ein Spannungsteiler R_1 und R_2 verwendet. Die gepunktete Brücke 2 bewirkt im Gegensatz zu 1 Gegenkopplung und Arbeitspunktstabilisierung. Bei angeschaltetem C bleibt nur die Stabilisierung des Arbeitspunktes erhalten.

integrierten Schaltungen, z.B. Abschnitt 7.3.3., zu betrachten sein. In der Regel ist es *vorteilhafter, nur die Wechselspannungen weiter zu verstärken.* Dadurch entfällt nämlich die verstärkte Übertragung von Driftgrößen bezüglich des Arbeitspunktes auf die Folgestufen. So entsteht eine *RC-Kopplung*, wie sie Bild 5.5.5a) zeigt. Es werde angenommen, daß beide Verstärkerstufen etwa gleiche Eigenschaften besitzen und die Kopplung so gestaltet ist, daß die zu berechnende obere Grenzfrequenz groß gegenüber der unteren ist. Dann soll zunächst die *untere Grenzfrequenz* berechnet werden. Für sie können alle Schaltkapazitäten und FET-Kapazitäten vernachlässigt werden. Es wird die Ausgangsspannung der ersten Stufe als u_{a1} eingeführt. Weiter ist ihr Ausgangswiderstand die Parallelschaltung von Arbeitswiderstand R_a und Innenwiderstand R_i, also

$$R_i^+ = \frac{R_a \cdot R_i}{R_a + R_i}. \tag{9}$$

So entsteht die Ersatzschaltung von Bild 5.5.5b). Werden nun die Reihenfolge von Kopplungskondensator C_k und R_i^+ vertauscht, so wird deutlich, daß die Reihenschaltung von R_i^+ und R_G als ohmscher Spannungsteiler wirkt. Für die untere Grenzfrequenz gilt

$$\boxed{f_u = \frac{1}{2\pi C_k (R_G + R_i^+)} \approx \frac{1}{2\pi C_k R_G}}, \tag{10}$$

wobei die Näherung dann gilt, wenn $R_G > R_i^+$ oder meist $R_G > R_a$ gilt. Diese Formel kann auch zur Berechnung des Koppelkondensators benutzt werden.

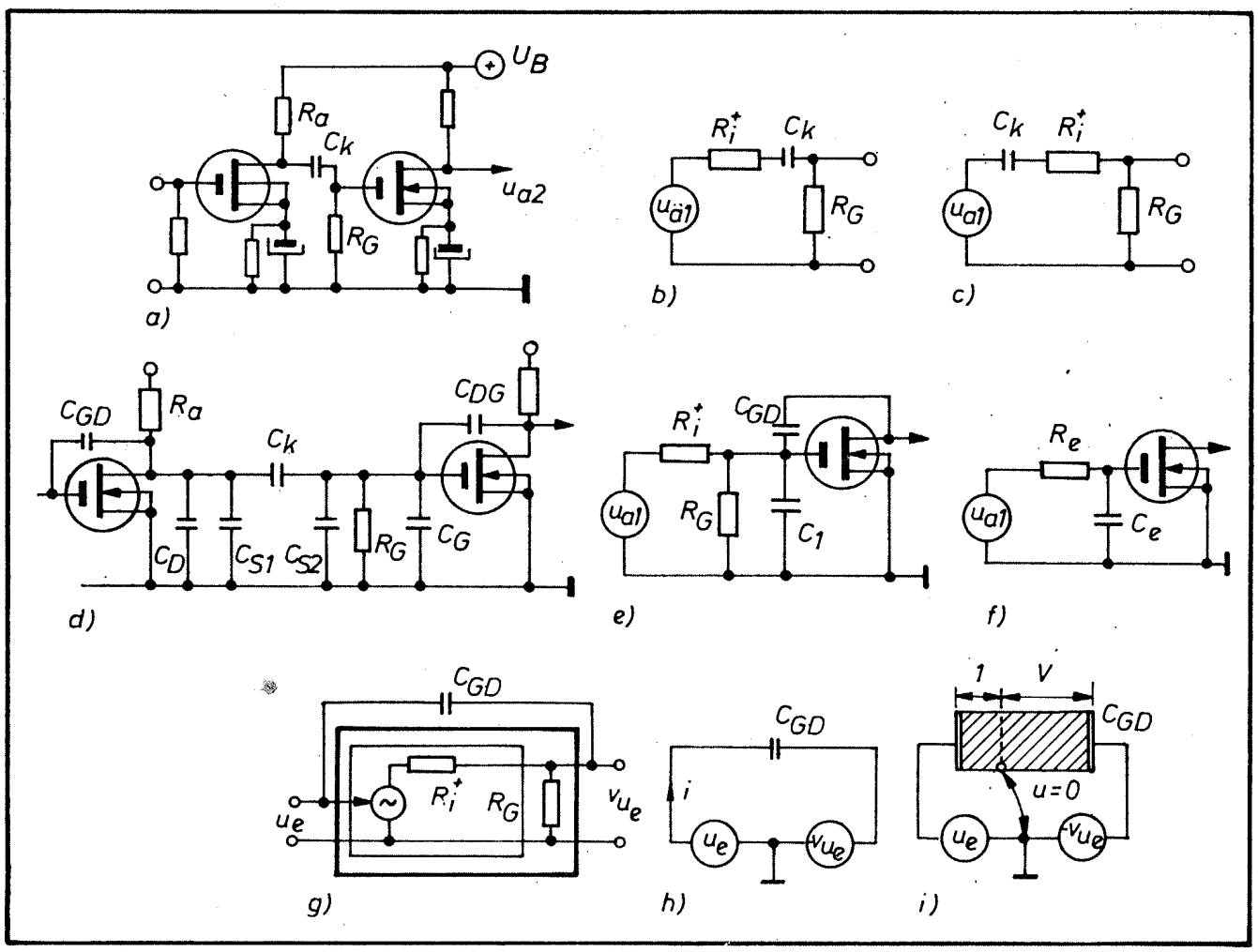


Bild 5.5.5. Zum Frequenzgang von *RC*-Verstärkern

a) Typische Kopplung zwischen zwei Verstärkerstufen
b) Ersatzschaltung bei tiefen Frequenzen (der Einfluß der Source-Elkos wird später erfaßt)
c) Abgeänderte Ersatzschaltung für tiefe Frequenzen, die deutlich den ohmschen Spannungsteiler erkennen läßt
d) Schaltung bei hohen Frequenzen mit sichtbarer Hervorhebung der Schaltkapazitäten C_{S1} und C_{S2}, der Drainkapazität C_D und der Gate-Drain-Kapazität C_{GD} des ersten FET sowie der Eingangskapazität C_G und der Gate-Drain-Kapazität C_{GD} des zweiten FET
e) Ersatzschaltbild unter Zusammenfassung der verschiedenen Kapazitäten mit Ausnahme von C_{GD}
f) Vollständig vereinfachte Ersatzschaltung bei hohen Frequenzen
g) Zur Berechnung der Wirkung von C_{GD} in der Schaltung
h) Zur Berechnung des Stromes durch C_{GD}
i) Zur Bestimmung des Nullpotentials innerhalb des Kondensators C_{GD}

Bei den *hohen Frequenzen* sind folgende Kapazitäten zu beachten:

C_{GD} — die Kapazität des FET zwischen Gate und Drain
C_D — die Kapazität des FET vom Drain zu Bulk und Source
C_G — die Gatekapazität des FET gegenüber Source und Bulk
C_{S1} — die Schaltkapazitäten vom Ausgang des ersten FET bis einschließlich der Leitungen und des Koppelkondensators gegenüber Masse
C_{S2} — die Schaltkapazitäten vom Koppelkondensator bis zum zweiten FET gegenüber Masse.

Sie sind in Bild 5.5.5d) deutlich herausgezeichnet. Infolge der Größe des Koppelkondensators können folgende Kapazitäten zusammengefaßt werden:

$$C_1 = C_{GD} + C_D + C_{S1} + C_{S2} + C_G \,. \tag{11}$$

So entsteht die Ersatzschaltung von Teilbild e). C_{GD} des ersten FET wird hier addiert, da die Spannung am Gate dieses FET vergleichsweise klein ist. Dies gilt dagegen nicht

für das C_{GD} des zweiten FET. Seine Wirkung muß gesondert betrachtet werden. Dazu sei zunächst Bild 5.5.5g) betrachtet. Es zeigt, daß am Eingang des FET u_e und am Ausgang $-Vu_e$ liegt. Das Minuszeichen stammt von der Phasenumkehr der Verstärkerstufe. Dieses zeigt prinzipiell Teilbild h). Der Strom durch den Kondensator beträgt also

$$i = \frac{(1 + V)\, u_e}{R_\sim} = u_e(1 + V)\, j\omega C_{GD} . \tag{12}$$

Dieser Blindstrom ist am Eingang des FET, also von u_e aufzubringen. Folglich liegt scheinbar eine Kapazität

$$C_{GD}^{+} = (1 + V)\, C_{GD} \tag{13}$$

vor. *Die Gate-Drain-Kapazität erscheint am Eingang um den Verstärkungsfaktor der Stufe vergrößert.* Dieser Effekt wird als MILLER-*Effekt* bezeichnet und eine außen entsprechend angeschaltete Kapazität (s. 7.5.5.) als MILLER-*Kapazität*.

Die Wirkung der MILLER-Kapazität soll auch noch zusätzlich durch Bild 5.5.5i) veranschaulicht werden. Im Dielektrikum der Kapazität wird die Potentiallinie des Wertes Null gesucht. Sie liegt genau dort, wo das Dickenverhältnis $1/V$ beträgt. Damit erscheint für u_e das Dielektrikum um den Faktor $1/(1 + V)$ dünner. Dies führt wieder zu Gl. (13). Umgekehrt ist das Dielektrikum für die Spannung $-Vu_e$ nur um $V/(1 + V)$, also unwesentlich dünner. Damit ist auch begründet, weshalb in (11) C_{GD} der ersten Stufe ohne Änderung übernommen werden konnte.

Jetzt können alle Kapazitäten zusammengefaßt werden,

$$C_e = C_1 + (1 + V)\, C_{GD} , \tag{14}$$

so entsteht mit

$$R_e = \frac{R_i^{+} \cdot R_G}{R_i^{+} + R_G} \tag{15}$$

die Ersatzschaltung von Bild 5.5.5f), und für die obere Grenzfrequenz gilt

$$\boxed{f_o = \frac{1}{2\pi C_e R_e}} \approx \frac{1}{2\pi R_a(VC_{GA} + C_D + C_G + 10\,\text{pF})} . \tag{16}$$

Für die Näherung ist $R_a < R_i \,||\, R_G$ vorausgesetzt und 10 pF als Schaltkapazität angesetzt.

5.5.4. Güte

Werden sehr hohe Frequenzen angestrebt, so müssen alle störenden Kapazitäten so klein wie möglich gehalten werden, und der Arbeitswiderstand ist entsprechend zu reduzieren. Wird weiter die untere Frequenz sehr niedrig gehalten, so kann f_o als *Bandbreite* B interpretiert werden. Infolge des kleinen Arbeitswiderstandes gilt weiter die Gl. (7), und es folgt

$$\boxed{BV \approx \frac{1}{2\pi} \cdot \frac{S}{C}} . \tag{17}$$

Das Produkt aus Bandbreite und Verstärkung ist also im wesentlichen durch die Steilheit und die Kapazitäten eines FET gegeben. Die *Zeitkonstante* S/C wird daher auch als *Güte* eines FET bezeichnet. Es ist bedeutsam, daß (17) für beliebige Verstärker

gilt, also auch für Verstärker mit Schwingkreisen. Da in diesem Fall alle Kapazitäten in die Schwingkreiskapazität einbezogen werden können, kann in einer ersten Näherung aus der Bandbreite und der Schwingkreiskapazität über die Steilheit die erreichbare Verstärkung berechnet werden (vgl. 5.5.6.).

5.5.5. Arbeitskennlinie bei komplexem Arbeitswiderstand

Bei vielen Schaltungen besitzt der Arbeitswiderstand Blindkomponenten (Kapazität und/oder Induktivität). Hierdurch ergeben sich *Phasenverschiebungen* zwischen den Wechselgrößen der Gatespannung, der Ausgangsspannung und des Drainstromes. Deshalb können nicht mehr die Arbeitskennlinien von Bild 5.5.3 gelten. Die neuen komplexen Kennlinien sind sowohl von der Aussteuerung als auch vom Phasenwinkel abhängig. Für eine bestimmte Aussteuerung und einen bestimmten Phasenwinkel zeigt Bild 5.5.6 die Konstruktion der *komplex-dynamischen Kennlinie.* Es enthält ausgezogen die statischen und gestrichelt die reellen dynamischen Kennlinien eines FET (sie können im Prinzip auch wesentlich anders liegen). Die Betrachtung der nicht phasenreinen Verstärkung geht am besten vom Drainstrom als Bezugsphase aus. Im Bild ist eine Verschiebung der Gatewechselspannung um —30° und der Drainwechselspannung um —60° angenommen. Die einzelnen Zeitpunkte innerhalb einer Drainstromperiode sind durch die Punkte 1 bis 12 gekennzeichnet. Die zugehörigen phasenverschobenen Zeitpunkte der Gatespannung bzw. Drainspannung zeigt ebenfalls das Bild. Durch Projektion der zusammengehörenden Punkte in die Kennlinienfelder ergeben sich die beiden *Ellipsen als komplexe Arbeitskennlinien.* Je nach der Phasenlage ändert sich ihr *Achsenverhältnis.* Grenzfälle sind Kreis und Gerade. Außerdem bestimmt die *Phase, in welcher Richtung die Ellipsen durchlaufen* werden. Wird die

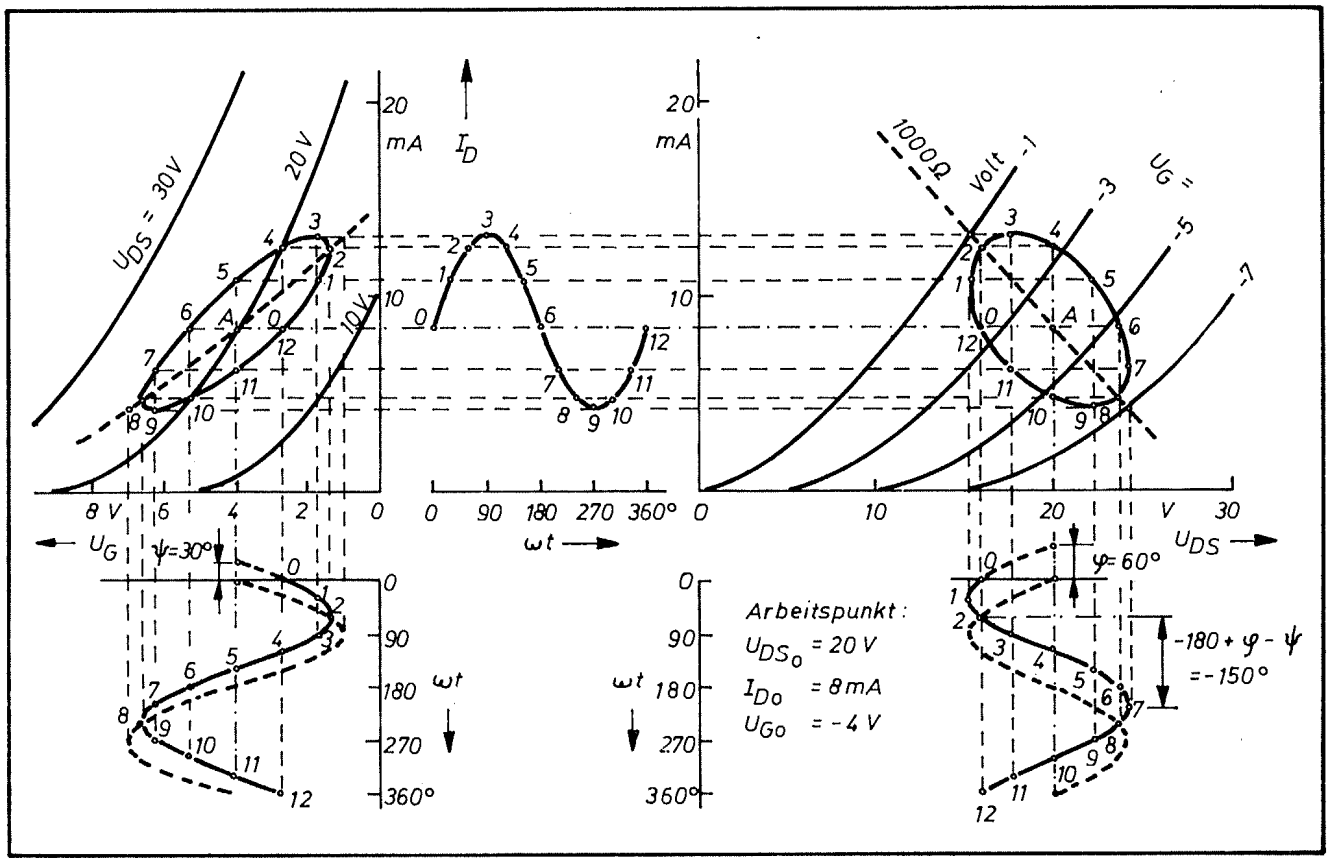


Bild 5.5.6. Konstruktion der Arbeitsellipse statt der Arbeitskennlinie für einen Verstärker mit komplexem Arbeitswiderstand

Amplitude der Schwingung verändert, so ändert sich nur die Größe der Ellipse, nicht jedoch ihr Achsenverhältnis. Sie schrumpft zusammen oder bläht sich auf. Wenn die reellen Arbeitskennlinien keine Geraden mehr sind, wird die Ellipse ebenfalls verformt. Dies ist im Ausgangskennlinienfeld des Bildes gerade zu erkennen.

5.5.6. Resonanz- und Breitbandverstärker

Gemäß Abschnitt 2.2. ist bei einem *Parallelschwingkreis* der Widerstand in Resonanz gleich dem ohmschen Widerstand R. Die Bandbreite berechnet sich aus der Güte

$$G = R\sqrt{C/L} \tag{18}$$

gemäß

$$B = f_m/G\,, \tag{19}$$

und wegen

$$f_m = \frac{1}{2\pi}\cdot\frac{1}{\sqrt{LC}} \tag{20}$$

folgt die Bandbreite zu

$$B = \frac{1}{2\pi}\,\frac{1}{RC} \tag{21}$$

und führt damit wieder zu (17), wenn nur die Schalt- und FET-Kapazitäten als Schwingkreiskapazität verwendet werden. Dies bringt aber praktische Probleme, da diese Kapazitäten und damit auch die Resonanzfrequenz vom Arbeitspunkt des FET abhängen (vgl. z.B. Bild 5.5.1g)). Dieser Einfluß läßt sich mindern, wenn eine zusätzliche Schwingkreiskapazität angewendet wird. Dann sinkt aber die Verstärkung. Aus dieser Betrachtung folgt ein zweites konstantes Produkt, das noch genauer im Abschnitt Rückkopplung (7.1.) untersucht wird, nämlich das Produkt aus Stabilität und Verstärkung.

Werden *Resonanzstufen hintereinandergeschaltet*, so engt sich die Bandbreite weiter ein. Der Abfall je Stufe darf an den Grenzen nur noch $1/\sqrt[2n]{2}$ betragen. Um so eine größere Bandbreite zu erreichen, ist es sinnvoll, mit verstimmten Schwingkreisen zu arbeiten. Dann sind in wenig abgewandelter Form die Ansätze aus Abschnitt 2.5., insbesondere bezüglich der GAUSS- und TSCHEBYSCHEFF-Filter zu finden. Ohne Welligkeit können in diesem Fall die Resonanzfrequenzen und Dämpfungen der Einzelkreise gemäß Bild 5.5.7b) graphisch bestimmt werden. Darin bedeuten f_o und f_u die beiden

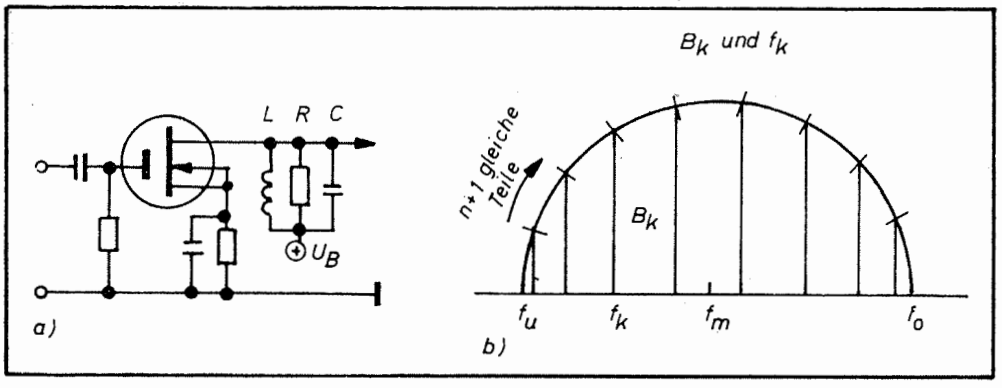


Bild 5.5.7. Verstärker mit Schwingkreis

a) Einstufiger Verstärker

b) Breitbandverstärker mit gestaffelten Schwingkreisen

Grenzfrequenzen und f_m das arithmetische Mittel. f_n sind die Resonanzfrequenzen und B_n die Bandbreiten der n Einzelkreise. Für die Gesamtverstärkung des n-stufigen Verstärkers gilt dann

$$V = \left(\frac{S}{C}\,\frac{1}{2\pi B}\right)^n. \tag{22}$$

Umgekehrt gilt für die Anzahl der notwendigen Stufen

$$n \geqq \frac{\log V}{\log\left(\frac{S}{C}\cdot\frac{1}{2\pi B}\right)}. \tag{23}$$

Dieses Prinzip der *Breitbandverstärker* wird meist bei den Videoverstärkern der Fernsehtechnik verwendet.

5.5.7. Kettenverstärker

Gemäß (17) gibt es eine *maximale Bandbreite*. Wird sie *überschritten*, so sinkt die *Verstärkung unter Eins*. Dann ist weder durch Hintereinanderschalten von Stufen noch durch Parallelschalten von FET eine Verstärkung möglich. Im letzten Fall addieren sich Steilheiten und Kapazitäten gleichermaßen. Mit dem Kettenverstärker gemäß Bild 5.5.8 besteht jedoch die Möglichkeit eines Ausweges. Das Entscheidende

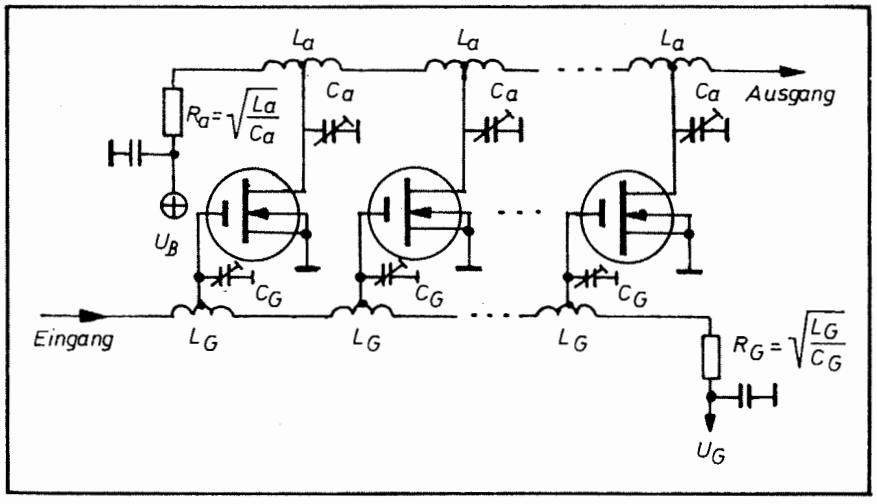


Bild 5.5.8. Prinzip des Kettenverstärkers

sind die *Laufzeitketten*, die zu den Gates führen bzw. von den Drains wegführen. Hierdurch werden einmal die *Parallelschaltung der Eingangs- und Ausgangskapazitäten des FET verhindert*. Sind beide Laufzeitketten abgeglichen, so werden andererseits die einzelnen Drainströme und damit die Steilheiten addiert, ohne daß sich die verschiedenen Kapazitäten addieren können. Für einen Kettenverstärker aus n Stufen folgt also

$$V_k = n\,\frac{S}{C}\cdot\frac{1}{2\pi B}. \tag{24}$$

Es besteht nun die *Optimierung* darin, *wieviel Stufen* in einem komplexen Verstärker *als Kette* und *wieviele* solcher *Ketten in Reihe* geschaltet werden müssen. Das Optimum wird dann erreicht, wenn die *Verstärkung der Kettenstufen bei* e, also etwa 2,71 liegt.

5.5.8. Technologien spezieller FET

Die Grundprinzipien der MOS-Herstellung sind bereits im Zusammenhang mit dem Bild 4.4.3 und in Ergänzungen mit den Bildern 4.4.4 sowie z.T. 3.3.4 und 3.3.7a) behandelt worden. Hier kommt es jetzt darauf an, wichtige Weiterentwicklungen und Sondertechnologien darzustellen, wobei keine Vollständigkeit angestrebt wird (weitere Details z.B. in [P 12]). Zwei Fakten bilden den Schwerpunkt: die *Verkürzung des Kanals*, um höhere Frequenzen (kürzere Schaltzeiten) zu erreichen, und der *Übergang zu Leistungs-MOSFET*.

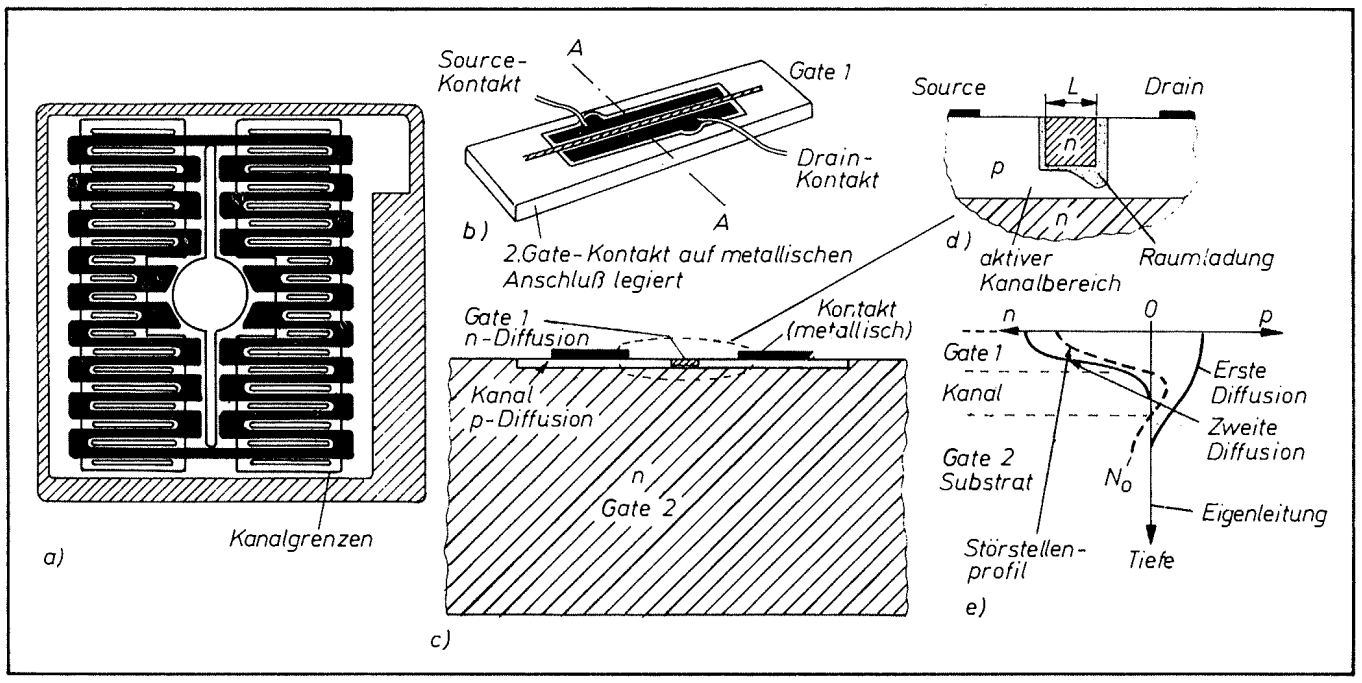


Bild 5.5.9. Aufbau eines Sperrschicht-FET

a) Topologie mit 2 × 18 Fingern
b) Struktur eines Fingers (vollständiger FET)
c) Maßstabgerechter Schnitt durch eine FET-Struktur längs Linie *AA* in b)
d) In vertikaler Richtung stark vergrößerter Ausschnitt aus c)
e) Diffusion und Dotierungsverlauf im Gate-1-Bereich

Zunächst sei mit Bild 5.5.9 die prinzipielle Technologie eines Sperrschicht-FET dargestellt. In ein *n*-Substrat wird zunächst mit der ersten Maske der Kanal großflächig durch Diffusion *p*-dotiert (Teilbild a), c) und e)). Mit einer zweiten Maske erfolgt die *n*-Diffusion für das Gate 1 im Kanalbereich (Teilbild a), c), d), schraffiert). Die Trennung von Kanal und Gate ist besonders deutlich in Bild e) durch das endgültige Störstellenprofil zu erkennen. Schließlich werden die metallischen Kontakte aufgebracht. Die Ausbildung der Raumladezone zeigt Teilbild d). Die *Vielzahl der einzelnen Finger* im Teilbild a) ist notwendig, um bei *kurzem Kanal* eine *größere Leistung* zu realisieren.

In Bild 4.4.3 ist der technologische Ablauf für den Metall- (Al-) Gate-*p*-Kanal-MOSFET beschrieben. Er ist für Einzeltransistoren und niedrige Integrationsgrade infolge seiner Einfachheit gut geeignet. In dieser Technologie sind z.B. die Schaltkreise U 101 bis U 808 des Kombinats Mikroelektronik Erfurt realisiert. Durch Substratorientierung [111] statt [100] kann die Schwellenspannung reduziert werden. Durch *Ionenimplantation* gemäß Bild 4.4.4 konnte die Kanallänge verkürzt und die MILLER-Kapazität reduziert werden. Außerdem gelang es so, eine Niedervolttechnik (geringe Betriebsspannungen) zu realisieren. Ein Übergang zur *n-Kanal-Technik* auf [100]-*p*-

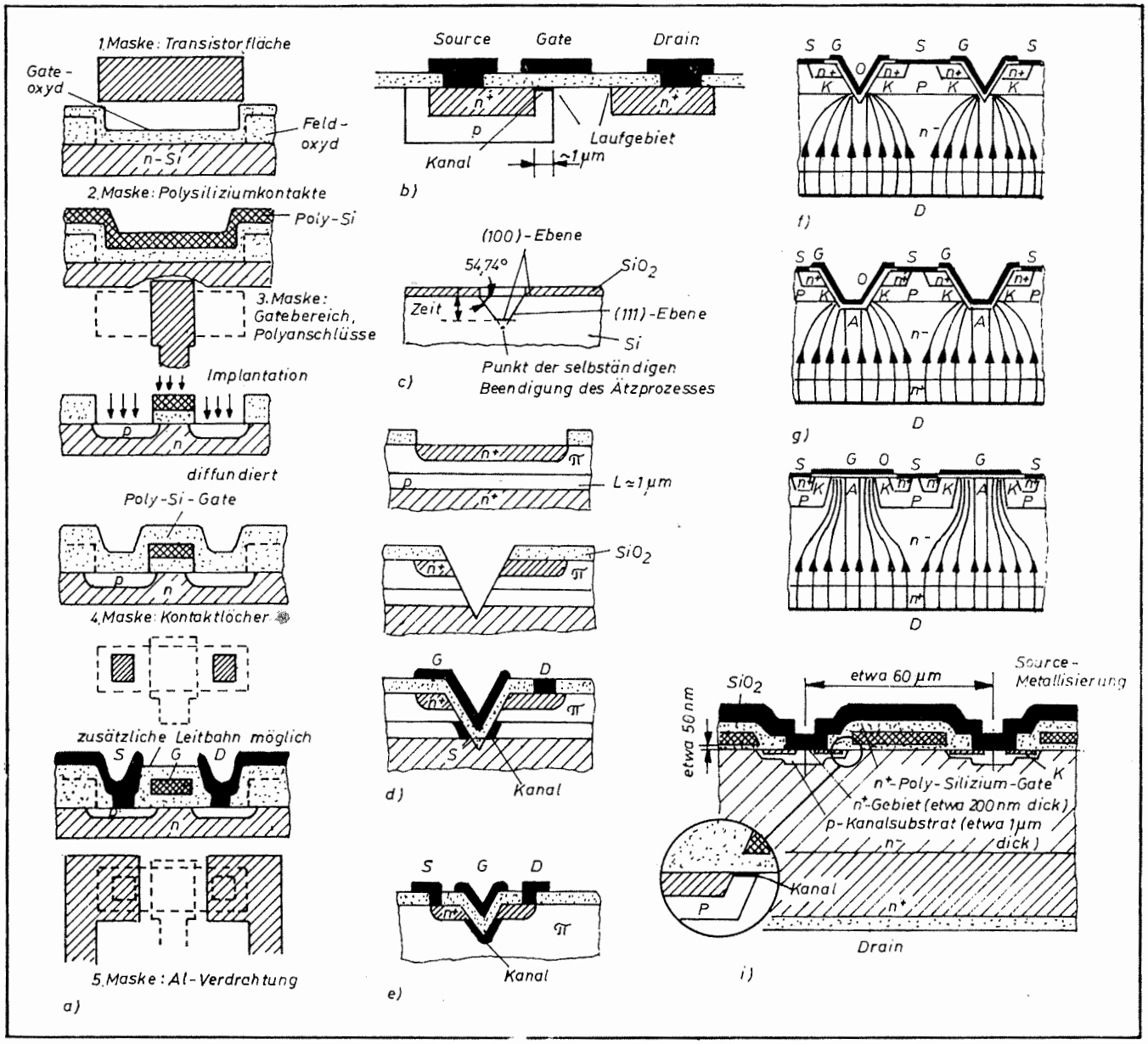


Bild 5.5.10. Ausgewählte MOS-Technologien. Die Standardtechnologie zeigt Bild 4.4.3 und 4.4.4.

a) Si-Gate-Technologieablauf
b) Aufbau eines DMOS-FET
c) Prinzip des vertikalen Ätzens
d) Technologieablauf für einen VMOS-FET
e) Aufbau des lateralen VMOS-FET
f) Stromlinien beim VMOS-Leistungs-FET
g) Stromlinien beim UMOS-Leistungs-FET
h) Stromlinien beim DMOS-Leistungs-FET
i) Detaillierte Struktur des DMOS-Leistungs-FET

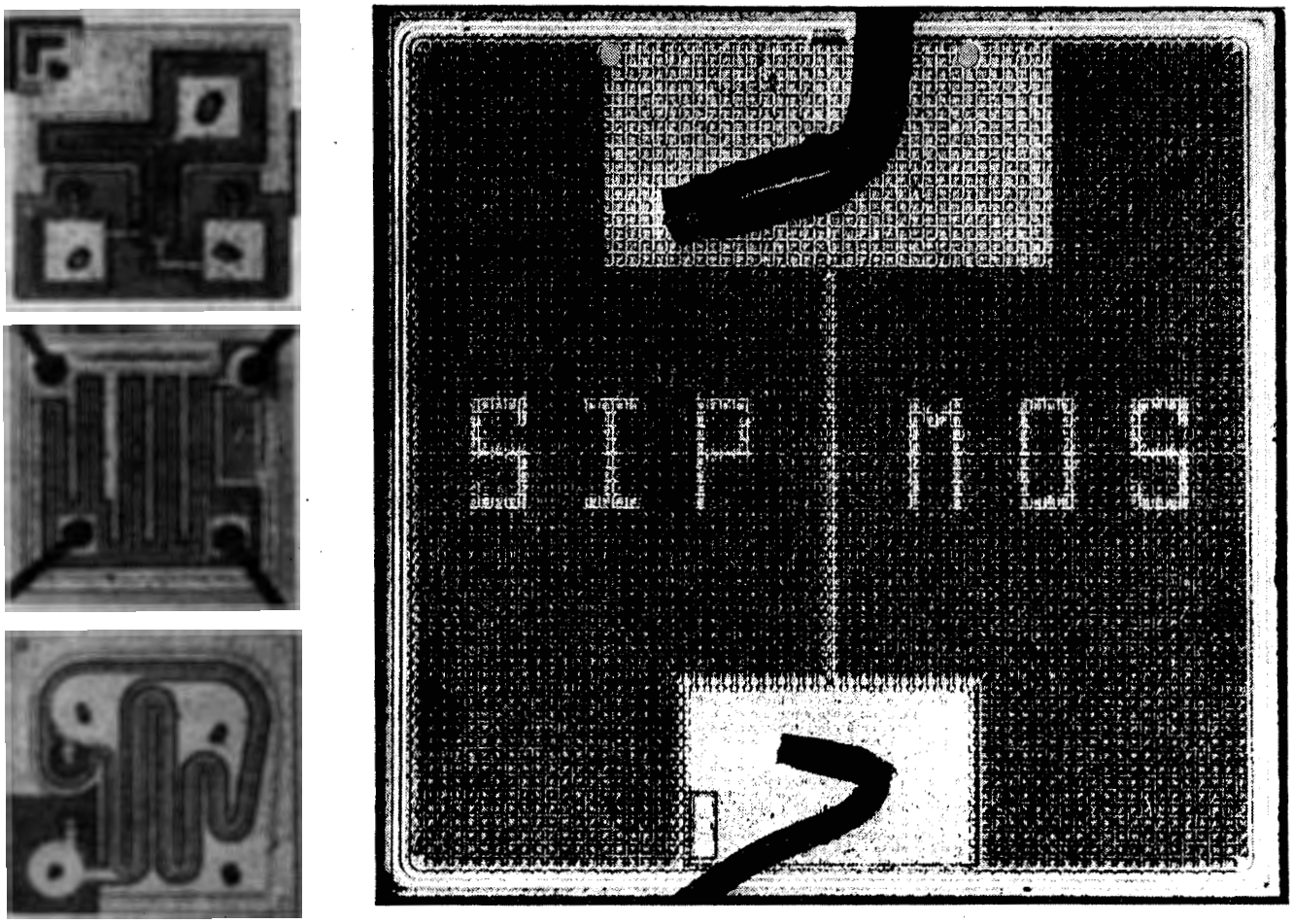
dotiertem Substrat ließ die höhere Beweglichkeit der Elektronen gegenüber den Löchern ausnutzen und die Geschwindigkeit erhöhen. Mit dem *polykristallinen* Si-*Gate* konnte dann zweierlei erreicht werden: Senkung der Schwellenspannung und eine *selbstjustierende Technik*. Den entsprechenden Technologieablauf zeigt Bild 5.5.10a).

Mit den bisher erwähnten Methoden werden meist Kanallängen im Bereich von 10 bis 3 µm realisiert. Diesen Langkanaltechniken stehen die Kurzkanaltechniken mit Längen unter 2 µm gegenüber. Sie lassen sich im Prinzip durch hohe Präzision analog herstellen. Effektiver ist es jedoch, Sondertechnologien zu verwenden. Sie sind vor allem durch die Begriffe **DMOS** und **VMOS** gekennzeichnet und ermöglichen

Kanallängen um 1 bis 2 μm. (Es sei erwähnt, daß bei den ersten FET um 1963 die Kanallänge bei 50 μm lag, 1971 erstmalig 2 μm erreicht wurden und heute 0,5 bis 0,3 μm angestrebt werden.)

DMOS bedeutet *doppelte Diffusion* durch ein Maskenfenster. Das Grundmaterial ist schwach p-leitend, oft mit π bezeichnet. Durch das Sourcefenster wird zunächst p und anschließend n^+ gemeinsam mit dem Drainfenster dotiert. So entsteht die Struktur von Bild 5.5.10b). Die Kanallänge ist bei diesem Prozeßablauf gut mittels der beiden Diffusionen zu steuern. Das π-Gebiet dient als Laufgebiet. Es ist in der Tiefe nicht mehr begrenzt. Dadurch tritt sein Widerstand nicht entscheidend in Erscheinung. Ähnlich wie bei *pin*-Dioden und den entsprechenden Transistoren ist es auch hier möglich, in diesem Gebiet hohe Spannungen abzufangen. So ist neben der Kurzkanaltechnik zugleich ein FET für hohe Betriebsspannungen (bis ca. 1000 V) möglich.

Beim **VMOSFET** wird mittels *anisotropen Ätzens* eine *V-förmige Grube* in der Kristalloberfläche erzeugt. Diese Technologie wurde ursprünglich für die Isolation von Einzelbauelementen auf einem Chip entwickelt, hat aber bald danach universelle Bedeutung erlangt (s. Bild 5.5.10c)). Den technologischen Ablauf für die Herstellung eines VMOS zeigt Bild 5.5.10d). Auf ein n^+-Substrat wachsen nacheinander großflächig eine p- und eine π-Schicht auf. In die π-Schicht wird dann lokal eine n^+-Schicht als Drain diffundiert. Danach erfolgt das anisotrope Ätzen. So entsteht der vertiakle VMOSFET. Gemäß Bild 5.5.10e) gibt es auch eine laterale Variante.



a) - c) d)

Bild 5.5.11. Beispiele von MOSFET-Strukturen

a) bis c) Vorstufen-FET
d) SIPMOS-Leistungstransistor (Kantenlänge 4 mm)

Für die Entwicklung von *Leistungs*-**MOSFET** ist es notwendig, die *entstehende Wärme gut abführen zu können*. Deshalb mußte erreicht werden, daß Ladungsträger großflächig zu der Stelle geführt werden, wo der Chip gut wärmeleitend aufgelötet ist. Der *Drainanschluß* mußte dazu bei nicht mehr wesentlichen Änderungen auf die *Unterseite des Chips* verlegt werden. So entstanden in nur geringem Zeitabstand die

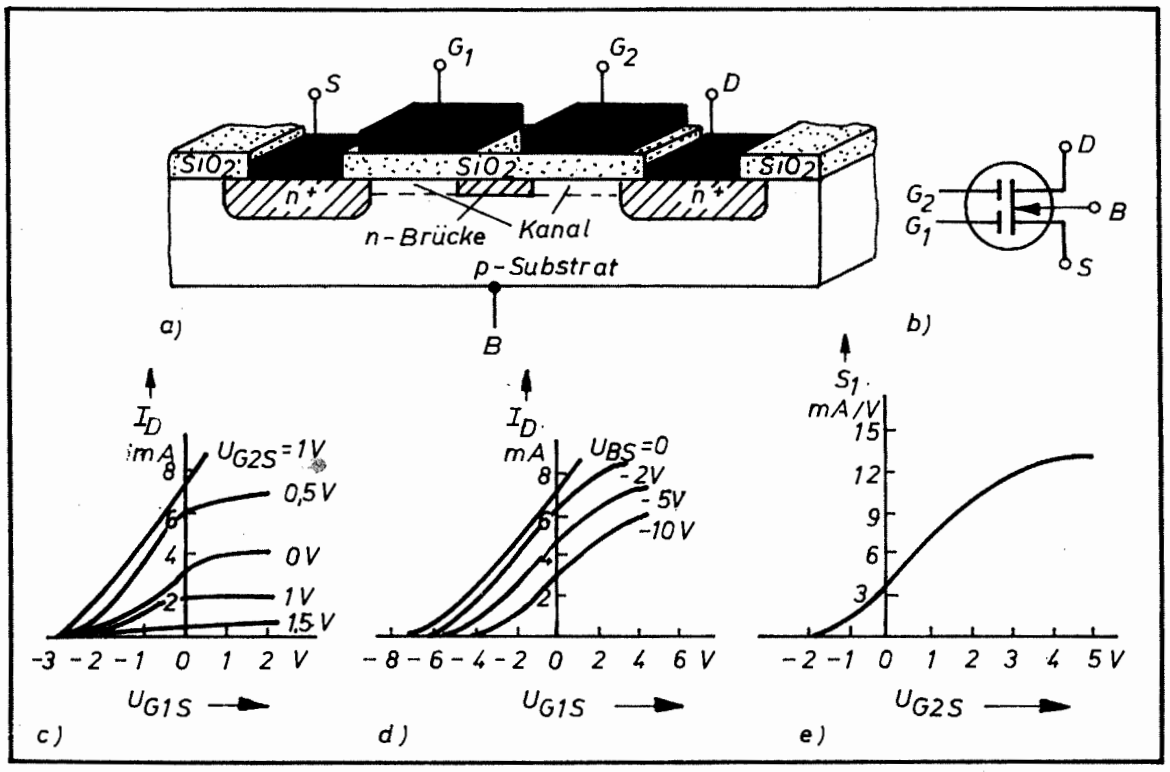


Bild 5.5.12. Doppel-Gate-MOSFET

a) Prinzipieller Aufbau
b) Schaltbild
c) Abhängigkeit des Drainstromes von der Spannung an Gate 1 mit der Spannung am Gate 2 als Parameter
d) Dito, jedoch mit zusätzlicher Steuerung vom Substrat
e) Abhängigkeit der Steilheit des Gate 1 als Funktion der Spannung an Gate 2

verschiedenen Firmenbezeichnungen für die entsprechende VMOS- und insbesondere für die DMOS-Technik (Bild 5.5.10f) bis i)). So deutet die Bezeichnung HEXFET an, daß die Strukturen des DMOS in regelmäßigen Sechsecken realisiert sind, SIPMOS enthält den Namen Siemens und RMOS könnte auf Transversal (Motorola) hinweisen. UMOS ist dagegen ein VMOS, bei dem die vertikale Ätzung nicht bis zu Ende geführt wurde und die Bezeichnung auf die Form des Grabens verweist. Dadurch entsteht eine etwas geringere Stromdichte beim Austritt der Ladungsträger aus dem Kanal. Infolge des engen und zu den sonstigen Abmessungen vergleichsweise kurzen Kanals tritt eine Geschwindigkeitssättigung der Ladungsträger auf. Dies bewirkt eine zusätzliche Linearisierung der Steuerkennlinien (s. Bild 5.5.10g)).

Es sei betont, daß die Leistungs-MOSFET von den meisten Firmen auch als weitgehend komplementäre Paare, also in *n*- und *p*-Kanal-Technik angeboten werden.

Bild 5.5.11 zeigt beispielhaft einige Strukturen von MOSFET.

Zum Schluß sei noch auf den *Doppel-Gate*-**FET** (auch FET-Tetrode genannt) hingewiesen. Er bringt zwei Vorteile besonders für hohe Frequenzen: Radikale Senkung der MILLER-Kapazität und zusätzliche Steuerung der Verstärkung. Aufbau und Kennlinien eines solchen FET zeigt Bild 5.5.12.

5.6. Bipolartransistor

Im Abschnitt 3.4. wurden die grundsätzlichen Eigenschaften des Bipolartransistors abgeleitet. Aus der Logik der Physik ergab sich dabei die *Basisschaltung*. Sie hat heute sehr untergeordnete Bedeutung. Am meisten angewendet wird die *Emitterschaltung*. Ihre theoretische Begründung erfolgt im Abschnitt 7.3. Hier werden ihre Eigenschaften im Vergleich zur Emitter- und zur *Kollektorschaltung* ohne Beweis dargestellt. Die Beziehungen dieser drei Schaltungen rühren daher, daß es möglich ist, jede der drei Elektroden als Bezugspunkt für Ein-Ausgang zu verwenden. Diese Elektrode bestimmt den Namen der Schaltung. Prinzipiell kann dann noch zwischen zwei Schaltungen unterschieden werden, je nachdem, welche der verbleibenden Elektroden für den Eingang verwendet wird. Wegen der *technologischen Unsymmetrie* von Kollektor und Emitter wird jedoch immer jene Schaltung verwendet, welche die besten Verstärkungseigenschaften ergibt. Die andere stellt die extrem seltene Anwendung im Inversbetrieb des Transistors dar. Auf sie wird hier generell verzichtet.

Im folgenden wird der *npn*-Transistor betrachtet. Die Ausführungen gelten auch für den *pnp*-Transistor. Es sind im Prinzip nur die Spannungen umzupolen.

Der FET ließ eine praktisch leistungslose Steuerung zu. Dadurch vereinfachten sich seine Betrachtungen gegenüber der Matrizendarstellung erheblich. Der Bipolartransistor verlangt dagegen immer eine Eingangsleistung, also Strom und Spannung. Er ist daher nur durch Matrizen effektiv zu beschreiben.

5.6.1. Die drei Schaltungen

Bild 5.6.1 zeigt die drei Grundverstärkerschaltungen des bipolaren Transistors. Für alle drei gilt das Ersatzschaltbild mit den h-Parametern gemäß Bild 5.6.1d). Die Werte der h-Parameter sind in den drei Schaltungen unterschiedlich. Deshalb ist es üblich, die jeweilige Schaltungsart als Index E, B, C anzuhängen. Die Werte können formal von jeder auf die anderen beiden umgerechnet werden. Dies ist in Tab. 5.6.1 zusammengefaßt. Beispielwerte für den Transistor SC 206 enthält Tab. 5.6.2. Hieraus

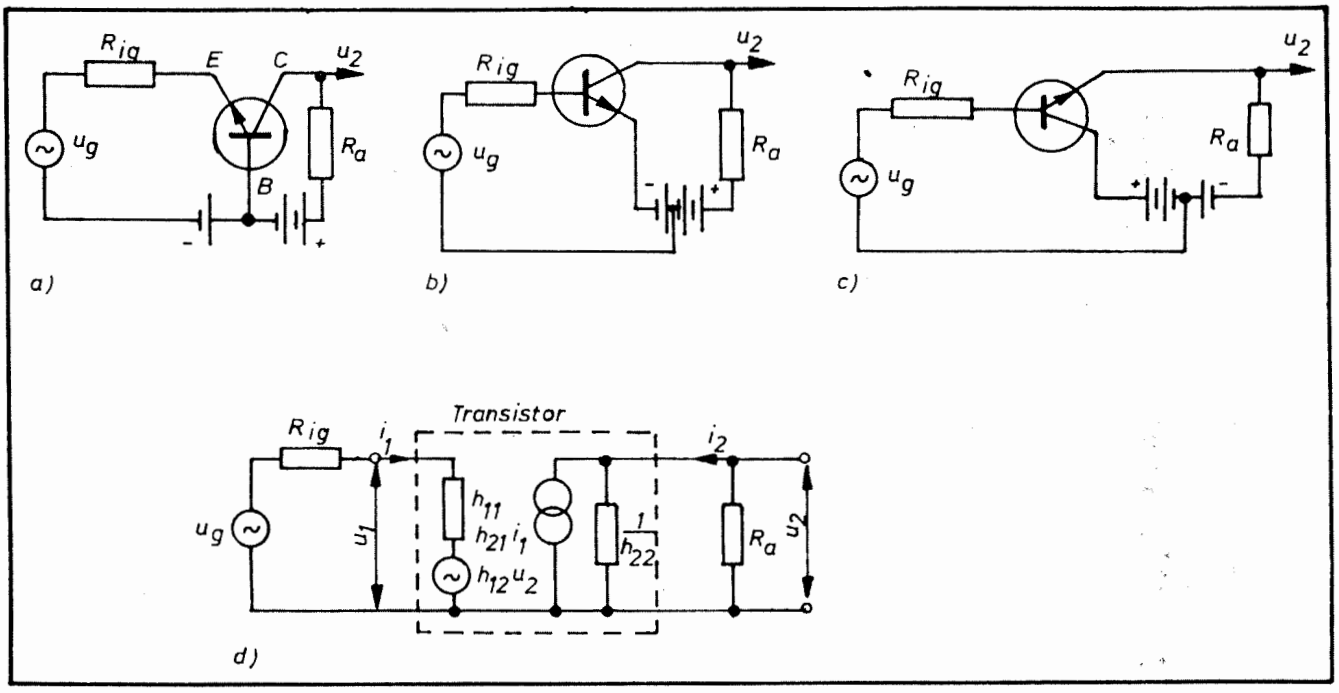


Bild 5.6.1. Die drei Schaltungen des Transistors

a) Basis-, b) Emitter-, c) Kollektorschaltung d) Allgemein gültiges h-Parameter-Ersatzschaltbild

Tabelle 5.6.1. Es bezeichnen E Emitter-, B die Basis- und C die Kollektorschaltung. Ferner wurden zur Abkürzung verwendet:
$\Sigma h = 1 + \Delta h + h_{21} - h_{12} \approx 1 + h_{21}$
$\Delta h = h_{11}h_{22} - h_{21}h_{12}$

	h_B	h_E	h_C
h_B	h_{11B} h_{12B} h_{21B} h_{22B}	$\frac{h_{11E}}{\Sigma h_E}$ $\frac{\Delta h_E - h_{12E}}{\Sigma h_E}$ $-\frac{\Delta h_E + h_{21E}}{\Sigma h_E}$ $\frac{h_{22E}}{\Sigma h_E}$	$\frac{h_{11C}}{\Delta h_C}$ $\frac{h_{21C} + \Delta h_C}{\Delta h_C}$ $\frac{h_{12C} - \Delta h_C}{\Delta h_C}$ $\frac{h_{22C}}{\Delta h_C}$
h_E	$\frac{h_{11B}}{\Sigma h_B}$ $\frac{\Delta h_B - h_{12B}}{\Sigma h_B}$ $-\frac{\Delta h_B + h_{21B}}{\Sigma h_B}$ $\frac{h_{22B}}{\Sigma h_B}$	h_{11E} h_{12E} h_{21E} h_{22E}	h_{11C} $1 - h_{12C}$ $-(1 + h_{21C})$ h_{22C}
h_C	$\frac{h_{11B}}{\Sigma h_B}$ $\frac{1 + h_{21B}}{\Sigma h_B}$ $\frac{h_{12B} - 1}{\Sigma h_B}$ $\frac{h_{22B}}{\Sigma h_B}$	h_{11E} $1 - h_{12E}$ $-(1 + h_{21E})$ h_{22E}	h_{11C} h_{12C} h_{21C} h_{22C}

ist ersichtlich, daß bei der Kollektor- und Emitterschaltung der Stromverstärkungsfaktor h_{21} größer als Eins ist. Ferner liegt bei beiden Schaltungen der Ein- und Ausgangswiderstand h_{11} bzw. $1/h_{22}$ in gleicher Größenordnung. Die Spannungsrückwirkung h_{12} ist jedoch bei der Kollektorschaltung unakzeptabel groß. So kommt es, daß die *Emitterschaltung folgende Vorteile auf sich vereint*:

- Sie bietet zugleich große Spannungs- und Stromverstärkung, also auch bezüglich der Leistung.

Tabelle 5.6.2. Richtwerte für den SC 206 bei einem Arbeitspunkt mit $U_{BE} = 0{,}6$ V, $U_{CE} = 6$ V und $I_C = 2$ mA

Parameter	Basis-	Emitter-	Kollektor-	Einheit
	Schaltung			
h_{11}	22,8	2300	2300	Ω
h_{12}	$7{,}2 \cdot 10^{-4}$	$3{,}8 \cdot 10^{-4}$	≈1	—
h_{21}	−0,99	100	−100	—
h_{22}	0,475	48	48	μS
$1/h_{22}$	≈2000	≈20	*P*20	kΩ

- Der Unterschied zwischen Eingangs- und Ausgangswiderstand ist meist gering, so daß relativ leicht mehrere Stufen unkompliziert hintereinandergeschaltet werden können.

Sie besitzt jedoch zwei Nachteile, auf die noch später einzugehen ist:

- Der *Reststrom* kann bei hohen Temperaturen stören.
- Der *Frequenzgang* ist schlechter als in den beiden anderen Schaltungen.

Wegen der großen Bedeutung der Stromversärkung sei eine wichtige Näherung aus Tab. 5.6.1 zwischen ihrem Wert in der Basisschaltung α und in der Emitterschaltung β angegeben.

$$\boxed{\beta = \frac{\alpha}{\alpha - 1}}\,. \tag{1}$$

Da α immer in der Nähe von Eins liegt, wird β also groß. Reale Werte liegen im Bereich von 50 bis 2000.

5.6.2. Betriebswerte

Die Berechnung der Betriebswerte der drei Schaltungen erfolgt mittels des Ersatzschaltbildes 5.6.1d). Es werden also primär die h-Parameter verwendet:

$$\left.\begin{aligned} u_1 &= h_{11} i_1 + h_{12} u_2\,, \\ i_2 &= h_{21} i_1 + h_{22} u_2\,. \end{aligned}\right\} \tag{2}$$

Umrechnungen zu anderen Parametern, insbesondere zu den y- und z-Parametern können gemäß Tab. 1.1.5 erfolgen. Für den Generator- und Arbeitswiderstand gilt ($u_g = 0$)

$$R_a = -\frac{u_2}{i_2}, \qquad R_{ig} = -\frac{u_1}{i_1}. \tag{3}$$

Die sich dann ergebenden Werte für den Eingangswiderstand R_e und Innenwiderstand R_i der Schaltung sowie für die *Spannungsverstärkung*

$$v_u = -\frac{u_2}{u_1} \tag{4}$$

und für die *Stromverstärkung*

$$v_i = \frac{i_e}{i_1} \tag{5}$$

enthält Tab. 5.6.3. Zur Vereinfachung bei den *z-Parametern* sind die Ein- und Ausgangswerte für *Kurzschluß* bzw. *Leerlauf* eingeführt. Es gilt:

$$\left.\begin{aligned} R_{eL} &= z_{11}\,, \\ R_{iL} &= z_{22}\,, \\ R_{eK} &= \frac{\Delta z}{z_{22}}\,, \\ R_{iK} &= \frac{\Delta z}{z_{11}}\,. \\ \mu &= \frac{z_{21}}{z_{11}} = -\frac{y_{21}}{y_{22}} = -\frac{h_{21}}{\Delta h}\,, \\ \alpha^* &= -\frac{z_{21}}{z_{22}} = \frac{y_{21}}{y_{11}} = h_{21}\,. \end{aligned}\right\} \tag{6}$$

Die Abhängigkeit der einzelnen Größen vom *Arbeitswiderstand* R_a bzw. für den *Innenwiderstand* R_i vom *Quellenwiderstand* R_{ig} zeigt Bild 5.6.2. Die Eingangs- bzw.

Tabelle 5.6.3. Wichtige Verstärkungswerte eines Transistors gemäß Bild 5.6.1.

Eingangswiderstand	$R_e = \dfrac{h_{11} + \Delta h R_a}{1 + h_{22} R_a}$	$\dfrac{1 + y_{22} R_a}{y_{11} + \Delta y R_a}$	$R_{eK} \dfrac{1 + R_a/R_{iK}}{1 + R_a/R_{iL}}$
Innenwiderstand	$R_i = \dfrac{h_{11} + R_{ig}}{\Delta h + h_{22} R_{ig}}$	$\dfrac{1 + y_{11} R_{ig}}{y_{11} + \Delta y R_{ig}}$	$R_{iK} \dfrac{1 + R_{ig}/R_{eK}}{1 + R_{ig}/R_{eL}}$
Spannungsverstärkung	$v_u = \dfrac{h_{21} \cdot R_a}{h_{11} + \Delta h R_a}$	$\dfrac{y_{21} R_a}{1 + y_{22} R_a}$	$\dfrac{\mu}{1 + R_{iK}/R_a}$
Stromverstärkung	$v_i = \dfrac{h_{21}}{1 + h_{22} R_a}$	$\dfrac{y_{21}}{y_{11} + \Delta y R_a}$	$\dfrac{\alpha^*}{1 + R_a/R_{iL}}$

Ausgangswiderstände nehmen je nach Betrieb und Schaltung Werte zwischen ca. 20 Ω und 2 MΩ an.

Die Spannungsverstärkung ist bei Emitter- und Basisschaltung etwa gleich groß. Sie steigt zunächst mit R_a linear an und nimmt für $R_a \approx 100\,\text{k}\Omega$ einen Grenzwert von 2000 an. Die Kollektorschaltung hat unter normalen Betriebsbedingungen eine Spannungsverstärkung von Eins.

Die Stromverstärkung liegt bei der Basisschaltung, sofern nicht sehr große Arbeitswiderstände verwendet werden, etwa bei Eins. Bei der Kollektor- und Emitterschaltung werden hier Werte um $\beta = 100$ erreicht. Bei großen Arbeitswiderständen über 10 kΩ nimmt die Stromverstärkung umgekehrt proportional zum Widerstand ab.

Da der Transistor *Leistungen verstärkt*, interessiert das Produkt

$$\begin{aligned} v_P &= v_u v_i \\ &= \frac{h_{21}^2 R_a R_{ig}}{[(1 + h_{22} R_a)\, R_{ig} + h_{11} + \Delta h\, R_a]^2} \\ &= \frac{y_{21}^2 R_a R_{ig}}{[1 + y_{22} R_a + R_{ig}(y_{11} + \Delta y\, R_a)]^2}. \end{aligned} \tag{7}$$

Es erreicht seinen höchsten Wert bei der Emitterschaltung. Nur mit ihr sind gleichzeitig echte Strom- und Spannungsverstärkungen zu erreichen. Die Maxima der einzelnen Leistungsverstärkungskurven liegen bei verschiedenen Werten des Arbeitswiderstandes. Das Maximum selbst läßt sich gut aus (7) berechnen. Zu ihm gehören *je ein optimaler Wert für den Arbeits- und Quellenwiderstand.* Es gilt

$$R_{a,\,max} = \sqrt{\frac{h_{11}}{h_{22}\,\Delta h}} = \sqrt{\frac{y_{11}}{y_{22}\,\Delta y}} = \sqrt{R_{iL} \cdot R_{iK}} \tag{8}$$

bzw.

$$R_{ig,\,max} = \sqrt{\frac{h_{11}\,\Delta h}{h_{22}}} = \sqrt{\frac{y_{22}}{y_{11}\,\Delta y}} = \sqrt{R_{eL} R_{eK}}. \tag{9}$$

Die größtmögliche Leistungsverstärkung folgt damit zu

$$v_{P,\,max} = \left(\frac{h_{21}}{\sqrt{\Delta h} + \sqrt{h_{11} h_{22}}}\right)^2 = \left(\frac{y_{21}}{\sqrt{\Delta y} + \sqrt{y_{11} y_{22}}}\right)^2. \tag{10}$$

Von besonderer Bedeutung bei Verstärkern ist die *Reihenschaltung von Verstärkerstufen.* Hier muß dann gelten

$$R_a = R_{ig} = R_{a,i}. \tag{11}$$

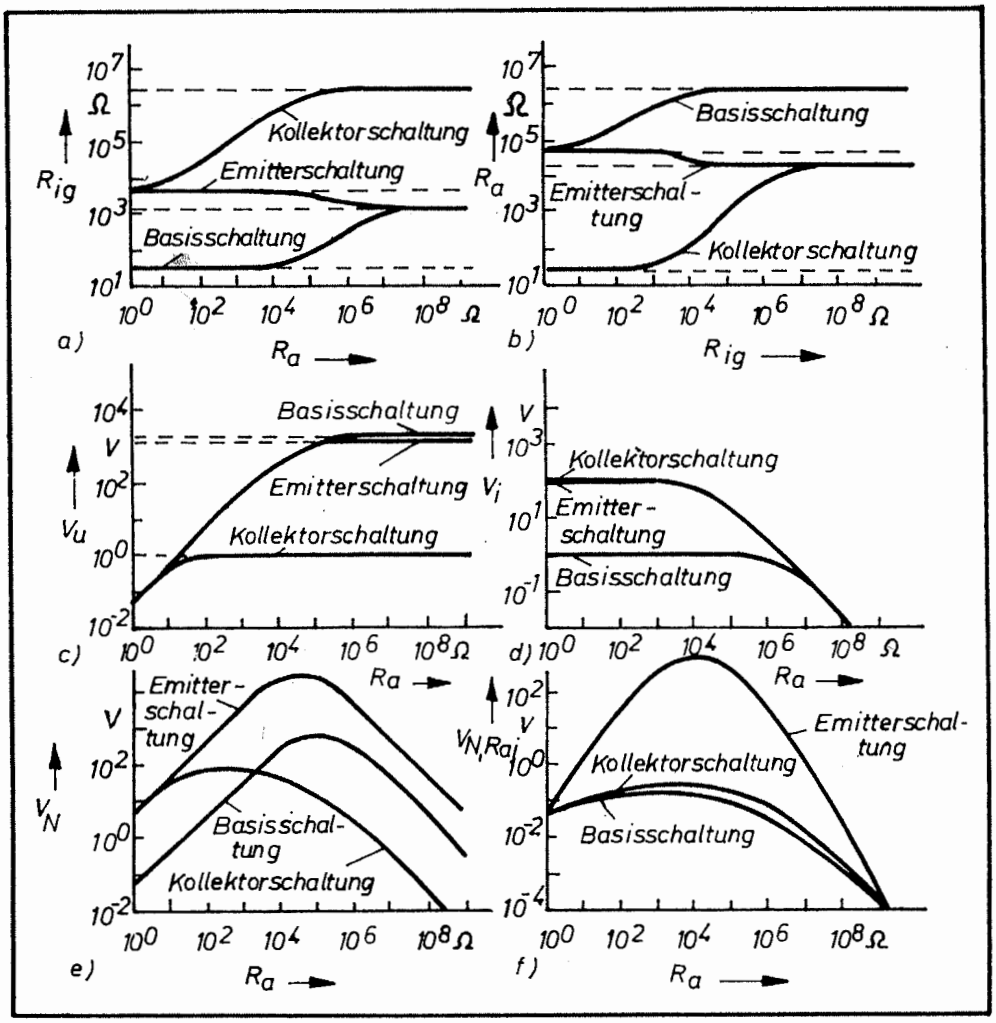


Bild 5.6.2. Betriebsgrößen als Funktion des Arbeitswiderstandes R_a bzw. beim Innenwiderstand R_i als Funktion des Quellenwiderstandes R_{ig}. Die Größen sind für die drei Grundschaltungen am Beispiel des Transistors SC 206 dargestellt.

Mit diesem Widerstand geht (7) über in

$$v_{P,R_{a,i}} = \frac{h_{21}^2 R_{a,i}^2}{[h_{11} + (1 + \Delta h)\, R_{a,i} + h_{22} R_{a,i}^2]^2}. \tag{12}$$

Für den gemeinsamen Widerstand $R_{a,i}$ gibt es ebenfalls eine größtmögliche Leistungsverstärkung. Zu ihr gehört der Widerstand

$$R_{a,i,max} = \sqrt{\frac{h_{11}}{h_{22}}}. \tag{13}$$

Er ist unabhängig von der Schaltungsart und beträgt bei dem gewählten Beispiel etwa 7 kΩ. Die *maximale Leistungsverstärkung* ergibt sich zu

$$v_{P,R_{a,i},max} = \left(\frac{h_{21}}{1 + \Delta h + 2\sqrt{h_{11}h_{22}}}\right)^2. \tag{14}$$

Hierbei ist lediglich mit der Emitterschaltung eine wirkliche Verstärkung möglich. Unter den gegebenen Bedingungen liegt sie bei etwa 5000, also rund $^1/_6$ des Wertes für den Fall, daß R_a und R_{ig} einzeln angepaßt sind. Zahlenwerte für die einzelnen Größen enthält Tab. 5.6.4.

Tabelle 5.6.4. Einige Werte zur optimalen Verstärkung bei Betrieb eines Transistors SC 206 in Ergänzung zur Tabelle 5.6.3

	Basis-	Emitter-	Kollektor-
	Schaltung		
Δh	$7{,}2 \cdot 10^{-4}$	$7{,}2 \cdot 10^{-2}$	100
$\sqrt{\Delta h}$	$2{,}7 \cdot 10^{-2}$	0,28	10
$h_{11} \cdot h_{22}$	$1{,}1 \cdot 10^{-5}$	0,11	0,11
$\sqrt{h_{11}h_{22}}$	$3{,}2 \cdot 10^{-3}$	0,32	0,32
$V_{P,\max}$	1000	$2{,}7 \cdot 10^{4}$	100
$V_{P,R_{a,i,\max}}$	0,86	$5{,}1 \cdot 10^{3}$	≈ 1

5.6.3. Kennlinien und Arbeitspunkt

Die Kennlinienfelder des SC 206 in Basisschaltung zeigt Bild 3.4.6b), die entsprechenden Kennlinienfelder in Emitter- und Kollektorschaltung ergänzt Bild 5.6.3a) und b). Weiter sind in dem Bild die Änderungen der h-Parameter in Emitterschaltung unter dem Einfluß des Kollektorstromes (c), der Kollektorspannung (d) und der Temperatur (e) gezeigt. Es ist auffällig, daß die *Kennlinienfelder* aller drei Schaltungen *nahezu gleich aussehen.* Die *Hauptunterschiede* leiten sich aus dem im vorigen Abschnitt Gesagten ab:

- Der Steuerstrom (linke Abszisse und Parameter im ersten Quadranten) ist bei der Basisschaltung rund 100mal größer.
- Die untere Ordinate trägt bei der Kollektorschaltung 20mal größere Werte.
- Der Verlauf im zweiten Quadranten weicht bei der Kollektorschaltung völlig ab (große Spannungsrückwirkung).

5.6.3.1. Näherungsbetriebswerte

Insbesondere bei der Emitterschaltung entspricht die Eingangskennlinie (dritter Quadrant) recht gut der Diodenkennlinie (Gl. (3.2.5)). Infolge der linearen Stromverstärkung und des Einflusses der Temperatur T und der Kollektorspannung U_{CE} gilt mit einem Sättigungsstrom I_S und der *Temperaturspannung*

$$U_T = kT/e_0 \tag{15a}$$

die Beziehung

$$I_C = I_S(TU_{CE})\, e^{U_{BE}/U_T}\,. \tag{15b}$$

Für kleine Aussteuerung läßt sich hieraus in Analogie zum Feldeffekttransistor eine Steilheit definieren:

$$S = \partial I_C/\partial U_{BE} = I_S e^{U_{BE}/U_T}/U_T\,. \tag{16}$$

Folglich gilt

$$S \approx I_C/U_T\,. \tag{17}$$

Mit der Stromverstärkung β folgt hieraus auch ein differentieller Eingangswiderstand

$$r_{BE} = \partial U_{BE}/\partial I_B \approx \beta/S\,, \tag{18}$$

also

$$r_{BE} \approx \beta U_T/I_C\,. \tag{19}$$

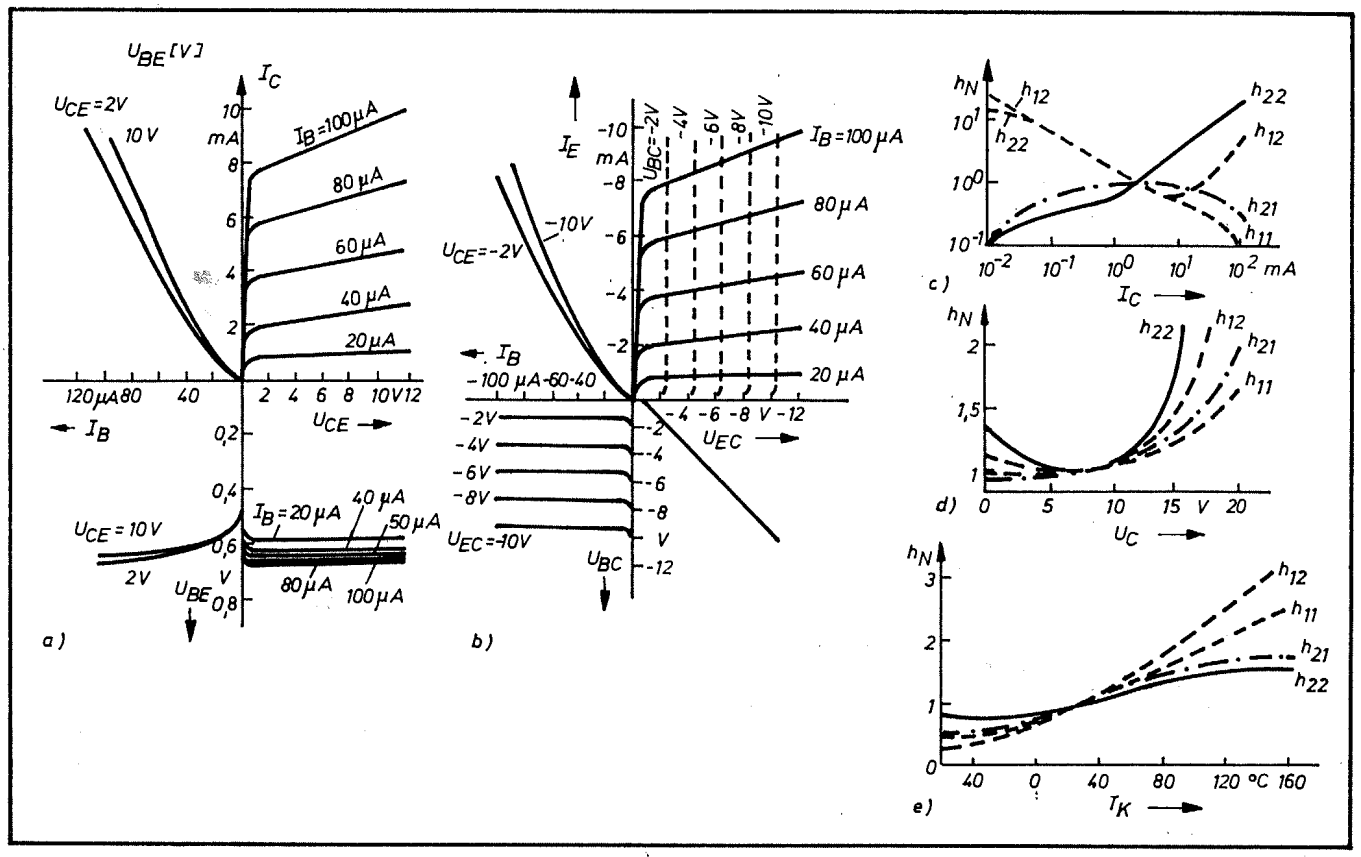


Bild 5.6.3. Kennlinien des Transistors SC 206

a) Kennlinienfeld der Emitter-, b) der Kollektorschaltung. Das der Basisschaltung zeigt Bild 3.4.6b)
c) Änderungen der h-Parameter bei der Emitterschaltung als Funktion des Kollektorstromes,
d) als Funktion der Kollektorspannung,
e) als Funktion der Temperatur

Die Werte sind normiert auf den Arbeitspunkt $T = 25$ °C, $U_{CE} = 6$ V und $I_C = 2$ mA

Schließlich kann noch ein differentieller Ausgangswiderstand eingeführt werden:

$$r_{CE} = \partial U_{CE}/\partial I_C \,. \tag{20}$$

Hierfür gibt es aufgrund der Kennlinien eine Näherung

$$r_{CE} \approx U_Y/I_C \,. \tag{21}$$

Darin ist U_Y die *Early-Spannung*. Sie läßt sich für den Sättigungsbereich aus einem Meßwert von r_{CE} bestimmen. Die anderen Werte werden dann recht gut genähert. Für *npn*-Transistoren liegt ihr Wert etwa bei 80 bis 200 V und für *pnp*-Transistoren bei 40 bis 150 V. *Mit diesen drei Kennwerten* S, r_{BE} und r_{CE} ist infolge der sehr geringen Rückwirkung der Kollektorwechselspannung *oft eine recht gute Beschreibung der Kleinsignaleigenschaften von bipolaren Transistoren möglich.*

Auf die h-Parameter bezogen gilt

$$r_{BE} \approx h_{11E} \,, \tag{22}$$

$$S \approx h_{21E}/h_{11E} \,, \tag{23}$$

$$r_{CE} \approx 1/h_{22E} \,. \tag{24}$$

Das Verhältnis U_T/U_Y kann weiter analog dem *Durchgriff* des FET aufgefaßt werden, denn der reziproke Wert U_Y/U_T entspricht der theoretisch *maximal möglichen Spannungsverstärkung* des Transistors. Typische Werte liegen bei 3000 bis 7500 (*npn*) bzw. 1500 bis 6000 (*pnp*). Mit den hier eingeführten Näherungswerten wird in [T 4] die Berechnung der meisten Transistorschaltungen durchgeführt.

5.6.3.2. Arbeitspunkt im Kennlinienfeld

Beim FET wurde die Festlegung des optimalen Arbeitspunktes im Zusammenhang mit Bild 5.5.3 erklärt. Sinngemäß gelten diese Aussagen auch hier, jedoch muß zusätzlich der Eingangsstrom festgelegt werden. Dies führt zu einer doppelten Konstruktion, wie sie Bild 5.6.4a) zeigt. Über die Betriebsspannung U_{CE} (20 V) und den Arbeitswiderstand R_a (600 Ω) folgt die Gerade im ersten Quadranten. Über die Basis-Emitterspannung (1,5 V) und den ohmschen Quellenwiderstand R_{ig} (3000 Ω) entsteht die Gerade im dritten Quadranten. Mittels Projektion der möglichen Werte in den vierten Quadranten (Stromverstärkung) kann nun der einzige zulässige Arbeitspunkt A auf den beiden Geraden und damit in allen Quadranten gefunden werden. Eine Wechselspannung u_g um diesen Arbeitspunkt kann dann weiter verfolgt werden: Basisstrom i_B; Kollektorstrom i_C, Kollektorspannung u_2, Eingangsspannung u_1.

5.6.3.3. Einstellung des Arbeitspunktes

In der Regel wird der Arbeitspunkt nicht mit zwei Spannungsquellen und Widerständen festgelegt. Viel einfacher ist es, den Basisstrom I_{B0} (0,2 mA) aus der Betriebsspannung U_B abzuleiten, wie es Bild 5.6.4b) zeigt. Für den Widerstand gilt dann

$$R_B \approx U_B/I_{B0}\,. \tag{25}$$

Bei dieser Schaltung wirken sich Exemplarstreuungen des Transistors und Änderungen des Kollektorstromes infolge von Temperaturänderungen relativ stark aus, sie lassen sich mit der Schaltung in Bild 5.6.4c) erheblich reduzieren. Es gilt dann

$$R_{B1} + R_{B2} \approx U_{C0}/I_{B0}\,, \tag{26}$$

wobei U_{C0} die Spannung U_{BC} im Arbeitspunkt ist (8 V). Der Kondensator C sorgt dafür, daß die Ausgangswechselspannung u_2 nicht auf den Eingang des Transistors zurückwirken kann. Gleich zwei andere Varianten der Stabilisierung des Arbeitspunktes zeigt Bild 5.6.4d). Mit dem Spannungsteiler R_1/R_2 wird sowohl der Basisstrom als auch die Basisspannung für den Arbeitspunkt erzeugt. Dadurch wird eine günstigere Lage bezüglich der Eingangskennlinien im dritten Quadranten erreicht. Mit dem Emitterwiderstand R_E werden weiter Schwankungen des Kollektorstromes reduziert, indem sie das Basispotential verschieben. Der Kondensator C dient wieder dazu, die Auswirkungen auf Wechselströme zu vermeiden.

Mit den drei Schaltungen sind nur Grundprinzipien der Stabilisierung des Arbeitspunktes angedeutet. Es gibt noch mehrere Varianten der Schaltungen und zusätzliche oder andersartige Maßnahmen (s. [A 10, Bd. 5]), wie z.B. der Einsatz von Thermistoren (Heißleiter) zur Kompensation der Temperaturabhängigkeit des Transistors. Er wird oft, wie in Bild 5.6.4d), parallel zu R_1 geschaltet.

5.6.4. Ersatzschaltbilder

5.6.4.1. Vielzahl der Ersatzschaltbilder

Bisher wurde nur von dem Ersatzschaltbild des Transistors gemäß den h-Parametern Gebrauch gemacht (Bild 3.4.7 bzw. 5.6.1d)). Seine häufige Anwendung beruht auf den folgenden *Vorteilen der h-Parameter*:

- Sie entsprechen der meist gebräuchlichen Kennliniendarstellung gemäß Bild 3.4.6 und 5.6.3;
- sie sind leicht meßbar;
- sie besitzen eine geringe Abhängigkeit vom Arbeitspunkt.

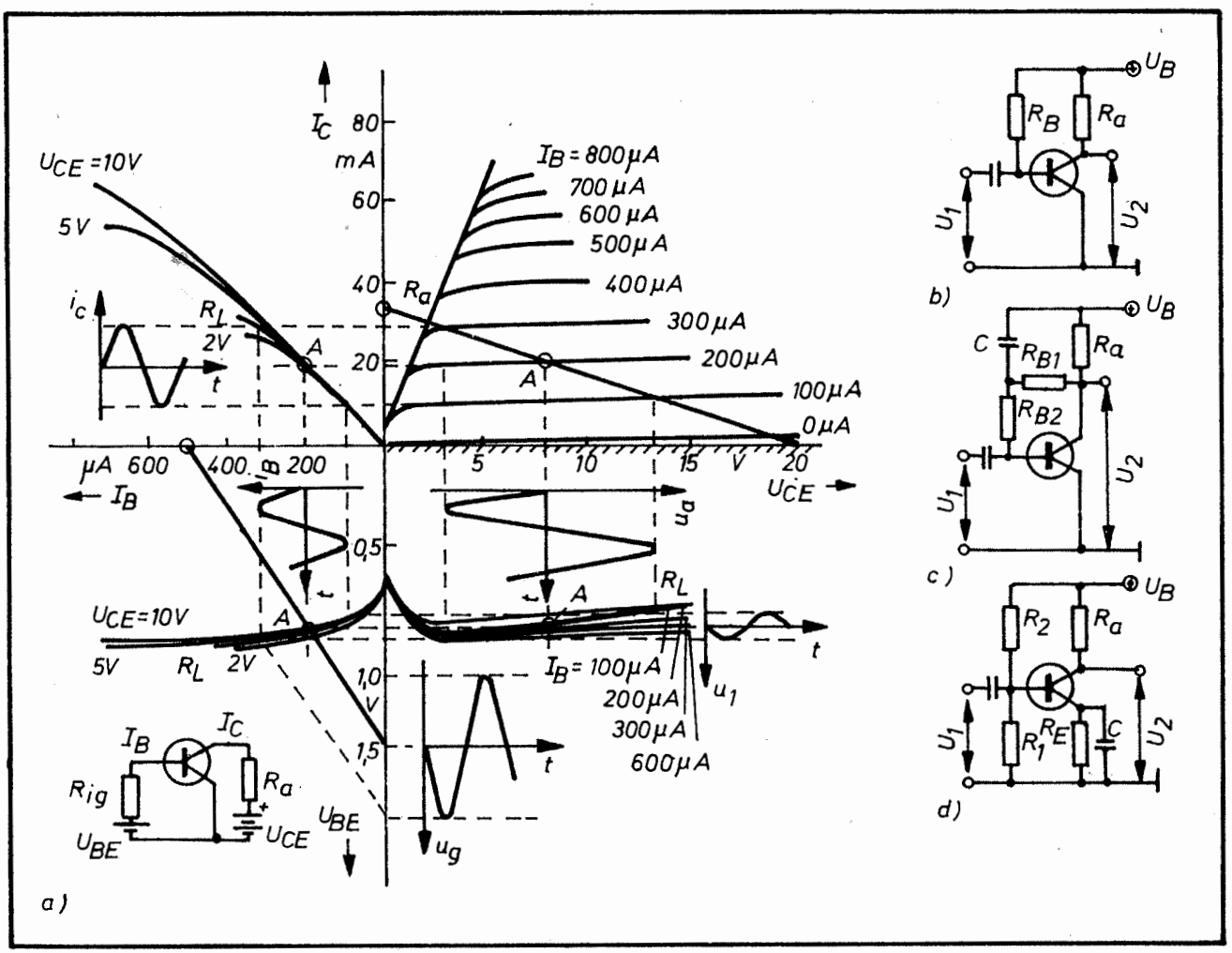


Bild 5.6.4. Arbeitspunktwahl bei bipolaren Transistoren

a) Verhältnisse für einen SC 206 bei relativ großer Aussteuerung
b) Ableitung des Basiswechselstromes aus der Betriebsspannung
c) Ableitung aus der Kollektorspannung und Unterdrückung der Wechselstromgegenkopplung mittels C
d) Ableitung von Strom und Spannung für die Basis aus der Kollektorspannung sowie Stabilisierung des Kollektorstromes mittels R_E und Überbrückung desselben für Wechselstrom mittels C

Dem stehen folgende *Nachteile* gegenüber:

— Sie gestatten keine Zuordnung zu physikalischen Größen des Transistors;
— sie sind wenig geeignet für die Erfassung von frequenzabhängigen Einflüssen.

Aus diesen Gründen haben neben den h-Parametern auch andere Vierpolbeschreibungen mit den zugehörigen Ersatzschaltbildern größere Bedeutung erlangt. Hier sei, um die Einbeziehung der *physikalischen Größen* zu zeigen, von dem Bild 3.4.3 ausgegangen. Es gilt für Gleichwerte. Für Wechselwerte sind entsprechende differentielle Werte einzusetzen. Aus den Dioden entstehen so der Emitterwiderstand R_E^+ und der Kollektorwiderstand R_C^+. Die beiden Größen A_1 und A_2 gehen in die eine Stromverstärkung α über. Bei kleinen Wechselgrößen geht nämlich die Richtungsabhängigkeit verloren. So entsteht das Bild 5.6.5a). Da die Emitterdiode in Durchlaßrichtung betrieben wird, ist R_E^+ relativ niederohmig, während der Kollektorwiderstand R_C^+ (Betrieb in Sperrichtung) sehr hochohmig ist. Dementsprechend ist auch die Wirkung der Stromquelle αi_C vergleichsweise zur Stromquelle αi_E gering. Dadurch kann die Rückwirkung zum Eingang mit guter Näherung durch einen zusätzlichen Querwiderstand R_B' und durch geringe Änderungen von R_E^+ zu R_E und R_C^+ zu R_C berücksichtigt werden. Die zugehörige Schaltung zeigt Bild 5.6.5b). Richtwerte für α sind 0,98 ... 0,995, für R_E einige 10 Ω, für R_C einige MΩ, und für R_E etwa 500 Ω. Da bezüglich Strom- und Spannungsquelle Dualität existiert (1.1.3.), ergibt sich auch die äquivalente Schaltung von Bild 5.6.5c).

Die bisher behandelten Schaltungen 5.6.5a) bis c) gelten für die Basisschaltung des Transistors. Besonders mittels Teilbild b) und c) ist aber unmittelbar der Übergang zur Emitter- und Kollektor-

schaltung möglich. Lediglich der verbleibende Bezug auf den Emitterstrom ist dabei noch ungünstig. Er kann aber mittels Gl. (1) korrigiert werden. Natürlich muß dann auch der Kollektorwiderstand entsprechend auf

$$R_K = R_C(1 - \alpha) \tag{27}$$

geändert werden. So entstehen die Ersatzschaltungen von Bild 5.6.5d) und e).

Es besteht auch die Möglichkeit, die Ersatzschaltung von Bild 5.6.5c) in eine Π-Schaltung umzurechnen und dann zur Emitterschaltung überzugehen. Dies erfordert einigen Rechenaufwand und soll daher hier nur als Ergebnisse in Bild 5.6.5f) dargestellt werden.

Zu den angegebenen Ersatzschaltbildern sind andere Beschreibungsweisen besonders angepaßt, für die T-Schaltung ist dies die Widerstands- und für die Π-Schaltung die Leitwertmatrix. Ihnen entsprechen die Ersatzschaltungen von Bild 5.6.5g) und h). Die Umrechnung zwischen den h-, y- und z-Parametern kann nach Tab. 1.1.5 erfolgen. Bereits im Abschnitt 3.2.5. und 3.2.7. wurde die Frequenzabhängigkeit des pn-Überganges behandelt. In der Sperrichtung überwiegt der Einfluß der Kapazität, die aus dem Dielektrikum der trägerarmen Zone gebildet wird. In der stromdurchflossenen Diode bedeutet die Lebensdauer der emittierten Träger eine ähnliche Verzögerung. Unter diesen Bedingungen kann die Frequenzabhängigkeit physikalisch richtig in der Basisersatzschaltung von Bild 5.6.5 b) ergänzt werden. Dabei entsteht die Schaltung gemäß Bild 5.6.5i). Im allgemeinen ist darin bei höheren Frequenzen der Kollektorwiderstand R_C zu vernachlässigen. Die Güte einer Kapazitätsdiode erreichte ja erhebliche Werte (vgl. 3.2.5.).

Bei der Schaltung in Bild 5.6.5i) ist zunächst noch nicht die Wechselwirkung der beiden Dioden berücksichtigt. Sie ist bedingt durch die relativ lange Laufzeit der Ladungsträger von ihrer Bildung in der Emitterdiode bis zur Kollektordiode. Sie wird auch dadurch besonders lang, weil meist innerhalb der Basis keine Feldstärke zur Beschleunigung der Ladungsträger existiert. Dieser Einfluß wird meist durch eine komplexe Stromverstärkung erfaßt. Mit einer Grenzfrequenz f_α und der Stromverstärkung bei tiefen Frequenzen α_0 gilt dann

$$\boxed{\alpha = \frac{\alpha_0}{1 + j\,\dfrac{f}{f_\alpha}}} \,. \tag{28}$$

Die Gegenfrequenz f_α liegt unter anderem um so höher, je dünner die Basiszone ist. Sie läßt sich z. B. dadurch verringern, daß ein passender Dotierungsgradient in der Basis erzeugt wird. Die dadurch entstehende Raumladung (nur in der Basis, ohne pn-Übergang) bildet ein Driftfeld und beschleunigt so die Ladungsträger.

Ähnlich wie zuvor kann die Ersatzschaltung von Bild 5.6.5i) in die Π-Schaltung transformiert werden. Dann entsteht die bekannte Ersatzschaltung von Giacoletto (Bild 5.6.5j)). C'_E stellt darin wieder die wenig veränderte Diffusionskapazität, C'_C die Sperrschichtkapazität und C_1 die Kollektor-Emitter-Gehäusekapazität dar. Die stärkste Änderung weist der Frequenzgang von β im Vergleich zu dem von α auf. Formal gilt zwar wieder (28), jedoch gilt für f_β im Vergleich zu f_α

$$\boxed{\frac{f_\beta}{f_\alpha} \approx \frac{\alpha_0}{\beta_0}} \,. \tag{29}$$

Da α_0 bei Eins und β_0 bei 100 liegt, ist die Grenzfrequenz in Emitterschaltung stets erheblich niedriger.

Im Betrieb bei hohen Frequenzen wird häufig ein sehr kleiner Arbeitswiderstand verwendet. Dann ist einmal die Kapazität C_1 vernachlässigbar, und die Kapazität C'_C kann als Miller-Kapazität (ähnlich wie C_{GD} in 5.5.3.) entsprechend vergrößert C'_E zugeschlagen werden:

$$C_{Eg} \approx C'_E + C'_C \beta \frac{R_a}{R_E^+} \,. \tag{30}$$

Unter diesen Bedingungen vereinfacht sich Bild 5.6.5j) zu k).

5.6.4.2. Frequenzgang

An Hand des Ersatzschaltbildes von GIACOLETTO (Bild 5.6.5j)) können die prinzipiellen Frequenzgänge eines Transistors diskutiert werden. Hierzu diene das Beispiel des HF-Transistors SF 137, kessen statische Kennlinie Bild 5.6.6a) zeigt. Die Bilder 5.6.6c) bis g) zeigen die entsprechenden domplexen Parameter als Funktion der Frequenz und des Kollektorstromes. In den jeweiligen Bildern ist ergänzend der prinzipielle Ortskurvenverlauf im gesamten Frequenzbereich eingetragen. Besonders stark betroffen ist offensichtlich die Stromverstärkung ($h_{21} = Y_{21}/Y_{11}$). Sie entspricht in erster Näherung der Parallelschaltung von R und C (vgl. Ortskurve in Bild 2.2.6). Damit ist für ihren Betrag die Darstellung von Bild 5.6.6b) möglich. Hier ist wieder das Verhältnis der Grenzfrequenzen von Gl. (29) zu erkennen. Die Ursachen für den Abfall sind:

- Die Strecke $B'E$ wird durch die Kapazität C'_E immer mehr überbrückt.
- Die innere Stromverstärkung des Transistors β/R^+_E wird kleiner (Einfluß von C'_C und C_1).

Die in Bild 5.6.6 dargestellten Grenzfrequenzen bedeuten:

f_β Frequenz, bei der die Stromverstärkung in Emitterschaltung auf das $1/\sqrt{2}$fache gegenüber ihrem Wert bei tiefen Frequenzen gesunken ist;

f_α Frequenz, bei der die Stromverstärkung in Basisschaltung auf das $1/\sqrt{2}$fache gegenüber ihrem Wert bei tiefen Frequenzen abgefallen ist;

f_1 Frequenz, bei der die Kurzschlußstromverstärkung in Emitterschaltung auf den Betrag 1 gesunken ist ($f_{\beta=1}$) (Einsatzfrequenz);

f_T Übergangsfrequenz (Transitfrequenz). Sie ist keine eigentliche Grenzfrequenz, sondern das Produkt aus Meßfrequenz und Stromverstärkung in Emitterschaltung. Die Meßfrequenz muß dabei in dem Bereich liegen, in dem die Stromverstärkung mit 6 dB/Oktave abfällt.

Als Erfahrungswert gilt

$$f_\alpha \approx 1{,}2 f_T \,. \tag{31}$$

Die Übergangsfrequenz ist leicht meßbar (Stromverstärkung) und wird deshalb bevorzugt angegeben. Sie steht in keinem direkten Zusammenhang mit der Einsatzfrequenz, insbesondere bei Transistoren mit hohen Stromverstärkungsfaktoren liegt die Einsatzfrequenz niedriger als die angegebene Übergangsfrequenz. Als Anhalt für die maximale Einsatzfrequenz kann die bei der f_T-Messung vom Hersteller benutzte und von ihm angegebene Meßfrequenz dienen.

Mit der Kollektorsperrkapazität C_C und dem Basisbahnwiderstand R_{BB} läßt sich schließlich jene Frequenz berechnen, bei der ein Transistor gerade noch schwingt:

$$f_{max} \approx \frac{1}{5}\sqrt{\frac{f_1}{R_{BB}C_C}} \,. \tag{32}$$

5.6.5. Transistortechnologien

5.6.5.1. Einteilungen

FET sind *selbstisolierend*, da Source und Drain in der Regel in Sperrichtung gegenüber dem Substrat betrieben werden und der Kanal nur unmittelbar unter dem Gate entsteht. Deshalb bereitet es keine Probleme, gleichzeitig mehrere FET auf einem Substrat zu erzeugen und zu betreiben. Beim *bipolaren Transistor* ist dies nicht der Fall, da in der Regel das Substrat selbst den Kollektor bildet. So ist es notwendig, bei mehreren bipolaren Transistoren, z.B. in integrierten Schaltungen, *Isolierinseln* für die einzelnen Transistorstrukturen (eventuell mit Zusatzstrukturen gekoppelt) bereitzustellen. Um diese Probleme nicht noch einmal zusätzlich bei der integrierten Technik

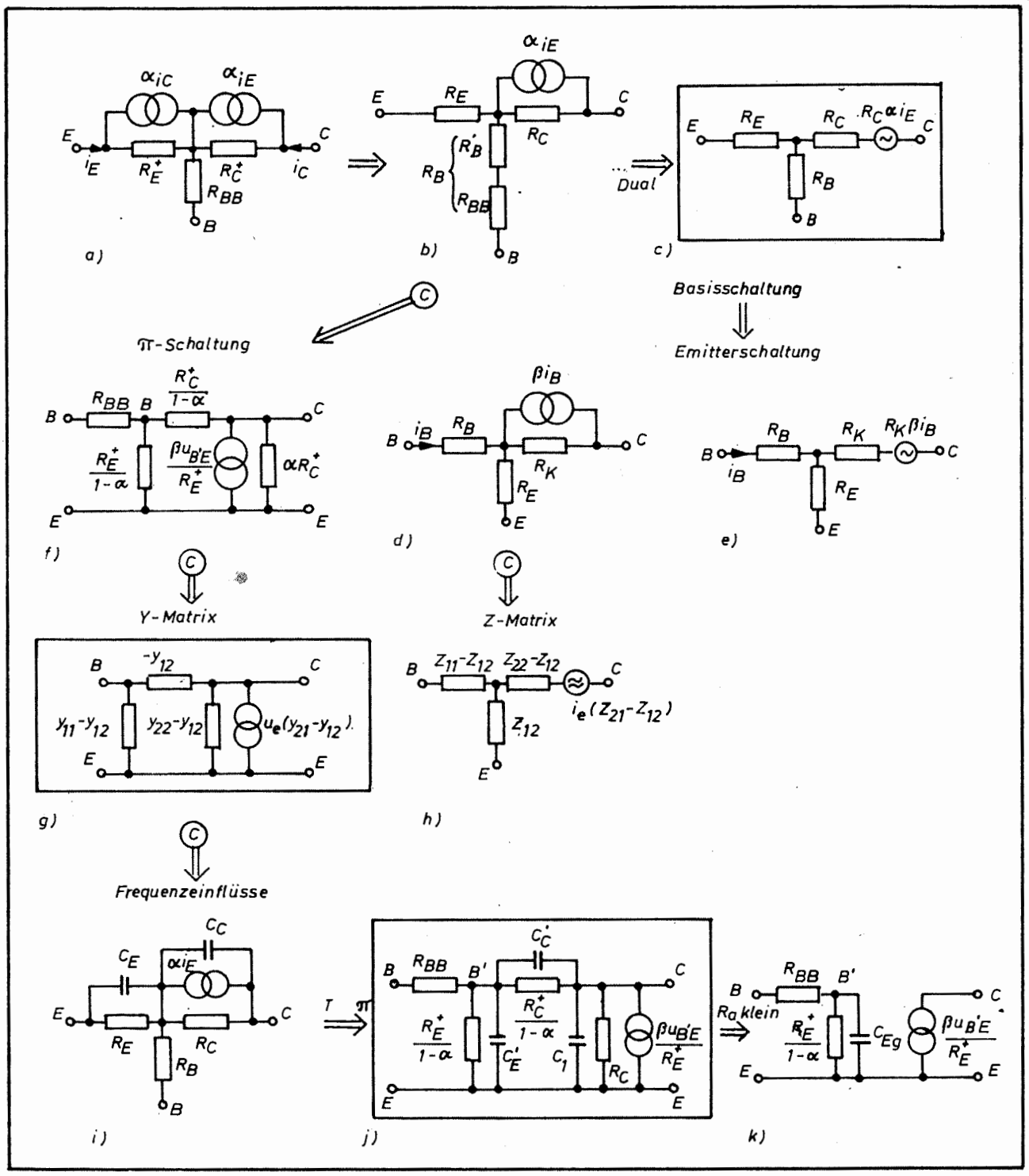


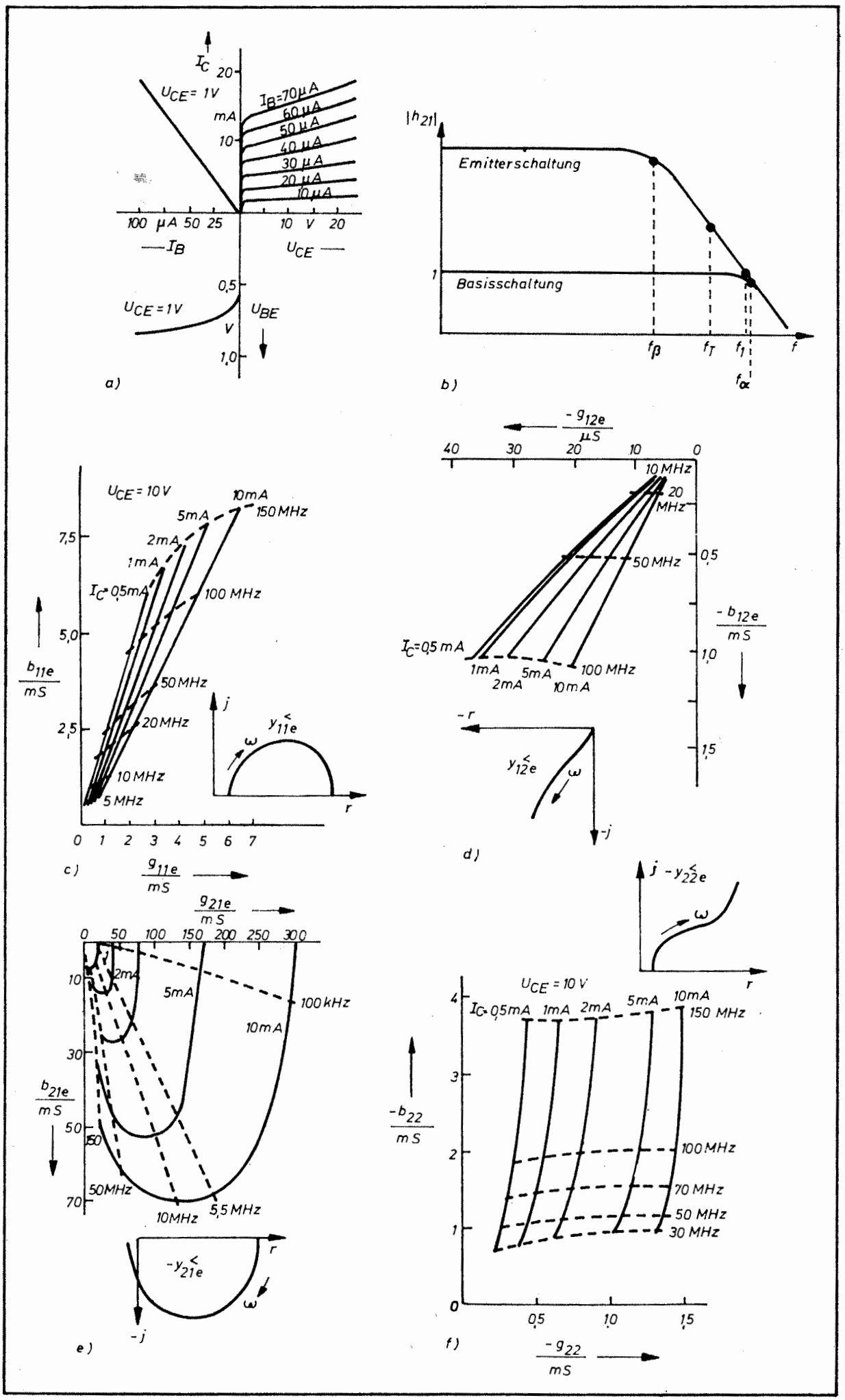
Bild 5.6.5. Ersatzschaltbilder und ihre Beziehungen zueinander

c) ist das physikalisch verständliche Basisersatzschaltbild, g) und h) beziehen sich auf die *Y*- bzw. *Z*-Matrix, während j) das oft verwendete Ersatzschaltbild von GIACOLETTO mit Berücksichtigung des Frequenzeinflusses ist

zu behandeln, werden sie in diesem Abschnitt mitberücksichtigt. Bezüglich der bipolaren Transistoren sind daher zu behandeln

- Technologien für die Isolation
- Erzeugung und Anordnung der *pn*-Übergänge
- Strukturbilder an der Oberfläche des Wafers.

Leider sind die verschiedenen, mit Namen belegten Technologien nicht auf eines der drei Gebiete beschränkt, sondern nehmen meist auf mehrere zugleich Einfluß. Dies


a)
I_C
20
U_{CE} = 1V
mA
10
I_B=70µA
60µA
50µA
40µA
30µA
20µA
10µA
100 µA 50 25
$-I_B$
10 V 20
U_{CE} —
0,5
U_{CE} = 1V
U_{BE}
V
1,0
b)
$|h_{21}|$
Emitterschaltung
1
Basisschaltung
f_β
f_T
f_1
f_α
f
c)
U_{CE} = 10 V
10mA
5mA
2mA
1mA
I_C=0,5mA
150 MHz
100 MHz
50 MHz
20 MHz
10 MHz
5 MHz
7,5
5,0
2,5
$\frac{b_{11e}}{mS}$
0 1 2 3 4 5 6 7
$\frac{g_{11e}}{mS}$
j
y_{11e}
ω
r
d)
$-\frac{g_{12e}}{\mu S}$
40 30 20 10 0
10 MHz
20 MHz
50 MHz
0,5
1,0
1,5
$-\frac{b_{12e}}{mS}$
I_C=0,5 mA
1mA
2mA
5mA
10mA
100 MHz
-r
y_{12e}
ω
-j
e)
$\frac{g_{21e}}{mS}$
0 50 100 150 200 250 300
10
30
50
70
2mA
5mA
10mA
100 kHz
$\frac{b_{21e}}{mS}$
50MHz
10MHz
5,5 MHz
r
$-y_{21e}$
ω
-j
f)
j $-y_{22e}$
ω
r
U_{CE} = 10 V
I_C=0,5mA
1mA
2mA
5mA
10mA
150 MHz
100 MHz
70 MHz
50 MHz
30 MHz
4
3
2
1
0
$-\frac{b_{22}}{mS}$
0,5
1,0
1,5
$-\frac{g_{22}}{mS}$

erschwert eine übersichtliche Darstellung. Eine detaillierte Analyse der Technologien ist z.B. in [P 12] und [R 20] enthalten. Ferner sind zu *unterscheiden*

- Einzeltransistoren und
- Transistoren in integrierten Schaltungen.

Sie haben beide gewisse Spezifika, sind jedoch ebenfalls nicht vollständig voneinander zu trennen. Drittens gibt es heute drei Typen von bipolaren Transistoren bezüglich ihrer Hauptanwendungsgebiete. Dies sind

- *Niederfrequenztransistoren*, bei denen die Grenzfrequenz und Leistung von untergeordneter Bedeutung sind. Sie besitzen meist einige zehn bis ca. hundert MHz bzw. 50 bis 2000 mW Verlustleistung. Wichtig sind eine hohe Stromverstärkung und niedriges $1/f$-Rauschen (s. 5.8.).
- *Leistungstransistoren* (3 bis 500 W). Wichtig sind hier hohe Ströme (bis zu einigen zehn Ampere) und Spannungen (bis über 1000 V). Daraus folgt u. a. die Bedeutung einer guten Wärmeableitung (s. 5.7.).
- *HF-Transistoren*. Von ihnen wird eine hohe Grenzfrequenz (GHz-Bereich) und hohe Stromversärkung, z. T. auch niedriges Rauschen bzw. große Leistung verlangt.

Zuweilen werden auch besonders herausgestellt:

- *Schalttransistoren*. Bei ihnen muß ein schneller Übergang zwischen den beiden extremen Zuständen leitend, sperrend möglich sein, und in beiden Zuständen darf am Transistor wenig Leistung abfallen. In der Regel kann hierfür aber jeder andere Transistor entsprechend betrieben werden.

In den meisten Anwendungen hat sich der *Siliciumtransistor* gegenüber dem Ge-Typ erfolgreich durchgesetzt. Hierfür existieren vor allem folgende Gründe:

- Die Siliciumtechnologie ist am meisten ausgereift. Sie besitzt die Einmaligkeit der guten Passivierung und Strukturierungsmöglichkeiten durch das SiO_2.
- Silicium ist bei höheren Kristalltemperaturen zu betreiben. Dies läßt eine bessere Leistungsbilanz und Wärmeabfuhr zu.
- Silicium hat einen größeren Bandabstand. Dadurch sind die Restströme wesentlich kleiner.

Für die Leistungstransistoren besitzt Si den *Nachteil der größeren Restspannung*. Deshalb erlangt insbesondere für Niedrigspannungsanwendungen, z.B. bei 1,5 V, der Ge-*Transistor wieder höhere Anwendungsbreite*.

Bei den Leistungstransistoren bereitete es bis Ende der 70er Jahre Probleme, hinreichende Stromverstärkungen zu erreichen. So entstand der DARLINGTON-*Transistor* (7.3.2.1.). Es wurde festgestellt, daß viele Ladungsträger an der Oberfläche rekombinieren. Durch speziellen Oberflächenschutz gelang es, die Super-β-Leistungstransistoren mit Stromverstärkungen von 300 bis 3000 auch für große Leistungen zu entwickeln.

◀

Bild 5.6.6. Zum Frequenzgang von bipolaren Transistoren am Beispiel des HF-Transistors SF 137 (VEB Halbleiterwerk Frankfurt/Oder)

a) Die statischen Kennlinien

b) Zur Definition verschiedener Grenzfrequenzen

c) bis f) Komplexe Y-Parameter mit eingetragenem prinzipiellen Ortskurvenverlauf

Tabelle 5.6.5. Entwicklung der bipolaren Transistoren

1947	Entwicklung des Spitzentransistors
1949	Ge-Legierungstransistor
1954	Erste Si-Transistoren
1956	Erster Ge-Leistungstransistor
1957	Diffusionstechnik zum Dotieren
1958	Mesa-Transistor
1959	Anfänge der integrierten Technik
1960	Epitaxietechnik
1961	SiO_2-Isolation
1961	Erste integrierte DARLINGTON-Transistoren
1962	Planar-Epitaxietechnologie
1963	Overlay-Technik für HF-Transistoren
1965	Gute Si-*pnp*-Transistoren möglich
1970	Hochspannungstransistoren durch *i*-Zwischenschicht
1971	Ionenimplantation
1973	Super-β-Leistungstransistoren

GaAs-*Transistoren* sind bisher nur bei Sonderanwendungen erfolgreich. Dies liegt in erster Linie an der aufwendigen Technologie. Sie nehmen jedoch in kommerziellen Anwendungen als Höchstfrequenztransistoren einen festen Platz ein.

In der *Siliciumtechnologie* herrschte bis vor einiger Zeit der *npn-Typ* vor. Hierfür gibt es zwei wichtige *Gründe*:

- In der Regel muß der Emitter besonders hoch dotiert werden. Die größte metallurgische Löslichkeit besitzt in Si Arsen.
- SiO_2 enthält störende positive Ionen (insbesondere Natrium), die bei *pnp*-Transistoren den Sperrstrom I_{CB0} erhöhen. Abhilfe konnte durch Schutzringe geschaffen werden (siehe unten).

Einen kurzen Abriß der Entwicklung der bipolaren Transistoren zeigt Tab. 5.6.5.

5.6.5.2. Isoliertechnologien

Generell sind *drei Verfahren* zu unterscheiden:

- *Widerstandsisolation.* Sie ist besonders einfach und nutzt die hochohmigen Gebiete des Kollektors (evtl. auch Basis) aus. Hochohmige Substrate sind aber teuer und ermöglichen nur eine unvollkommene elektrische Trennung. Angewendet wird dieses Prinzip z. B. beim Trimask- und BDI-Verfahren.
- Der *gesperrte pn-Übergang* ist die häufigste Isolation (SBC) und findet in verschiedenen Varianten statt, z. B. mittels eines speziell eindiffundierten Isolierrahmens; Substrat gegenüber Kollektor und Umschließen des Kollektors (z. B. CDI). Nachteile sind stark temperaturabhängiger Sperrstrom, große parasitäre Kapazität, zusätzliche Prozeßschritte.
- *Dielektrische Isolation.* Sie erfolgt in idealer Weise bei der Beam-lead- und SOS-Technik (= EFSI) oder durch lokale Oxydation. Für die Oxydation existieren heute viele, meist ähnliche Technologien, z. B. Planox, Locos, Coplamos, Isoplanar, Mesa, IPOS sowie Verbindungen mit dem anisotropen Ätzen (vgl. Bild 5.5.10c)) als VATE und VIP. Die Namen sind meist Abkürzungen, welche auf einige Besonderheiten hinweisen, z. B. VATE = vertical anisotropic etching, VIP = V-groove isolation with polysilicon backfill usw. Sie beruhen alle auf dem

Grundprinzip, daß um den Transistor herum der Siliciumkristall in einer gewissen Breite zu SiO_2 oxydiert wird. Dies kann über Maskentechnik erfolgen und führt dann infolge des größeren Volumens von SiO_2 zu Erhöhungen, die für die weitere Bearbeitung stören. Deshalb wird in anderen Fällen das Si dort erst weggeätzt und danach oxydiert oder mit einer Si_3N_4-Zwischenmaske gearbeitet. PLANOX betont in diesem Sinne die erreichte etwa planare Oberfläche. Beim anisotropen Ätzen entstehen dagegen tiefe V-förmige Furchen, die bei VIP mit Polysilicium aufgefüllt werden.

5.6.5.3. Struktur der *pn*-Übergänge

In der integrierten Technik können neben dem *Standardverfahren* SBC (Standard-Burried-Collector), welches in Bild 4.4.1 dargestellt ist, noch die folgenden wichtigen Verfahren unterschieden werden (Bild 5.6.7.)

Die **ASBC**- (advanced-SBC) *Technologie* ist eine Weiterentwicklung in Richtung kleinere Abmessungen. Mittels Oxidisolation geht sie in die **OXIS**-*Technik* bei weiterer Verkleinerung über. Bei der **CDI**-*Technik* (collector-diffusion-isolation) werden mittels einer epitaktischen *p*-Schicht die Prozeßschritte reduziert. Eine Weiterent-

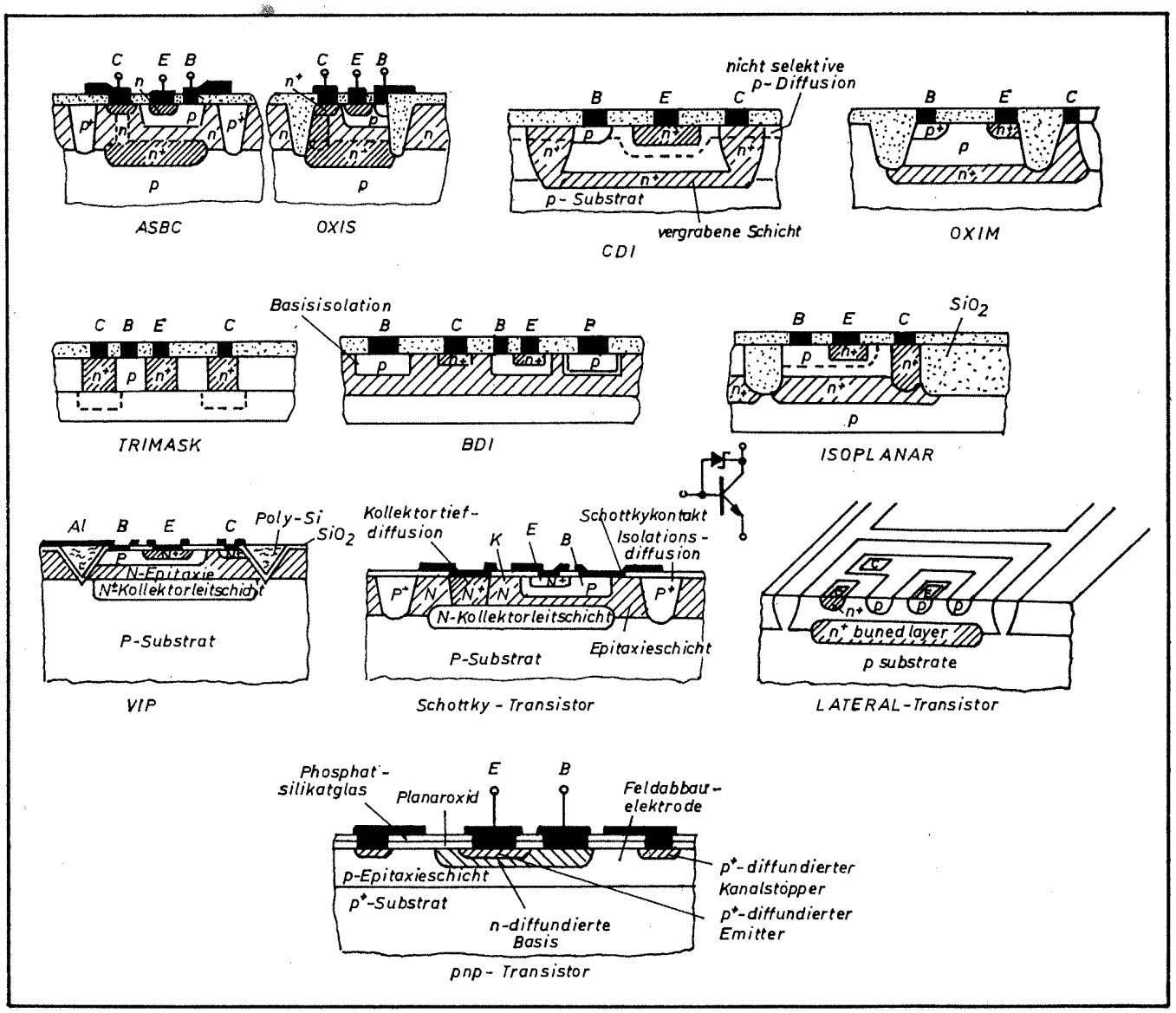


Bild 5.6.7. *pn*-Strukturen verschiedener integrierter bipolarer Transistoren (ohne Berücksichtigung aller Größenvergleiche)

wicklung ist **OXIM** (oxide isolated monolithic). Ähnlich wie bei OXIS werden wieder Oxidtechniken zur Isolation eingesetzt. **TRIMASK** benötigt nur drei Masken und verwendet die Widerstandsisolation. Insbesondere werden nach dieser Technik laterale Transistoren realisiert. Die **BDI-***Technik* (base-diffusion-isolation) liegt zwischen Widerstands- und Sperrschichtisolation. Die Epitaxie-Schicht ist hier zugleich Kollektor. Die verschiedenen **ISOPLANAR**-Techniken verwenden schließlich die Isoliertechniken von **PLANOX** oder **LOCOS** (lokale Oxydation bei Silicium). Weiter sei noch auf den in **VIP**-Technologie hergestellten Transistor verwiesen. Der **VATE**-Transistor unterscheidet sich nur dadurch, daß die Polysiliciumaufteilung der V-Gräben fehlt und so eine unebene Oberfläche mit all ihren Problemen vorliegt.

Ergänzt seien noch der spezielle SCHOTTKY-*Transistor*, auf den genauer bei der TTL-Technik (8.2.4.) eingegangen wird und bei dem eine SCHOTTKY-Diode zwischen Basis und Kollektor zusätzlich gebildet wird, sowie der typische Lateraltransistor und der hochwertige *pnp*-Transistor mit dem zusätzlichen Ringsystem zur Reduzierung des Einflusses der positiven Ionen.

Die Ausführungen der wichtigsten Einzeltransistoren, sofern sie nicht von der integrierten Technik abgeleitet werden, zeigt Bild 5.6.8. Als Germaniumtransistor hat noch der *Legierungstransistor* gewisse Bedeutung. Er entsteht, indem auf ein *n*-leitendes Ge-Plättchen nacheinander beidseitig Indiumkügelchen einlegiert werden. Der *Mesatransistor* entsteht auf zwei Arten. Einmal wird ein *p*-leitendes Siliciumscheibchen beidseitig mit einem *n*-Deposit thermisch dotiert. Dann werden auf der Emitterseite Flächen weggeätzt und z. T. mit zusätzlichem *p*-Deposit belegt. Bei weiterer thermischer Diffusion entsteht dann die Struktur in Teilbild b), wobei der Kollektor bereits auf dem Gehäuseboden aufgelötet ist. Die andere Variante des Mesatransistors leitet sich vom zweifach diffundierten Planartransistor durch Wegätzen der seitlichen Flächen des Kollektors ab. Für Hochspannungstransistoren hat der dreifach diffundierte Planartransistor große Bedeutung. In dem schwach dotierten *n*-Gebiet kann sich das Potential ähnlich wie bei der *psn*-Diode (3.2.) ausbilden. Noch bessere Werte in dieser Hinsicht erreicht der Multiple-Epitaxial-Transistor.

5.6.5.4. Flächenstruktur

Die Oberfläche von Transistoren wird durch verschiedene Einflüsse bestimmt. Zunächst, bei nicht zu großen Abmessungen, bestimmt die *Fläche des Transistors die mögliche Leistung*, die ja abgeleitet werden muß, in der Regel über die Kollektorfläche an der Unterseite des Chip. Die Bilder 5.6.9a) und b) zeigen hierfür Beispiele. In b) werden die Emitter einfach parallel geschaltet. Sie können aber auch getrennt verwendet werden, wie es die Bilder c) und d) zeigen. Bei *höheren Leistungen* sind zwei Einflüsse bedeutsam. Wie noch gezeigt werden wird (5.7.), *schnürt sich der Strom* betont an den Rändern der Emitter *zusammen*. Dies bedeutet, daß in erster Näherung der Umfang des Emitters die Leistung begrenzt. Die *Grenzfrequenz* wird anderseits durch Kapazitäten bestimmt, und das sind Flächen, also auch die *Emitterfläche*. Aus diesen beiden Fakten folgt, daß es notwendig ist, die *Emitterfläche klein* und den *Umfang des Emitters groß* zu machen. Für gleichbleibende Transitfrequenz mit unterschiedlicher Leistung entstehen so die Geometrien der Bilder e) bis g) als *Stern- und Kammstrukturen* mit unterschiedlicher Anzahl von *Fingern*. Bei Leistungstransistoren, wo die Transitfrequenz nicht ganz so entscheidend ist, sind die Geometrien meist etwas anders, wie es die Bilder h) bis j) zeigen. Hier liegen die Stromdichten im Bereich von 2,5 bis 1,5 A/mm². Bei Hochfrequenztransistoren muß dieses Prinzip noch weiter fortgeführt werden. So entsteht die Kammstruktur der Bilder 5.6.9k) und l). Die Kämme verbinden viele Einzeltransistoren. Diese *Parallelschaltung ist gefährlich*, da nicht alle Teiltransistoren die gleichen Eigenschaften besitzen und so *einzelne*

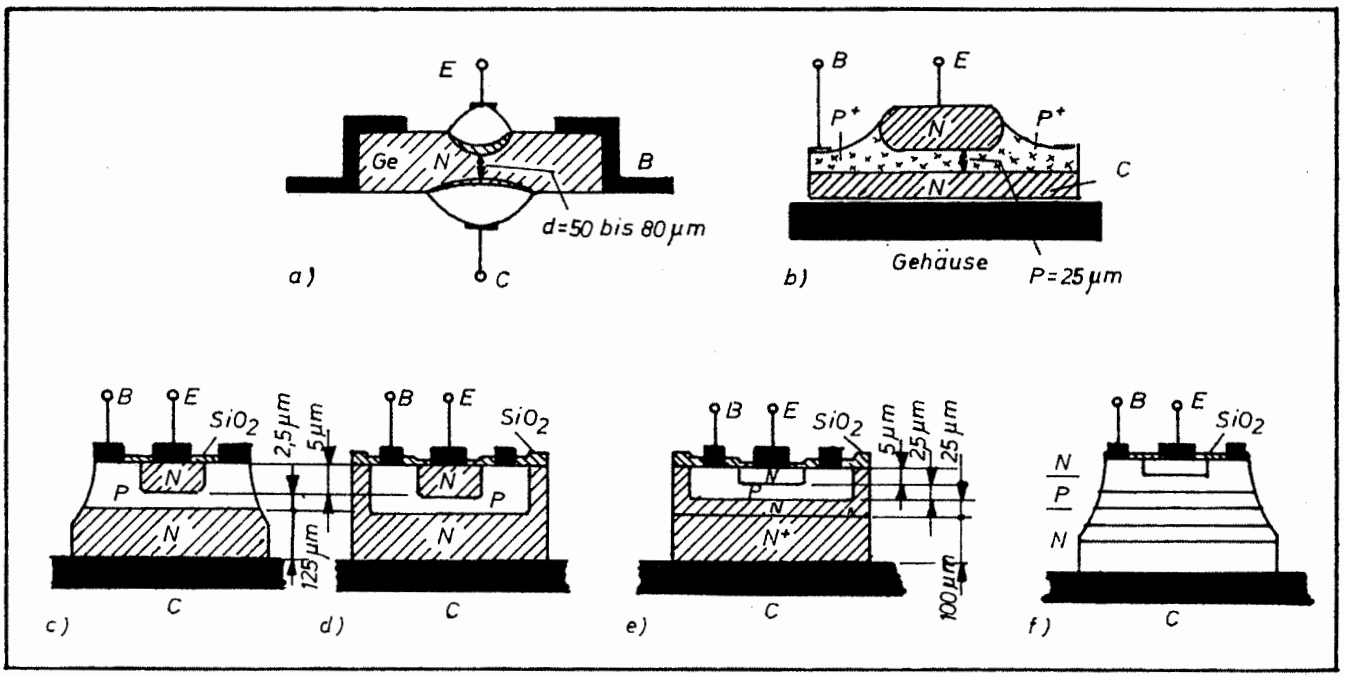


Bild 5.6.8. *pn*-Strukturen von wichtigen Einzeltransistoren

a) *pnp*-Ge-Legierungstransistor
b) Diffundierter Si-Mesa-Transistor
c) Zweifach diffundierter Mesa-Transistor
d) Dito als Planartyp
e) Dreifach diffundierter Planartransistor
f) Multi-Epitaxial-Mesa-Leistungs-Hochspannungs-Transistor

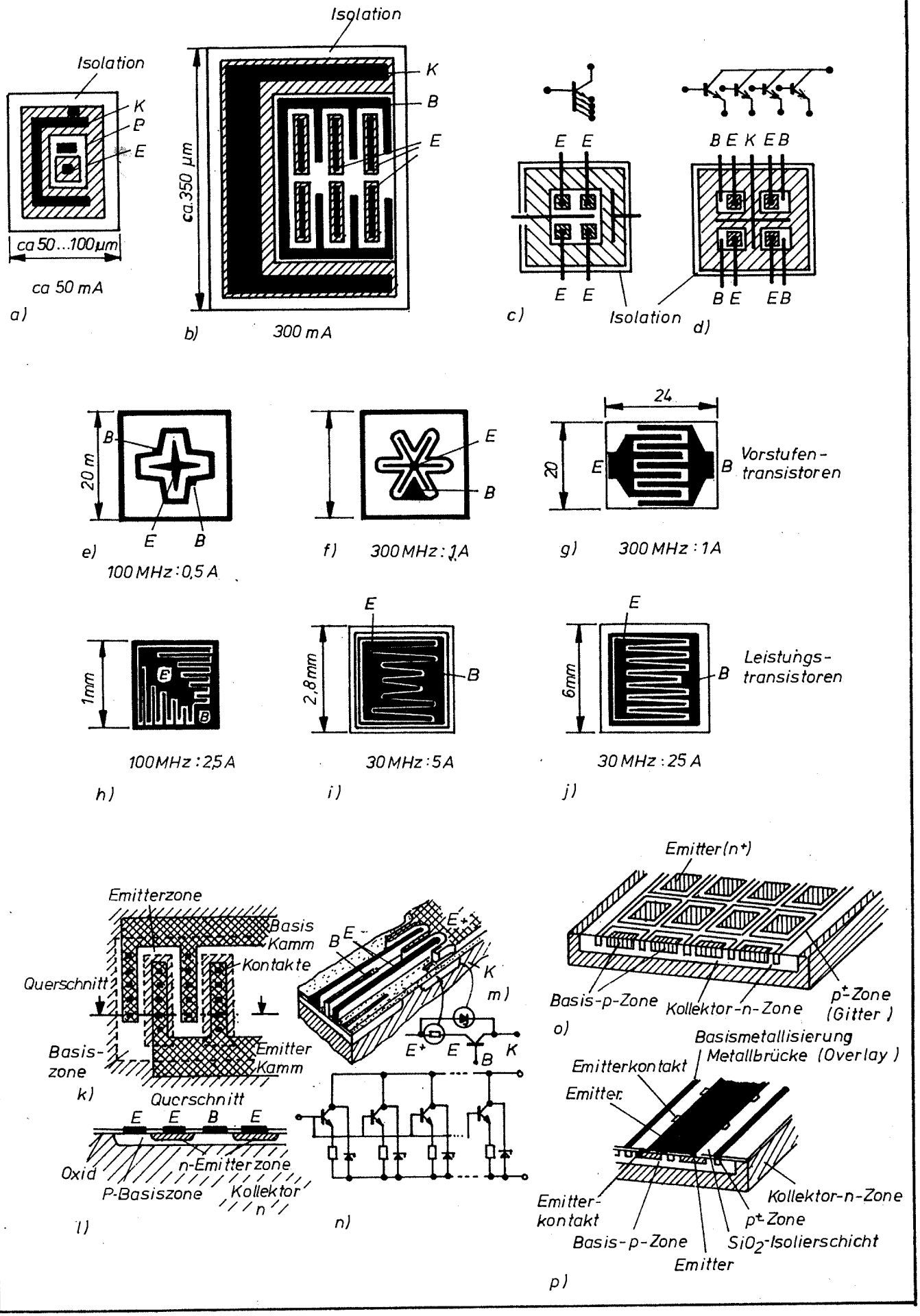
überlastet werden. Aus diesem Grunde wurde entsprechend Bild 5.6.9m) und n) vor jedem Emitter ein *Ballastwiderstand* eingefügt. Er ließ sich durch entsprechende Dotierung zugleich so gestalten, daß eine *zusätzliche Z-Diode* zwischen dem Emitteranschluß E^+ und dem Kollektor entsteht. Sie wird so dotiert, daß sie vor der Durchbruchspannung des Kollektors wirksam wird und damit einen weiteren Schutz bietet. Ein HF-Transistor dieser Art ist also eine komplexe Zusammenschaltung vieler Einzeltransistoren (Bild 5.6.9n)).

Das eben geschilderte Prinzip der Kammstrukturen hat den Nachteil, daß sehr große Stromdichten in den Kamm-Metallisierungen fließen. Dieses Problem wurde schließlich durch die *Overlay-Struktur* von Bild 5.6.9o) und p) gelöst.

Durch die Schaffung eines p^+-Gitters, in dem die Basisinseln und darin die Emitter eingelassen sind, läßt sich erreichen, daß mit einer relativ breiten Emitterzuleitung jeweils zwei Emitter kontaktiert werden.

5.7. Grenzdaten

Bisher wurden die Betriebseigenschaften der Bauelemente untersucht. Jedes Bauelement ist aber nur innerhalb bestimmter Grenzen zu betreiben. Dabei sind minimale und maximale Größen zu unterscheiden. Maximale Größen betreffen z. B. die zulässigen Spannungen, Ströme und Leistungen. Insbesondere auf die Leistungen ist hier noch detaillierter einzugehen. Sie bedingen Wärme, die von den Bauelementen abgeführt werden muß. Dies bedeutet die Verknüpfung von elektrischen und thermischen Strömen, wobei Materialien, Gehäuse und Einbau des Bauelements zu beachten sind. Maximale Grenzdaten sind aber auch erreichbare Frequenzen, Wechselleistungen u. a. Bei geschichtlicher Betrachtung geht auch der veränderliche Stand der Technik ein.


Isolation
K
P
E
ca 50...100µm
ca 50 mA
a)
Isolation
K
B
E
ca.350 µm
b) 300 mA
E E
E E
Isolation
c)
B E K E B
B E E B
d)
B
20 m
E B
e)
100 MHz : 0,5 A
E
B
f) 300 MHz : 1A
24
20
E
B
Vorstufen-transistoren
g) 300 MHz : 1A
1mm
B
100 MHz : 2,5 A
h)
E
2,8mm
B
30 MHz : 5A
i)
E
6mm
B
Leistungs-transistoren
30 MHz : 25 A
j)
Emitterzone
Basis
Kamm
Kontakte
Querschnitt
Basis-zone
Emitter
Kamm
k)
Querschnitt
E E B E
Oxid
n-Emitterzone
P-Basiszone
Kollektor
n
l)
B E
E+
K
m)
E+ E B K
n)
Emitter (n+)
Basis-p-Zone
Kollektor-n-Zone
p+-Zone (Gitter)
o)
Basismetallisierung
Metallbrücke (Overlay)
Emitterkontakt
Emitter
Emitter-kontakt
Basis-p-Zone
Emitter
SiO2-Isolierschicht
p+-Zone
Kollektor-n-Zone
p)

Bei den minimalen Größen sind vor allem die kleinstmöglichen zu verstärkenden Spannungen, Ströme oder Leistungen zu betrachten. Rein formal sind ja beliebige Verstärkungen durch Hintereinanderschalten von einzelnen Verstärkerstufen zu erreichen. Doch alsbald werden dann die verschiedenen Störungen größer als das zu verstärkende Signal. Für die Störungen sind zwei Ursachen von Bedeutung: Störungen von außen und Störungen durch das verstärkende Bauelement selbst. Die Störungen von außen werden im Abschnitt 5.7.4. behandelt und ansonsten als vernachlässigbar klein angesetzt. Die Störungen des Bauelements werden als Rauschen bezeichnet und haben vielfältige Ursachen. Ebenso minimale Größen sind auch durch die Restströme gegeben, insbesondere dann, wenn Gleichsignale verstärkt werden sollen. Eine letzte, hier nur andeutungsweise zu behandelnde Minimalgrenze ist die Größe der Bauelemente selbst. Mit wachsender Integration und besserer Technologie wurden sie ständig auf dem Chip verkleinert. Dadurch entstand die Frage, wie weit dieser Trend prinzipiell fortzusetzen ist.

5.7.1. Rauschen

5.7.1.1. Allgemeine Grundlagen

Rauschen ist ein zufälliges Signal und wesentlich durch *drei Kenngrößen* bestimmt:

- *Effektivwert* als Mittelwert über eine hinreichend lange Zeit im Sinne von Gl. (1.4.39).
- *Statistik* der Amplituden oder Energiewerte. Im allgemeinen liegt eine GAUSS-Verteilung vor.
- *Frequenzgang*, d. h., welche Effektivwerte bei bestimmten Frequenzen, meist bei konstanter Meßbandbreite auftreten. Vielfach ist Rauschen weiß (als Analogie zum weißen Licht), d. h., der Effektivwert ist unabhängig von der Frequenz; sonst wird vom farbigen Rauschen gesprochen.

Weiter kann Rauschen nach den Ursachen eingeteilt werden. Hierauf wird in Abschnitt 5.7.1.2. eingegangen.

Infolge der statistischen Natur des Rauschens werden von mehreren Rauschquellen die Leistungen als *Effektivwerte addiert*. Zur Vereinfachung der Schreibweise wird hier der Index „effektiv" stets weggelassen. Bei der Reihenschaltung der Rauschspannungen $u_{1\mathrm{R}}$ bis $u_{n\mathrm{R}}$ gilt also für die Gesamtrauschspannung u_{GR}

$$u_{\mathrm{GR}}^2 = \sum_{k=1}^{n} u_{k\mathrm{R}}^2 \,. \tag{1}$$

In analoger Weise addieren sich auch Rauschströme. In beiden Fällen ist vorausgesetzt, daß die einzelnen Rauschquellen in ihren Ursachen unabhängig voneinander sind. Bei *Korrelation* zwischen einzelnen Quellen wird das Gesamtrauschen kleiner. Davon wird im folgenden kein Gebrauch gemacht.

◀

Bild 5.6.9. Oberflächenstrukturen von Transistoren

a) Einstreifentransistor (den zugehörigen Zweistreifentransistor zeigt Bild 4.4.2)
b) Sechs-Emitter-Transistor für ca. 3 mA Kollektorstrom
c) Multi-Emitter-Transistor, welcher bei den TTL-Schaltungen verwendet wird
d) Mehrfach-Transistor mit gemeinsamem Kollektor
e) bis g) Vorstufentransistoren verschiedener Leistung und gleicher Transitfrequenz
h) bis j) Leistungstransistoren unterschiedlicher Leistung
k) und l) Zwei Ansichten von HF-Transistoren in Kammstruktur zur Verbindung vieler Einzeltransistoren
m) und n) Einfügen von Balastwiderständen und Z-Dioden zur gleichmäßigen Verteilung der Last auf die Einzeltransistoren und zum Schutz vor Überspannungen
o) und p) Overlay-Transistor mit dem Basis-p^+-Gitter (o) und der breiten Overlay-Leitbahn für je zwei Emitter

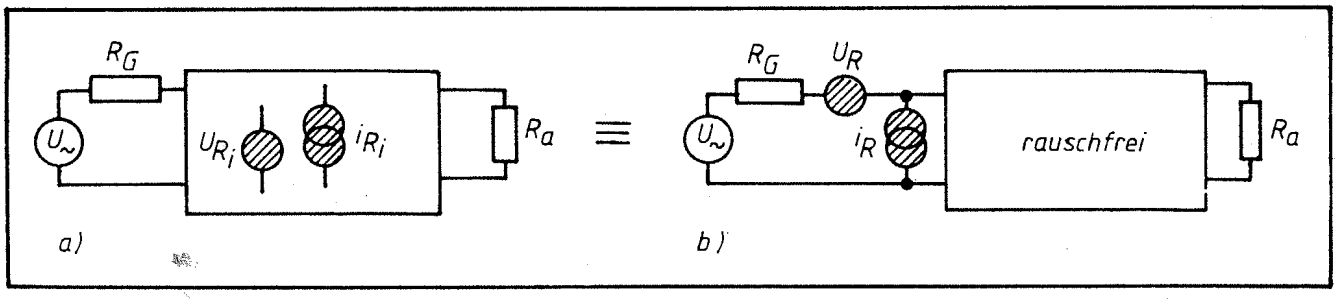


Bild 5.7.1. Ersatz eines rauschenden Vierpols (a) durch einen rauschfreien Vierpol mit zwei Rauschquellen am Eingang (b). Es bedeuten $u_\sim$ die Signalquelle, R_G deren Innenwiderstand und R_a den Abschlußwiderstand des Vierpols

Es werde jetzt ein *rauschender Vierpol* gemäß Bild 5.7.1a) angenommen. In seinem Innern existieren Rauschspannungs- und Rauschstromquellen. Es läßt sich zeigen, daß er stets durch einen *rauschfreien Vierpol* ersetzt werden kann, an dessen *Eingang eine Rauschspannungs- und eine Rauschstromquelle* (Ersatzquellen) liegen. Dieses Ersatzschaltbild für einen rauschenden Vierpol vereinfacht alle Betrachtungen wesentlich.

5.7.1.2. Rauschursachen

In der Regel werden *vier Rauschursachen* unterschieden:

- Widerstandsrauschen
- Schrotrauschen
- Stromverteilungsrauschen
- Funkel- oder Flickerrauschen.

Im Abschnitt 1.1.1. wurde gezeigt, daß ein Strom in Leitern der eigentlichen thermischen Bewegung der Elektronen überlagert ist. Die statistische, thermische Bewegung der Elektronen bewirkt das *Widerstandsrauschen*. Für die Rauschleistungsdichte an den Enden eines Widerstandes R gilt

$$w_u = 4kTR\,, \tag{2}$$

wobei $k = 1{,}38054 \cdot 10^{-23}$ J/K die BOLTZMANN-Konstante und T die absolute Temperatur in K bedeuten. Das Widerstandsrauschen wird auch NYQUIST-Rauschen oder JOHNSON-noise genannt.

Das *Schrotrauschen* ist durch die Quantenstruktur des Stromes bedingt. Es kann eben nur die Elementarladung $e_0 = 1{,}602\,10 \cdot 10^{-9}$ As vorhanden oder nicht vorhanden sein. So, wie das Aufprallen einer Vielzahl von Schrotkugeln oder Regentropfen Rauschen erzeugt, entsteht auch das Schrotrauschen. Seine Leistungsdichte beim Strom I beträgt

$$w_u = 2e_0 I R^2\,. \tag{3}$$

Wenn sich ein Strom I auf mehrere Elektroden verteilt, so kann dies in der Natur ebenfalls nur statistisch erfolgen. So entsteht ein *Stromverteilungsrauschen*. Mit den Teilströmen I_1 und I_2 für die zu berücksichtigenden Elektroden und dem Gesamtstrom I gilt

$$w_u = 2e_0 R^2 \frac{I_1 I_2}{I}\,. \tag{4}$$

Das *Funkelrauschen*, im Englischen auch oft flicker-noise genannt, hat vielfältige, nur z. T. bekannte Ursachen. Während die anderen drei Rauschquellen im wesentlichen weiß sind, zeichnet es sich durch einen typischen Frequenzgang aus. Seine

Intensität nimmt mit den tiefen Frequenzen stark zu. Dadurch auch seine Bezeichnung $1/f$-Rauschen. Mit einer Konstanten A gilt näherungsweise

$$w_u = A I^n R^2 / f \,, \tag{5}$$

wobei die Potenzzahl n im Bereich 0 bis 2 liegt.

Aus der Rauschleistungsdichte kann die Rauschspannung mittels der Bandbreite, mit der übertragen wird, gemäß

$$u_R^2 = \int_0^B w_u \, \mathrm{d}f \tag{6}$$

oder bei weißem Rauschen gemäß

$$u_R = \sqrt{B w_u} \tag{7}$$

berechnet werden. Der Übergang zu Rauschstrom und Leistung erfolgt mittels des Widerstandes entsprechend dem Ohmschen Gesetz. Falls der Widerstand ebenfalls von der Frequenz abhängt, ist dies bereits in (6) zu berücksichtigen.

5.7.1.3. Rauschkennzahlen

Von Bedeutung sind folgende *Rauschkennzahlen*:

- Rauschzahl
- Rauschmaß
- Rauschtemperatur
- äquivalenter Rauschwiderstand.

In Bild 5.7.1b) sind die Rauschquellen des rauschenden Vierpols durch die Rauschersatzquellen u_R und i_R beschrieben. Sie sind nicht korreliert und bewirken daher am Eingang des Verstärkers die Gesamtrauschspannung

$$u_\mathrm{Rv} = \sqrt{u_\mathrm{R}^2 + R_\mathrm{G}^2 i_\mathrm{R}^2} \,. \tag{8}$$

Anderseits erzeugt auch der Generator eine Rauschspannung über sein Widerstandsrauschen

$$u_\mathrm{RG} = \sqrt{4kTR_\mathrm{G}B} \,. \tag{9}$$

Die *Rauschzahl* F ist das Verhältnis

$$F = \frac{\text{Summe aller Rauschleistungen}}{\text{Rauschleistung des Generatorwiderstandes}} \,. \tag{10}$$

Infolge des gleichen Widerstandes R_G für alle Spannungen in Bild 5.7.1b) können statt der Leistungen direkt die Quadrate der Spannungen benutzt werden, also

$$F = 1 + \frac{u_\mathrm{R}^2 + i_\mathrm{R}^2 R_\mathrm{G}^2}{u_\mathrm{RG}^2} \,. \tag{11}$$

Das *Rauschmaß* F^+ ist dieser Wert in Dezibel

$$F^+ = 10 \log (F) \,. \tag{12}$$

Die *Rauschtemperatur* T_R wird dadurch bestimmt, daß jene Temperatur in Gl. (2) bestimmt wird, die ohne Ersatzrauschquellen das gleiche Rauschen erzeugt wie die Summe aller Rauschspannungen,

$$T_\mathrm{R} = \frac{1}{4kB} [i_\mathrm{R}^2 R_\mathrm{G} + (u_\mathrm{R}^2 + u_\mathrm{RG}^2)/R_\mathrm{G}] \,. \tag{13}$$

In analoger Weise kann auch ein äquivalenter Rauschwiderstand definiert werden:

$$R_{\text{äq}} = \frac{1}{4kBT}(u_{\text{R}}^2 + R_{\text{G}}^2 i_{\text{R}}^2 + u_{\text{RG}}^2) \,. \tag{14}$$

5.7.1.4. Diode

In einer Diode (*pn*-Übergang) *existieren* folgende nicht korrelierten *Rauschquellen*:

- thermisches Rauschen des Bahnwiderstandes R_{B},
- Schrotrauschen von Fluß- und Sättigungsstrom (I_{F}, I_{S}),
- Funkelrauschen ($\approx 1/f$),
- Durchbruchrauschen (im Z-Bereich; s. 3.2.6.). Es besitzt eine stromabhängige Resonanzfrequenz bei etwa 1 bis 10 GHz.
- Generation-Rekombination-Rauschen (meist im Schrotrauschen miterfaßt).

Im stromlosen Fall rauscht eine Diode nur thermisch. Das Durchbruchrauschen ist Null. Das Funkelrauschen werde vernachlässigt. Dann muß das thermische Rauschen aus dem Schrotrauschen entstehen. Der Flußstrom ist hier entgegengesetzt dem Sättigungsstrom. Für eine Strom-Ersatz-Rauschquelle parallel zum *pn*-Übergang gilt hier also gemäß (3) mit Faktor 2

$$i_{\text{R}}^2 = 4e_0 I_{\text{S}} B \,. \tag{15}$$

Wenn die Diode gesperrt ist, tritt nur das halbe Rauschen auf, da nur I_{S} wirkt. In Flußrichtung ist $I_{\text{F}} > I_{\text{S}}$ und folglich

$$i_{\text{R}}^2 = 2e_0 I_{\text{F}} B \,. \tag{16}$$

Weitere Details sind in [M 20] und [P 9] enthalten.

Das Rauschstromquadrat als Funktion der Frequenz zeigt beispielhaft Bild 5.7.2a).

5.7.1.5. FET

Den prinzipiellen Frequenzgang des Rauschens von FET zeigt Bild 5.7.2b). Im *mittleren Bereich* überwiegt das *thermische Rauschen* des Kanalwiderstandes. Bei den *tiefen Frequenzen* kommt ein *Funkelrauschen* und bei den *hohen* ein *influenziertes Gaterauschen* hinzu. Das Funkelrauschen hat wieder vielfältige Ursachen, unter anderem Generation und Rekombination von Trägern, beim MOSFET Oberflächeneffekte. Das influenzierte Gaterauschen entsteht durch Rückwirkung des thermischen Rauschens über die MILLER-Kapazität zum Gate. Es kann ganz analog auch als Verstärkungsabfall des FET bei hohen Frequenzen interpretiert werden. In erster Linie ist also das Rauschen des differentiellen Kanalwiderstandes r_{DS} entscheidend. Für die entsprechende Spannungs-Ersatz-Rauschquelle gilt

$$u_{\text{R}}^2 = 4kTBr_{\text{DS}} \,, \tag{17}$$

und so ergibt sich in erster Näherung die Rauschzahl zu

$$F = 1 + (r_{\text{DS}}/R_{\text{G}}) \,. \tag{18}$$

Beim FET kommt es also vor allem darauf an, den Quellenwiderstand sehr groß zu machen, wobei aber die in der Formel nicht berücksichtigten Kapazitäten Begrenzungen bewirken. Dies zeigen die Meßkurven für je einen SGFET und MOSFET in Bild 5.7.2c) und d). Der SGFET zeichnet sich durch sehr kleine Rauschzahlen im NF-Gebiet und der MOSFET im HF-Gebiet aus. Das Funkelrauschen ist beim MOSFET meist erheblich größer als beim Sperrschicht-FET.

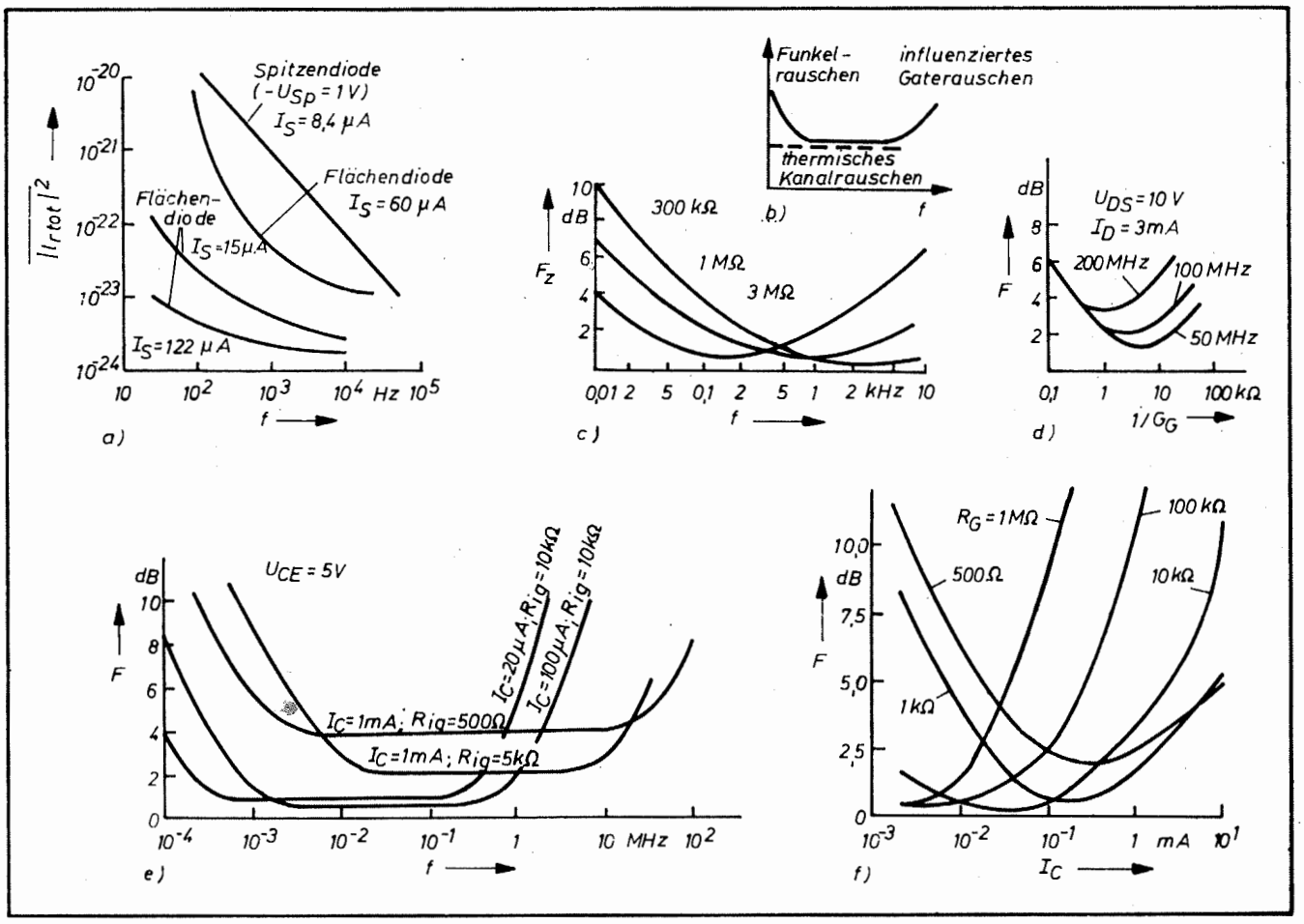


Bild 5.7.2. Daten zum Rauschen von Bauelementen

a) Beispielwerte von Dioden-Rauschen
b) Frequenzgang des Rauschens von FET
c) Beispielwerte von einem Sperrschicht-FET,
d) von einem MOSFET
e) Frequenzgang bei einem bipolaren Transistor (Silicium-Planar),
f) Optimierung über Kollektorstrom und Generatorwiderstand

Tabelle 5.7.1. Richtwerte für Rauscheigenschaften von FET

	MOSFET		SGFET	
	NF	HF	NF	HF (Gate-schaltung)
Optimaler Generatorwiderstand in Ω	$2 \cdot 10^7$ ... $2 \cdot 10^8$	500 ... 1000	10^5 ... 10^7	200 ... 300
Rauschzahl F in dB	6 ... 15	2 ... 4	1,5 ... 3	1 ... 4

Immer häufiger werden in Tabellen die frequenzbezogenen Spannungsrauschwerte angegeben. Sie betragen etwa 5 bis 10 nV/$\sqrt{\text{Hz}}$ (SGFET) und 0,1 bis 1 µV/$\sqrt{\text{Hz}}$ (MOSFET). Einige Richtwerte für die Rauschoptimierung bei FET enthält Tab. 5.7.1.

Weitere Details zum Rauschen von FET enthalten u. a. [M 20] und [P 10].

5.7.1.6. Bipolarer Transistor

Das Rauschen bei bipolaren Transistoren zeigt den gleichen prinzipiellen Verlauf wie bei FET (Bild 5.7.2e)). Das Ansteigen bei den oberen Frequenzen ist wieder durch

den Verstärkungsabfall des Transistors und das bei unteren Frequenzen durch den Funkeleffekt bedingt. Als *Rauschursachen* gelten

- thermisches Rauschen des Bahnwiderstandes R_B
- Schrotrauschen der drei Ströme
- Stromverteilungsrauschen
- Funkelrauschen.

Das *Funkelrauschen* des bipolaren Transistors läßt sich weitgehend auf den *Mitlaufeffekt* zurückführen. Die Basis-Emitter-Spannung eines Si-Transistors besitzt einen negativen Temperaturkoeffizienten von ca. -2 mV/K. Danach können selbst kleinste Temperaturänderungen in der Emitter-Basis-Sperrschicht Störspannungen erzeugen (10^{-4} K bewirken bereits 0,2 μV). Solche Temperaturänderungen entstehen bereits durch die Rauschströme selbst (Mitlaufeffekt). Sie können sich nur relativ langsam ausgleichen ($1/f$). Daraus wird verständlich, daß die Größe des Sperrschichtvolumens für das Funkelrauschen wesentlich ist. Das Funkelrauschen kann so durch eine Strom-Ersatz-Rauschquelle über den Emitterstrom I_E und einen Funkelstrom I_F, der experimentell etwa im Bereich von $3 \cdot 10^{-19}$ A bis $6 \cdot 10^{-11}$ A liegt, berechnet werden:

$$i_{Rf}^2 = I_F I_E B/f \,. \tag{19}$$

Das *Stromverteilungsrauschen* ist wegen der hohen Stromverstärkung sehr klein. Für das Schrotrauschen des Basisstromes gilt

$$i_{RS}^2 = 2e_0 B T_B \,. \tag{20}$$

Für das *thermische Rauschen* des Bahnwiderstandes und das Kollektor- sowie Emitterrauschen sei die Temperaturspannung

$$u_T = kT/e_0 \tag{21}$$

mit etwa 25 mV bei Zimmertemperatur eingeführt. Dann gilt

$$F = 1 + \frac{R_B}{R_G} + \frac{u_T}{2R_G I_E} + K_1 \frac{R_G I_E}{2u_T} \,. \tag{22}$$

Darin ist zur Vereinfachung

$$K_1 = \frac{1}{\beta^+} + \frac{I_f}{2e_0 f} \tag{23}$$

mit dem Gleichstromverstärkungsfaktor β^+ zur Umrechnung des Basisstromes von Gl. (20) in den Emitterstrom eingeführt.

Mit (22) läßt sich nun die Rauschanpassung bez. des optimalen Emitterstromes oder Generatorwiderstandes berechnen. Aus $\mathrm{d}F/\mathrm{d}I_E = 0$ folgt

$$I_{E,\,opt} = \frac{u_T}{R_G \sqrt{K_1}} \tag{24}$$

und die minimale Rauschzahl

$$F_{min} = 1 + \frac{R_B}{R_G} + \sqrt{K_1} \,. \tag{25}$$

Dieses Ergebnis entspricht den experimentellen Werten von Bild 5.7.2f).

Es sei erwähnt, daß sich für die Datenblätter immer mehr die Angaben über Strom- und Spannungsersatzquellen anstelle der Rauschzahlen durchsetzen.

5.7.2. Restströme

Restströme sollen hierbei in zweifacher Hinsicht verstanden werden. Eine Diode oder ein Transistor kann als Schalter verwendet werden. Dann sollte das Bauelement zwei extreme Zustände einnehmen können: Widerstand Null und unendlich. Dies ist beides praktisch nicht möglich (vgl. Bild 5.4.3e) bei der Diode). Hier interessiert vor allem der möglichst große Widerstand. Er kann auch durch einen *Rest-* (*bzw. Leck-*) *Strom* beschrieben werden. Der zweite Fall unterscheidet sich eigentlich nur in der Interpretation von der obigen Auffassung. Hier wird der Strom als funktionelle Größe des Halbleiterbauelements verstanden, und dann ist der Begriff *Sperrstrom* gebräuchlicher. Die Sperr- bzw. Restströme bei der Diode sind in den Abschnitten 3.2. und 5.4. behandelt.

Auf das Gate des SGFET sind diese Ergebnisse unmittelbar übertragbar. Bei allen FET-Typen treten darüber hinaus noch Source-Drain-Restströme auf, wenn diese Strecke als Schalter verwendet wird. Hierauf wird in Abschnitt 5.7.5. eingegangen.

Der bipolare Transistor ist bezüglich seiner Sperrströme komplexer. Daher ist er im folgenden fast ausschließlich behandelt.

Schließlich ist es bei allen Rest- und Sperrströmen notwendig, ihre meist beachtliche Temperaturabhängigkeit zu berücksichtigen. Auch darauf wird hier eingegangen. Diese Ergebnisse sind dann leicht auf Dioden und FET übertragbar.

5.7.2.1. Restströme des bipolaren Transistors

Im Abschnitt 3.4.3. wurden bereits die vier Betriebszustände des bipolaren Transistors behandelt. Dies erfolge hier nun aus der Sicht der Restströme für die Emitterschaltung. Bild 5.7.3a) zeigt die Abhängigkeit des *Kollektorstromes in der Umgebung des Nullpunktes*. Sowohl für die offene Basis, $I_B = 0$, als auch für die kurzgeschlossene Basis, $U_{BE} = 0$, ist das Verhalten sofort verständlich. Es fließt ein kleiner Kollektorstrom I_{CEO} bzw. I_{CES} (O = offen; S = geschlossen). Er erreicht bei einer nicht sehr großen Spannung U_{CE} seinen konstanten Sättigungswert. Bei einem positiven Basisstrom nimmt für größere Kollektorspannungen der Kollektorstrom in der z.B. aus Bild 5.6.3a), 5.6.4a) oder 5.6.6a) bekannten Weise zu. Dies gilt jedoch genaugenommen nur für den im Bild 5.7.3a) als normal aktiv bezeichneten Bereich. Zu jedem Basisstrom I_B gehört eine Spannung U_{BE}, und sofern U_{CE} kleiner als diese Spannung, wird der Kollektorstrom abnehmen und kann sogar negative Werte erreichen. Deshalb ist die Grenze des normal aktiven Bereichs dann gegeben, wenn U_{CB} Null wird. Der Kollektorstrom Null wird bei einer Spannung $U_{CE}(0)$ angenommen. Dieser Wert ist meist kleiner als 100 mV und fällt daher in den oben genannten

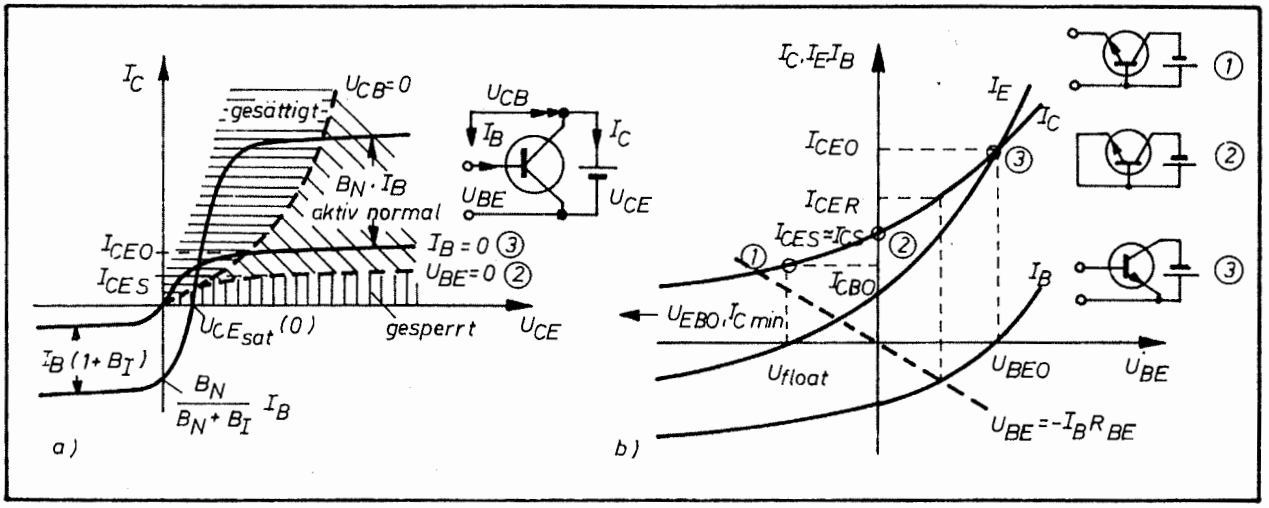
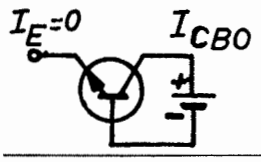
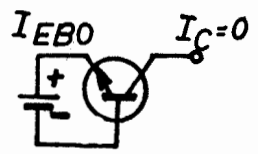
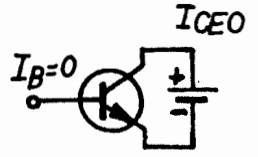
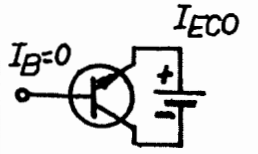
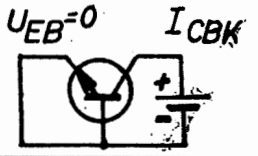


Bild 5.7.3. a) Kennlinien, b) Verlauf der drei Teilströme des bipolaren Transistors in der Umgebung des Nullpunktes. B_N und B_I sind die Gleichstromverstärkungsfaktoren im Normal- und Inversbetrieb bei der Emitterschaltung (vgl. Bild 3.4.3. und Gl. (5.6.1)).

Tabelle 5.7.2. Definition von Restströmen

Bezeichnung	Schaltung	Wert	Spannung
Kollektorreststrom I_{CEO}		$I_{CBO} = -I_{CS}\,(1-A_N A_I)$	$U_{EB} = U_T \ln\,(1-A_N)$
Emitterreststrom I_{EBO}		$I_{EBO} = -I_{ES}\,(1-A_N A_I)$	$U_{CB} = U_T \ln\,(1-A_I)$
Kollektor-Emitterreststrom I_{CEO}		$I_{CEO} = -I_{CS}\,\frac{1-A_N A_I}{1-A_N} = \frac{k_{CBO}}{1-A_N}$	$U_{EB} = U_T \ln\left(1-\frac{I_{CS}\,(1-A_I}{I_{ES}\,(1-A_N}\right.$
Kollektor-Emitterreststrom I_{ECO}		$I_{ECO} = -I_{ES}\,\frac{1-A_N A_I}{1-A_I} = \frac{I_{EBO}}{1-A_I}$	$U_{CB} = U_T \ln\left(1-\frac{I_{ES}\,(1-A_N}{I_{CS}\,(1-A_I}\right.$
Kollektorreststrom I_{CBK}		$I_{CBK} = -I_{CS}$	

üblichen Kennliniendarstellungen nicht auf. Die Kennlinie $U_{CB} = 0$ entspricht in etwa der Durchlaßkennlinie der Emitterdiode.

Allgemeiner sind die verschiedenen Restströme der drei Elektroden als Funktion der Spannung U_{BE} bei hinreichend großer Kollektorspannung im Bild 5.7.3b) gezeigt. Die drei typischen Fälle — von denen zwei auch in Teilbild a) enthalten sind — sind als Schaltungen in dieses Bild eingetragen. Zwischen den Punkten ② und ③ kann der jeweilige Arbeitspunkt über die Widerstandsgerade des Basis-Emitter-Widerstandes für U_{BE} als Schnittpunkt mit I_B konstruiert werden. Tab. 5.7.2 faßt schließlich die entsprechenden Werte zusammen und zeigt die entsprechenden Rechenwerte aufgrund des EBERS-MOLL-Modells von Bild 3.4.3. Weitere Details enthalten u. a. [P 10 und S 27]. Einen Überblick zu den Größenordnungen der Restströme gibt Tab. 5.7.3. Daran wird deutlich, daß die *Restströme bei* Ge-*Transistoren oft sehr störend* sind. Ergänzt sei, daß die Stromverstärkung bei integrierten Si-Transistoren ihr Maximum bei 5 bis 30 A/mm² besitzt.

Tabelle 5.7.3. Größenordnungen von Restströmen bei etwa 300 K

Wert	Germanium		Silicium		
	Vorstufen	Leistungsst.	Vorstufen	Leistungsst.	integriert
I_{CBO}	1 ... 50 µA	0,1 ... 5 mA	0,1 ... 10 nA	0,1 ... 50 µA	0,1 µA/mm²
I_{CEO}	0,1 ... 1 mA	10 ... 100 mA	1 ... 100 nA	0,1 ... 10 mA	0,1 ... 1 µA/mm

5.7.2.2. Temperatureinflüsse

Der Temperaturgang der Sperr- und Restströme ist im wesentlichen durch die Diodengleichung (3.2.5) gegeben. Sind die Ströme bei der Umgebungstemperatur T_U gegeben, so lassen sich ihre Änderungen bezüglich der Temperaturänderung ΔT relativ einfach berechnen:

$$I_{CBO}(\Delta T) = I_{CBO}(T_U) \exp C_B \Delta T\,, \tag{26}$$

$$I_{CEO}(\Delta T) = I_{CEO}(T_U) \exp C_E \Delta T\,, \tag{27}$$

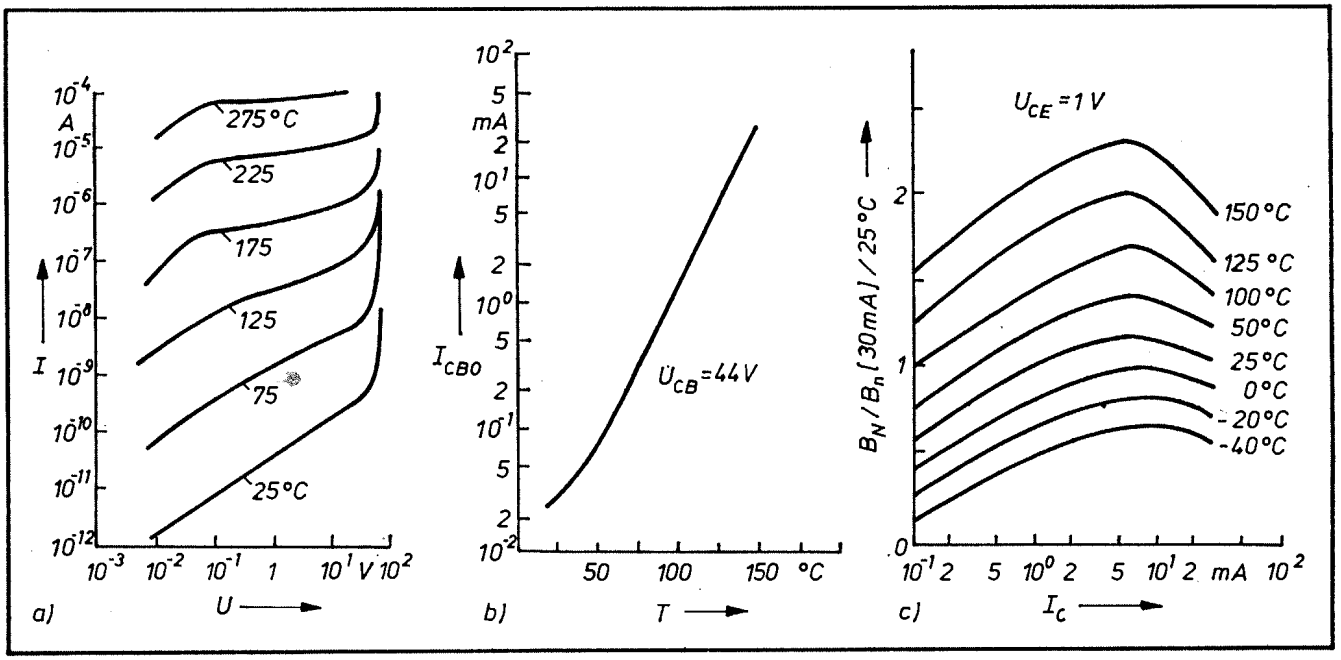


Bild 5.7.4. Temperatureinfluß auf Bauelemente

a) Sperrstrom einer guten Si-Planar-Diode bei verschiedenen Temperaturen
b) Kollektorstrom I_{CBO} des Si-Transistors SF 137
c) Dito, Änderung der statischen Stromverstärkung in Emitterschaltung bei verschiedenen Kollektorströmen

wobei für die Konstanten näherungsweise pro Grad gilt:

	Si	Ge
C_B	0,08	0,09
C_E	0,15	0,15

Die Auswirkungen dieses Einflusses zeigt Bild 5.7.4. Die Restströme ändern sich also um Zehnerpotenzen. Doch auch die Stromverstärkung ändert sich mit der Temperatur, wie Bild 5.7.4c) zeigt, beachtlich. Auf den Einfluß der Temperatur auf die Basis-Emitter-Spannung wurde bereits im Abschnitt 5.7.1.6. eingegangen. Hierfür gilt näherungsweise bei Si-Transistoren

$$\boxed{\frac{dU_{BE}}{dT} \approx -(2 \ldots 3)\ \text{mV/K}}\,. \tag{28}$$

5.7.3. Minimalabmessungen

Um 1960 lag die minimale Fläche eines integrierten Transistors zwischen 10^4 und $10^5\ \mu m^2$, heute liegt sie um $100\ \mu m^2$. Voraussetzungen waren hierfür vor allem die verbesserten Technologien. Lagen um 1970 die beherrschbaren Strukturmaße um $10\ \mu m$, so werden heute $2\ \mu m$ in vielen

Fällen genutzt, und wie im Kapitel 4. gezeigt wurde, geht diese Entwicklung weiter. Daher erhebt sich die Frage nach den kleinstmöglichen Abmessungen von Bauelementen auf dem Chip. Hierzu gibt es verschiedene Abschätzungen, z. B. kann die minimale Emitterfläche gemäß

$$A_E = \frac{I_E R_B W_B^2}{U_T} \tag{29}$$

berechnet werden. Darin bedeutet W_B die Basisweite, sie dürfte infolge der minimal realisierbaren Epitaxieschichten im Bereich von 1 bis 0,1 μm liegen. Weiter ist R_B der Basiswiderstand, I_E der maximale Emitterstrom und U_T die Temperaturspannung gemäß (21). Aus einer größeren Anzahl ähnlicher Formeln und Plausibilitätsbetrachtungen folgten die Ergebnisse der WALLMARK-Studie von 1975. Einige Daten faßt Tab. 5.7.4 zusammen, während Bild 5.7.5 zwei wichtige Abhängig-

Tabelle 5.7.4. Ausgewählte Daten zum minimalen Transistor

Randunschärfe	10 nm
Gesamttoleranz	1 μm
Bauelementfläche	$10 \times 10\ \mu m^2$
Spannungshub (digital)	$10 U_T \approx 0{,}25$ V
Kleinste Stromdichte, 100 MHz	$2 \cdot 10^3$ A/cm²
Max. Stromdichte (Leistung, Leiterbahn)	10^5 A cm²
Min. Leistung je Transistor	$2 \cdot 10^{-11}$ W

keiten zeigt. Die weißen Flächen stellen die realisierbaren Parameter dar und betreffen etwa 10^5 Transistoren auf einem Chip (bei einer mittleren, fehlerfreien Betriebszeit von 1000 Stunden). Inzwischen sind auch Studien veröffentlicht worden, die für 1990 über eine Million Bauelemente je Chip prognostizieren. Solche Schaltungen dürften aber notwendig mit Redundanz arbeiten.

Bei MOS-Schaltungen wird häufig eine Konstante K gewählt, welche z. B. Frequenz, Störabstand und Pegel enthält. Für die Breite b, die Länge l des Kanals, die Ladungsträgerbeweglichkeit μ, Oxiddicke d und Dielektrizitätskonstante ε muß dann gelten

$$K = \frac{\mu \varepsilon b}{dl}. \tag{30}$$

Andere Betrachtungen verwenden einen Skalierungsfaktor und weisen aus, wie sich Kenndaten gemäß diesem Faktor mitändern müssen.

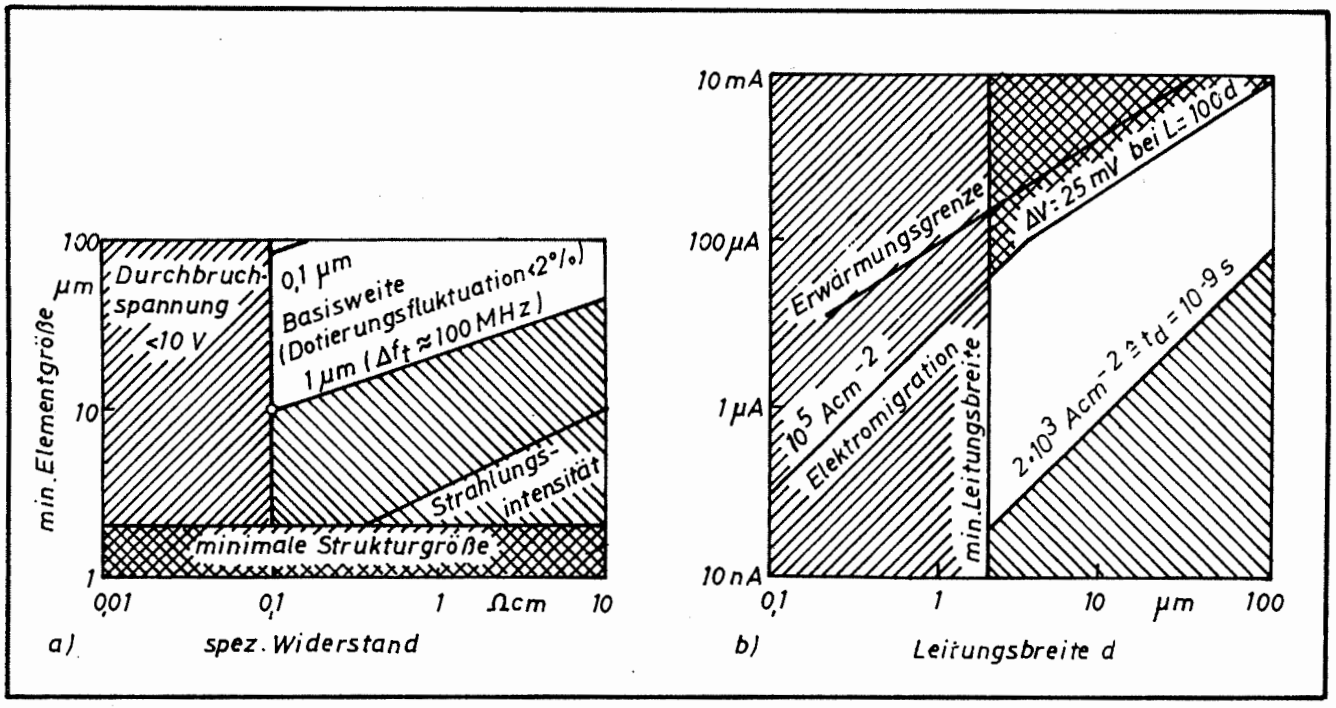


Bild 5.7.5. Zu den Minimalgrenzen von bipolaren Transistoren

a) Einflüsse und Grenzen der Flächenabmessungen für 10^5 Transistoren und MTBF 10^{-3}/h

b) Leiterbreite und Stromdichte

5.7.4. Beseitigung äußerer Störungen

Äußere *Störungen treten in vielfältiger* Form auf. Beispiele sind:

- atmosphärische Störungen (Gewitter usw.),
- HF-Störungen von Sendern aller Art (Rundfunk, Fernsehen, Radar, HF-Öfen, Medizintechnik usw.),
- impulsförmige Störungen (von Schaltern, Zündkerzen, Elektrobahnen, Schweißen usw.),
- Netzstörungen (50 Hz und Oberwellen),
- Störungen aus dem Weltraum.

Die Störungen sind hier bewußt auf elektrische, magnetische und elektromagnetische eingeschränkt. Einen Überblick zu den Störungen gibt Bild 5.7.6. Derartige Störungen werden in der Regel durch drei Maßnahmen *von den Eingängen der Verstärker ferngehalten*:

Erdung, Schirmung und Symmetrie.

Sie sind jedoch nur effektiv zu gestalten, wenn sie in ihrer Wechselwirkung erkannt und gemeinsam, miteinander ausgewogen eingesetzt werden.

5.7.4.1. Erdung

Vor einiger Zeit war es üblich, elektronische Baugruppen auf ein leitendes Chassis aufzubauen. Dieses Chassis war zugleich Bezugspotential für alle übrigen Spannungen. Es wurde auch mit Masse oder Erde bezeichnet. Die Bezeichnung Erde geht darauf zurück, daß es zuweilen mit der Erdleitung verbunden wurde. Ein solches Chassis diente häufig zugleich zur Abschirmung von Störungen. Zu seiner vollen Wirkung kommt es jedoch erst dann, wenn es keinen Gleich- oder Wechselstrom führt. Hierbei kann nämlich zwischen zwei Punkten ein Spannungsabfall entstehen, der zu einer ungewollten Verkopplung in der Schaltung führt. Die Induktivität zwischen zwei solchen Punkten kann sich noch stärker als der Widerstand auswirken. Über sie kann auch eine Störspannung magnetisch eingekoppelt werden. Um derartige ungewollte Massekopplungen zu vermeiden, ist es am günstigsten, für jede Verstärkerstufe einen isolierten Bezugspunkt festzulegen. Dies gilt natürlich auch für Schaltungen, die nicht auf ein Chassis, sondern z.B. als gedruckte Schaltung aufgebaut sind. Die einzelnen Bezugspunkte lassen sich dann so zusammenfassen, daß durch die auftretenden Gleich- und Wechselströme keine Verkopplung erfolgt. Das Chassis bzw. die Abschirmung ist am wirkungsvollsten mit dem Bezugspunkt der ersten Stufe zu verbinden. Das Beispiel einer entsprechend aufgebauten Schaltung zeigt Bild 5.7.7.

5.7.4.2. Schirmung

Schirmung und Symmetrie sind im allgemeinen noch schwerer voneinander zu trennen als die Erdung von ihnen. Sie ergänzen einander und sind beide sowohl für magnetische als auch für elektrische Felder anwendbar. Die zuweilen anzutreffende Meinung, daß elektrische Felder abzuschirmen und magnetische Felder zu symmetrieren seien, ist nur bedingt richtig. Sie wird meist daraus gefolgert, daß bei der elektrischen Leitfähigkeit zwischen Leiter und Isolator ein größerer Unterschied besteht als bei der magnetischen Permeabilität zwischen Vakuum und magnetischen Materialien.

Bei der Schirmung ist es im Prinzip gleichgültig, ob der oder die Störer einerseits bzw. die Signalquellen und der Verstärker andererseits abgeschirmt werden. Ent-

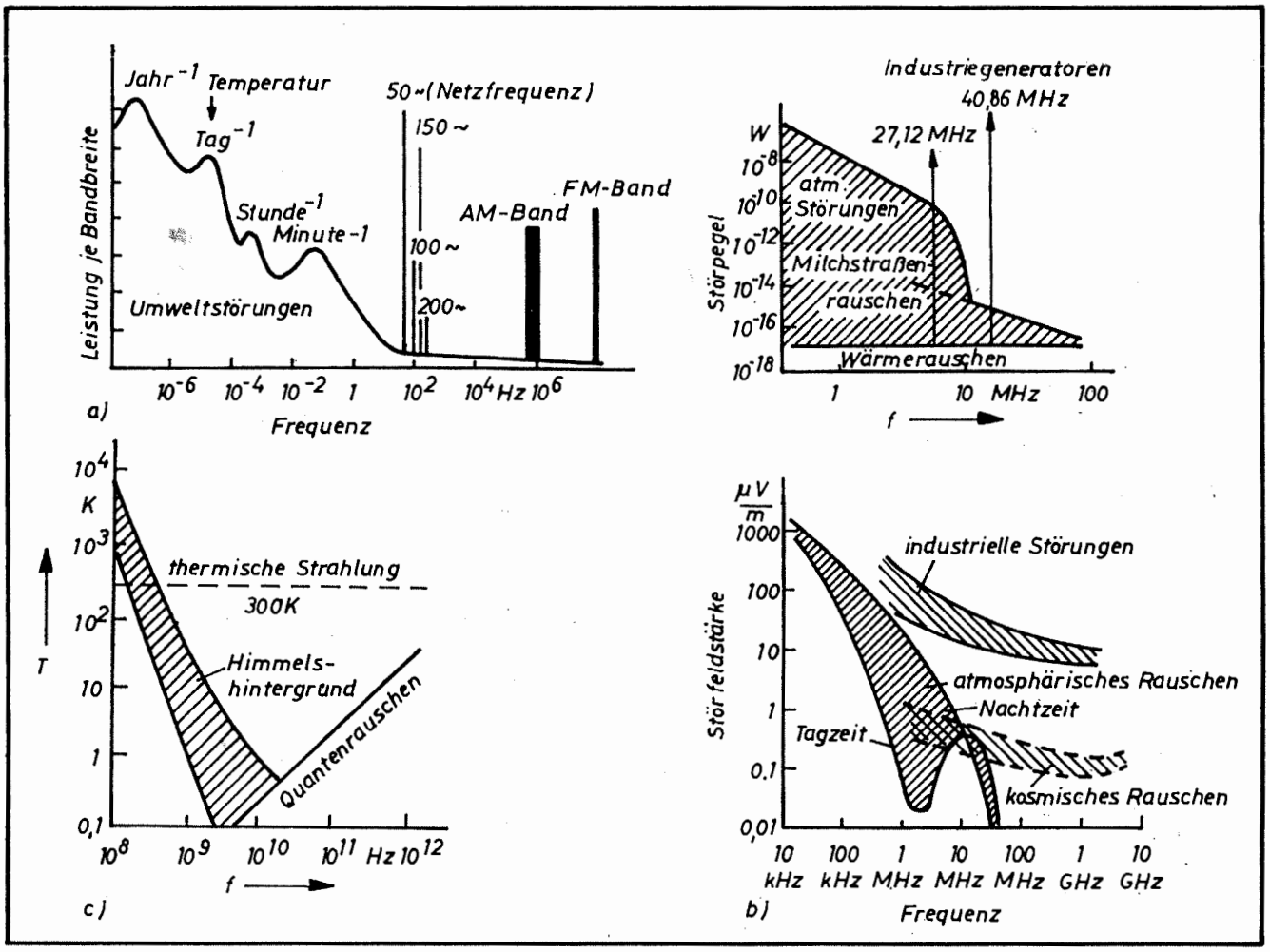


Bild 5.7.6. Drei Ausschnitte aus dem Störspektrum

a) Bei tiefen und tiefsten Frequenzen. Hierin sind auch Temperaturschwankungen usw. enthalten. Es sind hier kaum absolute Größen angebbar, deshalb fehlt der Ordinatenmaßstab

b) Hervorhebung jenes Teils, in dem Rauschstörungen aus dem Weltall unter sonst nahezu idealen Bedingungen auftreten

c) Übergang zwischen der 3-K-Reststrahlung des Weltalls und dem dann folgenden Teil des Quantenrauschens. Die Rauschtemperatur ist über $h\nu$ gemäß $T_Q/\mathrm{K} \approx 0{,}07 f/\mathrm{GHz}$ berechnet

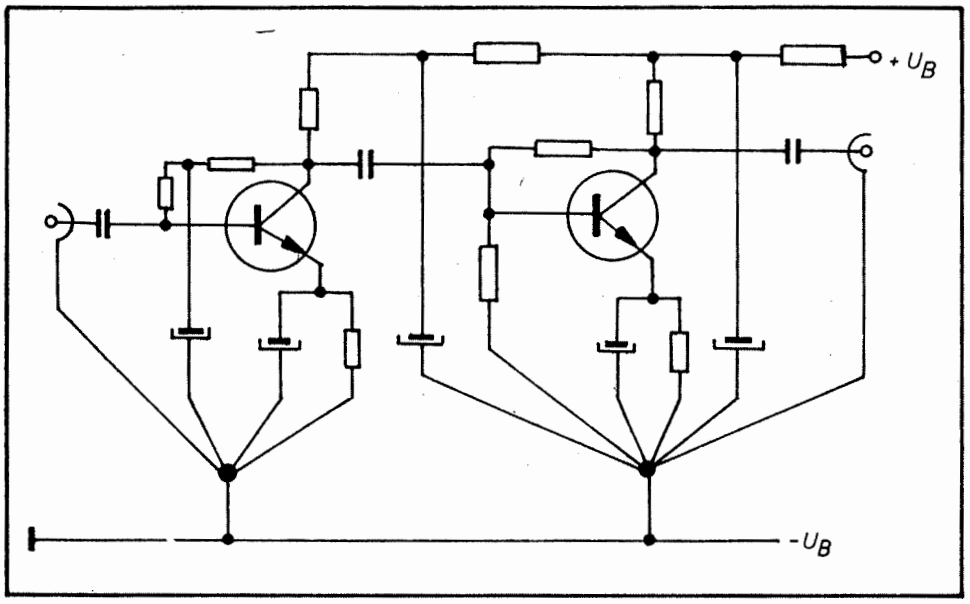


Bild 5.7.7. Beispiel der Leitungsführung bei einer zweistufigen Verstärkung. Es sind Erdschleifen zu vermeiden, und durch richtige Zusammenfassung von Erdpunkten muß erreicht werden, daß die eigentliche Masseleitung möglichst frei von Störungen gehalten wird

scheiden hierüber kann nur der notwendige Materialaufwand sowie das Volumen und die Lage von Störern bzw. Gestörten.

Bei der Schirmung ist es wichtig, daß der Störer bzw. das gestörte Gerät vollständig mit dem elektrisch bzw. magnetisch leitenden Material umgeben werden. Selbst durch relativ kleine Löcher kann oft ein ganz erheblicher Störanteil hereinbzw. hinausdringen und so den Aufwand der Abschirmung fragwürdig machen. Die Schirmwirkung nimmt in großer Näherung exponentiell mit der Dicke des Materials d und dem Verhältnis k aus den elektrischen bzw. magnetischen Leitfähigkeiten vom Material und der Umgebung zu. Zwischen der Intensität I_s hinter und I_0 vor dem Schirm gilt näherungsweise

$$\frac{I_s}{I_0} \sim e^{-dk} . \tag{31}$$

Da aber jeder Schirm zusätzlich die Felder durch den unstetigen Übergang zur Umgebung schwächt, ist es meist wirksamer, eine Mehrfachabschirmung bei gleichem Materialaufwand zu verwenden. Die Verbindung der einzelnen Schirme darf dabei jeweils nur in einem Punkt erfolgen. Andernfalls werden Ströme in der Abschirmung auf den nächstfolgenden Abschirmmantel übertragen, so wie es bei der Erdung beschrieben wurde. Die exponentielle Wirkung des Schirmes führt zwar zur schnellen Abnahme des Störfeldes, macht sie jedoch prinzipiell nie zu Null.

5.7.4.3. Symmetrierung

Eine Symmetrierung kann ebenfalls beim Störer und beim Gestörten erfolgen. Typisch hierfür ist es, das störende Feld in zwei gleich große, aber entgegengesetzte Felder zu zerlegen. Für einen Störer zeigt dies symbolisch Bild 5.7.8a). Die beiden Kondensatoren bewirken, daß die Störquelle in der Mitte geerdet ist und so zwei gegenphasige Spannungen auf den Verstärker einwirken und sich dort aufheben. Eine ähnliche Wirkung kann aber auch erreicht werden, wenn ein zum Störsender gegenphasig schwingender Hilfssender mit passender Amplitude aufgestellt wird. Mit der Symmetrie kann daher prinzipiell ein Störeinfluß vollständig unterdrückt werden. Der Nullabgleich ist hierbei jedoch nach Amplitude und Phase äußerst kritisch. Bereits kleine Änderungen in der Umgebung können zu einer vollständig veränderten Situation führen. Um dies zu vermeiden, ist es sinnvoll, neben der Symmetrie die Abschirmung zu verwenden.

Bei einem gestörten Verstärker läßt sich die Symmetrie durch einen Gegentakteingang erzwingen. Bild 5.7.8b) zeigt dies am Beispiel eines Operationsverstärkers. Derartige Zusammenhänge werden genauer in Abschnitt 7.3.3. und 7.5. behandelt.

Als weiteres Beispiel für die Symmetrie sei ein abgeschirmtes Kabel betrachtet (Bild 5.7.8c)). Eine andere Form der Symmetrierung ist die verdrillte Leitung. Hierbei wird mit großer Sicherheit erreicht, daß beide Adern ein gegenphasiges Feld so ausstrahlen, daß es in ganz kurzer Entfernung bereits verschwindet. Beim Verstärker mit Gegentakteingang nehmen dagegen beide Adern mit großer Sicherheit ein gleichphasiges Feld auf, das dann im Verstärker unterdrückt wird.

5.7.4.4. Netzstörungen

Das elektrische Starkstromnetz ist praktisch überall vorhanden. Damit gelangt auch das entsprechende 50-Hz-Signal sowie die verschiedenen Oberwellen mehr oder weniger in alle Schaltungen. Entscheidend ist dabei die Größe der Schaltung und die Länge der Leitungen. Aus diesem Grunde sind integrierte Schaltungen generell weniger

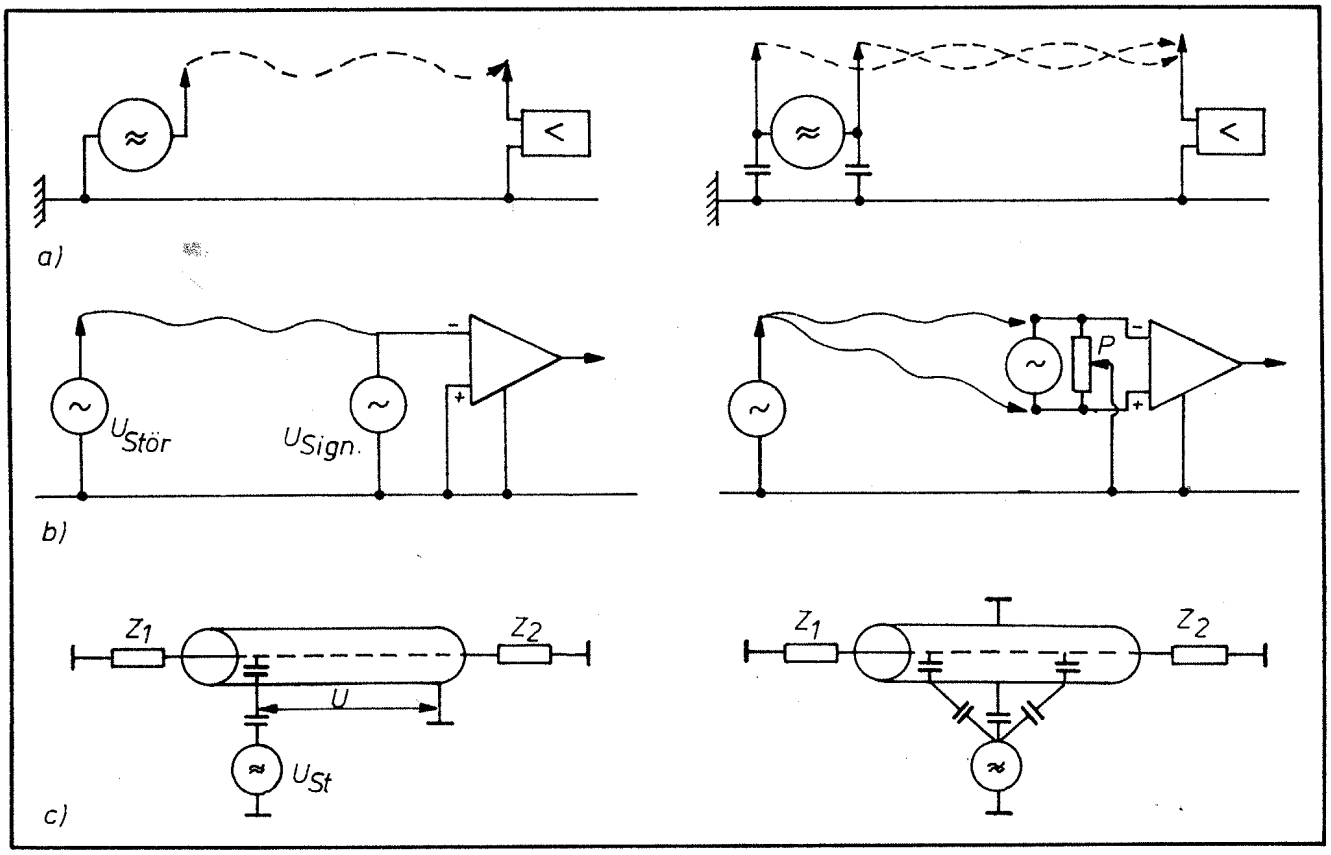


Bild 5.7.8. Zur Symmetrierung bei Störungen; jeweils links der ungünstige und rechts der symmetrierte Fall

a) Symmetrierung eines Senders durch Mittelpunkterdung infolge der beiden Kondensatoren
b) Symmetrierung mittels Gegentakteingang (Differenzverstärker)
c) Zur Schirmung und Symmetrie bei einem abgeschirmten Kabel. Links wirkt der Störer unsymmetrisch auf den Kabelmantel. Dadurch entsteht an ihm der Spannungsabfall U. Er überträgt sich auf die Widerstände Z_1 und Z_2. Ist dagegen das Kabel symmetrisch geerdet, so heben sich die beiden Teilkomponenten auf

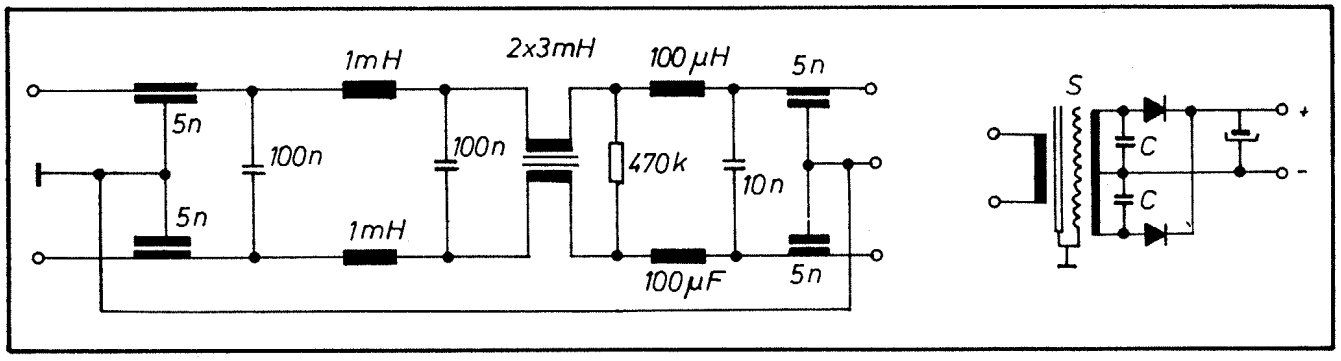


Bild 5.7.9. Zur Vermeidung von Störungen über das elektrische Netz

a) Filter mit Tiefpaßverhalten und hoher Symmetrie. Die vier Kondensatoren zu 5 nF sind Durchführungskondensatoren, denn das ganze Filter ist oft abgeschirmt aufgebaut. Die Gegeninduktivität von 2 × 3 mH soll eine zusätzliche Symmetrierung und Entkopplung erzwingen
b) Trafo mit Schutzwicklung S zwischen Primär- und Sekundärwicklung sowie Kondensatoren C zur Überbrückung der Sekundärwicklungen, um zu verhindern, daß evtl. vorhandene Hochfrequenz von den Gleichrichtern brummoduliert wird

empfindlich als die größeren, diskret aufgebauten. (Dies gilt übrigens für alle von außen einwirkenden Störungen.) Besonders betroffen sind natürlich alle Geräte, die aus dem elektrischen Netz mit Energie versorgt werden.

Das Problem der Siebung der gleichgerichteten Spannungen wurde bereits im Abschnitt 5.4. behandelt und wird noch einmal bei der Spannungsstabilisierung im Abschnitt 7.6. aufgegriffen. Es steht hier also nicht zur Diskussion. Bedeutsam sind

aber noch jene *Störungen, die mittelbar aus dem Netz ins Gerät eindringen können bzw. umgekehrt vom Gerät ins Netz.* Dies betrifft alle Hochfrequenzen und starke Impulse. Daher müssen empfindliche Schaltungen und Schaltungen, die viele Störungen erzeugen, vom Netz durch *spezielle Filter* entkoppelt werden. Im Grunde handelt es sich dabei um ein- oder mehrstufige Tiefpässe (Bild 5.7.9a)). In einfacheren Fällen genügt es zu beachten, daß die *Netzgleichrichter auch Hochfrequenz* gleichrichten bzw. mit der Netzfrequenz bei der Gleichrichtung modulieren (*Modulationsbrumm*). Diese Störungen lassen sich weitgehend durch eine Schutzwicklung auf dem Netztrafo (geerdet zwischen Primär- und Sekundärwicklungen) und/oder durch Kurzschluß der Hochfrequenz mit Kondensatoren reduzieren (Bild 5.7.9b)).

5.7.5. Obere Grenzen, Schaltverhalten

Für jedes Bauelement gibt es maximale Werte, z.B. von Strom, Spannung und Leistung. Die umgesetzte Leistung muß in jedem Fall als Wärme abgeführt werden. Hierauf wird im Abschnitt 5.7.6. eingegangen. Hier sei nur mitgeteilt, daß die Darstellung der Leistungsgrenze in zwei Arten üblich ist. Wird Strom und Spannung linear aufgetragen, so tritt die *Leistungshyperbel*

$$N = U \cdot I \tag{32}$$

als Hyperbel deutlich in Erscheinung (z.B. in Bild 5.1.3b), 5.5.1b), 5.5.3); werden dagegen Strom und Spannung *logarithmisch* aufgetragen, so geht die Hyperbel in der Darstellung in eine *Gerade* über:

$$\log N = \log U + \log I \,. \tag{33}$$

Dies ist andeutungsweise in Bild 5.5.1h) zu erkennen. Dort ist aber auch bereits zu erkennen, daß ein bestimmter Kollektorstrom und eine gegebene Kollektorspannung nicht überschritten werden dürfen. Diesen Werten werden wir uns jetzt zuwenden, um dann noch auf das Schaltverhalten einzugehen.

5.7.5.1. Spannungsgrenzen

Beim Feldeffekttransistor wurde bereits im Zusammenhang mit Bild 3.3.3b) auf die maximale Spannung eingegangen, bei welcher der Durchbruch erfolgt. In ähnlicher Weise besitzen auch die beiden Dioden des bipolaren Transistors ein Durchbruchverhalten, das dem einer Z-Diode (3.2.6.) entspricht. Da der Emitter meist relativ hoch dotiert ist, liegen die Werte der Emitterdiode bei ca. 5 bis 10 V. Der Kollektor ist dagegen niedrig dotiert, und bei Leistungstransistoren sind zusätzlich zur Erhöhung der Durchbruchspannung oft noch ein oder zwei schwach leitende Gebiete eingefügt (Bild 3.2.8, 5.6.8). Hierdurch existiert bei der Kollektordiode nur der Avalanche-Effekt mit der Trägervervielfachung

$$m = \frac{1}{1 - (U_{CB}/U_{CBO})^n} \,. \tag{34}$$

Darin bedeutet U_{CBO} die Durchbruchspannung, und der Exponent n besitzt Werte zwischen 3 und 9.

Im Transistor treten die Eigenschaften der beiden Dioden nicht einzeln auf, sondern sind in das Gesamtverhalten des Transistors integriert. Dadurch sind die Verhältnisse, ähnlich wie bei den Restströmen, komplexer. Dies wird deutlich, wenn das Durchbruchverhalten der Kollektor-Emitter-Strecke bei unterschiedlichen Basisbedingungen betrachtet wird.

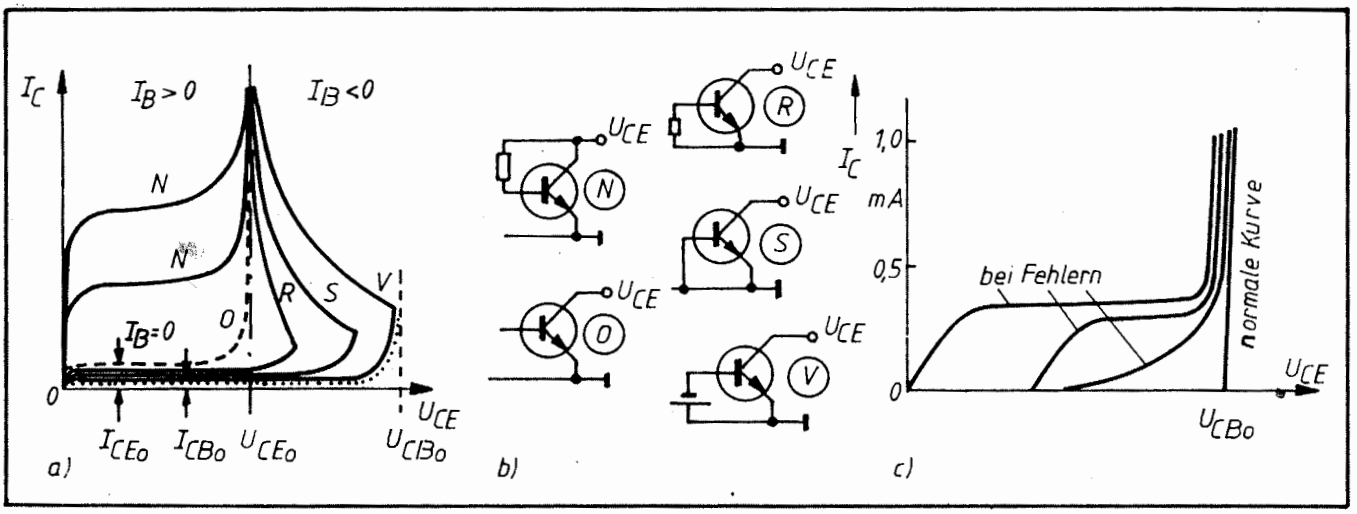


Bild 5.7.10. Sperrspannungen des bipolaren Transistors

a) Die verschiedenen Betriebsarten in Emitterschaltung
b) Die zugehörigen Durchbruchkennlinien
c) Abweichungen vom normalen Durchbruchverhalten, bez. des Falls (*O*) im Teibilld b)

Für die offene Basis ($I_B = 0$) zeigt sich der gestrichelte Verlauf (Kurve *O*) in Bild 5.7.10 a). Es wird mit steigender Kollektor-Emitter-Spannung zunächst der Reststrom I_{CEO} (vgl. Bild 5.7.3a)) angenommen, und bei der Spannung U_{CEO} erfolgt der Durchbruch analog zu Gl. (34), wobei mit dem Gleichstromverstärkungsfaktor $B \approx h_{21E}$ in Emitterschaltung gilt:

$$U_{CEO} \approx U_{CBO} \sqrt[n]{1 + B} < U_{CBO} \,. \tag{35}$$

Wird der Emitter mit positivem Strom betrieben, so befindet sich der Transistor bei Spannungen unter U_{CEO} im normal aktiven Betrieb (Kurven *N*). Hier entstehen also die normalen Ausgangskennlinien. In der Nähe von U_{CEO} erfolgt stets der entsprechende Durchbruch.

Die nächsten *drei Sonderfälle* treten auf, wenn

- die Basis-Emitter-Strecke mit einem Widerstand R_{BE} überbrückt wird (*R*),
- dieser Widerstand Null wird, also Kurzschluß der Basis-Emitter-Strecke vorliegt (*S*),
- die Basis-Emitter-Strecke negativ vorgespannt wird (*V*).

Diese drei Fälle haben einen ähnlichen Kennlinienverlauf bezüglich des Durchbruchs und zeichnen sich dadurch aus, daß ein negativer Basisstrom während des Durchbruchs wirksam wird. Dabei wird die Grenze U_{CBO} nur dann asymptotisch erreicht, wenn der gesamte Kollektorstrom über die Basis abfließt.

Normalerweise werden Transistoren nicht im Durchbruchbereich betrieben. Hierfür geben die Hersteller auch keine Daten oder Garantien. Für Sonderfälle sind die Kennlinien mit dem negativen differentiellen Widerstand jedoch von Bedeutung. Dies ist nützlich, weil die Durchbrüche innerhalb von ns stattfinden und sich auf diese Weise extrem kurze Impulse erzeugen lassen. Man spricht dann vom *Avalanche- bzw. Lawinenbetrieb des Transistors* [T 3, Bd. 2].

Es sei erwähnt, daß es bei einigen Transistoren (z.B. durch *Fertigungsmängel*) Abweichungen gegenüber dem normalen Sperrverhalten gibt. Die Kurve *O* in Bild 5.7.10 a) zeigt dann Abweichungen gemäß Bild 5.7.10 c) [S 27].

5.7.5.2. Schaltverhalten

Im Abschnitt 5.5. wurde bereits erwähnt, daß FET und damit auch Leistungs-MOSFET sehr gute Schalteigenschaften haben. Die Gründe sind: kleiner Durchlaß- und großer Sperrwiderstand sowie hohe Schaltgeschwindigkeit. Dies ist schematisch in Bild 5.7.11 angedeutet. Im Sperrbetrieb fließt nur ein sehr kleiner Reststrom, der auf der gesperrten Draindiode gegenüber dem Substrat beruht. Mit der Fläche der Senke A, der Störstellenkonzentration n_A, der Eigenleitungsdichte n_i, dem ε des Halbleitermaterials und der Speicherzeitkonstanten τ gilt

$$I_{SO} = \frac{A n_i}{\tau} \sqrt{\frac{2 e \varepsilon U_{SD}}{n_A}} \,. \tag{36}$$

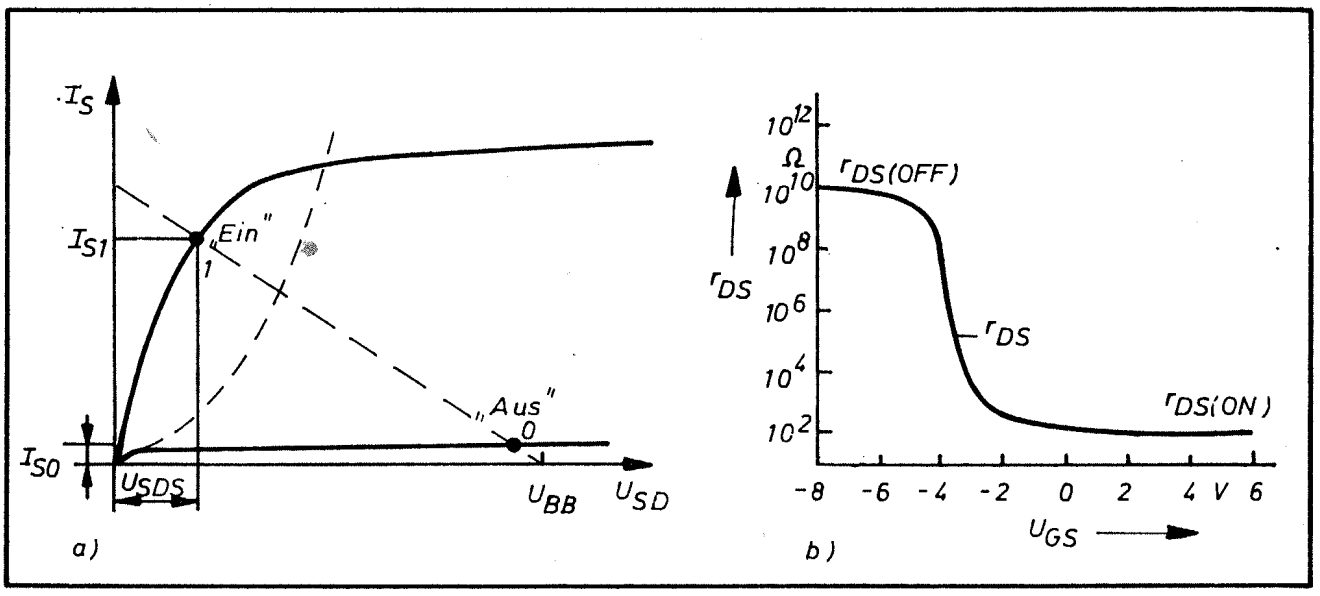


Bild 5.7.11. Schaltverhalten von FET

a) Die beiden Betriebszustände

b) Kennlinie des differentiellen Widerstandes r_{DS} bis an die Aussteuerungsgrenzen

Wie klein dieser Strom ist, wird deutlich durch den differentiellen Widerstand eines Vorstufen-n-Kanal-FET in Bild 5.7.11b). Im Gegensatz zu den Bildern 5.1.2 und 5.5.1f) ist er hier bis an die Grenzen der Aussteuerbarkeit gezeigt.

Beim bipolaren Transistor sind auch diese Verhältnisse wieder komplizierter. Neben dem hier nicht zu behandelnden Lawinenbetrieb (avalanche mode) existieren *zwei Sonderfälle*. Sie sind in Bild 5.7.12 dargestellt:

1. *Sättigungsprinzip* (saturated mode). Der eingeschaltete Zustand liegt im Sättigungsbereich und wird mit positivem Basisstrom erreicht. Der ausgeschaltete Zustand liegt je nach dem Basisstrom ($\leqq 0$) auf der entsprechenden Sperrkennlinie.
2. *Stromschalterprinzip* (current switching mode). Hier wird gemäß Bild 5.7.12c) eine Stromquelle zwischen zwei Transistoren umgeschaltet. Dadurch wird der Sättigungsbereich vermieden.

Das Sättigungsprinzip hat die Vorteile: kleine Verlustleistung, kleiner Durchlaßwiderstand, großer Lastwiderstand, großer Spannungshub und einfache Schaltung, dafür aber den Nachteil einer langen Ausschaltzeit, da im Sättigungsfall (Ein) viele Minoritätsträger im Basis- und Kollektorgebiet gespeichert sind.

Das Stromschalterprinzip vermeidet die Sättigung und ist dadurch wesentlich schneller, hat aber dafür die Nachteile des kleinen Spannungshubes, des kleinen Arbeitswiderstandes und der komplizierten Schaltung.

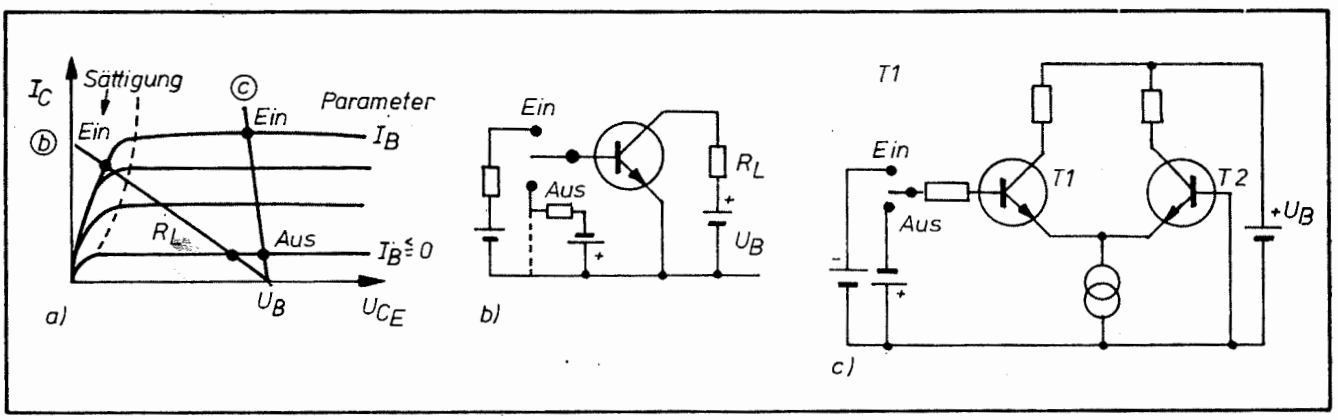


Bild 5.7.12. Schalterprinzipien von bipolaren Transistoren

a) Kennlinienfeld mit den beiden Prinzipien
b) Sättigungsbetrieb (saturated mode)
c) Stromschalterbetrieb (current mode)

Im Kapitel Digitaltechnik werden noch Lösungen zwischen diesen beiden Extremen gezeigt. Ein wichtiges Beispiel ist der SCHOTTKY-Transistor (8.2.4.).

5.7.5.3. Schalten von Blindlast

Im Zusammenhang mit Bild 5.5.6 wurde bereits gezeigt, wie sich die Arbeitskennlinie bei Blindlast verändert. Dieser Effekt tritt verstärkt beim Schalterbetrieb auf. Hierfür gibt es zwei Gründe:

- Die Kennlinien werden auch im nichtlinearen Bereich ausgesteuert.
- Die Übergänge erfolgen zwischen zwei Zuständen in möglichst kurzer Zeit.

Sind bereits in Bild 5.5.6 *Strom- und Spannungsüberhöhungen* gegenüber dem statischen Betrieb zu erkennen, so werden diese beim Schalterbetrieb wesentlich größer sein. Die genannten Effekte zeigt dementsprechend Bild 5.7.13 deutlich. Bei *kapazitiver Last wird eine Stromüberhöhung beim Einschalten* wirksam, denn es muß zusätzlich die Ladung des Kondensators verändert werden. Bei *induktiver Last* tritt dagegen *Spannungsüberhöhung beim Ausschalten* auf. Es muß die in der Spule gespeicherte Energie abgebaut werden. Auf die Beseitigung dieser Spannungsspitze mittels Diode wurde bereits im Zusammenhang mit Bild 5.4.10 hingewiesen.

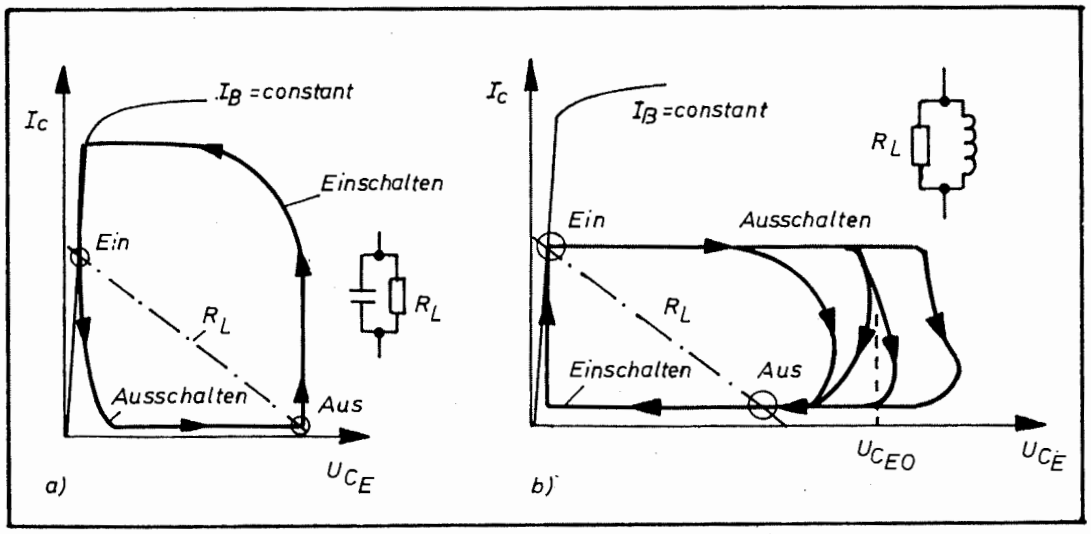


Bild 5.7.13. Schalten von Blindlast

a) bei zusätzlicher Kapazität,
b) bei zusätzlicher Induktivität mit unterschiedlicher Energie (bezüglich u_{CEO} vgl. Bild 5.7.10b))

5.7.5.4. Zweiter Durchbruch

Der zweite Durchbruch ist eine typische Erscheinung für den bipolaren Transistor. Dabei sind gemäß Bild 5.7.14a) zwei von der Ursache her verschiedene, in der Auswirkung aber nur schwer zu unterscheidende zweite Durchbrüche möglich:

A. im aktiven Bereich (forward second breakdown)
B. im Sperrbereich (reverse second breakdown).

Im aktiven Betrieb ist die Stromverteilung nie ganz gleichmäßig. Neben den Effekten des Emitterrandes (Abschnitt 5.6.6.4.) treten auch stets kleine Unregelmäßigkeiten bei der Herstellung, z. B. Kristall-, Masken-, Justier- oder Fotofehler, auf. Beim Auflöten des Si-Blättchens auf die Unterlage können Lunker aufgetreten sein. Hierdurch bilden sich lokal höhere Ströme und damit im Kollektorsperrraum erhöhte Temperaturen

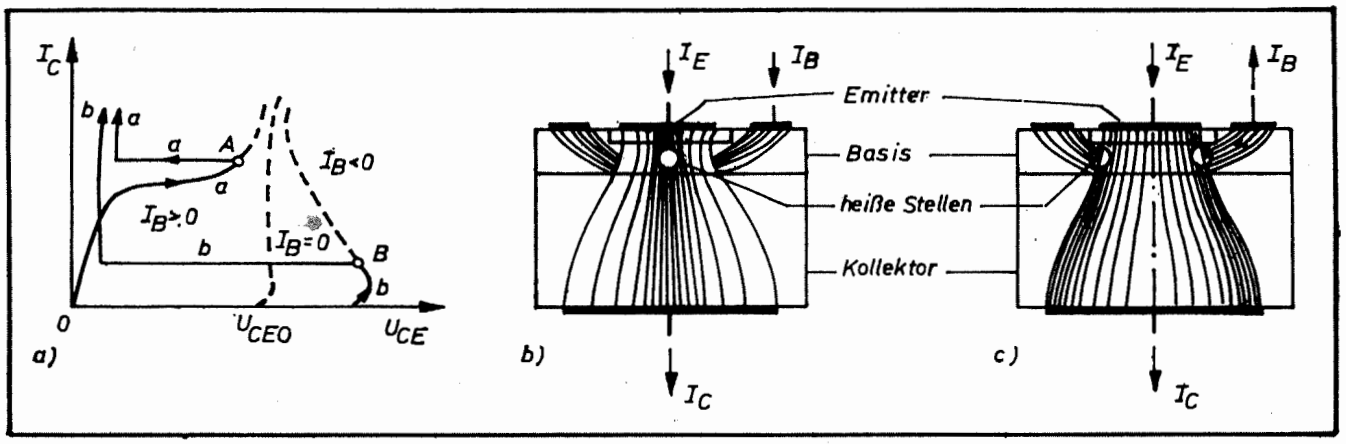


Bild 5.7.14. Zum zweiten Durchbruch

a) Kennlinienverläufe für die beiden möglichen Fälle; schematische Stromliniendarstellung für die beiden Fälle
b) Im Sperrbetrieb (Punkt *B* im Teilbild a))
c) Im aktiven Bereich (*A*)

(hot spot) aus. Überschreiten diese eine bestimmte Temperatur, so werden die Gebiete infolge des negativen Temperaturkoeffizienten instabil. Es erfolgt eine Einschnürung dieser Bereiche, die sehr schnell fortschreitet und den Transistor zerstört. Den zugehörigen Stromverlauf zeigt Bild 5.7.14a) als Kurve *a* mit dem Beginn in Punkt *A*.

Der andere Fall des zweiten Durchbruchs erfolgt im Sperrbetrieb, sofern die Spannung U_{CEO} überschritten ist. Hier setzt in Punkt *B* eine plötzliche Lawinenmultiplikation ein. Im Gegensatz zum zuvor geschilderten Fall ist dieser Durchbruch keine Folge von lokalen thermischen Einflüssen.

In Bild 5.7.14b) und c) sind die kritischen Stellen der hohen Stromdichten (heiße Stellen) für beide Fälle schematisch dargestellt.

5.7.5.5. Sicherer Arbeitsbereich

Den sicheren Arbeitsbereich eines Leistungs-Feldeffekttransistors zeigte Bild 5.5.1h). Gegenüber dem eines bipolaren Transistors gemäß Bild 5.7.15a) unterscheidet er sich nur durch das Fehlen des zweiten Durchbruchs im aktiven Bereich (Kurve III). Deshalb genügt es hier, den bipolaren Transistor zu betrachten. Es werden folgende Grenzen unterschieden:

I — maximal zulässiger Strom
II — maximal zulässige Leistung
III — Grenzen zur Sicherheit gegenüber dem zweiten Durchbruch
IV — maximale Spannung.

Hiervon ist lediglich noch nicht die Stromgrenze besprochen. Sie ist unter anderem durch die maximale Stromdichte im Halbleitermaterial und in den Leiterbahnen gegeben. In den Leiterbahnen kann durch den Strom unerwünschter Materialtransport erfolgen (Elektromigration). Den Zusammenhang mit der Lebenserwartung zeigt Bild 5.7.15 b).

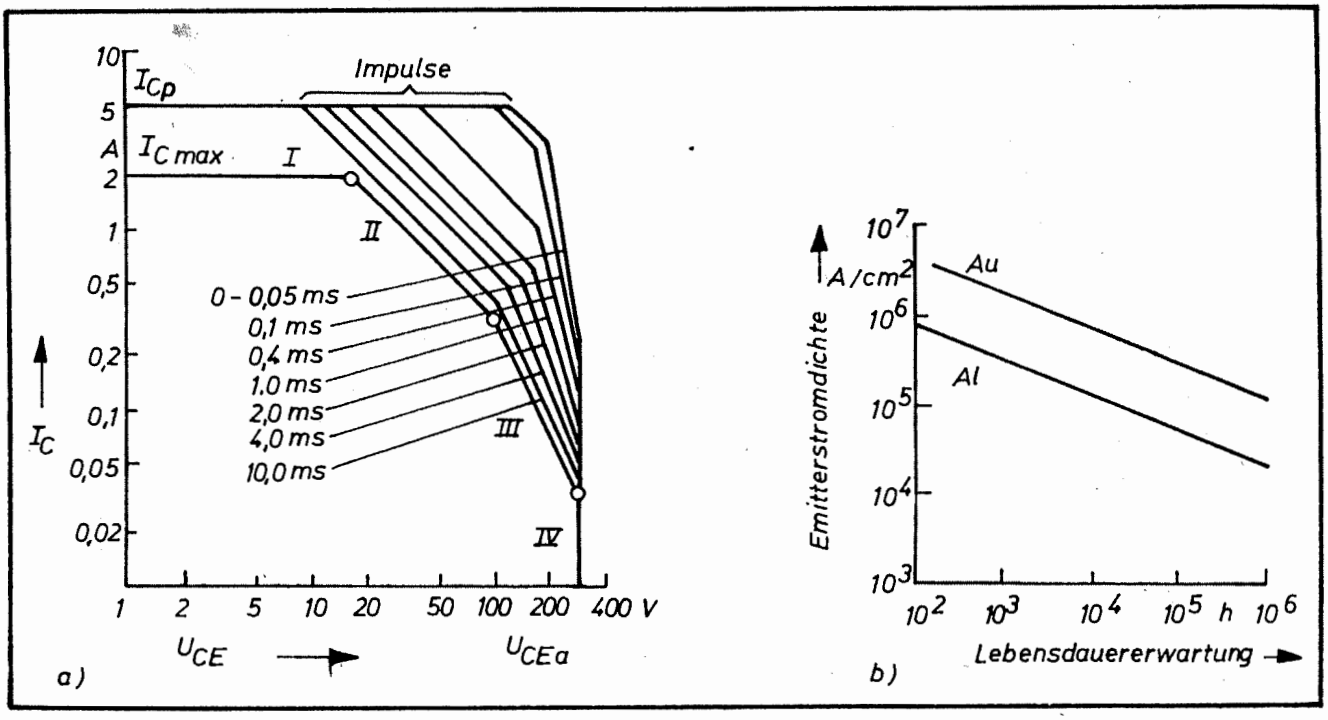


Bild 5.7.15. Zum sicheren Arbeitsbereich eines Leistungstransistors

a) Beispiel für einen Leistungstransistor 2N 3585

b) Abhängigkeit der Lebenserwartung von Leitbahnen von der Stromdichte

Im Bild 5.7.15a) sind neben den Grenzen der Gleichstrombelastung auch noch Grenzen für Impulsbelastungen gezeigt. Da hier nur kurzfristig Leistung erzeugt wird, liegen sie mit Ausnahme der Spannungsgrenze bei größeren Werten. Die Bereiche I, II und III bleiben weiter deutlich trennbar.

5.7.6. Wärmeableitung

5.7.6.1. Thermische Grundlagen

Jeder Körper besitzt mit seiner Temperatur T eine bestimmte Wärmeenergie, die in Joule (früher in Kalorien) gemessen wird. Bei Temperaturänderungen ΔT ist diese Energie zu ändern. Die ab- und zuzuführende Energie hängt von der Masse m und einer Materialkonstante c (spezifische Wärme in J/K kg) ab:

$$W_{th} = cm\,\Delta T\,. \tag{37}$$

Das Produkt $c \cdot m$ wird auch als Wärmekapazität C_{th} bezeichnet. Für die Änderung der Temperatur in einer bestimmten Zeit muß den Körper ein Wärmestrom (das ist eine Energie, gemessen in J) erreichen. Dieser *Wärmestrom* kann nach *drei physikalischen Mechanismen* erfolgen:

- Strahlung
- Konvektion
- Wärmeleitung.

Für die *Strahlung* gilt das STEFAN-BOLTZMANN-Gesetz, wonach der ideal schwarze Körper (vgl. 6.1.1.2.) mit der Oberfläche A und der Temperatur T die Energie

$$P = \sigma A T^4 \tag{38}$$

an die Umgebung abstrahlt. Darin bedeutet σ die STEFAN-BOLTZMANN-Konstante mit $5{,}6697 \cdot 10^{-8}\,\mathrm{W/m^2\,K^4}$. Umgekehrt wird jedem Körper aus der Umgebung mit der Temperatur T_u die dieser Temperatur entsprechende Energie zugestrahlt. Mit dem Absorptionskoeffizienten ε_A ($\leqq 1$) gilt daher für die insgesamt abgestrahlte minus zugestrahlte Energie

$$P_S = \varepsilon_A \sigma A (T^4 - T_u^4)\,. \tag{39}$$

Im allgemeinen wird diese Energie wegen der niedrigen Betriebstemperaturen der Halbleiter klein im Vergleich zu den anderen beiden Wärmeableitungen sein. Werte für ε_A enthält Tab. 5.7.5.

Bei *Konvektion* erfolgt der Wärmeaustausch durch Transport mittels eines Mediums (Gas, Flüssigkeit), das die Wärme gespeichert hat. Mit einer Wärmeübergangszahl α gilt

$$P_K = \alpha A (T - T_u)\,. \tag{40}$$

Tabelle 5.7.5. Absorptionskoeffizienten einiger Materialoberflächen

Kupfer, poliert	0,05
Kupfer, oxydiert	0,64
Aluminium, poliert	0,04
Aluminium, blank (unbearbeitet)	0,08
Aluminium, eloxiert (schwarz)	0,7 ... 0,8
Lackschichten (schwarz)	0,8 ... 0,9

Die Konvektion besitzt dann bei Halbleitern Bedeutung, wenn die Wärmeleitung nicht zu groß ist bzw. wenn spezielle Maßnahmen ergriffen werden, um sie zu vergrößern, z. B. mittels Gebläsen oder gezielt angewendeter Wärmeströmung bzw. in den Wärmerohren (5.7.6.4.).

Die *Wärmeleitung* ist die am meisten benutzte Wärmeableitung bei Halbleitern. Für einen prismatischen Körper mit dem Querschnitt A und der Länge l gilt mit einer Materialkonstanten λ, der Wärmeleitfähigkeit,

$$P_L = \lambda \frac{A}{l} (T - T_u)\,. \tag{41}$$

Tab. 5.7.6 enthält die genannten Materialkonstanten c und λ sowie die Dichte für ausgewählte Materialien. Da sie auch zumindest etwas von der Temperatur abhängen, sind die Werte bei Zimmertemperatur gewählt. Werte für Wärmeübergangszahlen ergänzt Tab. 5.7.7.

Mittels Gl. (41) läßt sich ein Vergleich von Halbleitermaterialien über das Produkt $\lambda(T - T_u)$ durchführen. Wenn man dazu $T_u = 25$; $T_{Ge} \approx 85$; $T_{Si} = 200$ und $T_{GaAs} = 400$ °C ansetzt, folgt für Ge ≈ 40; Si ≈ 255 und GaAs ≈ 180 sowie für GaP (mit $T = 400$ °C und $\lambda \approx 1{,}1$) ≈ 420 W/cm. Je größer die Zahl ist, desto mehr Wärme kann bei gleichen Dimensionen abgeführt werden. So folgt wiederum der große Vorteil von Si und die energetisch günstige Anwendung von GaP.

5.7.6.2. Elektrische Analogie

Die zuvor gegebenen Formeln legen eine elektrische Analogie der Wärmeableitung nahe, und zwar besonders wenn die Wärmestrahlung nicht berücksichtigt zu werden

Tabelle 5.7.6. Wärmeleitfähigkeit λ, spezifische Wärme c und Dichte ϱ einiger Materialien

Material	λ in W cm^{-1} K^{-1}	c in J K^{-1} g^{-1}	ϱ in g cm^{-3}
Ag	4,1	0,23	10,5
Al (99%)	4,1	0,9	2,7
Au	3,1	0,13	19,3
Cu (99%)	3,5	0,38	8,9
Fe	0,4 ... 0,7	0,45	7,8
Ge	0,64	0,31	5,32
Mo	1,4	0,25	10,2
Pb	0,33	0,13	11,3
Si	1,45	0,76	2,33
BeO	2,30		2,88
GaAs	0,46	0,32	5,32
Diamant	20	0,49	3,51
Messing	0,8 ... 1,2	0,38	8,4
Flußstahl	0,42	0,42	7,8
Grauguß	0,5	0,54	7,2
Glimmer	$5 \ldots 6 \cdot 10^{-3}$	0,88	ca. 3
Glas	10^{-2}	0,8	ca. 2,5
Porzellan	$1{,}5 \cdot 10^{-2}$	0,85	2,4
Al_2O_3	0,36	0,75	3,9
Papier	$\approx 10^{-3}$	ca. 1,5	0,7 ... 1,2
Kunststoffe	$\approx 2 \cdot 10^{-3}$	ca. 1,5	ca. 1,5
Vaseline	$1{,}9 \cdot 10^{-3}$	2	ca. 0,9
Wasserstoff	$1{,}8 \cdot 10^{-3}$	14,4	$8 \cdot 10^{-5}$
Luft	$2{,}5 \cdot 10^{-4}$	1,0	$1{,}3 \cdot 10^{-3}$

Tabelle 5.7.7. Beispiele für Wärmeübergangszahlen α in W/mK (v in m/s)

Wasser	$600 + 2100\sqrt{v}$
Siedendes Wasser	3500 ... 10000
Kondensierter Wasserdampf	12000
Luft (Gase) an glatten Metallflächen bei $v < 5$ m/s	$6 + 4\sqrt{v}$
Luft bei 10 m/s	100
Eigenkonvektion von Luft	5 ... 10

braucht. Sie zeigt Tab. 5.7.8. Neu ist dabei der Wärmewiderstand R_{th}. Mit ihm lassen sich die Gl. (40) und (41) in der Form des Ohmschen Gesetzes schreiben:

$$U = IR \quad \text{bzw.} \quad \Delta T = T - T_u = P_{th} R_{th}\,, \tag{42}$$

wobei für den eingeführten thermischen Widerstand gilt

$$R_{th} = \frac{l}{\lambda A} \tag{43a}$$

$$= \frac{1}{\alpha A}. \tag{43b}$$

In der Regel wird eine Wärmeableitung mehrstufig erfolgen. Dies wird schematisch im Beispiel von Bild 5.7.16 gezeigt: In einem Chip wird die elektrische Energie in die Wärme P_{th} (Stromquelle) umgesetzt. Sie wird in der Wärmekapazität $C_{th,G}$ z.T.

Tabelle 5.7.8. Analogie zwischen Wärmeableitung und stationärem elektrischen Strom

Wärme			Elektrische Ersatzschaltung		
Verlustleistung	P	W	Strom	I	A
Temperaturdifferenz	ΔT	K	Spannung	U	V
Wärmeleitfähigkeit	λ	W/mK	Spezifische Leitfähigkeit	$1/\varrho$	$1/\Omega$ m
Wärmekapazität	C_{th}	J/K	Kapazität	C	As/V
Wärmewiderstand	R_{th}	K/W	Widerstand	R	Ω

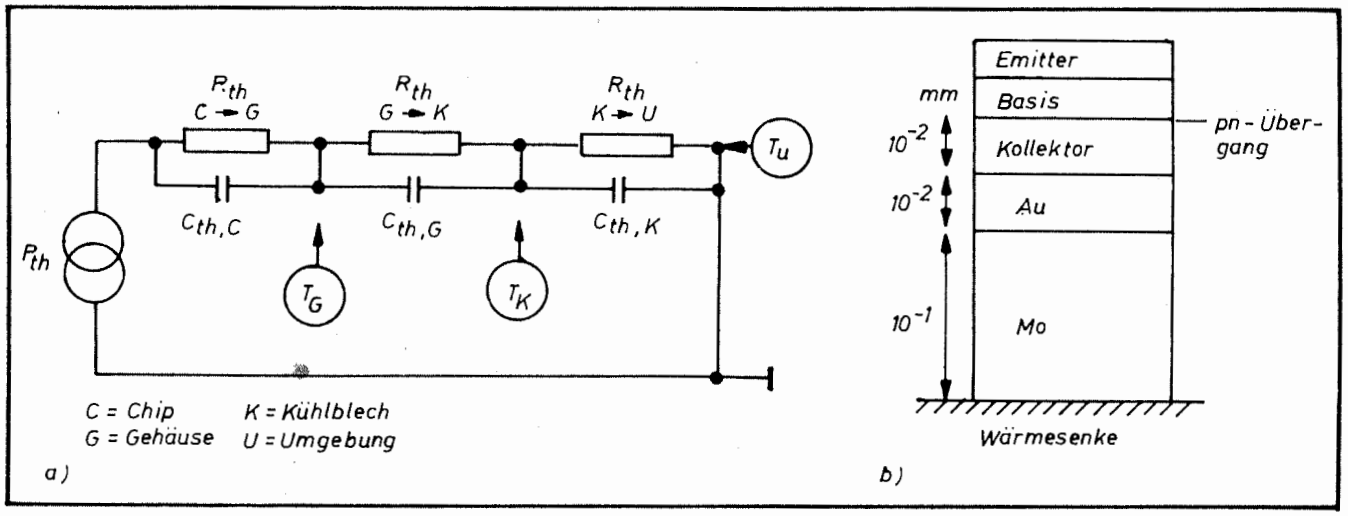


Bild 5.7.16. a) Beispiel für die elektrische Analogieschaltung des Wärmeleitungsvorganges. Die elektrische Energie wird im Chip in Wärme umgesetzt und über das Gehäuse und Kühlblech an die Umgebung weitergeleitet. Die Wärmekapazitäten von Chip, Gehäuse und Kühlblech speichern Wärme zwischendurch. Die Temperaturen entsprechen Potentialen
b) Beispiel zur Berechnung eines Wärmewiderstandes

gespeichert und z. T. über den Wärmewiderstand R_{th} vom Chip an das Gehäuse, in dem der Chip sich befindet, weitergeleitet. Das Gehäuse hat wieder eine Wärmekapazität ($C_{th,G}$) und Wärmekontakt zu einem Kühlblech (R_{th}) mit der Kapazität $C_{th,K}$. Letztlich gibt das Kühlblech über seinen Wärmewiderstand R_{th} Wärme an die Umgebung mit der Temperatur (Potential) T_u ab. Auf dieser Basis können nun die Temperaturen an den verschiedenen Stellen als Spannungen gegenüber T_u berechnet werden. Es ist sofort ersichtlich, daß bei Änderungen der elektrischen Energiezufuhr das System Einschwingvorgänge zeigen muß. Die Berechnung erfolgt allgemein in komplexer Schreibweise, z.B. [P 10]. In vielen Fällen werden die Kapazitäten vernachlässigt bzw. nur der eingeschwungene, stationäre Zustand betrachtet. Dadurch haben die Betrachtungen um die Wärmewiderstände große Bedeutung erlangt.

5.7.6.3. Beispiele für Wärmewiderstände

In der Regel werden *vier Arten von Wärmewiderständen* unterschieden:

- Gehäuse für Halbleiter
- Kühlkörper bzw. Kühlprofile
- Kühlbleche
- Übergangswiderstände.

Der Wärmewiderstand von *Gehäusen* läßt sich *unter zwei Bedingungen* berechnen:

- Es wird der Konvektions- und Strahlungswiderstand gegenüber der Luft mit berücksichtigt. Dieser Wert wird vor allem bei Transistoren angegeben, die keine

zusätzlichen Kühlmaßnahmen erfordern, also insbesondere Vorstufentransistoren.

- Es wird jener Wärmewiderstand angegeben, der gegenüber einem Gehäuse mit außen konstant gehaltener Temperatur gilt, bei dem also eine ideal gute Wärmeableitung vom Gehäuse vorausgesetzt wird. Diese Gehäuse sollen mit Kühlkörpern zusammengebaut betrieben werden, und alle folgenden Wärmewiderstände sind dann noch zu addieren.

Beispiele hierzu faßt Tab. 5.7.9 zusammen. Die Berechnung für ein Beispiel erfolgt anhand von Bild 5.7.16b). In dem *pn*-Übergang Basis-Kollektor entsteht die Wärme, und bis zur Wärmesenke, dem Gehäuse, sind drei Schichten zu durchlaufen. Folglich gilt bei konstantem Querschnitt

$$R_{\mathrm{th}} = \frac{1}{A} \sum_{i=1}^{3} (l_i/\lambda_i) \,. \tag{44}$$

Mit den Werten der Tabellen folgen für Germanium 0,09 und für Silicium 0,086 K/W cm².

Infolge der breiten Anwendung von elektronischen Bauelementen sind eine Anzahl von *Kühlkörpern* und *Kühlprofilen* genormt und damit als Bauteil erhältlich. Die Kühlkörper werden auf den Transistor oder die integrierte Schaltung aufgesetzt. Die Kühlprofile gibt es dagegen als Meterware. Sie werden in der gewünschten Länge

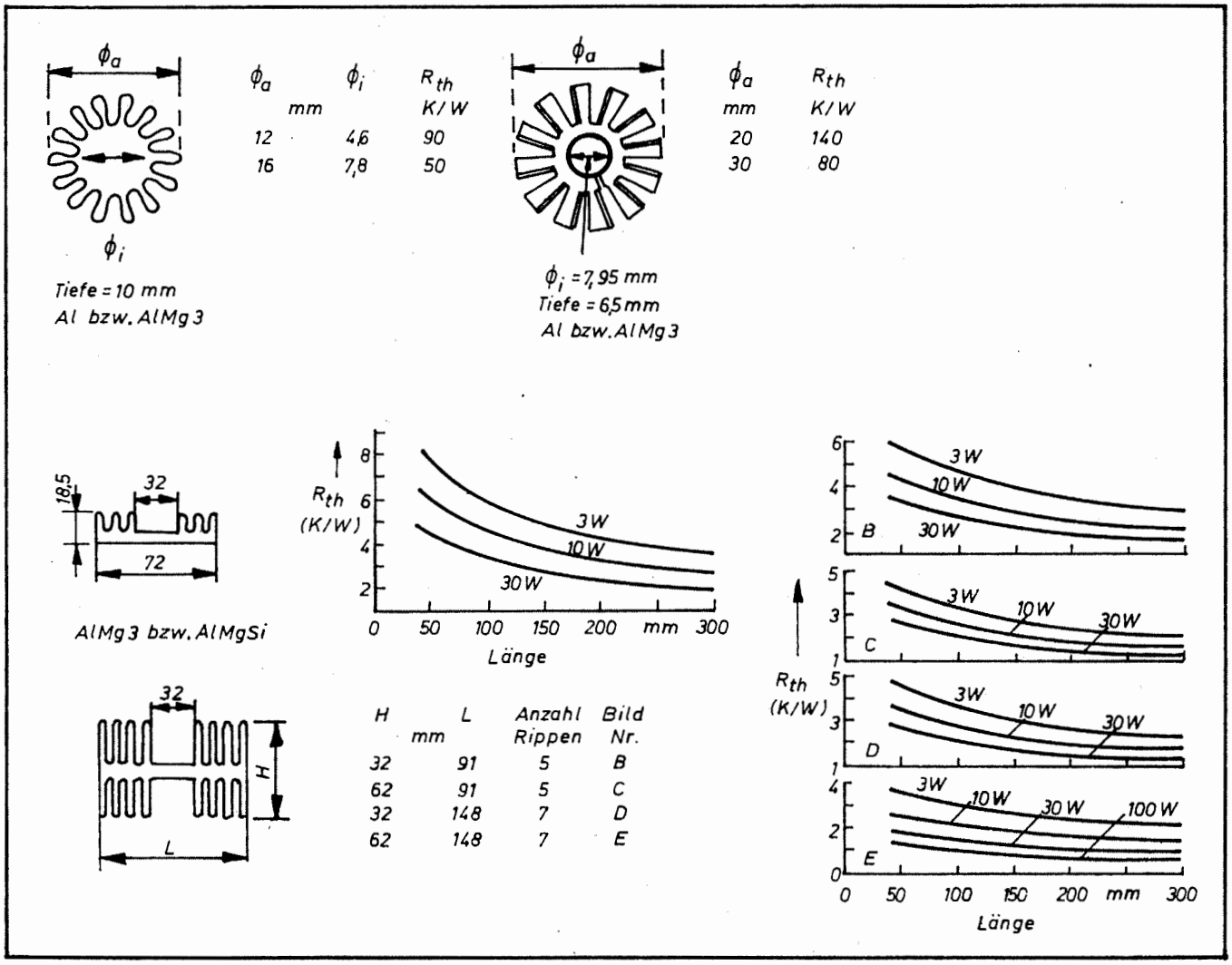
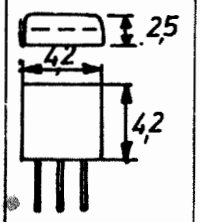
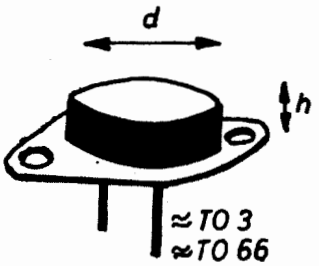
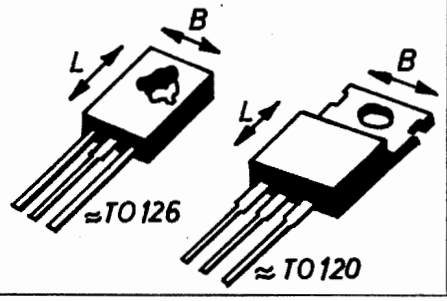
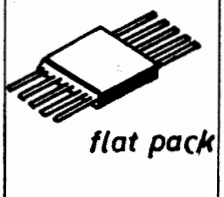
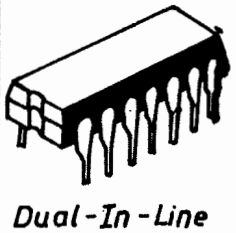


Bild 5.7.17. Wärmewiderstände von genormten Kühlkörpern und Kühlprofilen

Tabelle 5.7.9. Beispiele von Gehäusen mit groben Richtwerten für den Wärmewiderstand. In Klammern gesetzte Werte gelten gegenüber Umwelt ohne zusätzliche Kühlkörper, sonst bis zum Gehäuse.

d/h bzw. L/B Abmessungen in mm	siehe Bild	4,8/ca. 6	8,5/6,6	16,5/8,5(10)	19,5/8,5(7,3)
Bauartbezeichnung	L	$B \sim$ TO-18	$A \sim$ TO-5	$\sim$ TO-66	$\sim$ TO-3
Wärmewiderstand R_{th} in K/W	— (500)	150 ... 300 (500 ... 700)	35 ... 60 (150 ... 400)	4 ... 8	1,5 ... 5
Ableitbare Leistung in W bei $T_u = 25\%$ und Si	— (0,2)	0,3 (0,2)		25 (1)	100 (2)
Gehäuse					

d/h bzw. L/B Abmessungen in mm	15-20/20-13	Bild 4.5.2h)	Bild 4.5.2 c) bis e)			
			pin-Zahlen			
Bauartbezeichnung	~TO 120/126	14 u. 16 *pin*	8	14	16	24
Wärmewiderstand R_{th} in K/W	1,5 ... 3,5	20 ... 500 (300)	95● (200)	(100*) 80● (160)	(80*) 70● (150)	60●
Ableitbare Leistung in W bei $T_u = 25\%$ und Si	20 ... 100	(ca. 0,3)		(ca. 0,3)		
						

* in Keramik statt Plast ● Plast im Sockel

zugeschnitten. Deshalb ist ihr Wärmewiderstand je cm angegeben. Der angegebene Wärmewiderstand ergibt sich, wenn der Kühlkörper oder das Blech auf konstanter Temperatur gegenüber der Umgebung ist (Bild 5.7.17).

Damit bleibt noch der *Übergangswiderstand* vom Gehäuse des Halbleiters bis zum Kühlkörper übrig. Er muß unter Umständen sogar elektrisch isoliert ausgeführt werden. In jedem Fall ist aber ein sehr guter Wärmekontakt erwünscht. Richtwerte für die Übergangswiderstände gibt Tab. 5.7.10 wieder.

In einigen Fällen ist es notwendig oder einfacher, das Halbleiterbauelement auf ein Kühlblech zu setzen. Für diesen Fall gibt es viele Formeln. Doch meist genügt eine *Faustformel.* Bewährt hat sich:

$$R_{\mathrm{th}}/\mathrm{K/W} = 3{,}3\sqrt{\frac{C_{\mathrm{k}}}{d/\mathrm{mm}}} + 650\,\frac{C_{\mathrm{k}}}{A/\mathrm{cm}^2}, \tag{45}$$

wobei A die einseitige Oberfläche, d seine Dicke und C_{k} eine Konstante gemäß Tab. 5.7.11 ist.

Tabelle 5.7.10. Richtwerte für Wärmeübergangswiderstände in K/W bei einer Kontaktfläche von ca. 2 cm²

	Wärmeleitpaste	
	ohne	mit
Kühlmaterial unbearbeitet	0,5	0,2
Kühlmaterial eloxiert	0,7	0,3
Glimmerscheibe 50 µm	1,2	0,4
Glimmerscheibe 100 µm	1,4	0,6
Hartpapier 100 µm	1,0	
Bleizwischenlage 1 mm	0,5	—
TO-3 nicht isoliert	0,3	0,1

Tabelle 5.7.11. Werte für den Faktor C_{k} in Gl. (45)

Lage	Oberfläche	
	geschwärzt	blank
horizontal	0,5	1
vertikal	0,43	0,85

Auch Leiterplattenmaterial trägt zur Kühlung bei. Sofern es nicht selbst zur Kühlung herangezogen wird, muß der entsprechende Wärmewiderstand zu den anderen parallel geschaltet werden. Für Leiterbasismaterial gelten die Richtwerte von Tab. 5.7.12.

Für den Betrieb von Geräten in großen Höhen sei erwähnt, daß die Luftkühlung über 1000 m je 1000 m um fast 10% abnimmt. Der Wärmewiderstand von Kühlblechen beträgt daher in 5000 m Höhe das reichlich 1,5fache.

Tabelle 5.7.12. Wärmewiderstände in K/W für Leiterplattenmaterial (Pertinax), einseitig kaschiert, 1,5 mm Gesamtdicke und Flächen mit Kantenlängen größer als 50 mm

Pertinax	Seite der Wärmequelle	
	Cu	Pertinax
35 µm Cu	26	48
70 µm Cu	16	36

5.7.6.4. Wärmeleitrohr

Das erste Patent für ein Wärmeleitrohr (heat-pipe) wurde 1944 erteilt. Die erste Erprobung erfolgte 1964 in elektronischen Einrichtungen beim Weltraumeinsatz. Seit Ende der siebziger Jahre wird es auch in normalen Geräten genutzt. Es bildet eine Verkopplung von Wärmeleitung, Konvektion sowie Verdampfungswärme. Sein Prinzip geht aus Bild 5.7.18 hervor. Es ist ein geschlossenes Rohr, in dem sich etwas Arbeitsflüssigkeit, oft Wasser, befindet. Links wird Wärme zugeführt, dadurch verdampft die Flüssigkeit und nimmt viel Wärme auf. Mittels des Dampfdrucks und der Art der Flüssigkeit kann die dann nicht mehr überschreitbare Siedetemperatur eingestellt werden, bei Wasser mit vermindertem Dampfdruck z.B. 80 °C. Der Dampf strömt nach rechts. Damit er nicht zu früh kondensiert, ist der lange Mittelteil des Wärmeleitrohres außen wärmeisoliert. Am rechten Teil wird die Wärme, z.B. mittels Kühlblech, abgeführt (Teilbild c)). Dadurch kondensiert das Gas (Wasserdampf) und gibt die zuvor aufgenommene Wärme ab. Damit der Kreislauf geschlossen ist, muß die kondensierte Flüssigkeit an den Wänden des Wärmeleitrohres wieder nach links, zum Verdampfungsteil, zurückgeführt werden. Hierzu werden Kapillarstrukturen verwendet (Teilbilder d) bis g)). Sie sind notwendig, um den Betrieb des Wärmeleitrohres lageunabhängig zu machen. Die hohe Effektivität eines Wärmeleitrohres zeigt Bild 5.7.18b) im Vergleich mit einem gleich dimensionierten Kupferstab. Die Wärmeleitung ist um ein Vielfaches besser. Der größte Vorteil des Wärmeleitrohres besteht

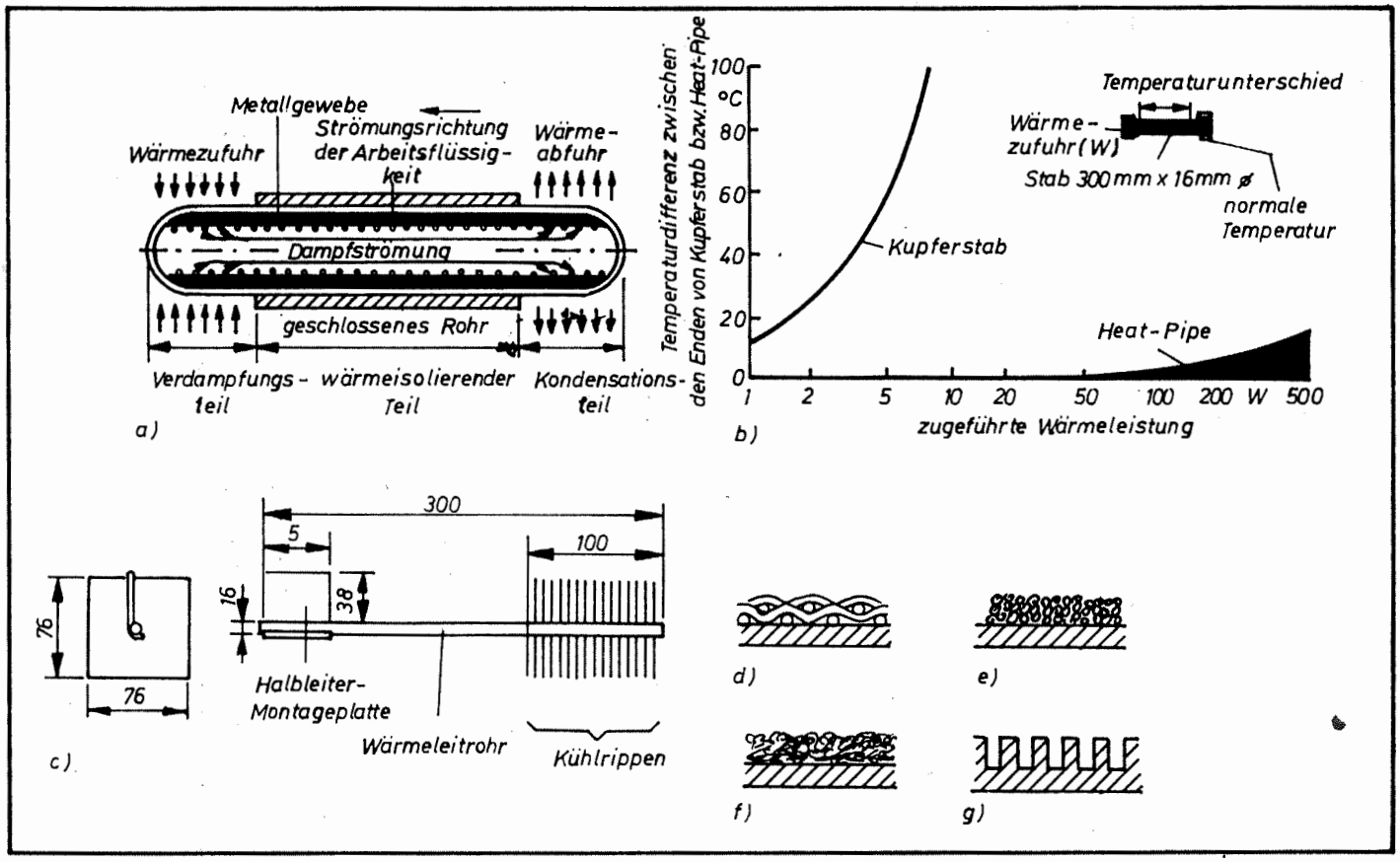


Bild 5.7.18. Zum Wärmeleitrohr

a) Prinzipielle Funktion
b) Vergleich der Wärmeleitfähigkeit mit einem gleich großen Kupferstab
c) Anwendungsprinzip mit Zahlen zur Dimensionierung
d) bis g) Kapillarstrukturen zur Rückführung der Flüssigkeit
d) Gewebtes Drahtgeflecht
e) Gesintertes Metallpulver
f) Gesinterte Metall- oder Glasfiber
g) Rillen in der Wand des Wärmeleitrohres

darin, daß die Wärme nicht am Entstehungsort, sondern an nahezu beliebig wählbarer Stelle abgeleitet werden kann. Dadurch kann die elektronische Schaltung z. T. wesentlich günstiger gestaltet werden. Das Wärmeleitrohr ist auch in gebogener Form realisierbar.

5.7.6.5. Impulswärmewiderstand

Bisher wurde die elektrisch zugeführte Energie als konstant angenommen. Das ist aber so gut wie nie der Fall. Deshalb müssen auch die Kapazitäten im Wärmeersatzschaltbild 5.7.16a) berücksichtigt werden. Die *RC-Glieder* besitzen Zeitkonstanten, die gemäß

$$\tau_{\mathrm{th}} = R_{\mathrm{th}} C_{\mathrm{th}} \tag{46}$$

berechnet werden können. Für einen prismatischen Stab gilt für R_{th} (43a), und die Wärmekapazität kann gemäß (37) als

$$C_{\mathrm{th}} = cm = c\varrho A l\,, \tag{47}$$

mit ϱ als Dichte, geschrieben werden. Daher folgt

$$\tau_{\mathrm{th}} = \frac{l^2}{\lambda c \varrho}. \tag{48}$$

Sie ist also eine auf die Länge bezogene Materialkonstante und kann aus Tab. 5.7.8 berechnet werden. Für komplette Bauelemente sind mehrere solche Zeitkonstanten in Reihe geschaltet. Deshalb hat z. B. die Abkühlung nicht genau einen exponentiellen Verlauf, wie er z. B. nach Abschnitt 1.5. zu erwarten wäre (Bild 5.7.19a) und b)). Dennoch genügt die Angabe einer Zeitkonstanten oft als erste Näherung. Sie liegt in der Regel bei einer Sekunde. Als Folge dieser Erscheinung kann ein Transistor mit erheblich stärkeren Impulsen belastet werden. Für einige Tastverhältnisse k zeigt dies anschaulich Bild 5.7.19c). Dies wurde auch bereits (ohne Begründung) in den Bildern 5.5.1h) und 5.7.15a) sichtbar. In der Praxis wird hier jedoch ein durch die thermischen Kapazitäten verminderter Impulswärmewiderstand eingeführt. Ihn zeigt beispielhaft Bild 5.7.19d). Auf dieser Ebene wird der im Abschnitt 5.7.1.6. behandelte Mitlaufeffekt für das Rauschen noch verständlicher.

5.7.6.6. Thermische Stabilität

Es sei von einer vereinfachten Ersatzschaltung nach Bild 5.7.20a) ausgegangen. Für den Knoten A gilt dann die Differentialgleichung

$$-P + C_{\mathrm{th}} \frac{\mathrm{d}T}{\mathrm{d}t} + R_{\mathrm{th}}(T_{\mathrm{C}} - T_{\mathrm{u}}) = 0\,. \tag{49}$$

In welche Richtung sich die Chiptemperatur unter gegebenen Bedingungen ändert, bestimmt der Ausdruck

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{1}{C_{\mathrm{th}}}\left[P - \frac{T_{\mathrm{C}} - T_{\mathrm{u}}}{R_{\mathrm{th}}}\right]. \tag{50}$$

In der Regel wird dem Kollektor nicht konstante Leistung, sondern konstante Spannung U (Kollektorspannung) zugeführt. Dann ist die von ihm aufgenommene Leistung von seinem elektrischen Widerstand R abhängig:

$$P = \frac{U^2}{R}. \tag{51}$$

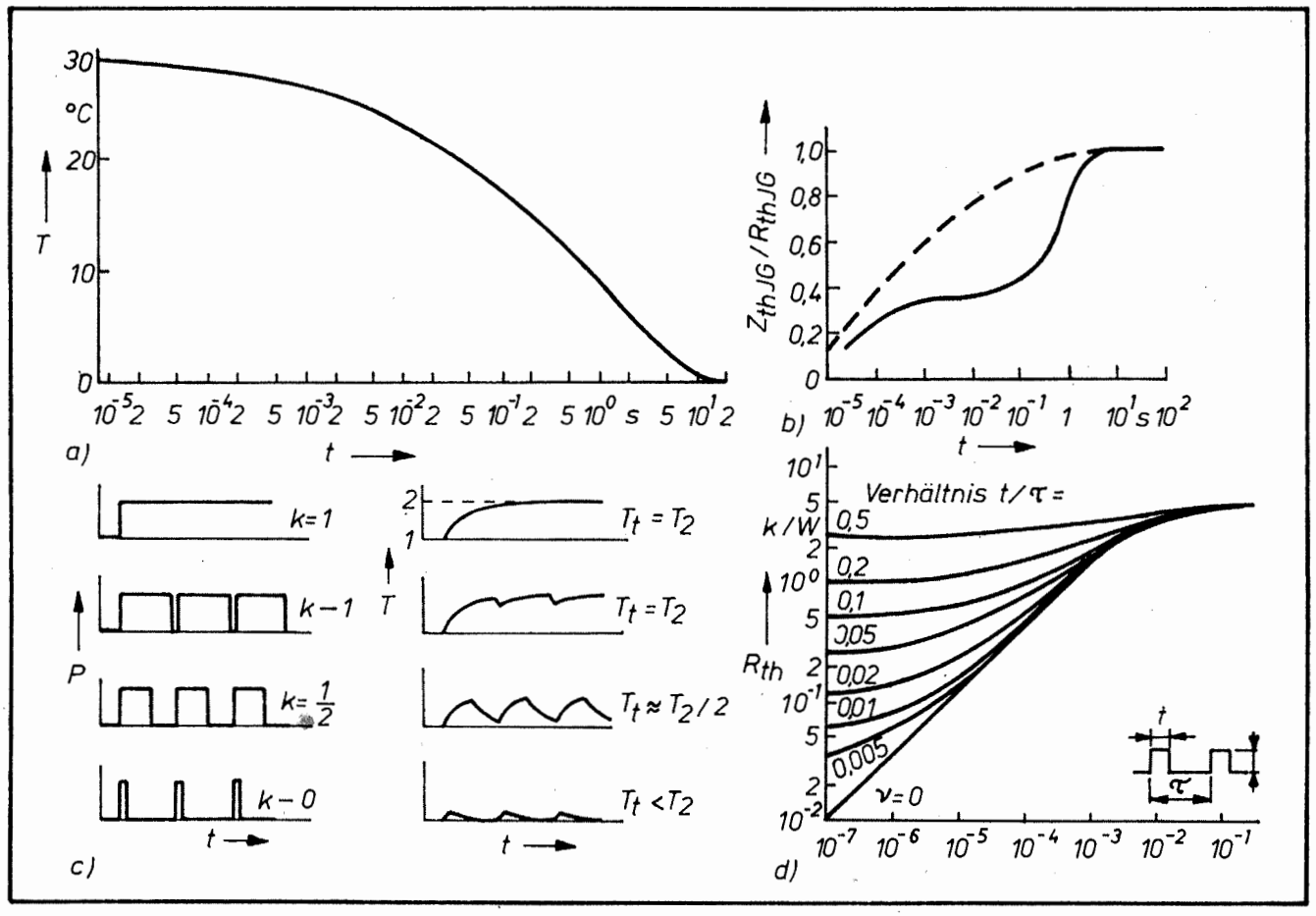


Bild 5.7.19. Zum Impulswärmewiderstand

a) Abkühlkurve eines Transistors SF 126 nach einer Belastung von 11 V und 50 mA
b) Einschwingvorgang eines Leistungstransistors (gestrichelt eine *RC*-Zeitkonstante)
c) Impulse und sich einstellende Sperrschichttemperatur bei einigen Tastverhältnissen $k = T_{ein}/T_{gesamt}$
d) Beispiel eines Impulswärmewiderstandes

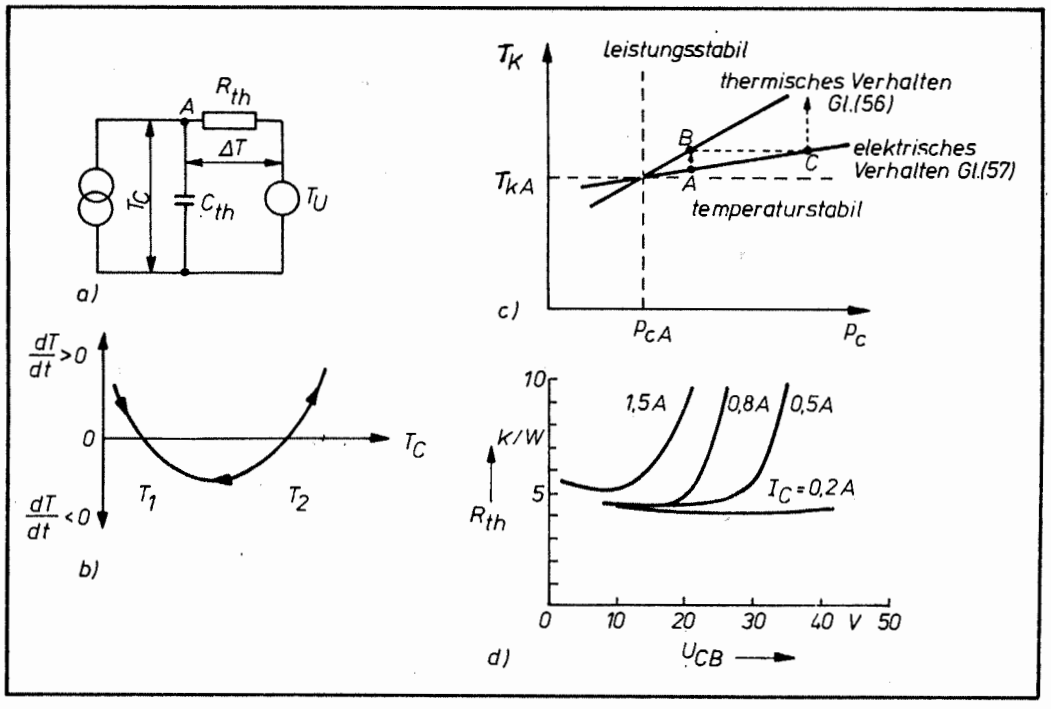


Bild 5.7.20. Zur thermischen Stabilität

a) Vereinfachtes Ersatzschaltbild
b) Mögliche Temperaturverläufe bei negativem Kollektorwiderstand
c) Zur Unterscheidung von elektrischer und thermischer Stabilität
d) Abhängigkeit des Wärmewiderstandes eines kleinen Leistungstransistors von Kollektorstrom und -spannung

Besitzt R nun einen negativen Temperaturkoeffizienten, also

$$R = R_0(1 - \alpha\,\Delta T)\,, \tag{52}$$

so geht (50) über in

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{1}{C_{\mathrm{th}}}\left[\frac{U^2}{R_0(1 - \alpha\,\Delta T)} - \frac{T_{\mathrm{C}} - T_{\mathrm{u}}}{R_{\mathrm{th}}}\right]. \tag{53}$$

Da

$$\Delta T = T_{\mathrm{C}} - T_{\mathrm{u}} \tag{54}$$

gilt, ergibt sich für die Chiptemperatur und ihre zeitliche Änderung ein ungefährer Verlauf, wie ihn Bild 5.7.20b) zeigt. Für Chiptemperaturen unterhalb von T_2 wird sich also stets der stabile Wert T_1 einstellen. Sobald jedoch durch irgendwelche Umstände (z.B. Impulse oder Erhöhung von T_{u}) auch nur für sehr kurze Zeit die Temperatur T_2 überschritten wird, heizt sich das Material bis zu seiner Zerstörung auf.

Um der soeben geschilderten Instabilität auszuweichen, werden u. a. die Ströme von Transistoren gegen Schwankungen der Umwelttemperatur stabilisiert (vgl. 5.6.3.3.). Infolge der Stabilisierung des Arbeitspunktes ist deutlich zwischen dem elektrischen und thermischen Verhalten eines Transistors zu unterscheiden. Im elektrischen Verhalten wird nämlich ausgedrückt, wie stark eine äußere Temperaturänderung den Arbeitspunkt beeinflußt. Mit der Kollektorverlustleistung P_{CA} im Arbeitspunkt bei der normalen Temperatur und der Stabilitätsgröße G gilt näherungsweise für die sich bei erhöhter Umgebungstemperatur einstellende Verlustleistung

$$\boxed{P_{\mathrm{C}} = P_{\mathrm{CA}} + G\,\Delta T}\,. \tag{55}$$

Für das rein thermische Verhalten gilt weiter Gl. (43). Werden beide nach der Temperatur differenziert, so folgt

$$\frac{\mathrm{d}P}{\mathrm{d}\Delta T} = \frac{1}{R_{\mathrm{th}}} \quad \text{bzw.} \quad \frac{\mathrm{d}P_{\mathrm{C}}}{\mathrm{d}\Delta T} = G\,. \tag{56) bzw. (57}$$

Beide Kurven sind in Bild 5.7.20c) als Geraden dargestellt. Bei idealer Wärmeleitung würde eine Leistungserhöhung zu keiner Temperaturänderung führen (waagerechte gestrichelte Gerade). Bei idealer Temperaturstabilisierung würde eine Temperaturänderung der Umwelt (d. h. des Chip) zu keiner Arbeitspunktverlagerung und damit zu keiner Leistungsänderung führen (senkrechte gestrichelte Gerade). In der Praxis ergeben sich aber wegen (56) und (57) zwei geneigte Geraden. Es mögen beide Geraden in einem speziellen Fall so liegen, wie es Bild 5.7.20c) zeigt. Eine Erhöhung der Umgebungstemperatur verlagert dann den Arbeitspunkt längs der Geraden für elektrisches Verhalten bis zum Punkt A. Dazu gehört aber der thermische Punkt B und zu diesem der elektrische Punkt C usw. Die Temperatur wächst also ständig. Die geschilderte Anordnung ist also instabil. Deshalb kann gefolgert werden, daß bei Stabilität für die beiden Steigungen gelten muß

$$\boxed{GK < 1}\,. \tag{58}$$

Andersartig geführte Stabilitätsbetrachtungen enthalten u. a. [M 20, M 25, P 9].

Alle bisherigen Betrachtungen sind in der Regel nur Näherungen. Denn die Erwärmung im Kristall kann örtlich unterschiedlich sein. Hierauf wurde u. a. schon beim zweiten Durchbruch (5.7.5.4.) hingewiesen. Bild 5.7.20d) zeigt dies in anderer Weise für einen Leistungstransistor. Bei ihm ist der gemessene Wärmewiderstand in der Nähe der Leistungsgrenzen von Kollektorspannung und Kollektorstromabhängig.

Weiter sei auf die schon bei den Widerständen behandelte, aber allgemeingültige Unterlastkurve gemäß Bild 1.1.14 hingewiesen.

5.7.7. Leistungsgrenzen

Die Vielzahl der Transistortypen ist sehr groß. Es wird geschätzt, daß 1982 auf dem Weltmarkt 25000 Transistortypen und 8000 Diodentypen existierten. Die Fülle macht es schwer, allgemeine Aussagen und erreichte Grenzdaten vorzugeben. Dies ist bei den folgenden Betrachtungen zu beachten. Wie unsicher dabei selbst die Einschätzung so einfach erscheinender Parameter wie maximaler Strom bzw. maximale Spannung ist, zeigen die Bilder 5.7.21 a) und b). In der Regel waren die später erreichten Werte besser als die prognostizierten. Grenzdaten bezüglich der Leistung zeigt Teilbild c) und bezüglich der Schaltleistung (wobei auch noch zu behandelnde Thyristoren einbezogen sind) Teilbild d). Unwahrscheinlich groß ist der Preisverfall bei den Einzeltransistoren. Bei den bipolaren Leistungstransistoren ist dieser Prozeß noch nicht abgeschlossen, und bei den Leistungs-FET beginnt er gerade (Teilbild e)). Für die Erzeugung und Verstärkung hoher Frequenzen und großer Leistungen ist das Produkt aus Frequenz und Leistung wichtig. Die zeitliche Entwicklung hierfür gibt Teilbild f) wieder. Für die z. Z. erfolgende Entwicklung der Leistungs-FET geben die Teilbilder g) und h) ergänzende Daten wieder.

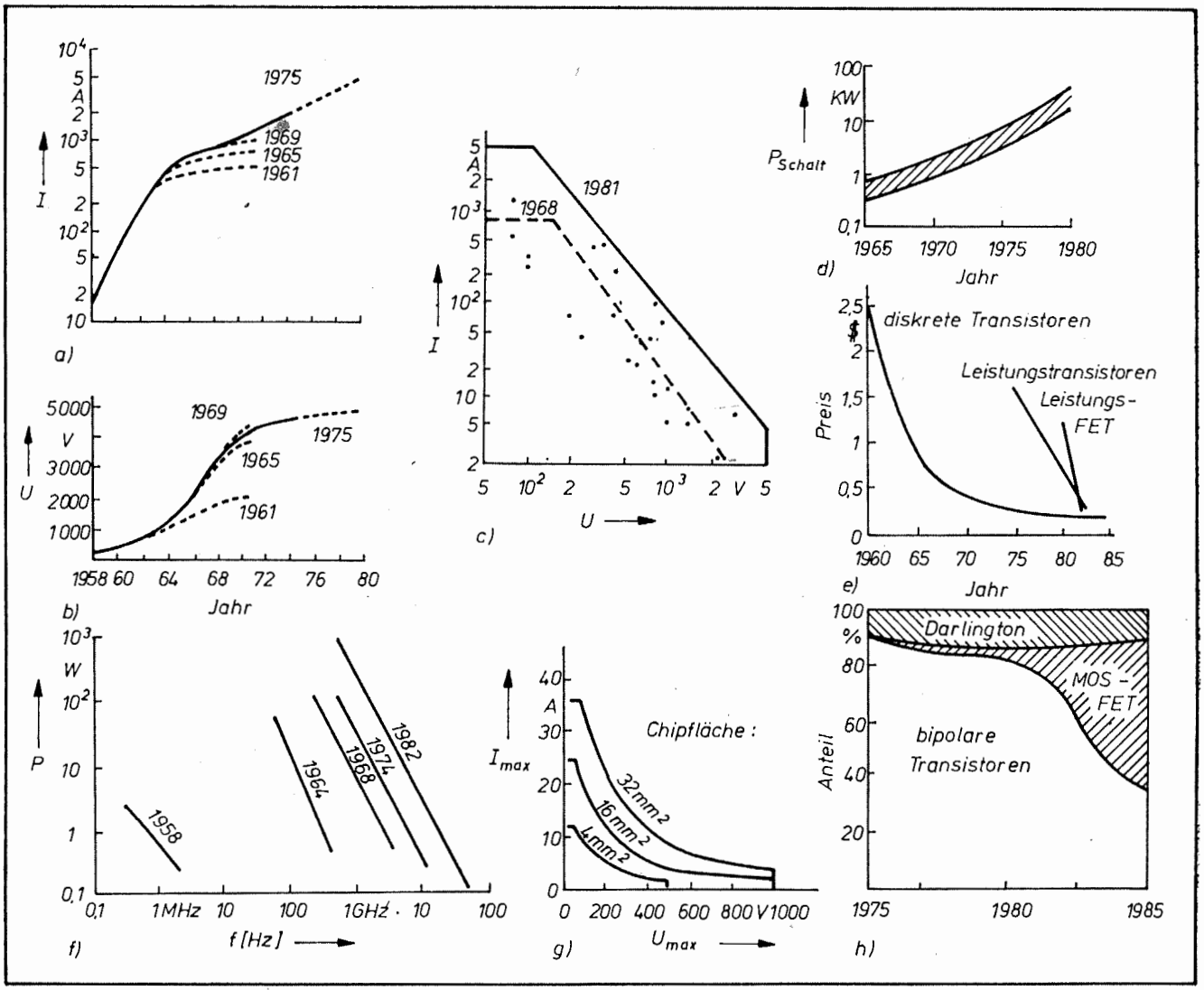


Bild 5.7.21. Zu den Leistungsgrenzen von Halbleiterbauelementen

a) Maximaler Strom, b) maximale Spannung von bipolaren Transistoren in verschiedenen Prognosen und den bis 1975 erreichten Werten

c) Leistungsgrenzen 1968 und 1981 sowie als Punkte eingetragene Beispiele bei bipolaren Transistoren

d) Erreichte Schaltleistungen von Leistungstransistoren

e) Entwicklung der Preise von Einzeltransistoren, bipolaren Leistungstransistoren sowie Leistungs-FET

f) Zu verschiedenen Zeiten erreichte Grenzen des Produktes aus Leistung und Frequenz

g) Erreichbare Spannungs- und Stromgrenzen als Funktion der Chipfläche

h) Entwicklung des Anteils bipolarer Transistoren, FET und DARLINGTON-Transistoren bei den Leistungstransistoren

6. Bauelemente spezieller Gebiete

6.0. Überblick

Die Elektronik ist nicht Selbstzweck. Sie hat Aufgaben zu lösen. Dadurch sind Verbindungen zum Menschen und zu technischen Geräten, Einrichtungen und Ausrüstungen notwendig. Es ist also eine Wandlung zu und von nichtelektrischen Signalen notwendig. Dieses Gebiet ist sehr umfangreich. Gehören zu ihm doch unter anderem die große Fülle der Sensoren oder Signalwandler, der optischen und akustischen Ausgabe, der Stellglieder, der peripheren Geräte der Rechentechnik, wie Drucker, Plotter, Speicher usw., ja im weitesten Sinne auch Teilgebiete der Elektroniktechnologie selbst, wie z. B. Fassungen, Schalter, Verbindungstechniken (wie Leiterplatten usw.). Hier kann nur eine kleine Auswahl getroffen werden. Sie beschränkt sich auf die wichtigsten Bauelemente entsprechend dem eingeschränkten Anwendungsanliegen dieses Buches, also nicht berücksichtigt werden z. B. Nachrichtenübertragung und Unterhaltungselektronik.

Eingeteilt ist dieses Kapitel in die Hauptabschnitte

- Optoelektronik
- Vakuumelektronik
- Mechano- und Akustoelektronik
- sonstige Bauelemente.

Die *Optoelektronik* wird immer mehr zu einem eigenen Zweig der Elektronik. Hierfür gibt es vor vor allem drei Gründe. Durch die hohen Trägerfrequenzen lassen sich große Bandbreiten und damit hohe Verarbeitungsgeschwindigkeiten realisieren. Die Übertragung optischer Signale, z. B. über Glasfasern, eröffnet neue effektive Leitungswege. Das wichtigste Sinnesorgan des Menschen ist das Auge. Die Optoelektronik stellt hierfür die Wiedergabemittel bereit. Hier wird vor allem die Signalausgabe und optoelektronische Signalwandlung dargestellt. Dabei werden jedoch jene Gebiete ausgeklammert, die in das traditionelle Gebiet der *Vakuumelektronik* gehören. Dies sind vor allem die offensichtlich einstweilen unersetzlichen Ausführungen der Bildröhren, z. B. für Oszillographen und Displaytechniken mit Fernsehbildröhren in Schwarz-Weiß und Farbe sowie mit Digitrons und Plasmapanels.

Kurz wird weiter auf die Bildaufnahme, die Bildwandler und SEV eingegangen. Nicht behandelt werden die Vielzahl der anderen industriellen Techniken, wie Elektronenmikroskop, Elektronen- und Plasmastrahlbearbeitung, Beleuchtungstechnik, Höchstfrequenztechnik usw. Einen geschichtlichen Überblick zur Entwicklung der Optoelektronik gibt Tab. 6.1.0.

Sehr umfangreich ist ebenfalls das Gebiet der *Mechanoelektronik*. Hier wird auf die folgenden ausgewählten Teilgebiete eingegangen: Tasten, Relais, Drehspulinstrumente, Registriergeräte, magnetomotorische Speicher (Magnetbandgeräte, Floppy disc, Plattenspeicher) und dann auf piezoelektronische Bauelemente, wie Quarz, Piezokeramik, Oberflächenwellen und -filter (Akustoelektronik). Nicht behandelt werden u. a. Motoren, Lautsprecher, Mikrofone. Auf die Verbindungstechniken wird im Zusammenhang mit der Schicht- und Hybridtechnologie (9.4.) eingegangen.

Unter den *sonstigen Bauelementen* werden zusätzliche Fakten aufgenommen, die eine gewisse Bedeutung für die Elektronik besitzen oder sonst schwer einzuordnen waren, also einige Sensoren, wie HALL-Elemente, Magnetfeldwiderstände, Dehnungsmeßstreifen, Tunnel- und Backward-Dioden, Thyristoren und dazugehörige Bauelemente, sowie etwas Tieftemperaturphysik, welche in der Zukunft eine größere Bedeutung erlangen könnte.

Tabelle 6.1.0. Geschichtsdaten zur Opto- und Vakuumelektronik (Details für Lumineszenz- und Laserdioden in [W12])

1810	DAVY entdeckt ersten Lichtbogen zwischen Holzkohle
1839	BECQUEREL entdeckt Fotoeffekt an mit Halogen bedeckten Silberplatten in Schwefelsäure
1842	Bogenlampe von FOUCAULT
1850	Leitung in Gasen durch PLÜCKERS
1854	GOEBEL-Lampe
1859	Glühlampe von GOEBEL
1873	MAY und SMITH beobachten Fotoleitung an Selen
1876	ADAMS und DAY entdecken Sperrschichtfotoeffekt
1876	GOLDSTEIN prägt Begriff Katodenstrahlen (Elektronenstrahl)
1876	Lichtbogenlampe
1879	Erste Serienglühlampe von EDISON
1879	Kohlefadenlampe
1881	Glühlampenfabrik bei Siemens
1883	EDISON entdeckt glühelektrischen Effekt
1884	NIPKOW-Scheibe für Fernsehen
1888	Entdeckung der ersten Flüssigkristalle
1888	HALLWACHS entdeckt äußeren lichtelektrischen Effekt (Fotozelle)
1893	Metallfadenlampe
1893	Fotozelle durch GEITEL und ELSTER erfunden
1897	Katodenstrahlröhre von BRAUN
1900	Strahlungsgesetz von PLANCK (Anfang der Quantentheorie)
1905	Photonenhypothese von EINSTEIN
1906	Erste Verstärkerröhre von LIEBEN
1907	ROUND stellt Lichterzeugung an einem SiC-Kristall fest
1910	Langdrahtlampe aus Wolframdraht
1910	Erste Neon-Gasentladungsröhre
1913	Glühwendelung
1917	EINSTEIN formuliert die stimulierte Emission
1918	Glimmlampe für Hausgebrauch
1920	Thalliumsulfid von CASE als Fotowiderstand angewendet
1920	und später: Untersuchungen von POHL und GUDDEN über Fotoleitung an Kristallen
1921	Magnetron
1923	Ikonoskop von ZWORYKIN
1923	LOSSEW entdeckt Lichtemission an Kristallen neu
1924	Fernsehröhre schwarzweiß
1925	Doppelwendellampe
1926	Begründung der Elektronenoptik durch H. BUSCH
1928	LADENBERG und KOPFERMANN beobachten die stimulierte Emission
1930	Um diese Zeit Anwendung der Selenzelle für Belichtungsmesser
1930	Mehrschichtfotokatoden
1930	SeTe-Zelle von MICHELSON
1933	Erstes Elektronenmikroskop
1933	Quecksilberdampflampe
1935	DESTRIAU entdeckt Lichtemission an ZnS
1935	Erstes Leuchtstoffpatent: ZnBe-Silikat
1936	Leuchtstofflampe
1937	Klystron
1937	Bildwandler
1939	Orthikon-Bildaufnahmeröhre
1940	Magisches Auge als Vorläufer der Digitronröhren
1943	Wanderfeldröhre
1944	PbS-Zelle von GUDDEN
1946	Si-Fotozelle

Fortsetzung Tabelle 6.1.0

1946	CdS-Zelle von FREDERICHS
1949	Maskenfarbröhre
1949	Fototransistor
1950	Vidicon-Aufnahmeröhre
1952	HAYNES und BRIGGE stellen Lichtemission an Ge- und Si-Dioden fest
1953	Erster Maser von WEBER
1954	Erste industriell produzierte Lochmaskenbildröhre: 53 cm Diagonale, 70° Ablenkung, 300000 Farbtripel
1955	Erste Ziffernanzeigeröhren (NIXIE) mit Gasentladung
1955	Begründung der Faseroptik
1956	Erste Index-Farbröhre
1957	Vidicon-Fernsehröhre
1959	Halogenglühlampe
1959	Lumineszenzdioden mittels der III-V-Verbindungshalbleiter von WELKER (1952)
1960	Erster Festkörperlaser von MAIMANN auf Rubinbasis
1962	Erste Halbleiter-Laser-Dioden
1964	Nobelpreise für Physik (Laser) an TOWNES, BASOW, PROKHOROW
1965	Entdeckung der isoelektrischen Zentren von N in GaP
1965	Natrium-Hochdrucklampe
1967	Plumbicon-Fernsehaufnahmeröhre
1967	Transparente PLZT-Keramik
1967	Digitron-Anzeigeröhre in Japan
1968	Japanische In-line-Farbbildröhre Trinitron
1968	Zn-O-Paar für Rotlichtdioden mit hohem Wirkungsgrad
1969	Laser mit Heterostrukturen für wichtige Schwellströme
1970	BOYLE und SMITH weisen auf CCD-Elemente für optischen Einsatz hin
1970	Erste Glasfasern mit Verlusten kleiner als 10 dB/km
1971	Mischkristalle mit N-Dotierung für Farbbereich grün bis rot
1972	Sichtspeicherröhre (Transferspeicherröhre)
1972	Präzisions-In-line-Farbbildröhre in Produktion
1972	Erste Halbleiterlaser bei Raumtemperatur
1973	Erste praktisch brauchbare LCD-Anzeigen auf Basis DSM
1974	Dünnschicht-Elektrolumineszenzanzeige von Sharp
1975	Lichtleiter mit wenigen dB/m
1976	LCD mit Feldeffekt in Anwendung
1976	Erste Lichtleiterübertragung in Japan, 14 km, 123 Mbit/s
1977	Erster LCD-Kleinfernsehschirm, 8 × 11 cm: 8339 Punkte
1978	Erste elekrochemische Anzeigen
1982	Elektronischer Standbildfotoapparat MAVICA

6.1. Optoelektronik

6.1.1. Lichttechnische Größen und allgemeine Grundlagen

6.1.1.1. Maßeinheiten

Das Licht ist ein Teil des Spektrums elektromagnetischer Wellen (Bild 6.1.1). Sie breiten sich mit der Lichtgeschwindigkeit $c = 2{,}997925 \cdot 10^8$ m/s aus. Daher besteht zwischen der Wellenlänge λ und der Frequenz f (zuweilen auch ν) die Beziehung

$$\lambda = c/f\,. \tag{1}$$

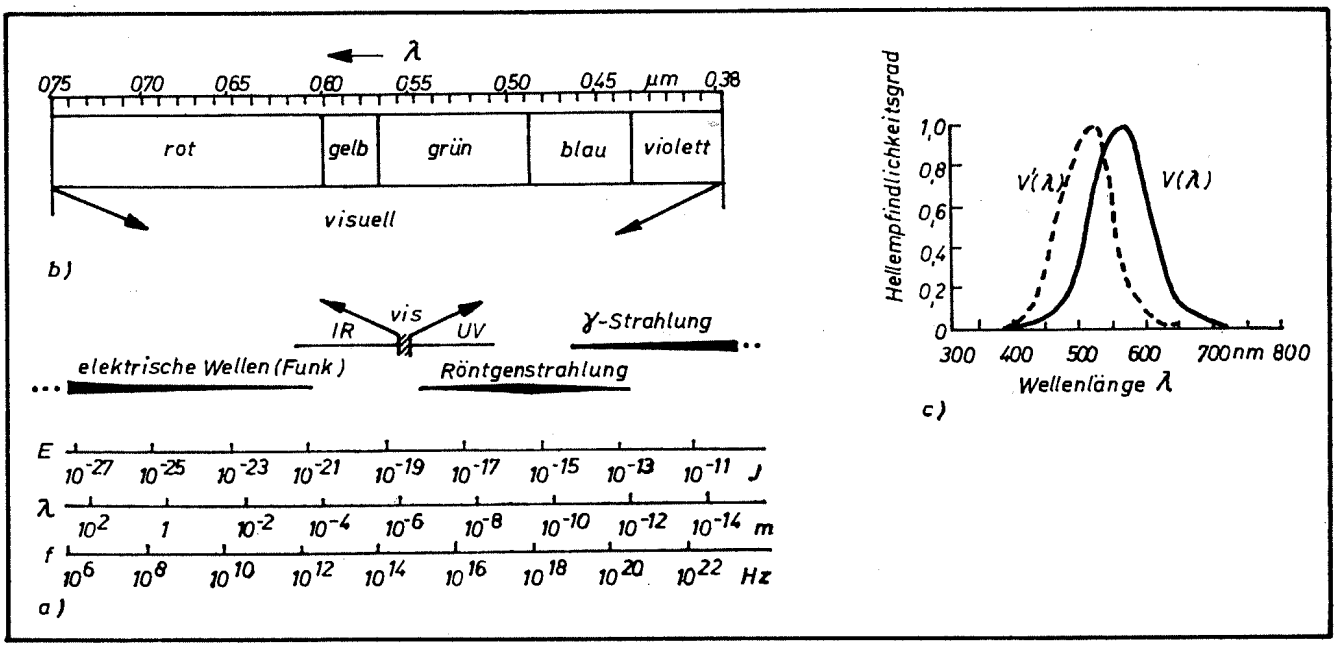


Bild 6.1.1. Licht als Teilgebiet der elektromagnetischen Wellen

a) Es wird im allgemeinen in die drei Gebiete IR = Infrarot; vis. = sichtbares (visuell wahrnehmbares) Licht und UV = Ultraviolett eingeteilt

b) Das sichtbare Licht kann weiter (als monochromatische Welle) die Spektralfarben von rot bis violett annehmen. Daher die nicht mehr für den Menschen sichtbaren Grenzbereiche IR und UV. Bezüglich der üblichen Farben siehe Bild 6.1.4

c) Zur Umrechnung von energetischen, physikalischen Größen in lichttechnische, physiologische dient die genormte Augenempfindlichkeitskurve für das Tagessehen $V(\lambda)$. Für das Nachtsehen gilt $V'(\lambda)$. Tab. 6.1.1 gibt genauere Werte

Licht hat im dualistischen Sinne auch Quantencharakter. Deshalb gehört zu jeder Frequenz die Energie der einzelnen Photonen mit

$$W_{ph} = hf = hc/\lambda\,, \tag{2}$$

wobei $h = 6{,}6256 \cdot 10^{-34}\ \mathrm{Ws^2}$ die PLANCKsche Konstante ist.

Im Sinne von *Strahlung* und *Licht* sind *zwei* unterschiedliche *Maßangaben* möglich:

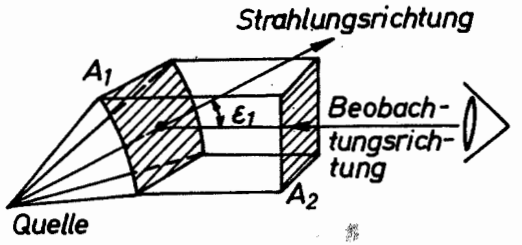
energetisch-physikalische (Index E)
lichttechnisch-physiologische (Index L).

Sie sind beide über die international genormte und als Mittelwert über viele Menschen gewonnene spektrale *Augenempfindlichkeit* verknüpft. Ihren Verlauf als Funktion der Wellenlänge λ zeigt Bild 6.1.1c) und Daten dazu Tab. 6.1.1.

Tabelle 6.1.1. Zahlenwerte für $V(\lambda)$ und $V'(\lambda)$

λ in nm	$V(\lambda)$	$V'(\lambda)$	λ in nm	$V(\lambda)$	$V'(\lambda)$
380	0	0,000589	560	0,995	0,329
400	0,0004	0,00929	580	0,870	0,121
420	0,0040	0,0966	600	0,631	0,033
440	0,023	0,3281	620	0,381	0,007
460	0,060	0,567	640	0,175	0,0015
480	0,139	0,793	660	0,061	0,0003
500	0,323	0,982	680	0,017	0,00007
513	0,564	1,000	700	0,0041	—
520	0,710	0,935	720	0,00105	—
540	0,954	0,650	740	0,00025	—
555	1,000	0,405	760	0,00006	—

Tabelle 6.1.2. Lichttechnische Einheiten und ihre physikalischen Entsprechungen
(bezüglich Umrechnungen siehe Text)



Allgemein		Energetisch-physikalisch (Index E)		Lichttechnisch-physiologisch (Index L)	
Begriff	Formelzeichen u. Gleichung	Bezeichnung	Einheit	Bezeichnung	Einheit
Energie	W	Strahlungsenergie	J	Lichtmenge	lm s
Leistung (P)	$\Phi = \mathrm{d}W/\mathrm{d}t$	Strahlungsleistung	W	Lichtstrom (Lumen)	lm
Senderseitige Größen (Index 1)					
Leistungsdichte	$\Psi = \mathrm{d}\Phi/\mathrm{d}A_1$	spezif. Ausstrahlung	W/m^2	spezif. Lichtausstrahlung	lm/m^2
Leistung je Raumwinkel	$I = \mathrm{d}\Phi/\mathrm{d}\omega_1$	Strahlstärke	W/sr	Lichtstärke (Candela)	1 cd = lm/sr
Leistung je Raumwinkel und projizierte Fläche	$L = \dfrac{\mathrm{d}^2\Phi}{\mathrm{d}\omega_1 \mathrm{d}A_1 \cos\varepsilon_1}$	Strahldichte	$W/sr\ m^2$	Leuchtdichte	cd/m^2 1 sb (Stilb) $= 10^4\ cd/m^2$
Empfangsseitige Größen (Index 2)					
Leistung je Fläche	$E = \mathrm{d}\Phi/\mathrm{d}A_2$	Bestrahlungsstärke	W/m^2	Beleuchtungsstärke (Lux)	1 lx = lm/m^2
Zeitintegral der einfallenden Leistung je Fläche	$H = \int E\,\mathrm{d}t$	Bestrahlung	Ws/m^2	Belichtung	lx s

Die Kurve $V(\lambda)$ gilt für das *Tagessehen*. Für sie sind die Zäpfchen (mit den drei verschiedenen Farbpigmenten) verantwortlich. Sie gilt für Leuchtdichtegrößen von etwa $\geqq 30$ cd/m². Die $V'(\lambda)$-Kurve geht auf die Stäbchen zurück, welche nur Helligkeitswerte aufnehmen können. Sie gilt bei *Dämmerungs-* bzw. *Nachtsehen*, also unterhalb etwa 0,003 cd/m². Im Zwischenbereich treten Übergangskurven auf. Zwischen den verschiedenen energetischen und lichttechnischen Größen gilt die Umrechnung

$$X_{\mathrm{L}} = K_{\mathrm{m}} \int_0^\infty X_{\mathrm{E}}(\lambda)\, V(\lambda)\, \mathrm{d}\lambda\,. \tag{3}$$

Darin ist K_{m} eine Konstante, die ebenfalls international festgelegt ist und infolge der Normierung der Empfindlichkeitskurve auf 1 notwendig ist. Sie besitzt den Wert

$$K_{\mathrm{m}} \approx 673\ \mathrm{lm/W}\,. \tag{4}$$

Seit 1979 gibt es eine neue, genauere Festlegung der *lichttechnischen Basiseinheit Candela* (cd). Sie geht von einer monochromatischen Strahlung von 540 THz ($\approx$ 555 nm) aus. Sie besitzt die Lichtstärke 1 cd, wenn sie in den Raumwinkel von 1 sr mit 1/683 W strahlt. 1 Steradiant (sr) ist der Raumwinkel, der auf einer Kugel mit einem Radius von einem m auf ihrer Oberfläche einen m² abgrenzt. Anschaulicher wird der Raumwinkel, wenn von der gesamten Kugeloberfläche $4\pi r^2$ ausgegangen wird. Sie besitzt also den Raumwinkel 4π sr $\approx$ 12,566 37 sr. Eine andere Vorstellung geht von einem Kegel als ihrem Mittelpunkt aus. Er grenzt dann 1 sr ein, wenn für seinen Öffnungswinkel α gilt $\cos(\alpha/2) = 1 - 1/(2\pi)$, also $\alpha \approx 65{,}5411°$ $\approx$ 1,144 908 rad beträgt. Ein monochromatischer Kugelstrahler (strahlt gleichmäßig in alle Raumrichtungen) mit $\lambda = 555$ nm besitzt also dann die Lichtstärke 1 cd, wenn er $4\pi/683$ Watt, ca. 18,4 mW, abstrahlt. Ein ideal schwarzer Körper mit 1/60 cm² Oberfläche strahlt bei der Temperatur des erstarrenden Platins von 1796 °C 1 cd ab (alte Definition). Für Glühlampen gilt grob, daß sie soviel cd abstrahlen, wie sie Watt verbrauchen.

Bezüglich der Maßeinheiten ist es günstig, von den Energiemaßen auszugehen und dann senderseitige und empfangsseitige Größen zu unterscheiden. So entsteht Tab. 6.1.2. Besonders für die *Leuchtdichte* existiert eine Vielzahl alter Einheiten. Es gilt 1 cd/m²

$$= 1/\pi\ \mathrm{asb}\ (\text{Apostilb}) = 10^4\ \mathrm{sb}\ (\text{Stilb}) = 10^4/\pi\ \mathrm{L}\ (\text{Lambert})$$
$$\approx 10{,}764\ \mathrm{cd/ft^2} \approx 3{,}426\ \mathrm{fL}\ (\text{foot lambert})$$
$$= 1550\ \mathrm{cd/in^2}\ (\text{candela per square inch}).$$

Hiervon ist nur das Stilb als inkohärente Einheit im Système International zugelassen.

Für die *Beleuchtungsstärke* hatte früher noch das ph (Phot) mit 1 lx = 10^{-4} ph und die foot candle (fc) mit 1 lx = 0,0929 fc Bedeutung.

Die Größenordnung von lichttechnischen Werten zeigt Bild 6.1.2.

6.1.1.2. Schwarzer Strahler

Körper, denen thermische Energie zugeführt wird, nehmen eine höhere Temperatur an. Heiße Körper strahlen einen Teil dieser Energie als sichtbares Licht ab. Kalte Körper nehmen aus der wärmeren Umgebung thermische Energie auf. So entsteht ein thermischer Strahlungsfluß zwischen Körpern unterschiedlicher Temperatur. Dieser Strahlungsfluß wird noch durch den meist wellenlängenabhängigen Absorptionskoeffizienten beeinflußt. Ein *ideal schwarzer Körper absorbiert alle auf ihn treffende Energie*. Diesen Zustand erfüllt am besten ein innen geschwärzter *Hohlraum* mit

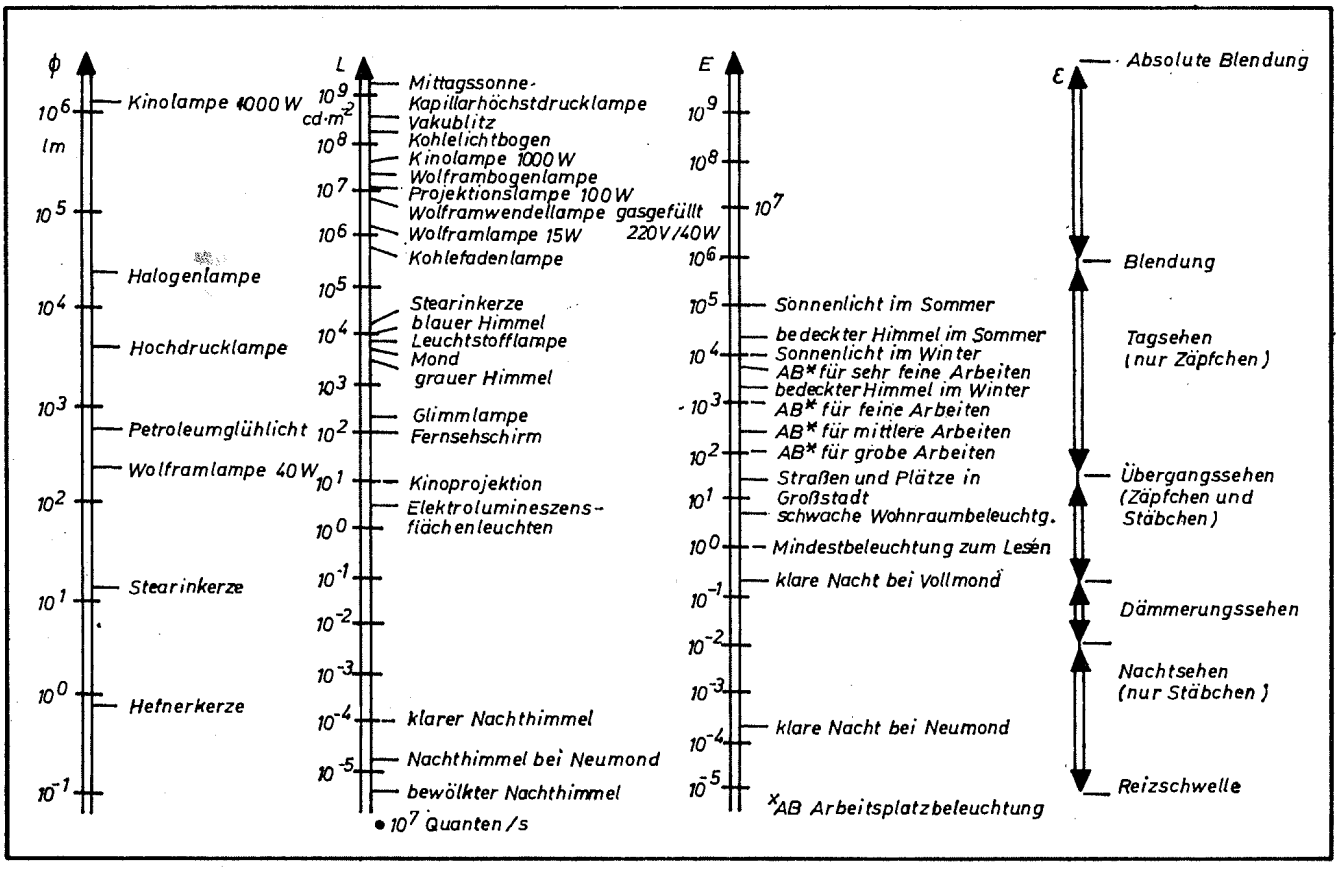


Bild 6.1.2. Größenbereiche der Größen Lichtstrom, Leuchtdichte und Beleuchtungsstärke. Den Werten der Beleuchtungsstärke sind auf unser Auge bezogene Werte zugeordnet (E in lx)

kleiner Öffnung. Es kann keine Strahlung wieder durch die kleine Öffnung hinausgestrahlt werden. Umgekehrt strahlt ein solcher Hohlraum bei erhöhter Temperatur in ihm auch ideal die thermische Energie über seine Öffnung ab. Deshalb wird der ideal schwarze Körper auch *Hohlraumstrahler* genannt. Seine spektral abgestrahlte Leistung hängt nur von der Frequenz bzw. Wellenlänge ab, und hierfür gelten die von PLANCK gefundenen Formeln. Da sie nur für quantisierte Photonen ableitbar sind, begründeten sie die Quantentheorie:

$$\frac{\mathrm{d}W}{\mathrm{d}\nu} = \frac{2h}{c^2}\nu^3 \exp\left(-h\nu/kT\right) \qquad (\text{in J/m}^2) \tag{5}$$

bzw.

$$\frac{\mathrm{d}W}{\mathrm{d}\lambda} = \frac{2hc^2}{\lambda^5}\exp\left(-hc/\lambda kT\right) \qquad (\text{in W/m}). \tag{6}$$

Darin bedeutet $k = 1{,}38054 \cdot 10^{-23}\,\mathrm{J/K}$ die BOLTZMANN-Konstante (die anderen Konstanten wie bei (2)). Den wellenlängenabhängigen Verlauf zeigt Bild 6.1.3a). Dabei fällt auf, daß sichtbares Licht erst ab etwa 500 bis 600 °C abgestrahlt wird, während Sonnenlicht mit ca. 6000 K alle Farben etwa gleich stark, also *weißes Licht*, enthält. Für einen schwarzen Strahler kann eine spezielle Empfindlichkeitskurve des menschlichen Auges theoretisch und experimentell bestimmt werden. Sie zeigt Bild 6.1.3b). Sie ist unterhalb von 2000 K gering und erreicht ihr Maximum bei 6000 K. *Glühlampen* strahlen bei etwa 3000 K mit 20 lm/W.

Weiter ist aus Bild 6.1.3a) ersichtlich, daß es für jede Temperatur eine Wellenlänge λ_m gibt, die maximale Energie besitzt. Dies ist das WIENsche *Verschiebungsgesetz*

$$\lambda_\mathrm{m} T = \frac{c\,h}{5\left(1 - \exp\left(-hc/kT\lambda\mathrm{m}\right)\right)} \approx 2{,}898 \cdot 10^{-3}\,\mathrm{m/K}\,. \tag{7}$$

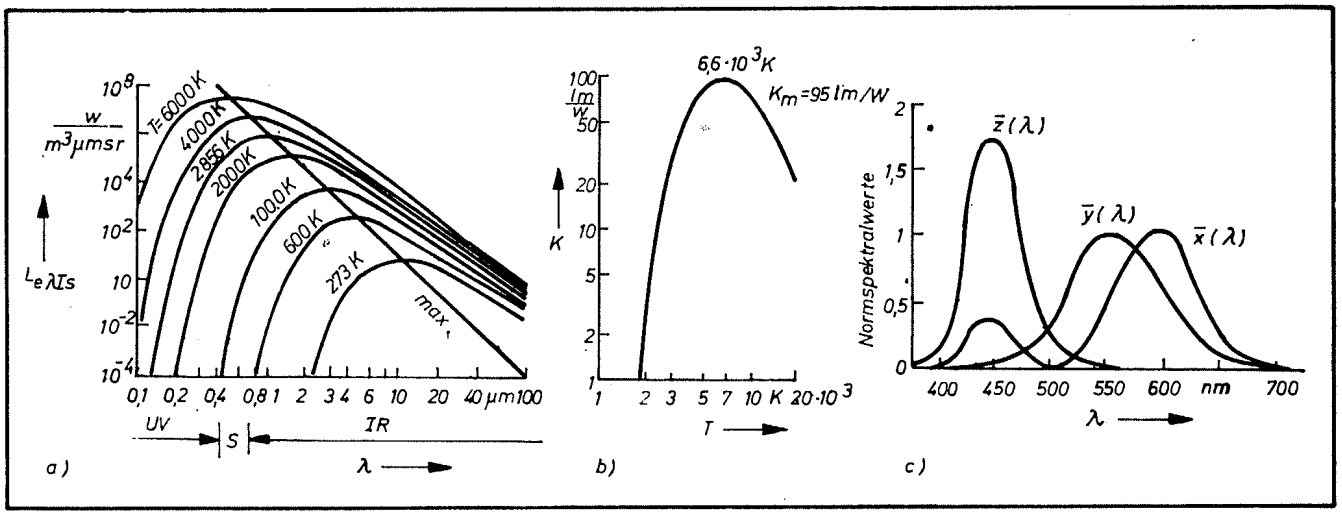


Bild 6.1.3. a) Spektrale Strahldichte des schwarzen Körpers als Funktion der Wellenlänge und der Temperatur
b) Empfindlichkeit des menschlichen Auges in bezug auf die Strahlung des schwarzen Körpers
c) Normspektralwertkurven des energiegleichen Spektrums bei 2° Gesichtsfeld

Wird die Energie von Gl. (5) bzw. (6) über das gesamte Spektrum integriert, so entsteht das STEFAN-BOLTZMANN-Gesetz gemäß Abschnitt 5.7.6.1.

6.1.1.3. Farbmetrik

Unser *Auge* enthält zwei Typen von *Sinneszellen*:

- Stäbchen: Sie sprechen nur auf Helligkeitswerte an und besitzen eine maximale Empfindlichkeit für $\lambda \approx 500$ nm. Sie sind generell etwa 1000mal so empfindlich wie die Zäpfchen (Bild 6.1.2).
- Zäpfchen: Hier gibt es drei Typen mit den maximalen Empfindlichkeiten bei
  rot: $\lambda \approx 675$ nm ,
  grün: $\lambda \approx 545$ nm ,
  blau: $\lambda \approx 475$ nm .

(Bei den verschieden farbtüchtigen Tieren treten z. T. andere Farbmaxima auf.) Mit diesen drei Typen sehen wir alle Farben. Vereinfacht dargestellt, wird dabei vom Nervensystem der Erregungsunterschied der drei Typen ausgewertet. Aus dieser Sicht ist es sofort verständlich, daß mit drei ausgewählten monochromatischen Quellen alle Farben darstellbar sein müssen. Hierauf baut die heutige *Farbmetrik* auf.

Die physiologische *Farbwahrnehmung* gliedert sich in:

- Helligkeit der Farbe (Leuchtdichte)
- Farbton
- Farbsättigung.

Der *Farbton* (z.B. rot oder blau) charakterisiert die Farbwahrnehmung in Bezug zu einer Wellenlänge der Strahlung. Die *Farbsättigung* entspricht der Intensität, mit der Farbe wirkt, z.B. haben rot oder blau höhere Sättigung als rötlich oder bläulich. Geringe Leuchtdichte verändert die Farben etwas, z.B. gelbgrün geht bei geringer Leuchtdichte in olivgrün über.

Farbton und Farbsättigung werden in der Farbmetrik unter dem Begriff *Farbart* zusammengefaßt. Sie *enthält* also *nicht die Leuchtdichte*. Eine Farbart kann entsprechend einer additiven Mischung aus mehreren Spektralkomponenten zusammengesetzt werden. Für die Farbmetrik geht man von drei *Primärvalenzen* aus:

$$\left.\begin{array}{ll} R\,(\text{rot}): & \lambda = 700{,}0\ \text{nm}: \text{relative Strahldichte } 73{,}0420 \\ G\,(\text{grün}): & \lambda = 546{,}1\ \text{nm}: \text{relative Strahldichte } 1{,}3971 \\ B\,(\text{blau}): & \lambda = 435{,}8\ \text{nm}: \text{relative Strahldichte } 1{,}0000\,. \end{array}\right\} \tag{8}$$

Mit den relativen Beträgen r, g, b folgt dann die *Farbvalenz* zu

$$F = rR + gG + bB\,. \tag{9}$$

Diese räumliche Darstellung ist zwar vollkommen, aber wenig praktikabel. Deshalb wird auf eine Maßzahl verzichtet: Eine Ebene konstanter Helligkeit verläuft u. a. durch die Punkte R, G, B. So entsteht das *Farbdreieck*.

Die Primärvalenzen R, G, B besitzen weiter einige praktische Nachteile: *Spektralfarben lassen sich nicht direkt mit voller Sättigung ermischen*, und die Verteilung der Farben im Dreieck ist sehr ungleichmäßig. Aus diesem Grunde wurden bereits 1931 gemischte Primärvalenzen als *Normvalenzsystem* genormt:

$$\left.\begin{aligned} X &= 2{,}364\,60R - 0{,}515\,15G + 0{,}005\,20B\,, \\ Y &= -0{,}896\,53R + 1{,}426\,40G - 0{,}014\,41B\,, \\ Z &= -0{,}468\,07R + 0{,}088\,75G + 1{,}009\,21B\,. \end{aligned}\right\} \tag{10}$$

Mit ihm können alle natürlichen Farben und die Spektralfarben mit nur positiven Faktoren x_r, y_r, z_r gemäß

$$F = x_r X + y_r Y + z_r Z \tag{11}$$

dargestellt werden. Im Mittelpunkt des neuen Farbdreiecks liegt das energiegleiche Spektrum (weiß, grau, schwarz) mit der Farbvalenz E. Die reinen Spektralfarben lassen sich als Werte (Normspektralwerte) $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ und $\bar{z}(\lambda)$ gemäß Bild 6.1.3c) bzw. Tab. 6.1.3 darstellen, wobei $\bar{y}(\lambda)$ identisch ist mit $V(\lambda)$ in Bild 6.1.1c) bzw. Tab. 6.1.1.

Tabelle 6.1.3. Normspektralwerte für 2° Gesichtsfeldgröße; $\bar{y}(\lambda)$ ist identisch mit $V(\lambda)$, s. Tab. 6.1.1

λ in nm	$\bar{x}(\lambda)$	$\bar{z}(\lambda)$	λ in nm	$\bar{x}(\lambda)$	$\bar{z}(\lambda)$
380	0,0014	0,0065	580	0,9163	0,0017
390	0,0042	0,0201	590	1,0263	0,0011
400	0,0143	0,0679	600	1,0622	0,0008
410	0,0435	0,2074	610	1,0026	0,0003
420	0,1344	0,6456	620	0,8544	0,0002
430	0,2839	1,3856	630	0,6424	0,0000
440	0,3483	1,7471	640	0,4479	0,0000
450	0,3362	1,7721	650	0,2835	0,0000
460	0,2908	1,6692	660	0,1649	0,0000
470	0,1954	1,2876	670	0,0874	0,0000
480	0,0956	0,8130	680	0,0468	0,0000
490	0,0320	0,4652	690	0,0227	0,0000
500	0,0049	0,2720	700	0,0114	0,0000
510	0,0093	0,1582	710	0,0058	0,0000
520	0,0633	0,0782	720	0,0029	0,0000
530	0,1655	0,0422	730	0,0014	0,0000
540	0,2904	0,0203	740	0,0007	0,0000
550	0,4334	0,0087	750	0,0003	0,0000
560	0,5945	0,0039	760	0,0002	0,0000
570	0,7621	0,0021			

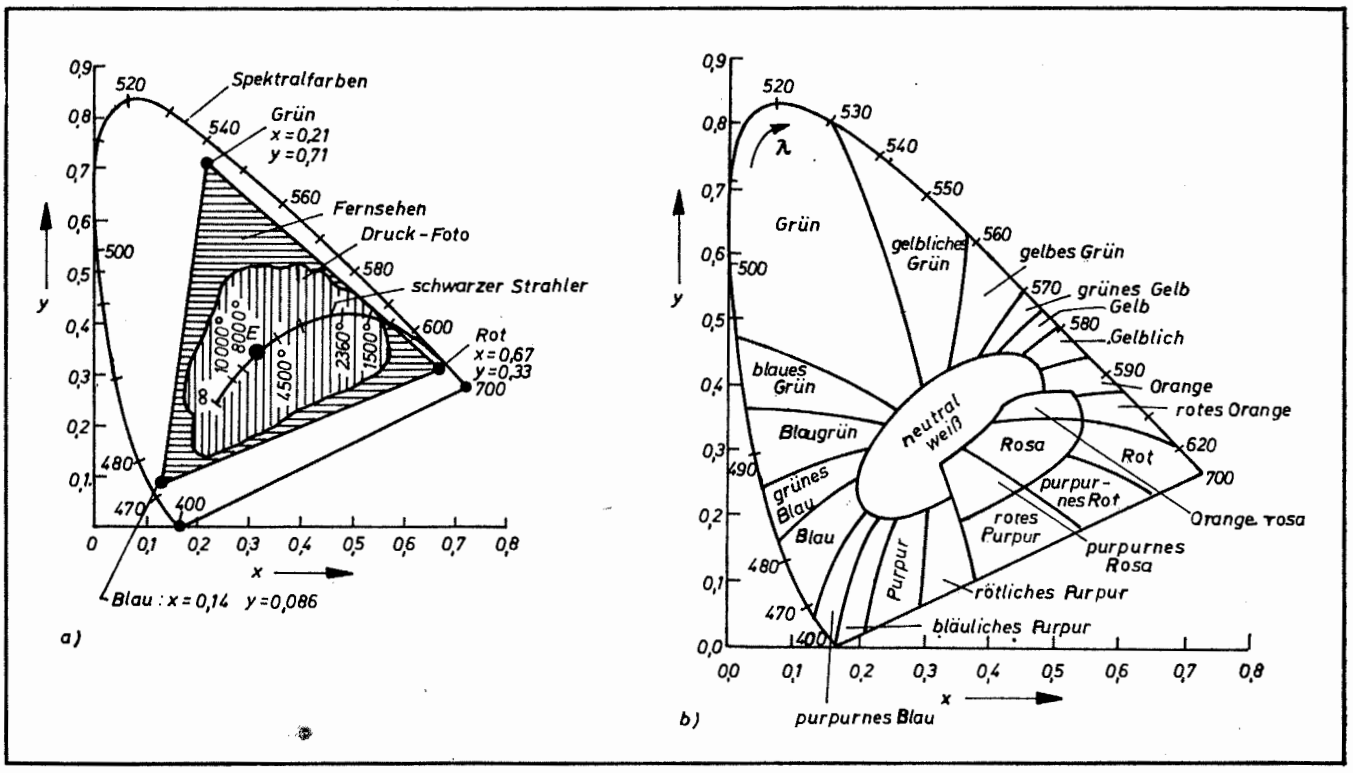


Bild 6.1.4. Farbtafel des Normalvalenzsystems

a) mit Farbarten für den schwarzen Körper und nutzbarer Fläche bei Farbdruck und Farbfotografie sowie beim Farbfernsehen,

b) mit Angabe wichtiger Farbarten

Für die *flächenhafte Darstellung* des *Farbdreiecks* werden die Werte

$$x = \frac{X}{X + Y + Z}; \qquad y = \frac{Y}{X + Y + Z} \tag{12}$$

verwendet. Die *Spektralfarben* erzeugen in dieser Fläche eine hufeisenförmige Kurve, die zugleich die Farbvalenzen einhüllt. Unten wird die Kurve durch eine Gerade zur Farbfläche abgeschlossen. Bild 6.1.4 zeigt die entsprechende Lage der *Farborte* des schwarzen Strahlers und der wichtigsten Farben.

Gesättigte Farben sind die Spektralfarben und die Purpurfarben (Farbsättigungsgrad 1). Alle übrigen Farben sind nicht voll gesättigt (Farbsättigungsgrad zwischen 0 und 1). Sie werden als weniger farbig wahrgenommen. Mit der Abnahme des Farbsättigungsgrades wird aus jeder Farbe bei genügender Leuchtdichte schließlich reines Weiß (sonst unbuntes Grau). Beispiel: Vermindert man den Farbsättigungsgrad von Rot, so ergibt sich damit Rosa, das immer blasser wird und schließlich in Weiß übergeht.

Mischungen wenigstens zweier einwelliger Strahlungen sind meistens nicht voll gesättigt. Ausnahmen sind die Purpurfarben sowie Mischungen aus Strahlungen mit Wellenlängen zwischen etwa 560 nm und 700 nm. Diese Mischungen nimmt man ebenfalls als voll gesättigte Farben wahr und bezeichnet sie deshalb auch als gesättigt. So werden beispielsweise Mischungen von gesättigt orangefarbener oder roter Lichtstrahlung mit gesättigt gelbgrüner Lichtstrahlung als gesättigt gelbe Lichtstrahlung wahrgenommen. Gesättigte Farben existieren auch in der Natur selten. Fast immer ist der Farbsättigungsgrad wesentlich kleiner als 1. Das gilt besonders für alle grünen Farbtöne und für die mittleren Purpurtöne. Somit stört es wenig, daß in der Reproduktionstechnik (Farbfotografie und Farbdruck) gesättigte Farben nicht darzustellen sind. Bei solchen Reproduktionen kann nur etwa das in Bild 6.1.4 a) senkrecht schraffierte Gebiet der Farbart-Tafel

ausgenutzt werden. Beim Farbfernsehen ist der Bereich größer (zusätzlich das waagerecht schraffierte Gebiet). Weitere Details sind u. a. in [A 10, Bd. 4] enthalten.

Bei *zeitlich schnell veränderlichen Farben*, wie z. B. beim Farbfernsehen, treten gewisse Einschränkungen des Farbsehens ein. Hierauf wird in Abschnitt 6.2.3.5. eingegangen.

6.1.1.4. Optische Darstellungen

Eine *Klassifikation* alles optisch Darstellbaren und Dargestellten ist wegen der großen Vielfalt schwierig. Eine Möglichkeit bietet die *Komplexität* über die *Anzahl* der benötigten oder verwendeten *Bildelemente* (in der Aufzählung in Klammern) und den Darstellungsbezug:

- *Signal*, z. B. Punkt oder Strich (1 — 10);
- *Zifferndarstellung* (7 — 100);
- *Alphanumerik*: Ziffern, Buchstaben und einige Sonderzeichen (10 — 3000);
- *Linienhafte Bilder* ohne Grauwerte, z. B. Strichätzungen, Kurven, Histogramme (100 — 30000);
- *Bilder mit Grauwerten*, z. B. Schwarzweißfoto (> 10000);
- *Bilder mit Farbwerten*, z. B. Farbfernsehen (> 20000);
- *Raumbild* (Stereobilder) (> 50000).

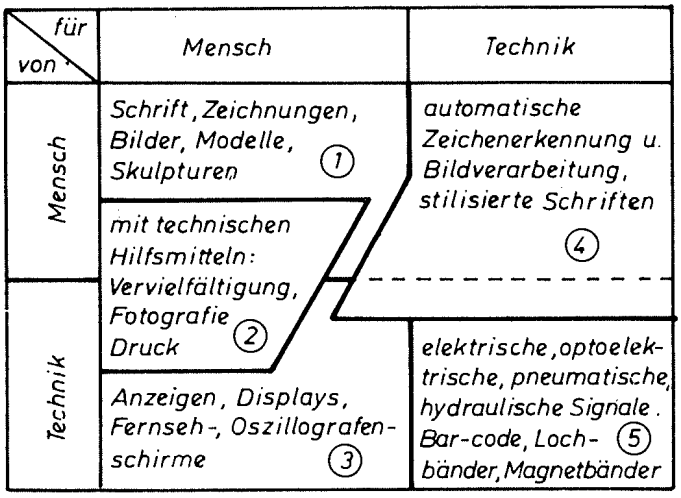


Bild 6.1.5. Einteilung optischer Darstellung in fünf Gruppen unter dem Gesichtspunkt: von wem für wen dargestellt

Eine *zweite*, hiervon unabhängige *Klassifikation* zeigt Bild 6.1.5. Das Gebiet ① ist das ursprüngliche, wird also schon so lange verwendet, wie der Mensch für sich etwas darzustellen versucht. Im Laufe der Zeit wurden technische Hilfsmittel einbezogen, so entstand ②. Die weitere Entwicklung der Technik führte zu relativ selbständigen Darstellungsprinzipien ③. Mit der Entwicklung der Technik wurde es immer schwieriger, die Daten in die Rechner bzw. Geräte einzugeben. So entstand das Gebiet ④. Eine Verkopplung von gut technisch und noch für den Menschen lesbaren Schriften, wie sie z. B. Bild 6.1.6 zeigt, war eine der Konsequenzen. Der Vollständigkeit halber sind noch die Technik-Technik-Signale ⑤ in Bild 6.1.5 eingetragen. Im Zusammenhang mit der obigen Einteilung und dem Teil ③ ist die Unterscheidung

- *Anzeige*: für Ziffern und Alphanumerik bis etwa 1000 Bildelemente;
- *Display*: für Bilddarstellungen und/oder umfangreiche Texte mit mehr als ca. 500 Bildelementen, wobei hochauflösende Displays über 100000 Bildelemente realisieren,

wichtig.

Bei den Anzeigen werden in unterschiedlicher Weise spezielle Punkt- bzw. Stricharten, wie sie Bild 6.1.7 zeigt, verwendet. Bei den Displays kommt in der Regel nur das Punktraster in Betracht. Insbesondere bei Bildröhren gibt es noch weitere Varianten, auf die später (6.2.3. und 7.7.4.) eingegangen wird.

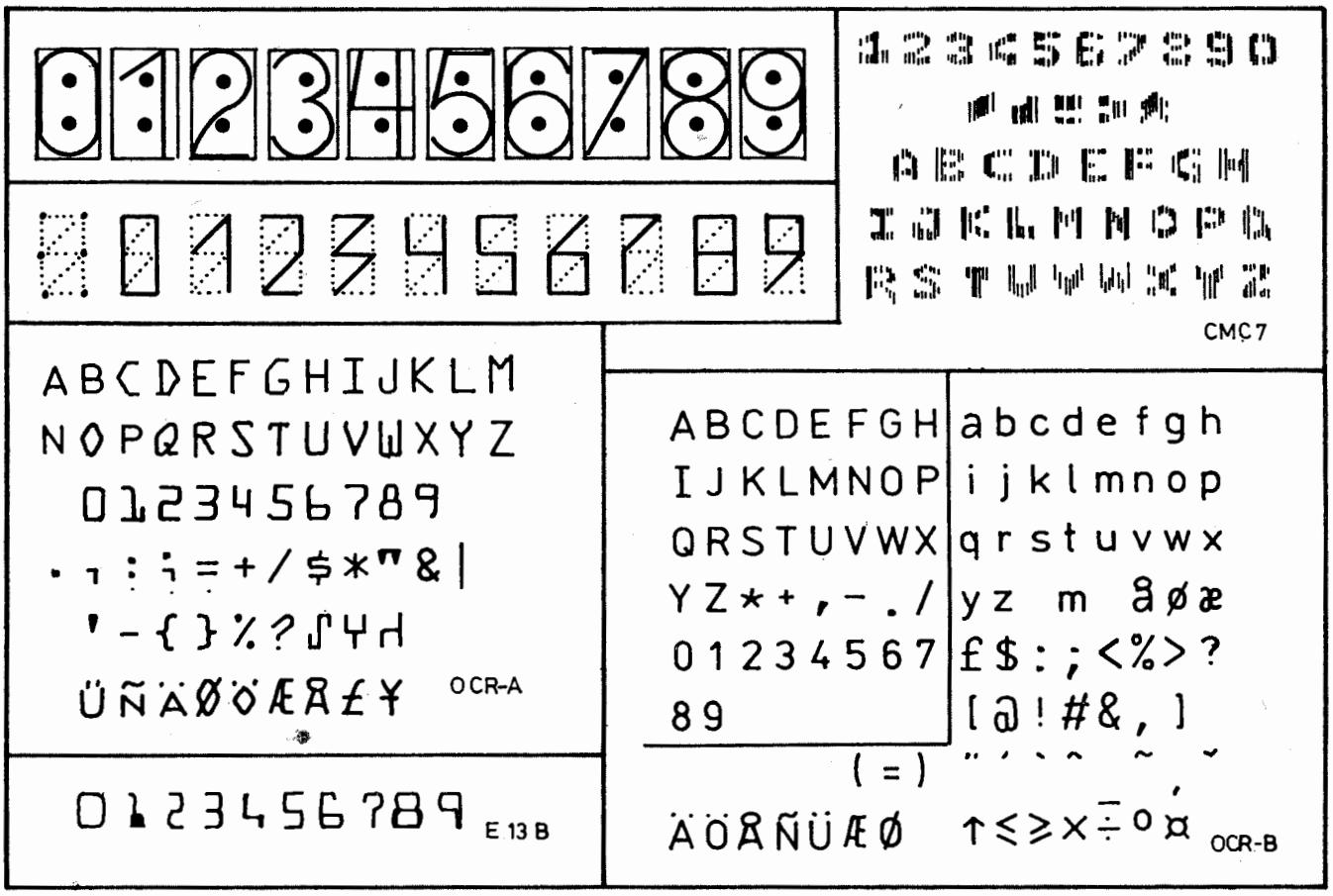


Bild 6.1.6. Stilisierte Schriften (links oben vom Menschen ins Punkt- bzw. Strichraster zu schreiben) und genormte Schriften, die für Mensch und Maschine gut lesbar sind. Bei OCR-B gibt es den abgegrenzten Subset (OCR bedeutet: optical character recognition)

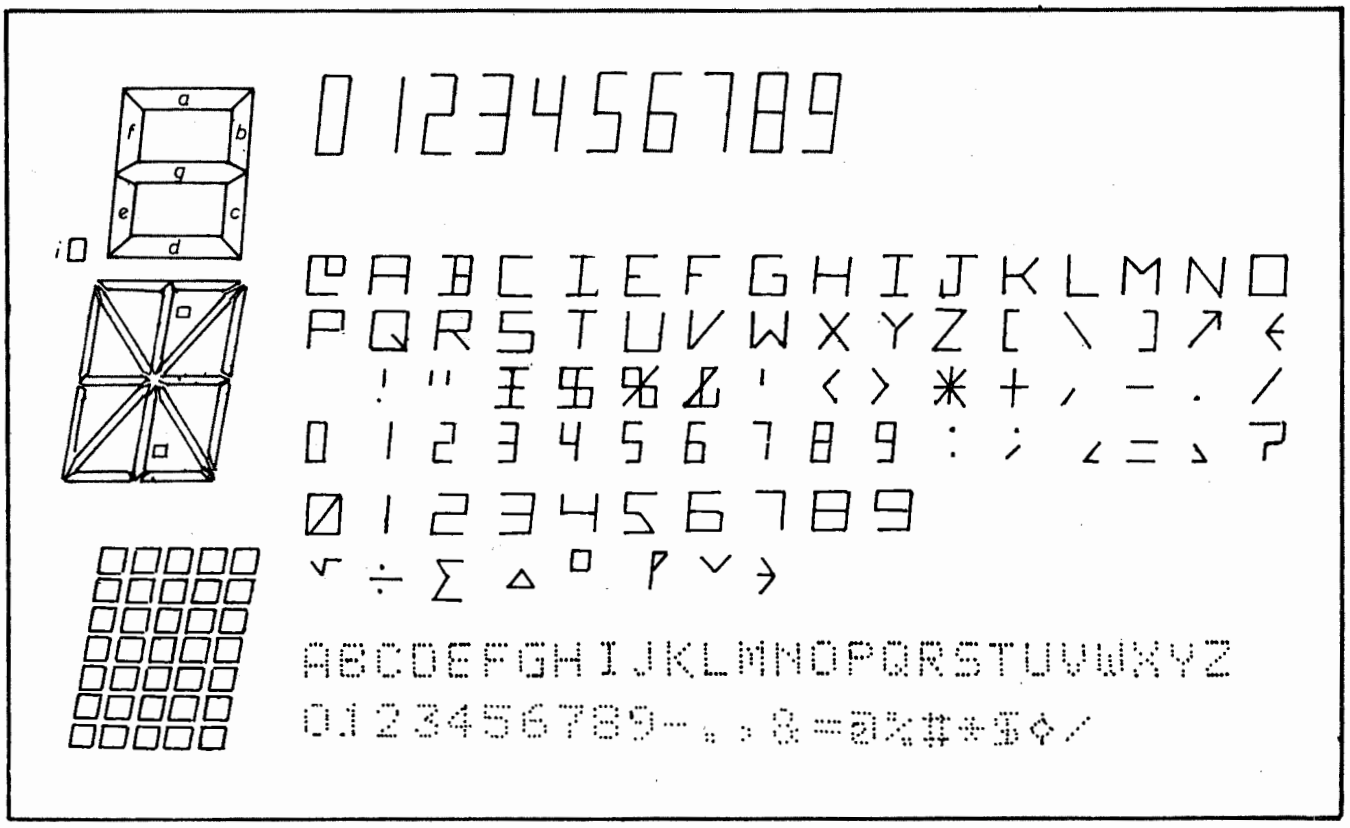


Bild 6.1.7. Die drei wichtigen Darstellungsformen von Ziffern, Buchstaben und Sonderzeichen

a) mit sieben Segmenten *a* bis *g* nur für Ziffern,

b) mit sechzehn Segmenten für Ziffern, große Buchstaben und einige Sonderzeichen,

c) mit 5 × 7 Punktraster

Die zusätzlichen Punkte in a) und b) und eventuelle zusätzliche Punkte sind unterschiedlich vorhanden

6.1.1.5. Ergonomische Grundlagen

Der Informationsfluß über unser Auge ist sehr groß. Dennoch gibt es für eine optimale Informationsaufnahme günstige Gestaltungsmöglichkeiten der einzelnen Zeichen, ihrer Größe und Anordnungen [V 11]. Einige ausgewählte Werte hierzu sind in Tab. 6.1.4 zusammengestellt. Bei fast allen Anzeigen und Displays wird die Information periodisch reproduziert. Dadurch kann ein Flimmern (Flicker) entstehen. Die Grenze hierzu zeigt Bild 6.1.8a). Bild b) ergänzt diese Betrachtungen bezüglich der Abhängigkeit der Arbeitsleistung und Ermüdung von der Beleuchtungsstärke.

Tabelle 6.1.4. Empfohlene (in Klammern gerade noch gut unterscheidbare) Richtwerte für optische Informationsdarbietungen

Wert	Größe (horizontal; vertikal)	
Bildfläche	25° ÷ 30°;	ca. 20°
Raster für Rasterdisplays	5 ÷ 8;	7 ÷ 9
Zeilenzahl	60 ÷ 100;	20 ÷ 50
Alphanumerische Zeichen	ca. 70	
Schrifttypen	1 ÷ 3	
abstrakte Symbole (Anzahl)	10 ÷ 20	
bildhafte Symbole (Anzahl)	200 ÷ 1000	
Symbolgrößen	2 ÷ 3	(5)
Höhe optimaler Zeichen	10′ ÷ 20′	
Größe komplexer Symbole	20′ ÷ 40′	
Gruppierung von Symbolen	2 ÷ 5	
Helligkeitsstufen	2 ÷ 3	(5)
Farbarten	5 ÷ 8	(11—15)
Schraffuren	ca. 4	
Skalenlängen	3° ÷ 5°	
Abstand von Skalenstrichen	10′ ÷ 15′	
Strichdicken bezogen auf Rasterbreite	10 ÷ 15%	
Stricharten	2 ÷ 3	

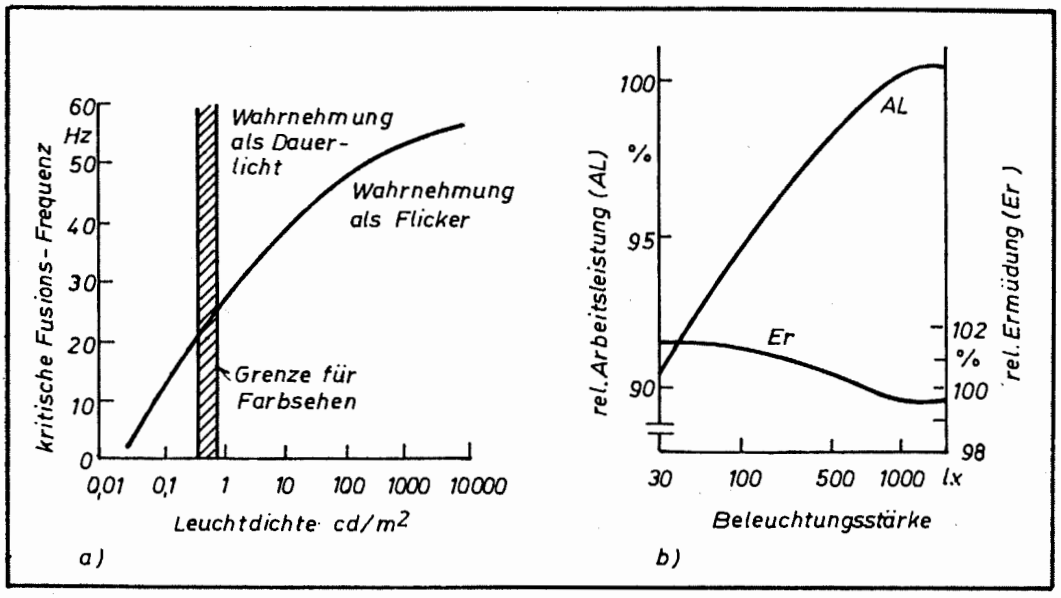


Bild 6.1.8. a) Einfluß der Leuchtdichte auf die Flimmergrenze
b) Einfluß der Beleuchtungsstärke auf die relative Arbeitsleistung AL bzw. relative Ermüdung Er

6.1.2. Halbleiter-Fotoempfänger

6.1.2.1. Allgemeine Grundlagen

Die Halbleiter-Fotoempfänger beruhen darauf, daß eindringende Photonen absorbiert werden und dabei je ein Ladungsträgerpaar, Elektron und Loch, generieren können. Erste Voraussetzung hierfür ist, daß einfallende Photonen nicht an der Oberfläche des Materials reflektiert werden. Mit dem wellenlängenabhängigen inneren Brechungsindex n_i und dem ihm gegenüberstehenden n_a gilt für den *Reflexionsanteil*

$$R = \left(\frac{n_i - n_a}{n_i + n_a}\right)^2 . \tag{13}$$

In Luft bzw. Vakuum ist $n_a \approx 1$. Für Si beträgt $n_i \approx 3{,}44$. So werden 30% des einfallenden Lichtes reflektiert. Dieser Verlust läßt sich durch eine $\lambda/4$ dicke *Antireflexschicht* fast vollständig vermeiden, sofern gilt

$$n_R = \sqrt{n_i n_a} . \tag{13a}$$

Hierfür sind z. B. Filme aus SiO, SiO_2 oder Si_3N_4 geeignet.

Von dem einfallenden Licht der Intensität I_0 wird je durchdrungene Schichtdicke stets der gleiche Anteil absorbiert. Für die in der Tiefe x vorhandene Intensität I_x gilt daher mit der material- und wellenlängenabhängigen *Absorptionskonstanten* α die Beziehung

$$I_x = I_0 \exp(-\alpha x) . \tag{14}$$

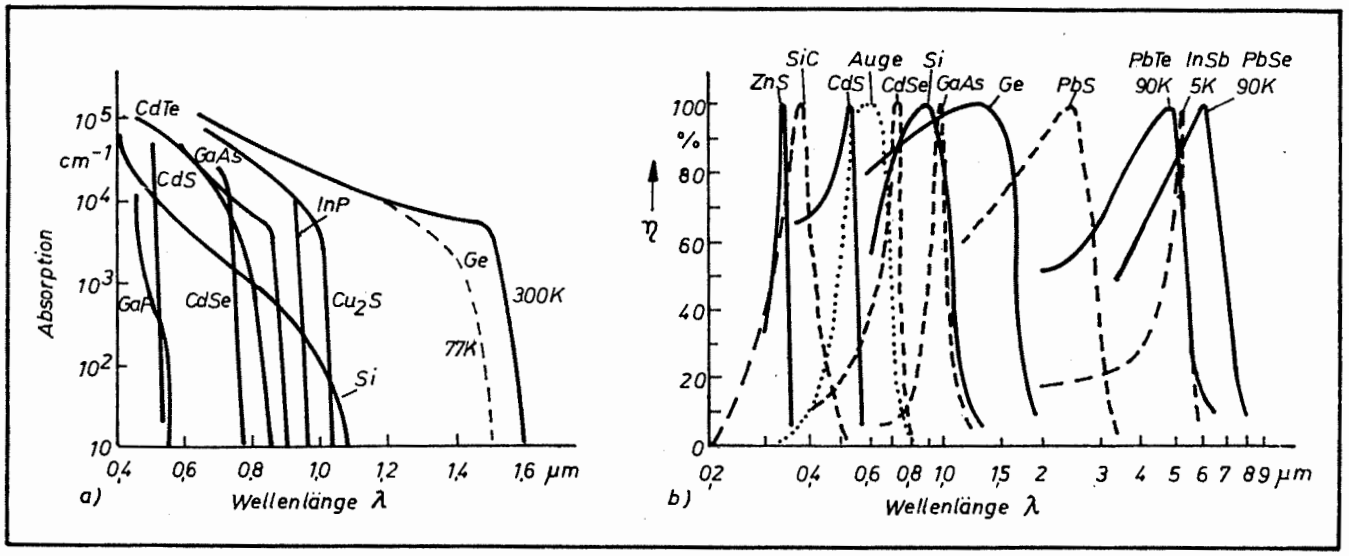


Bild 6.1.9. a) Absorption und b) relativer Wirkungsgrad verschiedener Halbleitermaterialien als Funktion der Wellenlänge

Der reziproke Wert von α gibt also jene Dicke an, in der vom Licht $1 - 1/e \approx 63{,}2\%$ absorbiert werden. Sie wird auch als *Eindringtiefe* bezeichnet. Den Verlauf von α zeigt Bild 6.1.9a). Besonders auffällig ist dabei die *Bandkante*. Beginnend von großen Wellenlängen (tiefen Frequenzen), wird das Material hier ziemlich steil undurchsichtig; es absorbiert stark. Diese Bandkante λ_g steht in Zusammenhang mit dem Bandabstand ΔW der Materialien. Aus (2) folgt unmittelbar

$$\lambda_g \approx hc/\Delta W \tag{15}$$

bzw.

$$\lambda_g/\mu\text{m} \approx 1{,}24/\Delta W/\text{eV} \tag{15a}$$

Die Energie eines Photons ist bei der Bandkante (Grenzwellenlänge) also gerade so groß, daß es einem Elektron die Energie zur Überwindung des Bandabstandes erteilen kann. Infolge der außerdem vorhandenen thermischen Energie verschiebt sich die Bandkante etwas mit der Temperatur. Dies ist im Bild 6.1.9a) für Ge angedeutet. Wenn das Photon eine kürzere Wellenlänge, also größere Energie, hat, kann das Elektron noch *Bewegungsenergie* gemäß

$$\tfrac{1}{2} m_e v^2 = h\nu - \Delta W \tag{16}$$

aufnehmen. Dieser Teil wird bei den Fotoelementen als nutzbare Spannung frei.

An der Oberfläche von Halbleitermaterialien sind stets Störungen unterschiedlicher Art wirksam. Anderseits dringen kurze Wellen nur sehr wenig tief in das Material ein. Dadurch wird die für das Material nutzbare Lichtenergie nach kurzen Wellen wieder geringer. So entsteht insgesamt ein relativer Wirkungsgrad bezüglich Lichteinstrahlung, wie ihn Bild 6.1.9b) zeigt.

6.1.2.2. Detektivität

Vor allem zwei *Kriterien* bestimmen die *Güte eines Fotodetektors*:

- Größe des Signals bezüglich der einfallenden Strahlungsleistung,
- Störabstand des Signals vom Rauschen.

Der erste Punkt führt zur *Empfindlichkeit R* des Empfängers und ist durch das Verhältnis

$$R = S/P \tag{17}$$

gegeben, wobei P die einfallende Leistung in W und S das Signal, je nach Art der Signalauswertung z.B. in V oder A, ist. Dieses Verhältnis wird später bei den konkreten Bauelementen genauer betrachtet. Für den zweiten Punkt sind die *Rauschquellen* des Fotoempfängers ähnlich wie im Abschnitt 5.7.1. zu untersuchen. Es treten wieder auf

- thermisches Rauschen des Widerstandes,
- Schrotrauschen des Stromes,
- Generations-Rekombinations-Rauschen,
- $1/f$-Rauschen,
- statistische Schwankung des Quantenflusses des Lichtes und des thermischen Hintergrundes.

Alle diese Rauschquellen (außer $1/f$-Rauschen) führen zu einem Rauschsignal (in Strom oder Spannung), das proportional der Quadratwurzel aus der Bandbreite B ist,

$$S_R = R_{R_0} \sqrt{B}\,. \tag{18}$$

Die für den Störabstand 1 (Signal gleich Rauschen) notwendige bezogene Leistung in $W/\sqrt{Hz}$ heißt in englisch *Noise Equivalent Power* (NEP). Für sie gilt folglich

$$\mathrm{NEP} = \frac{P}{\sqrt{B}} \frac{S_R}{-S}. \tag{19}$$

Bei vielen Empfängern ist sie darüber hinaus der Wurzel aus der *aktiven Fläche A* proportional. Deshalb wurde als spezifische Größe die *Detektivität* gemäß

$$D^* = \frac{\sqrt{A}}{\mathrm{NEP}} = \frac{R}{S_R} \sqrt{BA} \tag{20}$$

eingeführt. Sie besitzt die Einheit $\mathrm{cm}\sqrt{\mathrm{Hz}}/\mathrm{W}$. Werte für die Detektivität enthält Bild 6.1.10. Der ideale Fotodetektor ist nur durch das Quanten- und thermische Rauschen begrenzt.

Weitere Details zu den ersten beiden Abschnitten enthält u. a. [B 30].

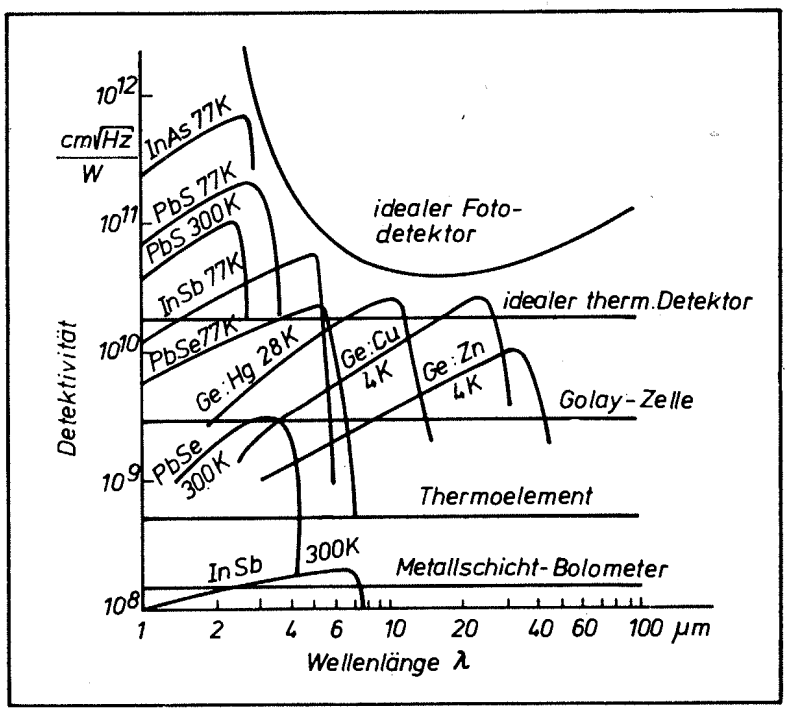


Bild 6.1.10. Detektivität von Infrarotstrahlungsempfängern in Abhängigkeit von der Lichtwellenlänge

6.1.2.3. Einteilung der Fotoeffekte

Es gibt eine Vielzahl von Lichtwirkungen. Hier interessieren nur spezielle elektrische Wirkungen. Sie werden in zwei Klassen geteilt:

- *Äußerer Fotoeffekt.* Bei ihm verlassen Elektronen die Oberfläche von Festkörpern unter Lichteinfluß. Er wird im Abschnitt 6.2., *Vakuumelektronik,* behandelt.
- *Innere Fotoeffekte* finden im Material statt und können weiter unterteilt werden:
  - Im *homogenen Material* wird unter Lichteinfluß der spezifische Widerstand geändert. Dies ist Grundlage der im Abschnitt 6.1.2.4. behandelten Fotowiderstände.
  - An *inneren Grenzflächen,* hier speziell *pn*-Übergängen. Dabei sind zwei Anwendungsfälle des gleichen Bauelementes Fotodiode zu beachten:
    - *Fotoelemente* liegen dann vor, wenn die Lichtenergie mit möglichst hohem Wirkungsgrad direkt in elektrisch nutzbare Energie verwandelt werden soll. Ein Sonderfall sind die Solarzellen.
    - *Fotoempfänger* nutzen dagegen die Lichtenergie nur zur Beeinflussung vorhandener Stromkreise. Sie existieren als Fotodioden, -transistoren und -thyristoren.
  - In der *Raumladezone* bei MOS-ähnlichen Bauelementen. Die wichtigste Anwendung ist hier der CCD-Sensor. Er wird im Abschnitt 6.1.5. behandelt.

6.1.2.4. Fotowiderstände

Es gibt zwei Typen von Fotoleitern, die aus *eigenleitendem* oder *störstellenleitendem* Material bestehen. Vom Aufbau her können sie polykristallin (z.B. gedampft, gesputtert, gesintert) und einkristallin sein. Durch den Lichteinfall werden zwei Mechanismen wirksam:

- Erhöhung der Ladungsträgerzahl durch Fotogeneration,
- Erhöhung der effektiven Beweglichkeit der Ladungsträger.

Tabelle 6.1.5. Eigenschaften einiger Fotowiderstände. Die mit * versehenen Materialien sind Störstellenhalbleiter, die anderen gehen von der Eigenleitung aus.

Material	Betriebstemperatur K	Empfindlichkeit µm λ_m bzw. Bereich bis 50% Abfall	Detektivität cm $\sqrt{\text{Hz}}$/W	Dunkelwiderstand Ω	Ansprechzeit s
PbS	295	2,1 (0,9 ÷ 2,8)	10^{11}	10^6	10^{-4}
CdS	295	0,5	$3 \cdot 10^{14}$	10^8	$10^{-4} \div 5 \cdot 10^{-2}$
PbSe	295	3,4 (0,5 ÷ 4,2)	$4 \cdot 10^{10}$	$2 \cdot 10^6$	$2 \cdot 10^{-6}$
ZnSe	300	0,5			10^{-2}
InAs	295	3,6	$1 \cdot 10^8$		$2 \cdot 10^{-2}$
InSb	295	3,6 ÷ 7,3	$3 \cdot 10^8$	20	10^{-8}
InSb	77	3,0 ÷ 5,6	10^{11}	10^4	10^{-5}
Se	295	0,7			10^{-3}
TlS	295	1,05			10^{-3}
Sb_2S	295	0,8			10^{-3}
* Ge : Au	77	3 ÷ 9	10^{10}	$4 \cdot 10^5$	$3 \cdot 10^{-8}$
* Ge : Hg	27	6 ÷ 14	$5 \cdot 10^{10}$	10^5	10^{-7}
* Ge : Cu	4,2	12 ÷ 27	$4 \cdot 10^{10}$	$2 \cdot 10^4$	$2 \cdot 10^{-9}$

Der zweite Effekt ist bei polykristallinem Material besonders groß. Zwischen den Körnern bestehen hier Barrieren, die durch den Lichteinfall gesenkt werden und so die effektive Beweglichkeit erhöhen. Besonders bei diesen polykristallinen Materialien ist die *Zeitkonstante* relativ groß. Vom CdS abgesehen (fällt etwa mit der Augenempfindlichkeit zusammen, deshalb u. a. Belichtungsmesser), werden Fotowiderstände vor allem im langwelligen IR eingesetzt. Hier ist zur Reduzierung der thermischen Generation oft Kühlung bis zu tiefen Temperaturen notwendig.

Einen Überblick zu wichtigen Materialien für Fotowiderstände gibt Tab. 6.1.5.

Bei der Einschätzung der *Leistungsfähigkeit* von Fotowiderständen sind *zwei Zeiten* entscheidend:

T_1: *Transitzeit*, in der die Ladungsträger unter der Feldwirkung der angelegten Spannung von der einen zur anderen Elektrode gelangen;

T_2: *Rekombinationszeit*, in der vorhandene Ladungsträger rekombinieren, also verschwinden.

Dabei ist zu beachten, daß die Elektroden so lange Ladungsträger nachliefern, wie eine Störstelle ionisiert bleibt. Deshalb ist die *Verstärkung*

$$M = T_2/T_1 \tag{21}$$

eingeführt. Sie gibt an, um wieviel der Fotoleitstrom größer als der primäre Fotostrom ist. Wenn die Elektroden sehr eng beieinander liegen und hohe Felder wirksam sind (dabei ist auch der Sättigungswert der Geschwindigkeit zu beachten), können durchaus für tiefe Frequenzen Verstärkungen von 10^5 und mehr realisiert werden. Es läßt sich zeigen, daß auch hier wieder ein *Bandbreiten-Verstärkungsprodukt* auftritt. Es liegt bei ca. 1 GHz [U 5].

Bild 6.1.11 zeigt schließlich den typischen Aufbau eines Fotowiderstandes, seine Schaltung und eine Kennlinie. In der Regel sind Spannungen von ca. 100 bis 500 V und Belastungen von ca. 0,05 bis 1 W möglich. Es werden dann Empfindlichkeiten von ca. 10^4 bis 10^5 V/W erreicht. Für ca. 50 lx soll der Widerstand etwa auf $^1/_{1000}$ des Dunkelwiderstandes abfallen. Es werden vielfach insgesamt 8 bis 12 Zehnerpotenzen vom Widerstand überstrichen. Mit einem Exponenten γ, der zwischen 0,5

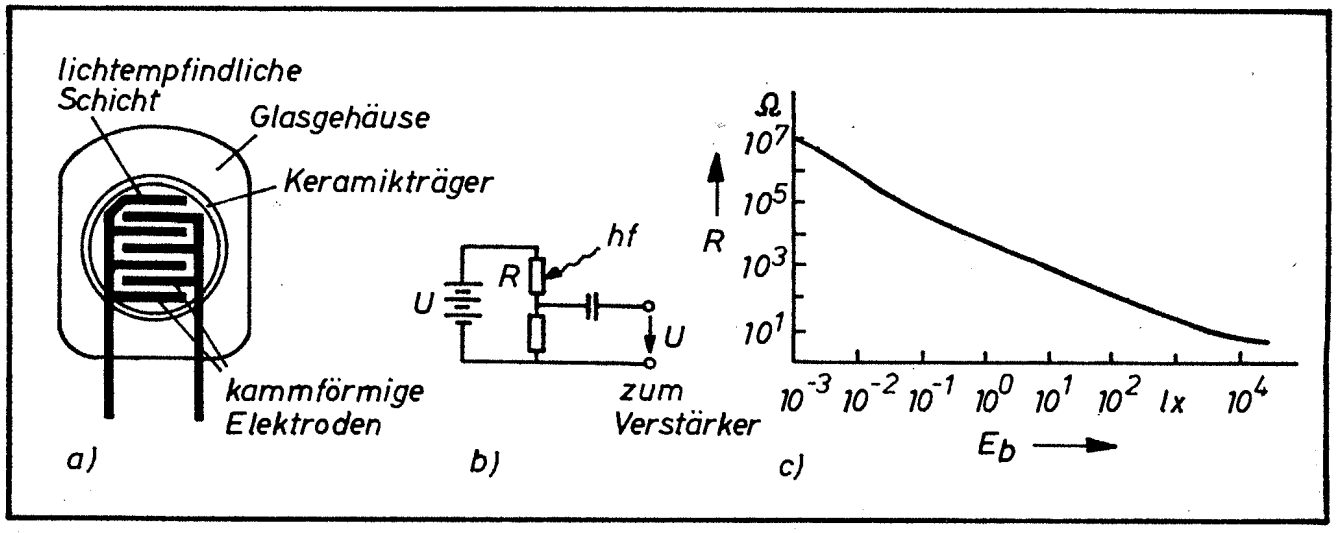


Bild 6.1.11. Zum Fotowiderstand

a) Prinzipieller Aufbau: In einem Gehäuse befindet sich ein Keramikträger, der mit der lichtempfindlichen Schicht z.B. bedampft ist. Darüber sind kammförmige Elektroden angebracht (Interdigitalstruktur), welche eine kurze Transitzeit realisieren. Ein durchsichtiger Schutz passiviert die Schicht

b) Prinzipielle Meßschaltung bei Fotowiderständen

c) Typischer Verlauf einer Widerstands-Licht-Kurve

und 1 liegt, gilt vielfach

$$R/R_D \sim E^{-\gamma}. \tag{22}$$

6.1.2.5. Fotoelemente

Bei Fotodioden und Fotoelementen wirkt der *pn-Übergang* so, daß die von Licht *generierten Ladungsträger getrennt* werden. Dadurch entsteht bei Kurzschluß ein *Fotostrom* I_K. Er addiert sich zum normalen Diodenstrom (vgl. (3.2.5)), so daß gilt

$$I = I_S \cdot \left(\exp\left(\frac{eU}{kT}\right) - 1\right) - I_K, \tag{23}$$

wobei der Kurzschlußstrom linear mit der Bestrahlungsleistung wächst. Für Si werden etwa $8 \cdot 10^{-9}$ A/lx mm² erreicht. Hierdurch entsteht eine additive Verschiebung, wie sie schon beim Transistor infolge des Emitterstromes im Zusammenhang mit Bild 3.4.5d) beschrieben wurde. (Dort ist $A_E I_E$ durch I_K zu ersetzen.) So entstehen die *Kennlinienfelder der Fotodiode* von Bild 6.1.12a), und es sind zwei wesentlich unterschiedliche *Betriebsfälle* zu betrachten:

- mit Hilfsspannung U_B als Lichtempfänger (I. Quadrant),
- ohne Hilfsspannung als elektrische Energiequelle (II. Quadrant).

Bauelemente, die für den ersten Betrieb ausgelegt sind, heißen Fotodioden, die für den zweiten Fotoelemente. Bei den *Fotoelementen* ist der *energetische Wirkungsgrad* entscheidend, bei den *Fotodioden* die Empfindlichkeit im Sinne der *Detektivität* bzw. der Wandlung, z.B. in μA/lm, und die höchste Übertragungsfrequenz. In einigen Fällen sind Bauelemente für beide Anwendungsfälle geeignet. Für die Fotoelemente interessiert weiter die *Leerlaufspannung*, sie ist aus (23) mit $I = 0$ zu berechnen, und es folgt

$$U_L = \frac{kT}{e} \ln\left(\frac{I_L}{I_S} + 1\right). \tag{24}$$

Sie nimmt also im Gegensatz zum Strom nur linear mit der Bestrahlungsleistung zu (Bild 6.1.12b)) und bleibt immer kleiner als der *Bandabstand*. Deshalb haben Foto-

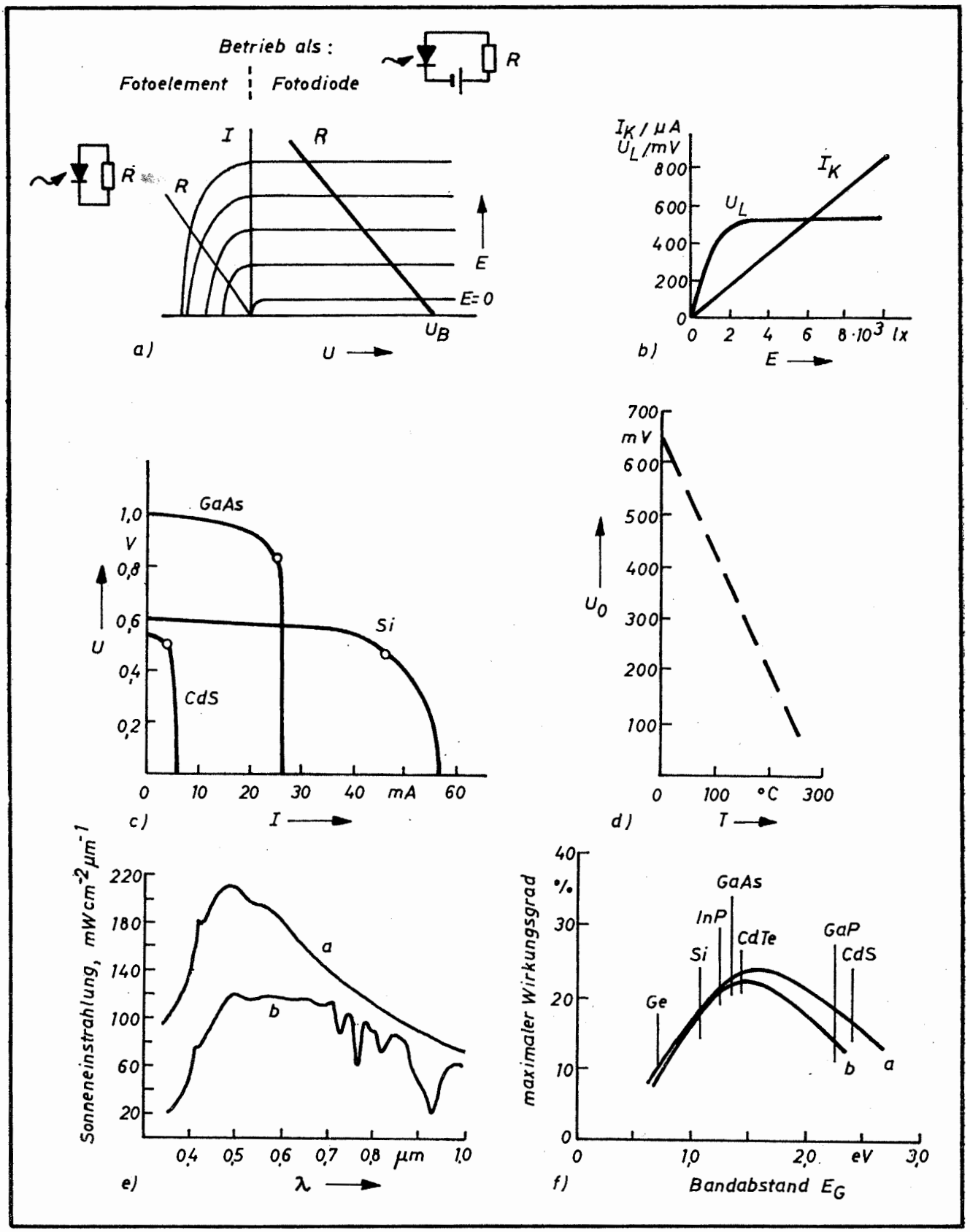


Bild 6.1.12. Zur Anwendung der Fotodiode als Fotoelement

a) Die beiden möglichen Betriebsweisen im Zusammenhang mit den Kennlinienscharen
b) Kurzschlußstrom und Leerlaufspannung als Funktion der Beleuchtung bei einem Si-Fotoelement
c) Kennlinienvergleich von drei verschiedenen Fotoelementen mit Punkten der maximalen Leistungsabgabe
d) Abhängigkeit der Leerlaufspannung einer Si-Fotodiode von der Temperatur
e) Sonnenspektrum außerhalb der Erdatmosphäre (a) und an der Erdoberfläche bei klarem Himmel (b)
f) Maximaler Wirkungsgrad von Solarzellen bei extraterrestrischer (a) und terrestrischer Sonnenbestrahlung (b) in Abhängigkeit vom Bandabstand des verwendeten Materials

elemente mit höherem Bandabstand Vorteile. Er darf jedoch nicht zu groß werden, da sich sonst die Empfindlichkeit zu sehr in den ultravioletten Bereich verschiebt.

Weiter ist der Temperatureinfluß des *Reststromes* bedeutsam. Dies geht aus Bild 6.1.12d) hervor. Unter diesen Bedingungen sind die Kennlinien der drei verschiedenen Fotoelemente gemäß Bild 6.1.12c) verständlich. Da die Hauptanwendung der Fotoelemente die *Stromversorgung* ist, interessiert vor allem die abgegebene *Leistung*. Es muß also auf den Kennlinien jener Punkt (Arbeitswiderstand R) bestimmt werden, wo $I \cdot U$ sein Maximum hat. Diese Punkte sind in Bild 6.1.12c) angedeutet. Als Ausgangspunkt für die Fotoelemente dient in der Regel weiter *Sonnenlicht*. Die entsprechenden Fotoelemente werden auch *Solarzellen* oder Sonnenbatterien genannt. Die spektrale Verteilung des Sonnenlichtes im Weltraum und auf der Erde zeigt Bild 6.1.12e). Die Leistungsdichte liegt im Weltraum bei 135 mW/cm² und auf der Erde, bei klarem Himmel, bei 106 mW/cm² und sinkt bei mittlerer Bewölkung auf 12 mW/cm² ab.

Tabelle 6.1.6. Ausgewählte Solarzellen. Stand 1980

Zelle	Struktur	Diodentyp	Maximal gemessener Wirkungsgrad in Prozent
Si	Einkristall	*pn*-Homokontakt	19
Si	Einkristall	SCHOTTKY-Kontakt	12
Si	Polykristall	*pn*-Homokontakt	14
Si	amorphe Schicht	SCHOTTKY-Kontakt	5,5
GaAs	Einkristall	*pn*-Homokontakt	15
GaAs-GaAlAs	Einkristall	*pn*-Heterokontakt	21
CdS-InP	Einkristall	*pn*-Heterokontakt	14
CdS-$CuInSe_2$	Einkristall	*pn*-Heterokontakt	12
CdS-$CuInSe_2$	dünne polykristalline Schicht	*pn*-Heterokontakt	6,9
Cu_2S-CdS	dünne polykristalline Schicht	*pn*-Heterokontakt	9,1
CdTe-CdS	dünne polykristalline Schicht	*pn*-Heterokontakt	5
CdSe	dünne polykristalline Schicht	SCHOTTKY-Kontakt	in Entwicklung

Unter diesen Voraussetzungen kann jetzt der Wirkungsgrad berechnet werden. Er ist für verschiedene Materialien in Bild 6.1.12f) dargestellt. Die größte Anwendungschance hat immer noch Si. Dies liegt u. a. an der gut beherrschten Technologie. Weiter ist von Vorteil, daß infolge der niedrigen Sperrspannungen auch niederohmiges Material angewendet werden kann. Jedoch sind die Einkristalle sehr teuer. Deshalb gibt es vielfältige Untersuchungen mit polykristallinem Material. Besonders ökonomisch scheinen hier die Heterokontakte, z. B. p-Cu_2S auf n-CdS, zu sein. Während im Weltraum die Fotoelemente bereits unentbehrlich sind, haben sich auf der Erde noch keine größeren Anwendungen durchgesetzt. Jährlich sollen z. Z. weltweit etwa 50 kW installiert werden. Die dazu benötigte Fläche ist relativ groß. 100 mW/cm² entsprechen etwa $6 \cdot 10^7$ Photonen/cm², und es wird dabei ein Fotostrom (Si) von ca. 35 mA/cm² realisiert. Tab. 6.1.6 gibt einen Überblick zu dem Stand der Solarzellen um 1980.

6.1.2.6. Fotodioden

Bei den Fotodioden kommt es im Gegensatz zu den Fotoelementen nicht auf die Leistungsbilanz, sondern vor allem auf *zwei* andere *Eigenschaften* an:

- die Empfindlichkeit im Sinne von Detektivität,
- eine möglichst hohe Grenzfrequenz.

Deshalb sind Fotodioden im Gegensatz zu den meist großflächigen Fotoelementen *mit kleiner Fläche* realisiert. Dadurch sinkt die Kapazität. Sie sinkt um so mehr, wie im Abschnitt 3.2.4. beschrieben, je stärker die Dioden in Sperrichtung betrieben werden (vgl. Bild 6.1.12a), rechte Seite). Ähnlich wie bei den Fotoelementen gibt es auch verschiedene Ausführungen von Fotodioden. Sie sind in Bild 6.1.13 dargestellt. Sie unterscheiden sich nicht nur in der Struktur, sondern auch darin, wo Photonen die Trägerpaare erzeugen und wie sie zu den Elektroden wandern. Konstruktiv besonders einfach ist die *pn*-Fotodiode. Bei der *pin-Diode* ist der Raumladebereich so weit ausgedehnt, daß praktisch *alle Photonen im Raumladegebiet absorbiert* werden. Mittels

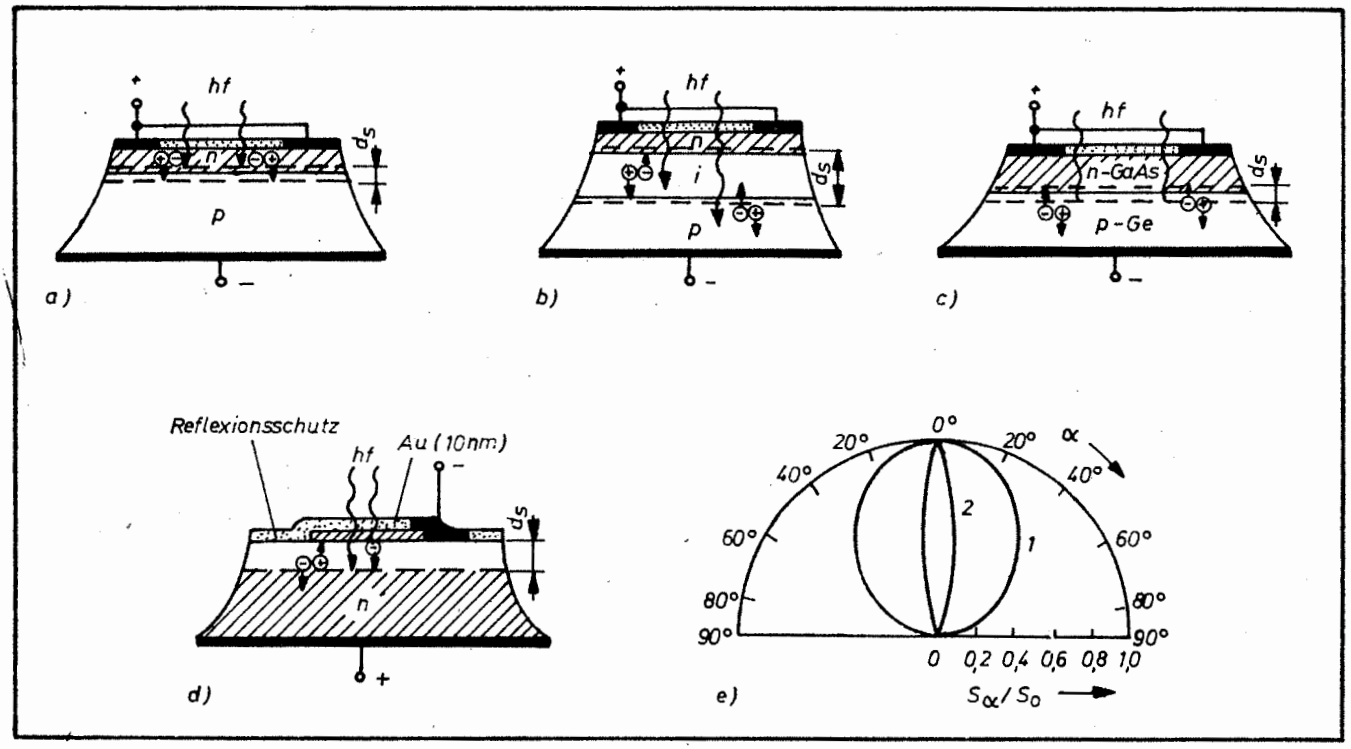


Bild 6.1.13. Fotodioden werden in der Regel von oben belichtet. Es ist aber auch möglich, sie von der Seite zu belichten (a), wenn dies entsprechend konstruktiv vorgesehen ist. Konstruktiv können folgende Typen unterschieden werden:

a) mit *pn*-Übergang, b) mit *pin*-Struktur,
c) Hetero-Fotodiode, d) mit Metall-Halbleiter-Übergang (SCHOTTKY),
e) Richtdiagramm zweier Fotodioden: ohne (1) und mit Linsen (2) (d_s ist die Raumladezone)

der Dicke der *i*-Schicht ist es darüber hinaus möglich, sie für sehr *hohe Übertragungsfrequenzen* anwendbar zu machen. Si-*pin*-Dioden haben oft Ansprechzeiten unter 1 ns, während *pn*-Dioden meist den über hundertfachen Wert besitzen. Mit den Hetero- und SCHOTTKY-Dioden erfolgt meist eine Anpassung an bestimmte Lichtwellenlängen. Für *Infrarot* hat die *Ge-Fotodiode* noch Bedeutung. Die *pin*-Diode besitzt weiter eine besonders gute Linearität zwischen Beleuchtung und Strom. Die Empfindlichkeit der Fotodioden liegt bei etwa 30 mA/lx und der Dunkelstrom bei 10 ... 100 nA, bei *pin*-Dioden 0,1 ... 1 nA.

6.1.2.7. Weitere Fotoempfänger

Die *Foto-Avalanche-* oder *Foto-Lawinendiode* wird in Sperrichtung in der Nähe des Durchbruchs betrieben. Hierdurch werden von Licht erzeugte Ladungsträger im

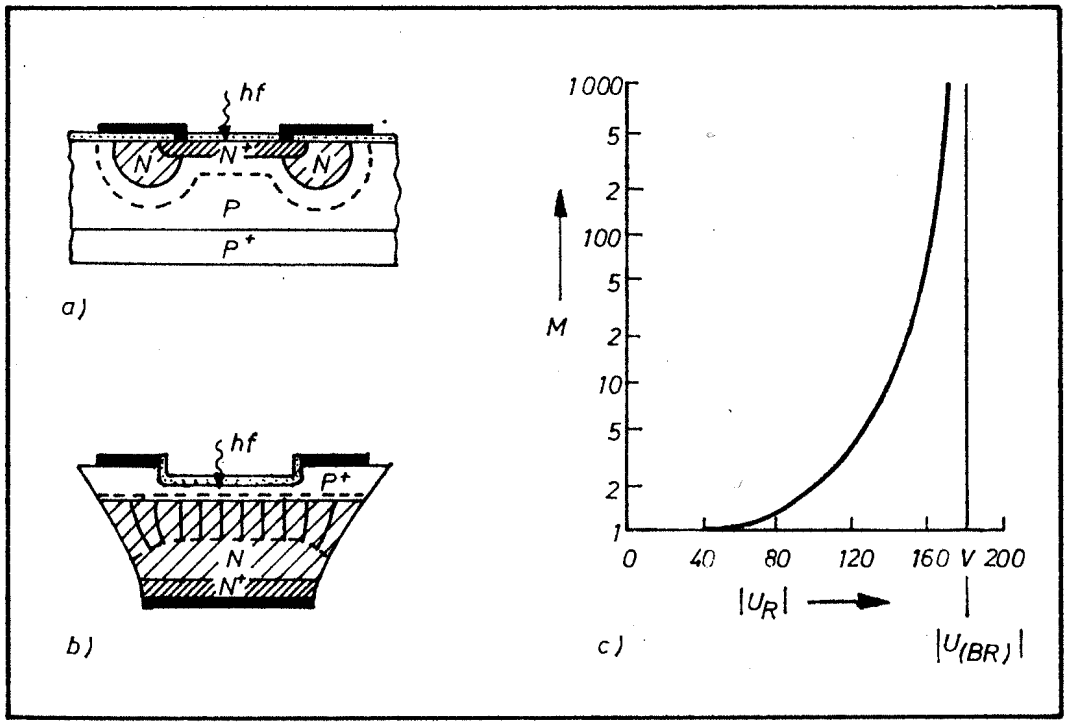


Bild 6.1.14. Lawinenfotodioden

a) mit *n*-Schutzring, b) in abgeschrägter Mesaform

c) Abhängigkeit des Durchbruch- oder Verstärkungsfaktors M von der Betriebsspannung bei einer Durchbruchspannung von 180 V

Raumladungsgebiet so beschleunigt, daß sie durch Stoßionisation weitere Ladungsträger aus den Gitterplätzen reißen. So entsteht ein Durchbruch- bzw. Vervielfachungsfaktor M mit Werten zwischen 50 und 500 (Bild 6.1.14c)). Die Empfindlichkeit wird dabei auf etwa 1 μA/lx erhöht. Gleichzeitig entsteht eine hohe Grenzfrequenz (kleine Sperrschichtkapazität und hohe Trägergeschwindigkeit). Es werden Bandbreiten-Verstärkungsprodukte um 100 GHz erreicht. Damit im gesamten *pn*-Übergang gleichzeitig der Durchbruch erfolgt, sind spezielle *pn*-Strukturen notwendig. Zwei Beispiele hierfür zeigt Bild 6.1.14a) und b).

Eine andere Art der Empfindlichkeitssteigerung ist mittels des *Fototransistors* möglich. Seinen prinzipiellen Aufbau zeigt Bild 6.1.15a). In vielen Fällen wird die

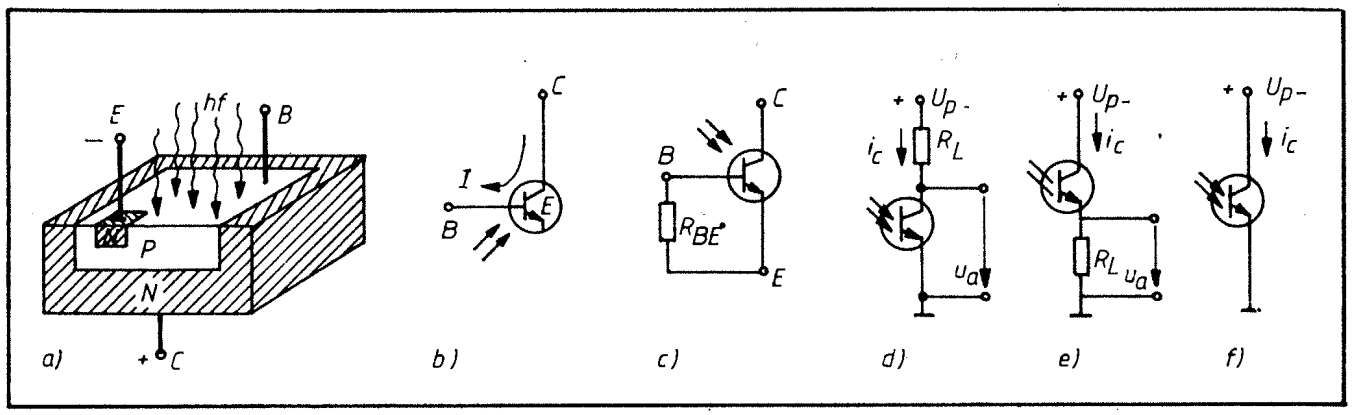


Bild 6.1.15. Zum Fototransistor

a) Struktureller Aufbau

b) Schaltung als Fotodiode

c) Mit überbrückter Basis-Emitter-Strecke

d) Mit R_L im Kollektorzweig, e) im Emitterzweig

f) Nur zur Strommessung

Die Verstärkung des primären Fotostromes nimmt zu von b) über c), e), d) nach f). Im selben Maße sinkt die Grenzfrequenz

Basis dabei nicht angeschlossen. Der primäre Fotostrom wird in der Basis-Kollektor-Strecke erzeugt. Er wird dann in der Emitterstrecke entsprechend der Stromverstärkung des Transistors bis über tausendfach vervielfacht, so daß Empfindlichkeiten erreicht werden, die über denen der Foto-Lawinendiode liegen. Typische Werte sind etwa 3 μA/lm. Im Gegensatz zur Lawinendiode ist die Grenzfrequenz des Fototransistors wesentlich geringer als bei einer normalen *pn*-Diode. Es treten hier dieselben Effekte auf wie beim Transistor in Emitterschaltung (vgl. Bild 5.6.6b)). Dementsprechend kann der Fototransistor auch in verschiedenen Schaltungen mit unterschiedlicher Empfindlichkeit und Verstärkung betrieben werden (Bild 6.1.15b) bis f)).

Es sei erwähnt, daß mit Fotowiderständen Stromempfindlichkeiten von ca. 50 μA/lm üblich sind.

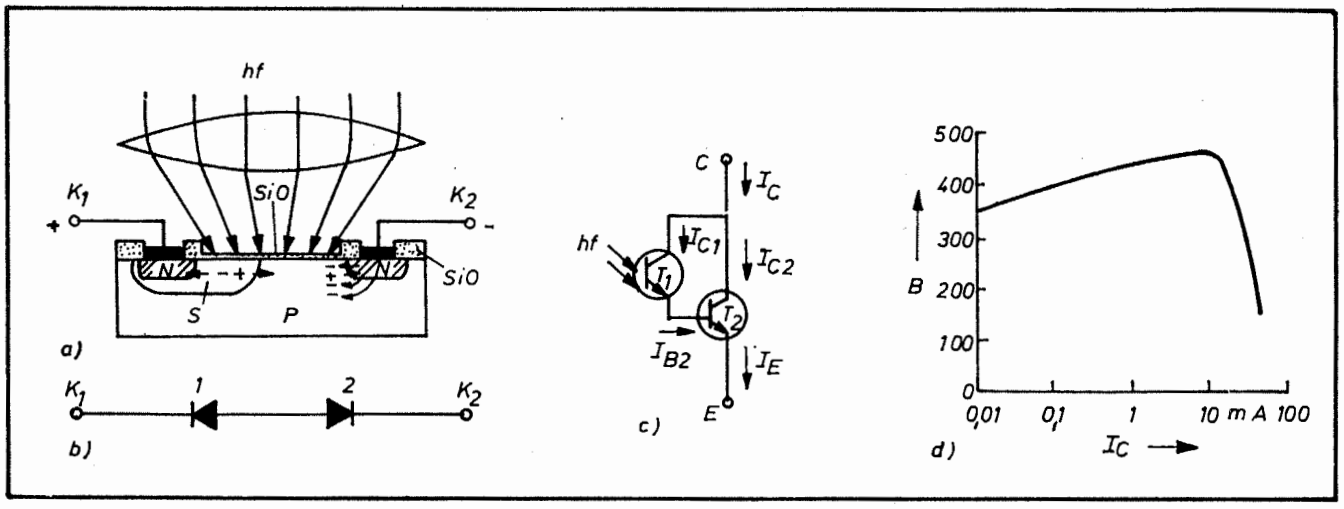


Bild 6.1.16. a) Struktur einer Fotoduodiode mit Linse und b) die Ersatzschaltung
c) Aufbau eines DARLINGTON-Fototransistors sowie d) Verlauf der Stromverstärkung

Ein Fotoempfänger besonderer Art ist die *Fotoduodiode* gemäß Bild 6.1.16. Hierbei handelt es sich um zwei Dioden, die gegeneinander geschaltet sind. Da die beiden *pn*-Übergänge weit voneinander entfernt sind, tritt keine Wechselwirkung zwischen ihnen auf. Im gewissen Sinne kann die Struktur auch als FET interpretiert werden. Die Wirkung des fehlenden Gates übernimmt das Licht. Darüber hinaus kann die Fotoduodiode mit beliebiger Polarität, ja selbst mit Wechselstrom betrieben werden. Die typische Empfindlichkeit liegt bei ca. 0,5 μA/lm.

Ein Fotoempfänger mit besonders hoher Verstärkung ist der DARLINGTON-*Fototransistor*. Bei ihm steuert der eigentliche Fototransistor einen weiteren Transistor (Bild 6.1.16c)), so daß nochmals eine Stromverstärkung B (Bild 6.1.16d)) hinzukommt, die allerdings meist vom Fotostrom selbst etwas abhängig ist. Deshalb ist der DARLINGTON-Fototransistor weniger linear. Es werden dafür aber Empfindlichkeiten um 300 μA/lm erreicht. Auf das allgemeine Prinzip der DARLINGTON-Transistoren wird im Abschnitt 7.3.2.1. eingegangen.

Schließlich sei noch auf den *Fotothyristor* verwiesen, der erst im Rahmen von Abschnitt 6.4.3. verständlich wird.

6.1.3. Leucht- und Laserdioden

Entsprechend dem Atom- und Bändermodell (3.1.1.) können Elektronen verschiedene Energieniveaus besitzen. Durch unterschiedliche Anregungsmechanismen, z.B. Paarbildung (Generation) über Wärme- oder Lichteinflüsse, kann ein *Elektron auf ein*

höheres Niveau angehoben werden. Dabei gilt

$$h\nu \geqq E_2 - E_1 , \tag{25}$$

wenn ν die Frequenz des Lichtes und E_2, E_1 die Energieniveaus sind. Umgekehrt kann auch ein Elektron vom Niveau E_2 auf das Niveau E_1 herabfallen (wenn dort ein Platz frei ist). Dann wird ein Photon der Frequenz

$$\nu = (E_2 - E_1)/h \tag{26}$$

abgestrahlt. Dieser Effekt heißt *Lichtemission* oder auch *Lumineszenz*. Insbesondere in den Fällen, wo der Lumineszenz eine Anregung durch Licht vorausging, werden *zwei Fälle* danach unterschieden, wie lange das Elektron im höher angeregten Zustand verweilt:

- Fluoreszenz: Verweilzeit $< 10^{-6}$ s
- Phosphoreszenz: Verweilzeit $> 10^{-6}$ s .

Beide Effekte werden noch genauer im Abschnitt 6.2., Vakuumelektronik, bei den Leuchtstoffen zu behandeln sein.

Sonnenlicht und die Strahlung von Körpern besteht, wie schon im Abschnitt 6.1. behandelt wurde, aus einer Vielzahl von Frequenzen. Dem gegenüber steht der *monochromatische Strahler*, der nur eine Frequenz abstrahlt. Im Farbdreieck (Bild 6.1.4) ist er auf die Randkurve der Spektralfarben festgelegt. Bei monochromatischem Licht ist der *Kohärenzgrad* eine weitere wichtige Größe. Infolge der Photonen wird im dualistischen Bild monochromatisches Licht als *Wellenpaket* ausgestrahlt. Das Licht besteht also aus einer Vielzahl *monochromatischer Wellenzüge*, die einen *Anfang* und ein *Ende* besitzen. Zwischen diesen Wellenzügen braucht zunächst keine Phasenbeziehung zu bestehen. Dann ist das monochromatische Licht inkohärent. Sind aber alle *Wellenpakete* zusätzlich starr miteinander in der *Phase verkoppelt*, so ist das Licht *kohärent*. Interferenz tritt ein, wenn das Licht, zerlegt über verschieden lange Wege, aufeinander einwirkt und sich z.T. aufhebt.

Nur die monochromatische, kohärente Lichtquelle entspricht also einer Wechselspannung, die dem technischen Sinuswechselstrom ähnelt. Bei inkohärentem Licht liegt dagegen nur ein Schmalbandrauschen vor. Die Güte des Schmalbandstrahlers wird in der Optik durch die *Kohärenzlänge* gemessen. Das ist jener Umweg für aufgeteiltes Licht einer Quelle, bei dem es noch gerade *interferenzfähig* ist.

6.1.3.1. Grundlagen der Lumineszenzdioden

Für die Lumineszenzdioden existieren auch die Namen LED (light emitting diode), Leuchtdiode, Injektionsdiode, Signallampe. Bei ihnen wird der *pn*-Übergang in Durchlaßrichtung betrieben, dadurch entsteht eine erhöhte Elektronendichte im Leitungsband und eine erhöhte Löcherdichte im Valenzband. Dieser Vorgang heißt auch *Injektion* oder elektrische Generation von Ladungsträgern. Das so erzwungene *thermodynamische Ungleichgewicht* versucht sich auszugleichen. Hierzu stehen verschiedene Mechanismen zur Verfügung, die grob in zwei Gruppen eingeteilt werden können:

- *strahlender Übergang*: Hierbei rekombinieren ein Loch und ein Elektron und geben die dabei frei werdende Energie als Photon ab.
- *nichtstrahlender Übergang*: Hierbei erfolgt die Rekombination von Loch und Elektron so, daß keine Lichtabstrahlung auftritt. Dies kann z.B. durch stufenweisen Abbau an Störstellen oder AUGER-*Effekte* erfolgen. Letztlich übernimmt der Halbleiter dann die Energie als Wärme.

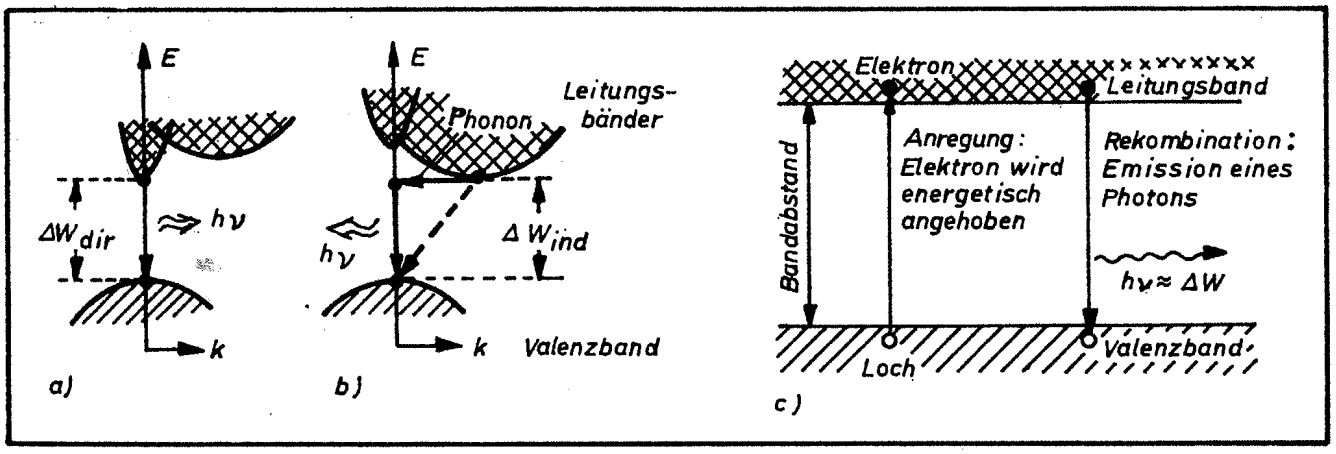


Bild 6.1.17. Zur Strahlung der Lumineszenzdiode

a) Halbleiter mit direkter und b) indirekter Bandstruktur

c) Anregung erfolgt durch Strom in Durchlaßrichtung. Dabei wird ein Elektron ins Leitungsband gehoben und zurück bleibt ein Loch im Valenzband. Strahlung erfolgt, indem ein Elektron aus dem Leitungsband mit einem Loch im Valenzband rekombiniert (direkter Übergang)

Die Lumineszenzdiode muß also so gestaltet werden, daß der *strahlende Übergang* überwiegt. Von großem Einfluß ist dabei, ob das Halbleitermaterial *direkte oder indirekte Übergänge* besitzt. In Anlehnung an Bild 3.1.2 gilt hierfür Bild 6.1.17. Beim direkten Übergang existieren die geringsten Bandabstände im Kristallgitter genau gegenüber, beim indirekten Übergang sind sie um eine bestimmte Wellenzahl k seitlich versetzt. Dadurch muß zusätzlich eine entsprechende *Gitterschwingung* (Phonon) bei der Rekombination mit wirksam werden. Deshalb ist bei indirekten Halbleitern der strahlende Übergang rund vier Zehnerpotenzen unwahrscheinlicher. Dieser Wert läßt sich mit *isoelektronischen Störstellen* z. T. verbessern. Das sind eingelagerte Atome, welche die gleiche Valenz haben und daher nicht dotierend wirken. Sie besitzen aber abweichende Atomradien und können so zeitweilig die beiden Rekombinationspartner binden, d. h. die Gitterschwingung aufbringen. Bei GaAs bzw. GaP kommen u. a. zur Anwendung Zn, O und N.

In der Regel erfolgt die Rekombination von Elektronen an der unteren Kante des Leitungsbandes und die von Löchern an der Oberkante des Valenzbandes. Deshalb gilt wieder (15) zwischen dem Bandabstand und der Wellenlänge des emittierten Photons (Bild 6.1.17c)). Die *Bandbreite des emittierten Spektrums* ist dadurch recht schmal. Sie liegt in der Größenordnung von kT, also um 10 nm. Ihre Farbwerte liegen damit direkt auf der Randkurve der Spektralfarben in Bild 6.1.4.

Um *sichtbares Licht* zu erzeugen, sind somit die relativ großen *Bandabstände* von 1,6 (rot) bis 3 eV (blau) notwendig. Dies und die Forderung nach möglichst direkten Bandübergängen hat zur Folge, daß die *Optoelektronik eine Vielzahl spezieller Halbleiter* fordert. Einen Überblick gibt Bild 6.1.18. Weitere Detaildaten enthält [W 12]. Im optischen Bereich herrschen deutlich die III-V-Verbindungen GaAs, GaP und daraus abgeleitete Mischkristalle vor. Generell sind Mischkristalle aus III-V-Verbindungen für Lumineszenzdioden aller Wellenlängen vorteilhaft. Dabei können — im Sinne der isoelektronischen Atome, wie Bild 6.1.18b) und c) zeigt — in einem gewissen Mischungsbereich direkte Übergänge genutzt werden.

Für die *Lichtabstrahlung* ist der *Brechungsindex* n des Halbleitermaterials bedeutsam. Wenn Licht flacher als unter einem *Grenzwinkel*

$$\varphi_g = \arcsin(1/n) \tag{27}$$

von innen auf die Materialkante gegenüber Luft auftrifft, erfolgt Totalreflexion. Für GaAs beträgt dieser Winkel nur 16,2° und für GaP 17,7° (n: 3,6 bzw. 3,3). Es kann

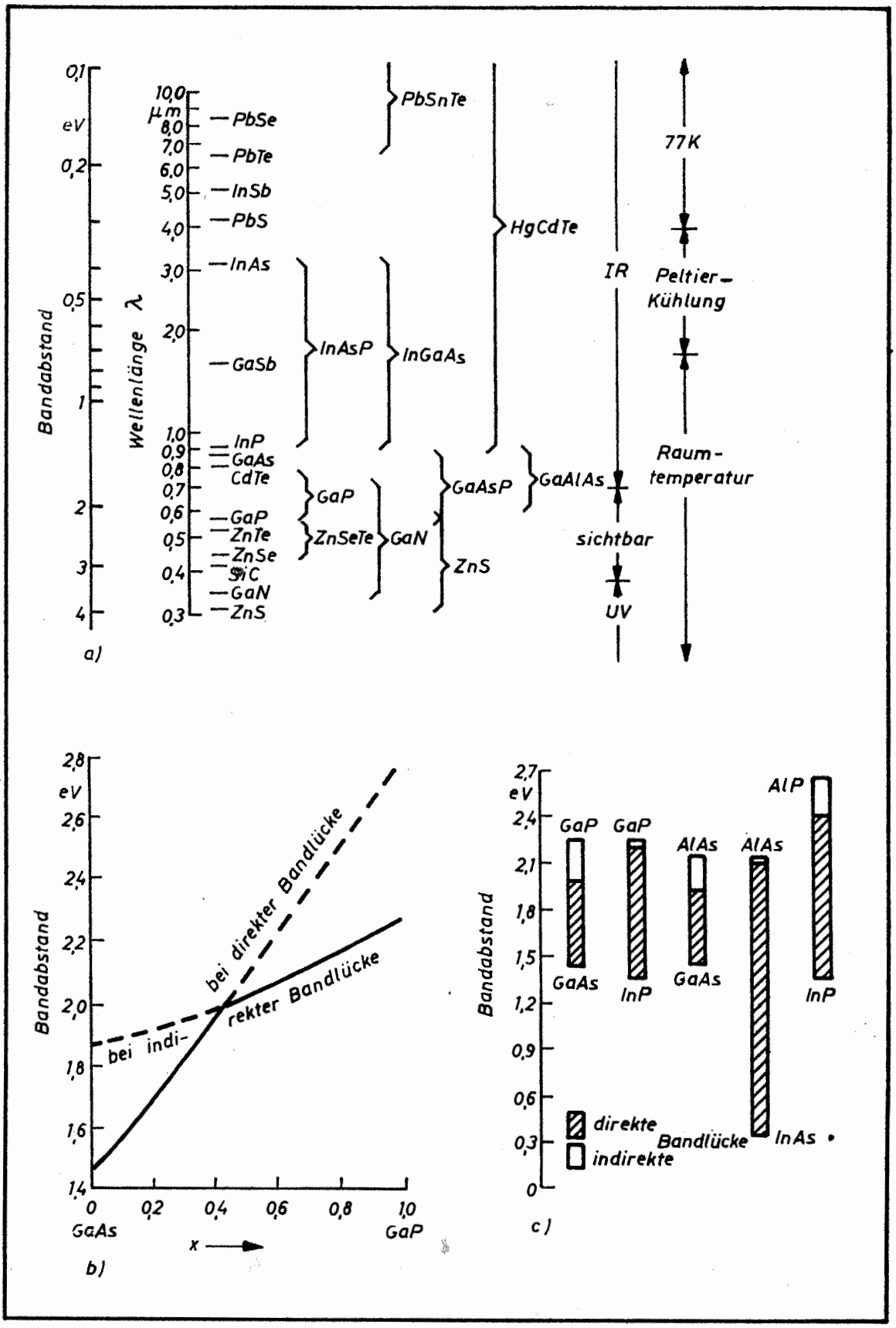


Bild 6.1.18. Substanzen für Lumineszenzdioden

a) Erzeugbare Wellenlängen und Bandabstand verschiedener nutzbarer Halbleitermaterialien

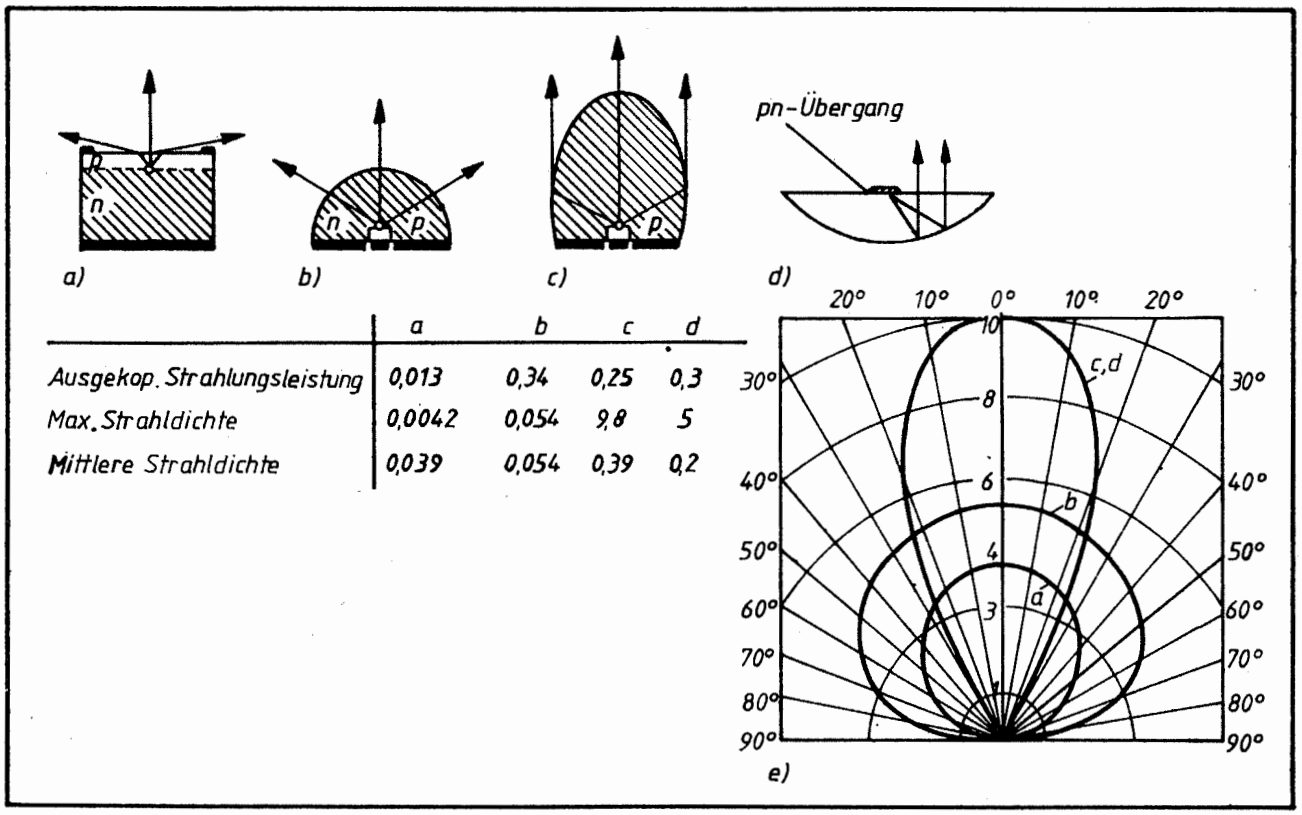
b) Abhängigkeit der Bandlücke bezüglich des direkten und indirekten Überganges bei Mischkristallen aus $GaAs_{1-x}P_x$

c) Direkter und indirekter Bereich bei den wichtigsten III-V-Mischkristallen

also nur jenes Licht den Halbleiter verlassen, das steiler ankommt. Aber selbst für diesen Anteil erfolgt noch *teilweise Reflexion*. Für senkrechte Strahlung beträgt der *Transmissionskoeffizient* nur

$$T_0 = \frac{4n}{(n+1)^2}, \tag{28}$$

also bei GaP 0,715. Da der Transmissionskoeffizient für φ_g verschwindet, ist der gemittelte Wert noch kleiner, bei GaP z.B. 0,7. Insgesamt kann so bei ebener Oberfläche von der im Halbleiter entstehenden Strahlung bei GaP nur etwa 1,7% den Kristall auf direktem Wege verlassen.



	a	b	c	d
Ausgekop. Strahlungsleistung	0,013	0,34	0,25	0,3
Max. Strahldichte	0,0042	0,054	9,8	5
Mittlere Strahldichte	0,039	0,054	0,39	0,2

Bild 6.1.19. Möglichkeiten zur Vermeidung von Reflexionsverlusten bei Lumineszenzdioden

a) Ebener Flächenstrahler mit Begrenzung der Auskopplung durch Totalreflexion
b) Halbkugelstrahler
c) Ellipsoidsegmentstrahler
d) Strahler mit parabolischer Grundfläche
e) Typische Richtdiagramme der Typen a) bis d)

Deshalb wurden *Oberflächenstrukturen* entwickelt, die effektiver sind. Sie zeigt mit einigen Daten Bild 6.1.19. Dabei wird auch zugleich das *Richtdiagramm* der Dioden verändert. Weiter kann der Transmissionskoeffizient durch eine angepaßte $\lambda/4$-*Schicht* gemäß (13a) erheblich verbessert werden. Für die senkrechte Strahlung wird er dann praktisch Eins. Über alle Winkel gemittelt, steigt die *Strahlungsausbeute* um 30 bis 45%.

Im Inneren des Halbleitermaterials wird weiter *Strahlung absorbiert*. Deshalb möchte der strahlende *pn*-Übergang möglichst dicht an der Oberfläche liegen. Andererseits kann er auch nicht zu dicht an die Oberfläche verlegt werden, weil dann an den *Oberflächentermen* zu viele nichtstrahlende Übergänge wirksam werden.

6.1.3.2. Aufbau und Daten von Lumineszenzdioden

Lumineszenzdioden werden heute im sichtbaren Bereich vor allem als Signalanzeigen und Segment- bzw. Punktanzeigen gemäß Bild 6.1.7 verwendet. Als Anzeigen gibt es kleine, die über Kunststofflinsen vergrößert werden, z. B. bei Taschenrechnern, und große, die bis zu Symbolhöhen von etwa 20 mm reichen. Dabei sind die Farben *rot — gelb — grün* mit Übergängen realisierbar. *Blaue Dioden* werden z. Z. noch nicht großtechnisch beherrscht. Sie benötigen eine Wellenlänge von $\lambda < 480$ nm bzw. einen Bandabstand $\Delta W > 2{,}6$ eV. Hierfür kämen nach Bild 6.1.18 u. a. GaN, ZnS, $ZnS_{1-y}Se_y$ und SiC in Betracht. Für GaN lassen sich keine *pn*-Übergänge erzeugen. Für Substanzen auf der Basis ZnS besteht Unklarheit über die Dotierung und günstige Herstellung des Substrates. Am weitesten gediehen sind die Untersuchungen bei SiC. Hier existiert jedoch eine große Anzahl von Modifikationen, die bei der Hochtemperaturtechnologie um 2000 °C noch nicht definiert beherrscht werden.

Von großer praktischer Bedeutung sind auch *Infrarotdioden*, vor allem auf der Basis von GaAs. Sie werden u. a. für Lichtschranken, Fernsteuerung und Optokoppler (6.1.4.) eingesetzt.

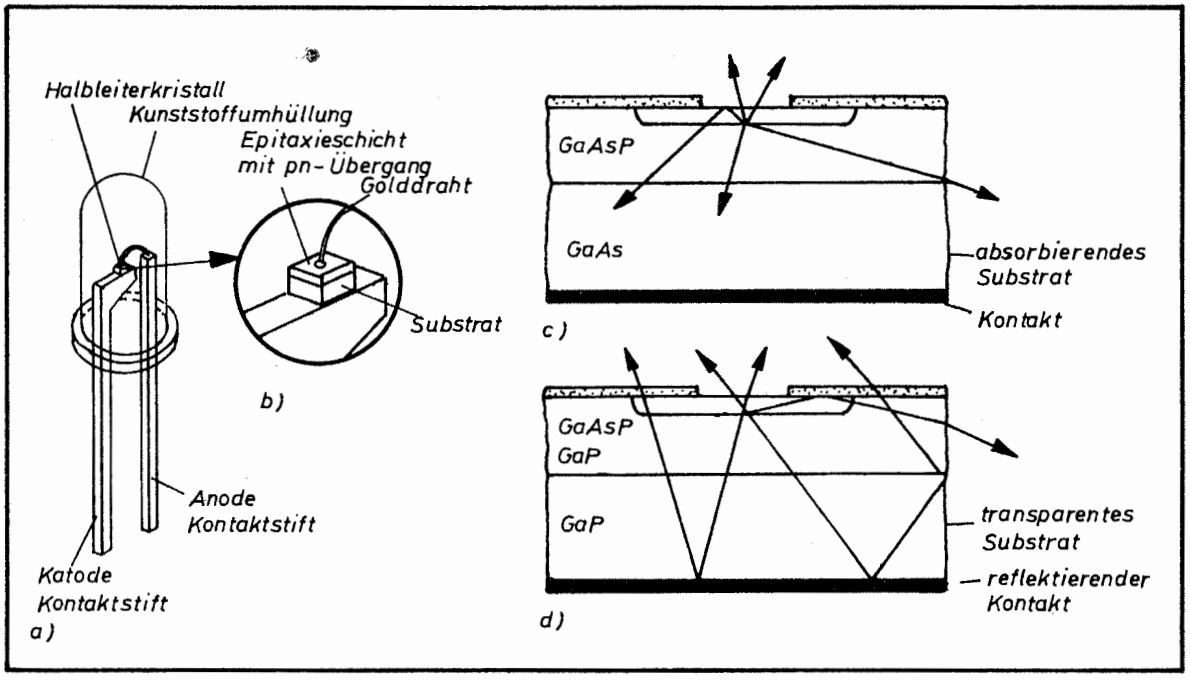


Bild 6.1.20. Zum Aufbau von Signalanzeigen

a) Grundprinzip

b) Befestigung des Halbleiterkristalls auf dem Kontaktstift

c) Aufbau und Lichtwege in lichtabsorbierendem GaAs- und d) lichtdurchlässigem GaP-Substrat

Den typischen Aufbau einer Signalanzeige zeigt Bild 6.1.20a) und b). Wenn man *n*-GaP als Substrat verwendet, kann der Vorteil genutzt werden, daß es für die emittierte Strahlung durchlässig ist. Mit einer Reflexschicht an der Unterseite (Bild 6.1.20d)) kann die nach unten gehende Strahlung zusätzlich genutzt werden. Es wird dann von der *TSN-Technologie* (transparent substrate nitrogen) gesprochen. Diese Dioden sind daher 5- bis 10mal heller als die üblichen mit dem absorbierenden GaAs-Substrat. Leider ist die Technologie des GaP-Substrates teurer.

Für die *Segmentanzeigen* sind die monolithischen (Bild 6.1.21a)) und die diskreten zu unterscheiden. Die monolithischen lassen sich nur bis zu gewissen Größen realisieren. Bei ihnen sind z. T. Sondermaßnahmen notwendig, damit Licht nicht auf Umwegen in den anderen Segmenten sichtbar wird. Bei den diskreten war es früher üblich,

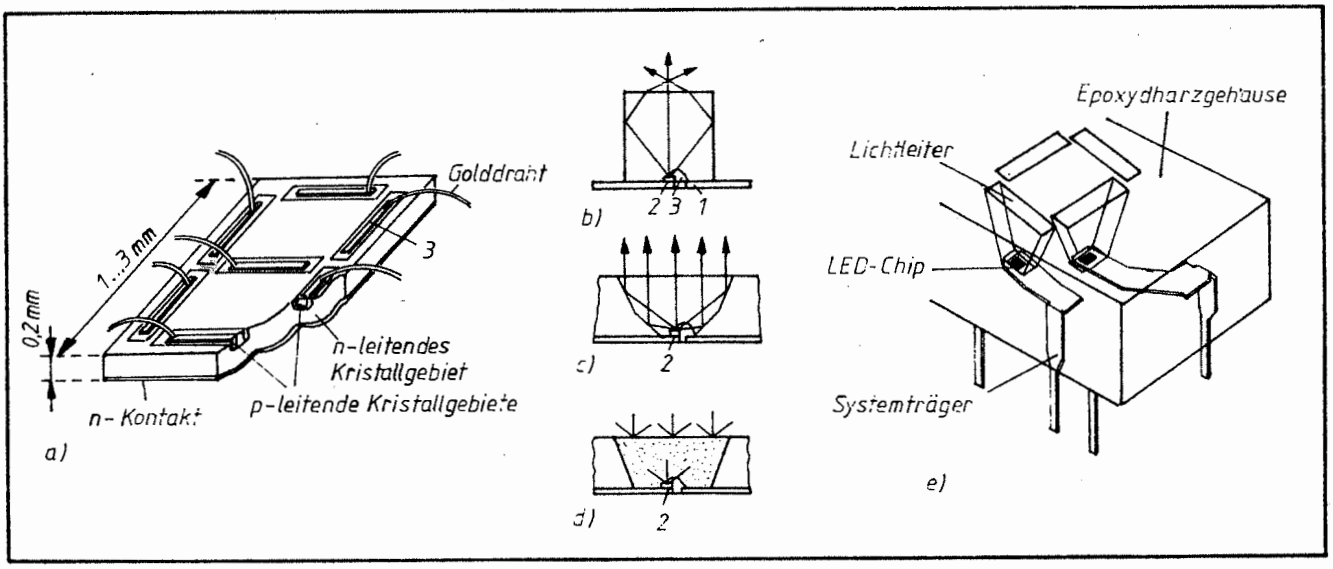


Bild 6.1.21. Aufbau von LED-Anzeigen

a) Monolithisch
b) bis e) Diskret
b) Lichtleiterprinzip
c) Mit verspiegelten Seitenflächen
d) Mittels diffuser Reflexion
e) Ausschnitt aus einem 7-Segment-Element mit Lichtleitern

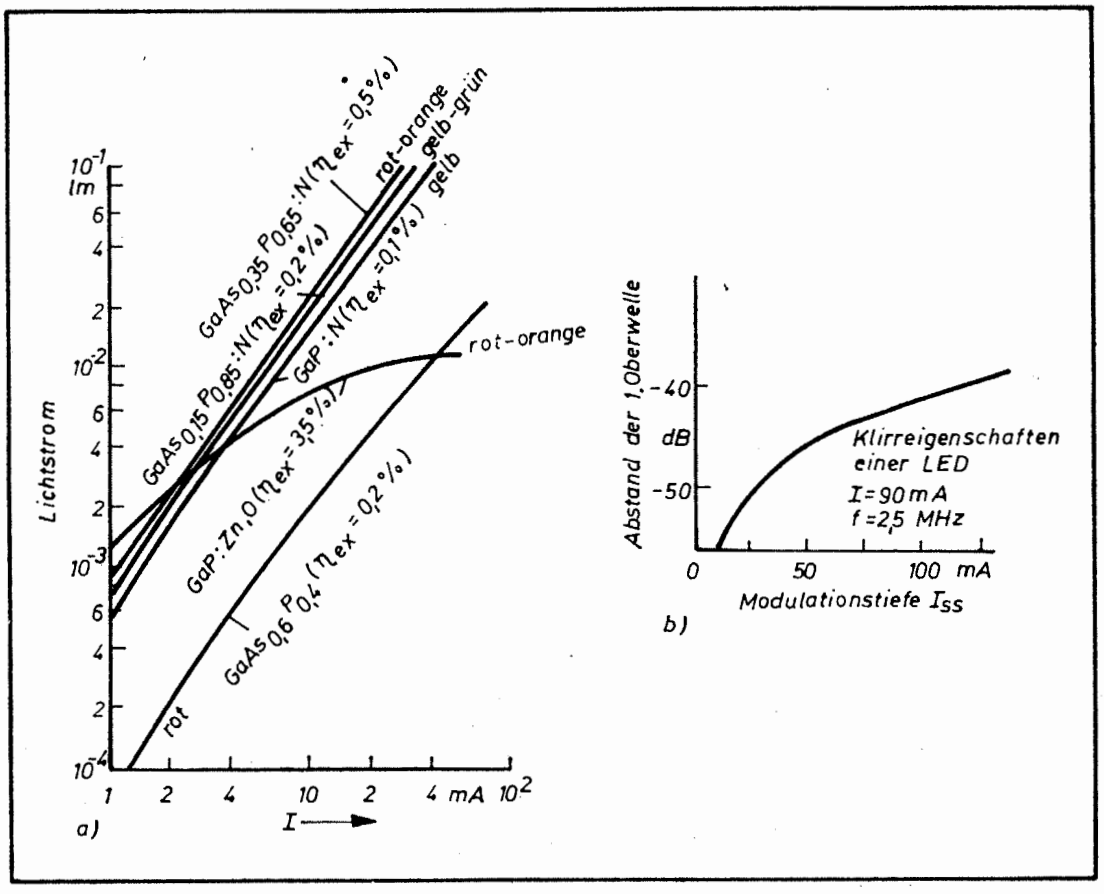


Bild 6.1.22. Kennlinien von Lumineszenzdioden

a) Typische Daten unterschiedlicher kommerzieller Typen
b) Nichtlinearität der Lichtmodulation in Abhängigkeit von der Modulationstiefe für eine typische Infrarotdiode

mehrere einzelne Leuchtpunkte zu einer Linie zu kombinieren. Dieses Verfahren ist durch die *Lichtschacht-* bzw. *Lichtleitertechnologie* gemäß Bild 6.1.21b) bis e) abgelöst. In der Regel wird an der Oberfläche der Strukturen noch eine Folie angebracht, welche möglichst nur das emittierte Licht durchläßt. Dadurch wird der Kontrast bei Tageslicht erheblich erhöht, so daß selbst bei hellem Sonnenlicht (10000 lx) noch gut sichtbare Anzeigen möglich sind.

Die technischen Daten von Lumineszenzdioden zeigen Bild 6.1.22 und Tab. 6.1.7.

Tabelle 6.1.7. Typische Daten von Lumineszenzdioden

Material	Substrat	Farbe	Äußerer Wirkungsgrad in %		Lichtausbeute in lm/W		Lichtstärke in mcd bei 10 mA
			Mittelwert	Spitze	Mittelwert	Spitze	
GaAs	Si	IR	15	35	—	—	—
$GaAs_{0,6}P_{0,4}$	GaAs	rot	0,2	0,5 bei 1 A/cm²	0,15	0,4	0,4 ... 1
GaP : ZnO	GaP	rotorange	4	15 bei 1 A/cm²	0,6	3	ca. 10
$GaAs_{0,35}P_{0,65}$:N	GaP	rotorange	0,4	0,6 bei 500 A/cm²	0,8	1,2	2 ... 4
$GaAs_{0,15}P_{0,85}$:N	GaP	gelb	0,2	0,3 bei 500 A/sm²	0,9	1,4	1 ... 3
GaP : N	GaP	gelbgrün	0,1	0,7 bei 100 A/cm²	0,6	4,5	0,5 ... 3

6.1.3.3. Laserprinzip

Laser gibt es heute in vielen *Varianten*:

- *Gaslaser* in Gasentladungen, z. B. CO_2, He-Ne, N;
- *Festkörperlaser* in Kristallen, z. B. Al_2O_3 mit Cr (Rubin), also Farbzentren;
- *Farbstofflaser* mittels spezieller Farbzentren;
- *chemische Laser* unter Ausnutzung chemischer Prozesse;
- *Halbleiterlaser.*

Hier soll nur der Halbleiterlaser genauer betrachtet werden. Bezüglich der anderen siehe z. B. [R 12, Bd. 5].

Die *Anwendung* des Lasers hatte vor allem wegen *drei Eigenschaften* große Bedeutung erlangt:

- monochromatische, kohärente Lichtstrahlung (auch UV und IR)
- große Strahlungsintensität
- extrem kurze Lichtimpulse.

Schließlich gibt es auch *durchstimmbare Laser.* Zur Lasertechnik gehören heute weiter Lichtmodulatoren, Lichtablenksysteme. Anwendungsbeispiele sind u. a. Materialbearbeitung, Analysenmethoden, Holographie und die optische Nachrichtentechnik. Hier interessiert der Laser wiederum nur als spezielle Lichtquelle.

Das *Grundprinzip des Lasers* wird am besten mit Bild 6.1.23a) verständlich. Mittels einer Pumpe als Energiequelle wird dem Laserwerkstoff Energie zugeführt, die er speichert. Durch eine gezielte oder zufällige Lichtquelle wird infolge des optischen Resonators und auftretender Wechselwirkungen die zuvor gespeicherte Energie als monochromatische (Resonator) und kohärente (Wechselwirkung) elektromagnetische Welle abgestrahlt. Die Wechselwirkung geht auf die bereits 1917 von Einstein formulierte *induzierte bzw. stimulierte Strahlungsemission* zurück. Sie wurde erstmals 1928 von Ladenberg und Kopfermann nachgewiesen. Diese induzierte Emission ist in Bild 6.1.23b) der Absorption und spontanen Emission von Licht gegenüber-

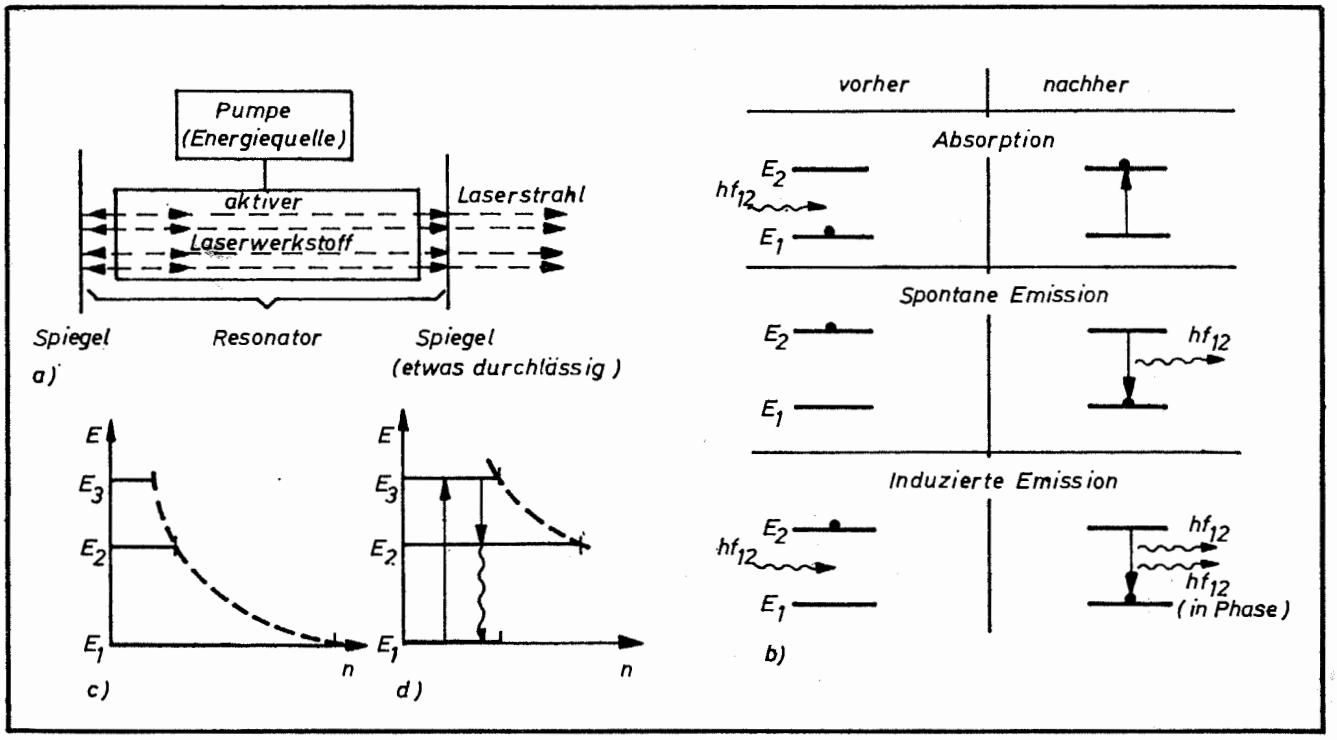


Bild 6.1.23. Grundlagen des Laserprinzips

a) Grundbausteine und deren Zusammenwirken beim Laserprinzip
b) Schematische Darstellung der drei Vorgänge Absorption, spontane und induzierte Emission
c) Thermische Belegung bei einem Dreiniveausystem
d) Das gleiche System im gepumpten Zustand (Inversion)

gestellt. Die Absorption kann für den Lasermechanismus als *Pumpen* betrachtet werden. Es erfolgt dann optisches Pumpen. Bei den Lumineszenzdioden erfolgt dieses Pumpen durch elektrische Generation von Ladungsträgern, indem ein *pn*-Übergang in Durchlaßrichtung betrieben wird. Die spontane Emission ist der Ausgleich des thermischen Ungleichgewichts durch Strahlung, wie sie z. B. bei Lumineszenzdioden unmittelbar erfolgt. Wird dieser Ausgleich durch spezielle Mechanismen verhindert, so kann die Emission durch ein Lichtquant eingeleitet werden. Dieses Lichtquant wird auf diese Weise (in anderer Interpretation) verstärkt. So ist der Name Laser gemäß *"light amplification by stimulated emission of radiation"* verständlich.

Die *Verhinderung der spontanen Emission* setzt zumindest ein energetisches *Drei-Niveau-Schema* gemäß Bild 6.1.23c) voraus. Hier sind entsprechend der gestrichelten Linie (vgl. Bild 3.1.6) die drei Terme gemäß der vorhandenen Temperatur belegt. Mittels Pumpen wird der Zustand E_3 auf Kosten von E_1 belegt. Weiter ist der Übergang $E_3 \rightarrow E_2$ wahrscheinlicher als der Übergang $E_3 \rightarrow E_1$. So ist es möglich, eine Inversion zu erreichen. Dies bedeutet, daß E_2 höher als E_1 belegt ist und nur zwischen E_2 und E_3 wieder die thermisch zu erwartende Verteilung vorliegt. Mittels einer Einstrahlung eines Quantes der Energie $E_2 - E_1$ kann dann die stimulierte Emission ausgelöst werden.

6.1.3.4. Halbleiterlaser

Der *Halbleiterlaser unterscheidet* sich *von der Lumineszenzdiode* vor allem durch zwei Fakten:

- In der aktiven Schicht muß eine so starke Besetzungsinversion erreicht werden, daß ein entstehendes Photon nicht mehr absorbiert wird, sondern ein weiteres induziert, stimuliert.

- Es muß ein optischer Resonator existieren, der die stimulierte Emission verkoppelt.

Der zweite Punkt hat unmittelbare Konsequenzen für den Aufbau. Der Halbleiterkristall wird so gestaltet, daß er ein FABRY-PEROT-*Interferometer* bildet. Dies ist in Bild 6.1.24a) angedeutet. Die Kanten werden entsprechend den Achsen des Kristalls bearbeitet und bilden die verspiegelten Wände im Abstand von l_1 bzw. l_2. Infolge des hohen Brechungsindex kann die zusätzliche Verspiegelung meist entfallen. Da die Längen des Kristalls im Bereich einiger hundert µm liegen, befinden sich im Resonator einige hundert bis tausend Wellenlängen. Dadurch kann es zur Ausbildung mehrerer *stehender Wellenlängen* kommen.

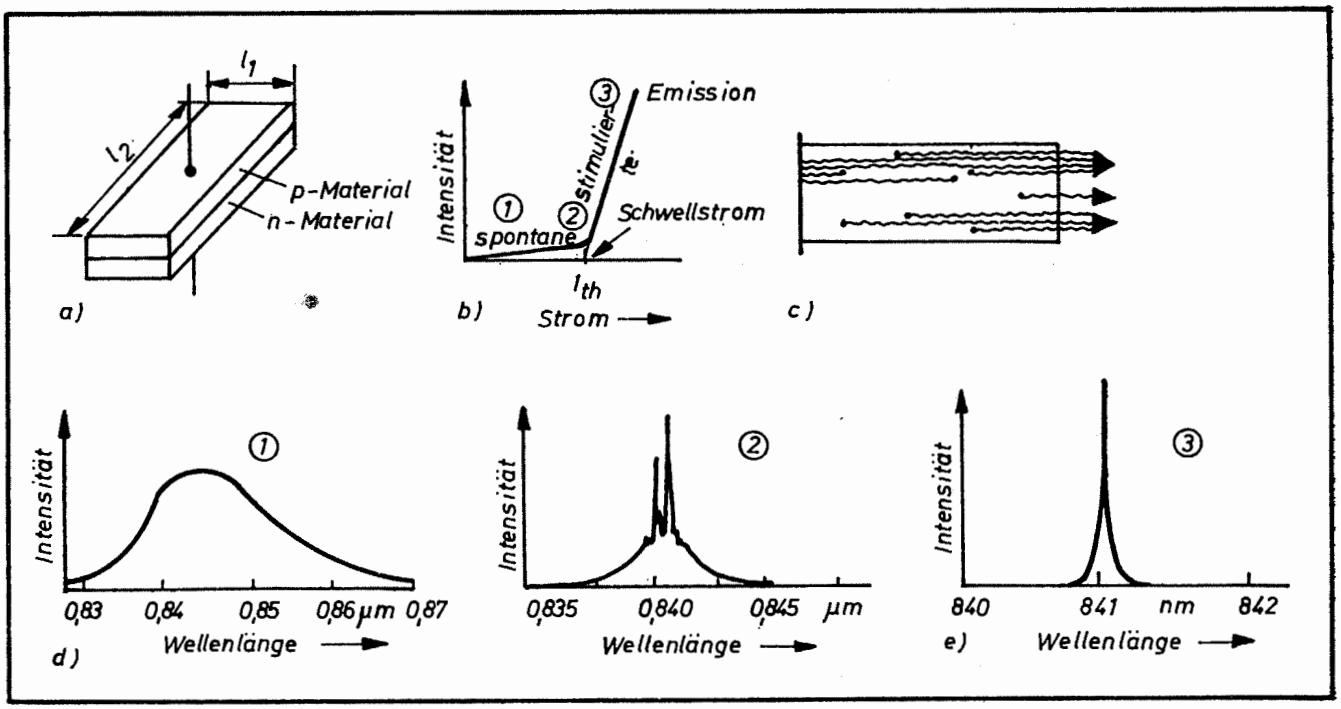


Bild 6.1.24. Zum Laserprinzip des Halbleiters

a) Resonater eines Halbleiterkristalls
b) Kennlinie: Bereich ① wie bei Lumineszenzdiode, ② Knick mit Anfang der stimulierten Emission, ③ Laserbetrieb
c) Prinzip der Superstrahlung im Halbleiterkristall
d) Wellenlängenverteilung im Lumineszenzbetrieb, dann im Übergangsgebiet, also etwa Superstrahlung
e) Laserstrahlung

Wird eine normale Lumineszenzdiode immer stärker angeregt, so zeigt die Emissionsintensität plötzlich einen steilen Anstieg (Bild 6.2.24b)). Zu diesem Zeitpunkt überwiegt die Generation die Absorption, und es kommt zur Ausbildung von *Superstrahlung* (Bild 6.1.24c)). Hierbei bewirken entstehende spontane Emissionen zusätzliche Strahlung. Es existieren aber noch mehrere ungekoppelte Gebiete im aktiven Gebiet der Lumineszenzdiode. In diesem Zustand (eventuell unter zusätzlicher Erhöhung des Stromes durch den *pn*-Übergang) kann durch die Wirkung eines *Resonators* eine Kopplung zwischen den Wellen, einschließlich zusätzlicher *Rückkopplungsmechanismen*, entstehen. Reduziert sich die Bandbreite der Strahlung beim Übergang zur Superstrahlung schon erheblich, so engt sie sich dann (beim Laserbetrieb) auf praktisch eine Wellenlänge ein (Bild 6.1.24d) bis f)). Mit der Einengung der Bandbreite erfolgt auch eine Verkleinerung des Abstrahlwinkels.

Das Hauptproblem bei der Entwicklung der Laser war die Reduzierung des *Schwellstromes* von ursprünglich 10^5 A/cm² auf heute fast 500 A/cm². (Bei der Lumineszenzdiode treten etwa 10 A/cm² auf.) Diese Entwicklung zeigt Bild 6.1.25a). Damit

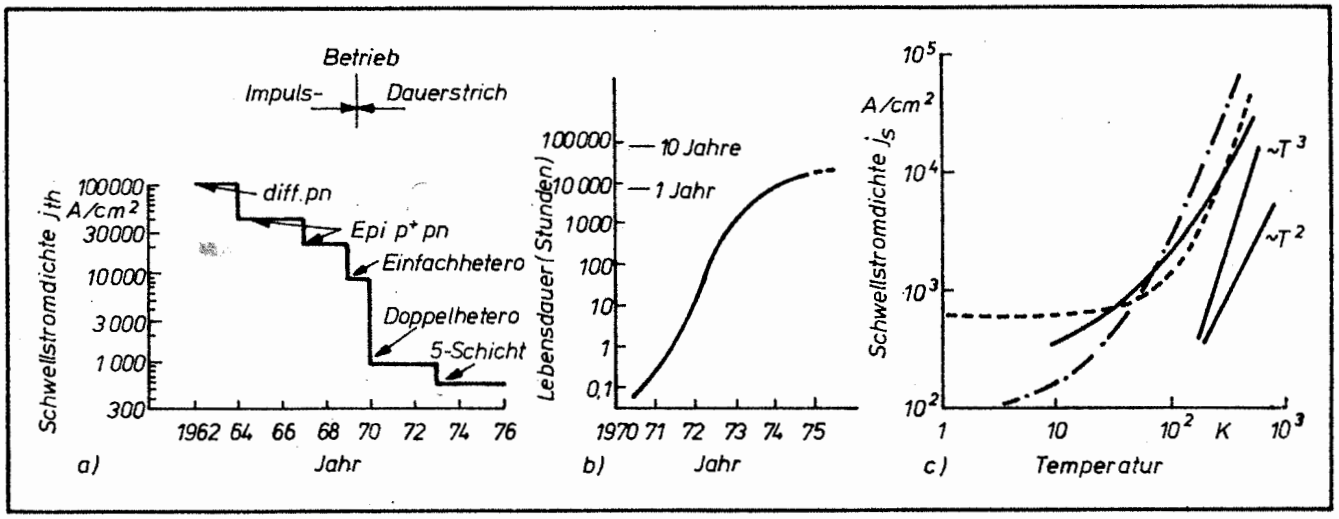


Bild 6.1.25. Zur Entwicklung des Lasers

a) Abnahme der Schwellstromdichte bei Zimmertemperatur
b) Zunahme der Lebensdauer von Lasern im Dauerstrichbetrieb
c) Abhängigkeit der Schwellstromdichte von der Temperatur unter verschiedenen Randbedingungen

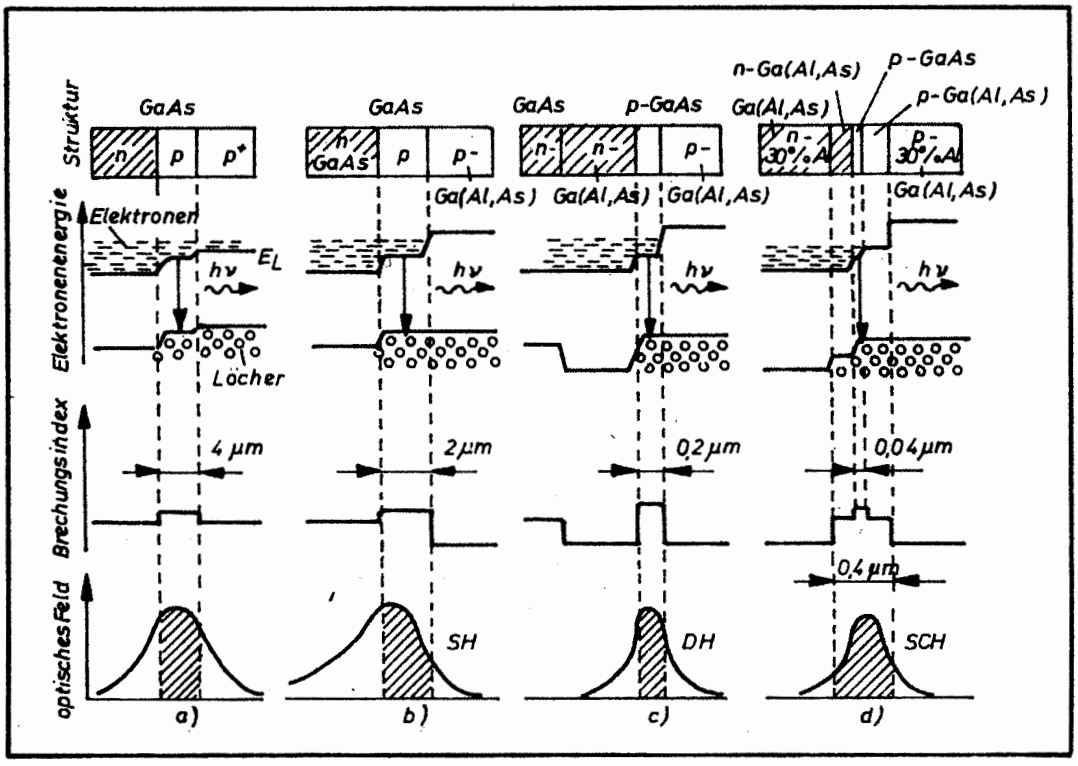


Bild 6.1.26. Aufbau, Bänderschema, Verlauf des Brechungsindex und einfachstes optisches Strahlungsfeld für die wichtigsten Laserdiodentypen mit FABRY-PEROT-Struktur

a) Homodiode, p^+pn
b) Diode mit einem Heteroübergang (SH)
c) Diode mit zwei Heteroübergängen und damit starker Lokalisierung des optischen Feldes (DH)
d) Fünfschichtdiode mit vier Heteroübergängen (SCH), die das optische Feld auf einen relativ breiten Streifen lokalisiert und die injizierten Träger auf eine darin enthaltene, sehr schmale Zone.

Die kleinen Brechungsindexsprünge an den Übergängen betragen 0,1% bis 1,0%, die großen liegen bei etwa 5%; dementsprechend ist die Lichteindringtiefe in die nicht angeregten Bereiche etwa 1 µm bzw. 0,1 µm. Der Bereich der angeregten Zone ist schraffiert dargestellt

einher geht die Entwicklung der Lebensdauer (Teilbild b)) sowie der Übergang vom Impulslaser bei Kühlung über den ohne Kühlung bis zum Dauerstrichlaser bei Zimmertemperatur.

In Bild 6.1.25a) sind verschiedene *Strukturen* genannt. Ihren Aufbau mit den entsprechenden Eigenschaften zeigt Bild 6.1.26. Hieraus wird deutlich, wie *im Laufe*

der Zeit ständig die Strukturen *komplizierter* wurden. Dabei sind *zwei Fakten* von Bedeutung:

- GaAs und AlAs haben praktisch die gleiche Gitterkonstante. Mit beiden Verbindungen ist daher ein fast idealer epitaktischer Übergang, auch bei Mischung, möglich (Bild 6.1.27b)).
- Im Mischkristall $Al_xGa_{1-x}As$ ändert sich der Brechungsindex erheblich (Bild 6.1.27a)).

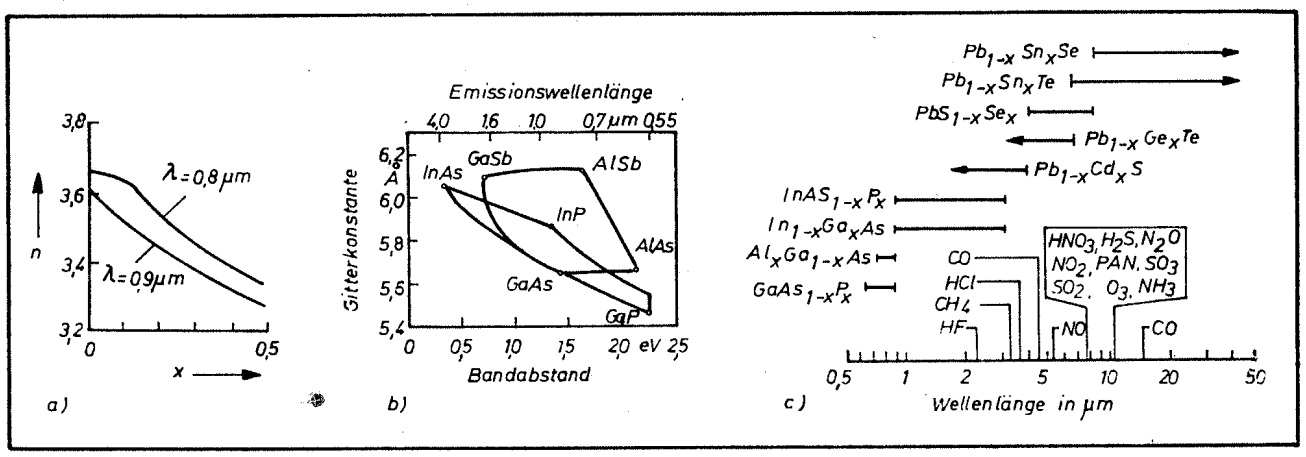


Bild 6.1.27. Daten zu Laserdioden

a) Brechungsindex $n = \sqrt{\varepsilon_r}$ der Verbindungen $Al_xGa_{1-x}As$ bei Zimmertemperatur
b) Bandabstand und Gitterkonstanten verschiedener Mischkristalle
c) Wellenbereich für Laser und absorbierende Substanzen

Auf dieser Basis ist es möglich, den *aktiven Bereich* sowohl bezüglich der Trägererzeugung als auch bezüglich der optischen Begrenzung immer *besser zu gestalten*: Während beim einfachen Heteroübergang (SH = single hetero) noch optischer und *pn*-Bereich zusammenfallen, wird beim doppelten Heteroübergang (DH = double hetero) bereits der optische Bereich zusätzlich erheblich eingeengt. Beim fünffachen Heteroübergang (SCH = separate confinement hetero) wird dagegen der elektronisch wirksame Bereich extrem und der optische Bereich nur weniger eingeengt. Durch den so immer enger werdenden wirksamen *pn*-Übergangsbereich konnte die Schwell-

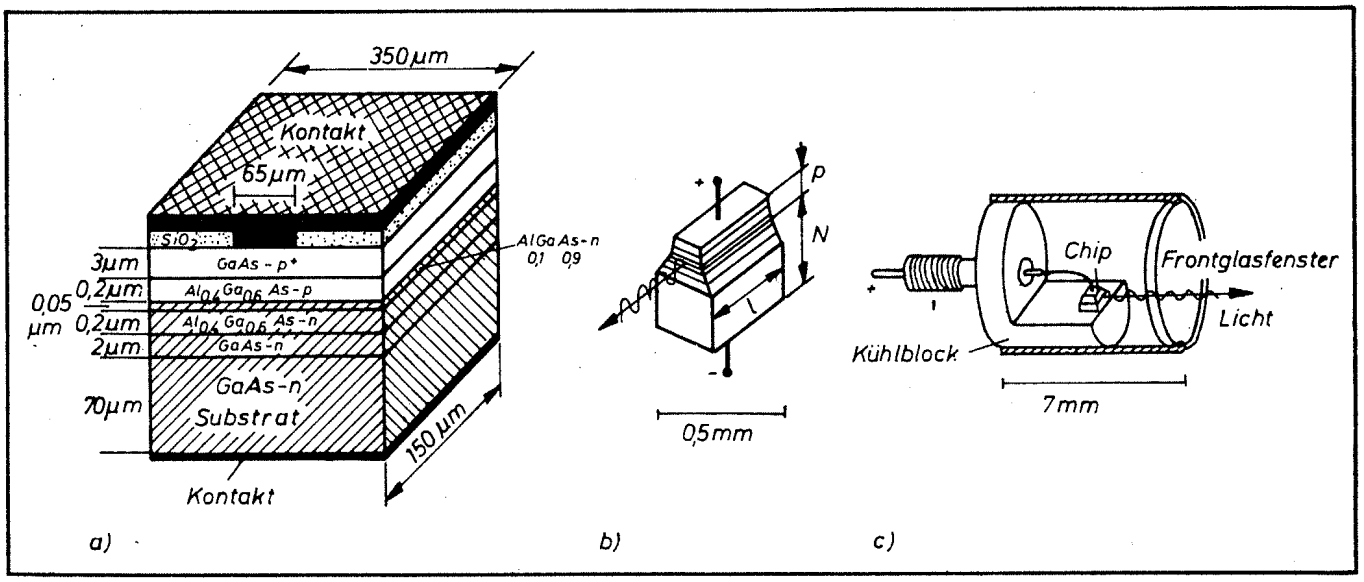


Bild 6.1.28. Aufbau eines Lasers (c), des darin enthaltenen Kristalls mit Mesastruktur (b) und der Struktur selbst (a)

stromstärke für die gleiche Inversion gesenkt werden. Durch optische Begrenzung wurde die Wellenführung verbessert; die Struktur wirkte immer mehr als Wellenleiter.

Den realen Aufbau eines Laserkristalls sowie eines Lasers zeigt Bild 6.1.28. Insbesondere aus Teilbild a) ist bereits ein weiterer Weg der Entwicklung zu erkennen, nämlich die zusätzliche laterale Einengung des aktiven *pn*- und des optischen Bereichs, in diesem Fall durch den verkürzten oberen Kontakt. Hier werden ebenfalls verschiedene Methoden angewendet, z. B. zusätzlicher seitlicher *pn*-Übergang, V-Nut usw. Dadurch wird neben weiter verbesserten Daten erreicht, daß auch eine optimale Anpassung an zylindrische Lichtleitfasern möglich ist.

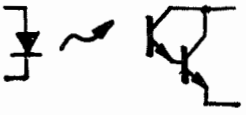
Weitere Details zu Lasern enthält u. a. [W 12].

6.1.4. Optokoppler und Lichtschranken

Das von GaAs-Lumineszenzdioden emittierte Licht hat gemäß Bild 6.1.18 eine Wellenlänge um 0,9 μm, liegt also im Infrarotbereich. Genau hier liegt gemäß Bild 6.1.9 die maximale Empfindlichkeit einer Si-Fotodiode. Deshalb läßt sich durch eine optische Kopplung beider Bauelemente ein spezielles elektronisches Bauelement realisieren, das ohne elektrische Kopplung Gleich- und Wechselströme überträgt. Es heißt Optokoppler oder auch Koppelelement, Optoisolator, optoelektronisches Relais bzw. OCI (= optically coupled insulator).

Statt der Fotodiode wird auch ein Fototransistor bzw. DARLINGTON-Fototransistor eingesetzt. So ergeben sich die Schaltungen und Daten von Tab. 6.1.8. Der entscheidende Vorteil des *Optokopplers* gegenüber *elektromechanischen Relais* ist das Fehlen bewegter Teile und die dadurch bedingte höhere Grenzfrequenz und Zuverlässigkeit sowie ihre volle Kompatibilität zu den elektronischen Bauelementen.

Tabelle 6.1.8. Typische Eigenschaften von Optokopplern
Generell gilt etwa:
Isolationswiderstand Sender : Empfänger $\geqq$ 100 GΩ,
Koppelkapazität ca. 1 pF; Isolationsspannung größer 1000 V,
Produkt Bandbreite · Verstärkung ca. 500 kHz.

Schaltung	Empfänger	Stromübertragungsfaktor	Bandbreite ca.	Rauschen
	Diode	0,1 ... 0,3%	10 MHz	gering
	Transistor	30 ... 100%	300 kHz	mittel
	DARLINGTON-Transistor	100 ... 1000%	30 kHz	erheblich

Das Prinzipschaltbild eines Optokopplers mit Fototransistor zeigt Bild 6.1.29a), eine Ersatzschaltung Teilbild b). Hieraus wird deutlich, daß vom Licht eine Ersatzstromquelle I gesteuert wird, die den Basisstrom des Fototransistors steuert (6.1.2.7.). Insbesondere die hohe Kollektor-Basis-Kapazität und ihre eventuelle Erhöhung mittels des MILLER-Effektes (vgl. 5.5.) reduzieren die Grenzfrequenz gegenüber einer Fotodiode. Aus diesem Grunde ist zuweilen auch die Basis als Anschluß herausgeführt.

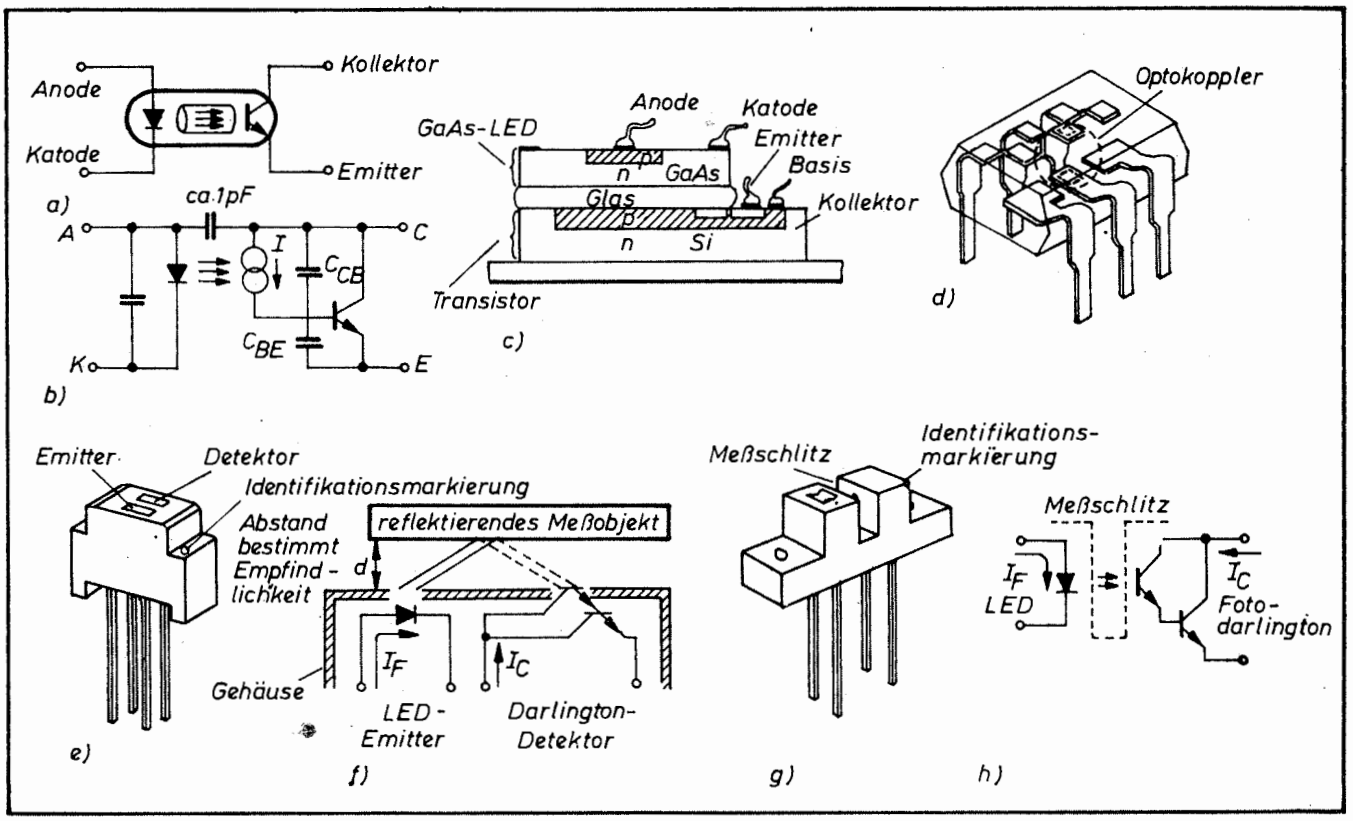


Bild 6.1.29. Bauelemente mit optischer Kopplung

a) Prinzip des Optokopplers mit Fototransistor
b) Ersatzschaltung
c) möglicher struktureller Aufbau
d) Anordnung im DIL-Gehäuse
e) Aussehen eines Reflexkopfes und
f) innerer Aufbau
g) Aussehen einer Gabellichtschranke und
h) innerer Aufbau

Dann sind wieder die Schaltungen gemäß Bild 6.1.15 möglich und so ein Kompromiß zwischen Empfindlichkeit und Bandbreite einstellbar.

Optokoppler sind auch als analoge Bauelemente verwendbar. Die *Linearität* ist vor allem bei Fotodioden mit 1 bis 5% ausreichend. Bei Fotodioden existiert darüber hinaus ein meist sehr kleiner Dunkelstrom von 1 bis 10 nA. Die Linearität kann im Bedarfsfall durch Sonderkoppler, die z.B. eine zweite Fotodiode enthalten, durch Gegenkopplung von dieser Diode erheblich verbessert werden.

Einen häufig verwendeten Aufbau für Optokoppler zeigt Bild 6.1.29c). Sie sind meist im DIL-Gehäuse (Teilbild d)) untergebracht. Sonderformen des Optokopplers sind der *Reflexkopf* und die *Gabellichtschranke* (Bild 6.1.29e) bis h)). Sie haben als Sensoren bezüglich bewegter Objekte Bedeutung.

6.1.5. Bauelemente mit Ladungskopplung

6.1.5.1. Drei- und Zweiphasenbetrieb

Die Grundlage der ladungsgekoppelten Bauelemente (CCD = charge coupled devices bzw. CTD = charge transfer devices) wurde im Abschnitt 3.3.8., insbesondere mit Bild 3.3.10 eingeführt. Unterhalb einer isolierten Elektrode wird in einem Halbleitermaterial eine ladungsträgerfreie Zone, ein Potentialtopf, erzeugt. In diesem Potentialtopf können Minoritätsträger für relativ kurze Zeiten gespeichert werden. Diese Ladungsträger lassen sich weiter wie eine Flüssigkeit in

andere, tiefere, benachbarte Potentialtöpfe überführen. Deshalb kommt es darauf an, viele solcher Potentialtöpfe in einem Material anzuordnen, welche dann die Ladung transportieren.

Das Grundprinzip der CCD ist die *Dreiphasenanordnung* gemäß Bild 6.1.30: Auf einem *n*-leitenden Substrat sind eine Oxidschicht und darüber in dichter Folge eine Vielzahl von Metallelektroden angeordnet. Die Länge der Elektroden liegt bei einigen zehn μm, ihr gegenseitiger Abstand unter 3 μm. Die jeweils dritten Elektroden sind zusammengeschaltet und an einen Dreiphasentakt angeschaltet. Für vier verschiedene Zeiten ist die Potentialtopftiefe und die Lage bzw. Bewegung der Löcher (Minoritätsträger) im Bild angedeutet.

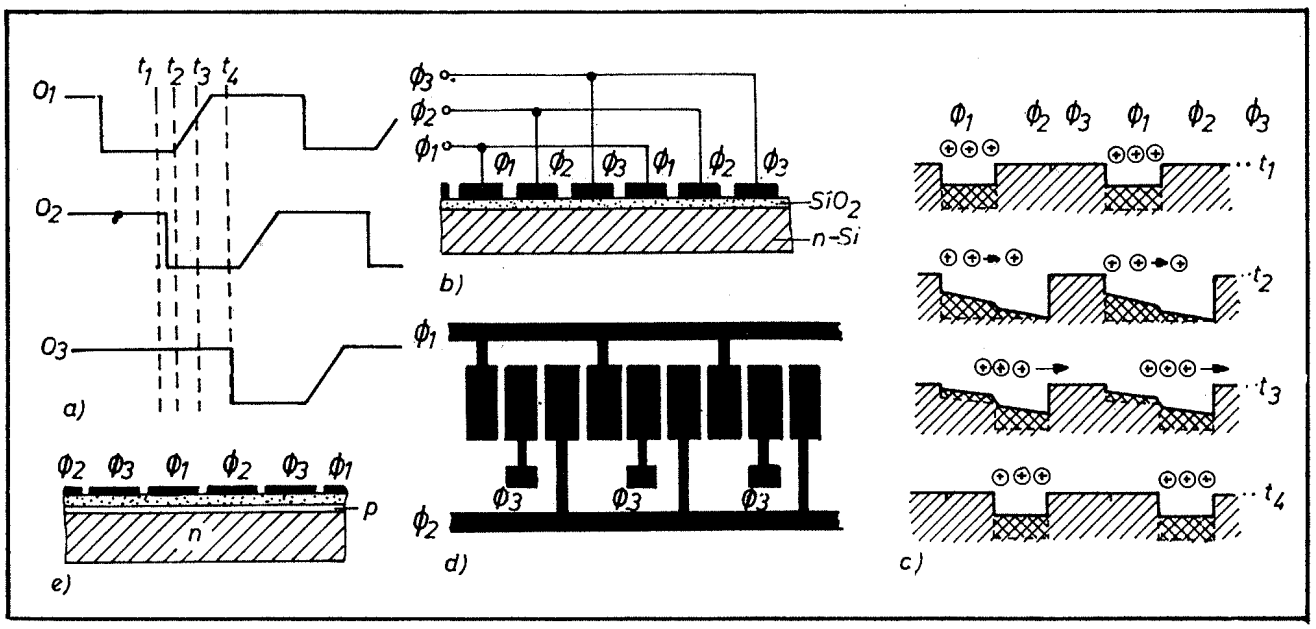


Bild 6.1.30. Aufbau und Wirkungsweise von Dreiphasen-CCD

a) Zuordnung der drei Taktsignale
b) Ausschnitt aus einem Bauelement
c) Potentialtöpfe und Bewegung der Ladungsträger zu den vier Zeiten im Teilbild a). Von t_1 bis t_4 sind die Ladungen genau um eine Elektrode weitertransportiert
d) Lay-out der Leiterstruktur. Die Anschlüsse für die Phase Φ_3 müssen gebondet oder in einer zweiten Leiterebene verbunden werden
e) Bauelement mit begrabenem Kanal

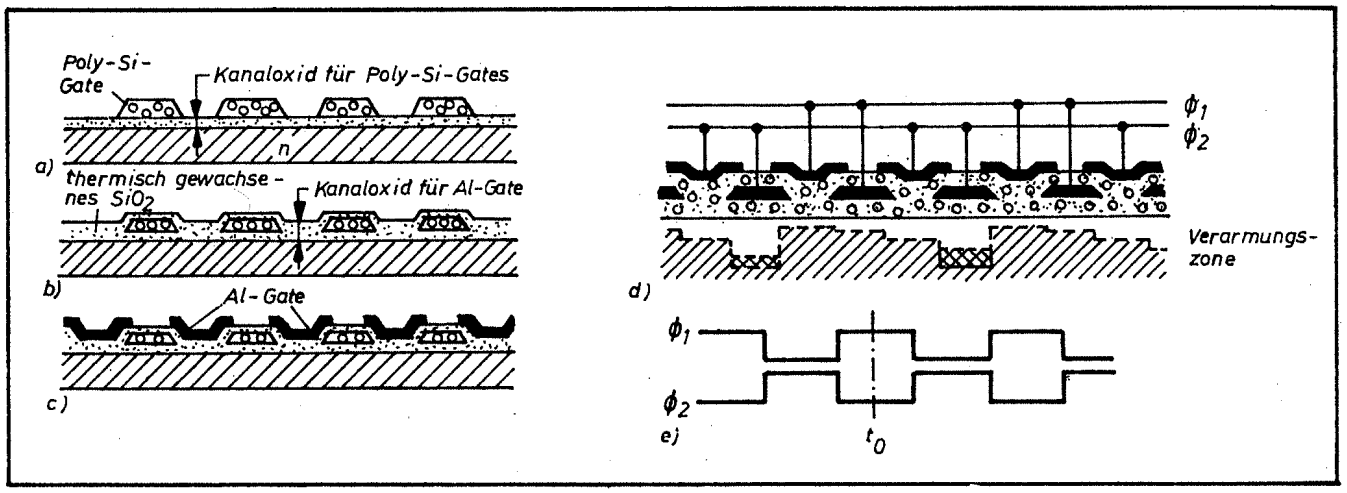


Bild 6.1.31. Zweiphasen-CCD

a) bis c) Herstellungsschritte
d) Schaltung und Verarmungszone zum Zeitpunkt t_0
e) Taktspannungen

Ein Problem der Dreiphasenanordnung wird über Bild 6.1.30d) deutlich: Die Verbindung einer der drei Taktleitungen muß gebondet oder in einer zweiten, isolierten Leiterebene realisiert werden. So wurden andere Strukturen entwickelt. Bedeutung hat hauptsächlich die *Zweiphasenstruktur* erreicht, wie sie Bild 6.1.31 zeigt. In den Teilbildern a) bis c) ist der technologische Ablauf angedeutet. Auf das mit SiO_2 überzogene *n*-Si-Substrat wird polykristallines Silicium abgeschieden (a)) und dann erneut mit SiO_2 überdeckt (b)). Zwischen die Poly-Si-Gates werden darüber hinaus dann Al-Gates aufgebracht (c)). Die Zusammenschaltung der Strukturen zum Zweiphasenbetrieb zeigt Teilbild d). Gleichzeitig enthält es die entstehende Verarmungszone zum Zeitpunkt t_0 im Taktschema (Teilbild e)). Infolge unterschiedlicher Schwellenspannungen von Al und Poly-Si sowie unterschiedlichen Abstands zum Substrat wirkt die gleiche Taktspannung an beiden unterschiedlich, und bezüglich der Potentialtopftiefe liegt ein *Vierphasenbetrieb* vor.

Nach diesem Prinzip *entfallen* die *Nachteile*

- des Dreiphasenbetriebes mit der komplexen Zusammenschaltung der Leiterstrukturen,
- des lithografisch schwer zu realisierenden Abstandes zwischen den Leiterstrukturen unter 3 μm.

Dafür *entsteht* aber ein *Nachteil*:

Es ist nicht mehr elektronisch die Transportrichtung der Ladungen umzuschalten.

6.1.5.2. Verluste, Grenzen und Anwendungen

Beim *Transport der Ladung* entstehen verschiedene *Verluste*:

- Bleiben die Ladungen zu lange unter einer Elektrode stehen, so erfolgt Rekombination mit thermisch generierten Majoritätsträgern.
- Haftstellen im Halbleitermaterial und besonders an dessen Oberfläche halten die Minoritätsträger länger fest. Dadurch gelangen einige Ladungen schneller zu einem Ort, als sie sich von dort entfernen.

Beide Fakten und weitere Störeinflüsse haben zur Folge, daß die *Taktfrequenz* weder zu groß noch zu niedrig sein darf. Der Bereich wird durch viele Einflüsse, wie z.B. Material, Trägerart (Löcher oder Elektronen als Minoritätsträger) und geometrische Abmessungen, bestimmt und liegt etwa je nach Ausführung im Bereich von *einigen kHz bis zu mehreren MHz*. Schneller und günstiger als die in den Bildern gezeigten *n*-Substrate sind *p*-Substrate, da dann Elektronen als Minoritätsträger auftreten. Sie sind jedoch technologisch schwieriger beherrschbar. Die besonders stark störenden *Oberflächenterme* lassen sich teilweise mittels eines *vergrabenen Kanals* (burried channel oder bulk-CCD) gemäß Bild 6.1.30e) vermeiden. Der *pn*-Übergang wird gesperrt betrieben, und die obere Grenze für die Ladungen ist der *pn*-Übergang. Hier existieren weniger Haftstellen als am Übergang $Si:SiO_2$. Eine andere Methode zur Vermeidung der Oberflächenterme besteht darin, generell eine *Grundladung* (fat-zero) von ca. 10% des Maximums zu transportieren, welche dann ohne störenden Einfluß an den Haftstellen verzögert verbleiben kann. Allerdings wird dadurch der Dynamikbereich für die zu transportierenden Ladungen erheblich eingeschränkt. Eine Weiterentwicklung sind die *Peristaltic-CCD* (PCCD). Hierbei wird erreicht, daß die schwer zu übertragende Restladung nicht an der Oberfläche, sondern in der Tiefe des Substrates liegt. Dieses Prinzip wurde noch weiter vervollkommnet durch zwei unterschiedlich dotierte Schichten für den Ladungstransfer (profiled PCCD = P^2CCD).

Ein wesentliches Kennzeichen für die Güte von CCD-Bauelementen ist der *Übergangsverlust je Stufe*. Er liegt bei ca. 10^{-3} ohne und 10^{-4} mit fat-zero, mit begrabenem Kanal bei 10^{-5}. Dieser Verlust ε bestimmt mit dem maximal zulässigen Gesamtverlust V die *Anzahl der* möglichen *Stufen* n:

$$1 - V = (1 - \varepsilon)^n \tag{29}$$

bzw.

$$n = \frac{\ln (1 - V)}{\ln (1 - \varepsilon)} \approx V/\varepsilon \,. \tag{30}$$

Angewendet werden heute Bauelemente bis zu ca. 1000 Stufen.

Die *Hauptanwendungen* der CCD liegen auf drei Gebieten:

- *Lichtempfindliche Sensoren*. Sie werden hier weiter behandelt.
- *Schieberegister* für analoge und digitale Schaltungen. Hierauf wird in den Abschnitten 7.8. und 8.3.6.1. eingegangen. Sie werden vielfach als dynamische Speicher hoher Kapazität verwendet.
- *Transversalfilter* für die Signalverarbeitung werden im Abschnitt 7.9.6. berücksichtigt.

Bei den Lichtempfängern werden die zu transportierenden Ladungen über den inneren Fotoeffekt erzeugt, während bei den beiden anderen Anwendungen ein in Durchlaßrichtung betriebener *pn*-Übergang die Ladungspakete erzeugt. Die Ausgangssignale werden in allen Fällen durch einen in Sperrichtung betriebenen *pn*-Übergang übernommen. Bei den Transversalfiltern werden schließlich einzelne Finger in unterschiedlicher Weise unterteilt (Bild 6.1.30d)).

Bei den CCD hatte man ursprünglich eine sehr schnelle Entwicklung und breite Anwendung erwartet. Technologische Probleme bewirkten jedoch, daß die Erfolge bis jetzt hinter den Erwartungen zurückblieben.

6.1.5.3. Lichtsensoren

Die *Lichtsensoren* werden in drei Gruppen eingeteilt:

- CCD-Zeilen
- CCD-Flächen (Bildaufnahme)
- Farbsensible CCD-Bildmatrizen.

Den Grundaufbau einer *CCD-Zeile* zeigt Bild 6.1.32a). Im oberen Teil des Bildes befinden sich die lichtempfindlichen Bereiche. Sie werden mit einem Dreiphasentakt ausgelesen und speichern während der Taktperiode die durch Licht generierten Minoritätsträger. Um genau definierte Lichtflecke zu erhalten, befindet sich zwischen den einzelnen Bildelementen die *Kanal-Stop-Diffusion*. Nachdem die der Lichtintensität entsprechenden Ladungspakete nach unten in den lichtgeschützten Transferbereich übergeben sind, werden sie im Dreiphasentakt zur Ausgangselektrode übertragen. Dort steht also sequentiell das Signal aller Sensoren zur Verfügung.

Die Wirkung des lichtempfindlichen Bereiches in zwei Zuständen zeigt Bild 6.1.32b) und c).

Diese soeben beschriebene Struktur ist heute meist etwas abgewandelt (Bild 6.1.32d)): Der sensorische Teil befindet sich in der Mitte des Chips, und die Ladungen werden *abwechselnd nach rechts und links übertragen*. Dadurch ist die lange Transferkette auf die Hälfte verkürzt, und die Verluste sind entsprechend kleiner. Ferner ergibt sich auf dem Chip eine günstigere Anordnung. Einen Ausschnitt

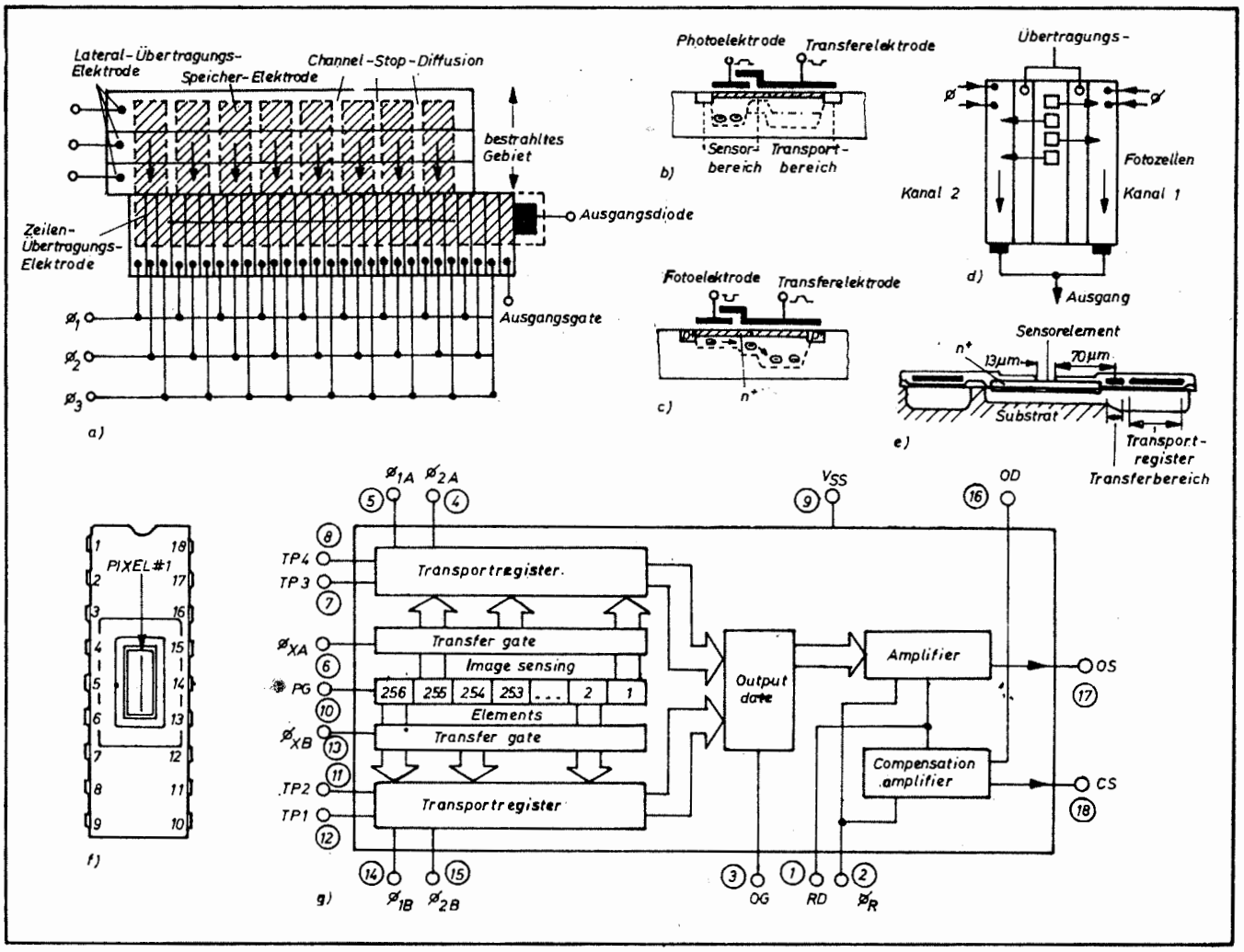


Bild 6.1.32. Zum Aufbau von CCD-Sensorzeilen

a) Prinzipielle Anordnung
b) Ladungserzeugung durch Licht
c) Übertragung der Ladung in den Transferbereich
d) Heute übliche Anordnung mit zwei Transferbereichen
e) Querschnitt durch den Chip d)
f) Aussehen einer vollständigen CCD-Zeile
g) Prinzipschaltbild der vollständigen Zeile

aus einem solchen Chip zeigt Bild 6.1.32e). Solche Zeilen werden bis ca. 3000 Bildpunkte hergestellt und finden bei vielen Erkennungsprozessen Anwendung. Sie enthalten in der Regel auf dem Chip zugleich Verstärker- und Takthilfselektronik, wie es beispielhaft Bild 6.1.32f) und g) zeigt.

Die Belichtung der *lichtempfindlichen Flächen* (pixel oder piksel) kann von der Vorder- oder Rückseite her erfolgen. Von der Vorderseite sind die Elektroden und das SiO_2 zu durchstrahlen. Bei Belichtung von der Rückseite muß der Chip auf etwa 10 bis 20 µm verdünnt werden. Beide Prinzipien haben ihre Vor- und Nachteile, so daß z. Z. keine Entscheidung für einen Fall vorliegt. Bei hoher *Blauempfindlichkeit* ist die Rückseitenbelichtung etwas im Vorteil.

Für die Bildaufnahme, z.B. im Sinne von Fernsehsignalen, müssen Bildpunkte und aus ihnen zusammengefaßte Zeilen unterschieden werden. Für diesen Zweck haben sich *drei Strukturen*, die Bild 6.1.33 schematisch zeigt, durchgesetzt:

- Die *Bildübertragung* ist besonders übersichtlich. Sie hat aber einen recht hohen Platzbedarf und neigt bei starker Belichtung zu Störungen im Bildaufbau.
- Bei der *Zeilenübertragung* existiert ein Abtastgenerator, der zeilenweise die Bildinformation in das Ausgangsregister schiebt.

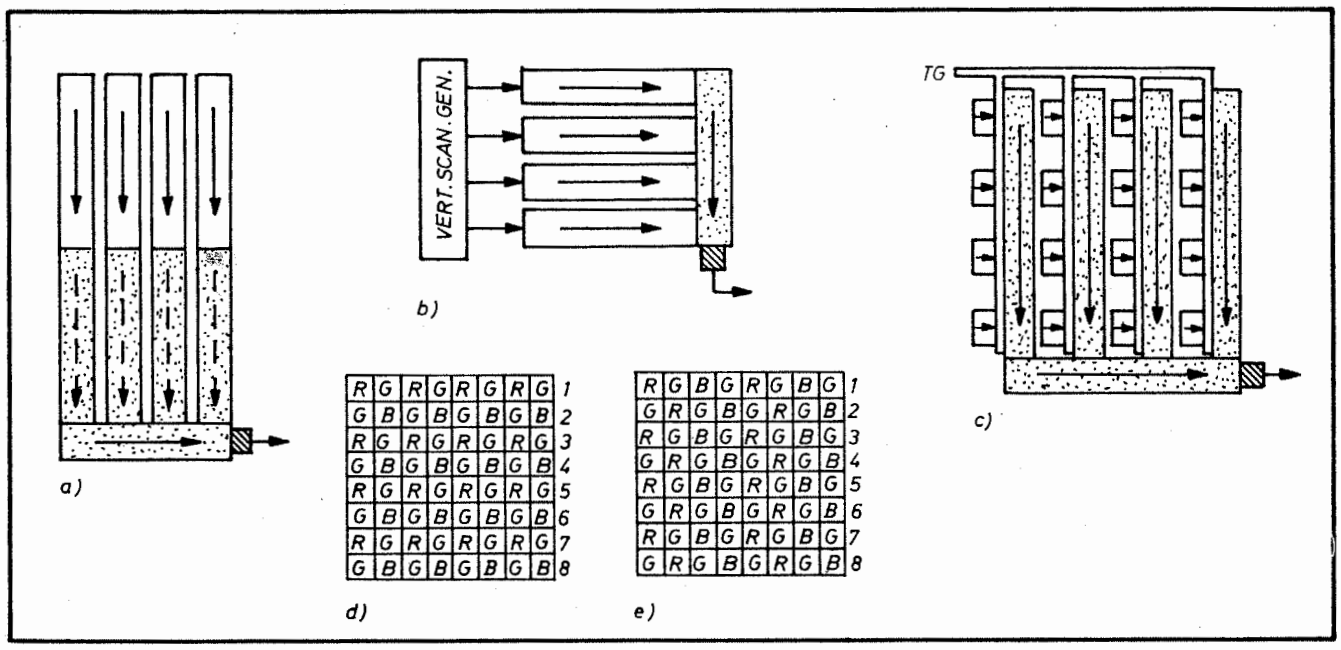


Bild 6.1.33. CCD-Bild-Matrizen mit a) Bildübertragung, b) Zeilenübertragung, c) Spaltenauswahlprinzip, d) Anordnung der Farbfilter nach BAYER sowie e) Interline-Geometrie

- Beim *Spalten-Auswahl-Prinzip* sind die einzelnen lichtempfindlichen Bildzeilen durch lichtverdeckte Zwischenbereiche unterbrochen. Es liegen eigentlich viele nebeneinandergelegte Zeilen vor. Der Nachteil der weniger kompakten Bildfläche wird dadurch aufgewogen, daß hier prinzipiell keine Bildverschmierung auftreten kann.

Mit den bisher beschriebenen Bildmatrizen kann nur ein Schwarzweißbild übertragen werden. Immer größer wird aber der Anwendungsbereich des Farbfernsehens. Deshalb mußten auch für diesen Zweck CCD-Matrizen geschaffen werden. Hierbei werden in der Regel die bisher beschriebenen *Matrizen mit Farbfiltern* versehen. Besonders häufig werden zwei Varianten verwendet (Bild 6.1.33d) und e)).

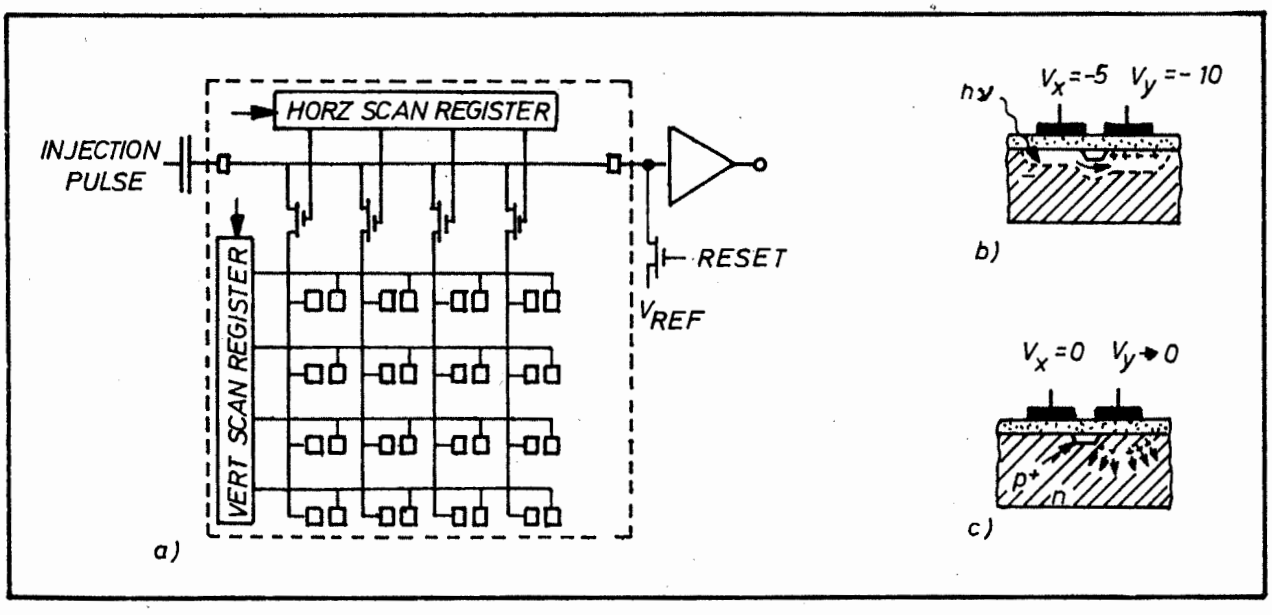


Bild 6.1.34. CID-Bild-Matrizen

a) Prinzipieller Aufbau

b) Einzelzelle während der Ladungsspeicherung unter Lichteinfluß und während des Ladungstransfers auf den Y-Teil

c) Zustand der Einzelzelle während der Ladungsinjektion

Seit einiger Zeit ist eine abgewandelte Bildsensorstruktur, die *Charge injection device* (CID), bekannt geworden. Hier wird echt das Matrizenprinzip mit spalten- und zeilenweiser Ansteuerung genutzt. Jedes lichtempfindliche Element ist durch zwei Elektroden, je eine für die Spalte und die Zeile, gekennzeichnet (Bild 6.1.34c)). Im Teilbild b) wird der Zustand der Ladungsspeicherung und des Ladungstransfers gezeigt. Durch Spalten- und Zeilenauswahl (c)) wird die gespeicherte Ladung einer Zelle im Substrat injiziert und kann damit am Verstärker registriert werden.

Die Bildpunktzahl der meisten Sensoren lag 1982 in der Größenordnung der Fernsehnorm, ca. 500 (vertikal) mal 600—700 (horizontal), rund 300000 Punkte. Bei den Farbmatrizen muß daher im Halbzeilenverfahren gearbeitet werden.

6.1.6. Flüssigkristalle

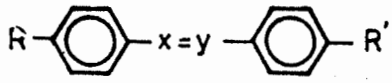
Die flüssigen Kristalle wurden 1888 entdeckt und galten bis etwa 1960 als wissenschaftliche Kuriosität. Bereits 1970 wurden etwa 120 Anwendungsmöglichkeiten registriert. Inzwischen dürften weitaus mehr hinzugekommen sein.

Der Begriff flüssiger Kristall scheint zunächst ein Widerspruch in sich zu sein. Um ihn zu verstehen, muß zwischen Nah- und Fernordnung unterschieden werden. In einem Kristall sind die Atome bzw. Moleküle über große Entfernungen im Vergleich zu ihrer Größe geordnet. Im Einkristall werden keine Fehler zugelassen.

Bei den Flüssigkeiten liegen oder bewegen sich die Atome bzw. Moleküle frei verschiebbar in einem festen Volumen. Ihre Verteilung ist zwar nicht ganz statistisch wie z. B. in einem Gas, aber die Nahordnung reicht nur bis zu wenigen Moleküldurchmessern und ergibt selbst dort nur eine gewisse Regelmäßigkeit.

In den Flüssigkristallen existiert dagegen eine gewisse Fernordnung bis etwa in den Bereich von zehntel Millimetern. Diese Ordnung ist außerdem leicht durch äußere Bedingungen, wie z. B. Temperatur, elektrisches und magnetisches Feld oder epitaktische Wirkungen der Gefäßwände zu beeinflussen. Es sind viele tausend organische Substanzen bekannt, die Flüssigkristalle bilden. Es wird angenommen, daß ca. 5% aller organischen Flüssigkeiten hierzu fähig sind.

Die *Mehrzahl der Substanzen* für Flüssigkristalle sind gemäß der chemischen Formel



(31)

aufgebaut. Dabei können für die Mittelgruppe $-x = y-$ unter anderem auftreten

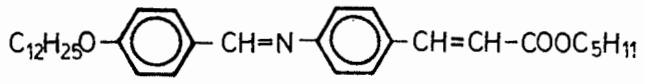
$-N=NO-$	Azoxybenzole,
$-CH=N-$	SCHIFFsche Basen,
$-CH=CH-$	Stilebene,
$-O-CO-$	aromatische Ester,
ohne	Biphenole.

Für die Flügelgruppen treten u. a. auf

$-(CH_2)_nCH_3$	Alkyl-Gruppe,
$-O(CH_2)_nCH_3$	Alkoxy-Gruppe,
$-OCO(CH_2)_nCH_3$	Acyloxy-Gruppe,
$-C=N$	Nitrilgruppe.

Als *Beispiele* von flüssigen Kristallen seien genannt:

4-(4-n-Dodecyloxybenzalamino)-zimtsäure-n-amylester

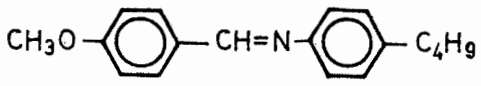


(32)

74 °C 95 °C 107 °C 134 °C

fest ↔ smektisch B ↔ smektisch C ↔ smektisch A ↔ isotrop;

4-(4-Methoxybenzalamino)-n-butylbenzol (MBBA)

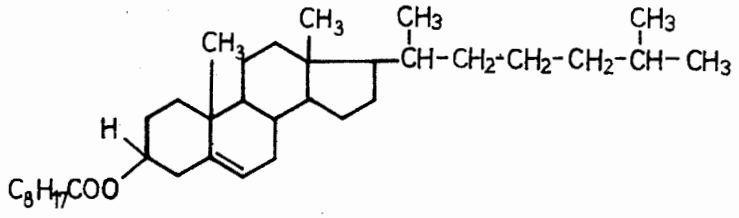


(33)

20 °C 47 °C
fest ↔ nematisch ↔ isotrop.

Es gibt auch Flüssigkeitskristalle mit *abweichender Struktur*. Hier zwei Beispiele:

Cholesterylnonanoat

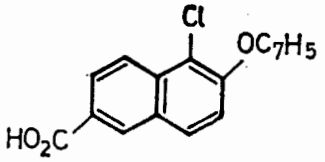


80 °C 90 °C
fest ↔ cholesterinisch ↔ isotrop;

↕ 76 °C

↑——— smektisch

5-Chlor-6-n-heptyloxy-2-naphthalinsäure



165,5 °C 176,5 °C 201 °C
fest ↔ smektisch ↔ nematisch ↔ isotrop.

Entsprechend ihrem Grundaufbau können die meisten Moleküle der Flüssigkristalle als *langgestreckte Toroide* aufgefaßt werden. Diese Struktur und die dadurch zwischen ihnen mögliche Wechselwirkung ist die Ursache des Auftretens von Flüssigkeitskristallen. Sie besitzen dann im Gegensatz zur isotropen Flüssigkeit eine *Anisotropie*.

Der flüssigkristalline Zustand liegt jeweils zwischen dem *Schmelzpunkt* einer Substanz und dem *Klärpunkt*. Vom Klärpunkt an liegt eine übliche isotrope Flüssigkeit vor. Da die flüssigkristallinen Substanzen in dicken Schichten stets trüb erscheinen, ist der Name Klärpunkt verständlich (Bild 6.1.35a)).

Es gibt drei verschiedene *Arten von Anisotropie* (Mesophasen) (Bild 6.1.35):

Bei den *smektischen* Phasen gibt es mehrere Zustände. Sie stehen dem festen Kristall am nächsten. Die Schichtstruktur stellt eine zweidimensionale Ordnung dar. Eine Parallelverschiebung zwischen den Schichten ist leicht möglich. Eine Verschiebung in die Längsrichtung der Kristalle ist nicht möglich. Derartige Flüssigkristalle sind hochviskose Pasten. Sie besitzen keine größere Bedeutung für Anwendungen.

Bei den *nematischen* Phasen sind die Moleküle mehr oder weniger gut parallel ausgerichtet. Ist Θ der Winkel eines Moleküls gegenüber der Symmetrieachse des ganzen Flüssigkeitskristalls, so kann der Ordnungsgrad

$$S = 1 - \tfrac{3}{2}\overline{\sin^2 \Theta} \tag{36}$$

eingeführt werden. Für die ideale nematische Phase beträgt er 1, für reale liegt er bei 0,7 ... 0,8, und bei Temperaturerhöhung folgt bei ungefähr 0,4 sprunghaft der Klärpunkt mit $S = 0$.

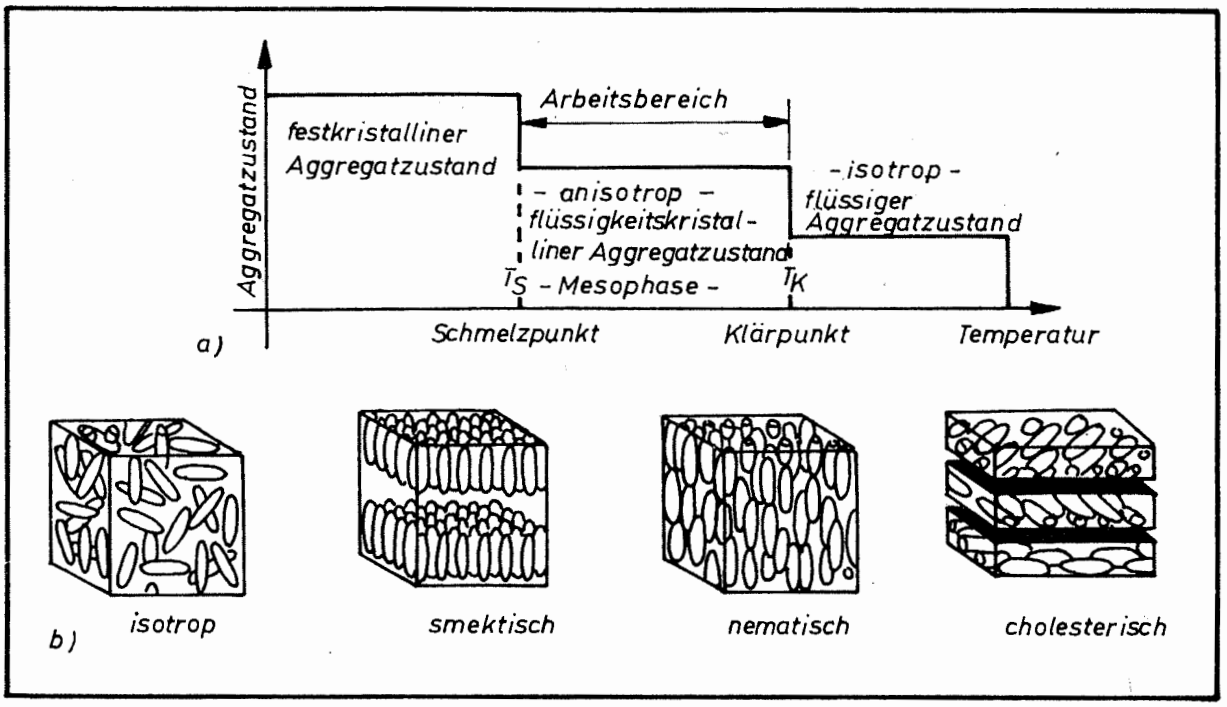


Bild 6.1.35. Zum Flüssigkeitskristall

a) Als Zustand vieler organischer Substanzen zwischen Festkörper und Flüssigkeit

b) Prinzipielle Anordnung der Molekülgruppen in der Flüssigkeit (isotrop) und bei den drei Anisotropien der Flüssigkeitskristalle

Die *cholesterische* Phase ist eine verdrillte nematische. Senkrecht zur Längsachse der Moleküle entspricht die Richtung der Molekülgruppe einer Schraube oder Wendeltreppe. Die Periodenlänge hängt von vielen Einflüssen ab und variiert beträchtlich. Werte zwischen 0,2 und 20 μm je Windung wurden beobachtet.

Infolge der Anisotropie ihrer Struktur sind verschiedene *Eigenschaften* der Flüssigkeitskristalle ebenfalls *richtungsabhängig*, u. a. sind dies:

Dielektrizitätskonstante; Brechungsindex; Leitfähigkeit; Elastizität.

Für die Anwendung der Flüssigkeitskristalle als Anzeigeelemente (LCD = liquid crystal display) ist vor allem die elektrische *Beeinflußbarkeit dieser Eigenschaften* wichtig. So können verändert werden:

Streuung, optische Dichte, Farbe.

Solche Änderungen beruhen auf *unterschiedlichen Effekten*, von denen hier die fünf bedeutsamsten ausgewählt sind. Die ersten beiden haben größere Bedeutung erlangt. Die anderen drei sind in der Erforschung und Erprobung (vgl. Tab. 6.1.9).

Die *dynamische Streuung* (DSM = dynamic scattering mode, auch elektrodynamischer Effekt) war die erste Anwendung von LCD. Sie erfolgte ab 1973 in Japan als Massenproduktion für Rechneranzeigen. Heute hat sie kaum noch Bedeutung.

Beim Anlegen einer Wechselspannung entstehen im nematischen Flüssigkeitskristall Turbulenzen. An ihnen wird Licht durch Doppelbrechung gestreut. Der zuvor klare Flüssigkeitskristall erscheint an diesen Stellen im Durchlicht schwarz, in Reflexion milchig-weiß. Die Hauptnachteile dieses Effektes sind die relativ große Spannung und die Leistungsaufnahme infolge der Energiezufuhr für die Aufrechterhaltung der Turbulenz.

Der *Feldeffekt* wird seit etwa 1976 angewendet. Er ist heute der am meisten angewendete Effekt. Für ihn sind auch die Namen: TN-FEN (= twisted nematic field effect mode), SCHADT-HELFRICH-Effekt, Steuerung der optischen Aktivität, Drehung der Polarisationsebene usw. gebräuchlich. Er wird im Abschnitt 6.1.6.1. genauer beschrieben.

Tabelle 6.1.9. Effekte in Flüssigkristallen

Effekte / Eigenschaften	Dynamische Streuung (DSM)	Feldeffekt (TN-FEM), auch SCHADT-HELFRICH-Effekt	Doppelbrechung (DAP) (ECB), auch: FREEDERICKSZ-Effekt	Guest-host-Effekt (G—H)	Phasenänderung (P—C)
Funktionsprinzip	Streuung	Polarisationsdrehung	Doppelbrechung Interferenz	Absorption	Streuung
Einschaltzeit	5 ... 40 ms	2 ms	1 ... 10 ms		
Ausschaltzeit	30 ... 400 ms	100 ms	100 ms		3 ... 12 h (Speicherzeit)
Kontrastverhältnis	20:1	30:1	4000:1	>10:1	
Steuerspannung	hoch (15 ... 30 V)	1 ... 10 V	5 ... 10 V	5 ... 20 V	hoch (12 ... 14 V)
Leistungsaufnahme	groß (0,5 mW/cm²)	niedrig (2 μW/cm²)		niedrig (10 μW/cm²)	
Polarisationsfilter nötig	—	+	+	—	—

Die *elektrisch gesteuerte Doppelbrechung* wird auch ECB = electrically controlled bifrequence oder DAP = Deformation aufgerichteter Phasen (= deformation of aligned phases) genannt. Hierbei sind die cholesterinischen Molekülstrukturen im Ruhezustand in der Ebene der Anzeigefläche ausgerichtet. Das elektrische Feld der Elektroden zwingt diese Strukturen dann mehr oder weniger in Richtung der Elektroden. Dadurch wird die Doppelbrechung der Struktur verändert. Dies kann mittels Polarisationsfolien sichtbar gemacht werden. Dabei sind auch Farbinterferenzen nutzbar.

Beim Übergang von Licht aus einem Medium mit dem Brechungsindex n_1 in ein Medium mit dem Brechungsindex n_2 verhalten sich die zugehörigen Wellenlängen und Lichtgeschwindigkeiten wie folgt:

$$\frac{n_1}{n_2} = \frac{c_2}{c_1} = \frac{\lambda_1}{\lambda_2}. \tag{37}$$

Wenn ein Medium anisotrop ist, so ist der Brechungsindex von der Richtung, in der sich das Licht bewegt, abhängig. Deshalb spaltet sich ein Lichtstrahl in zwei senkrecht zueinander polarisierte Richtungen auf, die sich mit unterschiedlicher Geschwindigkeit und Wellenlänge fortbewegen. Dieser Effekt heißt Doppelbrechung und verschwindet nur in der optischen Achse des anisotropen Mediums. Durch Polarisationsfilter ist es möglich, beide Wellen zu trennen. Eine Beeinflussung der Anisotropie verändert folglich unter sonst konstanten Bedingungen die Intensitätsverteilung auf beide Strahlen und kann so über Polarisationsfilter sichtbar gemacht werden.

Ein wenig abgewandelter Effekt mit Doppelbrechung ist der FREEDERICKSZ-Effekt. Im Gegensatz zum oben genannten Effekt sind hier die Moleküle parallel zu den Oberflächen ausgerichtet.

Die *Steuerung der Absorption*, auch G-H-Effekt (guest-host, Gast-Wirt) oder Bistabilitätseffekt genannt, geht von Flüssigkeitskristallen (Wirt) aus, die Farbstoffe (Gast) enthalten. Gesteuert wird die Wechselwirkung zwischen diesen beiden Komponenten. Hierdurch ändert sich die Lichtabsorption, so daß keine Polarisationsfolien

notwendig sind. Durch die Wahl der Farbstoffe sind viele Farben realisierbar. Im feldlosen Zustand absorbieren die Farbstoffe recht stark. Durch das Feld wird eine Ordnung der nematischen Struktur erzwungen, und es verbleibt wesentlich weniger Absorption.

Die *Phasenänderung zwischen cholesterisch und nematisch* (P—C = phase changing) nutzt die Änderung der Lichtstreuung bei der Phasenänderung aus. Der Übergang cholesterisch → nematisch wird durch das elektrische Feld bei konstanter Temperatur bewirkt. Es ist also ein Flüssigkeitskristall notwendig, bei dem beide Phasen nebeneinander bestehen können. Dieser Effekt hat den Vorteil, daß er auch als Speichereffekt auftreten kann. Die gesteuerte Phase ist auch dann noch vorhanden, wenn das Feld wieder verschwindet.

6.1.6.1. Aufbau und Arbeitsweise von TN-Zellen

Den *prinzipiellen Aufbau* einer Flüssigkristall-Anzeige für eine Siebensegmentziffer zeigt Bild 6.1.36a) bis c). Auf zwei Glasplatten sind die entsprechenden *Elektroden* mit Zinn- oder Indiumoxid aufgebracht. Es genügt, wenn die eine Seite die einzelnen Segmente enthält und die Gegenelektrode die ganze Fläche trägt. Oberhalb der gut halbleitenden Strukturen befindet sich *SiO zur Isolation*. Hierdurch soll vermieden werden, daß bei ungewolltem *Gleichstrombetrieb* in der Flüssigkeit bzw. an den Elektroden elektrolytische Prozesse ablaufen. Die beiden Glasplatten werden mit den

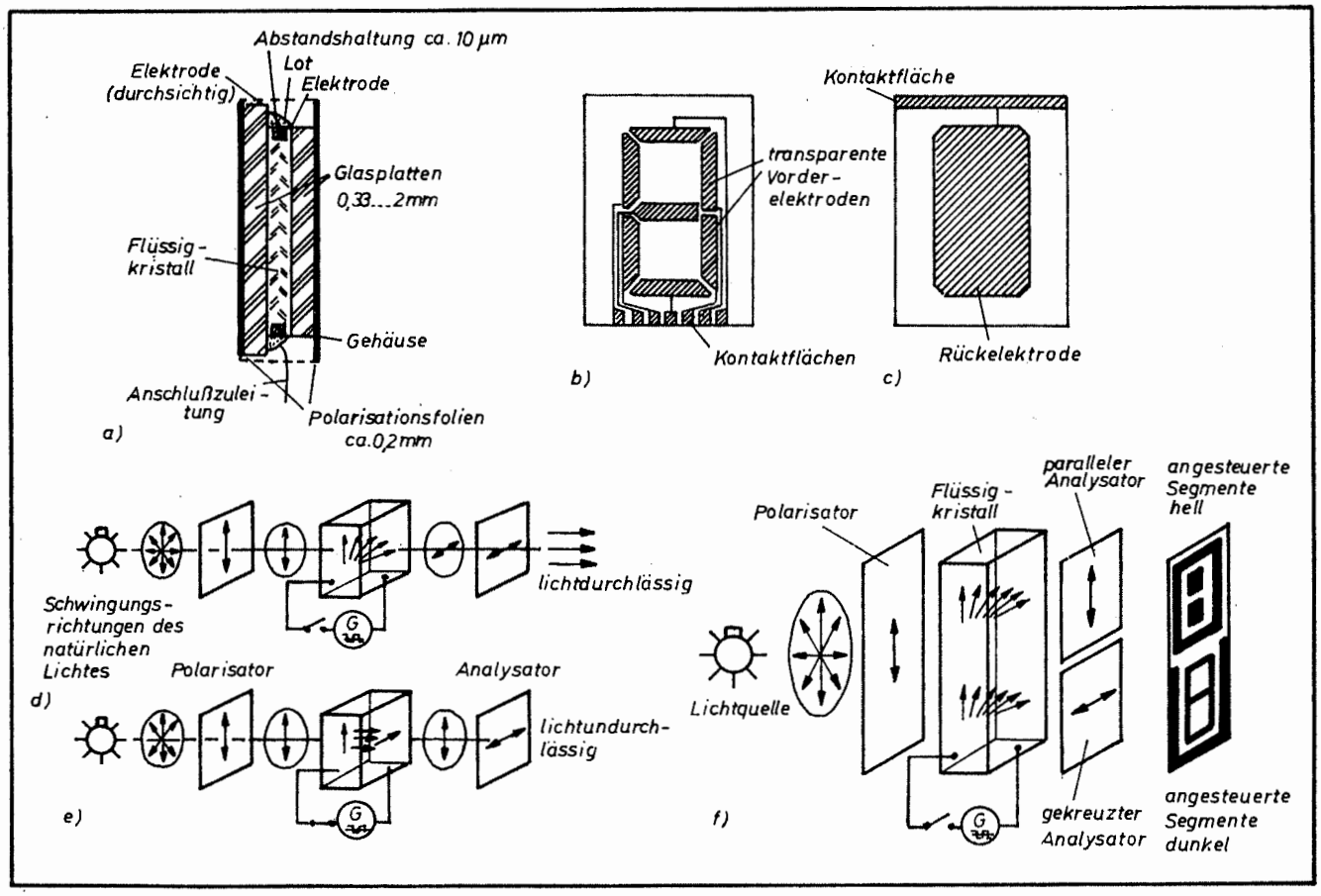


Bild 6.1.36. Aufbau und Funktion eines TN-LCD

a) Querschnitt durch die Zelle
b) Durchsichtige Maske mit den Strukturen
c) Durchsichtige Maske mit voller Fläche
d) Wirkungsmechanismus der Zelle ohne Spannung,
e) mit angelegter Spannung(die Zelle in der Mitte ist in d) und e) ohne Strukturierung angenommen)
f) Schema der Anzeige mit parallelen (oberer Weg) und gekreuzten (unterer Weg) Polarisationsfiltern

Elektrodenflächen nach innen auf etwa 10 µm *Abstand* montiert. In den Hohlraum wird der Flüssigkristall (er besteht meist aus einer Mischung vieler Substanzen) eingefüllt und nach außen hermetisch abgeschlossen. Für extreme Umweltbedingungen erfolgt dies durch Glaslot. Bei den normalen Anwendungen hat sich weitgehend eine hochentwickelte Klebetechnik durchgesetzt. Von der Qualität dieses Abschlusses hängt nämlich wesentlich die Lebensdauer der LCD-Zelle ab. Heute sind ca. 50000 Stunden der Normalfall. Abschließend wird auf die beiden Außenseiten *Polarisationsfolie* geklebt, in der Regel Folie mit zueinander senkrechter Orientierung.

Das SiO ist vor dem Einbau noch *in spezieller Weise bearbeitet.* Es erhält mittels unterschiedlicher Methoden (z.B. Schrägbedampfen oder Reiben) eine spezielle *Textur,* die so gestaltet ist, daß die nematischen Partikeln hieran *zwangsorientiert* sind, und zwar an beiden Seiten um 90° verdreht. Dadurch nehmen die *Teilchen eine stetige Drehung* ein, wie es in der Mitte von Bild 6.1.36d) gezeigt wird. Ohne Spannung wird in der Zelle die Ebene des Lichtes auf Grund der Anisotropie um 90° gedreht. Dadurch kann das Licht bei gekreuztem Polarisator und Analysator die Zelle passieren. Wird nun eine *Wechselspannung* an die Elektroden der Zelle angelegt, so wird die *kontinuierliche Drehung* der Partikeln in der Zelle weitgehend dadurch *aufgehoben,* daß sich die Partikeln in Längsrichtung des Feldes stellen. Das Licht passiert die Flüssigkristalle folglich ohne Drehung und wird daher im Analysator absorbiert (Teilbild e)). Teilbild f) zeigt unter diesen Gesichtspunkten die Abbildung der Elektrodenstrukturen bei parallelen und gekreuzten Polarisatoren. Das Symbol erscheint hell auf dunklem Grund oder umgekehrt.

Neben dem soeben behandelten Transmissionsbetrieb kann die LCD-Zelle auch in *Reflexion* betrieben werden. Auf der nicht zu betrachtenden Rückseite befindet sich hinter der Polarisationsfolie dann zusätzlich eine Reflexfolie. In diesem Fall müssen die beiden Polarisationsfolien gekreuzt sein.

Entsprechend der Trägheit der Zellen werden für die Steuerspannung der Zellen Rechteckimpulse mit Frequenzen im Bereich von 25 Hz bis 1 kHz verwendet.

Den Aufbau einer LCD-Zelle *auf einer Leiterplatte* zeigt Bild 6.1.37a). Um mechanische Spannungen, die zum Bruch der Gläser der Zelle führen könnten, und um eine einfache, problemlose Kontaktgabe zu erreichen, wurde ein spezieller *Zebragummi* entwickelt. Er besteht aus sich abwechselnden Abschnitten von leitendem und isoliertem Gummi. Der Abstand dieser Teile liegt unterhalb eines Millimeters. Dadurch wird einerseits eine Verbindung von LCD zur Leiterplatte über mehrere Leitstücke gebildet, andererseits ist die Breite so gering, daß sie noch keinen Kurzschluß von benachbarten Kontakten bewirkt. Beim Einbau ist also lediglich auf die Zuordnung der Kontakte auf der Leiterplatte zu den Kontakten der LCD zu achten. Dies geschieht mittels der durchsichtigen Hülle. So ist ein einfacher, weitgehend automatischer Einbau möglich.

In Bild 6.1.37b) bis f) sind beispielhaft einige Strukturen bei LCD-Zellen zusammengestellt. Sie zeigen eine große Anzahl von Kontakten. Bei Zellen, welche z.B. 40 oder gar 80 Zeichen mit jeweils einer Matrix von 5 × 7 Punkten enthalten, können dies sogar 1401 bzw. 2801 Kontakte sein. Dies ist sowohl technisch als auch von der Ansteuerung her nicht realisierbar. So wurde es schon sehr viel früher notwendig, *Multiplexbetrieb* zu verwenden. Hierzu werden entsprechend Bild 6.1.38 Zeilen und Spalten zusammengefaßt und die Kreuzungspunkte zeitlich nacheinander angesteuert (Details in 8.6.5.). Dies bedeutet aber, daß die einzelnen Punkte auch nur kurzzeitig sichtbar werden. Deshalb waren u. a. besondere Bedingungen an den Kontrastverlauf und die Speichereigenschaften der Kapazitäten der einzelnen Punkte zu stellen. Eine immer häufiger angewendete Variante benutzt an den Kreuzungspunkten der Spalten- und Zeilenleitungen Dünnfilmtransistoren (9.4.3.3.) (TFT; Bild 6.1.38c)). Sie schalten

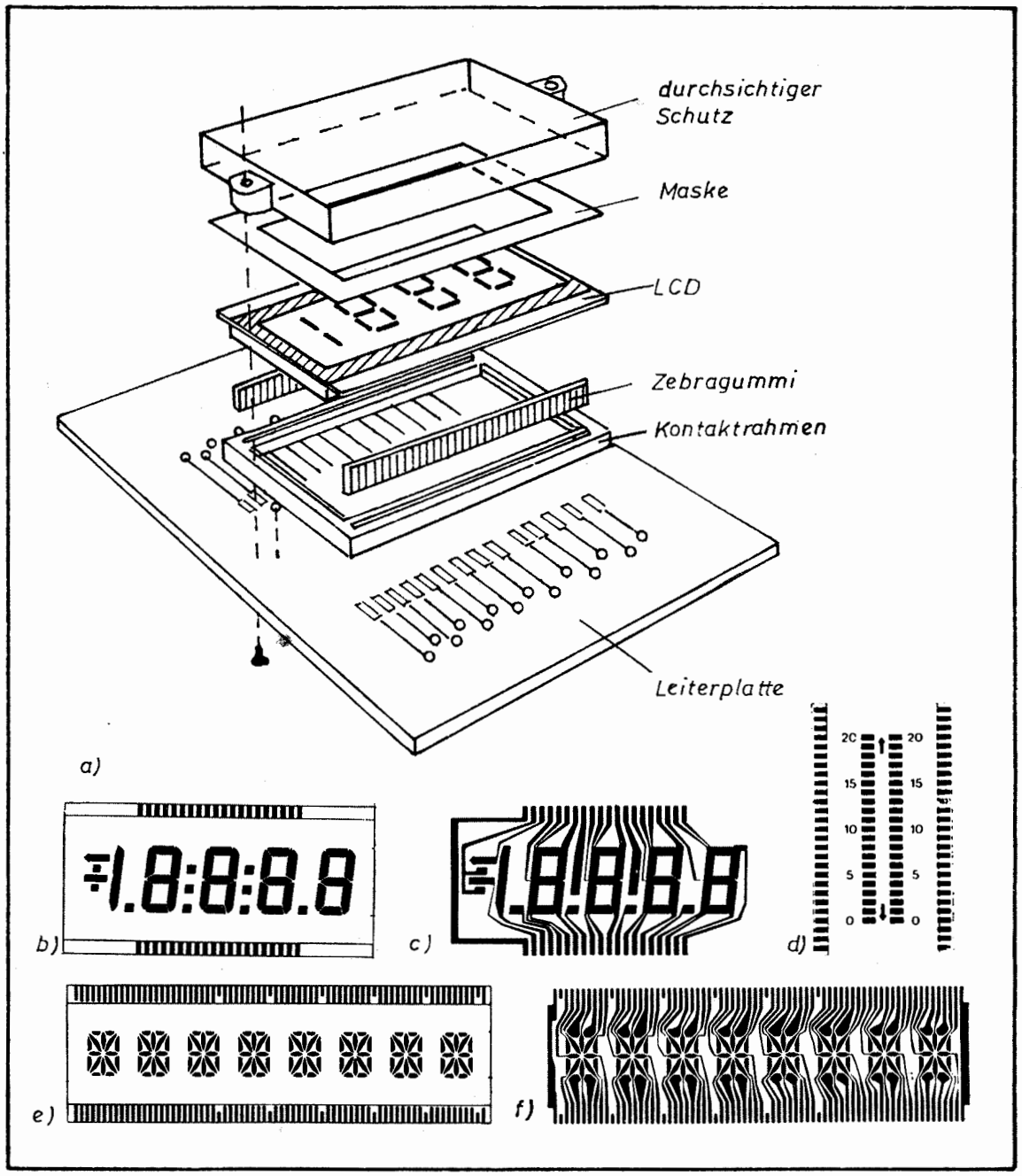


Bild 6.1.37.

a) Explosionsdarstellung einer kompletten und auf die Leiterplatte montierten LCD-Anzeige
b) bis f) Beispiele für Anzeigen und Masken bei LCD
b) Anzeige mit Anschlüssen (oben und unten) für ein Digitalvoltmeter,
c) Maske dazu
d) Doppelte Linearanzeige (rechts und links Kontakte)
e) 16-Segment-Anzeige,
f) dazugehörige Maske

kurzfristig die LCD an. Ihre Kapazität oder eine eventuell hinzugefügte Kapazität hält den eingestellten Zustand bis zum nächsten Einschaltzeitpunkt aufrecht.

In den letzten Jahren erfolgten auch Anwendungen der LCD-Technik für *flache Bildschirme* (6.1.8.). Hierbei ist der Multiplexbetrieb noch komplexer. Daher wird z. T. die Ansteuerschaltung gleich in bzw. an die LCD-Zelle integriert. Werden dabei komplexe Bausteine, z.B. gemäß Bild 6.1.38b), einbezogen, so spricht man von einem intelligenten Terminal. Es benötigt nur noch relativ wenig Anschlüsse, über die kodiert die Information zugeführt wird.

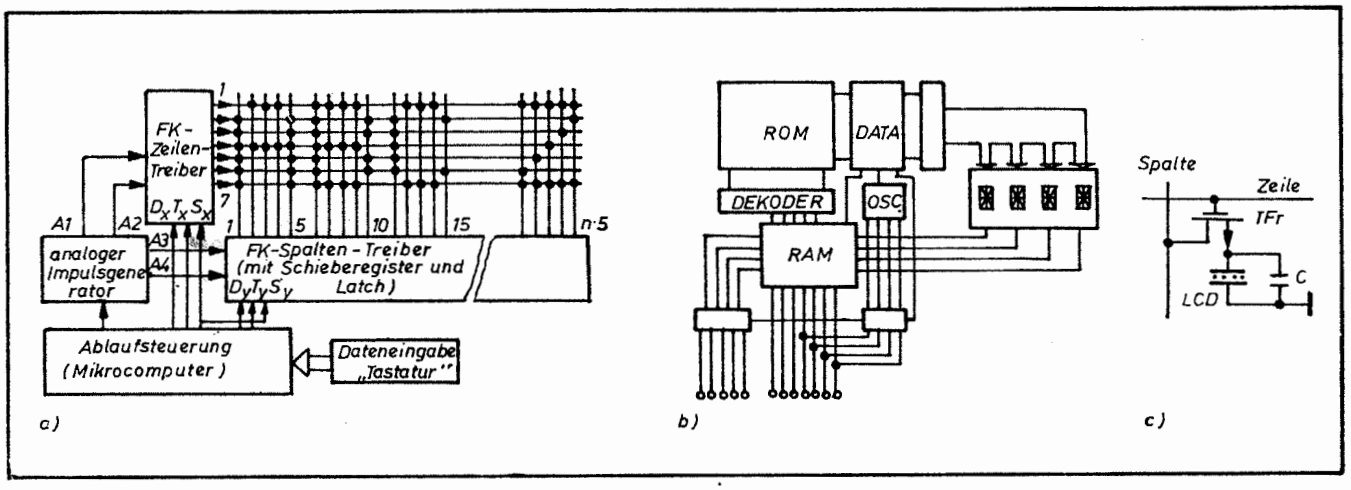


Bild 6.1.38. a) Prinzipschaltung des Multiplexbetriebs einer LCD-Zeile mit 5 × 7 Matrixelementen und b) Erweiterung der Schaltung zum intelligenten Terminal

Der *Betriebstemperaturbereich* von LCD-Bauelementen ist in erster Linie durch den Schmelzpunkt und Klärpunkt der Flüssigkristalle (Bild 6.1.35) bestimmt. In der Regel schadet es einer Zelle nicht, wenn beide Punkte zeitweilig überschritten werden. Die Substanz verhält sich in gewissem Umfang diesbezüglich reversibel. Bei tiefen Temperaturen in der Nähe des Schmelzpunktes wird allerdings die Schaltzeit der Zelle sehr groß. Aus diesen und ähnlichen Gründen geben die Hersteller Temperaturbereiche an, die enger sind. Typische Werte sind −15 bis +60 °C. Für Sonderfälle stehen aber auch Zellen im Bereich −25 bis +85 °C oder −15 bis +100 °C zur Verfügung.

6.1.6.2. Fluoreszenzaktiviertes Display

Bei den üblichen LCD geht ca. 70% des Lichtes an den Polarisationsfolien verloren. Deshalb sind sie unterhalb von 100 Lux kaum mehr ablesbar. Eine wesentliche Verbesserung ist hier mittels Fluoreszenz möglich. So entstand das fluoreszenzaktivierte Display FLAD. Seinen Aufbau zeigt Bild 6.1.39a). Das Umlicht gelangt auf eine *Fluoreszenzplatte*. Das ist ein Kunststoff, der mit fluoreszierenden Molekülen versehen ist. Sie absorbieren z.B. entsprechend der gestrichelten Kurve in Bild 6.1.39b) und emittieren Licht gemäß der ausgezogenen Kurve. Das Licht wird entsprechend der Fläche der Platte und der Absorption der Farbstoffe gesammelt und in der Platte mehrfach reflektiert. Dies fördern die Spiegel an den beiden Stirnflächen. Ausgekoppelt wird das Licht an den beiden verspiegelten Kerben gegenüber den Anzeigestellen des Displays. Die rauhen Stellen gegenüber sorgen für eine Lichtstreuung, damit die Anzeige unter einem ausreichenden Winkel beobachtet werden kann. Da hier das Licht nur bei den Zeichen passieren kann, erscheinen sie hell auf dunklem Grund. Der Vorteil des FLAD wird aus Bild 6.1.39d) sichtbar. Generell ist der Kontrast erhöht, und erkennbar werden die Zeichen bis herab zu etwa 5 Lux. Damit werden Ergebnisse erreicht, die LED und Fernsehröhre in den meisten Situationen übertreffen. Für vollständige Dunkelheit bietet sich bei der Fluoreszenzplatte schließlich noch eine weitere Möglichkeit an, nämlich ihre Anregung durch *Tritiumquellen* (Bild 6.1.39c)). Dies sind z.B. kleine Glasröhren, in denen unter erhöhtem Druck Tritiumgas eingeschlossen ist und deren Wände mit Leuchtphosphor (s. 6.2.1.6.) überzogen sind. Beim radioaktiven Zerfall eines Tritiumatoms (H_3) wird ein Elektron frei, das den Leuchtstoff anregt. Es gibt auch Tritiumleuchtfarbe, in der H_3 chemisch gebunden ist. Die Halbwertsdauer dieser Quellen liegt bei einigen Jahren. Bei Anwendung dieser Quellen sind besondere *Sicherheitsmaßnahmen* zu beachten.

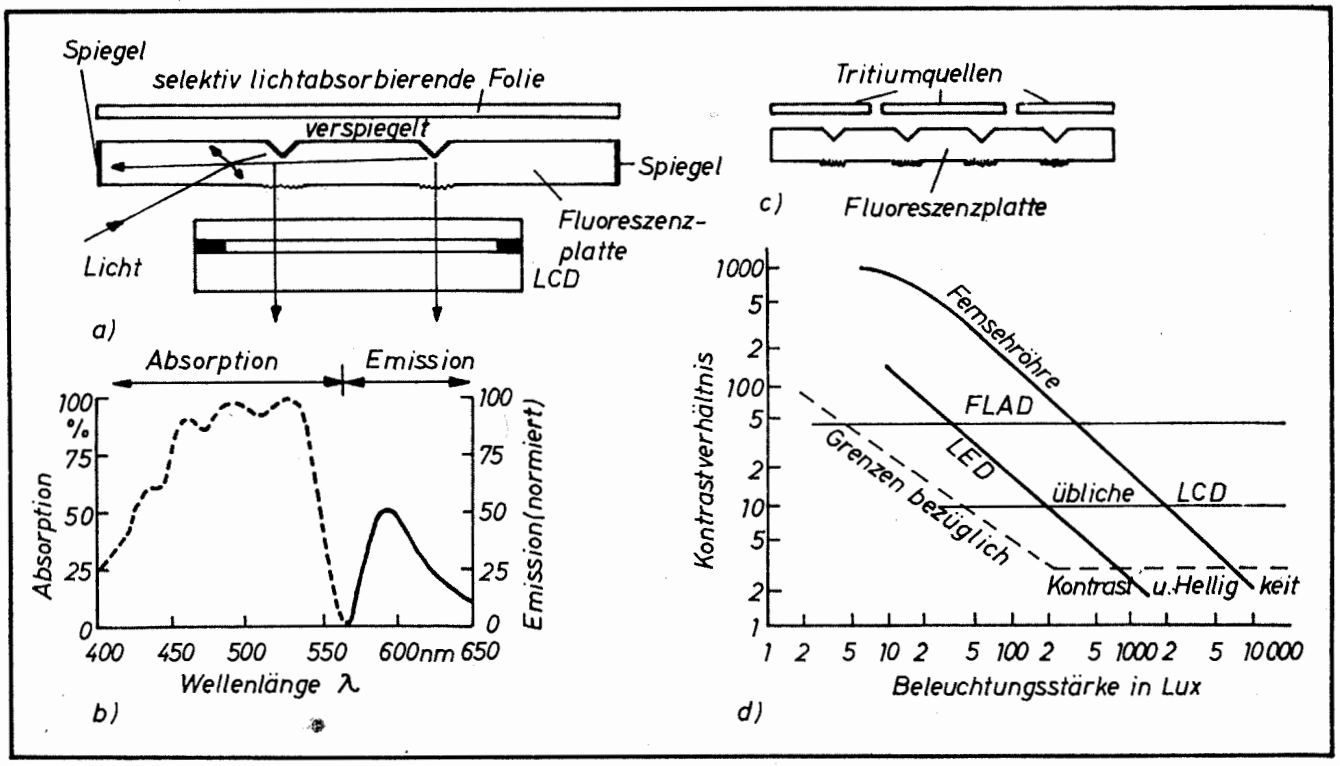


Bild 6.1.39. Fluoreszenzaktiviertes Display

a) Prinzipieller Aufbau
b) Absorption und Emission günstiger Farbstoffmoleküle, die in der Fluoreszenzplatte eingelagert sind
c) Anregung der Fluoreszenzplatte bei Dunkelheit durch Tritiumquellen
d) Vergleich verschiedener Anzeigen bezüglich Beleuchtungsstärke und Kontrastverhältnis. Als Anhalt für das Kontrastverhältnis können die folgenden Werte dienen: 2:1 klein, 3:1 Bleistift Nr. 2 auf Zeichenpapier, 5:1 bequem lesbar, 10:1 sehr gut lesbar, 14:1 Zeitungsschrift

6.1.7. Weitere Anzeigetechniken

Neben den bisher behandelten Anzeigetechniken LED und LCD gibt es eine Vielzahl anderer, die jedoch nicht diese Bedeutung erlangt haben. Im folgenden werden davon behandelt:

- Lumineszenzzellen
- elektrochemische Anzeigen
- ferroelektrische Anzeigen
- magnetische Anzeigen.

Hier nicht behandelt sind dann vor allem

- *Vakuumelektronik* (6.2.), also unter anderem
  - Fernsehröhren: schwarzweiß und Farbe
  - Fluoreszenzanzeigen (Digitron)
  - Plasmadisplays
  - Glimmlichtanzeigen
- *Mechanik und Elektromechanik* (6.3.).

Generell nicht behandelt wird das *Eidophorverfahren*. Es verwendet Prinzipien der Schlierenoptik, wobei in der Regel eine Anregung durch Elektronenstrahlröhren erfolgt. Das Verfahren hat Bedeutung bei der Großproduktion von Fernsehbildern.

6.1.7.1. Lumineszenzzellen

Lumineszenz existiert auch bei den LED im Abschnitt 6.1.3. Die dortige Lumineszenz wird zur Abgrenzung zuweilen auch Injektionslumineszenz genannt. Dann heißt

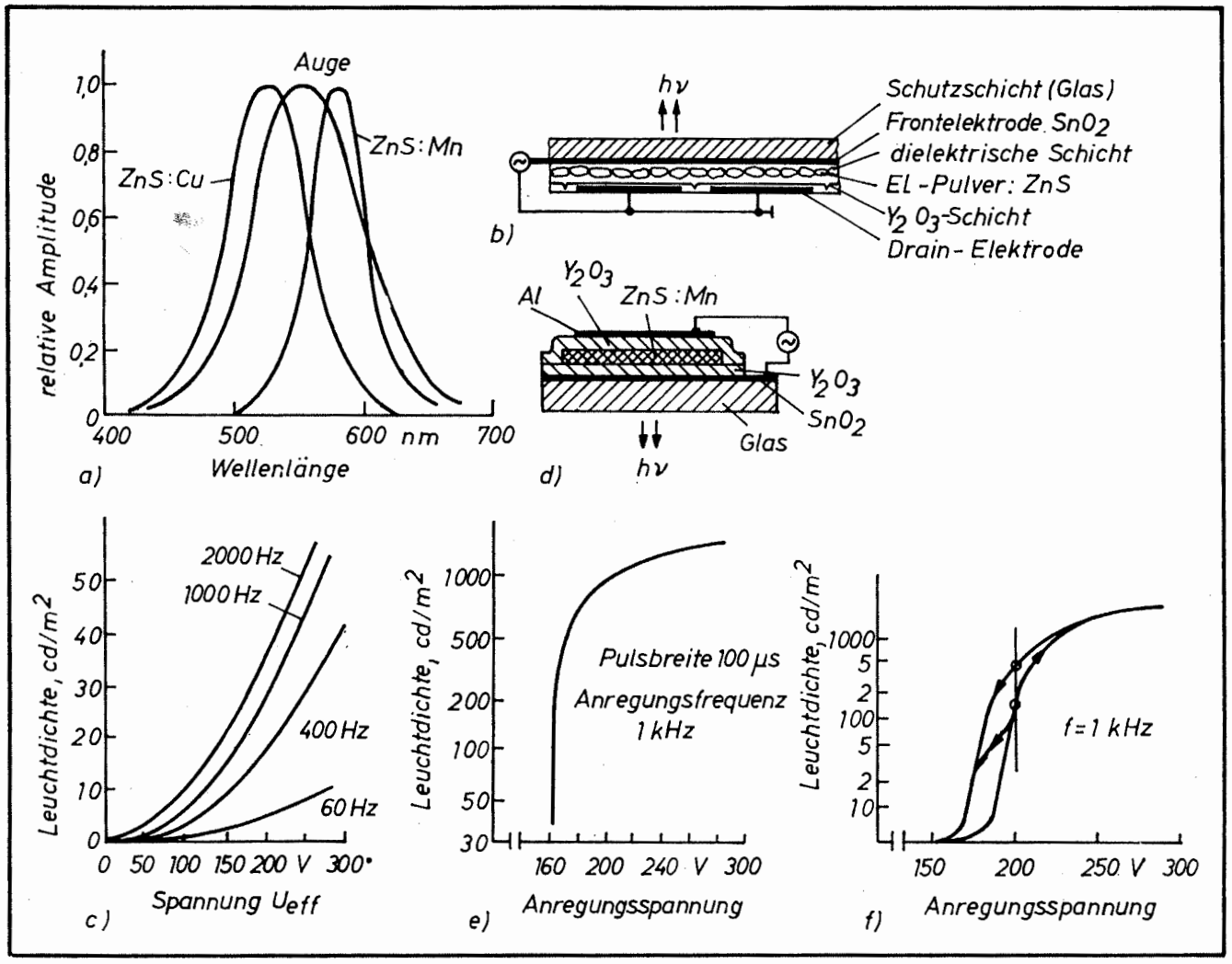


Bild 6.1.40. Lumineszenzzellen (Leuchtkondensatoren)

a) Emissionskurven von mit Cu bzw. Mn dotiertem ZnS im Vergleich zur Augenempfindlichkeit
b) Aufbau einer wechselstrombetriebenen Pulver-Lumineszenzzelle
c) Leuchtdichte einer Zelle gemäß b) bei verschiedenen Betriebsbedingungen
d) Aufbau der wechselstrombetriebenen Dünnschichtzelle von Sharp
e) Leuchtdichte der Zelle gemäß d)
f) Leuchtdichte einer Dünnschichtzelle mit Hystereseeigenschaften

die hier zu behandelnde *Feldlumineszenz*. Weitere Namen für die Feldlumineszenzzellen sind *Leuchtkondensator, Elektrolumineszenzzelle* (EL) und LEF (= light emitting film). Sie verwenden, von Ausnahmen abgesehen, mit Cu bzw. Mn dotiertes ZnS. Das Emissionsspektrum dieser Leuchtstoffe im Vergleich zur Augenempfindlichkeit zeigt Bild 6.1.40a). Weitere Substanzen sind z.B. CaS und SrS. Auf Leuchtstoffe wird generell im Abschnitt 6.2.1.6. eingegangen.

Der Farbstoff wird in zwei Arten verwendet, als

- *Pulver* mit 5 bis 50 µm Teilchendurchmesser, meist in ein Bindemittel mit hoher Dielektrizitätskonstante, z.B. Cyanostärke, eingebettet;
- *dünne Schicht*, u. a. durch Aufdampfen hergestellt.

Betrieben werden die Zellen mit Wechsel- bzw. Gleichstrom, so daß es insgesamt vier Typen von Lumineszenzzellen gibt, wobei die Gleichstrom-Schichtzelle bisher keine Bedeutung erlangte.

Die *wechselstrombetriebenen Pulver-Lumineszenzzellen* sind am bedeutsamsten. Ihren Aufbau zeigt Bild 6.1.40b). Sie wird meist mit ca. 100 V (effektiv) von 50 bis 1000 Hz betrieben. Dabei fließt infolge der hohen Feldstärke von 10^6 bis 10^7 V/cm in dem

Kondensator ein großer Verschiebestrom. Freie Elektronen werden dadurch so stark beschleunigt, daß sie das ZnS-Pulver zum Leuchten anregen. Die Abhängigkeit der Leuchtdichte von Spannung und Frequenz zeigt Bild 6.1.40c). Die Kennlinien sind nichtlinear, da sie aber keine ausgeprägte Schwelle besitzen, sind nur Multiplexfaktoren von 10 bis 100 möglich. Dies ermöglicht einen Matrixaufbau mit gekreuzten Leiterzügen an der Ober- und Unterseite der Zelle bis etwa 100 × 100 Bildpunkte.

Es war und ist z. T. auch heute noch schwierig, Lumineszenzzellen mit großer Lebensdauer herzustellen. Einmal müssen sie gut hermetisch abgeschlossen werden, da eindringende Feuchtigkeit und Sauerstoff die Leuchtschichten schnell altern lassen. Weiteren Einfluß haben Diffusion der Dotierungsstoffe und Gitterfehlstellen.

Wesentlich stabilere Eigenschaften weisen die *wechselstrombetriebenen Dünnschicht-Lumineszenzzellen* auf. Sie werden durch Aufdampfen von Elektroden und dielektrischen Schichten (z.B. Y_2O_3, Al_2O_3, Si_3N_4) sowie der dazwischen befindlichen Mn-dotierten ZnS-Schicht hergestellt (Bild 6.1.40d)). Die Dicke aller Schichten beträgt etwa 2 μm. Bei den seit 1974 vorhandenen, gelb-orange leuchtenden ($\lambda \approx 580$ nm) Zellen von Sharp werden über 20000 Betriebsstunden erreicht. Sie besitzen, wie Bild 6.1.40e) zeigt, eine stark ausgeprägte Schwelle und sind daher hervorragend für Multiplexbetrieb geeignet. Unter speziellen Bedingungen lassen sich bei der Herstellung auch Hystereseeigenschaften (Bild 6.1.40f)) realisieren. Sie werden mit einer konstanten Impulsspannung (im Bild 200 V) betrieben. Es sind dann zwei unterschiedlich helle Zustände möglich, zwischen denen durch einen einmaligen Impuls umgeschaltet werden kann. Dies führt zu erheblichen Vereinfachungen, da der Bildwiederholspeicher entfallen kann.

Gleichstrombetriebene Lumineszenzzellen existieren nur auf Pulverbasis. Sie sind ähnlich wie Bild 6.1.40b) aufgebaut. Die Mn-dotierten ZnS-Kristallite werden jedoch mit einer dünnen Kupferschicht bedeckt, und es fehlt der organische Binder, d. h., die Kristallite haften direkt aneinander. Die so aufgebauten Zellen müssen vor dem Betrieb formiert werden. Dabei wird die leitende Schicht zwischen den Kristalliten zerstört, und es bilden sich unregelmäßig angeordnete *pn*-Übergänge mittels Cu_2S (*p*-leitend) und ZnS (*n*-leitend) aus, ähnlich, wie es schon bei den VDR-Widerständen, Abschnitt 5.1.5., gezeigt wurde. Diese Zellen verfügen z. Z. über eine Lebenserwartung von ca. 5000 Stunden und besitzen eine ausgeprägte Schwellwertcharakteristik.

6.1.7.2. Elektrochemische Anzeigen

Die *Vorteile* der elektrochemischen Anzeigen bestehen in ihrem *Speichereffekt*, dem meist hohen und dazu oft *farbigen Kontrast*, ihrer möglichen *Großflächigkeit* und z. T. in der einfachen Herstellbarkeit. Ihre Nachteile sind die bisher *geringe Lebensdauer* und die große *Schaltzeit* von meist mehreren zehntel Sekunden. Es gibt eine Vielzahl von Möglichkeiten in der *Arbeitsweise*:

- reversible Abscheidung dünner Schichten
- Erzeugung und Löschung von Farbzentren
- Elektrophorese
- chemische Reaktionen
- chemisch erzeugte Lichtstrahlung (Chemolumineszenz).

Diese Fülle kann hier aufgrund der bisher erreichten geringen Anwendung nicht behandelt werden. Es sollen nur kurz einige *Beispiele* beschrieben werden.

Auf der Basis eines noch nicht genau bekannten Mechanismus läßt sich WO_3 mit H^+-Ionen *blauviolett* verfärben. Dabei läuft etwa folgender Vorgang ab:

$$n\mathrm{H}^+ + \mathrm{WO}_3 + e^- \rightleftarrows \mathrm{H}_n\,\mathrm{WO}_3\,. \tag{38}$$

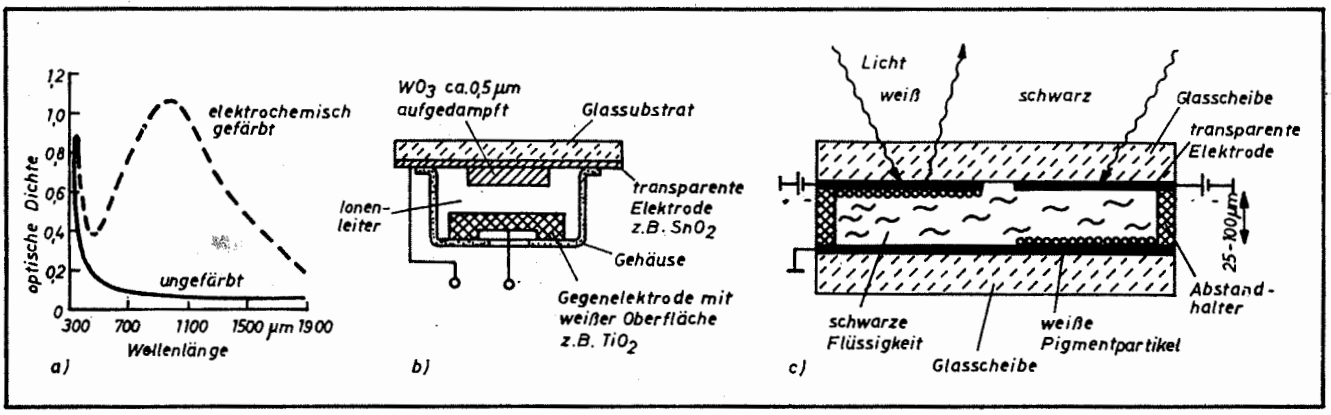


Bild 6.1.41. Elektrochemische Anzeigen

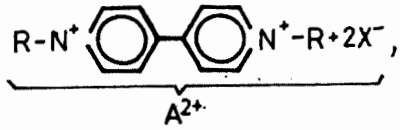
a) Absorptionskurven von WO_3

b) Prinzipaufbau einer elektrochromen Anzeige

c) Prinzipaufbau und Funktion einer elektrophoretischen Anzeige

Die Absorptionskurven beider Wolframsubstanzen zeigt Bild 6.1.41a), den Aufbau einer Zelle nach diesem Prinzip Teilbild b). Die aufzubringende elektrische Leistung je Schaltzyklus beträgt etwa 10 mW/cm². Die Betriebsspannung muß größer als 0,5 V (ca. 500 V/cm) sein. Ursprünglich wurde bei diesem Zellentyp als Elektrolyt z.B. eine Mischung aus schwefliger Säure und Alkohol bzw. Schwefelsäure und Glycerin verwendet. In der letzten Zeit erfolgte ein Übergang zu festen Elektrolyten, z.B. Natrium-Aluminium-Keramik. Statt des WO_3 wurde z. T. auch Iridiumoxid verwendet, wobei dann Grautöne bis schwarz möglich sind.

Eine zweite elektrochrome Anzeige beruht auf den *Viologenen*.



R Alkylgruppe, z.B. C_7H_{15},
X^- Anion, z.B. Br^-, (39)

Das A^{2+}-Ion ist ungefärbt, das reduzierte A^+ dagegen tiefblau. Da beide Ionen relativ große Lebensdauer besitzen und mittels Stromzufuhr reversibel ineinander überführbar sind, läßt sich auf dieser Basis eine Anzeige aufbauen. Dabei scheidet sich das A^+ als Farbfilm an der Katode ab.

Silber-Natrium-Jodid befindet sich in einem organischen Lösungsmittel, z.B. Methanol oder Azetonnitrid. Durch Elektrolyse wird eine dünne (ca. 5 nm) Silberschicht abgeschieden.

Den Aufbau eines *elektrophoretischen Displays* zeigt Bild 6.1.41c). In einer schwarzen Flüssigkeit sind kleine weiße oder farbige Pigmentpartikeln von unter 1 µm Durchmesser enthalten, die elektrisch geladen sind. Besitzen sie eine positive Ladung wie im Bild, so werden sie an der jeweils negativen Elektrode angelagert. Die Oberfläche der Zelle erscheint dann entweder schwarz oder in der Farbe der Partikeln. Es werden etwa 20 bis 50 µW/cm² s während der Schaltzeit benötigt.

6.1.7.3. Ferroelektrische Keramik

1967 gelang es G. HAERTLING, *Blei-Lanthan-Zirkonat-Titanat* als transparente Keramik (PLZT) herzustellen. Diese Keramik hat mehrere wichtige optische Eigenschaften, die durch ein elektrisches Feld steuerbar sind. Sie betreffen vor allem

Doppelbrechung und Lichtstreuung.

Weiter existiert eine ferroelektrische Hysterese, so daß auch ein

Speichereffekt

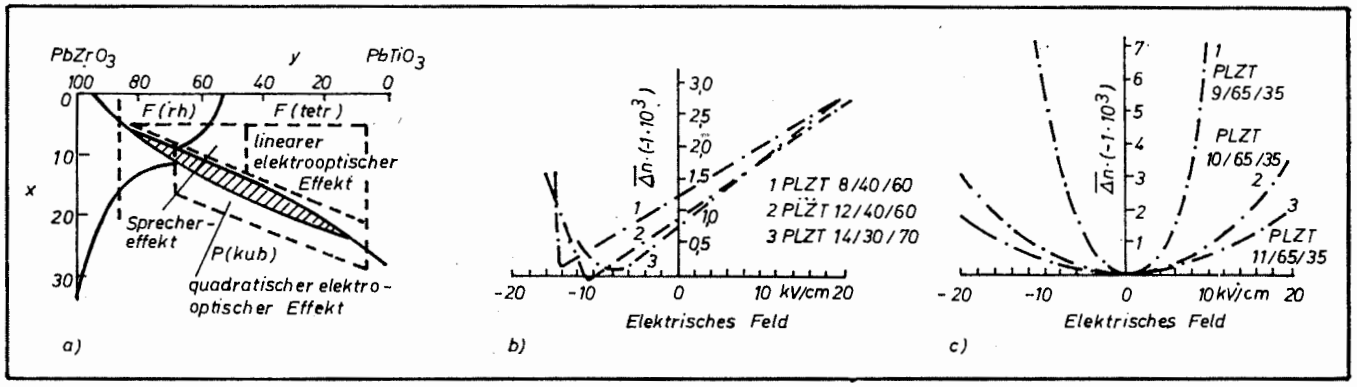


Bild 6.1.42. Eigenschaften von PLZT-Keramik

a) Eigenschaftsgebiete von $(PbZr_yTi_{100-y}O_3)_{1-x}La_x$. Die drei Kristallstrukturen rhombisch, tetraedrisch und kubisch sind ebenfalls eingetragen

b) Beispiele für lineare und

c) für quadratische Doppelbrechung

Die Zahlen geben der Reihe nach die Atomprozente von Lanthan, Zirkon, und Titan an

auftritt. Die entsprechenden Gebiete zeigt Bild 6.1.42a) und Beispiele für den linearen und quadratischen elektrooptischen Effekt Teilbild b). Es ist naheliegend, diese Effekte für Anzeigen zu verwenden. Zur Zeit bestehen noch verschiedene Probleme, insbesondere wegen der relativ hohen Feldstärken. Doch da es bereits gelang, die PLZT durch Sputtern in sehr dünnen Schichten herzustellen, dürfte hier die Möglichkeit bestehen, zu brauchbaren Spannungen zu kommen. Die Hauptanwendung liegt noch bei optischen Modulatoren und steuerbaren Lichtverschlüssen. Hier kommt die für passive Anzeigen kurze Schaltzeit von 3 µs zum Tragen.

6.1.7.4. Magnetische Anzeigen

Eine *magnetische Flüssigkeit* enthält in einem Lösungsmittel feinstverteilte magnetische Partikeln. Die Partikeln lassen sich unter Einfluß eines magnetischen Gradientenfeldes dichter und lockerer anordnen. Auf dieser Basis sind im Prinzip sehr einfache Anzeigen möglich, insbesondere wenn die Hysterese des Materials im magnetischen Steuerkreis als Speichereffekt ausgenutzt wird. Größere praktische Bedeutung hat dieses Prinzip noch nicht erlangt.

Weiter wäre es möglich, daß Materialien gefunden werden, die einen großen *magnetooptischen Effekt* zeigen. Dann wären auf dieser Basis ebenfalls Anzeigen möglich.

6.1.8. Vergleich von Anzeigen

Mit der rasanten Entwicklung der Mikroelektronik wurde es immer dringender, die Daten und Programme auf einfache und effektive Weise dem Menschen optisch anzubieten. So entstanden insbesondere in der letzten Zeit eine Vielzahl von Anzeigen und Displays. Diese Entwicklung ist z. Z. noch nicht abgeschlossen, deshalb ist ein Vergleich unsicher. Wenn vom Fernsehschirm abgesehen wird, *dominieren heute* drei Anzeigen:

LED, LCD und Plasmaanzeigen.

Ihren jeweiligen Marktanteil für 1976 und 1981 zeigt Bild 6.1.43. Es ist auffällig, wie dabei der relative Anteil des LCD-Marktes auf Kosten von LED und Plasmaanzeigen zugenommen hat. Noch mehr sind die LCD bei Uhren im Vorteil. Hier bestand bereits 1976 ein Verhältnis bez. LED und LCD von 70:30, welches 1977 schon 50:50 betrug und 1980 auf 35:65 anstieg. Mit Bildröhren zusammen betragen für alle Anzeigen die Verhältnisse:

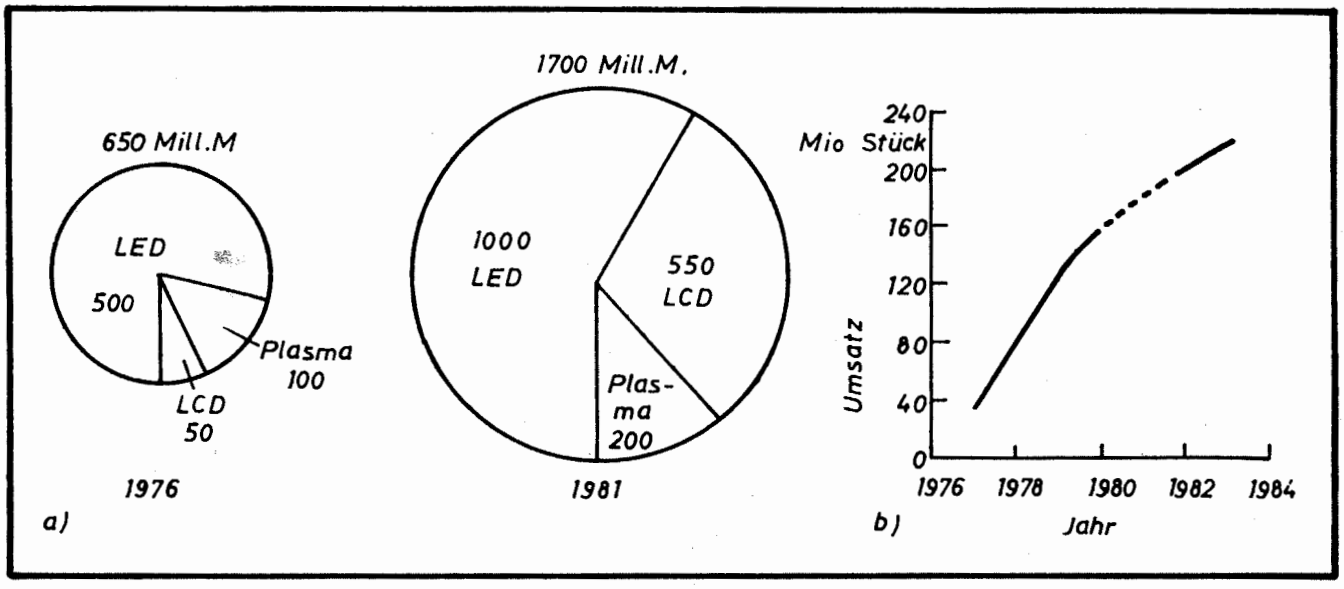


Bild 6.1.43. a) Weltumsätze und relative Verhältnisse bei Anzeigen, jeweils ohne Bildröhren. Andere, nichtaufgeführte Techniken hatten so wenig Umsatz, daß sie nicht darstellbar waren
b) Stückzahlentwicklung bei LCD-Anzeigen

Bildröhren	= 78,5%
LED	= 9,3%
LCD	= 7,8%
Plasmapanal	= 2,4%
Vakuumfluoreszenz	= 1,6%
Sonstige	= 0,4%.

Zum Anwendungsvergleich von optischen Anzeigen können eine Fülle von Eigenschaften herangezogen werden:

- gute Lesbarkeit bei verschiedener Umgebungshelligkeit
- großer Sichtwinkelbereich
- Multiplexfaktor (Möglichkeit Bildschirm)
- kostengünstig
- Kompaktheit (Volumen)
- kleines Gewicht
- Speicherfähigkeit
- lange Lebensdauer
- großer Betriebstemperaturbereich
- geringer Leistungsbedarf
- farbtüchtig
- kompatibel mit MOS-Schaltkreisen
- kurze Schaltzeit.

Ferner lassen sich Anzeigen einteilen in
aktive und passive
je nachdem, ob sie selbst Licht erzeugen oder nur Licht beeinflussen. Die passiven Anzeigen sind bezüglich des geringen Energieverbrauchs im Vorteil, bezüglich der kurzen Schaltzeit aber im Nachteil.

Einen Überblick zu groben Daten der wichtigsten Anzeigen und Displays gibt Tab. 6.1.10.

6.1.9. Lichtleitfasern

Lichtleitfasern werden in erster Linie für die *Nachrichtenübertragung* — also ein Gebiet, das in diesem Buch nicht behandelt wird — verwendet. Sie haben daneben aber auch für andere Gebiete Bedeutung erlangt und werden hier künftig immer mehr eingesetzt werden. Diese Gebiete sind u. a.

Tabelle 6.1.10. Vergleich einiger Anzeigen und Displays, bezogen auf etwa einen cm^2
Bei Farbe bedeutet: b — blau; g — grün, gelb; o — orange; r — rot; w — weiß. Alle Daten sind Richtwerte.

Typ/Eigenschaft	Bildröhre	Plasmaanzeige	Digitron	Glühlampe	LED	LEF	LCD
Strom, mA	ca. 10^{-2}	0,3 ... 6	0,5 ... 6	20 ... 200	5 ... 50	0,02 ... 5	10^{-3}
Spannung, V	10^3 ... 10^4	100 ... 300	15 ... 25	2,5 ... 25	1 ... 5	20 ... 250	1 ... 20
Leistung, mW/cm^2	10	70	50	—	500	300	10^{-2} ... 10^{-3}
Leuchtdichte, cd/m^2	500	3 ... 70	300	—	200	10 ... 300	passiv
Ausbeute, lm/W	5 ... 7	0,07 ... 0,6	3 ... 5	10	2000 ?	0,1 ... 1,5	passiv
Schaltzeit, s	10^{-7}	10^{-5}	10^{-6}	10^{-3}	10^{-7}	10^{-5}	10^{-1}
Kontrastverhältnis	50	30	50	50	50	30	20
Graustufen	150	60	50	50	50	15	6 ... 8
Farben	alle	g, o, r, b	grün	weiß	r, g	g, o	weiß
Multiplexverhältnis	10^6	10^3	10^4	10^2	10^3	200	100
Display	+	+/—	—	—	+/—	+/—	+/—
Lebensdauer	10^4	10^5	10^4	$5 \cdot 10^3$	10^6	10^4	10^5
Temperatur, °C		0 ÷ 70		—50 ÷ 200	—55 ÷ 125	—30 ÷ 70	0 ÷ 70

- *Informationsübertragung* in Geräten und Einrichtungen, wo elektrische Leiter nicht sinnvoll, effektiv zuverlässig oder zulässig sind;
- *Lichttransport* an Stellen, wo eine Lichtquelle aus verschiedenen Gründen, z. B. Volumen, Wärmeabstrahlung oder Auswechselbarkeit nicht möglich ist;
- *Übertragung von Leistungen*, z. B. in der Chirurgie oder zur Materialbearbeitung.

Um diese Beispiele zu vertiefen, seien hier *Vorteile von Glasfasern* bei einigen Anwendungen aufgezählt, die nicht zum Nachrichtenweitverkehr (dem Schwerpunkt der Glasfasern) gehören:

- Gewichts- und raumsparender Aufbau,
  z. B. bei der Verkabelung in Flugzeugen;
- Hohe Übertragungsbandbreite,
  z. B. bei extrem breitbandigen Signalen in Computern;
- Potentialtrennung von Sender und Empfänger,
  z. B. in Kraftwerken;
- Unempfindlichkeit gegenüber elektromagnetischen und optischen Störfeldern,
  z. B. in Flugzeugen, Schwermaschinen und Computern.

Bei den Lichtleitfasern müssen unterschieden werden

- Materialien für die Lichtleitung
- Aufbau der Fasern.

Die wichtigsten Daten sind:

- Dämpfung
- Dispersion
- Numerische Apertur.

6.1.9.1. Dämpfung

Für die *Fortpflanzung einer Lichtwelle* in einem Medium gilt Gl. (14), die hier in folgender Form geschrieben sei:

$$V = \exp(\alpha x) . \tag{40}$$

Dabei bedeuten V das Leistungsverhältnis bezüglich einer Strecke der Länge x und α die *Absorptionskonstante*, die gewöhnlich in cm^{-1} angegeben wird. Für die *praktische Anwendung* bei Lichtleitern hat sich ein *Wert* D in dB je km durchgesetzt (bez. dB = Dezibel = dezi-Bell = 10 log V s. 7.7.1.4.). Daher gilt

$$D/(\mathrm{dB/km}) = 10^{-6} \cdot \alpha \cdot \mathrm{cm/log\,e} \approx 2{,}3026 \cdot 10^{-6}\, \alpha/\mathrm{cm^{-1}}. \tag{41}$$

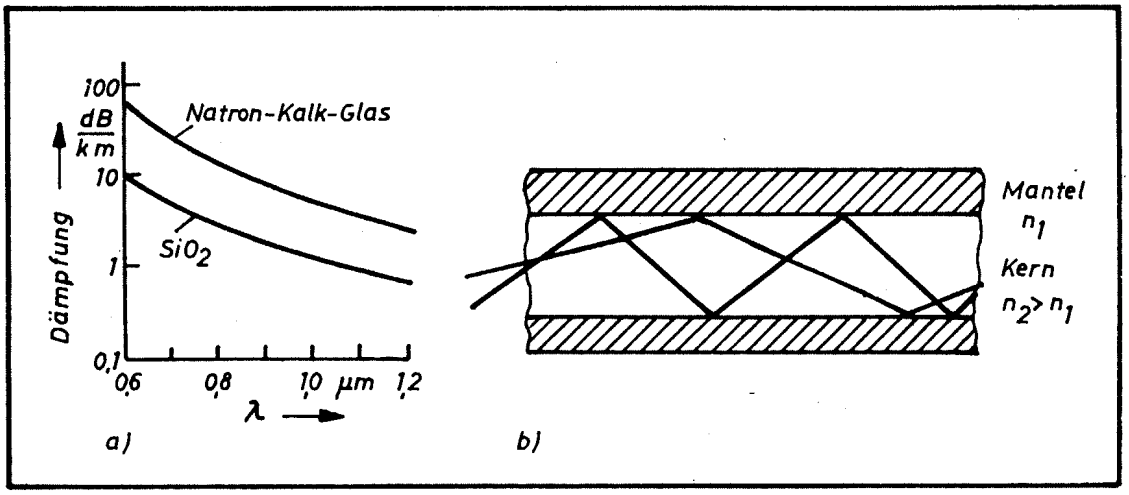


Bild 6.1.44. Zu Lichtleitfasern

a) Grundverluste kompakter Gläser durch Absorption und Streuung

b) Grundprinzip der Lichtleitfaser

Tabelle 6.1.11. Dämpfungswerte einiger kompakter Materialien für Lichtleitfasern bei verschiedenen Wellenlängen

Material	SiO_2			AgBr		AgCl	KBr	KCl	TlBr
Wellenlänge in μm	0,85	1,3	1,55	5	10,6	10,6	10,6	10,6	10,6
Dämpfung in dB/km	3	0,5	0,2	0,001	20	5	0,01	0,08	1

Als *Material* für die gebräuchlichsten Lichtleiter hat sich reines Glas (SiO_2) und Natrium-Kalk-Glas durchgesetzt. Die Dämpfung der kompakten Materialien zeigt Bild 6.1.44a).

Für die Anwendung über *kurze Entfernungen* haben auch *Kunststoffasern* und im Infrarotbereich, z.B. für Leistungsübertragung, Fasern aus Chloriden und Bromiden Bedeutung. Tab. 6.1.11 zeigt einige Werte kompakter Materialien.

6.1.9.2. Typen von Fasern

Das Grundprinzip einer Lichtleitfaser geht aus Bild 6.1.44b) hervor. Ein *Kern* mit dem Brechungsindex n_2 ist von einem *Mantelmaterial* mit dem Brechungsindex n_1 umgeben, wobei gilt $n_2 > n_1$. Lichtwellen, die unter dem Grenzwinkel auf die Wände des Kerns treffen, werden dann total reflektiert und damit im Kern fortgeleitet. Gleichzeitig ist auf dem Bild erkennbar, daß es Lichtstrahlen im Kern gibt, die bei gleicher Länge der Lichtleitfaser unterschiedlich lange Wege zurücklegen. Hierdurch erscheint ein Lichtimpuls am Ausgang um die Wegdifferenz der Lichtstrahlen verlängert. Diesen Effekt nennt man *Modendispersion*. Für sie gilt etwa mit der Länge L der Faser und der Lichtgeschwindigkeit c

$$\Delta T \approx \frac{L}{c} \cdot \frac{n_2^2 - n_1^2}{2n_2}. \tag{42}$$

Weiter gibt es eine *Materialdispersion*. Sie berücksichtigt die Änderung des Brechungsindexes bei verschiedenen Wellenlängen und wird daher wirksam über die Bandbreite der eingestrahlten Lichtwelle (Bild 6.1.45d)). Diese Dispersion wird in ns/km (Impulsverlängerung) angegeben.

Bedeutsam ist weiterhin der Winkel 2Θ, unter dem Licht in die Faser eintreten kann. Er wird durch die *numerische Apertur* N_a beschrieben,

$$N_a = \sqrt{n_2^2 - n_1^2} = \sin(\Theta/2). \tag{43}$$

Je nach technologischen Möglichkeiten und Anwendungszwecken werden heute *vier Faserarten* unterschieden, wie sie Bild 6.1.45a) mit einigen Daten zeigt. Im Teilbild b) ist gezeigt, wie weit es möglich ist, mit einer Monomodenfaser an die theoretischen Werte heranzukommen. Die *Absorptionsmaxima* stammen von schwer zu beseitigenden OH^--Anteilen. Das Gegenbeispiel zeigt Bild c) für ein ganz einfaches und billiges Lichtleitkabel auf der Basis von Kunststoffen. Die Dämpfung ist hier rund tausendmal so groß. Beispiele für 1980 in Japan produzierte Lichtleitfasern hoher Qualität faßt Tab. 6.1.12 zusammen.

Für die Möglichkeiten der *Leistungsübertragung* sei folgendes Beispiel genannt. Mit einer Faser aus einem polykristallinen Material von KBr, KCl, AgCl und TlBr können bei 11 μm Wellenlänge und bei 0,5 mm Dicke ohne Probleme 2 W übertragen werden.

Tab. 6.1.13 gibt Hinweise auf die *ökonomische Gestaltung* von Lichtübertragungen bezüglich der Sender, Fasern und Empfänger. Bild 6.1.45e) zeigt geschichtliche Daten und Teilbild f) Leistungsdaten zur Übertragung von Licht.

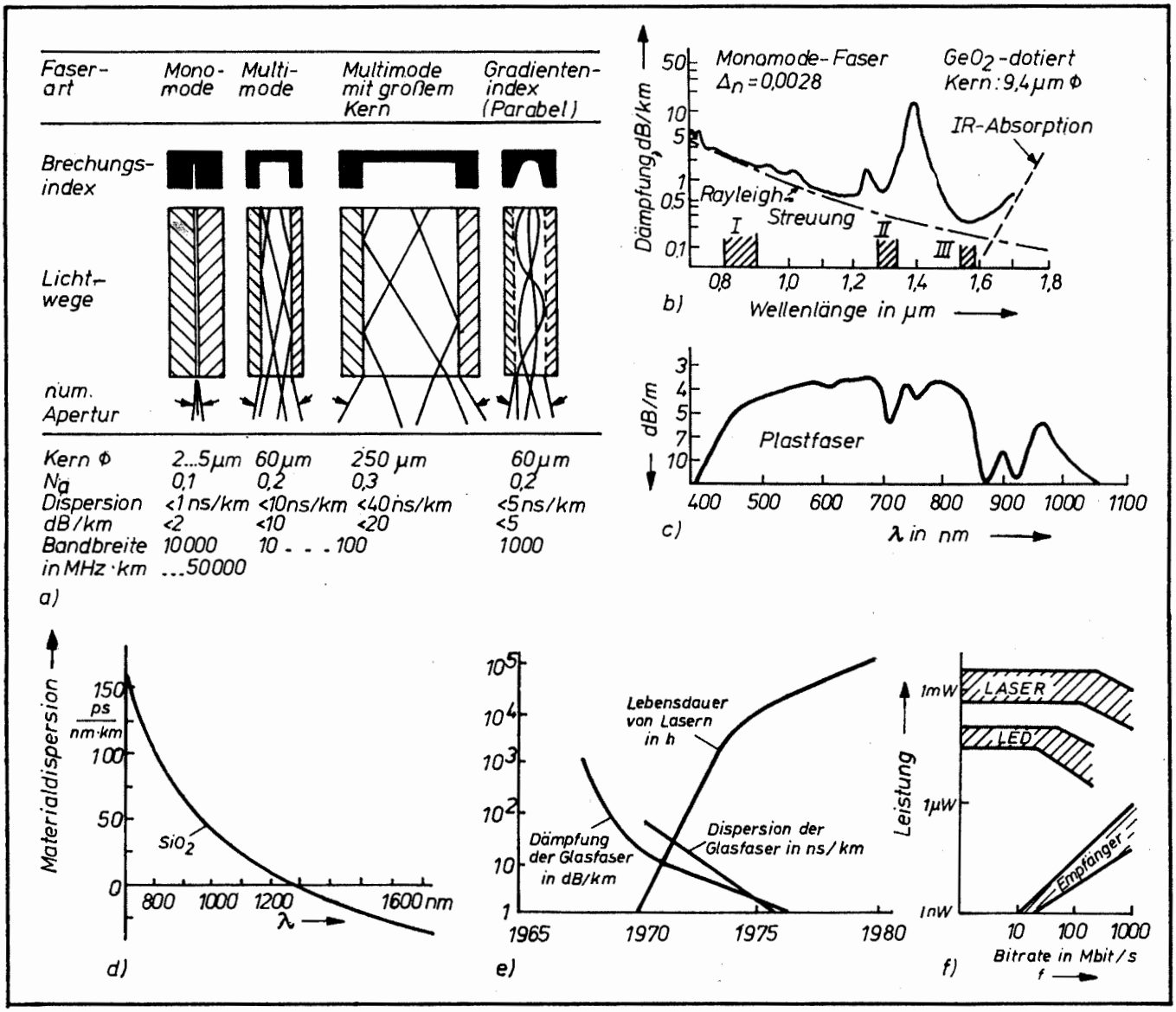


Bild 6.1.45. Eigenschaften verschiedener Glasfasern

a) Die Multimodenfaser mit großem Kern wird vor allem für Leistungsübertragung verwendet. Bei den Multimodenfasern ist zuweilen der Mantel auch aus Kunststoff. Schließlich gibt es für sehr einfache Anwendungsfälle, für kurze Lichtwege, auch Multimodenfasern nur aus Kunststoff

b) Dämpfung einer Monomodenfaser und Vergleich mit theoretischen Werten. Die hohen Absorptionsmaxima der realen Faser werden vor allem durch schwer zu beseitigende OH^--Gruppen bewirkt. Weitere störende Stoffgruppen enthält Bild 6.1.27c). Die römischen Zahlen I bis III beziehen sich auf übliche Übertragungskanäle. Der Kanal I ist durch GaAs-Leuchtdioden bzw. -Laser und Si-Empfänger bestimmt. Der Kanal III ist durch das absolute Dämpfungsminimum gegeben

c) Dämpfungswerte eines sehr einfachen Lichtkabels aus nur Plastmaterialien

d) Materialdispersion von Quarzglasfasern. Der Nulldurchgang bestimmt den Übertragungskanal II in Bild b)

e) Geschichtliche Entwicklung der besten Daten von Lichtleitfasern und der Lebensdauer von Lasern

f) Leistungswerte von Lichtsendern und -empfängern als Funktion der Bitrate

6.2. Vakuumelektronik

Die Vakuumelektronik basiert heute auf *zwei Hauptgruppen* von Bauelementen

- *Elektronenröhren* mit dem Untergebiet *Elektronenstrahlröhren,*
- *Gasentladungsröhren* mit dem Untergebiet *Plasmaanzeigen* bzw. -displays.

Im ersten Fall wird die Steuerbarkeit von im Vakuum frei beweglichen Elektronen durch elektrische bzw. magnetische Felder ausgenutzt. Im zweiten Fall existiert ein Plasma, welches ein Gas ist, das aus Elektronen, Ionen und Molekülen besteht. Die Bewegung der Teilchen dieses Gases und ihre Wechselwirkung wird durch äußere Felder beeinflußt.

Tabelle 6.1.12. Typische Werte der optischen Dämpfung und 6-dB-Übertragungsbandbreite von in Japan entwickelten Glasfasern

Mode Faser Material	Multistufenprofil			Gradientenprofil		Monostufenprofil
	Quarz	Quarz-Plastik	Mehrkomp.	Quarz	Mehrkomp.	Quarz
Kerndurchm. (μm)	60	130 ... 200	80	60	45 ... 60	5 ... 8
Manteldurchm. (μm)	125 ... 150	350	125	125 ... 150	150 ... 200	120 ... 150
Dämpfung (dB/km) bei 0,85 μm	2 ... 5	3 ... 6	5 ... 10	3 ... 5	5 ... 10	2 ... 5
6-dB-Bandbreite (MHz bei 1 km)	40 ... 65	18 ... 30	30	400 ... 1200	400 ... 1000	3000

Tabelle 6.1.13. Leistungsdaten und Verfügbarkeit optischer Bauelemente

Bandbreite ↓ Verstärkerfeldlänge ↓

Sender	Fasertyp	Empfänger
Lumineszenzdiode (LED)	Stufenprofil-Faserbündel	Fototransistor
High-radiance-LED	Stufenprofil-Multimodefaser	*pin*-Fotodiode
Laserdiode	Gradientenprofil-Multimodefaser Monomodefaser	Avalanche-Fotodiode

Verfügbarkeit ↑

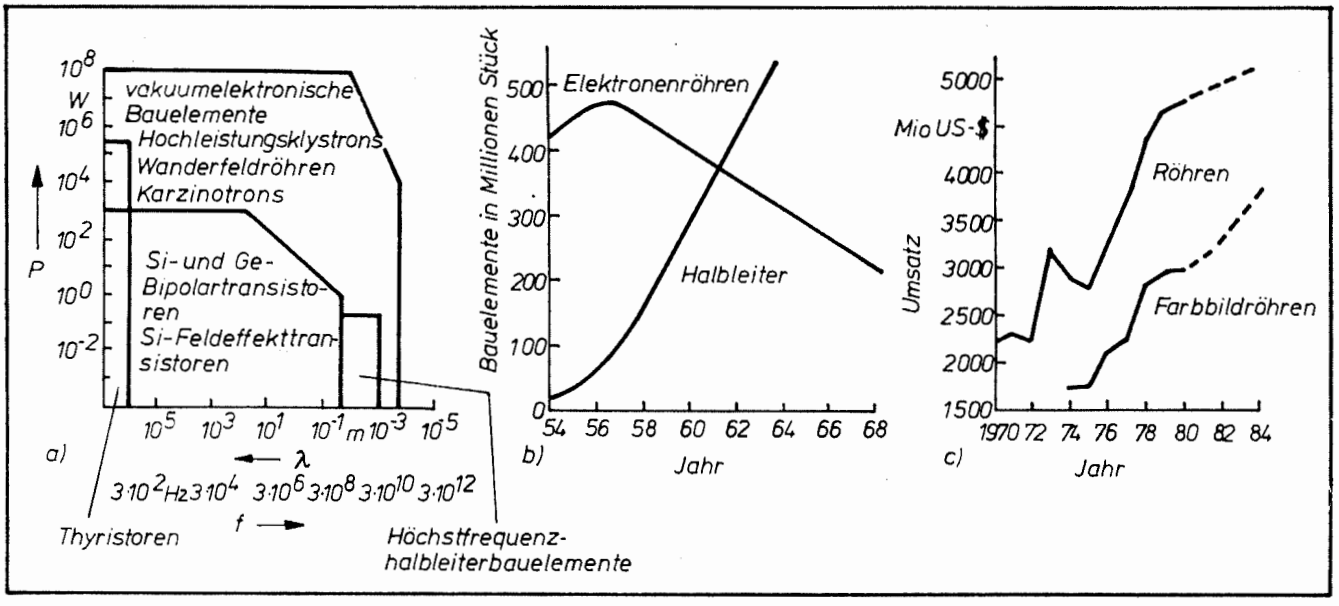


Bild 6.2.1. Zum Wandel der Anwendung von Verstärkerröhren

a) Einsatzgebiete in der Frequenz-Leistungsebene
b) Verlauf der Transistor- und Röhrenproduktion in den USA
c) Verlauf des Umsatzes bei allen Röhren und bei Farbbildröhren in den USA

Im Gegensatz zu den bisher hauptsächlich behandelten Halbleiterbauelementen sind die *Bauelemente* der Vakuumelektronik *in der Regel groß*. Durch besonders große Gestaltung und spezielle Anwendung haben sie sogar *Dimensionen des Maschinenbaus* und damit auch großtechnische Anwendungen erreicht. Als Beispiel seien nur Bearbeitungen mit Elektronen- bzw. Plasmastrahlen

genannt, wie Schmelzen, Schneiden und Schweißen von Materialien. Als *Großgeräte* seien die vielfältigen Typen von Elektronenmikroskopen und die Großgeräte der *Kern- und Plasmaphysik*, wie z. B. Protonenbeschleuniger und Magnetohydrodynamischer Generator (MHD), genannt.

In der hier zu behandelnden Elektronik haben die Bauelemente der Vakuumelektronik infolge der Entwicklung der *Halbleiterelektronik* einen *beachtlichen Wandel* durchgemacht. Noch um 1950 war die Verstärkerröhre das wesentliche aktive Bauelement. Heute ist sie in der klassischen Form nicht mehr vorhanden. Sie existiert, stark abgewandelt (was auch ihre Wirkungsmechanismen betrifft), nur noch bei sehr hohen Frequenzen (über 10 GHz) und Leistungen (über 5 kW) (Bild 6.2.1 a)). Dies sind also die Höchstfrequenz- und Senderöhren. Sie machen z. Z. weltweit einen Produktionswert um 3 Milliarden Mark aus, stehen in diesem Buch jedoch nicht zur Diskussion. Dennoch ist die Produktion der *klassischen Verstärkerröhre* noch nicht auf Null abgesunken. Aber seit etwa 1970 wird sie faktisch nur noch zur *Ersatzbestückung* produziert. 1975 waren es weltweit noch ca. 80 Millionen Stück. Die drastische Übergangsperiode zeigt Bild 6.2.1 b).

Steigende Produktionsziffern besitzen dagegen zwei Anwendungen von Vakuumbauelementen, nämlich die der

- Anzeige- und Bildröhren und
- optischen Informationswandlung.

Diese Typen gehören eigentlich zur *Optoelektronik*. Sie werden jedoch — und das geschieht auch hier — infolge ihrer spezifischen Wirkungsmechanismen getrennt behandelt. Die Bedeutung dieses Gebietes ergibt sich u. a. aus folgenden Zahlen. 1975 wurden weltweit ca. 400 Millionen *Fernsehgeräte* benutzt. Das entspricht einer Bildfläche von einigen Quadratkilometern und einem Leistungsverbrauch von 10^{10} bis 10^{11} Watt. 1982 wurden ca. 40 Millionen *Farbbildröhren* produziert. Beispielhaft ist diese Entwicklung in Bild 6.2.1 c) dargestellt. Die Röhrenproduktion hat sich also insgesamt auf spezielle, hochwertige Typen verlagert.

6.2.1. Grundlagen

In diesem Abschnitt werden die Grundlagen der *Erzeugung freier Elektronen* und ihrer *Bewegung unter Feldeinfluß*, also die Gebiete

Emission, Bündelung, Ablenkung und Abbildung

von bzw. mittels Elektronen im Hochvakuum behandelt. Der Begriff *Hochvakuum* wird hier in dem Sinne verwendet, daß die Wechselwirkungen von Elektronen mit Ionen bzw. Molekülen vernachlässigbar bleiben. Ein ideales Vakuum ist weder prinzipiell noch technisch herstellbar. In dem abgeschlossenen Gefäß verbleibt immmer ein Restgas. Es soll eine so geringe Konzentration besitzen, daß die obige Näherung keine Einschränkung bedeutet. Die Praxis zeigt, daß dann die Stoßzahlen zwischen den Gasmolekülen bzw. -ionen untereinander etwa zehnmal so groß sind wie die mit Elektronen.

Die Gasmenge im Vakuum wird durch den Druck (die Anzahl und Stärke der Stöße) auf die Wände des Gefäßes gemessen. Die Maßeinheit ist das Pascal (Pa). Es entspricht etwa dem 10^{-5}-ten Teil der Atmosphäre (Normalluftdruck). Das Vakuum in den technischen Röhren liegt zwischen 10^{-4} und 10^{-6} Pa (10^{-6} und 10^{-8} Torr). Die freie Weglänge der Moleküle beträgt dabei etwa 0,1 ... 10 km. Sie ist also in jedem Fall groß gegen die Abmessungen der Gefäße. Trotzdem enthält eine Röhre von etwa 3 cm^3 Volumen zwischen 10^9 und 10^{10} Gasatome.

Das technische Vakuum wird in zwei Stufen erzeugt. Mit einfachen Vakuumpumpen ergibt sich ein Gasdruck von etwa 10^{-2} Pa. Die weitere Reduzierung erfolgt mittels eines Getterstoffes. Das Gas wird an ihm physikalisch (Lösung) oder chemisch (Reaktion) gebunden. Heute gibt es eine Vielzahl von Getterstoffen mit unterschiedlichen Eigenschaften. Alle wirken jedoch erst ab einer bestimmten Temperatur. Nachdem also die Röhre durch Auspumpen auf das Vorvakuum gebracht wurde, wird sie abgeschmolzen und dann einmalig auf die Arbeitstemperatur des Getterstoffes erhitzt. Sie liegt zwischen 40 und 1000 °C.

Neben den *Vakuumröhren* gibt es auch Röhren, deren Wirkung gerade auf der Wechselwirkung zwischen den Gasatomen und -ionen beruht. Sie verlangen also gerade einen bestimmten Gasdruck, wobei eine größere Anzahl der Atome bzw. Moleküle ionisiert ist. Dann wird dieses Gas als *Plasma* bezeichnet. Seine Grundlagen werden im Abschnitt 6.2.4.1. behandelt.

6.2.1.1. Emission von Elektronen

Damit ein Elektron ein einzelnes, freies Atom verlassen kann, muß ihm die *Ionisierungsenergie* W zugeführt werden (vgl. Tab. 6.2.5). Mit der Elementarladung $e_0 = 1{,}602\,189 \cdot 10^{-19}$ As läßt sie sich in Elektronenvolt (eV) umrechnen

$$U = W/e_0 . \tag{1}$$

Im Festkörper besteht entsprechend dem Bändermodell eine zusätzliche Wechselwirkung zwischen den Atomen. Daher ist die Energie zum Verlassen des Festkörpers (*Austrittsarbeit*) meist geringer als die Ionisierungsenergie. Sie wird bei den Alkalimetallen um so kleiner, je größer die Gitterkonstante ist. Für Metalle liegt die in Spannung umgerechnete Austrittsarbeit U_0 im Bereich von 0,5 bis 6 eV (Tab. 6.2.1). Diese Energie kann einem Elektron auf unterschiedliche Weise zugeführt werden. Danach unterscheidet man die *verschiedenen Emissionsarten*:

- Feldemission
- Stoßemission (Ionenstoß und Sekundäremission)
- Fotoemission (6.2.1.2.)
- Thermoemission (6.2.1.3.).

Das Material, aus dem die Elektronen emittiert werden, heißt *Katode*.

Für die *Feldemission* sind unmittelbar an der Katodenoberfläche Feldstärken von 10^7 V/cm notwendig. Sie sind nur an sehr kleinen Krümmungsradien erreichbar. Deshalb hat dieser Effekt für die hier zu behandelnden Röhren keine Bedeutung. Direkten Gebrauch hiervon macht das Feldelektronenmikroskop. Dabei wirken sich die geringsten Geometrie-Abweichungen und Verunreinigungen an der Spitze auf die

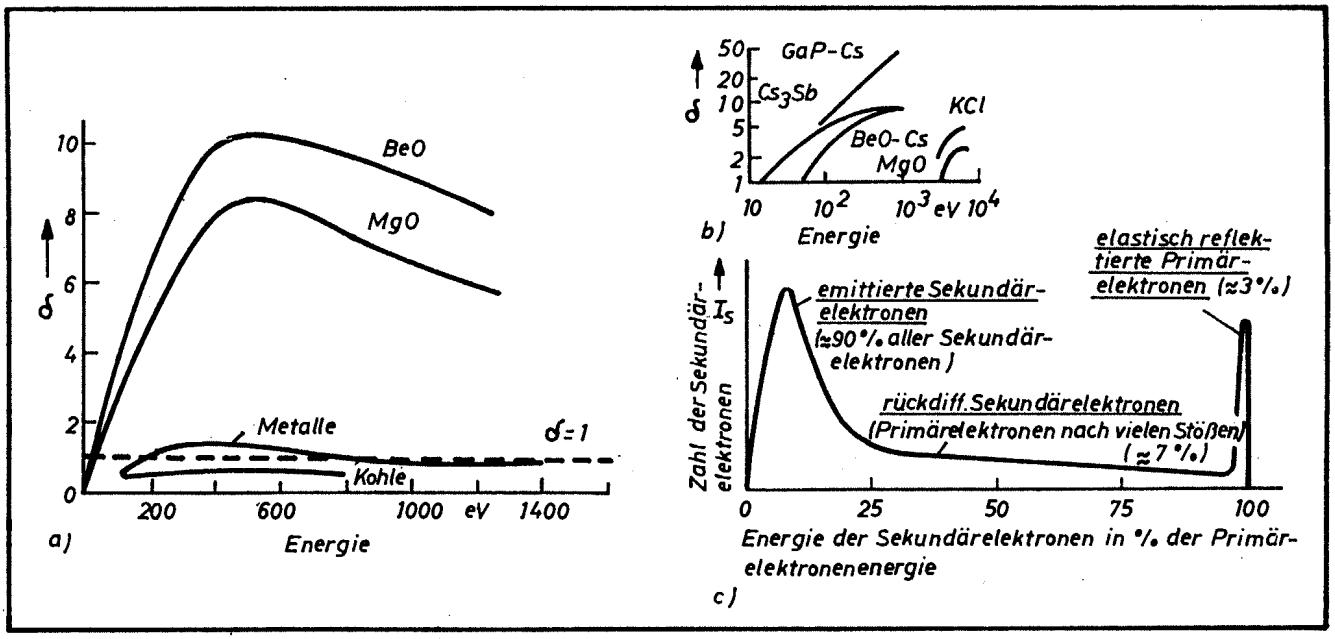


Bild 6.2.2. Zur Sekundäremission

a) Ausbeutefaktor einiger Materialien in linearer und
b) in logarithmischer Darstellung
c) Energiespektrum der Sekundärelektronen

Tabelle 6.2.1. Austrittsarbeiten W, RICHARDSON-Konstanten A, Betriebstemperaturen T_B, Schmelztemperaturen T_S und Elektronenausbeuten $A_e = i_S/P$

Metalle	W/eV	Katodenmaterial	W eV	A A/cm²K²	T_B K	T_S K	A_e mA/W	Heizleistung W/cm²	Emission A/cm²
Rb	2,13	Ba	2,5	60	800	1140			
Sr	2,25	Cs	1,9	162	293	322			
Ka	2,26	Mo	4,2	55	2300	2910			
Na	2,30	Nb	4,0	57		2800			
Li	2,36	Ni	5,0	1380					
Ca	2,96	Pt	6,27	17000					
Cr	3,72	Ta	4,1	40 ... 60	2100	3140	2 ... 18	34 ... 50	0,17 ... 0,65
Mg	3,96	Th	3,4	60	1500	2130			
Al	4,25	W	4,5	60 ... 100	2500	3660	3 ... 10	70 ... 84	0,3 ... 0,7
Zn	4,25	BaO + SrO auf Ni	1	10^{-2} ... 10^{-3}	1100	—			
Cu	4,29	BaO + SrO auf W	1,6	10^{-2} ... 10^{-3}	1400	—	100 ... 10000	2,8 ... 5	0,15 ... 30
Hg	4,53	ThO auf W	1 ... 1,5		1800				
Ge	4,55	W + Th	2,6	3,0	1900	—	35 ... 70	14 ... 22	0,7 ... 1,5
Ag	4,73	W + Ba	1,6	1,5	1000	—	5 ... 100		
Fe	4,76	W-O-Ba	1,3	0,18	1000	—			
Si	4,80	Cs-O-Ag	0,75	0,098		—			

Emission aus. Sie werden infolge der geradlinigen Bewegungen der Elektronen bei entsprechendem Abstand von der Spitzenkatode riesenhaft vergrößert.

Bei der *Stoßemission* überträgt sich die kinetische Energie von Ionen oder Elektronen, die auf die Oberfläche der Katode prallen, auf einzelne Elektronen, die dann die Katode verlassen können. Ähnlich wie beim Wassertropfen, der ins Wasser fällt und dabei viele Wassertropfen herausschleudert, können auch hier mehrere Elektronen durch einen Stoß den Festkörper verlassen. Beim Elektronenstoß spricht man dann von *Sekundärelektronen* und Sekundärelektronenemission. Das Verhältnis von auftreffender zu austretender Elektronenzahl heißt Ausbeutefaktor δ. Er ist von der Spannung, d. h. von der Primärenergie der auftreffenden Elektronen abhängig und durchläuft für einige 100 eV ein Maximum (Bild 6.2.2a)). Bezüglich des Wertes dieses Maximums können drei Materialgruppen unterschieden werden:

- Stoffe mit $\delta < 1$. Dazu gehören die Alkali- und Erdalkalimetalle sowie Kohlenstoff (Graphit).
- Stoffe mit etwa $1 < \delta < 2$, dazu gehören die meisten Metalle.
- Stoffe mit $\delta > 2$. Hierzu gehören vor allem Metalloxide und -chloride (BeO: ≈ 10, Mg: ≈ 8, NaCl: ≈ 6) und einige Metall-Legierungen (GaP-Cs: ≈ 50, Cu-Mg: ≈ 13, Ni-Be: ≈ 12, Al-Cu: ≈ 10).

Weiter tritt bei den Sekundärelektronen eine typische Energieverteilung auf, wie sie Bild 6.2.2b) zeigt. Die Verzögerung der Sekundärelektronen ist mit ca. 10^{-10} s sehr gering.

6.2.1.2. Fotoemission

Bei der Fotoemission ist von der Energie eines Photons, der Lichtwellenlänge λ auszugehen.

$$W = hv\,. \tag{2}$$

Dabei bedeutet h das PLANCKsche Wirkungsquantum mit $6.6256 \cdot 10^{-34}$ Ws². Ein Photon kann seine Energie vollständig auf ein Leitungselektron übertragen. Entsprechend (2) können also Photonen mit einer hinreichend hohen Frequenz Elektronen aus dem Metallverband lösen. Dieser Minimalfrequenz entspricht eine maximale Wellenlänge, die *Grenzwellenlänge* genannt wird:

$$\lambda_m = \frac{c \cdot h}{W} = \frac{c \cdot h}{e_0 U} \approx \frac{1250}{W/\mathrm{eV}}\,\mathrm{nm}\,. \tag{3}$$

Für langwelliges rotes Licht von etwa 800 nm ergibt sich hiernach eine Energie von 1,5 eV. Zum blauen Licht mit 400 nm gehören 3 eV. Aus Tab. 6.2.1 ist ersichtlich, daß bei den meisten Metallen zur Auslösung von Elektronen durch den Fotoeffekt ultraviolettes Licht notwendig ist.

Wird die Energie der Photonen größer als die Austrittsarbeit, so können die Elektronen nach ihrem Austritt die Differenzenergie als kinetische Energie besitzen. Diese versetzt sie in den Stand, gegen ein *negatives Potential*

$$U \leqq \frac{h}{e_0}(\nu - \nu_g) \quad \text{mit} \quad \nu_g = \frac{c}{\lambda_m} \tag{4}$$

anzulaufen. Gemäß Bild 6.2.3a) lädt sich deshalb die Anode ohne Stromentnahme maximal auf dieses Potential auf. Für verschiedene Wellenlängen ergibt sich so ein Verlauf zwischen Spannung und Frequenz, wie ihn Bild 6.2.3b) zeigt. Der Durchstoßpunkt für die Spannung Null entspricht der *Austrittsarbeit*

$$W = h\nu_g\,. \tag{5}$$

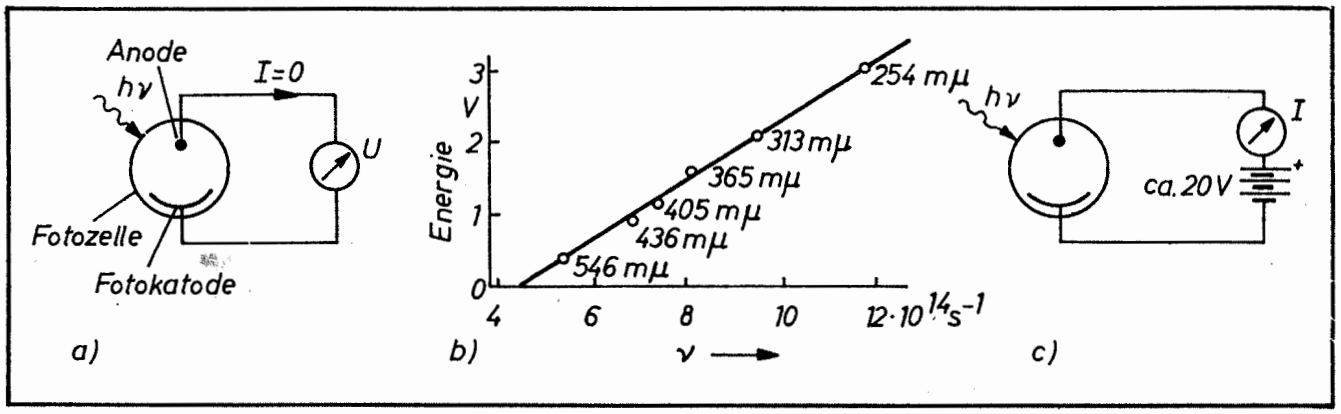


Bild 6.2.3. Zu Kenndaten von Fotozellen

a) Meßprinzip für den Zusammenhang von Gl. (4)

b) Darstellung der Meßergebnisse mit der EINSTEIN-Geraden. Der Durchstoßpunkt der Geraden bestimmt die Austrittsarbeit der Fotokatode gemäß Gl. (5)

c) Messung des Fotostromes als Funktion der eingestrahlten Lichtintensität (Quantenausbeute)

Der Zusammenhang zwischen Spannung und Frequenz in Gl. (4) ist unabhängig von der eingestrahlten Lichtintensität. Sie bestimmt die Zahl der erzeugten Elektronen. Das Verhältnis aus eingestrahlter Lichtintensität (Anzahl von Photonen) und erzeugten Elektronen wird Quantenausbeute genannt. Sie ist kleiner als 1 und von der Frequenz abhängig. Sie ist bei der Grenzwellenlänge sehr klein und nimmt zu kürzeren Wellen hin zu. Bei sehr kurzen Wellen nimmt sie wieder ab, weil nicht mehr alle Photonen absorbiert werden. Aus diesem Grunde hat jede Fotokatode eine maximale *Quantenausbeute* in einem Frequenzbereich, der etwas über der Grenzwellenlänge liegt. Dieser Zusammenhang ist mit einer Schaltung gemäß Bild 6.2.3c) meßbar. Dabei muß die Spannung so groß gewählt werden, daß alle Elektronen, die aus dem Metall herausgeschlagen werden, zur Anode gelangen.

Die *Spektralempfindlichkeit* reiner Metalle zeigt Bild 6.2.4a). Diese Materialien haben heute keine Bedeutung mehr. Es haben sich allgemein Mehrstoffkatoden durchgesetzt, die u. a. auf GÖRLICH, BRAINBRIDGE und KOLLER zurückgehen und folgende Substanzen verwenden: Cs-Cs_2O-Ag; K-Bi; K-Sb; Cs-Bi; Cs-Sb. Viele *Katoden* dieser Art sind *heute genormt*. Einige Beispiele zeigen Bild 6.2.4b) sowie Tab. 6.2.2. Den Spektralverlauf einer sehr breitbandigen Fotokatode auf der Basis Sb-K-Na-Cs zeigt Bild 6.2.4c). Hier wird auch der Einfluß des Glasfensters bei etwa 180 nm deutlich.

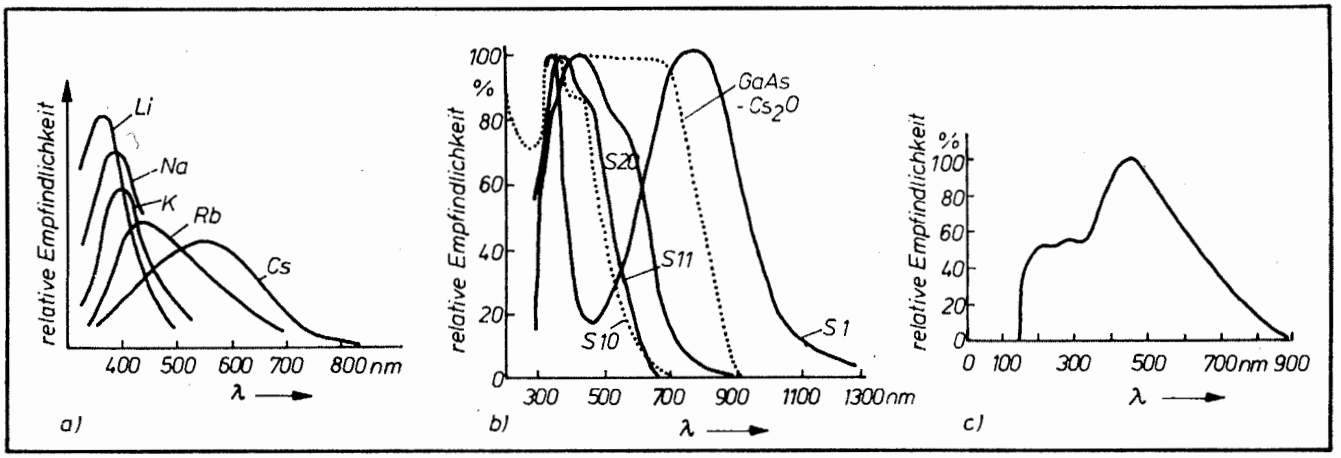


Bild 6.2.4. Spektrale Empfindlichkeit von Fotokatoden

a) Typischer Verlauf mit relativer Empfindlichkeit der verschiedenen reinen Metalle

b) Relative Empfindlichkeit von wichtigen genormten Fotokatoden. Daten dazu (insbesondere die Quantenausbeut ergänzt Tab. 6.2.2

c) Verlauf einer breitbandigen Fotokatode des Typs ERMA (extended red multi-alkali) mit einem Fenster aus Kieselgla

Tabelle 6.2.2. Werte einiger typischer technischer Fotokatoden

Katode	Chem. Zusammensetzung	λ_m µm	η_E	Empfindlichkeit mA/W
S20	Na_2KSb-Cs	0,38	0,22	67,5
S11	Cs_3Sb-O	0,39	0,19	59,8
Super S11	Cs_3Sb-O	0,41	0,23	76,0
S10	BiAgOCs	0,42	0,068	23,0
S1	AgOCs	0,8	0,004	2,58
Bialkali	K_2CsSb	0,38	0,27	82,8
—	GaAs-Cs_2O	0,4	0,35	

Um diesen UV-Bereich zu erreichen, ist bereits spezielles Kieselglas notwendig. In der letzten Zeit sind *neue Fotokatoden* hinzugekommen, die auch Halbleitermaterialien, wie z.B. Si oder spezielle GaAs-Cs- und $InAs_{0,15}P_{0,85}$-Cs_2O-Verbindungen verwenden. Sie zeichnen sich unter anderem durch einen relativ gleichmäßigen Spektralverlauf, hohen Quantenwirkungsgrad und geringen Dunkelstrom aus.

Neben der Quantenausbeute wird bei Fotoempfängern auch oft die *maximale Empfindlichkeit* in µA/lm angegeben. Sie liegt meist bei Werten von 30 bis 60, Extremwerte sind 5 und 250. Schließlich ist noch der Dunkelstrom wichtig. Seine Werte liegen bei ca. 10^{-13} bis 10^{-16} A/cm². Durch Kühlung der Katode läßt er sich wesentlich senken.

Für den *technischen Aufbau* von Fotokatoden sind zwei Lösungen zu unterscheiden:

- *Lichteinfall und Emission auf der gleichen Seite.* Dann befindet sich die Fotokatode auf einem reflektierendem Körper, z.B. Ag.
- *Lichteinfall und Emission auf den beiden entgegengesetzten Seiten.* Dann befindet sich die Fotokatode auf durchsichtigem Material und wird durch dieses Material hindurch belichtet.

6.2.1.3. Thermoemission

Die thermische Geschwindigkeit der Leitungselektronen in einem Festkörper gehorcht der FERMI-Verteilung gemäß Bild 6.2.5a) (vgl. Bild 3.1.6). Die *mittlere Geschwindig-*

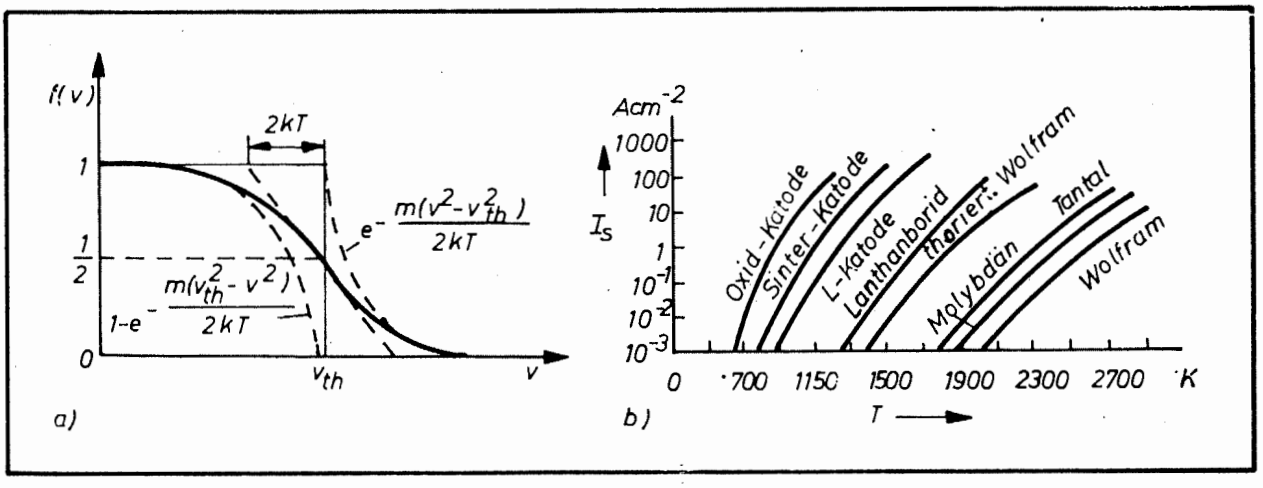


Bild 6.2.5. Zur Thermoemission

a) Geschwindigkeitsverteilung von Elektronen im Leiter. Die gestrichelten Linien stellen Näherungen der wirklichen FERMI-Verteilung durch BOLTZMANN-Verteilungen dar. Die Ordinate trägt die relative Anzahl der Elektronen, deren Geschwindigkeit größer als die auf der Abszisse aufgetragene Geschwindigkeit v ist

b) Sättigungsströme verschiedener Katoden als Funktion der Temperatur

keit ist durch folgende Gleichung gegeben:

$$\boxed{W = \tfrac{1}{2} m_0 v_{\mathrm{th}}^2 = kT = e_0 U_{\mathrm{T}}} \,. \tag{6}$$

Hierbei bedeuten

m_0 = Masse des Elektrons $9{,}11 \cdot 10^{-22}$ g,
k = BOLTZMANN-Konstante $1{,}38 \cdot 10^{-23}$ J K^{-1},
T = absolute Temperatur in K,
c_0 = Ladung des Elektrons $1{,}60 \cdot 10^{-19}$ As,
U_{T} = Spannung (Temperaturspannung), die bei Elektronen genau die Geschwindigkeit v_{th} erzeugt.

Für die *Temperaturspannung* gilt die Näherung

$$U_{\mathrm{T}} \text{ in Volt} \approx 8{,}6 \cdot 10^{-5} T \text{ in Kelvin} \,. \tag{7}$$

Einer Temperatur von 800 °C (Rotglut), wie sie bei Oxidkatoden gebräuchlich ist, entsprechen 0,1 eV. Für die Weißglut der Wolframkatode ergeben sich 2000 °C bzw. 0,2 eV. Die thermische Energie der Elektronen ist somit klein gegen die Austrittsarbeit U_0. Infolge der Energieverteilung von Bild 6.2.5a) gibt es jedoch wenige Elektronen, die eine ausreichende Energie besitzen, um die *Austrittsarbeit* zu überwinden. Merklich viele werden es gerade dann, wenn

$$U_{\mathrm{T}} \approx \left(\frac{1}{20} \cdots \frac{1}{100}\right) U_0 \tag{8}$$

gilt. Zur Emission bei erhöhter Temperatur tragen nur jene Elektronen bei, die in der Verteilungskurve von Bild 6.2.5a) ganz rechts liegen. Wie aus diesem Bild weiter hervorgeht, genügt in diesem Gebiet näherungsweise die MAXWELL-Verteilung. Die *relative Anzahl der austretenden Elektronen* folgt somit gemäß Gl. (8) zu etwa

$$\frac{n}{n_{\mathrm{ges}}} = \mathrm{e}^{-U_0/U_{\mathrm{T}}} \approx 10^{-4} \ldots 10^{-9} \,. \tag{9}$$

Können, wie bei einer Glühlampe, keine Elektronen das Vakuumgefäß verlassen, so bildet sich um die Glühwendel eine Elektronenwolke aus. Entsprechend einem dynamischen Gleichgewicht werden dann gleichviele Elektronen bezüglich des Metalls aus- und eintreten.

Werden alle austretenden Elektronen mittels einer positiven Elektrode (Anode) abgesaugt, so entsteht der maximale *Emissionsstrom* einer Röhre. Er wird Sättigungsstrom genannt und ergibt sich aus der RICHARDSON-*Gleichung*

$$\boxed{I_{\mathrm{s}} = A \cdot T^2 \, \mathrm{e}^{-e_0 U_0 / kT}} \tag{10}$$

Die Temperaturabhängigkeit dieser Gleichung zeigt Bild 6.2.5b). Es gibt also eine gewisse Temperatur, von der ab der Sättigungsstrom steil zunimmt.

Die Konstante A in Gl. (10) liegt theoretisch bei 120 A cm^{-2} K^{-2}. Werte für einige praktische Katoden zeigt Tab. 6.2.1. Bei dünnen Drähten (Katoden) überlagert sich der Thermoemission die Feldemission. Daher wird hier praktisch nie die Sättigung erreicht. Aus (10) bzw. Bild 6.2.5b) geht die starke Änderung des Sättigungsstromes mit der Temperatur hervor. In grober Näherung gilt, daß eine Temperaturänderung von $\pm 10\%$ den Sättigungsstrom auf das Zehnfache steigen oder auf ein Zehntel sinken läßt.

Bei *technischen Katoden* sind zwei Unterscheidungsmerkmale wichtig. Zunächst die Einteilung in

direkt und indirekt geheizte Katoden.

Bei *direkt geheizten Katoden* dient der stromdurchflossene heiße Draht zugleich zur Elektronenemission. Bei der *indirekten Katode* ist der Draht mit Al_2O_3 gegenüber der eigentlichen Katode aus Nickel und der darauf aufgetragenen Emissionsschicht isoliert. Hierdurch entsteht der Vorteil, daß Heizstromkreis und Emissionsstromkreis elektrisch getrennt sind und ein konstantes Potential für die gesamte Katode existiert. Der Nachteil ist die höhere Heizleistung und die zeitliche Verzögerung, bis die Katode die richtige Temperatur besitzt.

Das zweite Unterscheidungsmerkmal ist das *Katodenmaterial*:

- reine Metallkatoden (nur direkt)
- thorierte Wolframkatoden (nur direkt)
- Oxidkatoden (BaO, SrO und evtl. CaO)
- Vorratskatoden (nur indirekt).

Die ersten beiden Katoden sind sehr robust und werden vor allem in Senderöhren benutzt. Sie sind nur als reine Drahtkatoden verwendbar und benötigen viel Heizenergie.

Bei den *Vorratskatoden* werden die Nachteile der üblichen Oxidkatoden dadurch vermieden, daß ständig das verbrauchte, verdampfte oder vergiftete Oxid durch neues ersetzt wird. Sie vermeiden so auch weitgehend die Ausbildung einer schlechtleitenden Zwischenschicht.

Drei Katoden zeigt beispielhaft Bild 6.2.6.

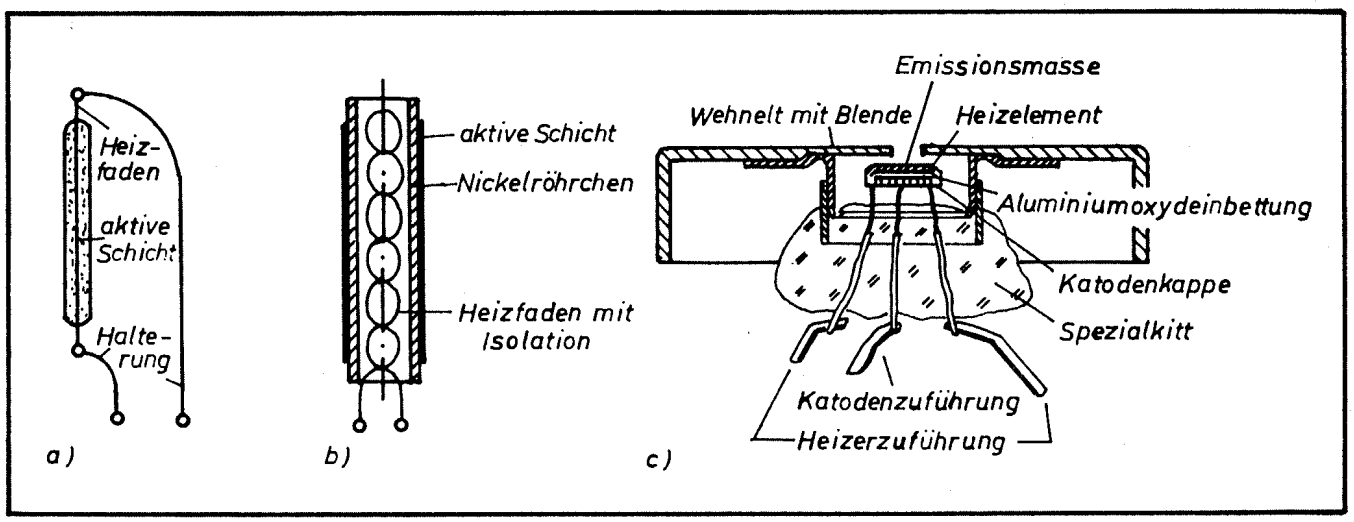


Bild 6.2.6. Beispiele von Katoden

a) Direkt geheizte BaO-Katode, typisch für Batterieröhren und Digitrons
b) Indirekt geheizte Katode (zylinderförmig), typisch für Verstärkerröhren
c) Flachkatode, indirekt geheizt, wie sie in etwa bei Fernseh- und Oszillographenröhren angewendet wird

6.2.1.4. Ströme einer Röhre

Um die Katode herum sei eine Anode angeordnet. Für die Messung des Katodenstromes wird eine Schaltung gemäß Bild 6.2.7a) verwendet. Die Ergebnisse in zwei unterschiedlichen Darstellungen (linear und halblogarithmisch) zeigt Bild 6.2.7b) und c). Es sind *drei Gebiete* bezüglich des funktionellen Zusammenhangs von Katodenstrom und Anodenspannung zu unterscheiden:

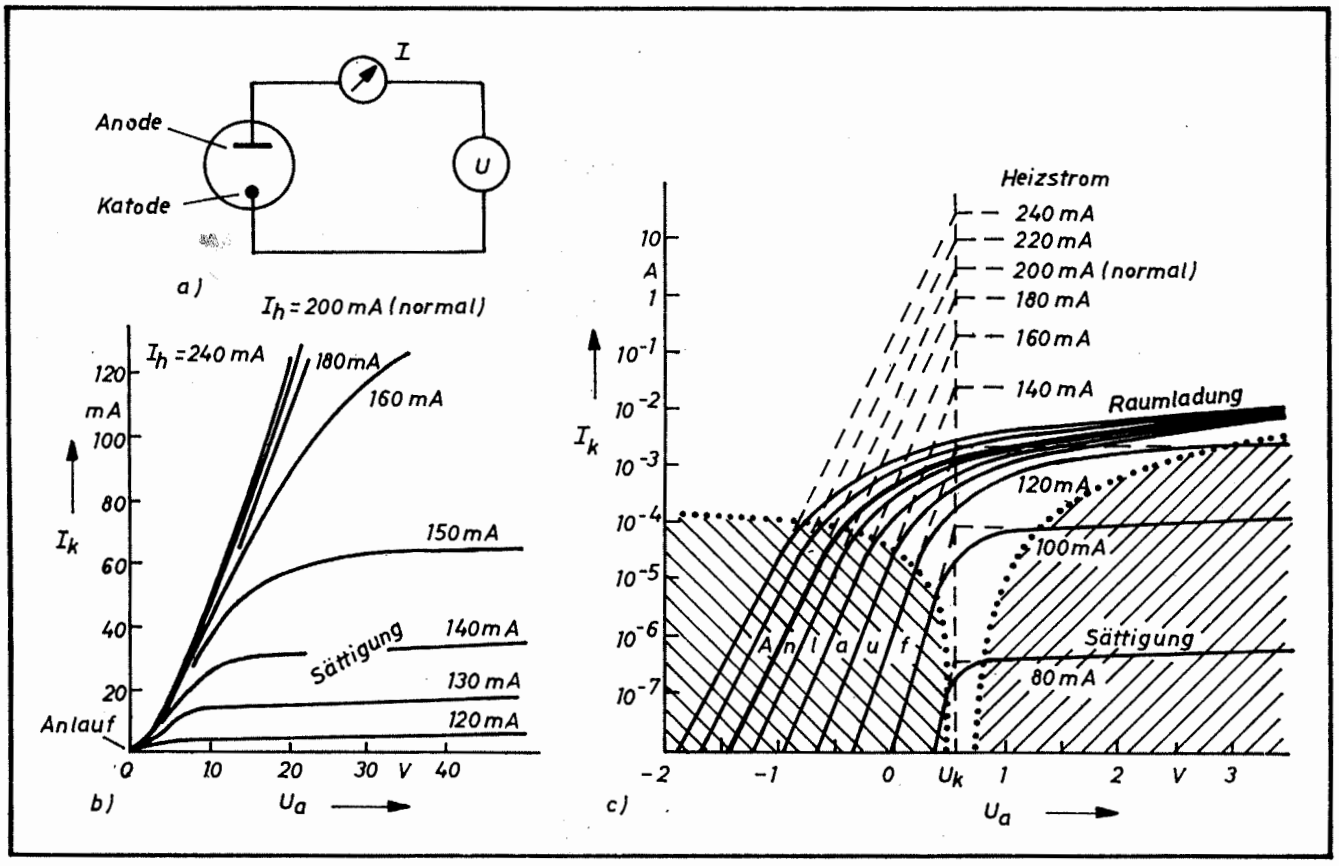


Bild 6.2.7. Katodenströme bei einer Zweipolröhre (Diode)

a) Schaltung

b) In linearer Darstellung

c) In halblogarithmischer Darstellung (beide Kennlinien von einer EF 12, bei der Gitter und Anode zusammengeschaltet sind)

Beim *Anlaufstrom* ist die Anodenspannung gegenüber dem Kontaktpotential Katode — Anode negativ. Es gelangen daher nur solche Elektronen zur Anode, die über eine entsprechend große Austrittsgeschwindigkeit verfügen. Dieser Fall entspricht dem der Fotozelle in Bild 6.2.3a). Infolge der MAXWELL-Verteilung gilt hier

$$\boxed{I_K = I_S \exp\left[(U_K - U_a)/kT\right]} \,. \tag{11}$$

Dieser Zusammenhang ist als Gerade deutlich in Bild 6.2.7c) zu erkennen (U_K = Kontaktpotential).

Ist die Anodenspannung sehr groß, so werden alle austretenden Elektronen voll von der Anode übernommen, und es entsteht der *Sättigungsstrom* I_S, der nur von der Katodentemperatur und vom Katodenmaterial abhängt.

Zwischen Anlauf- und Sättigungsgebiet liegt das wichtige Gebiet mit dem *Raumladungsstrom*. Hier wirken die zur Anode fliegenden Elektronen als Raumladung und begrenzen so den Strom. Mit der *Raumladekonstanten*

$$k_R \approx \frac{4}{3}\varepsilon_0 \frac{A}{a^2}\sqrt{\frac{2e_0}{m_0}} \tag{12}$$

gilt in diesem Gebiet

$$\boxed{I_K = k_R U_a^{3/2}} \tag{13}$$

Der Strom geht also parabelförmig in den Sättigungsstrom über. Die Konstanten bedeuten:

ε_0	Dielektrizitätskonstante des Vakuums	8,854185 As/Vm
A	Fläche der Katode, Anode	
a	Abstand Anode — Katode	
e_0	Ladung des Elektrons	$1{,}602189 \cdot 10^{-19}$ As
m_0	Masse des Elektrons	$9{,}109853 \cdot 10^{-31}$ kg.

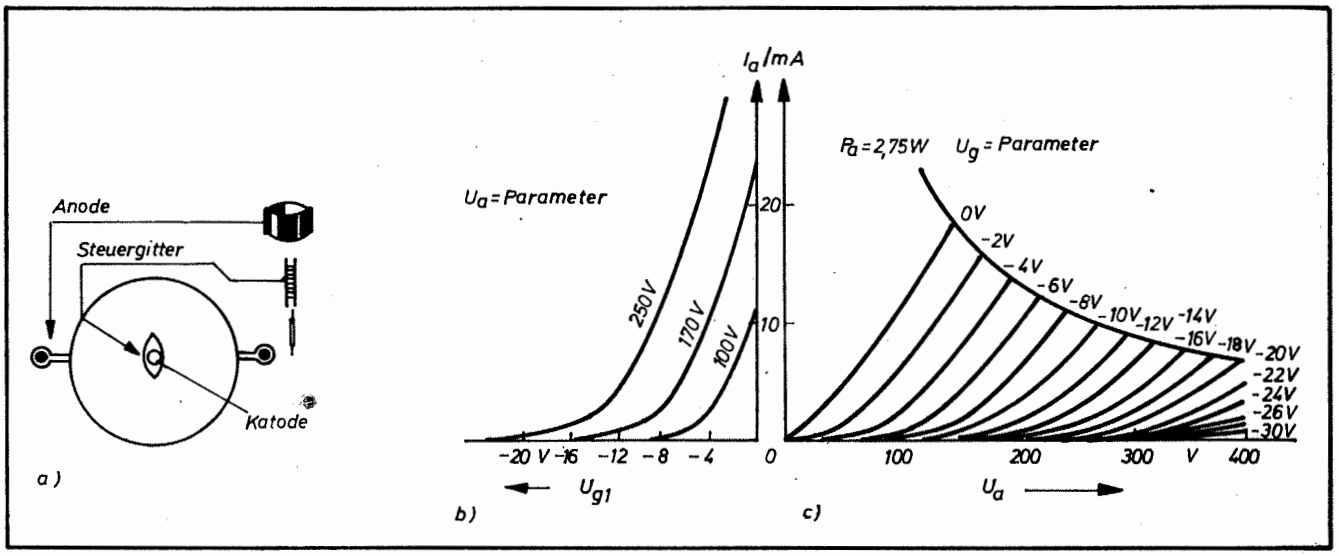


Bild 6.2.8. a) Aufbau und b) und c) Kennlinien einer Röhrentriode

Wird zwischen Anode und Katode ein *Gitter eingefügt*, so entsteht eine *Triode* (Bild 6.2.8a)). Sie wird so betrieben, daß dem Gitter eine negative Spannung gegeben wird und deshalb nur ein extrem kleiner Gitterstrom (ca. 10^{-14} A) fließen kann. Durch die Maschen des Gitters gelangt aber ein Teil des Anlaufstromes zur Anode. So entstehen Kennlinien, wie sie Bild 6.2.8b) und c) darstellt. Sie gehorchen in brauchbarer Näherung dem $U^{3/2}$-Gesetz. Über das Gitter kann somit der *Anodenstrom effektiv gesteuert* werden. Bei Oszillographen und Fernsehröhren ist das Gitter nur eine *Lochblende* (Bild 6.2.6c)) und wird dann WEHNELT-*Zylinder* genannt.

6.2.1.5. Bewegung und Ablenkung von Elektronen

Werden ruhende Elektronen einem Feld der Stärke U ausgesetzt, z.B. Katode — Anode, so bewegen sie sich in Richtung dieses Feldes und nehmen dabei kinetische Energie auf. *Nach dem Durchlaufen dieses Feldes besitzen sie die Geschwindigkeit*

$$v = \sqrt{\frac{2e_0}{m_0} U} \approx 5{,}93 \cdot 10^7 \sqrt{\frac{U}{\mathrm{V}}} \frac{\mathrm{cm}}{\mathrm{s}} \tag{14}$$

Die Geschwindigkeit eines Elektrons ist damit nur von dem durchlaufenen Spannungsunterschied abhängig. Auf *Äquipotentiallinien* besitzen Elektronen, die von gleichem Potential ausgingen auch *gleiche Geschwindigkeit*. Die *Flugzeit* bis zu den Äquipotentiallinien beträgt

$$t_{\mathrm{Fl}} = \frac{2a}{v}. \tag{15}$$

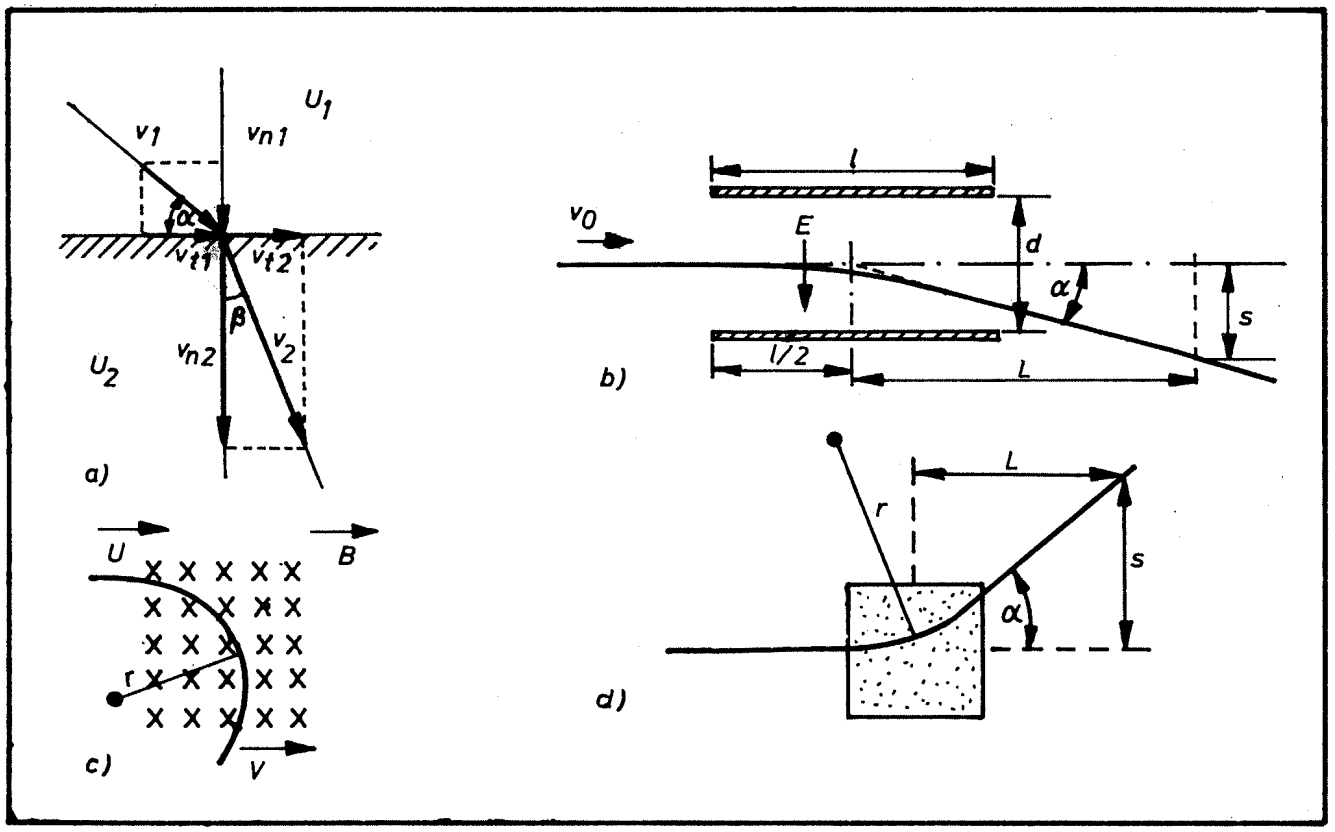


Bild 6.2.9. Ablenkung von Elektronenstrahlen

a) bei einem Spannungssprung,
b) zwischen einem Ablenkplattenpaar
c) Elektronenstrahl im homogenen Magnetfeld
d) Magnetische Ablenkung

Durchfliegen Elektronen ein Feld, so werden sie in ihrer Bahnrichtung *abgelenkt*. Für zwei Potentiale U_1 und U_2 mit sprunghafter Änderung (Bild 6.2.9a)) bleibt die Tangentialkomponente erhalten,

$$v_{t_1} = v_{t_2}\,, \tag{16}$$

während die Normalkomponente um den Betrag $U_2 - U_1$ gemäß (14) vergrößert wird. Folglich gilt das (dem optischen entsprechende) *Brechungsgesetz*

$$\frac{\sin\alpha}{\sin\beta} = \sqrt{\frac{U_2}{U_1}}\,. \tag{17}$$

Durchfliegen die Elektronen mit der Geschwindigkeit v ein *Querfeld* (Bild 6.2.9b)) der Länge l und der Feldstärke

$$E = U_1/d\,, \tag{18}$$

so werden sie um einen *Winkel*

$$\tan\alpha = \frac{e_0}{m_0}\cdot\frac{E}{v^2} \tag{19}$$

abgelenkt. Sie erreichen dabei die Geschwindigkeit

$$v_E = \frac{e_0}{m_0}\cdot E\cdot\frac{l}{v_0}. \tag{20}$$

Nach dem Durchlaufen einer Länge L beträgt die *Ablenkung*

$$s = L\tan\alpha = \frac{lLU_1}{2dU}. \tag{21}$$

Durchfliegen die Elektronen ein homogenes *Magnetfeld* (Bild 6.2.9c)), so werden sie auf eine *Kreisbahn* mit dem Radius

$$r = \frac{m_0 v}{e_0 B} \approx 3{,}37 \frac{\sqrt{U/\mathrm{V}}}{B/\mathrm{T}} \cdot \mathrm{cm} \tag{22}$$

gezwungen. Verlassen sie das Feld unter einem Winkel α, so beträgt die Ablenkung im Magnetfeld (Bild 6.2.9d))

$$s = L \cdot \tan\alpha = BlL \sqrt{\frac{e_0}{m_0} \cdot \frac{1}{2U}}$$

$$\approx 0{,}298\, B/\mathrm{T} \cdot l/\mathrm{cm} \cdot L/\mathrm{cm} \cdot \sqrt{\mathrm{V}/U}\,\mathrm{cm}\,. \tag{23}$$

Ihre Geschwindigkeit wird im Magnetfeld nicht geändert.

Auf der Grundlage dieser Betrachtungen ist es verständlich, daß mit magnetischen und elektrischen Feldern eine *Elektronenoptik* möglich ist. In diesem Sinne gibt es magnetische und elektrische Linsen als Analogie zu optischen Linsen. Im Prinzip zeigt dies Bild 6.2.10.

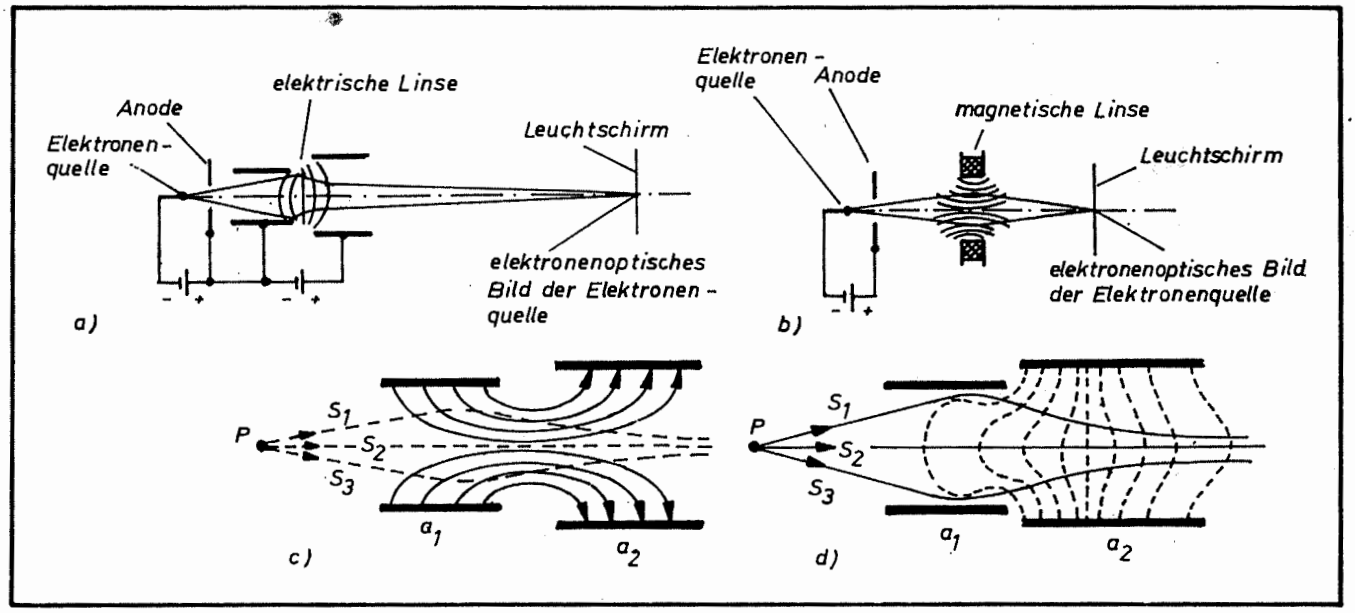


Bild 6.2.10. Elektronenoptische Linsen

a) Elektrostatische,
b) magnetische Variante
c) Verlauf der Feldlinien und
d) der Äquipotentiallinien

6.2.1.6 Leuchtstoffe

Einige allgemeine Aussagen zur Grunderscheinung der Lumineszenz und zu ihrer Einteilung in Fluoreszenz und Phosphoreszenz wurden bereits im Abschnitt 6.1.3. in allgemeinerem Zusammenhang behandelt. Hier sind vor allem die spezifischen Prinzipien und Eigenschaften der technischen Leuchtstoffe zu behandeln. Sie werden für viele *Anwendungen* benötigt. Hierzu gehören u. a.

Leuchtstoffröhren
Leuchtschirme der Oszillographen- u. Fernsehtechnik (6.2.3.)
Lumineszenz-Zellen (6.1.7.1.)
FLAD (6.1.1.2.)
Digitrons (6.2.3.1.)
Plasmapanals (6.2.4.).

Die meisten *Leuchtstoffe* sind *Kristallite* aus Oxiden, Silikaten, Sulfiden oder Fluoriden der Elemente Zink, Calcium, Cadmium oder Magnesium. Diese Materialien müssen nahezu die in der *Halbleitertechnik* geforderte *Reinheit* besitzen. Sie werden *dotiert* (verunreinigt) mit *Aktivierungen*, wozu meist Metalle, insbesondere Mangan, Kupfer, Silber, Blei und Zink, verwendet werden. Die Leuchtmechanismen können gut im *Bändermodell* (3.1.) beschrieben werden, hierzu dient Bild 6.2.11a). Von einem Aktivatorniveau wird ein Elektron ins Leitungsband gehoben. Unter Lichtausstrahlung kann dieses Elektron zu einer leeren Aktivatorstelle zurückkehren. Dann liegt *Fluoreszenz* vor. Bei der *Phosphoreszenz* ist dies nur auf Umwegen möglich. Hierbei besitzen *Haftstellen D* (Zwischenterme, metastabile Terme, Traps, Elektronenfallen — das sind nur andere Bezeichnungen) große Bedeutung. Sie lagern das freie Elektron ein und geben es erst nach längerer Zeit ($> 10^{-6}$ s, auch Sekunden und Minuten) wieder frei.

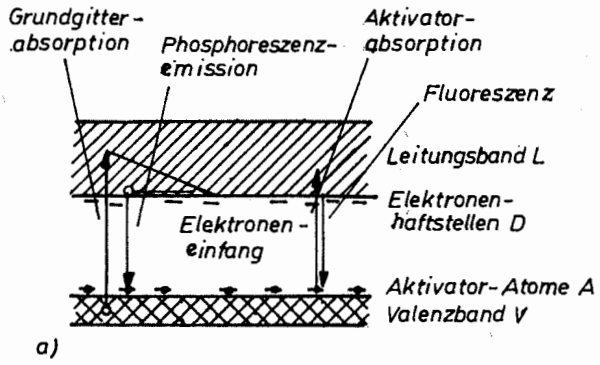
Die *Leuchtstoffe* und auch Leuchtschirme, die aus mehreren Leuchtstoffen zusammengesetzt sind, werden nach *Nomenklaturen* gekennzeichnet. In Europa erfolgt dies durch Pro Electron in Brüssel mit Buchstaben, z.B. GH; in den USA durch **JEDEC** (Joint Electron Device Engineering Committee) mit dem Buchstaben P und einer laufenden Nr., z.B.: P11. Diese Normung bestimmt nur wesentliche Fakten. Der Leuchtstoff kann daher von Firma zu Firma in gewissen Grenzen abweichen. Eine genaue Auskunft kann nur die Kurve der spektralen Energieverteilung aus den Applikationsdaten geben (z.B. Bild 6.2.11b)). Aus dem Spektrum ist zu erkennen, daß der Leuchtschirm GM (P7) zwei Schichten enthält, die übereinander angebracht sind. An der Seite, wo die Elektronen auftreffen, befindet sich ein blauer Leuchtstoff und darüber, zur Glasseite, von wo aus beobachtet wird, ein gelb-orangefarbiger. Dieser wird vor allem vom Licht des blauen angeregt, er leuchtet daher verzögert und darüber hinaus mit großer Verzögerung (Phosphoreszenz).

Für viele Anwendungen ist die *Nachleuchtdauer* eines Schirmes bedeutsam. Sie wird durch halblogarithmische Nachleuchtkurven beschrieben (Bild 6.2.11c) und d)). Bei dem Zweikomponentenschirm kann als spezielle Anwendung durch Filter eine Farbe ausgeschaltet werden. Die Nachleuchtzeiten liegen im Bereich von 100 µs bis zu mehreren Sekunden.

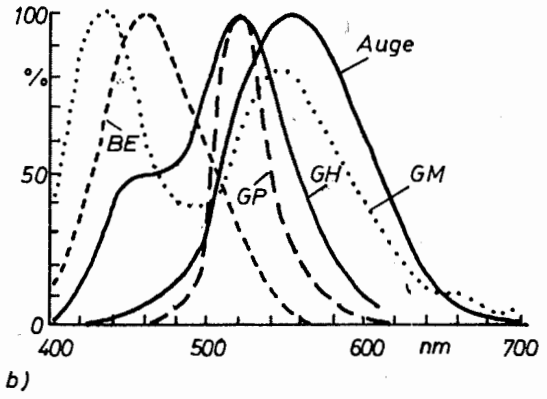
Die *Intensität* eines Leuchtstoffes kann mittels der *Beschleunigungsspannung* (Geschwindigkeit der auftreffenden Elektronen) und der *Strahlstärke* (Anzahl der Elektronen) verändert werden. Dies zeigt Bild 6.2.11d) und e). Der *energetische Wirkungsgrad* liegt dabei bei wenigen Prozent, maximal 10%.

▶

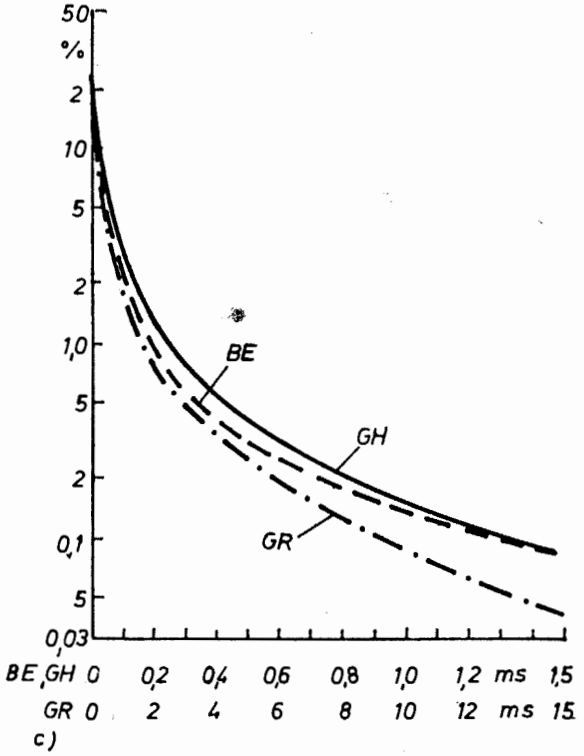
Bild 6.2.11. Leuchtstoffe
a) Schematisches Verhalten im Bändermodell
b) Spektrale Energieverteilung einiger Bildschirme und Empfindlichkeit des menschlichen Auges zum Vergleich:
P7 = GM: blau und gelblich-grün nachleuchtend, ZnS:Ag auf ZnCdS:Cu
P11 = BE: blau, ZnS:Ag(Ni)
P31 = GH: grün, ZnS:Cu
P39 = GR: grün (im Bild fälschlich GP)
c) Nachleuchten einiger Schirme
d) und e) Leuchtdichte als Funktion der Beschleunigungsspannung und des Schirmstromes
f) und g) Nachleuchten der beiden Komponenten des Schirmes GM
h) Lichtausbreitung bei normalem und
i) bei Al-hinterlegtem Leuchtschirm
j) Leuchtdichten bei beiden Typen von Leuchtschirmen


Grundgitter-absorption
Phosphoreszenz-emission
Aktivator-absorption
Fluoreszenz
Leitungsband L
Elektronen-haftstellen D
Elektronen-einfang
Aktivator-Atome A
Valenzband V

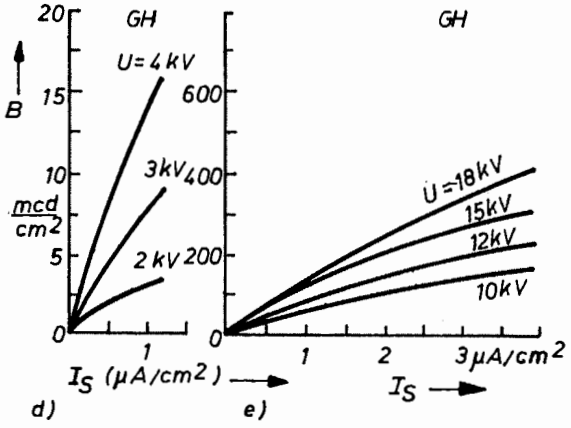
a)


100
%
50
0
Auge
BE
GP
GH
GM
400
500
600
nm
700

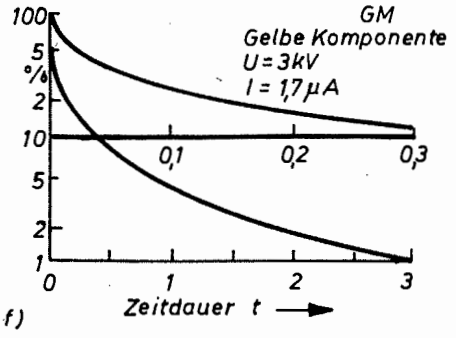
b)


50
%
2
10
5
2
1,0
5
2
0,1
5
0,03
BE
GH
GR
BE, GH 0 0,2 0,4 0,6 0,8 1,0 1,2 ms 1,5
GR 0 2 4 6 8 10 12 ms 15

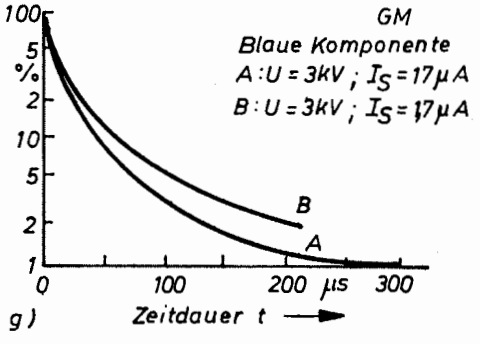
c)


20
B
15
10
mcd/cm²
5
0
GH
U=4 kV
3kV
2 kV
600
400
200
1
I_S (μA/cm²)
GH
U=18kV
15kV
12kV
10kV
1
2
3μA/cm²
I_S

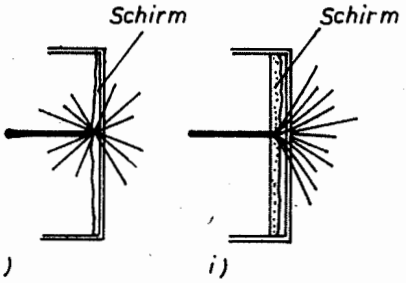
d) e)


100
5
%
2
10
5
2
1
GM
Gelbe Komponente
U=3kV
I=1,7μA
0,1
0,2
0,3
0
1
2
3
Zeitdauer t

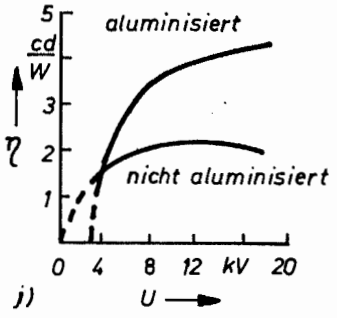
f)


100
5
%
2
10
5
2
1
GM
Blaue Komponente
A: U=3kV; I_S=17μA
B: U=3kV; I_S=17μA
B
A
0
100
200 μs
300
Zeitdauer t

g)


Schirm
Schirm

h) i)


5
cd/W
3
η
2
1
aluminisiert
nicht aluminisiert
0
4
8
12
kV
20
U

j)

Bei intensivem Elektronenbeschuß besteht die Gefahr, daß die Kristallite verändert werden und weniger gut leuchten. Es wird davon gesprochen, daß der *Leuchtschirm einbrennt.* Dies erfolgt eigenartigerweise bei Beschleunigungsspannungen unter 1000 V besonders leicht. Es wird vermutet, daß dann die Elektronen den Leuchtstoff nur an der Oberfläche belasten und so überlasten. Bei zu hohen Beschleunigungsspannungen (über 10 kV) dringen die Elektronen zu tief ein und erzeugen so wieder weniger Licht.

Unter normalen Bedingungen leuchtet der Leuchtstoff nach allen Seiten, also auch ins Innere der Röhre (Bild 6.2.11h)). Diese Intensität geht einmal verloren und mindert dann den möglichen Kontrast, indem sie von rückwärts den ganzen Schirm aufhellt. Dies konnte durch mit *Aluminium hinterlegte Leuchtschirme* geändert werden (Bild 6.2.11 i)). Die Schichtdicke (ca. 0,1 μm) und die Technologie sind kritisch, so daß derartige Schirme erheblich aufwendiger sind. Den Gewinn zeigt in etwa Bild 6.2.11j). Weiter wird damit die unerwünschte *Schirmaufladung* verhindert. Leuchtstoffe sind nämlich Isolatoren. Die ständig auf sie auftreffende Ladung der Elektronen muß abgeführt werden, sonst bilden sich Aufladungen, welche den Schirm für die Elektronen unzugänglich machen (weiße Wolken). Dieses Abführen der Ladung kann entweder über Sekundärelektronenemission oder durch die Al-Hinterlegung erreicht werden.

Beispielhaft sei erwähnt, daß Leuchtstoffe für die *Schwarzweiß-Fernsehröhren* aus komplementär strahlenden Komponenten zusammengesetzt sind: ZnS Ag und ZnS, CdS Ag. Für *Farbfernsehschirme* werden z.B. verwendet ZnS Ag (blau); (Zn, Cd)S Cu (grün); ZnP Eu oder Y_2O_3 Eu (rot).

Für *Oszillographen-* und *Digitronröhren* werden vorrangig ZnO, ZnS für grüne Schirme, für blaue Schirme Cadmiumwolframat und bei nachleuchtenden Schirmen Zink-Cadmiumsulfid verwendet. Die Lichtausbeute beträgt etwa 100 bis 500 lm/W. Weitere Details siehe [G 26, W 6].

Für die Anzeigen in Flugzeugen wurden wegen der sehr hellen Umgebung (Sonneneinstrahlung) *Spezialschirme* entwickelt, die Leuchtdichten bis etwa 75000 cd/m² erreichen. Sie verwenden Leuchtstoffe aus seltenen Erden und spezielle Schirmtechnologien.

6.2.2. Lichtempfindliche Röhren

Vier Typen von lichtempfindlichen Röhren sind hier zu behandeln:

Fotozellen
SEV
Bildwandlerröhren
Bildaufnahmeröhren.

6.2.2.1. Fotozellen und SEV

Die *Fotozellen* sind in ihren wesentlichen Eigenschaften bereits bei der Fotoemission (6.2.1.2.) behandelt. Sie unterscheiden sich vor allem durch ihre Katode und die spektrale Durchlässigkeit des Glases. Weiter können Vakuumfotozellen und Fotozellen mit Gasfüllung unterschieden werden. Bei *gasgefüllten Fotozellen* läßt sich analog zur Avalanche-Diode (6.1.2.), jedoch auf dem Prinzip der *Ladungsträgervervielfachung* in Gas gemäß Abschnitt 6.2.4., eine Erhöhung der Empfindlichkeit erreichen (Bild 6.2.12a)). Dabei wird allerdings die Grenzfrequenz wesentlich niedriger. Durch die Entwicklung der Halbleiter-Fotoempfänger (6.1.2.) sind sie heute praktisch vollständig verdrängt, so daß weitere Ausführungen entfallen können.

Der *Sekundärelektronenvervielfacher* (*SEV*) geht ebenfalls vom Prinzip der Fotozelle aus. Dabei werden jedoch die durch das Licht freigesetzten Elektronen weiter über

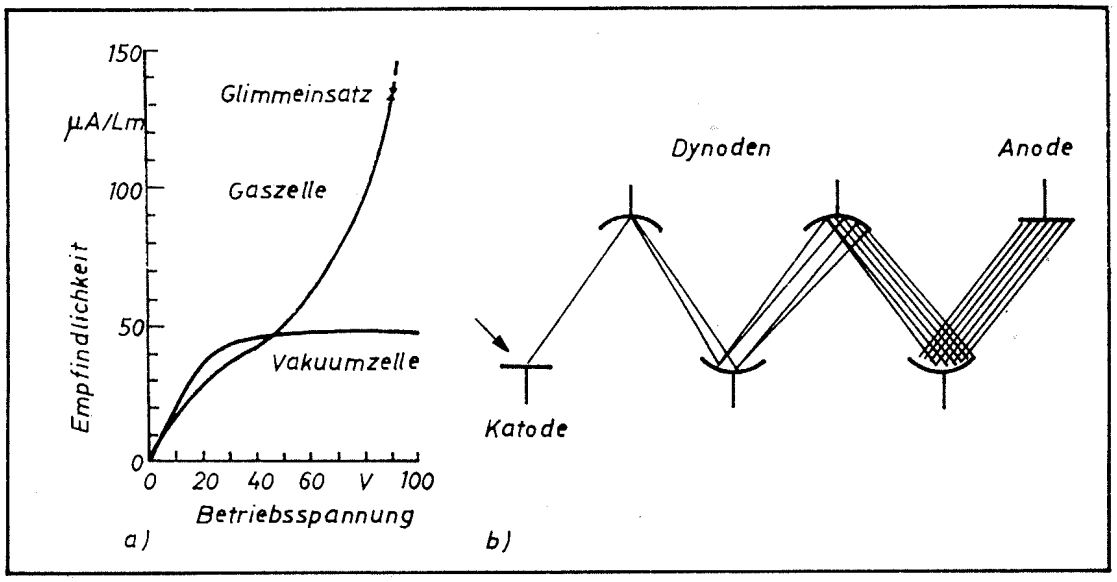


Bild 6.2.12. a) Kennlinien von Fotozellen (Vakuum- und gasgefüllte Zelle)
b) Prinzipaufbau von Sekundärelektronenverstärkerröhren (SEV)

Sekundärelektronen verstärkt. Das Grundprinzip zeigt Bild 6.2.12b). Die auf hohe Sekundärelektronenausbeute ausgelegten Dynoden können dabei in sehr unterschiedlicher Anordnung und Gestalt realisiert werden. Die Anzahl der Dynoden ist unterschiedlich, beträgt maximal 9 bis 14. Dann wird eine Vervielfachung bis zu 10^9 erreicht. Der SEV hat sich infolge seiner *extrem hohen Empfindlichkeit* bis zu 1000 A/lm und seiner großen Empfängerfläche trotz der starken Entwicklung der Halbleiterempfänger weiter behaupten können.

6.2.2.2. Bildwandler, Bildverstärker

Ein *Bildwandler* liegt vor, wenn ein optisches Bild bestimmter Frequenz, z.B. IR, in einen anderen, z.B. in den sichtbaren Bereich gewandelt werden soll. Bildwandler sind u. a. die *Nachtsichtgeräte*, welche die *thermische Strahlung* ausnutzen und *in ein sichtbares Bild wandeln*. In der Regel wird hierbei das Primärbild optisch auf einer

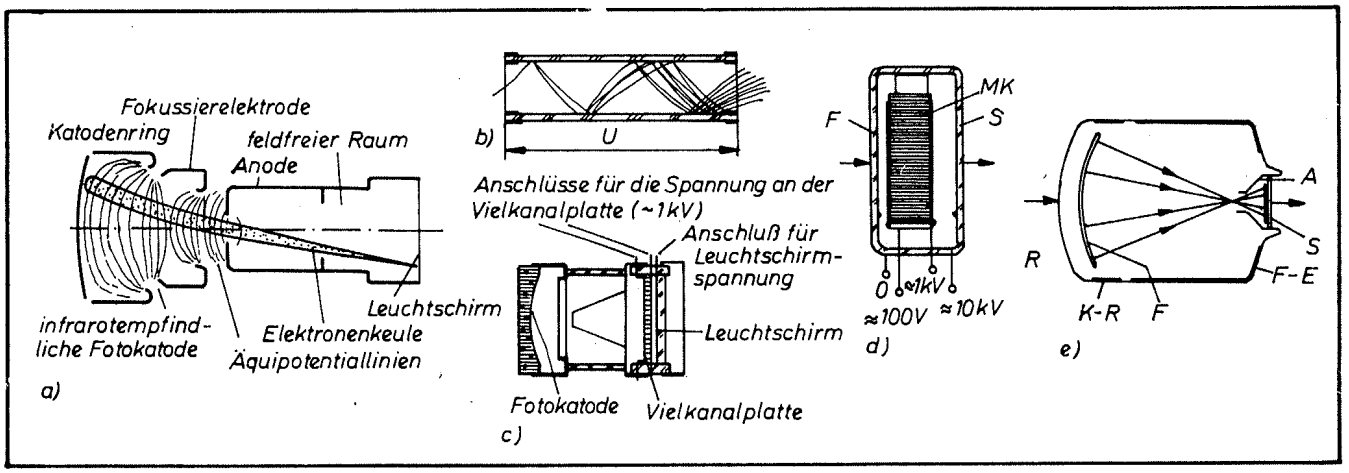


Bild 6.2.13. Bildverstärker und Bildwandler

a) Prinzipaufbau und -funktion einer Bildwandlertriode. Es ist nur der Elektronenweg für einen Bildpunkt gezeigt
b) Prinzip der Sekundärelektronenverstärkung in einem engen Kanalverstärker. Der Innendurchmesser beträgt etwa 10 bis 20 μm
c) Aufbau eines Bildverstärkers mit Vielkanalplatte
d) Aufbau eines Bildverstärkers, der nur mit Vielkanalverstärkung arbeitet
e) Schematischer Aufbau eines Röntgenbildverstärkers. Es bedeuten *R* Röntgenfluoreszenzschirm, *F* Fotokatode, *F-E* Fokussierelektrode, *A* Anode, *S* Leuchtschirm, *K-R* Katodenring

Fotokatode abgebildet. Die dabei ausgelösten Elektronen werden mittels einer elektrischen oder magnetischen Linse auf einem Leuchtschirm abgebildet. Gleichzeitig werden sie dabei durch ein Längsfeld beschleunigt und gewinnen so Energie für eine sichtbare Anzeige (Bild 6.2.13a)).

Ein *Bildverstärker* hat die Aufgabe, ein *schwaches Bild* heller und damit besser sichtbar zu machen. Es besteht also zunächst gegenüber dem Bildwandler mit Verstärker nur der Unterschied bezüglich der Ein- und Ausgangswellenlängen, also in erster Linie bezüglich der spektralen Empfindlichkeit der Fotokatode. Früher wurden Bildverstärker *kaskadiert* und erhielten dazu auf beiden Seiten der einzelnen Stufen *Bündel aus Glasfasern* (Glasfaseroptiken), um die Krümmung von Bildschirm und Katode ohne zusätzliche Unschärfe auszugleichen. In der letzten Zeit sind in die Röhre *Mikrokanalverstärker* eingesetzt worden. Dies sind Röhren von etwa 10 ... 20 µm Innendurchmesser, die innen mit einem Material hoher Sekundärelektronenausbeute versehen sind (Bild 6.2.13b)). Sie wurden zunächst in dichter Packung vor dem Leuchtschirm angeordnet und ermöglichten so Lichtverstärkungen von mehr als zehntausendfach. In letzter Konsequenz entstanden schließlich Bildverstärker, die nur die *Mikrokanalverstärker* enthielten (Bild 6.2.13c)). Sie sind äußerst kompakt bei hoher Leistungsfähigkeit.

Der *Röntgenbildverstärker* ist eine Spezialröhre der Bildwandler — Bildverstärker (Bild 6.2.13d)). Die Besonderheiten bestehen darin, daß zunächst die Röntgenstrahlen einen speziellen, Röntgenstrahlung stark absorbierenden Leuchtschirm anregen, der unmittelbar mit einer Fotokatode gekoppelt ist. Dann schließt sich eine Verstärkung der freigesetzten Elektronen an, im Bild eine elektrostatische Linse.

6.2.2.3. Bildaufnahmeröhren

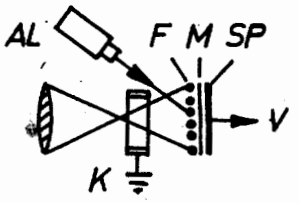
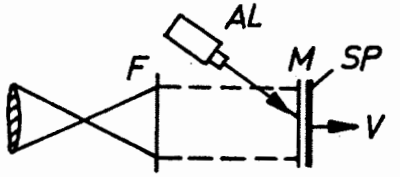
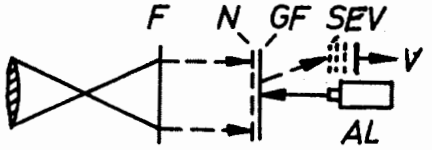
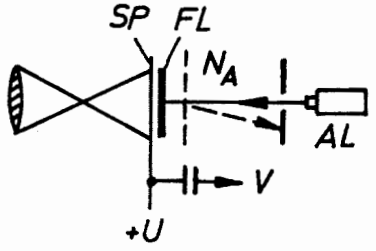
In vielen Fällen, z.B. beim Fernsehen, ist es notwendig, ein optisches Bild in ein zeitliches elektrisches Signal zu verwandeln. Das zugehörige allgemeine Prinzip solcher Aufnahmeröhren zeigt Bild 6.2.14. Mittels einer Optik wird das Bild auf eine lichtempfindliche Schicht abgebildet. Dort erfolgt eine Umwandlung der Lichtintensität in eine elektrische Größe. Mittels des inneren oder äußeren lichtelektrischen Effektes entstehen so elektrische Größen. Sie werden noch bildgetreu, also parallel gleichzeitig auf eine Speicherschicht überführt. Dort erfolgt eine zeitliche Lichtsummation für die Einzelbildpunkte. Dies führt zu einer ganz beträchtlichen Steigerung der Lichtempfindlichkeit. Wird z.B. in der lichtempfindlichen Schicht der äußere lichtempfindliche Effekt ausgenutzt, so werden in der Speicherschicht bildpunktweise die einzelnen ausgelösten Elektronen gesammelt. Mit der elektronischen Abtasteinrichtung werden dann sequentiell, also zeitlich nacheinander, die einzelnen Bildpunkte in der Speicherschicht angesteuert und die dort gesammelten Elektronen zum Verstärker übertragen.

Die Wirkungsweise der wichtigsten Bildaufnahmeröhren ist in Tab. 6.2.3 als Überblick dargestellt. Für die jeweils notwendige Beleuchtungsstärke gelten etwa folgende Richtwerte:

Superikonoskop	$> 10^2$ lx
Superorthikon	> 5 lx
Vidikon, Plumbikon	> 1 lx
Vidikon, Orthikon mit Bildverstärker	$> 10^{-3}$ lx
Bildverstärker zum Vergleich	$> 10^{-6}$ lx
Nachtsichtgerät zum Vergleich	$> 10^{-7}$ lx

Das Beispiel einer extrem lichtempfindlichen Bildaufnahme mit Lichtverstärker ist das Ebsikon gemäß Bild 6.2.15. Mit ihm sind Bilder im Bereich von 10^{-4} lx möglich. Sein Name leitet sich aus der Bezeichnung EBS = electron-bombarded silicon ab. Mit solchen Röhren wird die Empfindlichkeit aller fotografischen Filme überboten.

Tabelle 6.2.3. Eigenschaften und Prinzipien der Bildaufnahmeröhren

Röhrenart	Photoempfindl. Schicht	Gedächtnis (Speicher)	Ablesemechanismus	Aufbauschema	Kurzzeichen
Ikonoskop	Glimmer, eine Seite mit Leitschicht, die mit V verbunden ist. Photoschicht auf der anderen Seite ist in kleine Inseln geteilt		Elektronenstrahl 1000 V stabilisiert Oberfläche auf Kollektorpotential		AL = Ablesemechanismus F = photoempfindl. Schicht FL = Photoleiter GF = Glasfolie K = Kollektor M = Mosaik N = Netz N_A = Netz auf Anodenpotential SEV = Sekundärelektronenvervielfacher SP = lichtdurchlässige Signalplatte V = Verstärker
Superikonoskop (Bildikonoskop)	auf dem Glaskolben (äußerer lichtelektr. Effekt)	Mosaik einer SE-Schicht (z. B. MgO) auf 10 bis 20 μm dickem Glimmer. Rückseite leitende Schicht t = SP	Umladung der Speicherelemente durch SE-Vorgänge bei Abtastung durch Elektronenstrahl		
Superorthikon (Bildorthikon)	auf dem Glaskolben (äußerer lichtelektr. Effekt)	Photoelektronen gelangen auf Glasfolie oder MgO-Membrane GF 3 bis 5 μm, die große Sekundäremission($\delta = 3$ bis 4) hat. Netz N fängt sie auf. Positives Ladungsbild auf GF	Neutralisation der positiven Bildpunktladungen durch langsamen ($\delta < 1$) Abtaststrahl von Rückseite. Rückkehr nicht gebundener Elektronen, die in einen SEV gelangen und Bildsignal bilden		
Vidikon, Plumbikon Resistron	Photoleiter Sb_2S_3, PbO, Se, Si (innerer lichtelektr. Effekt) auf lichtdurchlässig., leit. Glasbelag	bildseitig Leitschicht SP mit Photoleiter, Leitfähigkeitsverteilung entspricht Helligkeitsverteilung. Positives Potential von SP erzeugt Potentialgebirge auf Rückseite von F	Ausgleich der Ladungsunterschiede durch Abtaststrahl. Aufladestromstöße liefern Bildsignal über SP U = 300 V		

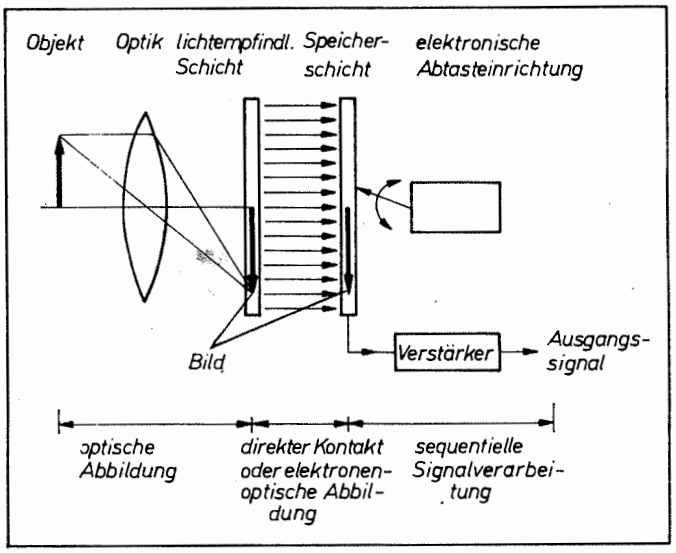


Bild 6.2.14. Prinzipieller Aufbau von Bildaufnahmeröhren

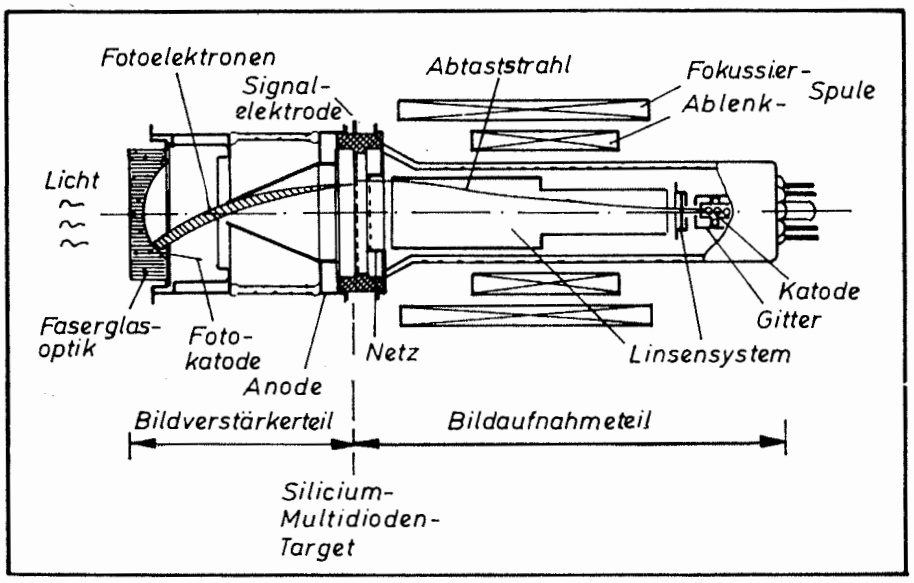


Bild 6.2.15. Schematischer Aufbau einer Ebsikon-Bildaufnahmeröhre mit Bildverstärkerteil

6.2.3. Anzeige- und Bildröhren

6.2.3.1. Digitron (Fluoreszenzanzeigeröhre)

Das Digitron, auch Fluoreszenzanzeigeröhre genannt, kann heute bereits auf vielfältige Wandlungen zurückblicken. Als Ausgangspunkt können die um 1940 entstandenen Abstimmanzeigen (magisches Auge) der Rundfunkempfänger angesehen werden. Um 1967 entstanden dann die ersten Digitronanzeigen in Japan. Sie waren noch in der Größe einer Miniaturröhre von ca. 20 mm Durchmesser und 60 mm Höhe und enthielten nur eine Siebensegmentanzeige. Anfang der 70er Jahre wurden bereits in einem weitaus dünneren, aber noch runden Glaskolben mehrere Ziffern zugleich untergebracht. Mitte der 70er Jahre wurde dann zum Flachkolben und zur direkt einlötbaren (bzw. steckbaren) Variante gemäß Bild 6.2.16a) übergegangen. Den detaillierten Aufbau zeigt Bild 6.2.16b), während die Funktion vor allem mittels eines Schnittes (Teilbild c)) erklärt werden soll.

Die Katode wird durch einen Wolframfaden von 10 bis 60 µm Durchmesser gebildet, der bis zu 20 µm Dicke mit einem Ba-Oxid beschichtet ist. Die Temperatur wird mit 650 bis 750 °C vergleichsweise gering gewählt (sonst für Oxidkatoden ca. 850 °C

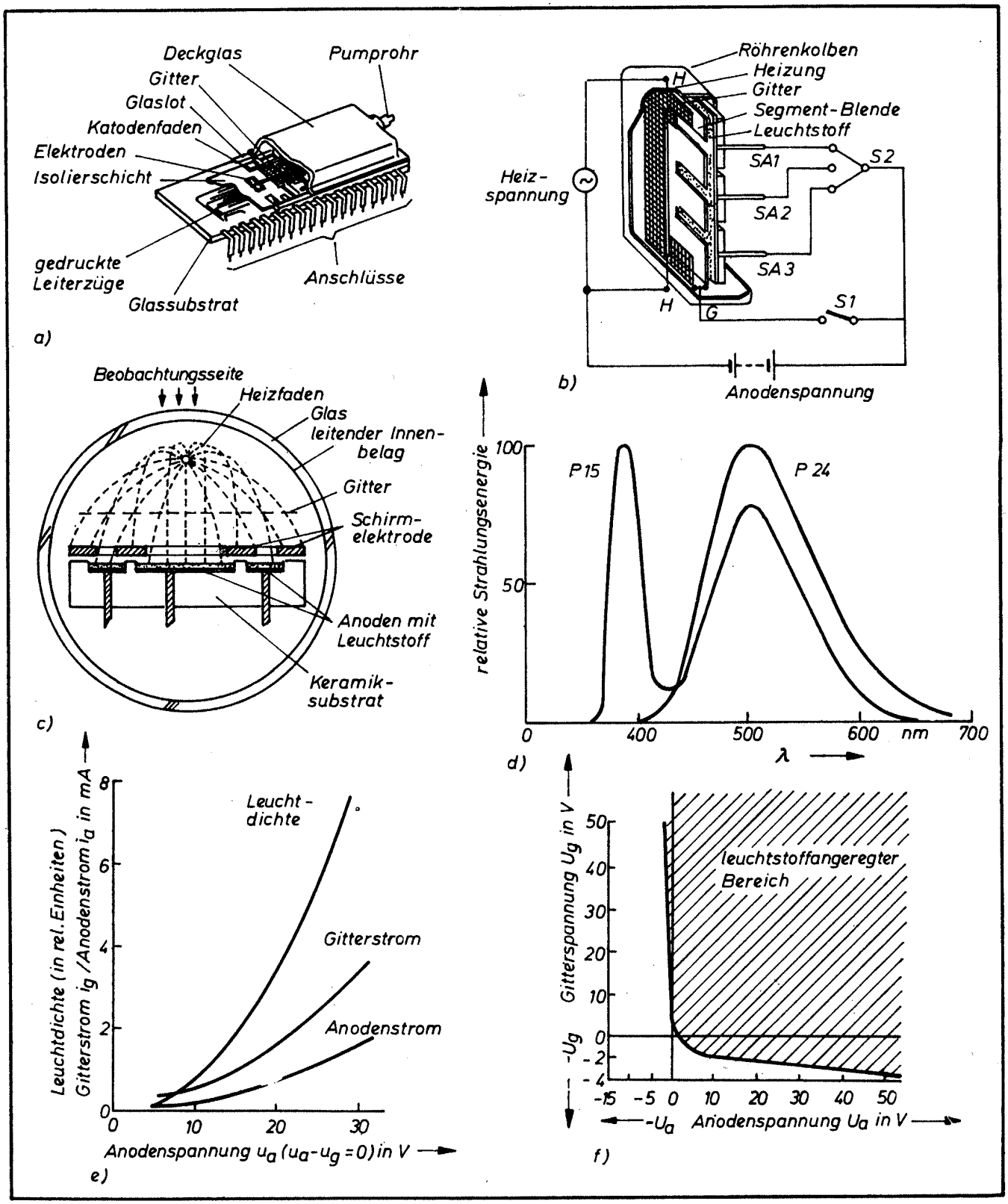


Bild 6.2.16. Digitron

a) Aufbau einer Röhre der dritten Generation
b) Schematischer Schnitt durch eine solche Röhre
c) Schnitt zur Beschreibung der Funktion
d) Spektralkurven oft verwendeter Leuchtstoffe
e) Leuchtdichte und Ströme als Funktion der Betriebsspannung bei $U_a = U_g$
f) Kennlinienverlauf zwischen Ein- und Ausbereich eines Leuchtelements

üblich). Hierdurch soll erreicht werden, daß die Rotglut des Fadens nicht zu sehen ist. Weiter soll dadurch eine unnötige Erhitzung des nahegelegenen Leuchtstoffes vermieden werden, weil dann sein Wirkungsgrad absinkt. Diese niedrige Temperatur bewirkt, daß der Heizstrom auf $\pm 10\%$ eingehalten werden muß, sonst sinkt die Elektronenemission, oder die speziell präparierte aktive Substanz verdampft. Es werden so Heizspannungen im Bereich von 1 bis 3 V, bei Heizströmen von 20 bis 50 mA erreicht.

Gegenüber dem Faden befindet sich ein sehr feines Gitter, üblich sind Nickeldrähte von 15 bis 30 μm Durchmesser, die zu Geweben mit 30 bis 200 μm Maschenweite verarbeitet sind. Es gibt auch Gitter, die aus massiven Blechen geätzt sind. Als nächstes schließt sich eine Schirmelektrode an, die in der Röhre elektrisch fest mit dem Gitter verbunden ist. Sie überdeckt alle Gebiete, die nicht leuchten dürfen, läßt also jene Gebiete frei, die z.B. den sieben Anzeigesegmenten entsprechen. Sie ist in der Regel schwarz, für eine geringe Sekundäremission mit Kohlenstoff versehen. Gleichzeitig hebt sie damit den Kontrast an. Unmittelbar unter der Schirmelektrode befinden sich die Anodensegmente. Sie sind aus Metall bzw. aus Leitpasten hergestellt. Auf ihnen befindet sich der Leuchtstoff — meist eine ZnO-Kombination mit extrem geringer Aktivierungsenergie.

Spektralverteilungen zweier oft benutzter Leuchtstoffe zeigt Bild 6.2.16d).

Die Umhüllung der Röhre ist aus Glas. Es ist innen mit einer Leitschicht, meist aus Sn_2O, versehen, die auf Katodenpotential liegt und das Innere der Röhre gegenüber äußeren Feldern abschirmt. Damit wird u. a. bewirkt, daß alle Elektronen in Richtung auf die Anode umgelekt werden (Teilbild c)).

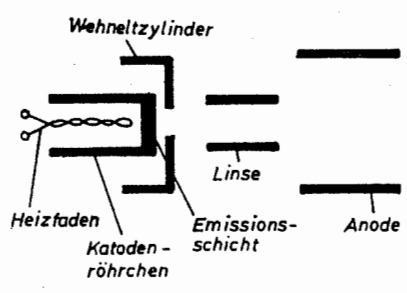
Betrieben wird die Röhre mit positivem Gitter (bei sonstigen Röhren ist es negativ) und positiver Anode. Auf das Gitter gelangen dabei etwa 10% der Elektronen, auf die Schirmelektrode 60 bis 70% und auf die Anode ca. 20 bis 30%. Die üblichen Betriebsspannungen liegen bei 20 V. Den Verlauf der Ströme (wobei Gitter- und Schirmströme nicht zu trennen sind, da beide in der Röhre zusammengeschaltet sind) und der Leuchtdichte zeigt Bild 6.2.16e). Zu sehen ist der typische Raumladeverlauf ($U^{3/2}$). Im Betrieb werden etwa Leuchtdichten von 300 bis 650 cd/m² erreicht.

Soll eine Anode nicht leuchten, so muß sie genügend negativ gemacht werden. Diese Zusammenhänge zeigt Bild 6.2.16f). Weiter ist zu beachten, daß die Ströme, die dann nicht zum Anzeigesegment gelangen können, von der Schirmelektrode übernommen werden. Sie nimmt auch alle Sekundärelektronen auf.

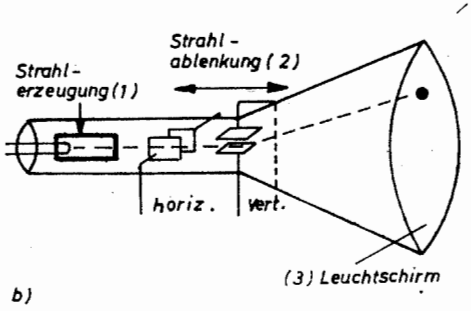
Die Röhren der dritten Generation werden in Dickschichttechnik (9.4.4.) hergestellt. Sie sind dem um 1980 vorhandenen Stand der Technik angepaßt. Beides bewirkt, daß diese Fluoreszenzröhren nur im Multiplexbetrieb verwendet werden können. Es sind daher die einzelnen Segmente aller Anoden verbunden. Bei Siebensegmentanzeigen gehen sieben Anodenanschlüsse aus der Röhre heraus. Gitter werden so viele herangeführt, wie die Röhre anzuzeigende Ziffern enthält. Es wird also das Matrix-Ansteuer-

Bild 6.2.17. Aufbau und Betrieb der Oszillographenröhre ▶

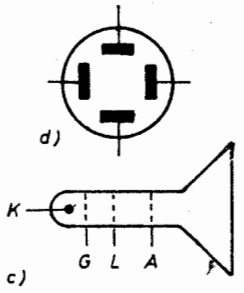
a) Strahlerzeugungssystem
b) Gesamtaufbau mit Hervorhebung des Ablenksystems
c) Vereinfachtes Schaltbild des Systems
d) Schaltbild des Ablenksystems
e) Unsymmetrischer Betrieb eines Ablenkplattenpaares
f) Stark übertriebener Trapezfehler bei unsymmetrischem Betrieb
g) Symmetrischer Betrieb eines Ablenkplattenpaares
h) Ablenkung mittels Gleich- und Wechselstrom
i) Dynamische Ablenkempfindlichkeit bei hohen Frequenzen
j) bis m) Beispiele für verschiedene Prinzipien der Nachbeschleunigung



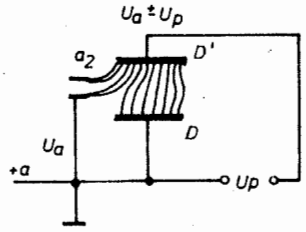
a)



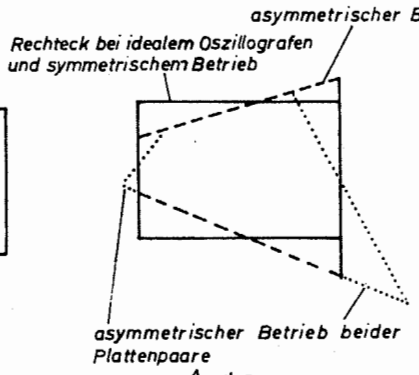
b)



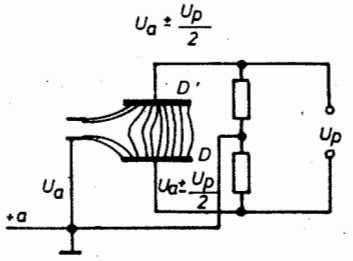
c)



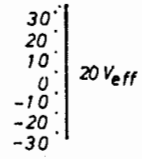
e)



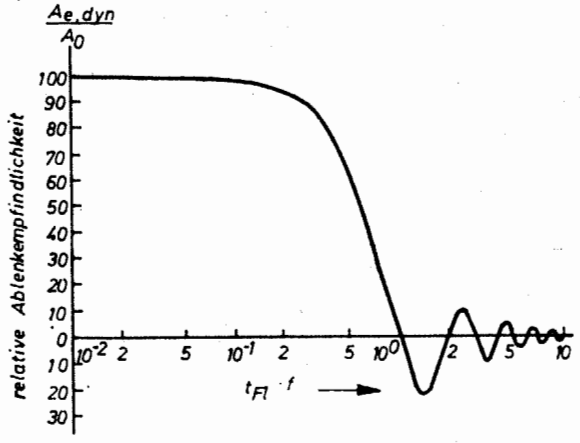
f)



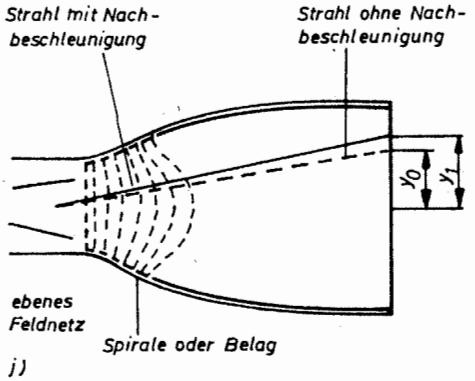
g)



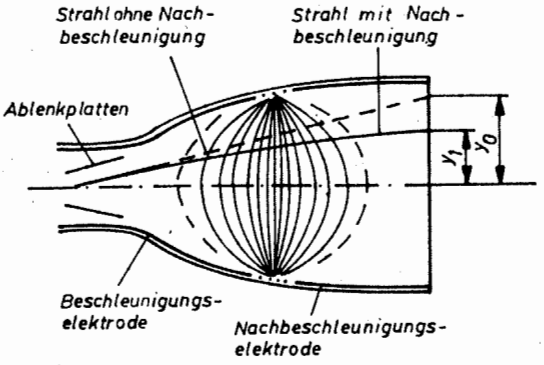
h)



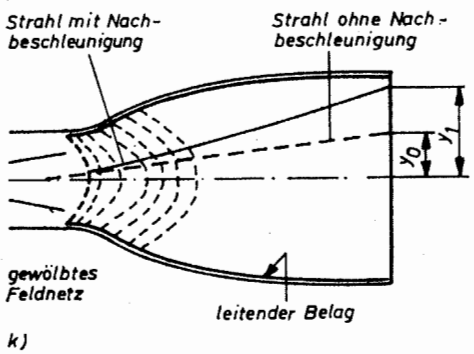
i)



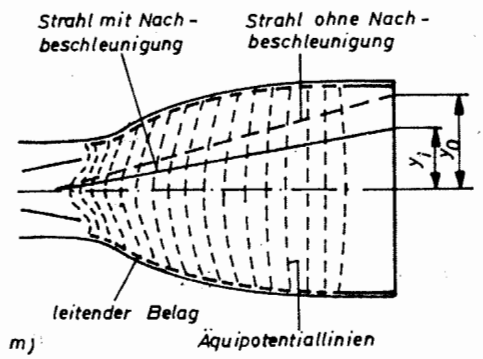
j)



l)



k)



m)

prinzip verwendet, wie es Bild 5.1.38 zeigt. Gemäß Bild 6.2.16f) müssen für das Leuchten das Gitter der jeweiligen Ziffer und die gewünschten Anodensegmente positiv sein. Für das Nichtleuchten genügt es, wenn eines von beiden etwa −2 V annimmt. Mit negativem Gitter wird darüber hinaus der gesamte Strom der entsprechenden Ziffer zu Null gemacht.

6.2.3.2. Oszillographenröhre

Neben der Bezeichnung Oszillographenröhre sind auch BRAUNsche Röhre, Elektronenstrahl- oder Katodenstrahlröhre üblich. Sie enthält gemäß Bild 6.2.17b) folgende *Teilsysteme* zur

- Erzeugung eines gebündelten Elektronenstrahls (1),
- Ablenkung des Elektronenstrahls (2) und
- Sichtbarmachung des Elektronenstrahls (3) (Leuchtschirm: Abschnitt 6.2.1.6.).

Das *Strahlungserzeugungssystem* ist detailliert in Bild 6.2.17a) dargestellt. Es besteht aus einer indirekt geheizten Katode (analog Bild 6.2.6c)) und einer Linse (analog Bild 6.2.10). Mit dem WEHNELT-Zylinder wird die Strahlstärke eingestellt, mit der Spannung an der Anode die Geschwindigkeit der Elektronen. Mit der Linsenspannung wird der Strahl auf dem Leuchtschirm fokussiert. Im Gegensatz zu optischen Linsen wird bei den elektrischen Linsen durch die Spannung die Brennweite eingestellt. Insgesamt wird für die Elektronenstrahlröhre ohne Ablenksystem ein Symbol ähnlich dem in Teilbild 6.2.17c) verwendet.

Das *Ablenksystem* besteht aus einem horizontalen und einem vertikalen Ablenkplattenpaar und wird gemäß Bild 6.2.17d) symbolisiert. Beide Ablenkplattenpaare müssen gegeneinander sehr genau senkrecht (max. Abweichung 1°) justiert sein.

Ein Teil der den WEHNELT-Zylinder passierenden Elektronen gelangt auf die Linsenelektrode bzw. die Anode. Sie führen daher Ströme, die deutlich vom Strahlstrom zu unterscheiden sind, der zum Leuchtschirm gelangt. Der *Strahlstrom* ist meist erheblich kleiner und liegt für gewöhnliche Oszillographenröhren bei einigen µA. Die Ablenkplatten werden so betrieben, daß sie keinen Strom aufnehmen. Die Ablenkung des Elektronenstrahles durch sie erfolgt rein elektrostatisch.

Die Ablenkung auf dem Leuchtschirm gehorcht Gl. (21). Sie ist also proportional der Spannung an den Platten und hängt außerdem von geometrisch konstruktiven Daten sowie von der *Beschleunigungsspannung* U_a ab. Für genaue Messungen muß sie daher stabilisiert werden. Bei den Berechnungen ist ferner ein *Planschirm* vorausgesetzt. Er ist vakuumtechnisch wegen *Implosionsgefahr* besonders bei großen Schirmen ungünstig. Infolge der Schwankungen der Beschleunigungsspannung und der maximalen Schirmgröße in bezug auf den kleinsten Durchmesser des Elektronenstrahls sind mit Oszillographenröhren nur in günstigen Fällen Genauigkeiten von etwa 1% zu erreichen, meist liegt der Wert bei etwa 10%.

Zur Unterstützung der Ablesegenauigkeit befindet sich vor dem Leuchtschirm meist ein *Raster*, das *beleuchtet* werden kann.

Die *Ablenkplatten* von Oszillographenröhren sollen *symmetrisch betrieben werden* (s. Bild 6.2.17g)). Andernfalls entsteht ein unsymmetrischer Feldverlauf, Bild 6.2.17e), und die Ablenkung wird unsymmetrisch. Dies bedingt den *Trapezfehler* (Bild 6.2.17f).

Bei der *Ablenkung* durch ein Plattenpaar sind zunächst Gleich- und *Wechselspannungen* zu unterscheiden. Gleichspannungen verlagern den Strahlpunkt gemäß Bild 6.2.17h). Eine Wechselspannung erzeugt einen Strich, der dem *Spitzen-Spitzenwert* des Wechselstromes entspricht. Wird die Frequenz stark erhöht, so kommt einmal

die Flugzeit t_{Fl} der Elektronen durch das Plattensystem in die Größenordnung der Frequenz. Zum anderen wirken auf die Elektronen beim Flug durch das Plattenpaar unterschiedlich große Spannungen und z. T. sogar Spannungen mit entgegengesetztem Vorzeichen ein. Dies bewirkt, daß die *wirksame Ablenkempfindlichkeit* abnimmt, durch Null geht und sich umkehrt usw. Dies zeigt Bild 6.2.17i). Für sehr hohe Frequenzen (meist über 100 MHz) müssen deshalb Sondermaßnahmen getroffen werden. Eine Möglichkeit ist die *Nachbeschleunigung*. Es wird mit relativ kleiner Anodenspannung gearbeitet. Die Ablenkung kann mit sehr kurzen Ablenkelektroden erfolgen, und die notwendige Helligkeit wird durch Beschleunigung der Elektronen nach der Ablenkung bewirkt. Mit dieser Nachbeschleunigung gelingt es unter speziellen Bedingungen sogar, die Ablenkung zu vergrößern. Beispiele hierzu gibt Bild 6.2.17j) bis m).

Eine andere Methode ist die Ablenkung ähnlich einer *Traveling-Wave-Röhre*. Die Ablenkplattenpaare werden dann in kleine Teile zerlegt, die genauso verzögert erregt werden, wie die Elektronen sich fortbewegen.

Für hohe Frequenzen bzw. schnelle Vorgänge ist die *maximale Schreibgeschwindigkeit* einer Oszillographenröhre wichtig. Sie gibt an, wie schnell der Strahl über den Leuchtschirm bewegt werden kann, damit gerade noch eine sichtbare Spur zurückbleibt. Für Sonderzwecke werden zur Erhöhung der Schreibgeschwindigkeit die im Zusammenhang mit Bild 6.2.13 beschriebenen Mikrokanäle verwendet.

6.2.3.3. Sichtspeicherröhre

Die Sichtspeicherröhre wird verwendet, um einmalige Vorgänge für längere Zeit stehend auf dem Schirm sichtbar zu machen. Dies ermöglicht ein spezielles *Speichergitter* in der Röhre. Es besteht aus einem feinmaschigen Netz, auf dem eine dielektrische Schicht aufgebracht ist, welche imstande ist, Ladungen anzunehmen und über Stunden beizubehalten. Die Ladungen werden durch Elektronenbeschuß und definierte Sekundärelektronenemission erzeugt.

Die Sichtspeicherröhre (Bild 6.2.18a)) enthält *ein übliches Katodenstrahlsystem* mit Ablenkung (Index s von schreiben) *und ein zweites* (oder mehrere solche) mit dem

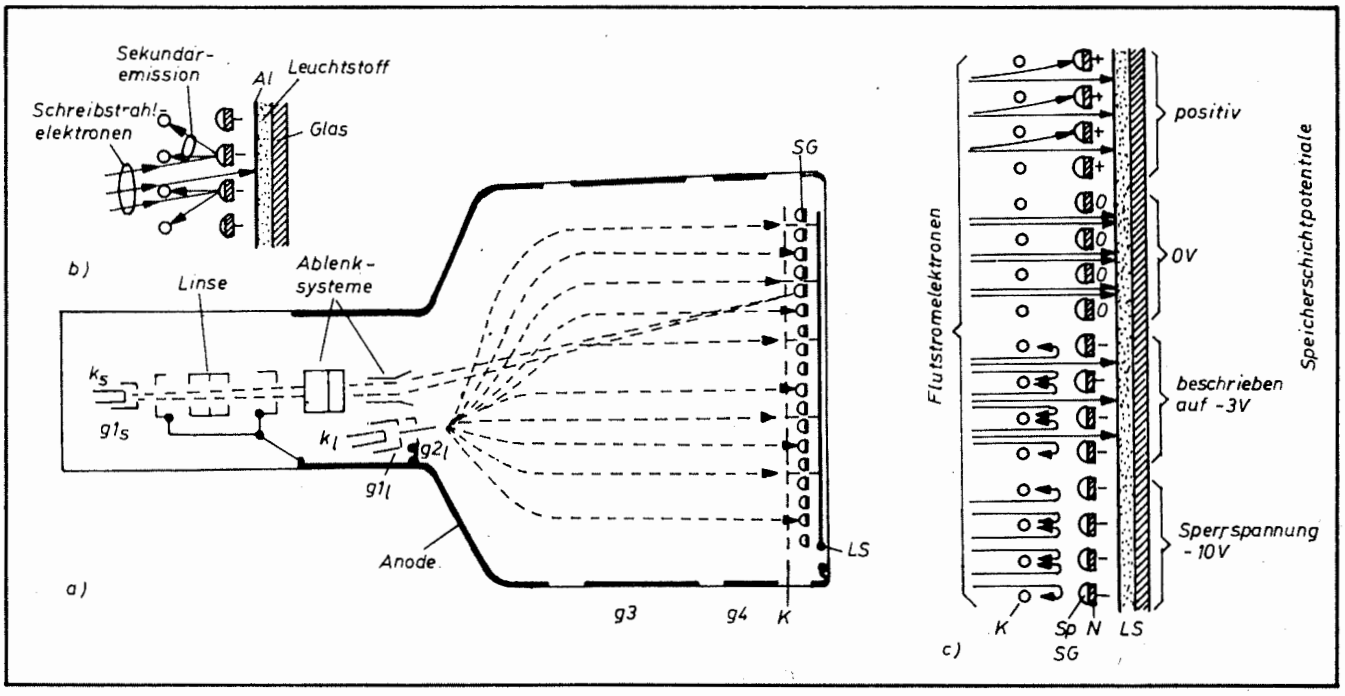


Bild 6.2.18. Prinzipieller Aufbau und Funktion einer Sichtspeicherröhre

Index l (lesen), welches mit langsamen Elektronen arbeitet und gleichmäßig den ganzen Leuchtschirm berieseln kann. Dies wird durch die Anschlüsse g_3 und g_4 bewirkt.

Die *Wirkungsmechanismen* in der Sichtspeicherröhre sind recht kompliziert. Für Details sei daher auf [R 12, Bd. 4] verwiesen. Entscheidend ist, daß dem Speichergitter durch den Schreibstrahl positive oder negative Ladungen aufgeprägt werden können, und zwar je nach den Potentialverhältnissen, die beim Schreiben vorhanden sind (Bild 6.2.18b)). Es können dadurch so viele Sekundärelektronen entstehen, daß sich das Speichergitter positiv auflädt. Es können aber auch weniger entstehen oder gar keine, wenn das Potential so hoch ist, daß sie tief eindringen und nicht mehr frei werden können. Dann lädt sich das Speichergitter negativ auf. Bei der *Wiedergabe* können die *langsamen Flutelektronen* den Leuchtschirm nur dort erreichen, wo beim Schreiben keine zu negative Spannung verblieben war (Bild 6.2.18c)).

6.2.3.4. Bildröhre

Die Bildröhre besitzt im Gegensatz zur Oszillographenröhre eine *magnetische Ablenkung*. Dadurch ergeben sich Besonderheiten, die aus Tab. 6.2.4 hervorgehen. Sie ist immer dort vorteilhaft einsetzbar, wo mit fest vorgegebenen Ablenkfrequenzen gearbeitet wird, wo also insbesondere ähnlich wie beim Fernsehen die gesamte Bildfläche zeilenweise gleichmäßig überstrichen wird.

Tabelle 6.2.4. Vergleich von Bildröhren mit magnetischer und elektrischer Ablenkung

Ablenkung	El.-magn.	El.-statisch
Innerer Aufbau	einfach	komplex
Rohrlänge	kurz	lang
Punkthelligkeit	hoch	mäßig
Ablenkempfindlichkeit	hoch	niedrig
Verstärkeraufbau	komplex	einfach
Ablenkgeschwindigkeit	niedrig	hoch
Gewicht	schwer	leicht
Fokussierung	gut über die Schirmfläche	Ecken unscharf

Die Bildröhre mit magnetischem Ablenksystem (Bild 6.2.19) besitzt ein *Strahlerzeugungssystem* im relativ dünnen Hals, welches dem der Oszillografenröhre entspricht. Um diesen Hals herum und bis zur Ausweitung des Halses zum Röhrenkolben schmiegen sich eng die Ablenkspulen an. Aus dem Bild ist ersichtlich, daß im Gegensatz zur Oszillographenröhre der abgelenkte Strahl einen weitaus längeren Weg als der unabgelenkte zum Schirm zurückzulegen hat. Dies erfordert besondere Maßnahmen zur *Bildkorrektur*, auf die hier nicht eingegangen werden kann.

Neben der Anwendung im Fernsehen hat die Bildröhre in den letzten Jahren große Bedeutung *als universelles Display*, insbesondere in der Rechentechnik bis hin zu den Mikro- und Personal-Computern, erreicht. International lag 1980 der Einsatz von Bildschirmarbeitsplätzen in Verwaltung und Wissenschaft bei 2 bis 15%. Diese Anwendungsbreite hat zur Weiterentwicklung der Bildröhre des Fernsehens geführt. Sind bei der Fernsehnorm 625 Zeilen, ca. 300 × 500 Bildpunkte, bei einer Videobandbreite von 5 MHz zur Helligkeitssteuerung gegeben, so werden heute Grenzwerte bis zu 5000 × 5000 Bildpunkte und Bandbreiten bis zu 100 MHz verwendet. Damit lassen sich z. B. ganze A4-Seiten mit voller Auflösung darstellen. Die Schirmauflösung ist

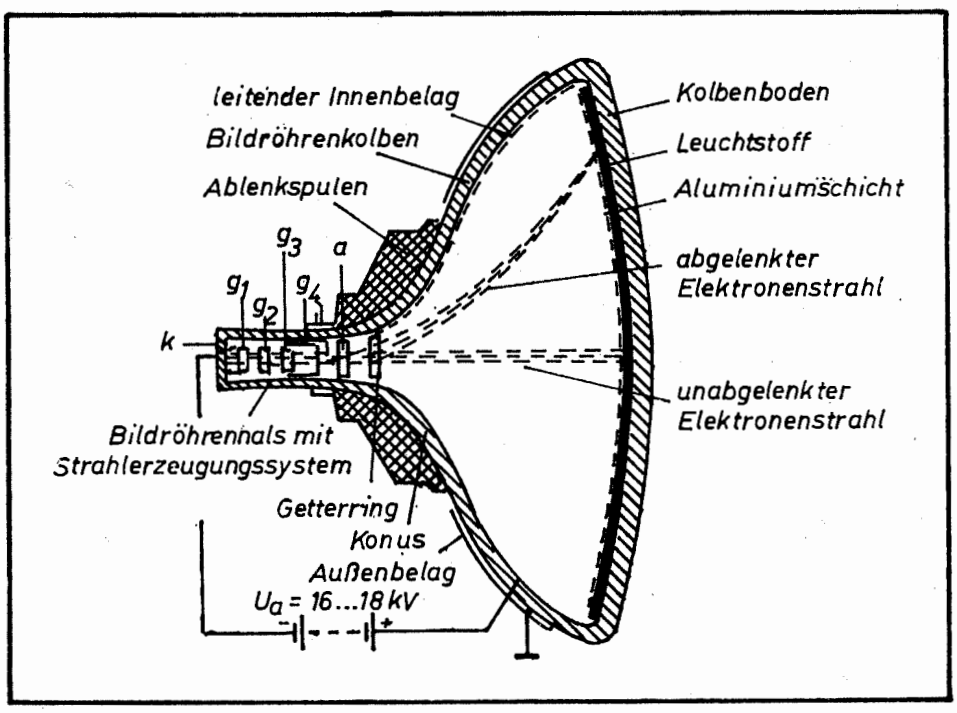


Bild 6.2.19. Aufbauprinzip einer Fernsehbildröhre

bei den Bildröhren nahezu von der Bildfläche unabhängig, da der erreichbare Fleckdurchmesser etwa proportional der Bildfläche ist. Es werden etwa 10 µm bei Schirmen von 10 cm Durchmesser erreicht.

Die große *Implosionsgefahr* von Bildröhren ist durch den speziellen Aufbau von Halterungen und die Gestaltung des Kolbens nicht mehr bedeutend.

Sonderausführungen der magnetischen Bildröhre werden in der *Radartechnik* verwendet. Hier erfolgt eine kreisförmige Ablenkung des Elektronenstrahls. Der Schirm ist mit einem Leuchtstoff von hoher Nachleuchtdauer versehen.

Weiter hat das Prinzip der Bildröhre viele *Sonderanwendungen* erfahren. Hier seien nur genannt: zur Darstellung bestimmter, fest definierter Zeichen, zur Herstellung von hochwertigen Druckerzeugnissen.

Der Bildröhre als Display gereicht es immer zum Nachteil, daß sie relativ lang und, insbesondere bei magnetischer Ablenkung, schwer ist. Dies hat zu den verschiedensten Versuchen geführt, eine *flache Variante* zu entwickeln. Hierbei entstehen jedoch vielfältige Probleme, z.B. Flächenkatode mit gleichmäßiger Emission über die gesamte Fläche, komplizierte Steuermechanismen bzw. Umlenkmechanismen für die Elektronenstrahlen. Obwohl es verschiedene Lösungen gibt (insbesondere für kleine Bildschirme), hat sich bisher keine durchsetzen können. Es werden vielmehr andere Varianten, z.B. auf der Basis von LCD, LEF (6.1.6. und 6.1.7.) bzw. Plasmapanal (6.2.4.) versucht. Dennoch wird allgemein eingeschätzt, daß eines Tages die heutige Bildröhre zumindest teilweise durch ein flaches Bildprinzip verdrängt wird.

6.2.3.5. Farbbildröhre

Für moderne, hochwertige Displays und für das Fernsehen ist in vielen Fällen die Farbe unverzichtbar geworden. Werden durch sie doch weitaus bessere Darstellungen möglich. Gemäß Bild 6.1.4 kommt es bei der Bildröhre darauf an, die drei Grundfarben rot (R), grün (G) und blau (B) in einem einstellbaren Verhältnis für jeden Bildpunkt zu erzeugen. Bis etwa Mitte der siebziger Jahre dominierte hier die *Lochmaskenröhre*, während z. Z. die *Schlitzmaskenröhre* die größte Anwendungsbreite besitzt. Beide Röhren beruhen auf dem gleichen Prinzip: Von drei in der Strahlstärke einzeln

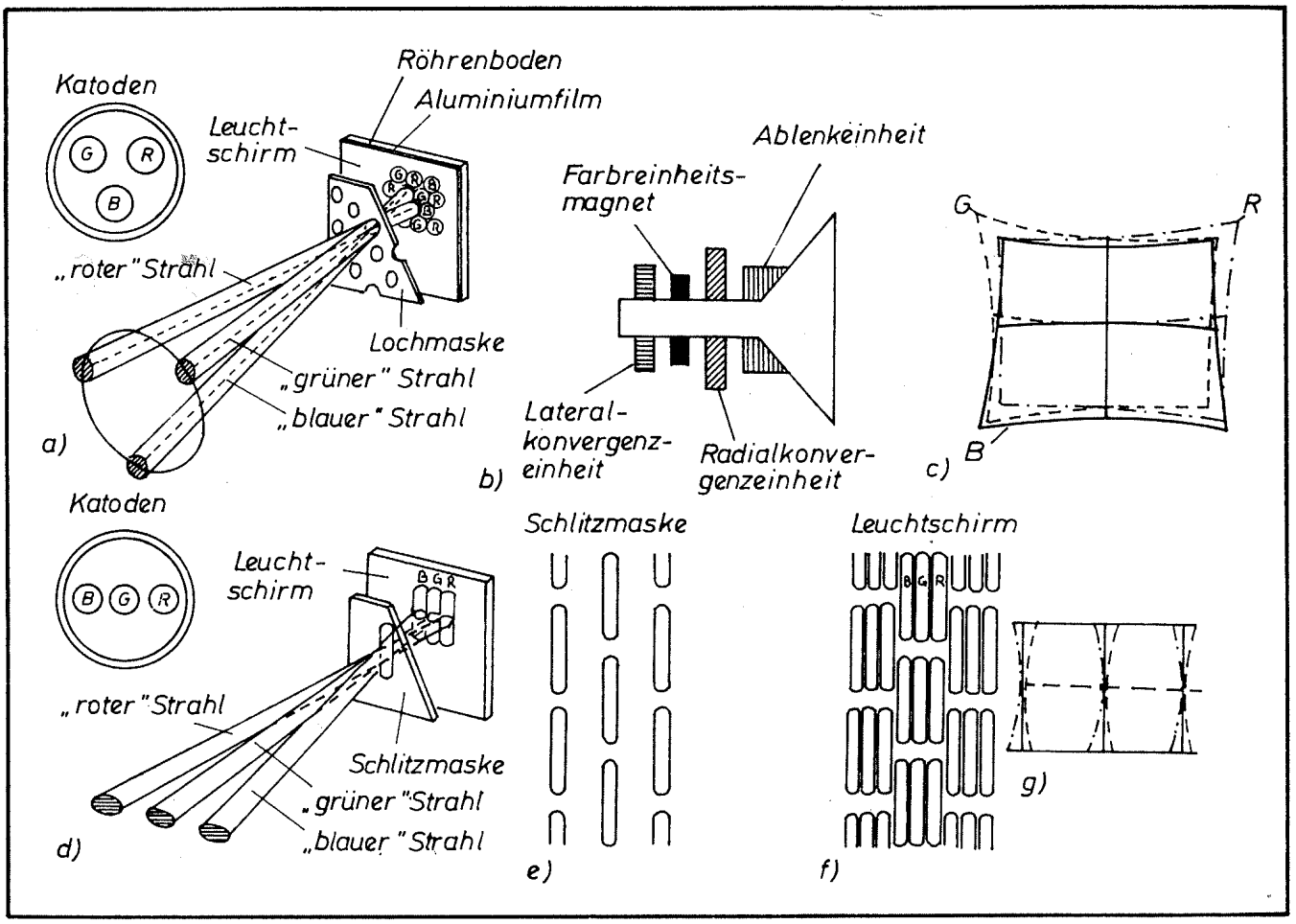


Bild 6.2.20. Farbbildröhren

a) Prinzip der Lochbildröhre
b) Ablenkeinheit und Konvergenzmagnete bei Lochbildröhre
c) Statische Verzeichnungen bei der Lochbildröhre
d) Prinzip der Schlitzmasken (In-Line-) Farbbildröhre
e) Schlitzmaske
f) Anordnung der Leuchtstoffe
g) Statische Verzeichnung bei der Schlitzmaskenröhre

gesteuerten *Elektronenstrahlsystemen*, die sich im Röhrenhals befinden, gehen drei Elektronenstrahlen, die gemeinsam abgelenkt werden, aus. Sie treffen wenige mm vor dem Bildschirm auf eine *Schattenmaske* und erreichen dahinter die den einzelnen Strahlen zugeordneten Leuchtstoffe R—G—B. Damit besteht jeder Bildpunkt — dafür gibt es auf der Maske ca. 300000 Löcher bzw. Schlitze — aus je drei beieinanderliegenden Farbpunkten. Eine Farbbildröhre ist also von höchster Präzision: Maske und Bildpunkte müssen, von den Elektronenstrahlsystemen aus gesehen, sich exakt ohne Farbfehler decken. Hierbei können *Temperaturänderungen* der Maske, *magnetische Streufelder*, ja selbst das Erdfeld Störungen hervorrufen. Deshalb wird die Maske z.B. von genau ausgelegten Bimetallstreifen getragen. Im Betrieb nimmt die Maske nämlich eine Elektronenstrahlenergie von ca 10 Watt auf. Ferner werden insbesondere die Masken bei jedem neuen Einschalten *automatisch entmagnetisiert*. Außerdem muß eine komplizierte statische und dynamische Korrektur aller Fehler vorgenommen werden. Dies deutet Bild 6.2.20b) und c) für die Lochmaske an.

Die *Schlitzmaske* (Teilbilder d) bis g)) hat den Vorteil, daß die drei Strahlen in einer Linie (*In-line*) liegen. Dadurch sind von vornherein alle *Fehler geringer*. Außerdem hat sie Vorteile bezüglich der Dichte der Leuchtstoffbelegung auf dem Schirm und der *höheren Durchlässigkeit* der Maske. Deshalb ist das Bild der Schlitzmaskenröhre stets brillianter, heller.

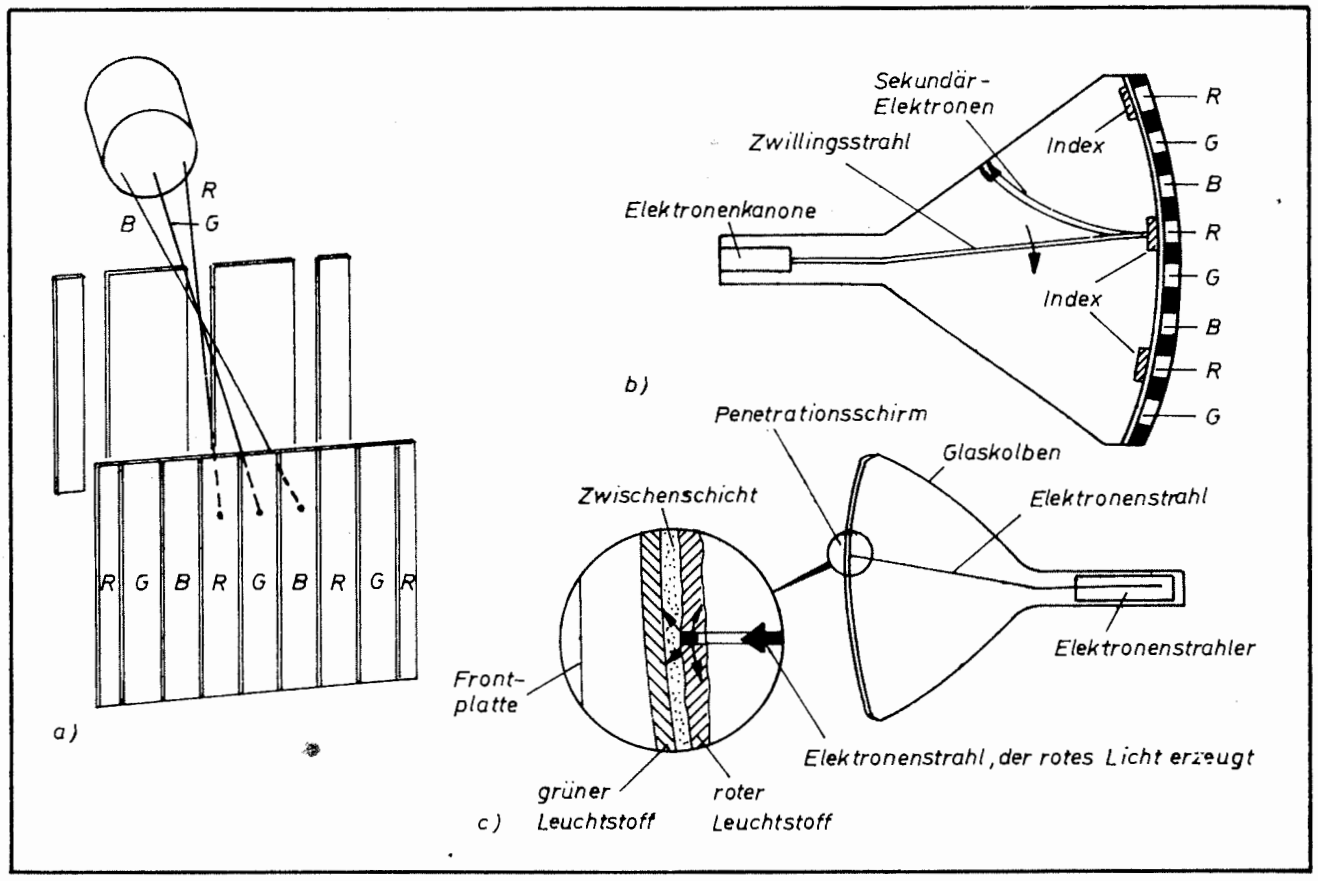


Bild 6.2.21. Weitere Farbbildröhren

a) In-Line-Gittermaskenröhre (Trinitron)
b) Indexröhre: im Beispiel über Sekundärelektronen (es gibt sie auch mit anderen Empfängern)
c) Penetrationsröhre mit Aufbau des Schirms

Neben den beiden geschilderten Prinzipien hatte zumindest eine Zeitlang die japanische *Trinitronröhre*, auch Gittermaskenröhre genannt, Bedeutung (Bild 6.2.21a)). Sie gibt im Prinzip die größte Helligkeit, verlor jedoch infolge internationaler Konzernbestrebungen und eventueller Mängel bezüglich der mechanischen Schwingstabilität der Gitter an Bedeutung.

Ein anderes Prinzip, das seit 1956 existiert, 1961 und dann erneut in den letzten Jahren wieder auftauchte, ist die *Indexröhre* (Bild 6.2.21b)). Sie hat den großen Vorteil, mit nur einem Elektronenstrahl zu arbeiten. Er überstreicht bei seiner Ablenkung nacheinander die drei Farben, welche ähnlich wie beim Trinitron streifenförmig angeordnet sind. Die Strahlstärke wird dabei je nach der erreichten Farbe gesteuert, also mit der dreifachen Frequenz wie bei den Maskenröhren. Um den Ort des Strahls festzustellen, existiert für jedes Farbtripel ein Indexstreifen, der z.B. über Sekundärelektronen, UV oder über eine Leuchtstoffarbe automatisch erkannt wird. Die Vorteile dieses Prinzips sind unmittelbar einzusehen, jedoch bestehen offensichtlich technologische Probleme.

Als letztes Prinzip sei die *Penetrations-Farbbildröhre* (Bild 6.2.21c)) erwähnt. Sie verwendet einen Leuchtschirm, bei dem zwei (oder mehrere) Farben übereinander aufgetragen sind. Je nach der Geschwindigkeit des Elektronenstrahls wird das Verhältnis der Anregung der Leuchtstoffe geändert. Bei einer französischen Röhre mit zwei Leuchtstoffen entsteht z.B. bei 8,5 kV rotes, bei 9,2 kV orangefarbiges, bei 10 kV gelbes und bei 11,5 kV grünes Licht. Mit solchen Röhren ist also nicht das gesamte Farbspektrum zu verwirklichen. Sie besitzen aber ein erheblich höheres Auflösungsvermögen als die zuvor genannten Farbbildröhren. Bei den Maskenröhren werden z. Z. maximal 1000 × 1000 Farbpunkte erreicht.

In vielen technischen Anwendungen werden Farbbildröhren mit den drei Signalen $R\,G\,B$ angesteuert. Dies bedeutet bei der Fernsehnorm eine Bandbreite von 3×5 MHz. Bereits 1947 wies HARTRIDGE nach, daß unser Auge bei schnellen Änderungen zunächst nur dem Helligkeitswert (Luminanz) Y folgen kann. Es entsteht kein sichtbarer Verlust, wenn die Farbsignale $R - Y$ und $B - Y$ statt mit 5 MHz nur mit je 2 MHz übertragen werden (Bild 6.2.22a)).

Durch eine weitere Verarbeitung der Signale kann die so schon reduzierte Bandbreite noch einmal eingeengt werden. Dies wiesen MIDDLETON und HOLMES 1949 nach. Es müssen dann die Signale

$$\left.\begin{aligned} I &= 0{,}877(R - Y)\cos 33° + 0{,}493(B - Y)\sin 33°\,, \\ Q &= 0{,}877(R - Y)\sin 33° + 0{,}493(B - Y)\cos 33° \end{aligned}\right\} \tag{24}$$

gebildet werden. Für diese Signale genügen die Bandbreiten gemäß Bild 6.2.22b)

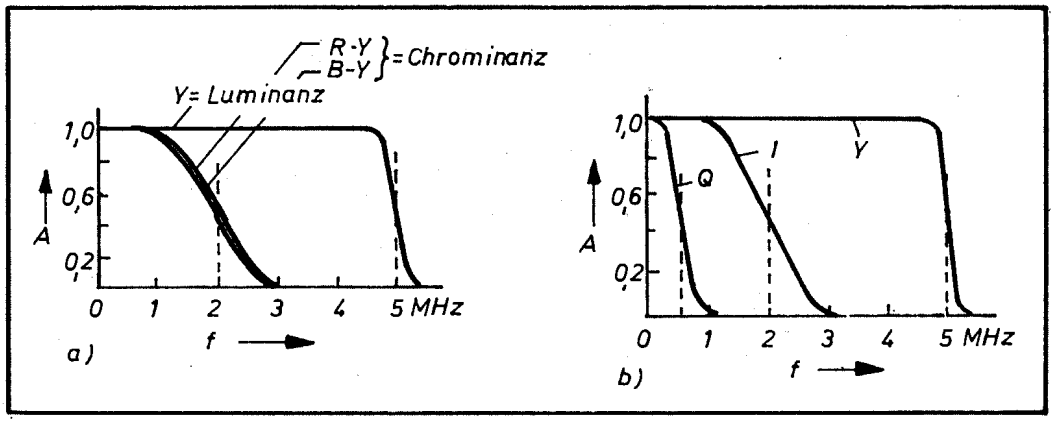


Bild 6.2.22. Abnahme der subjektiv wahrnehmbaren Bandbreite bei verschiedener Matrizierung der RGB-Signale

a) Aufbereitung bei SECAM und PAL
b) bei NTSC

6.2.4. Plasmabauelemente

6.2.4.1. Grundlagen

Gase sind unter normalen Bedingungen Nichtleiter. Erst wenn die *Gasmoleküle* ionisiert sind und/oder freie Elektronen enthalten, können sie Strom transportieren, d. h. leiten. In geringem Umfang sind in jedem Gas immer eine gewisse Anzahl von Ladungsträgern (Ionen, Elektronen) vorhanden. Sie werden durch die Höhenstrahlung, Lichteinstrahlung und andere vorhandene Energien erzeugt.

Eine Gasentladungsröhre ist ein hermetisch abgeschlossenes Gefäß, in dem zumindest zwei Elektroden, Anode und Katode, eingelassen sind und in der sich ein Gas befindet. Legt man an die Elektroden (über einen Vorwiderstand) eine Spannung von ca. 500 Volt an und *verändert den Gasdruck* in der Röhre, so geschieht folgendes: Bei einer Verdünnung von ca. 1 Pa ($\approx 7{,}5 \cdot 19^{-3}$ Torr) zeigt sich in der Röhre eine Gasentladung, wie sie etwa Bild 6.2.23 zeigt. Wird der Druck weiter abgesenkt, so erlischt die Gasentladung bei ca. 0,01 Pa wieder. Mit dieser Druckänderung variiert das Erscheinungsbild der *positiven Säule*. Sie zeigt z. T. eine geschichtete Struktur, wobei der Abstand der Schichten in etwa der gerade vorhandenen freien Weglänge l der Moleküle entspricht:

$$l/\mathrm{cm} \approx \frac{1}{p/\mathrm{Pa}}\,. \tag{25}$$

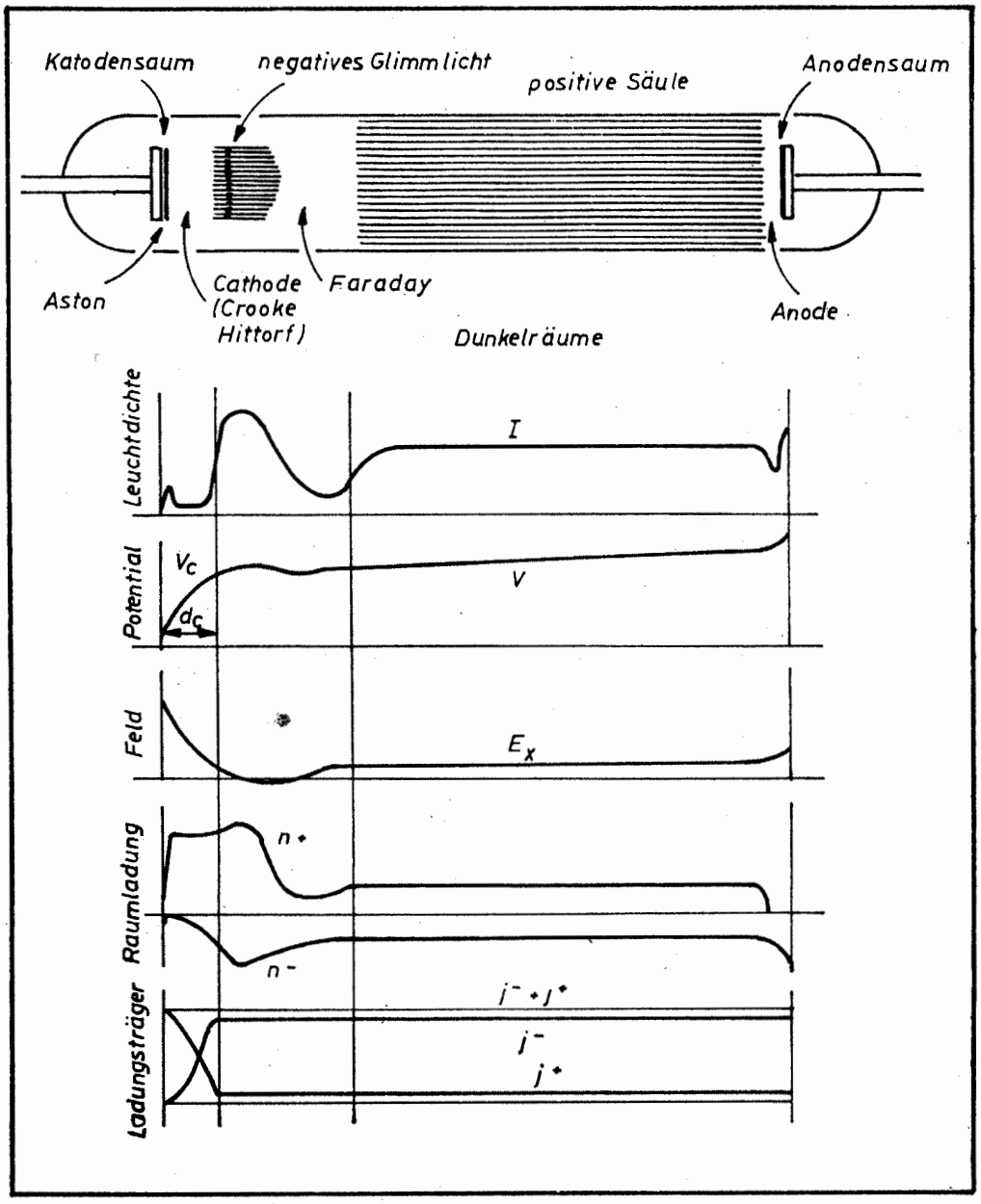


Bild 6.2.23. Prinzipieller Verlauf von Lichtern und Dunkelräumen sowie anderer Parameter bei einer langen Gasentladungsröhre

Diese Länge benötigen die Moleküle, um gerade so viel Energie aufzunehmen, daß sie beim Auftreffen auf das nächste Molekül dieses ionisieren können. Die dabei entstehenden Ladungsträger werden im Feld erneut beschleunigt, um nach der freien Wegstrecke wieder ein Molekül ionisieren zu können. Das leuchten der Säule kommt durch Rekombination von Ladungsträgern zustande. Der eben beschriebene Prozeß der Trägererzeugung heißt *Stoßionisation.*

Wesentlich für eine Gasentladung ist das *negative Glimmlicht* und der CROOKEsche oder HITTORFsche *Dunkelraum.* Hier existiert die größte Feldstärke, wie Bild 6.2.23 zeigt. In dem Dunkelraum werden positive Ionen beschleunigt (Katodenfall) und prallen mit so hoher Energie auf die Katode, daß sie Elektronen aus der Katode herausschlagen (Stoßemission). Die Elektronen sind entscheidend für die Aufrechterhaltung der Gasentladung. An der Katode muß also die *Austrittsarbeit* für die Elektronen *durch mechanischen Stoß* aufgebracht werden.

Im FARADAY-*Dunkelraum* und bei der geschichteten Säule ist nur die Ionisierungsenergie der Moleküle (Tab. 6.2.5) zu erreichen. Die Länge der positiven Säule (einschließlich ihrer eventuellen Schichtung) kann mit der Länge der Gasentladungsröhre

Tabelle 6.2.5. Ionisierungsspannung der chemischen Elemente und einige Farben bei Gasentladungen

Chem. Element	eV	Chem. Element	eV	Chem. Element	eV
Kalium	4,32	Nickel	7,60	Quecksilber (purpur)	10,41
Natrium (gelb)	5,09	Magnesium	7,63	Schwefel	10,42
Barium	5,21	Kupfer	7,67	Kohlenstoff	11,24
Lithium	5,37	Kobalt	7,80	Chlor	13,01
Gallium	5,97	Eisen	7,83	Wasserstoff	13,54
Aluminium	5,98	Germanium	8,10	Sauerstoff	13,57
Calcium	6,25	Silicium	8,14	Krypton	13,94
Chrom	6,70	Platin	8,80	Stickstoff	14,51
Vanadium	6,74	Cadmium	8,94	Argon (blau)	15,68
Titan	6,81	Gold	9,20	Neon (orange)	21,47
Blei	7,37	Zink	9,37	Helium (gelb)	24,48
Mangan	7,41	Selen	9,75		
Silber	7,58	Arsen	10,05		

verändert, ja sogar bei kurzen Röhren unterdrückt werden. Das *negative Glimmlicht* und damit der *Katodenfall* sind unabdingbar für jede Gasentladung. Die positive Säule hat praktische Bedeutung bei den *Leuchtstoffröhren* und den *farbigen Röhren* der Werbung. Das negative Glimmlicht ist für die folgenden Anwendungen wichtig, bei ihnen ist jedoch die positive Säule unterdrückt.

Die *Kennlinie einer Gasentladungsröhre* ist relativ kompliziert (Bild 6.2.24). Sie enthält mehrere Äste mit negativem Verlauf (fallende Spannung bei steigendem Strom) und wird daher am besten *als Funktion des Stromes* betrachtet und gemessen.

Zunächst steigt der Strom mit der Spannung an (A). In diesem Gebiet werden die immer vorhandenen Ladungsträger abgesaugt. Dadurch tritt ein großer Spannungsbereich (B) mit dem konstanten Strom dieser Ladungsträger auf. Dieses Gebiet wird

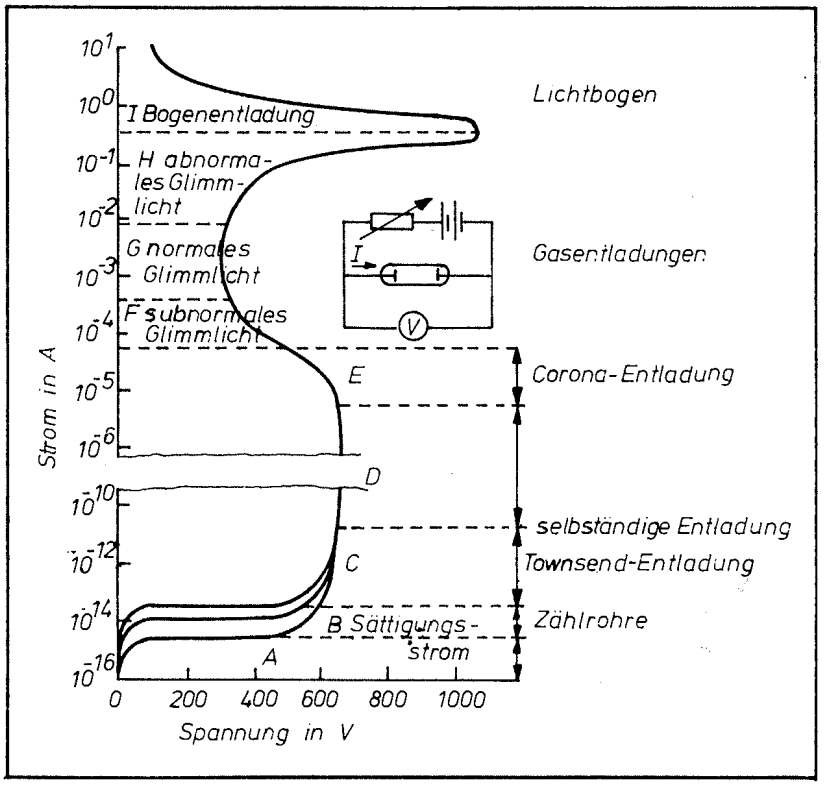


Bild 6.2.24. Prinzipieller Verlauf der Stromcharakteristik einer Gasentladungsstrecke

für *Zählrohre* genutzt, da es die Ladungsträger registriert, die im Gasraum auftreten. Deshalb die drei unterschiedlichen Höhen im Bild. Wird die Spannung über eine gewisse Grenze erhöht, so können sich die einzelnen *Ladungsträger durch Stoßionisation vervielfachen.* Es folgt also ein steiler Stromanstieg (*C*), die TOWNSEND-*Entladung.* Sie bricht jedoch zusammen, sobald die äußeren Ursachen für die Erzeugung von Ladungsträgern fehlen. Deshalb wird auch von *unselbständiger Entladung* gesprochen. Es folgt dann ein Gebiet, das nur bei Strombegrenzung existiert (*D*) und das in die *Corona-Entladung* (*E*) übergeht. Hier erfolgt eine *Emission an einzelnen Spitzen* der Katode (Spitzenentladung), und es bricht die Spannung an den Elektroden zusammen. Für höheren Strom (kleineren Außenwiderstand) folgt das instabile Gebiet der *subnormalen Glimmentladung* (*F*), wo erstmals das *negative Glimmlicht* (evtl. instabil auf der Katode umherspringend) auftritt. Während der *normalen Glimmentladung* (*G*) existiert eine *konstante Spannung,* die von Katode, Gasfüllung und Gasdruck abhängt und bei ca. 70 bis 300 V liegen kann. Sie wurde früher für *Glimmstabilisatoren* verwendet. Für dieses Gebiet gelten die bei Bild 6.2.23 beschriebenen Verhältnisse. In diesem Gebiet ändert sich die *Bedeckung der Katode* mit dem negativen Glimmlicht. Ist die Katode bei steigendem Strom voll mit dem Glimmlicht bedeckt, beginnt die *anormale Glimmentladung* (*H*). Die Spannung steigt wieder an, die Intensität des negativen Glimmlichtes wächst ebenfalls. Die *Ladungsträgerzahl steigt* in diesem Gebiet stark an. Die *Katodentemperatur wächst* infolge Erwärmung. Dies geht so weit, bis eine thermische und/oder optisch induzierte *Elektronenemission* an der Katode auftritt. Dann bricht der Katodenfall zusammen (er ist infolge der neuen Emissionsart nicht mehr notwendig). Die Gasentladung geht in die *Bogenentladung* (*I*) über, die u. a. beim *Kohlelichtbogen* angewendet wird. Die Gefahr des Lichtbogens ist der Hauptgrund, weshalb alle Gasentladungen mit einem Vorwiderstand bzw. bei Leuchtstoffröhren mit einer *Strombegrenzung* durch eine Drossel betrieben werden müssen.

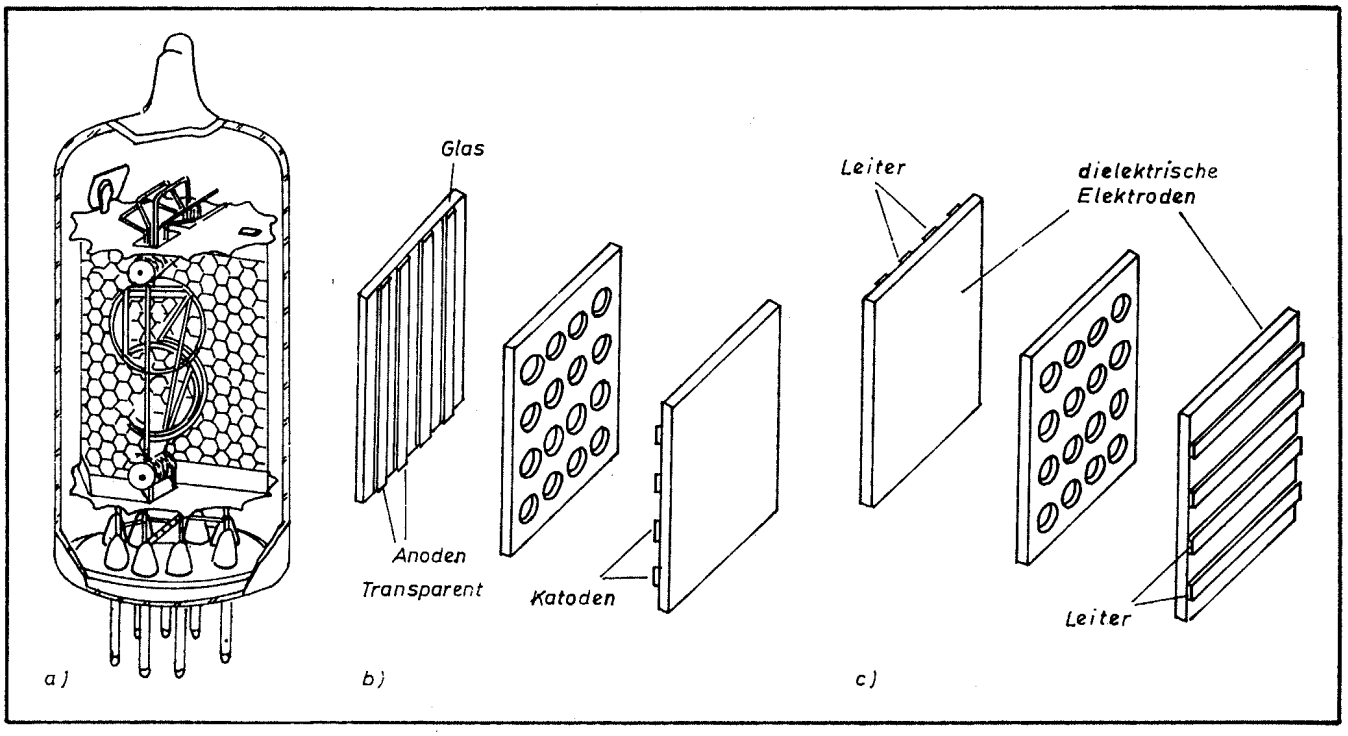


Bild 6.2.25. Gasentladungsanzeige

a) Ziffernanzeigeröhre
b) Plasmapanel für Gleichstrom- und
c) Plasmapanel für Wechselstrombetrieb

6.2.4.2. Ziffernanzeigeröhren und Plasmapanels

Das Prinzip der Ziffernanzeigeröhre zeigt Bild 6.2.25a), Die einzelnen Katoden befinden sich als gebogene Drähte in Form der einzelnen Ziffern vor einer Anode (Netz). Wird einer der Drähte als Katode gegenüber der Anode betrieben, so umgibt ihn das *negative Glimmlicht*, und es leuchtet nur diese Ziffer auf. Dieses Prinzip wurde ursprünglich als NIXIE-*Röhre* eingeführt, hat aber heute wegen der relativ hohen Spannungen (ca. 150 V) kaum noch Bedeutung.

Eine gewisse Chance, als flacher Bildschirm Breitenwirkung zu erreichen, besitzen die *Plasmapanels* (Plasmadisplays). Sie existieren in zwei Formen (Bild 6.2.25b) und c)), für Gleichstrom und Wechselstrom. Sie bestehen aus einer Lochscheibe von ca. 50 bis 100 µm Dicke, in der die einzelnen Gasentladungen brennen können. Sie werden vakuumdicht durch die beiden Deckscheiben mit den Streifenelektroden abgeschlossen. Die einzelnen Miniaturröhrchen werden nach dem *Matrixprinzip* angesteuert. Bei der Gleichstromvariante besteht die Gefahr, daß die Elektroden durch das ständige Ionenbombardement verdampfen. Dies wird bei der Wechselstromvariante vermieden. Sie hat jedoch die Nachteile der komplizierten Ansteuerung und der geringeren Leuchtdichte von z.B. 3 cd/m² statt bei Gleichstrom von ca. 50 cd/m².

Ähnlich wie bei Leuchtstofflampen werden auch bereits Leuchtstoffe eingesetzt, um unterschiedliche Farben (je Röhrchen eine Farbe) darzustellen. Obwohl erste Plasmapanels seit ca. 1970 existieren und heute (1982) vielfältige Plasmapanels in Betrieb sind, ist der Durchbruch noch nicht erfolgt. Das Gebiet ist noch stark in Entwicklung.

6.3. Mechano- und Akustoelektronik

6.3.1. Überblick

Die Mechanik ist in den letzten Jahren in den Ruf gekommen, sehr teuer, technologisch aufwendig, langsam und schwer handhabbar zu sein. Sie gilt deshalb als unmodern, und man ist bemüht, sie soweit wie irgend möglich auszuschalten. Diese Tendenz hat das Problem mit sich gebracht, daß mechanische Geräte und Geräteteile immer schlechter beherrscht werden. So sind u. a. hierdurch die peripheren Geräte der elektronischen Datenverarbeitung zu einem leistungsbegrenzenden Faktor geworden. Obwohl z.B. Tasten und Drucker wohl fast immer einen mechanischen Teil enthalten werden.

Wie bedeutsam die Mechanik auch heute noch und künftig ist, geht z.B. aus Bild 0.1 hervor, wonach die Summe von Kontaktelementen, Schaltern und Relais nicht nur absolut, sondern sogar relativ ansteigt. Dabei sind in dieser Darstellung nicht einmal solche Bauelemente wie Mikrophone, Tonabnehmer, Lautsprecher, Motoren, Drucker und mechanische Anzeigen enthalten. Die Entwicklung der Elektronik hat natürlich auf diese Bauelemente ebenfalls einen großen Einfluß ausgeübt. Dieser läßt sich relativ deutlich bei den Motoren zeigen, wo sich u. a. elektronisch kommutierte Motoren, Motoren mit Rotoren in quasi gedruckter Technik, Mikromotoren — also Motoren in völlig neuer Qualität herausbildeten. Die Entwicklung zeigt sich z.B. extrem bei den digital angesteuerten Schrittmotoren der Armbanduhren, aber auch bei den Spezialmotoren in den kleinsten und höchstpräzisen Magnetbandgeräten. Dennoch ist die Tendenz nicht zu verkennen, daß immer mehr (nicht funktional bedingte) Mechaniklösungen durch Elektronik ersetzt werden. Dies ist gerade bei der Armbanduhr deutlich, aber auch Schreibmaschine, Fernschreiber und Drucker zeigen diesen Trend. Die Mechanik wird in diesen und anderen Fällen immer mehr auf das funktionell unbedingt Notwendige zurückgedrängt, und gleichzeitig werden dabei für diesen „Mechanik-Rest" oft eine höhere Präzision und speziell neue Verfahren notwendig. Andererseits entstehen im Bereich der Mechanik, genauer der Schwingungstechnik bzw. Elektroakustik, auch völlig neue Gebiete, u. a. Akustoelektronik. Beispiele hierzu sind die mechanischen Filter und noch mehr die Oberflächenwellen-Bauelemente, die eigentlich erst im Kommen sind und neue elektronische

Tabelle 6.3.1. Geschichtsdaten zur Mechanoelektronik

50	PLINIUS schreibt vom Hirten MAGNES, dessen eiserner Stab von einem speziellen Stein angezogen worden sei
um 1100	Seefahrer und Bergleute verwenden Kompaßnadeln aus Magneteisenstein (Vorkommen in Magnesia — Griechenland)
1819	OERSTED findet, daß Magnetnadel von stromdurchflossener Spule abgelenkt wird
1822	FARADAY erfindet Elektromotor
1825	STURGEON erfindet Elektromagnet
1837	WAGNER führt seinen elektromagnetischen Hammer vor
1838	WHEATSTONE baut Relais für Telegrafie
1850	GAUSS erforscht Magnetfeld der Erde
ca. 1850	JOULE entdeckt Magnetostriktion
1851	MORSE-Schreiber
1856	Doppel-T-Anker von SIEMENS
1860	Ringanker für Motoren
1879	Piezoelektrischer Effekt von CURIE entdeckt
1881	Drehspulmeßwerk von DEPREZ und D'ARSONVAL
1884	KOHLRAUSCH verwendet Anziehung von Spule und Eisen zur Strommessung als Grundlage für das Dreheiseninstrument
1885	Lord RAYLEIGH beschreibt Oberflächenwellen
1888	Erstes technisches Drehspulinstrument von WESTON
1889	Prinzip des elektrostatischen Multizellular-Voltmeters von Lord KELVIN
1898	Telegraphon von POULSON (Magnetdrahtspeicher)
1903	Saitengalvanometer
1904	Zungenfrequenzmesser von HARTMANN-KEMPF
1908	Erstes Relais für Fernsprechtechnik
1912	Entwicklung von Haftmagneten
1920	Automatik-Uhr (mechanischer Aufzug).
1921	SEIGNETTE-Salz als ferroelektrisch erkannt (1655 von SEIGNETTE synthetisiert)
1921	Erste Versuche mit Quarz, Frequenz zu stabilisieren
1923	Erster Kompensationsschreiber, Patent von USIGLI
1924	Erstes Hitzdrahtinstrument von CARDEW
1927	Erste Flachrelais
1928	PFLEUMER erfindet Magnetband
1929	Blattfernschreiber
1929	Magnetostriktive Stäbe zur Frequenzstabilisieerung durch PIERCE verwendet
1934	Quarzuhr von SCHEIBE und ADELSBERGER
1935	Magnetbandgerät
1937/1940	HF-Vormagnetisierung bei Magnetbandspeicherung
1938	ELLWOOD erhält Patent für REED-Kontakt
1939	Spannbandaufhängung von Meßinstrumenten
1940	Kaliumdihydrogenphosphat KDP von BUSCH und SCHERRER
1944	$BaTiO_3$ als Piezomaterial erkannt
1945	$BaTiO_3$ als ferroelektrisch erkannt
1947	Erste Metallresonanzfilter.
1947	Magnetischer Trommelspeicher
1951	SIXTUS entdeckt Magnetostriktion bei Ferritringen
1955	Blei-Zirkonat-Titanat = PZT
1956	Magnetplattenspeicher
1956	Künstlich gezüchteter Quarz für Schwinger und Filter
1958	Mechanische Einseitenbandfilter
1960	Mechanische Filter werden in größerem Umfang verwendet
1961	Beginn der Akustoelektronik = Entdeckung der Verstärkung von Ultraschall bei CdS
1963	Elektromechanische Uhr mit Aufzug aus Batterie
1965	Interdigitalwandler für Oberflächenwellen durch WHITE und VOLTMER

Fortsetzung Tabelle 6.3.1

1968	Geschliffener Stimmgabelquarz für Uhren
1969	Erste Quarzarmbanduhr mit analoger Anzeige (Schrittmotor)
1972	Quarzarmbanduhr mit digitaler Anzeige
1974	Stimmgabelquarz für Uhren nach STAUDTE
1976	Elektronischer Fernschreiber
1977	Billige 4-MHz-Quarze für Armbanduhren
1978	Piezokunststoff Polyvinylidenfluorid PVF_2
1979	Quarzarmbanduhr mit digital gesteuerter, analoger Anzeige (LCD)

Prinzipien realisieren lassen. Die Summe dieser Fakten war der Grund, diesen Abschnitt mit der (noch) ungebräuchlichen, zusammenfassenden Überschrift einzuführen. Folgende Teilgebiete werden *nicht behandelt*:

- elektroakustische *Wandler*, wie Lautsprecher, Mikrophone, Tonabnehmer usw.,
- *Motoren* aller Art,
- *Schrift- und Bildausgabe auf Hardkopien*, also Schreibwerke und Drucker aller Art (Ausnahme: registrierende und XY-Schreiber, 6.3.3.),
- *mechanische Anzeigen*, wie Zähler, Tableaus usw.,
- *Kontaktbauelemente*, wie Schalter, Stecker,
- *Leitungen und Verbindungstechniken* (Ausnahme: Schicht- und Hybridtechnik, 9.4.).

Behandelt werden die Teilgebiete:

Relais, Drehspul- und Registrierinstrumente, Piezoeffekt und Oberflächenwellen (Akustoelektronik) sowie daraus abgeleitete Bauelemente wie Schwingquarze, Filter und Signalverarbeitungen, weiter elektronische Uhren und magnetomotorische Speichergeräte.

Einen Überblick zur Mechanoelektronik gibt Tab. 6.3.1.

6.3.2. Relais

Das elektromechanische Relais — *mehrfach totgesagt — existiert noch immer* und nicht zu wenig, wie Bild 6.3.1 zeigt. Das gilt sowohl für den relativen als auch für den ab-

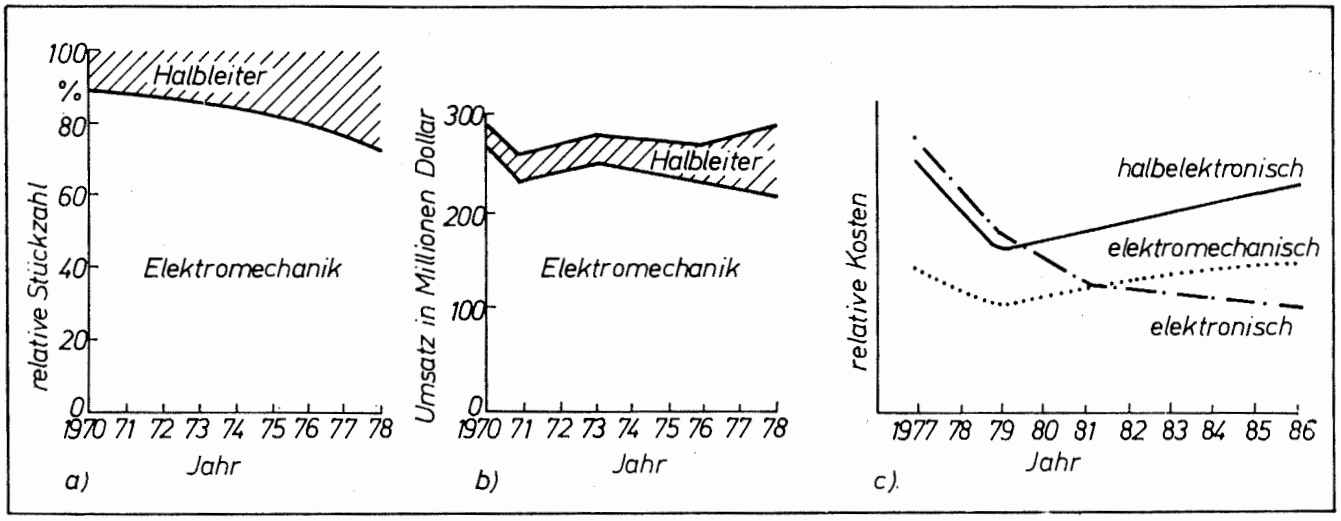


Bild 6.3.1. Zur Entwicklung der Relaistechnik

a) Stückzahlanteile von elektromechanischen und elektronischen Relais
b) Umsatz
c) Relative Kosten bei Automatisierungsanlagen

Tabelle 6.3.2. Gegenüberstellung von elektromechanischen Relais und Schaltern auf Halbleiterbasis; + bedeutet: im Vorteil

Eigenschaft	Relais	Halbleiter
Lebensdauer	10^6 bis 10^9 Schaltspiele	+ unabhängige Schaltzahl; lang
Schaltzeit	1 ... 50 ms (— Kontakt prellen)	+ ns bis µs
Schüttelfest	10 bis 1000 m/s² (1 bis 100 g)	+ gut
Überbelastung	+ unempfindlich	empfindlich gegen Spitzenspannungen und -ströme
Trennung von Primär- und Sekundärkreis	+ sehr gut bei kleinen bis großen Leistungen	nur mit Optokopplern erfüllbar
Strahlungsempfindlichkeit	+ unempfindlich	teilweise sehr hoch
Störbeeinflußbar	+ sehr gering	empfindlich
Schaltleistung	+ sehr hoch, extrem hoch bei Schützen	nur bei Thyristoren hoch, sonst klein bis mittel
On-Widerstand	+ sehr klein; mΩ — veränderlich durch Abrand — Kontakt prellen	— nicht so klein + stabil
Wechselstrombetrieb	+ auf Primär- und Sekundärseite möglich	schwierig (über Gleichrichter)
Vielzahl von Verknüpfungen	+ 3 bei Miniaturrelais, sonst viele Kontakte	nur 1 Schaltweg je Bauelement
Betriebstemperaturbereich	+ sehr groß	eingeengt, Temperaturkoeffizient vorhanden
Gewicht und Raum	meist größer als bei Halbleitern	+ klein
Primärleistung	mittel (Verstärkung 3 bis 50)	+ klein
Wartung notwendig	oft	+ nein
Alterung	+ gering	mäßig
Streufeld	magnetisch	+ nicht vorhanden
Speicherung des Schaltzustandes	+ möglich auch bei Stromausfall	schwierig bis unmöglich

Tabelle 6.3.3. Kennwertbereiche von Relais

Erregerseite		
Nennspannung	V	1,5 ... 220
Ansprechleistung (pro Wechsler)	mW	0,05 ... 1000
Ansprechzeit	ms	0,1 ... 15
Kontaktseite		
Dauerstrom, max.	A	0,2 ... 15
Schaltspannung, max.	V	24 ... 380
Schaltleistung, max.		
bei Gleichstrom	W	5 ... 400
bei Wechselstrom	VA	5 ... 10000
Durchgangswiderstand	mΩ	5 ... 200
Isolationswiderstand	Ω	10^9 ... 10^{12}
Kontaktkapazität	pF	0,2 5
Thermospannung		
(bei Betriebsleistung)	μV	2 ... 250
Prellzeit, max.	ms	0 ... 10
Kontaktkraft	cN	1 ... 50
Allgemein		
Prüfspannung Wicklung — Kontakt	V	150 ... 2500
Min. Einsatztemperatur	°C	−65 ... −25
Max. Einsatztemperatur	°C	50 ... 125
Schwing- und Stoßfestigkeit	m/s^2	10 ... 1000
	(g	1 ... 100)
Mechanische Lebensdauer	Schaltspiele	10^6 ... 10^9

soluten Anteil an relaisartigen Schaltungen. Es gilt sowohl für die Stückzahlen als auch die Umsatzpreise. Gewiß nimmt der Halbleiteranteil zu, aber eben viel langsamer als im allgemeinen angenommen wird. Ein Grund ist z. B. aus den relativen Kosten bei Automatisierungsanlagen gemäß Bild 6.3.1c) zu erkennen. Daneben führen auch die stark tradierten Erfahrungsträger beim Entwurf von Automatisierungsanlagen zur konservativen Entwicklung dieses Gebietes. Hinzu kommt, daß sich die Relais der Elektronik ebenfalls angepaßt haben. Sie wurden u. a. *miniaturisiert*, und so gibt es heute Relais, die in üblichen Halbleitergehäusen untergebracht sind. Und schließlich besitzt das Relais eine Reihe objektiver Vorteile für verschiedene Anwendungen. Einen Überblick hierzu gibt Tab. 6.3.2., während in Tab. 6.3.3 grob die Daten von Relais zusammengestellt sind. Bild 6.3.2 gibt einen Überblick zu wichtigen Relaistypen. *Rund-* und *Flachrelais* sind in der bei Telefonvermittlung um 1950 üblichen Form dargestellt. Sie existieren heute in vielen, meist stark verkleinerten Formen. Dadurch wird natürlich die Leistungsaufnahme und noch mehr das Streufeld kleiner. Das *polarisierte Relais* existiert in vielen Abwandlungen. Es hat den Vorteil, daß es mit kurzen Impulsen geschaltet werden kann und danach seinen Zustand speichernd (ohne Energiezufuhr) beibehält. Außerdem schaltet es sehr schnell (< 1 ms).

Der *Schutzgaskontakt* (auch REED-*Kontakt* genannt) hat eine Vielzahl von Vorteilen gegenüber anderen Relaisausführungen:

- hermetischer Abschluß gegen äußere Einflüsse,
- keine mechanischen Betätigungsglieder,
- hohe Lebensdauer bei völliger Wartungsfreiheit,
- große Schaltzuverlässigkeit,
- geringe Leistungsaufnahme,
- kurze Schaltzeiten (< 1 ms),
- kapazitätsarmer Kontaktaufbau,
- geringes Volumen und Gewicht.

Die Eigenschaften und Lebensdauer von Schutzgaskontakten nähern sich denen von elektronischen Bauelementen (Halbleitertechnik) an. Zusätzlich haben sie die guten Kontakteigenschaften. Eine Kombination von Halbleiterbauelementen und Schutzgasrelais kann also zu optimalen technischen Lösungen führen.

Soll ein *Relais mit Wechselstrom* gespeist werden, so muß sein Magnetkreis speziell ausgelegt werden. Zur Vermeidung von Wirbelströmen ist er zu lamellieren. Zur Ver-

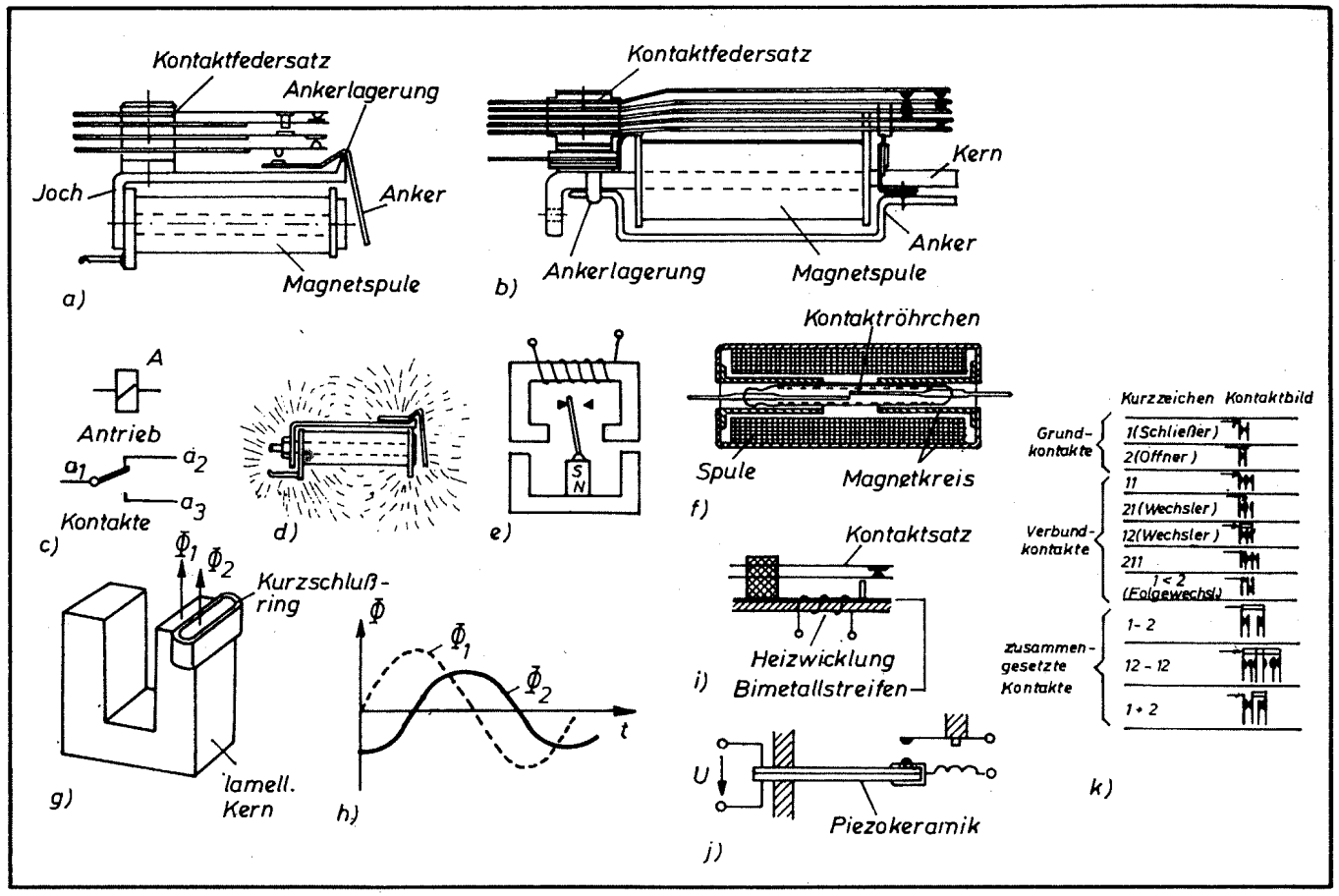


Bild 6.3.2. Beispiele von Relais

a) Rundrelais
b) Flachrelais
c) Schaltsymbol
d) Magnetisches Streufeld eines Relais
e) Polarisiertes (speicherndes) Relais
f) Relais mit Schutzgaskontakt (REED-Relais)
g) Kernbeispiel eines Wechselstromrelais (Spaltpolrelais) und
h) zugehöriger Fluß, der Flattern des Relais verhindert
i) Thermorelais
j) Piezorelais
k) Beispiele für Kontaktsätze an Relais

meidung von Flattern der beweglichen Teile, eventuell bis zu den Kontakten, müssen zwei phasenverschobene magnetische Flüsse erzeugt werden. Dies kann besonders einfach durch einen Spaltpol mit Kurzschlußring realisiert werden.

Es gibt neben den elektromagnetischen Relais eine Vielzahl anderer Typen. In Bild 6.3.2 ist ein *Thermorelais* (mittels Bimetallstreifen) und ein *Piezorelais* (wird elektrostatisch angesteuert) gezeigt. Tab. 6.3.4 zeigt in Ergänzung zu Bild 6.3.1 eine detaillierte Darstellung des Relaismarktes für das Jahr 1970.

Tabelle 6.3.4. Marktanteile für 1970 (in Klammer für 1980) bei verschiedenen Relaistypen (R.) mit einem Gesamtumsatz von 300 (500) Millionen Dollar

	Mill. Dollar		%	
Halbleiter-R.	22	(42)	7,3	(8,4)
Elektromagnetische Universal-R.	126,5	(150)	42,2	(30)
Sonstige R.	156	(308)	52	(61,6)
Anteilmäßig spezielle R.:				
Kristall-Dosen-R.	29,8		9,9	
Elektrodynamische R.	6,3		2,1	
Schutzgas-R.	26,1	(128)	8,7	(26,5)
Thermische R.	4,5		1,5	
Quecksilber-R.	15,0		5,0	
Stufenschalt-R.	20,0	(8,7)	6,7	(1,7)
Telefon-R.	24,8	(36,7)	8,3	(7,3)

6.3.3. Meßinstrumente

6.3.3.1. Drehspulinstrument

Für die Anzeige analoger Meßwerte hat das Drehspulinstrument (Bild 6.3.3a)) noch immer große Bedeutung, und das, obwohl sein Bau große Präzision verlangt. In einem kreisförmigen Spalt existiert ein homogenes, konstantes Magnetfeld H

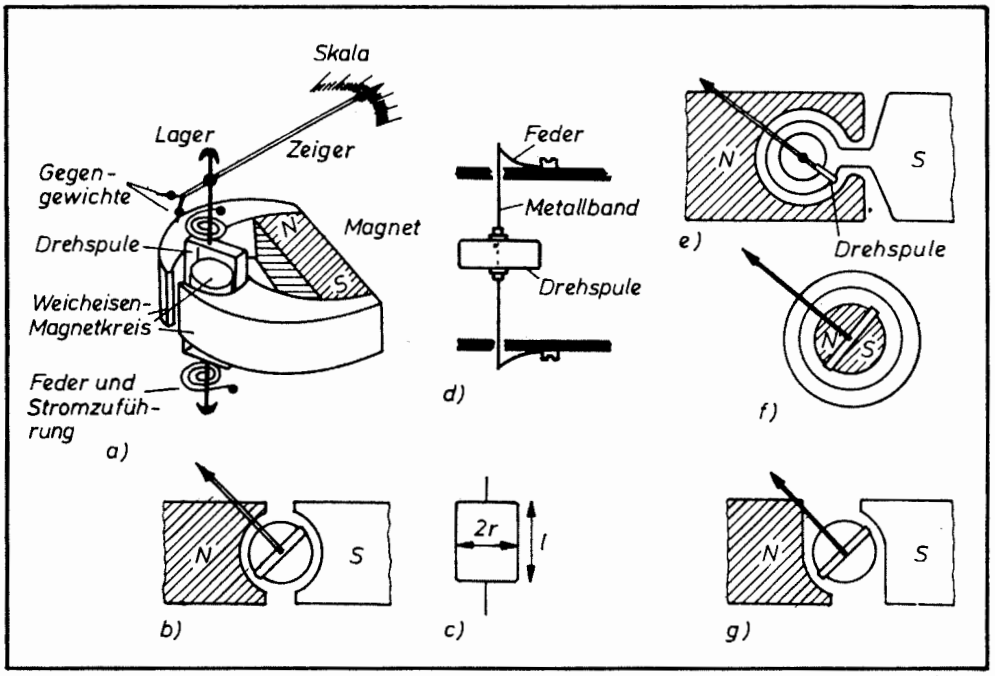


Bild 6.3.3. Drehspulinstrument

a) Prinzipaufbau
b) Erzeugung des homogenen Spaltfeldes, in dem sich die Drehspule bewegt
c) Abmessungen der Drehspule
d) Prinzip der Spannbandaufhängung
e) Magnetkreis für ein Drehspulinstrument mit großem Anzeigewinkel
f) Magnetkreis mit Kernmagnet
g) Kernkreis für nichtlineare Skala

(Bild 6.3.3b)). Darin ist eine Zylinderspule mit n Windungen, dem Radius r und der Länge l drehbar angeordnet. Durch einen Strom I wird also in zwei Spalten ein *Drehmoment* erzeugt, so daß gilt

$$M_I = 2InlH/r \,. \tag{1}$$

Dieses Drehmoment muß der mechanischen *Direktionskraft* D der beiden Federn entgegenwirken. Mit dem Ablenkungswinkel α gilt

$$M_F = 2D\alpha \,. \tag{2}$$

Zwischen beiden Drehmomenten stellt sich so ein *Gleichgewicht* für den Winkel α ein:

$$\alpha = \frac{nlH}{rD} I \,. \tag{3}$$

Der Ausschlag ist also direkt proportional dem Strom durch die Spule. Wichtig dafür ist, daß die Spaltweite im Drehbereich der Drehspule konstant ist, damit H konstant bleibt.

Für den Aufbau von *Drehspulinstrumenten* gibt es *viele Varianten*. Sie betreffen u. a. Lagerung der Drehspule, Stromzuführung und Magnetkreis. Eine hochwertige Lagerung geschieht mit einem *Spannband*, das bei Drehung tordiert wird (Bild 6.3.3d)). Dabei entfallen alle Lagerfehler, wie z. B. Umkehrspanne und Reibungsverluste.

Mit hochwertigen Magneten hat das *Kernmagnetinstrument* gemäß Teilbild f) Bedeutung erlangt. Hierbei bereitet jedoch die Homogenität des Feldes Probleme. Anderseits lassen sich hier relativ leicht *nichtlineare Skalen* realisieren, z. B. etwa logarithmische Anzeige, wie sie u. a. bei Belichtungsmessern benötigt wird. Bei Drehspulinstrumenten mit äußerem Magnet erfordert dies sonst kompliziert gestaltete Polschuhe, wie sie z. B. Teilbild g) zeigt. Für Rundskalen wurde der Magnetkreis gemäß Teilbild e) entwickelt, mit dem ein Drehwinkel von reichlich 270 realisiert werden kann.

6.3.3.2. Dynamometer

Alle Drehspulinstrumente sind nur für Gleichstrom geeignet. Wechselstrom kann über Gleichrichter gemessen werden. Für Sonderfälle hat es aber Vorteile, wenn der Dauermagnet durch einen *Elektromagneten* — durch den der Strom I_2 fließt — ersetzt wird. Dann gilt nämlich in einem gewissen Bereich für den *Ausschlagwinkel*

$$\alpha \sim I \cdot I_2 \,. \tag{4}$$

Der Wert ist vom Vorzeichen unabhängig, wenn beide Ströme aus dem gleichen Stromkreis entnommen werden. Es entsteht eine echte Effektivwertmessung ($\sim I^2$). Anderseits kann die multiplikative Verknüpfung beider Ströme auch für andere Zwecke, z. B. zur *Leistungsmessung* ($U \cdot I$), genutzt werden. Solche Instrumente heißen *Dynamometer*. Mit der Entwicklung der Elektronik werden sie aber jetzt kaum mehr genutzt bzw. produziert.

6.3.3.3. Dreheiseninstrument

Ein einfacheres Meßinstrument ist das Weicheisenmeßwerk (Bild 3.3.4a)). Seine Weiterentwicklung führte zum Dreheiseninstrument gemäß Teilbild b). Hierbei stoßen sich die beiden dünnen Weicheisenbleche D und E ab. Durch Formgebung und Abstand können definierte Skalenverläufe erreicht werden. Das Dreheiseninstrument ist für Gleich- und Wechselstrom geeignet. Es ist weitaus weniger empfindlich und auch ungenauer als ein Drehspulinstrument.

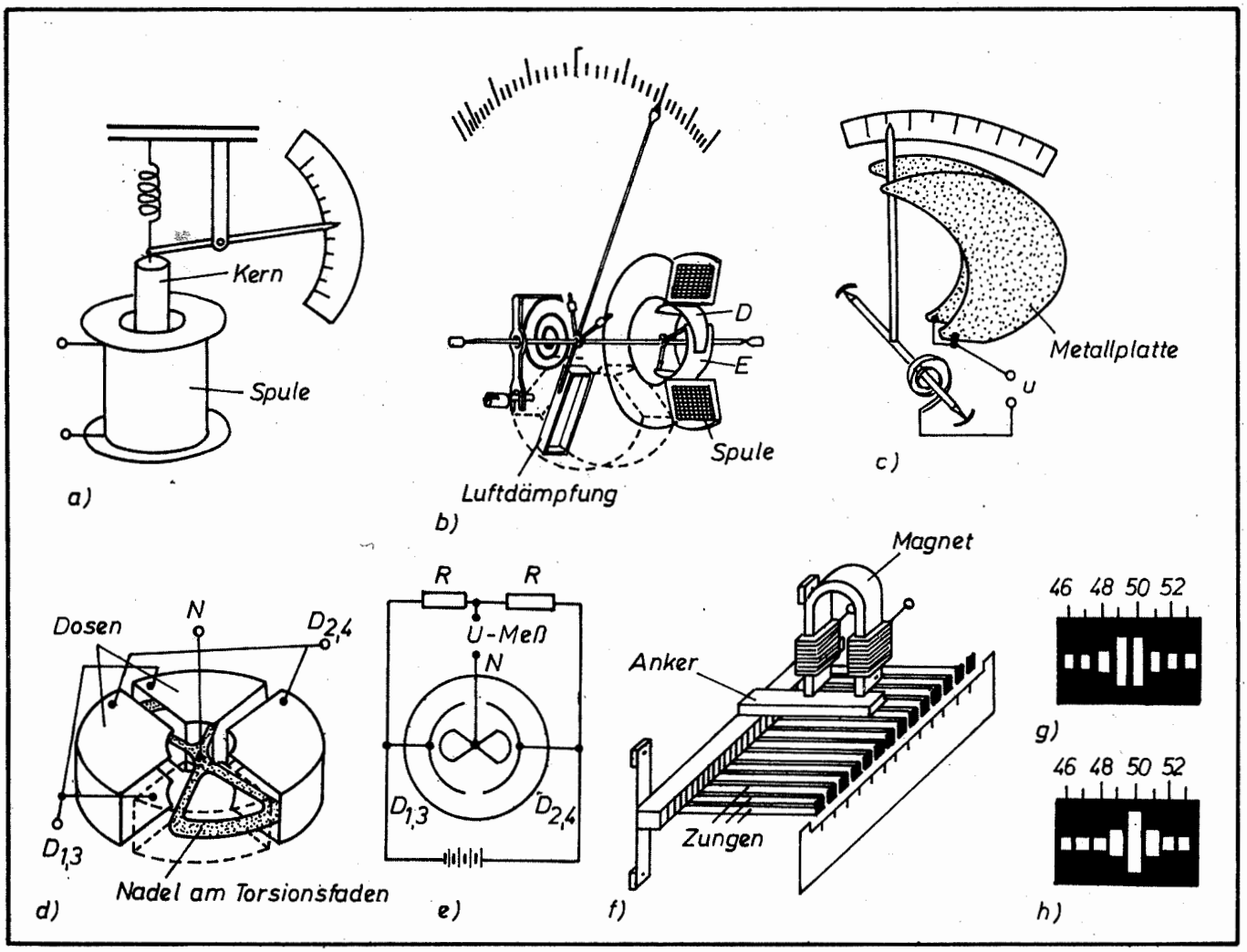


Bild 6.3.4. Mechanische Anzeigeinstrumente

a) Prinzip des Weicheiseninstruments
b) Daraus abgeleitetes Dreheiseninstrument; zwischen den beiden Magnetblechen D, E erfolgt im Feld Abstoßung
c) Prinzip des elektrostatischen Instruments
d) Dosen- bzw. Vierquadrantenelektrometer. Die etwa achtförmige Nadel ist an einem Torsionsfaden aufgehängt und kann sich in den vier Dosenteilen drehen. Die Anzeige erfolgt meist über einen am Faden befindlichen Drehspiegel optisch. Für das Quadrantenelektrometer existieren mehrere Schaltungen, u. a. die mit Hilfsspannung (e). Hierbei wird eine sehr hohe Empfindlichkeit erreicht. Bei anderen Schaltungen, z. B. der isostatischen, sind auch Wechselspannungen meßbar
f) Prinzip des Zungenfrequenzmessers und Anzeigebilder bei 49,5 Hz (g) und 50 Hz (h)

6.3.3.4. Elektrostatische Instrumente

Ein weiteres Meßprinzip ist das elektrostatische gemäß Bild 6.3.4c). Entgegen der Federkraft zieht die elektrostatische Anziehung zwischen Zeiger und den beiden gepunkteten Blechen den Zeiger möglichst weit in dieses Feld hinein. Solche Instrumente, die in vielfältigen Abwandlungen existieren, haben besonders wegen *zwei Eigenschaften* Bedeutung:

- Sie sind gut für Hochspannungen geeignet.
- Sie benötigen keine elektrische Leistung zur Anzeige.

Von Hochspannungsmessungen abgesehen, sind sie aber infolge ihrer geringen mechanischen Stabilität heute durch elektronische Anordnungen über Verstärker *praktisch völlig verdrängt*. Dies gilt sowohl für die an sich sehr robusten *Multizellularinstrumente* (die wie ein Dreko aufgebaut waren) als auch für das sehr empfindliche *Vierquadranten- bzw. Dosenelektrometer* (Teilbild d)). Mit einer Hilfsspannung (Teilbild e)) wurde hier

eine solche Empfindlichkeit erreicht, daß Spannungen unter 1 V elektrostatisch gemessen werden konnten.

6.3.3.5. Zungenfrequenzmesser

Als letztes mechanisches Anzeigeinstrument sei der Zungenfrequenzmesser gemäß Bild 6.3.4f) erwähnt. Er besteht aus einer Vielzahl von Metallzungen, die auf bestimmte Frequenzen abgestimmt sind. Im Bild ist dies mit Abstufungen von 1 Hz der Bereich von 46 bis 54 Hz. Diese Zungen werden von einem Elektromagneten über den Anker gemeinsam angeregt. Es zeigen sich dann Resonanzbilder wie in den Teilbildern g) und h) mit 49,5 bzw. 50 Hz.

6.3.3.6. Registrierinstrumente, *XY*-Schreiber und Plotter

Für viele Auswertungen ist es nützlich, ja notwendig, die Meßwerte auf Papier festzuhalten. Große Vorteile hat hier die Digitaltechnik mit ihren verschiedenen Druckern gebracht. Dennoch besitzen analoge Schreiber große Bedeutung. Dabei sind *zwei* prinzipiell verschiedene *Typen* zu unterscheiden:

- *Registrierinstrumente*, die nur den zeitlichen Verlauf des Meßwertes als Kurve auf Papier bringen;
- *XY-Schreiber* bzw. *Plotter*, die einen funktionellen Zusammenhang von zwei Meßwerten darstellen.

Für alle Registrierwerke existieren verschiedene *Methoden der Schreibtechnik*, z.B. mittels Tinte, Druck, Faserstiften, thermisch, fotografisch, mittels Funkendurchschlag, Abkratzen von Wachs usw. Diese Methoden erfordern z. T. spezielles Papier und spezielle konstruktive Auslegungen.

Der *Anzeigewert* kann

- *direkt* vom Meßwerk (Drehspule) abgeleitet werden oder
- *indirekt* über automatische Motorkompensation gewonnen werden.

Für das direkte Prinzip zeigt Bild 6.3.5a) das Beispiel des *Fallbügelschreibers*. In definierten Zeitabständen fällt der Bügel herab und druckt über die Stellung des Zeigers dann einen Punkt. Oft bewegt das Meßwerk auch direkt den Tintenstift oder die entsprechende Registriertechnik. Dazu ist mehr Leistung erforderlich.

Das indirekte Prinzip des *Kompensationsschreibers* ist schematisch in Bild 6.3.5b) und c) dargestellt. Die Meßspannung U_X wird mit einer Spannung U_V verglichen, die aus einer Normalspannung U_N über ein Potentiometer von einem Motor verstellt wird. Mittels Komparator und Verstärker wird dafür gesorgt, daß immer $U_X = U_V$ gilt, dann steht der über ein Seil geführte Zeiger und der Schreibstift genau an der entsprechenden Stelle.

Wird das Prinzip des Kompensationsschreibers auf eine *X*- und *Y*-Achse übertragen, so entsteht der *XY-Schreiber* (Bild 6.3.5d)). Wird die Stellung mit digitalen Methoden gewonnen, so liegt ein Plotter bzw. bei großen Abmessungen ein *Koordinatenschreiber* vor.

6.3.4. Wandler, Resonanzen und Filter

Folgende Gebiete werden behandelt:

- Wandler zwischen mechanischen Schwingungen und elektrischen Größen,
- mechanische Resonanzen und Filter,
- Oberflächenwellen (6.3.5).

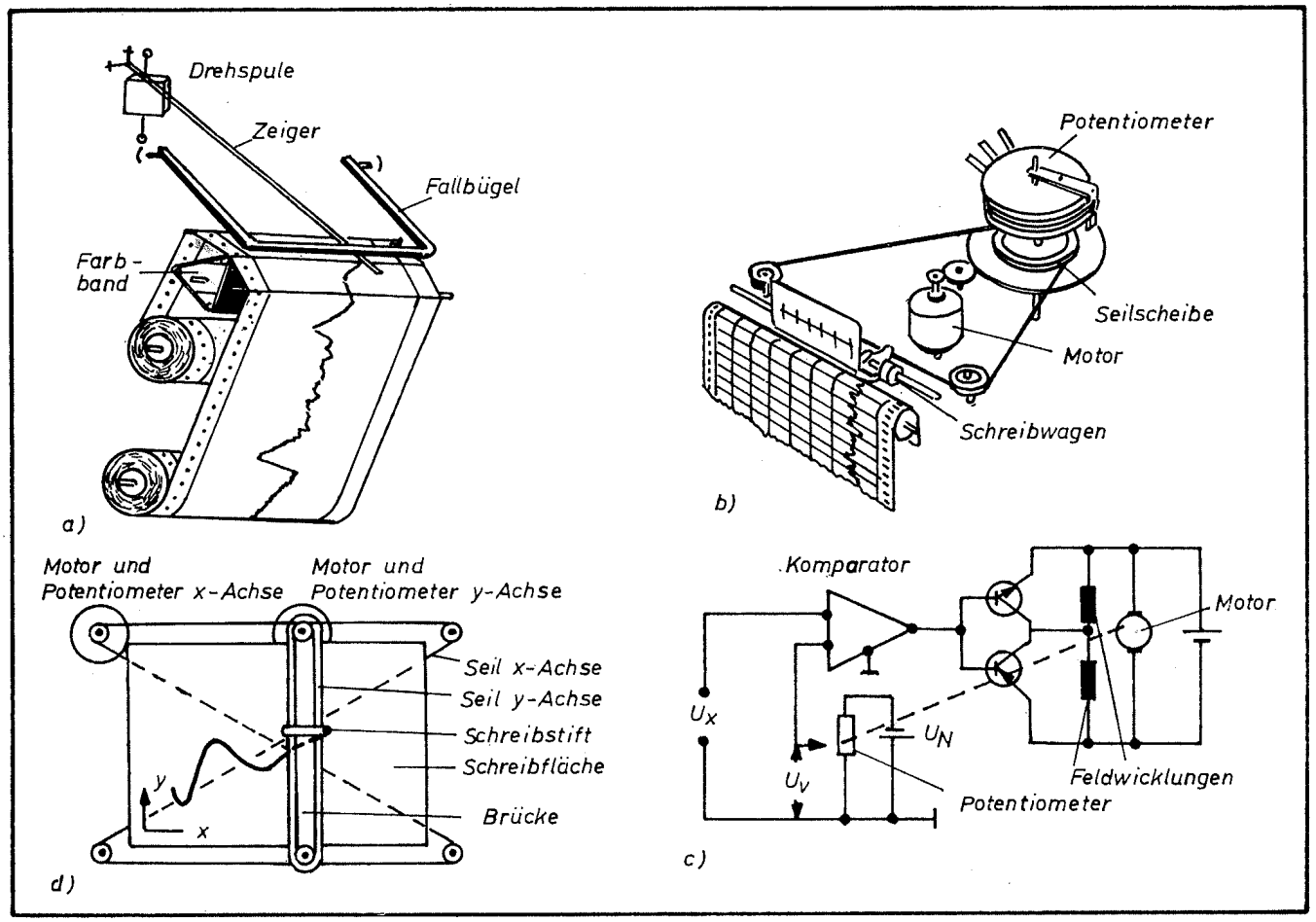


Bild 6.3.5. Registriergeräte

a) Fallbügelschreiber
b) Kompensationsschreiber
c) Schaltung dazu
d) Prinzipaufbau eines XY-Schreibers

Sie bilden die Grundlagen für die Anwendung in Resonanzschwingern, Filtern und Informationsverarbeitungen. Die Wandler haben darüber hinaus auch eigenständige Bedeutung, z. B. in Ultraschallsendern, Schwingungsaufnehmern, Piezotasten usw. Darauf wird hier nicht eingegangen.

6.3.4.1. Magnetostriktion

Die Magnetostriktion existiert in zwei Richtungen:

JOULEscher *Effekt*	VILLARI-*Effekt*
Unter dem Einfluß eines magnetischen Feldes entstehen Längenänderungen in magnetischen Materialien.	Durch eine mechanische Spannung wird in einem magnetisierbaren Material die Magnetisierung geändert.

Beide Effekte sind energetisch verkoppelt und treten daher meist gemeinsam auf. Der VILLARI-Effekt wird bei *Kraftmeßdosen* genutzt. Der JOULEsche Effekt (= magnetostriktiver Effekt im engeren Sinne) wird u. a. für *Ultraschallschwinger* ausgenutzt. Bei mechanischen Filtern werden beide Effekte in den Wandlern benötigt. Vor einiger Zeit hatten auch noch magnetostriktive Laufzeitspeicher Bedeutung.

So wie die Hysterese von magnetischen Materialien nichtlinear ist, so ist auch die Längenänderung Δl und das *Magnetostriktionsmaß*

$$\lambda = \Delta l / l \tag{5}$$

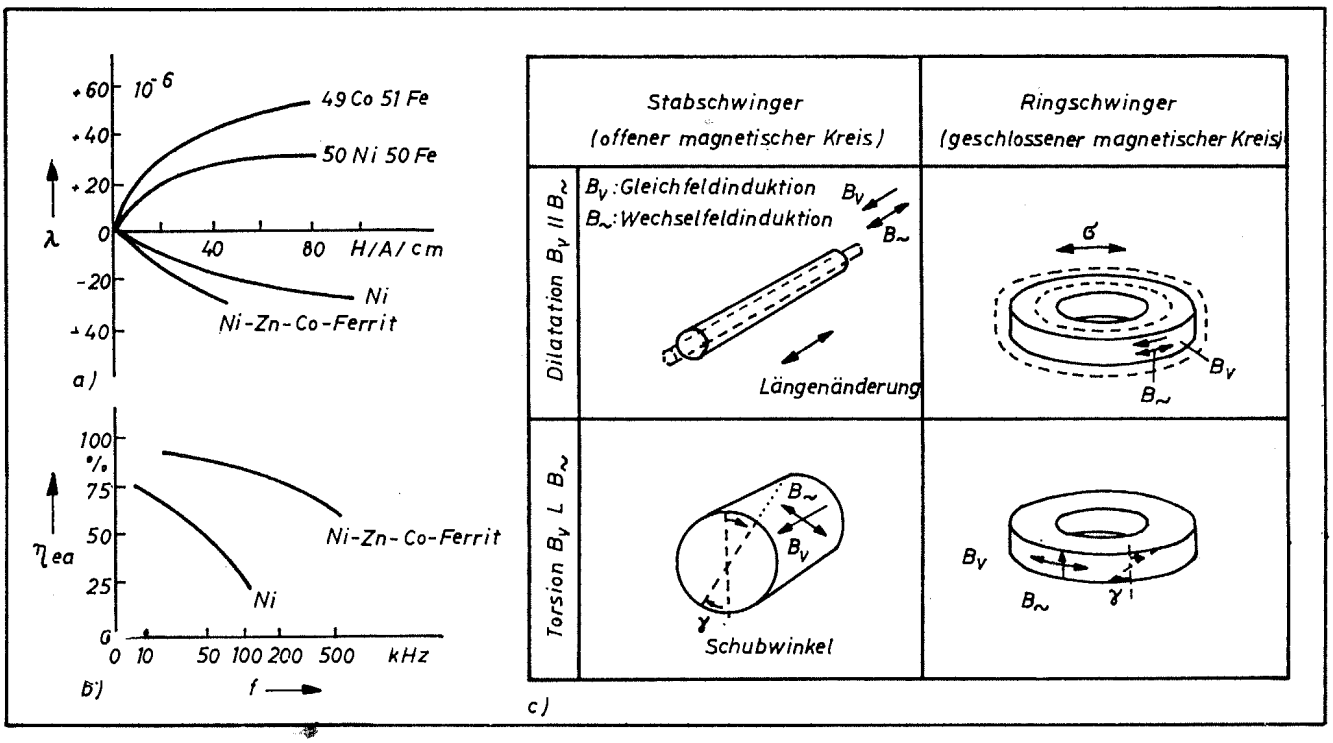


Bild 6.3.6. Zur Magnetostriktion

a) Relative Längenänderung als Funktion der Feldstärke
b) Wirkungsgrad eines Schwingers als Funktion der Frequenz
c) Schwingungsarten in Abhängigkeit vom magnetischen Kreis (offen bzw. geschlossen) und Orientierung des Vormagnetisierungsfeldes B_V zum Wechselfeld $B_\sim$

vom Arbeitspunkt abhängig (l = Länge der Probe). Für einige stark magnetostriktive Materialien ist dies in Bild 6.3.6a) dargestellt. Dabei ist zu beachten, daß die eingetretene Längenänderung unabhängig vom Vorzeichen der Feldstärke ist. Um bei den Wechselfeldern eine *Frequenzverdopplung* zu vermeiden, ist eine *Vormagnetisierung* notwendig, mit der ein Arbeitspunkt auf den Kennlinien festgelegt wird. Diese Vormagnetisierung kann parallel oder senkrecht zum Wechselfeld erzeugt werden. Weiter kann das magnetostriktive Material als offener oder geschlossener magnetischer Kreis (Stab bzw. Ring) ausgeführt werden. So entstehen die *vier Schwingervarianten* von Bild 6.3.6c). Der Schwinger kann also Längenänderung (Dilatation) oder Torsion zeigen. Mit der Längenänderung ist immer eine gewisse Querschnittsänderung verbunden.

Während früher als magnetostriktives Material nur Ni effektiv einzusetzen war, werden heute fast ausschließlich Ferrite verwendet. Durch ihre geringen Wirbelstromverluste sind sie bis zu sehr hohen Frequenzen anwendbar. Außerdem wird dadurch das Problem der Lamellierung umgangen. Diese Zusammenhänge machen die Kurven im Bild 6.3.6b) deutlich.

6.3.4.2. Piezoeffekt

Piezo stammt aus dem Griechischen und bedeutet „ich drücke". Piezoeffekte gibt es genau wie bei der Magnetostriktion zwei, jedoch werden sie nicht von der Bezeichnung her unterschieden. Sie werden beide in einem *Gleichungssystem* beschrieben:

$$\begin{pmatrix} D \\ S \end{pmatrix} = \begin{pmatrix} d & \varepsilon \\ s & d \end{pmatrix} \begin{pmatrix} T \\ E \end{pmatrix}. \tag{6}$$

Hierin bedeuten:

E elektrische Feldstärke in V/m,
D elektrische Verschiebung in As/m^2,
T mechanische Spannung (oft σ) in N/m^2,
S mechanische Dehnung (oft δ),
d Piezo-Ladungskonstante in m/V,
ε Dielektrizitätskonstante (bei der mechanischen Spannung Null) in $s/\Omega\,m$,
s Elastizitätskonstante (bei dem elektrischen Feld Null) in m^2/N.

Weiterhin wird als abgeleitete Größe noch die Spannungskonstante

$$g = dE/dT \tag{7}$$

in Vm/N verwendet. Sie gibt den Zusammenhang von mechanischer Spannung und elektrischer Feldstärke an.

Da im allgemeinen die Materialien anisotrop sind, besitzen die Konstanten von der Richtung abhängige Werte, sie sind Tensoren. Als *Materialien* sind zwei Typen zu unterscheiden:

Einkristalle, wie z. B. Quarz,
Polykristalle mit Orientierung wie Keramiken.

Polykristalle können nur dann piezoelektrisch sein bzw. gemacht werden, wenn sie *ferroelektrisch* sind, also in den Kristalliten eine *eingefrorene Polarisation* und, damit verbunden, insgesamt eine elektrische Hysterese vorhanden ist. Einige solcher Substanzen sind darüber hinaus *pyroelektrisch*, d. h., sie ändern das an den Elektroden vorhandene Potential mit der Temperatur. Weiter gibt es noch *piezoresistive* Materialien, auf sie wird im Abschnitt 6.5.3. eingegangen. Primär sind Ferroelektrika in der Regel nicht piezoelektrisch. Dies ist darin begründet, daß die Kristallite im Material regellos angeordnet sind und das Material pauschal isotrop erscheint. Durch Feldeinfluß oberhalb der CURIE-Temperatur und Abkühlung unter Feldeinfluß läßt sich bei verschiedenen Materialien eine Orientierung einfrieren, und es ist ein piezoelektrisches Material entstanden. Typisch hierfür sind die *Mischkeramiken* auf *Blei-Zirkonat-Titanat-Basis* (PZT-Keramiken).

Bild 6.3.7a) zeigt die Grundstruktur des *Quarzes* (SiO_2), wobei die Anordnung der Silicium- und Sauerstoffatome in einer Elementarzelle (von oben auf die Spitze gesehen = z-Achse) gemäß Bild 6.3.7b) vorliegt. Wird so ein Kristall in x-Richtung

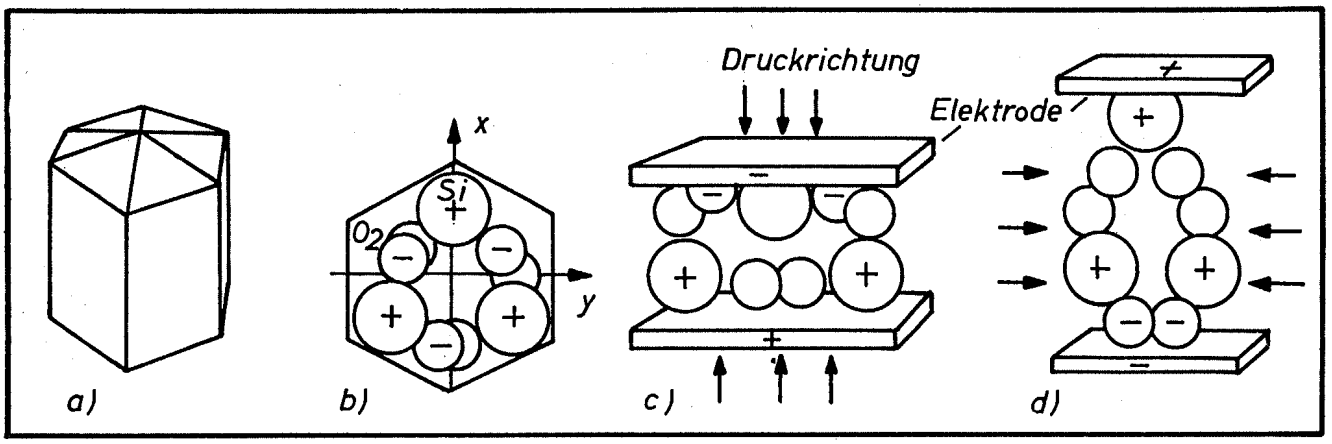


Bild 6.3.7. Piezoelektrische Effekte am Beispiel von Quarz

a) Vereinfachtes Schema eines Quarzkristalls. Die z-Achse zeigt nach oben
b) Elementarzelle von Quarz, mit Zuordnung der sechseckigen Form (Bild a) von oben gesehen) und den x- und y-Achsen
c) Longitudinaler Piezoeffekt
d) Transversaler Piezoeffekt

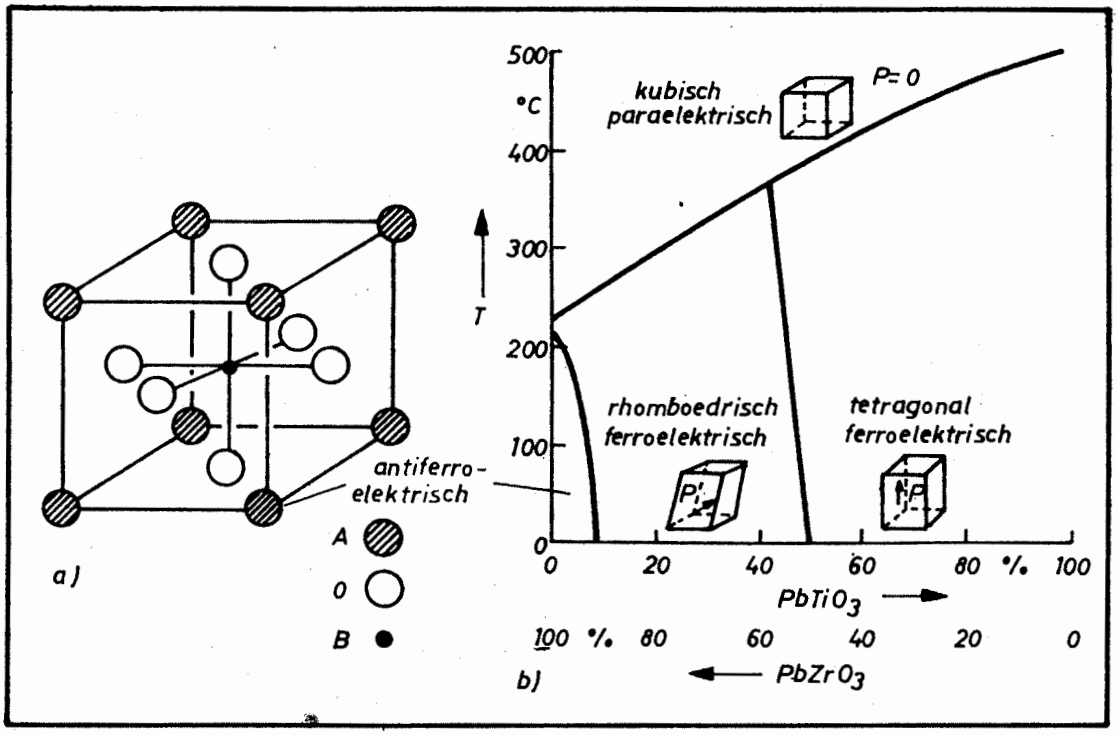


Bild 6.3.8. a) Elementarzelle und b) Phasendiagramm von PZT-Keramiken
a) Die chemische Formel lautet allgemein ABO_3, wobei die A-Plätze durch Pb und Ba und die B-Plätze durch Ti und Zr eingenommen werden können

mit Druck beauflagt und sind in *y-Richtung* Flächenelektroden angelegt, so entsteht der longitudinale Piezoeffekt gemäß Bild 6.3.7c). Er ist in diesem Schema leicht durch die elastische Verschiebung der Atome zu verstehen. Erfolgt an derselben Anordnung ein Druck in y-Richtung, so liegt der transversale Piezoeffekt gemäß Bild 6.3.7d) vor.

Die meisten *Piezokeramiken* gehen auf das kubische *Perowskitgitter* gemäß Bild 6.3.8a) zurück. Die Polarisation ist hierbei in verschiedenen Richtungen möglich Dies zeigt z.B. das Phasendiagramm von $PbZrO_3$-$PbTiO_3$-Mischkristallen gemäß Bild 6.3.8b).

Für das pauschale Material wird die z-Achse in Analogie zum Quarzkristall in Richtung der eingefrorenen Polarisation gelegt. Als Indizes für die verschiedenen *Konstanten* in Gl. (6) wurden die Ziffern 1 bis 6 gewählt. In der Form der *Doppelindizes* verweisen

erster Index: auf die Lage der Spannung bzw. Elektroden (nur 1 bis 3 möglich),

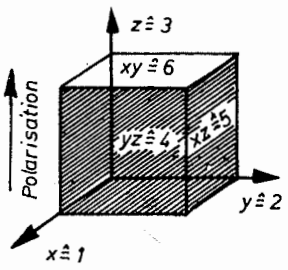
zweiter Index: auf die Deformation (1 bis 3) bzw. Torsion (4 bis 6).

Dabei gilt:

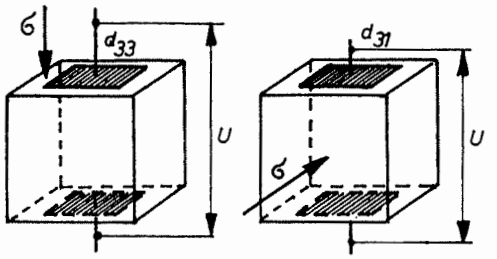
1: x-Achse	4: yz-Ebene
2: y-Achse	5: xz-Ebene
3: z-Achse	6: xy-Ebene.

Die entsprechenden Veranschaulichungen zeigt Bild 6.3.9. In dem Bild sind außerdem Beispiele einiger Schwinger mit den dabei wirksamen Konstanten enthalten. Dabei tritt noch der *Kopplungsfaktor k* zwischen den mechanischen und elektrischen Größen auf. Allgemein kann er aus Dielektrizitätskonstanten für konstante Spannung ε^T (dem bisher verwendeten ε) und konstanter Verformung ε^S gemäß

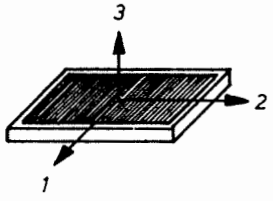
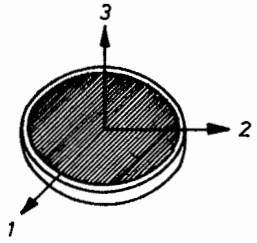
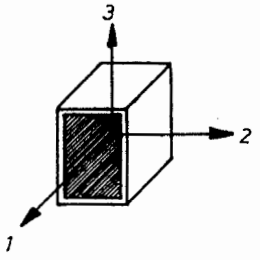
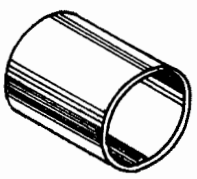
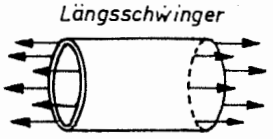
$$k = \sqrt{1 - \varepsilon^S/\varepsilon^T} \tag{8}$$



a)



b)

Form	elektromechanische Effekte	Konstanten
Platten, Stäbe	Dickenschwinger Feld Deformation	d_{33}, g_{33}, k_{33} s_{33} ε_{33}
	Längsschwinger	d_{31}, g_{31}, k_{31} s_{11} ε_{33}
Scheiben	Radialschwinger	d_{31}, g_{31}, kp s_{11}, ϱ ε_{33}
	Dickenschwinger	d_{33}, g_{33}, k_{33} s_{33} ε_{33}
Quader	Scherschwinger	d_{15}, g_{15}, k_{15} s_{44}, s_{55} ε_{11}
		
Rohre, Ringe	Dickenschwinger	d_{33}, g_{33}, k_{33} s_{33} ε_{11}
	Ringschwinger	d_{31}, g_{31}, k_{31} s_{11} ε_{33}
	Längsschwinger	

bestimmt werden. Aus den Konstanten in Gl. (6) kann er gemäß

$$k = \frac{d}{\sqrt{\varepsilon s}} \tag{9}$$

berechnet werden. Für den transversalen Fall (Längsschwinger) gilt folglich

$$k_{31} = \sqrt{\frac{d_{31}^2}{\varepsilon_{33} s_{11}}}. \tag{10}$$

Der planare Kopplungsfaktor beträgt

$$k_p = k_{31} \sqrt{\frac{2}{1 + s_{12}/s_{11}}}. \tag{11}$$

In vielen Fällen wird das Quadrat des Kopplungsfaktors, also k^2 verwendet. Es gibt den energetischen Wirkungsgrad eines Wandlers an.

Tab. 6.3.5 und 6.3.6 fassen typische Werte für Piezomaterialien zusammen. *Turmalin* und *Seignettesalz* haben heute keine Bedeutung mehr. In den letzten Jahren sind zu den genannten Materialien noch *Piezofolien* auf der Basis von z.B. Polyvinyleden-Fluorid (PVF_2 oder PVDF) hinzugekommen. Sein Piezokoeffizient liegt etwa bei $^1/_{10}$ desjenigen der Piezokeramiken. Dieses Material besitzt vielfältige Anwendungen, z.B. als Blutdruckmesser, Füllstandsensor (Resonanzlänge durch Eintauchtiefe) und schwingende Folie als Ventilator.

Tabelle 6.3.5. Chemische Zusammensetzung und Richtwerte einiger piezoelektrischer Materialien. Die Daten hängen von Dotierungen und der Orientierung des Materials ab.

Material			ε_r	k	Q	ϱ	v in m/s	TK: 10^{-6}/K
	kristallin	Quarz (SiO_2)	4,58	0,09	5 ... 7	2,65	3158	3
		$Bi_{12}GeO_{20}$	38	0,12	groß		1600 ... 1800	−100
		$Bi_{12}SiO_{20}$	mittel	0,13	groß		ca. 1600	+100
		$LiNbO_3$	29 bzw. 44	0,22	sehr groß		3500 ... 4000	−80
		$LiTaO_3$	ca. 42	0,1	sehr groß		3100 ... 3300	±50
ferroelektrisch	Keramik	$BaTiO_3$	1000	0,15 ... 0,4	350	5,4	3000 ... 5000	−80
		$PbTiO_3$	150	0,04 ... 0,4	7,7 1200	7,7	2700 ... 4000	
		$Pb(Ti-Zr)O_3$	800 ... 2000	0,25 ... 0,7	70 ... 1000	7,8	1500 ... 3400	
		$PbNb_2O_6$	250	0,04 ... 0,38	10	5,8	1800 ... 3000	

6.3.4.3. Mechanische Resonanzen, Quarze

Magnetostriktive und piezoelektrische Materialien werden für drei *Hauptanwendungen* genutzt:

- Erzeugung von zeitlich *konstanten Frequenzen* (Quarzgenerator, Quarzuhr),
- *Signalfilterung* (Analogie zum Abschnitt 2.5.),
- *Erzeugung* energiereicher mechanischer *Schwingungen* (z.B. Ultraschall).

◀

Bild 6.3.9. Bezeichnung der Achsen und Flächen für piezokeramische Materialien a) allgemein und b) für zwei spezielle Piezokonstanten sowie Beispiele von Piezoschwingern mit zugehörigen Parametern

Tabelle 6.3.6. Typische Daten von PZT-Keramiken und Quarz, Turmalin ($NaMg_3Al_6B_3Si_6O_{27}(OH)_4$) und Seignettesalz ($NaK \cdot C_4H_4O_6 \cdot 4H_{27}$) zum Vergleich

Konstante	$PbTiO_3$	PZT hohe Kopplung	PZT hohe Güte	$BaTiO_3$	Quarz	Turmalin	Seignettesalz
ε_{33}	130	2000	950	1000	4,58	1,6	
ε_{11}	180	1800	850	1000			
$\tan\delta$ (10^{-3})	3	20	10	10			
v_3 (km/s)	4,0	2,6	3,3	5,0	3,158	4,4	
v_1 (km/s)	4,1	2,8	3,4	4,6	3,159		
d_{33} (10^{-12} m/V)	39	370	190	100	$d_{11} = 2{,}26$	$d_{15} = 3{,}6$	$d_{14} = 2{,}3$
d_{31} (10^{-12} m/V)	−4,2	−170	−80	−40	$d_{14} = -0{,}66$	$d_{25} = -0{,}22$	$d_{25} = -5{,}6$
d_{15} (10^{-11} m/V)		580	260	220	$d_{25} = 0{,}66$	$d_{31} = 0{,}24$	$d_{36} = 1{,}16$
g_{33} (10^{-3} Vm/N)	33	25	23	14	ca. 60		
g_{31} (10^{-3} Vm/N)	−3,5	−11	−10	−5			
g_{15} (10^{-3} Vm/N)		38	35	22			
k_{33}	0,4	0,7	0,6	0,4	ca. 0,09		
k_{31}	0,04	0,34	0,27	0,14			
k_{15}	0,33	0,68	0,55	0,4			
k_p	0,07	0,60	0,5	0,25			
s_{33} (10^{-12} m²/N)	8,4	19	12	9		6,2	
s_{11} (10^{-12} m²/N)	7,8	16	11	8,5	13		ca. 3
s_{44} (10^{-12} m²/N)		47	30	20			
Güte, Q	1200	70	1000	350	10^5 ... 10^7		
ϱ (Ω cm)	10^{11}	10^{14}	10^{11}	10^{12}	10^{16}		e

Diese Anwendungen setzen *Resonanzgebilde* voraus, die eine Analogie zu den elektrischen Schwingkreisen darstellen. Statt Ladungen, die periodisch einen Kondensator umladen, ist im mechanischen Bereich ein periodischer Wechsel zwischen kinetischer und potentieller Energie notwendig. Ein besonders anschauliches Beispiel ist das Pendel; ein anderes die Orgelpfeifen, wo Luftdruck und Luftströmung sich abwechseln. Ähnlich verhalten sich mechanische Festkörper, wenn die *Schallgeschwindigkeit* in ihnen in einem passenden Verhältnis zu *geometrischen Abmessungen* steht. Die Abmessungen sind dabei in Richtung der Schwingung, z.B. bei den Wandlern gemäß Bild 6.3.6 bzw. 6.3.9, zu wählen. Für die Schallgeschwindigkeit v sind Anhaltswerte in Tab. 6.3.5 und 6.3.6 enthalten. Weitere Werte ergänzt Tab. 6.3.7. Generell gilt mit dem Elastizitätsmodul E und der Dichte ϱ die Beziehung

$$v \approx \sqrt{E/\varrho}\,. \tag{12}$$

Mit einer Frequenz f, die das Material anregt, entsteht im Material eine Wellenlänge λ gemäß.

$$\lambda = v/f\,. \tag{13}$$

Tabelle 6.3.7. Richtwerte von Schallgeschwindigkeiten in m/s

Luft	343	Aluminium	5080	Messing	3640
Wasser	1440	Kupfer	3710	Stahl	5000
Glas	4500	Nickel	4790	Ferrit	5650

Resonanzen treten immer dann auf, wenn die Abmessungen des Materials in Schwingungsrichtung ein ganzzahliges Vielfaches von $\lambda/2$ betragen. In der Regel wird die Länge $\lambda/2$ selbst verwendet. Da die meisten Schwingkörper in verschiedenen Richtungen verschiedene Abmessungen besitzen, verfügen sie auch über *mehrere Resonanzen*. Die Gestalt wird entsprechend den Materialeigenschaften meist so gewählt, daß eine Frequenz betont ausgeprägt ist. Hinzu kommt, daß durch unterschiedliche Orientierung zur Polarisations- bzw. Kristallachse spezielle Eigenschaften, wie z.B. günstige Temperaturkoeffizienten, erhalten werden. So sind die verschiedenen Quarzschnitte zu verstehen, für die Bild 6.3.10 einen kleinen Überblick gibt.

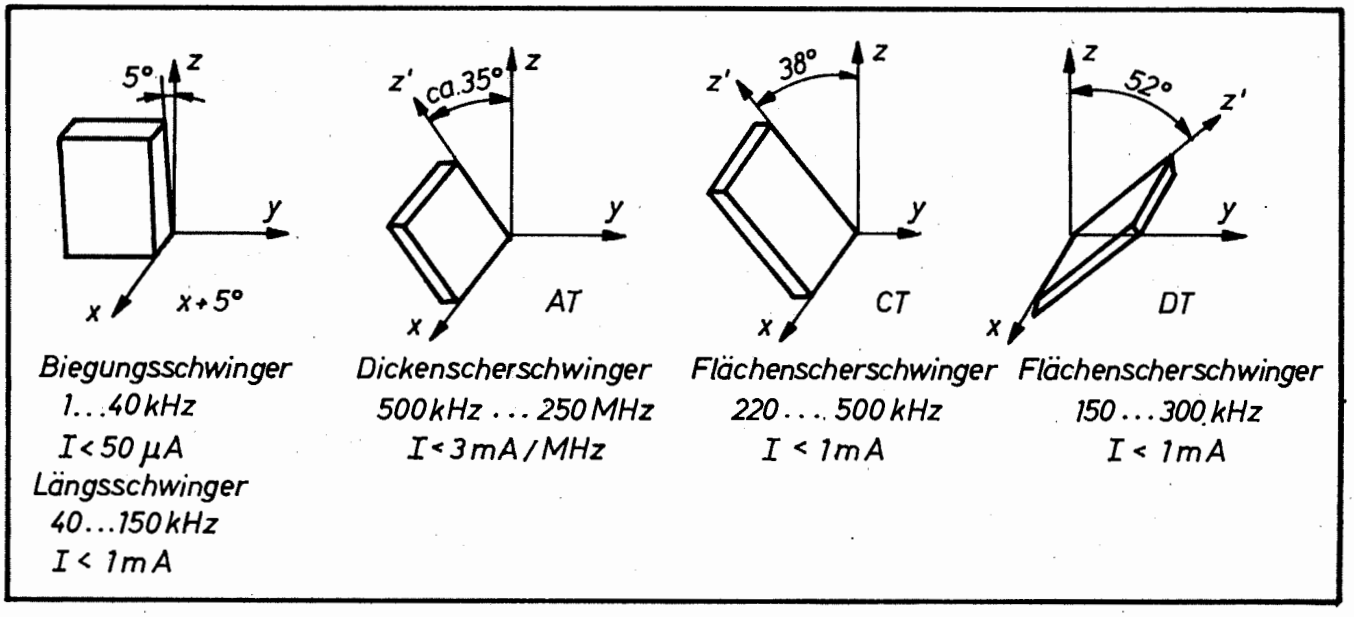


Bild 6.3.10. Schnitte, Schwingungsarten, Frequenzbereich und maximal zulässiger Strom für einige Quarztypen

Für die Elektronik ist zu beachten, daß *Quarz eine doppelte Bedeutung* hat. Einmal kann das *fertige Bauelement* gemeint sein und zum anderen nur das *Material.*

Beim Quarz — wie bei allen Schwingern — sind *zwei Frequenzen* zu beachten:
Im *Leerlauf*, d. h. bei offenen Klemmen, ruft die mechanische Schwingung eine Piezospannung hervor. Sie besitzt ihr Maximum bei der Parallelresonanz f_p.
Im *Kurzschlußbetrieb* entsteht kein elektrisches Feld, es fließt nur ein Strom. Die mechanische Spannung im Material ist dabei geringer. Der Strom besitzt ein Maximum bei der tieferen Serienresonanz $f_s < f_p$.

Dieses Verhalten legt die Ersatzschaltung gemäß Bild 6.3.11a) mit der Ortskurve gemäß Teilbild b) und der Impedanz gemäß Teilbild c) nahe. Bezüglich des *Kopplungsfaktors* k gilt in etwa

$$k = \sqrt{2{,}5 \cdot \frac{f_p - f_s}{f_s}} \,. \tag{14}$$

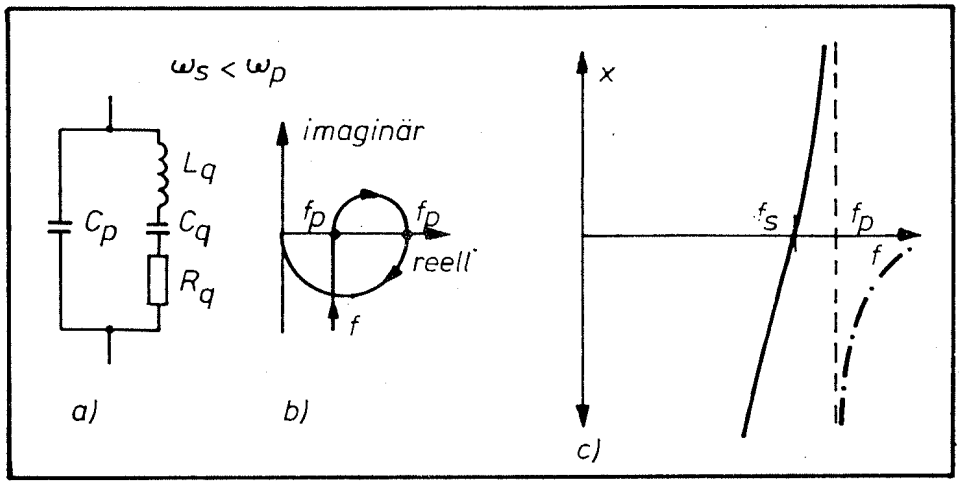


Bild 6.3.11. Eigenschaften von Quarzen

a) Ersatzschaltbild des Quarzes
b) Ortskurve, wobei zu beachten ist, daß die Parallelresonanz f_p in Wirklichkeit viel weiter vom Ursprung entfernt liegt
c) Relativer Impedanzverlauf als Funktion der Frequenz

Hieraus ist ersichtlich, daß ein Quarz mit dem kleinen Kopplungsfaktor viel genauer in der Frequenz definiert ist als keramische Schwinger mit großem Kopplungsfaktor. Deshalb werden die keramischen Massen je nach Anwendungszweck auf hohe Güte (kleines k) und hohen Kopplungsfaktor hin optimiert (vgl. Tab. 6.3.6). Für die *Serienresonanz* gilt

$$f_s = \frac{1}{2\pi} \sqrt{\frac{1}{L_q C_q}} \tag{15}$$

und für die *Parallelresonanz*

$$f_p = \frac{1}{2\pi} \sqrt{\frac{1}{L_q C_q} \left(1 + \frac{C_q}{C_p}\right)} \,. \tag{16}$$

Der Verlustwiderstand setzt sich aus inneren Verlusten im Material, Strahlungsverlusten und Verlusten durch Auskopplung von Energie über die Zuleitungen oder bewußte Auskopplung zusammen. Für die *Güte* gilt etwa

$$Q = \frac{1}{4\pi (f_p - f_s)\, R_q C_0} \,. \tag{17}$$

Theoretisch läßt sich aus den inneren Verlusten des Quarzes ableiten

$$Q \cdot f \leqq 16 \cdot 10^{12}\ \mathrm{Hz}\,. \tag{18}$$

Dieser Wert wird näherungsweise für Quarze ab 30 MHz erreicht. Dies zeigt Bild 6.3.12a).

Die verschiedenen in Bild 6.3.10 dargestellten *Quarzschnitte* bestimmen nicht nur den optimalen Frequenzbereich für die einzelnen Schwingungsformen, sondern viel stärker den Temperaturgang der Frequenz. Dies zeigt Bild 6.3.12b) und c).

Der *XY-B-Schnitt* weist z.B. eine Parabel auf, die man so legen kann, daß der Scheitelpunkt auf die gewünschte Temperatur fällt (Bild 6.3.12c)). Der *AT-Schnitt*

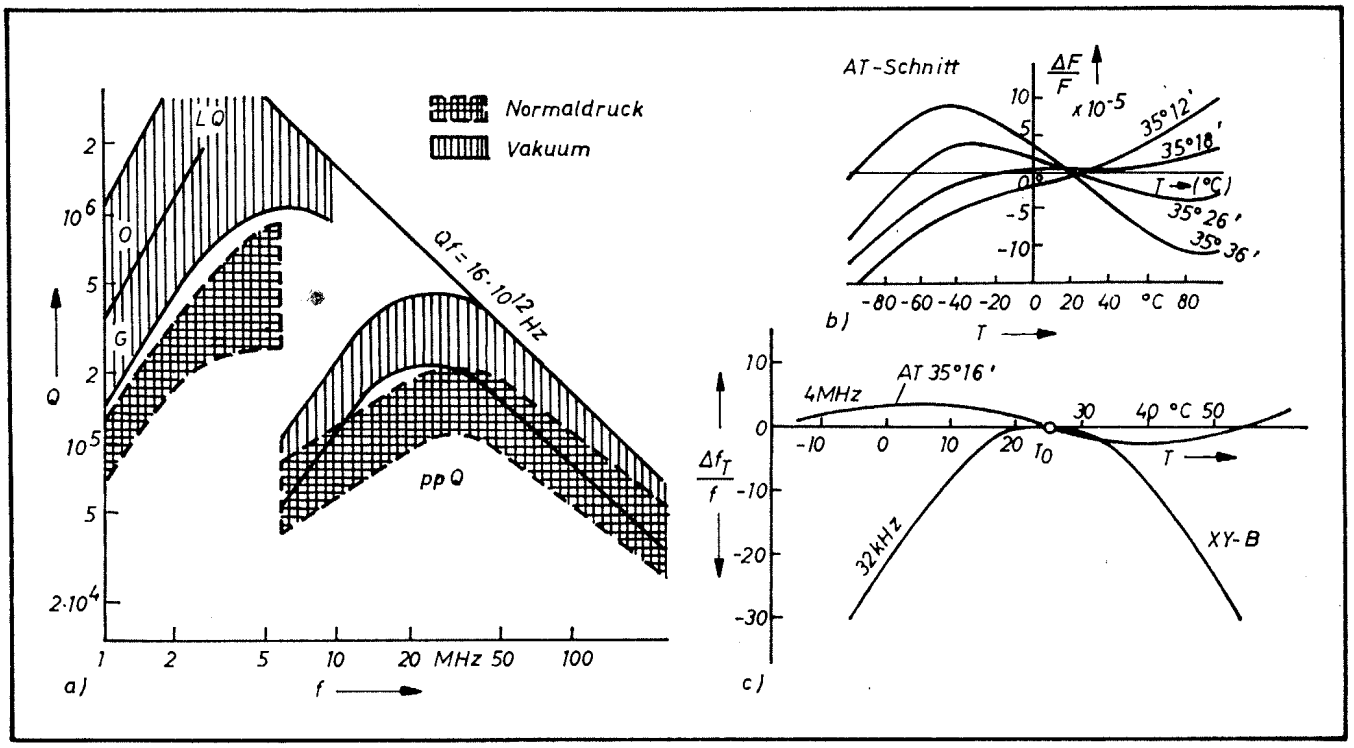


Bild 6.3.12. Güte (a) und Temperaturkoeffizienten (b) und c)) von Quarzen

b) Für AT-Schnitte

c) Vergleich von optimal gewähltem AT- und XY-B-Schnitt bei Uhrenquarzen

In a) bedeuten: LQ = Linsenquarze in O = Oberton- und G = Grundtonbetrieb, ppQ = planparallele Quarze

hat infolge des Wendepunktes im Temperaturgang seine große Bedeutung erlangt. Bei ca. 35° 16′ wird eine fast geradlinige Kennlinie erreicht. Dies gilt jedoch nur statisch, d. h. für langsame Temperaturänderungen. Bei schnellen Änderungen treten im Quarz selbst örtliche Temperaturdifferenzen auf, die zu einer Frequenzverwerfung führen. Hier wurde nun in den letzten Jahren ein neuer Schnitt — **TCC** bzw. **SC** (stress compensated) genannt — gefunden, der doppelte Winkelabweichung gegenüber den Quarzachsen anwendet (Bild 6.3.13a)). Mit dem Winkel Θ wird in erster Näherung die Kurve dritten Grades bestimmt, während Φ die Umkehrtemperatur festlegt. Deshalb muß für einen optimalen Temperaturgang zwischen beiden ein Zusammenhang gewahrt sein (Bild 6.3.13b)). Einen für 90 °C optimal kompensierten Temperaturgang stellt Bild 6.3.13c) dar. Den Vorteil dieser dynamischen Temperaturkompensation zeigt das Einschaltverhalten eines im Thermostaten befindlichen Oszillators (Bild 6.3.13d)).

Folgende Daten weisen zahlenmäßig den Gewinn für einen *typischen* 10-MHz-Quarz aus:

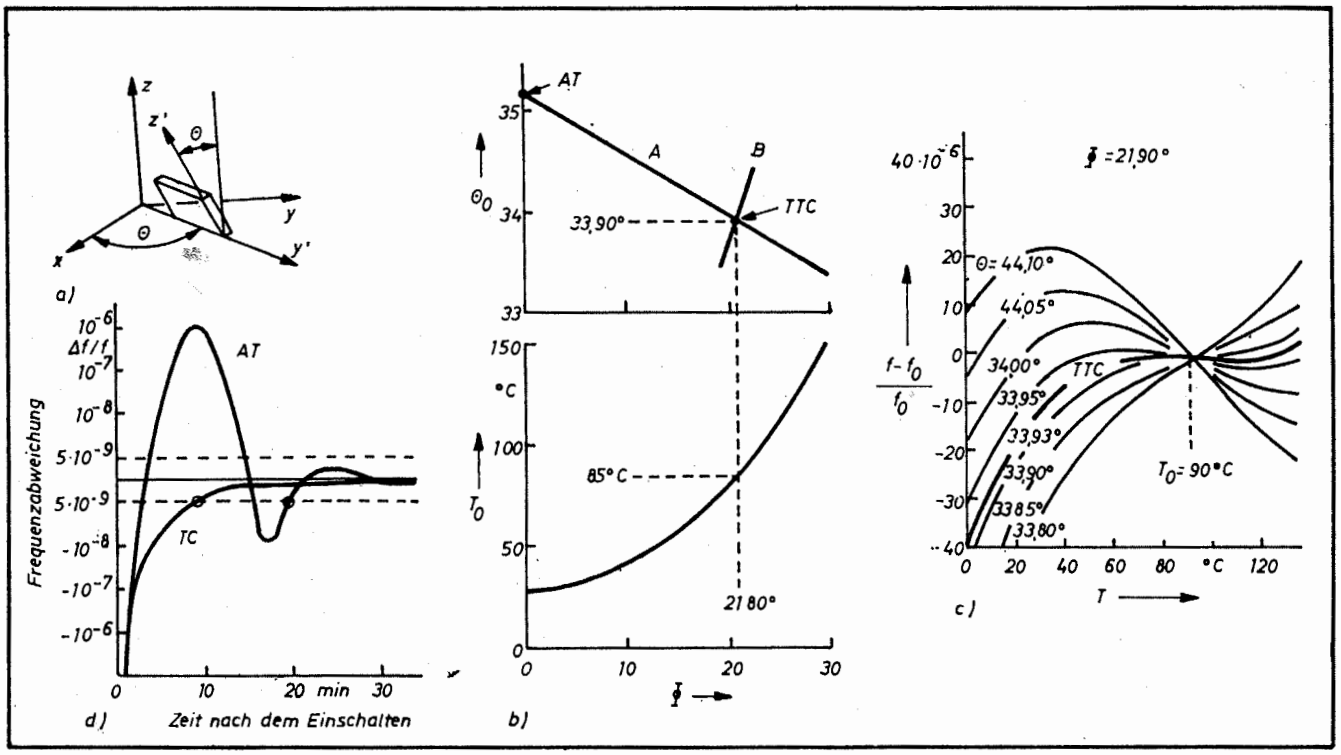


Bild 6.3.13. TTC- bzw. SC-Schnitt eines Quarzes

a) Lage der beiden Drehwinkel (zwischen x und y' liegt Φ)
b) Abhängigkeit der beiden Winkel für eine gemeinsame Kompensationstemperatur T_0
c) Kurven in der Umgebung eines auf 90 °C kompensierten TTC-Schnittes
d) Einschwingverhalten einer Quarzfrequenz in einem Thermostaten bei AT- und TTC-Schnitt

Frequenzänderung im Bereich von 70 bis 100 °C	2 statt $20 \cdot 10^{-6}$/K
Auswirkung eines Wärmeimpulses von $\Delta T = 1$ K als relative Frequenzänderung	2 statt $25 \cdot 10^{-8}$
Alterung bezüglich der Frequenz	2 statt $5 \cdot 10^{-9}$/d
Güte	4 statt $2 \cdot 10^5$
C_q	7 statt 30 fF
L_q	35 statt 8 mH
R_q	5 statt 2,5 Ω
Koppelfaktor	5 statt 8,8 %.

Aus Bild 6.3.11 ist ersichtlich, daß in einem gewissen kleinen Umfang die *Resonanzfrequenzen* eines Quarzes beeinflußt (gezogen) werden können. Hierbei sind zwei Fälle zu unterscheiden:

- *Kapazität in Reihenschaltung* zum Quarz

  Die Serienresonanzfrequenz wird zu höheren Werten verschoben. Das ist nur in geringem Maße möglich:

$$\frac{\Delta f}{f_s} \leqq 10^{-4}.$$

- *Kapazität parallel zum Quarz* (Lastkapazität)

  Die Parallelresonanzfrequenz wird zu tieferen Werten verschoben. Auch hier ist nur eine geringe Änderung möglich:

$$\frac{\Delta f}{f_p} \leqq 10^{-4}.$$

Durch das *Ziehen* werden die elektrischen *Eigenschaften* des Quarzes *verschlechtert*.

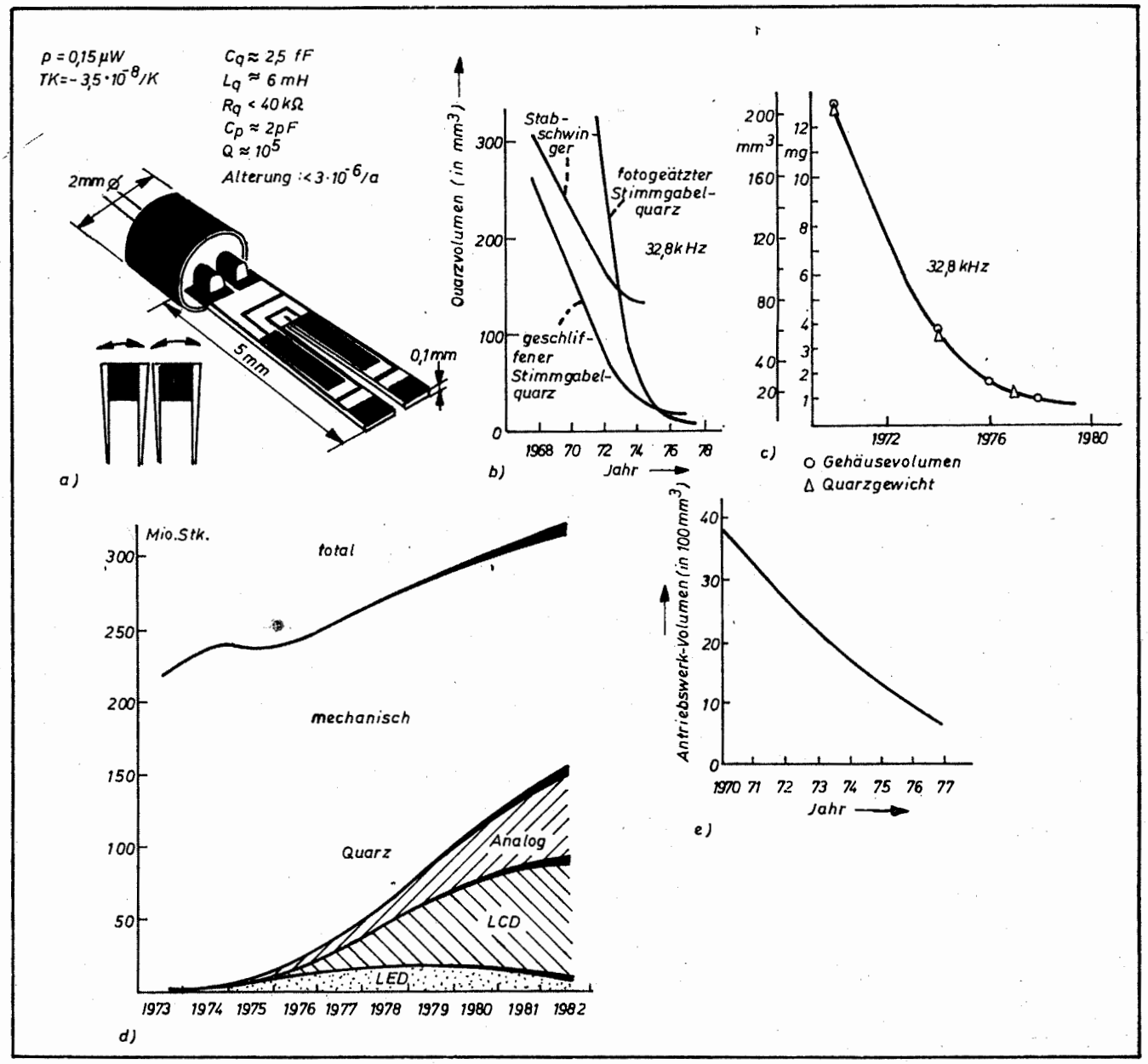


Bild 6.3.14. Zur Entwicklung der Quarzarmbanduhr

a) Fotolithografisch geätzter Stimmgabelquarz für 2^{15} Hz = 32768 Hz
b) Die drei Varianten von Uhrenquarzen im Laufe der Entwicklung
c) Über alles gemittelte Entwicklung der Quarzvolumina und -massen
d) Verteilung der Weltuhrenproduktion
e) Entwicklung des mittleren Antriebswerkvolumens von Quarzarmbanduhren

Eine große Weiterentwicklung auf dem Gebiet der Quarze hat die *elektronische Armbanduhr* gebracht. Entsprechende Fakten dazu zeigt Bild 6.3.14. Aus Teilbild b) sind die drei verschiedenen Technologien zu erkennen, welche hierbei entscheidend waren. Zunächst wurde ein einfacher Stab-Biege-Schwinger verwendet. Er ließ sich mit bester Technologie bis auf etwa 100 mm³ verkleinern. Gegen Ende der 60er Jahre erfolgte dann ein Übergang zur *Stimmgabel*, die mühsam aus dem Rohmaterial geschliffen wurde. Den entscheidenden Schritt brachte jedoch die von STAUDTE entwickelte Fotolithografie. Sie geht von einem 0,1 mm dicken *Dünnschliff* des Quarzes aus. Dieser ist so orientiert, daß *fotolithografisch* am besten gearbeitet (vor allem geätzt) werden kann. Die Anregung des Quarzes wird mit einer neuentwickelten Struktur so realisiert, daß sie ebenfalls von einer Seite fotolithografisch aufgebracht werden kann. Insgesamt entsteht dann die Struktur von Teilbild a). Sie ist wie bei der Halbleitertechnik für viele, z. B. 100, Quarze zugleich auf einem Blättchen zu realisieren. So entsteht eine sehr hohe Effektivität bei der Herstellung. Die *Metallbelegungen* an der Spitze der Stimmgabel werden für den *Feinabgleich* der Frequenz genutzt.

Im Laufe der Zeit entstand im Mittel über die drei Technologien ein Quarzvolumen bzw. eine Quarzmasse, wie sie Teilbild c) zeigt. Das gesamte Antriebswerk einer elektronischen Uhr reduzierte sich dabei gemäß Teilbild e). Teilbild d) zeigt anteilmäßig die weltweite Produktion von Armbanduhren. Interessant ist dabei auch das Verhältnis der Anzeigetechniken LED:LCD: Analog mit Schrittmotor.

Inzwischen wurde mehrfach versucht, den *Quarz durch Edelstahl* oder andere Materialien *zu ersetzen.* Hierbei konnten bisher jedoch nur Teilerfolge erzielt werden. Eine Stahlstimmgabel wird z. B. mit 20 µm dickem ZnO (piezoelektrisch) beschichtet, und darauf folgen dann die Leiterzüge (Gold). Es werden Güten um 30000 erreicht. Die Antriebsleistungen sind jedoch um ein Vielfaches größer als bei Quarz.

Quarz wird seit einigen Jahren *künstlich gezüchtet.* Dies erfolgt durch Lösung von Quarzbruch in Natronlauge bei ca. 1000 Atmosphären Druck und 450 °C. Die Wachsgeschwindigkeit beträgt etwa 2,5 mm/d.

Die *Vorteile der elektronischen Uhren* liegen nicht nur bei der stark vereinfachten Technologie, sondern auch bei den Käufern. Im Vergleich zur mechanischen Uhr sind, abgesehen vom Preis, die wesentlichsten Vorteile:

- *weniger Unterhalt,* die Uhr funktioniert auch ohne dauernd getragen werden zu müssen. Heute gibt es Quarzuhren mit Batterien, die für drei Jahre oder länger ausreichen.
- *höhere Zuverlässigkeit.* da weniger Teile vorhanden und diese weniger kritisch sind.
- Die *Genauigkeit* ist etwa hundertmal besser und immer noch verbesserungsfähig (1 min/a statt ca. 1 min/d).
- Besonders digitale Quarzuhren eignen sich für die Aufnahme *neuer Funktionen,* wie Wecker, Stoppuhr, Datum, Zeitzonen und vieles mehr.

Insgesamt kann man sagen, daß die Quarzuhr ein besseres Verhältnis von Leistung zu Preis aufweist als die mechanische.

Es könnte sein, daß der Quarzuhr eine weitere Entwicklung bevorsteht. Bisher wird die *Frequenz* 2^{15} Hz = *32768 Hz* verwendet. Sie erlaubt nur einen parabelförmigen Frequenz-Temperaturgang gemäß Bild 6.3.12c). Beim Übergang zur Frequenz 2^{22} Hz = *4,1943 MHz* ist ein AT-Schnitt mit einem kubischen Temperaturgang und damit noch höhere Ganggenauigkeit möglich. Dies wird z. Z. noch durch den dann erheblich höheren Stromverbrauch der Schaltungen erschwert. Bei den jetzigen Schaltungen liegt der Stromverbrauch bei ca. 1 µA.

6.3.4.4. Mechanische Filter

Ein magnetostriktiver oder piezoelektrischer Wandler kann ähnlich wie ein Quarz (im vorigen Abschnitt) auf Resonanz betrieben werden. Dann erzeugt er eine beträchtliche Schallenergie. Sie kann über rein mechanische Resonatoren weitergeleitet und damit gefiltert zu einem Ausgangswandler geführt werden. Als *Resonatoren* haben sich *kleine, zylinderförmige Metallproben* bewährt. Sie können zumindest in *drei Modi* betrieben werden (Bild 6.3.15):

Kompressionsschwinger (Dichtewellen)
Torsionsschwinger (Verdrehung von Anfang und Ende)
Biegeschwinger (Verbiegung der Form).

Neben den Eigenresonanzen der Materialien (Schwingkreise) müssen auch *Koppelelemente* vorgesehen werden, die einen meist kleinen Teil an das nächste Resonanzelement weiterleiten. Sie sind $\lambda/4$ lang. Die hohe *Güte* der mechanischen Resonatoren (ca. 20000) und ihre relativ einfache Herstellung erlauben es, so hoch wirksame und preiswerte Filter zu realisieren. Es werden etwa folgende Frequenzbereiche erreicht:

Biegefilter	10 ... 500 kHz
Torsionsfilter	100 ... 300 kHz
Kompressionsfilter	300 ... 600 kHz

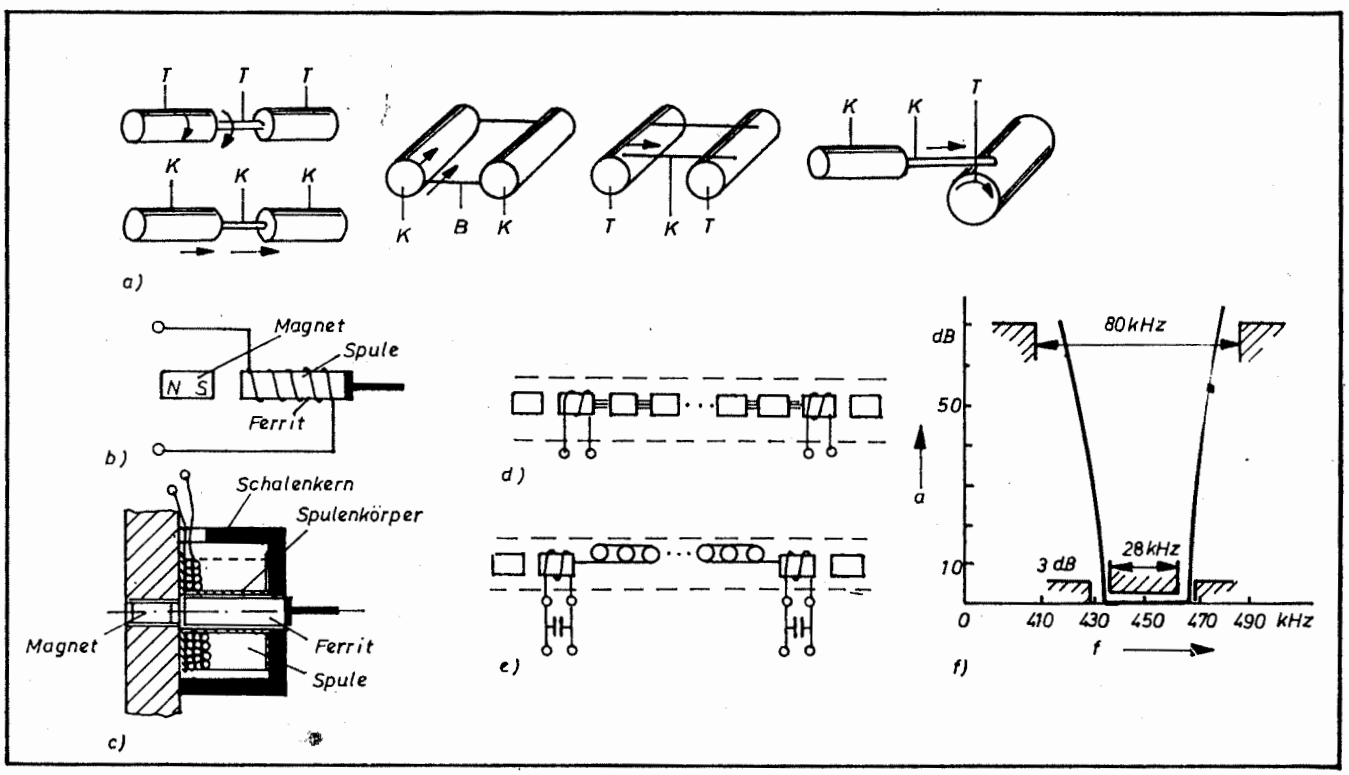


Bild 6.3.15. Mechanische Filter

a) Verschiedene Resonanz- und Koppelelemente. Es bedeuten K = Kompressions-, T = Torsions- und B = Biegeschwinger

b) Prinzipaufbau und

c) konstruktiver Aufbau eines magnetostriktiven Wandlers

d) Schematischer Aufbau eines Torsionsfilters und

e) eines Kompressionsfilters

Frequenzgang und Toleranzfeld eines neunkreisigen mechanischen Filters

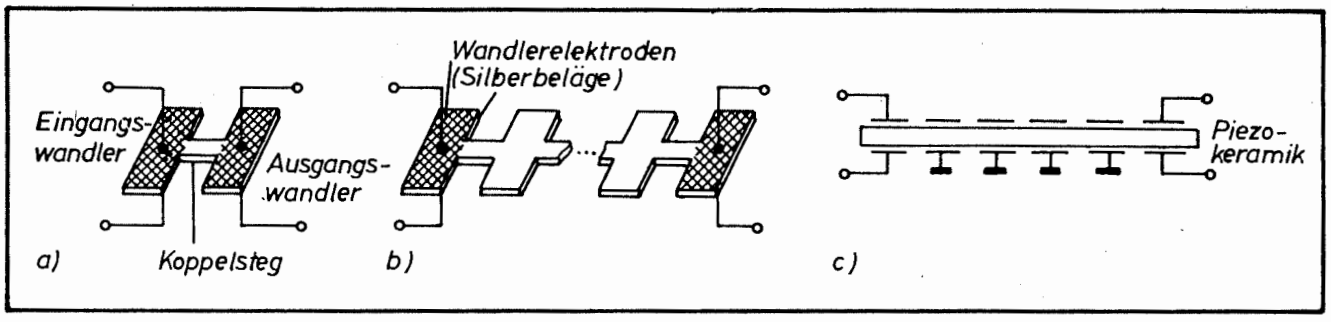


Bild 6.3.16. a) Zwei- und b) mehrkreisige piezoelektrische Koppelfilter, c) Prinzipaufbau eines monolithischen Filters

Eine zweite Variante mechanischer Filter leitet sich vom *Piezoeffekt* ab (Bild 6.3.16). Die Wandler, Resonatoren und Koppelelemente sind aus dem gleichen Piezomaterial. Wandler und Resonatoren unterscheiden sich durch die Belegung mit Kontaktmaterial. Die Koppelelemente sind wieder $\lambda/4$ lang und gegenüber den Resonatoren verjüngt. Der Aufbau solcher Filter ist also noch einfacher. Nachteilig ist jedoch die relativ geringe Güte der Resonatoren (s. Tab. 6.3.6). Der Vorteil der Piezokeramikfilter zeigt sich besonders deutlich, wenn sie monolithisch ausgeführt werden. Hierbei werden auf einem piezielektrischen Substrat die Resonatoren durch (auch nicht angeschlossene) Elektroden realisiert. Das Substrat zwischen den Elektroden bildet das Koppelelement (Bild 6.3.16c)). Mit *Piezofiltern* ist etwa folgender *Frequenzbereich* verfügbar:

400 kHz bis 4 MHz diskreter Aufbau
4 MHz bis 20 MHz monolithischer Aufbau.

6.3.5. Oberflächenwellen (OFW)

Oberflächenwellen (SAW = surface acoustic waves bzw. AOW = akustische Oberflächenwellen) sind typisch für *Flüssigkeiten* und wurden hier von LORD RAYLEIGH beschrieben. Sie treten aber auch bei *Festkörpern* auf. Nur ist hier die potentielle Energie *nicht die Schwerkraft, sondern die Elastizität* des Mediums. Das typische Verhalten einer Oberflächenwelle im Festkörper zeigt Bild 6.3.17. Die Verformungen des Materials an der Oberfläche sind besonders groß und *nehmen zur Tiefe hin schnell ab.* Bei einer Tiefe von 3λ ist die Bewegung vernachlässigbar. An der Oberfläche bewegt sich der Gitterpunkt in einer Ellipse der Ausbreitungsrichtung entgegen. (Bei Flüssigkeitswellen existiert die umgekehrte Richtung.) Bei ca. $0{,}14\lambda$ existiert eine lineare, vertikale Schwingung. Noch tiefer ist die Umlaufrichtung der Ellipse umgekehrt. Der Grund für das geringe Eindringen von OFW in die Tiefe liegt darin, daß die Ausbreitungsgeschwindigkeit von Oberflächen- und Volumenwellen erheblich verschieden ist. Dies bedingt eine sehr geringe Dämpfung der OFW. Bei den typischen Geschwindigkeiten der Wellen (vgl. Tab. 6.3.5 und 6.3.6) ergeben sich für 1 GHz *Wellenlängen* von ca. 3 µm. Infolge der geringen Dämpfung sind dabei *Güten* um 10^4 bis 10^5 zu erreichen. Damit entstehen mit OFW Schaltungsmöglichkeiten, die keine andere Methode effektiv erlaubt.

Angeregt und detektiert werden OFW mit *Interdigitalwandlern* (= IDW) (Fingerstrukturen) gemäß Bild.6.3.18a). Unter dem Einfluß der angelegten Spannung ver-

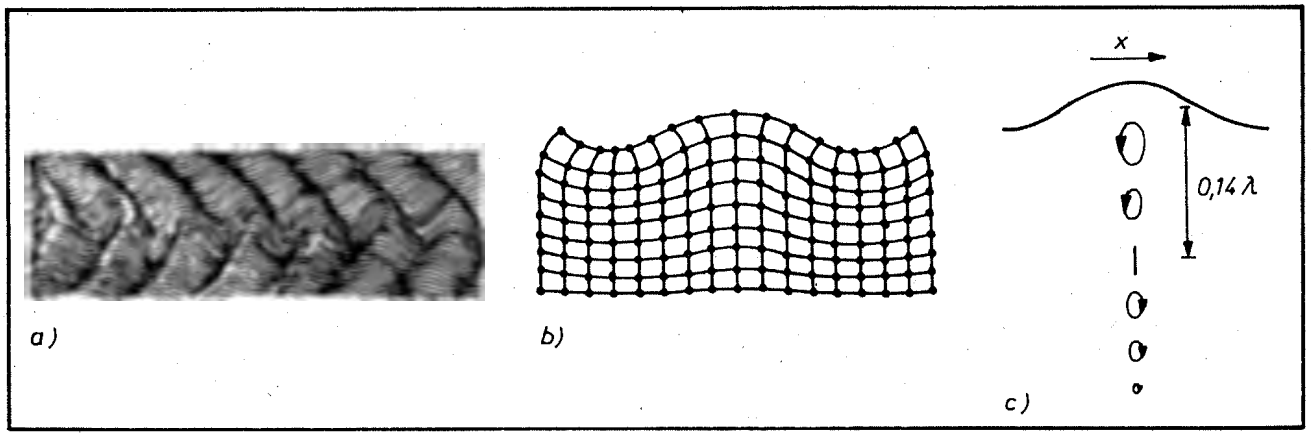


Bild 6.3.17. a) Schema einer Oberflächenwelle
b) Gitterverzerrung an der Oberfläche und in der Tiefe
c) Bewegung einzelner Gitterpunkte in verschiedenen Tiefen

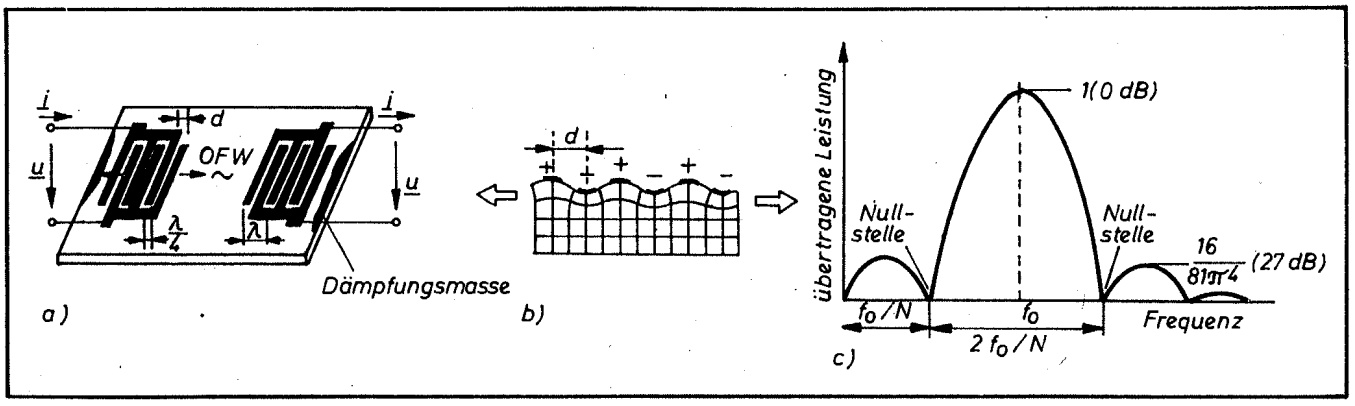


Bild 6.3.18. Interdigitalstrukturen der Wandler
a) Prinzipaufbau
b) Wirkungsmechanismus zwischen OFW und Spannungen,
c) Frequenzgang

formt sich das piezoelektrische Material entsprechend Bild 6.3.18b), und die so angeregte OFW breitet sich nach links und rechts aus. Gelangt sie zu einem zweiten IDW, so ruft sie umgekehrt in ihm die entsprechende elektrische Spannung hervor. Damit die Wellen an den Kanten nicht reflektiert werden, bringt man dort Dämpfungsmassen an.

Infolge des Kristallaufbaus ist die Ausbreitungsgeschwindigkeit einer OFW nicht in alle Richtungen gleich. Deshalb müssen die einzelnen Finger genau der *Orientierung des Kristalls* entsprechend der optimalen Ausbreitungsrichtung angepaßt sein. Weiter muß die Oberfläche des Materials extrem glatt sein. Jede Unebenheit und Inhomogenität führt zu *Reflexionen* und Störungen der OFW.

Seit etwa 1973 ist bekannt, daß auch *magnetostatische OFW* existieren (MSSW = magnetostatic surface waves). Ihre Ausbreitungsgeschwindigkeit ist höher, so daß derartige Filter im UHF-Bereich (mehrere GHz) wirksam werden können. Anwendungen liegen bisher kaum vor.

6.3.5.1. Eigenschaften von OFW-Filtern

Allgemein haben bei Schaltungen mit OFW die *einzelnen Finger* des IDW

- unterschiedliche Länge (Gewicht a_n)
- unterschiedlichen Abstand (Frequenz f)
- unterschiedliche Breite (Bandbreite)
- unterschiedliche Anzahl (N).

Darüber kann eine OFW

- mehrfach an verschiedenen Stellen angezapft werden.

Auf diese Weise ergibt sich eine sehr *große Anzahl von Varianten* und eine *Anpassung an viele Anwendungsmöglichkeiten.*

Das Prinzip des OFW-Filters kann somit gemäß Bild 6.3.19 dargestellt werden. Es entspricht den im Abschnitt 7.9. genauer behandelten Transversalfiltern. Hier erfolgt eine vereinfachte Betrachtung. Auf der Senderseite wirkt jeder Finger wie ein Sender mit der Erregerstärke a_k. Je nach seinem Anschluß an die obere bzw.

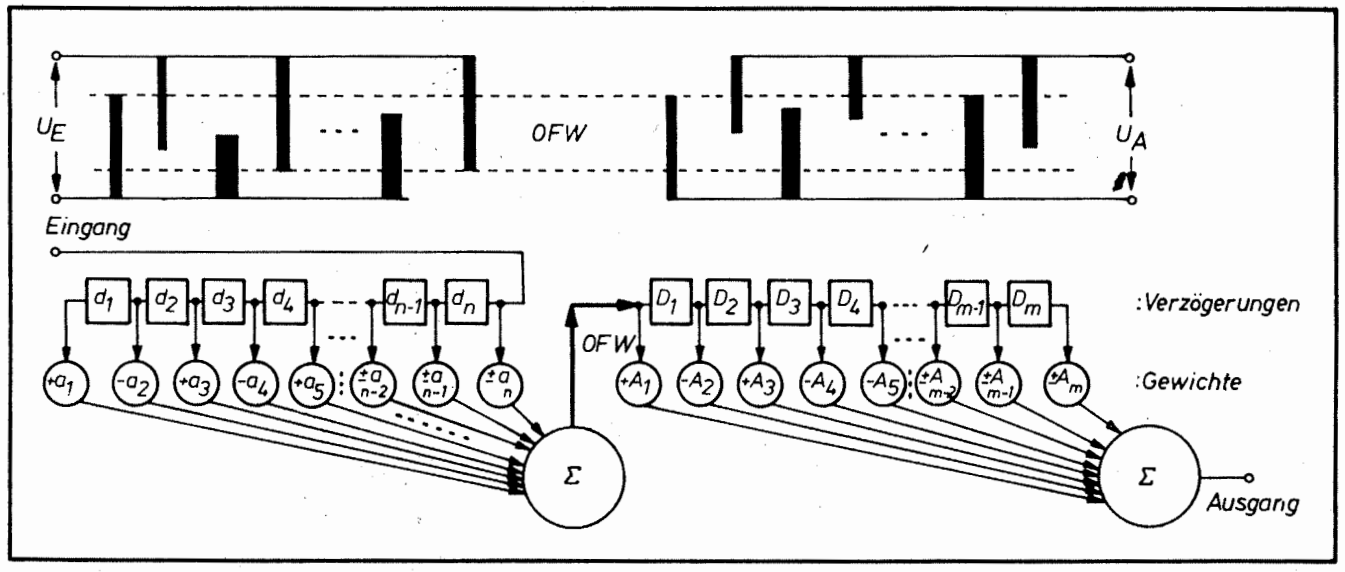


Bild 6.3.19. Schema und Ersatzschaltbild eines OFW-Filters. Die Größen d_k bzw. D_k sind die Abstände zwischen den einzelnen Fingern und damit, elektrisch gesehen, Verzögerungen; die Größen $\pm a_k$ bzw. $\pm A_k$ sind die Gewichte der Finger (entsprechend ihrer Größe). Die kleinen Buchstaben beziehen sich auf den Sende- und die großen auf den Empfangswandler

untere Seite erhält er das positive oder negative Vorzeichen. Entsprechend der Ausbreitungsrichtung auf den Empfänger wirkt jeder Sender um einen Abstand d_k verzögert. Die einzelnen Schallwellen (OFW) werden dann aufsummiert. Jeder Empfängerfinger wandelt die OFW in elektrische Spannungen zurück, die dann entsprechend verzögert (gemäß der Lage $= D_k$) mit dem Gewicht $\pm A_k$ aufsummiert werden.

Zur Vereinfachung werde angenommen, daß eine harmonische Schwingung

$$E(t) = \cos \omega t \tag{19}$$

die Senderwandler erregt. Jeder einzelne Finger erzeugt dann eine Schallwelle

$$S_k(t) = -(-1)^k a_k \cos(\omega t + \varphi_k), \tag{20}$$

wobei für den Winkel φ_k gilt:

$$\varphi_k = \sum_{p=k}^{n} \omega d_p . \tag{21}$$

Damit gilt

$$S(t) = -\sum_{k=1}^{n} (-1)^k a_k \cos \omega (t + \sum_{p=k}^{n} d_p) . \tag{22}$$

Hieraus läßt sich der Frequenzgang bezüglich der Frequenz $f(\omega = 2\pi f)$ berechnen. Zur Vereinfachung werden dabei mit d äquidistante Finger angenommen, dann gilt in Analogie zur FOURIER-Transformation (1.4.6.) und mit der Geschwindigkeit v der OFW

$$F(f) = \sum_{k=1}^{n} \pm a_k \exp(-\mathrm{j}2\pi f k d/v) . \tag{23}$$

Hieraus folgt die Frequenz f_0 der maximalen Überlagerung zu

$$f_0 = v/2d . \tag{23}$$

Bezogen auf diese Frequenz, läßt sich mit der Fingerzahl n eine Größe

$$x = n \frac{f - f_0}{f_0} \tag{24}$$

einführen, und für die weitere Vereinfachung $a_k = \pm 1$ folgt die Näherung in der Umgebung von f_0:

$$F(f) = \frac{\sin x}{x} . \tag{25}$$

Diese Funktion besitzt periodische Nullstellen bezüglich f, und zwar im Abstand

$$\Delta f = f_0/n . \tag{26}$$

Für gleichartige, äquidistante, gleichgewichtete Wandlerstrukturen bei Sender und Empfänger gemäß Bild 6.3.18a) ergibt sich eine Übertragungsfunktion mit dem Quadrat von Gl. (25). Sie zeigt Bild 6.3.18 c).

Auffällig bezüglich der *Frequenzcharakteristik* von OFW-Filtern ist, daß sie nur *Nullstellen* und *keine Pole* besitzen (vgl. 2.4.3.). Sie können weiter *keine Gleichwerte übertragen*. Es lassen sich folglich auch *keine Tiefpässe realisieren*.

6.3.5.2. Bauelemente mit OFW

Die vorangegangenen Betrachtungen zeigen mit geringen Abwandlungen. daß zwischen Zeit- und Frequenzverhalten von OFW-Filtern die FOURIER-Transformation

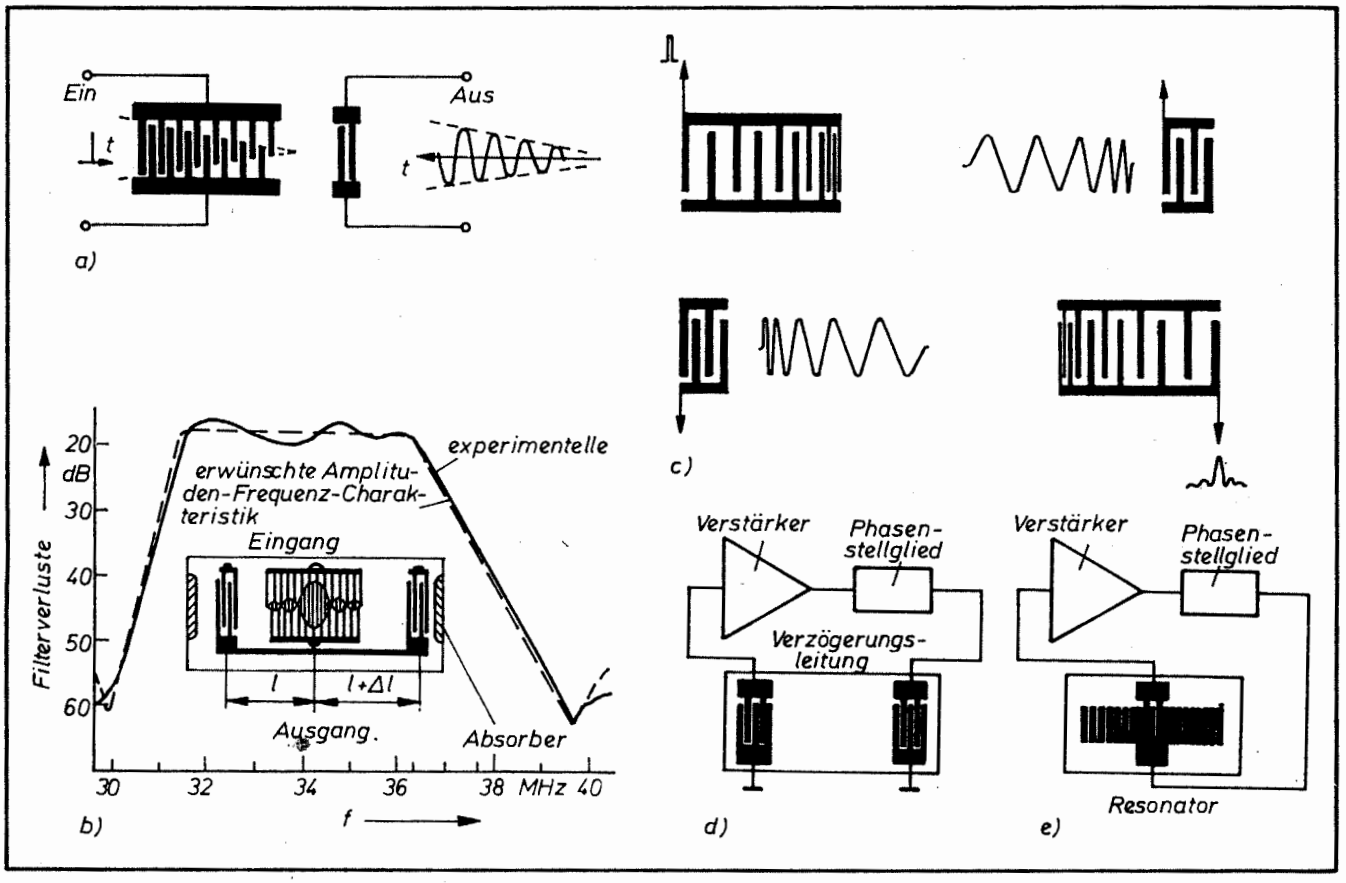


Bild 6.3.20. Anwendungsbeispiele von OFW

a) Die Fingerstruktur bestimmt den Ausgangsimpuls eines Filters. So ist über die FOURIER-Transformation der Zusammenhang von Frequenzgang und Fingerstruktur unmittelbar anschaulich verständlich

b) Beispiel eines Fernsehfilters, wobei der IDW des Senders gemäß $(\sin x)/x$ gestaltet ist und etwa eine Rechteckbandbreite bewirkt. Die beiden Empfänger-IDW besitzen einen um Δl unterschiedlichen Abstand. Dadurch entsteht das Minimum bei 39,5 MHz

c) Prinzip der Impulskompression; oben Erzeugung des Chirp-Signals, unten Rückwandlung mit Impulsverkürzung

d) Oszillator mittels Verzögerungsleitung

e) Oszillator mit Resonator

gilt, wobei in etwa die *Gewichte der Finger* den FOURIER-*Koeffizienten* entsprechen. Dies ermöglicht einen anschaulichen, näherungsweisen Entwurf, wie ihn Bild 6.3.20a) zeigt. Bei äquidistanten, gleich breiten Fingern (etwa $\lambda/4$) wird die Überlappung der Länge so gewählt, wie die Ausgangsfunktion des OFW-Filters bei Eingabe eines kurzen Einzelimpulses sein soll. Nach diesem Prinzip ist z. B. die *Filterstruktur für einen Fernseh-ZF-Kanal* gemäß Bild 6.3.20b) erkennbar. Der Senderwandler enthält eine Struktur $(\sin x)/x$, die einem etwa rechteckigen Frequenzbereich von 31 bis 37 MHz entspricht. Rechts und links befinden sich breitbandige Empfängerwandler. Sie sind jedoch um den Abstand Δl unterschiedlich weit entfernt. Durch ihre additive Zusammenschaltung wird das gewünschte Minimum bei 39,5 MHz bewirkt.

Für *verschiedene Anwendungen*, u. a. Rauschbefreiung, ist es sinnvoll, Einzelimpulse breitbandig und zeitverlängert zu übertragen. Dies leistet eine Senderstruktur, wie sie links oben Bild 6.3.20c) zeigt. Es wird eine frequenzmodulierte Welle abgestrahlt, die den Fingerabständen entspricht und mit hoher Frequenz beginnt und mit tiefer aufhört. Dieses Signal kann an Stelle des Einzelimpulses am breitbandigen IDW, rechts oben in Teilbild c), abgenommen und übertragen werden. Wird dieses Signal nun einem zweiten OFW-Filter (unten in c)) zugeführt, so kann der *komprimierte Impuls* wieder zurückgewonnen werden. Die Breitbandsignale entsprechen in etwa den Fledermaussignalen, die wie Zirpen klingen (engl.: chirp). Deshalb wird auch von *Chirp-Signalen* bzw. Chirp-Modulation in diesem Zusammenhang gesprochen. Die

Rückwandlung zum zeitkomprimierten Impuls wird auch Zeitkompression genannt. Sie hat u. a. in der Radartechnik zur genaueren Festlegung des Zeitpunktes eines Impulses Bedeutung. Mit OFW-Filtern sind *Impulsverkürzungen* bis 10000 erreichbar.

In Bild 6.3.20d) und e) sind zwei Arten von *Oszillatoren* mit OFW dargestellt. Der eine (d)) nutzt die Laufzeit aus. Der zweite (e)) die Resonanzverschärfung vieler IDW gemäß Gl. (24) bis (26). Mit solchen Schaltungen werden Güten von mehreren Tausend erreicht. Die Temperaturstabilität liegt in der Nähe der von Quarzen.

Neben den wenigen hier dargestellten *Anwendungen* der OFW gibt es viele weitere. Sie betreffen u. a.

- Verzögerungsleitungen und darauf beruhende Speicher,
- Multiplexer (durch Umschalten zwischen IDW),
- programmierbare Filter (Umschalten von Fingern im IDW),
- Wandlung von Ultraschallbildern,
- akustische Verstärker (durch Wechselwirkung von Elektronen, die mit den OFW laufen),
- Lichtablenkung (Wechselwirkung von Licht mit stehenden OFW),
- Bildaufnahme (durch Wechselwirkung von Abbild und OFW).

6.4. Leistungselektronik

Hier werden nur jene ausgewählten Gebiete behandelt, welche die Bauelemente betreffen, dagegen nicht jene, welche die spezifischen anwendungstechnischen Probleme, z. B. der Starkstromtechnik, HF-Erwärmung, Leistungsröhren, Teilchenbeschleuniger, Motorensteuerung, Regelungstechnik, KFZ-Elektronik usw. betreffen.

6.4.1. Geschichtliche Entwicklung und Anwendungsgebiete

Der Begriff Leistungselektronik existiert seit etwa 1975. Vorher gab es eine Vielzahl anderer Begriffe für dieses Gebiet, u. a.: Gleichrichtertechnik (ab 1902), Trockengleichrichter (1930), Stromrichtertechnik (1936), statische Umformertechnik (1948), industrielle Elektronik (1965). Die Zahlen geben den ungefähren Zeitpunkt der Einführung des jeweiligen Begriffes an. Einen Überblick zur Technikentwicklung gibt Tab. 6.4.1. Die Zunahme der mit Thyristoren maximal schaltbaren Leistungen sowie der maximalen Spannung zeigt Bild 6.4.1a).

In der Energieerzeugung und -verteilung sind Umformung von Spannungen, Frequenzen, Gleich- und Wechselstrom, Anzahl der Phasen sowie Erzeugung von kapazitiven und induktiven Blindlasten die wesentlichen Aufgaben. Verdichtet faßt dies Bild 6.4.1b) schematisch zusammen. Der zentrale Oberbegriff ist dabei Stromrichter mit folgender Einteilung:

Stromrichter:
- Gleichrichter
- Wechselrichter
- Umrichter für Wechselströme
- Umrichter für Gleichströme.

Eine weitere Einteilungsmöglichkeit ergibt sich aus der Herkunft der *Taktfrequenz*:

- fremdgeführt von:
  - Netzfrequenz
  - Last
- selbstgeführt, d. h. eigenständige Löschung (Zwangskommutierung)
- ohne Kommutierung, also Schalter und Steller für Wechsel- bzw. Drehstrom.

Entsprechend dieser Einteilung gelten die Diagramme in Bild 6.4.1c) bis f) für verschiedene Anwendungsgebiete. Die meisten Begriffe dürften dabei allgemein verständlich sein. Für einige sind jedoch Erklärungen notwendig.

Tabelle 6.4.1. Entwicklung der Leistungselektronik

1898	GRAETZ-Brückenschaltung
1902	COOPER — HEWITT entdecken, daß Strom in einer Hg-Gasentladung nur von Stahl nach Quecksilber und nicht umgekehrt fließt (Quecksilberdampf-Eisen-Gleichrichter). Anwendungen bis ca. 1950
1911	Einführung des Quecksilberdampf-Gleichrichters in Deutschland durch BÉLA SCHÄFER
1914	Erfindung der Gitterstreuung durch LANGMUIR
1923	Steuerung bei Quecksilberdampf-Gleichrichtern wird eingeführt.
1923	Erfindung der Phasenanschnittsteuerung durch TOULON und MÜLLER
1923	Beginn der Versuche von ALAXANDERSON und PRINCE für Wechselrichter mit Gegentakt
1927	Thyratron
1929	Berliner S-Bahn wird mit 131 gesteuerten Hg-Gleichrichtern und einer Gesamtleistung von 170 MW bei 800 V ausgerüstet. Anlagen sind z. T. noch heute in Betrieb.
1930	Kupferoxidul-Gleichrichter erlangt Bedeutung
1930	Selengleichrichter
1932	Ignotron
1932	Rückstromdiode für Wechselrichter durch PETERSEN
1934	Pumpenlose Hg-Dampfgleichrichter in mehranodigen Eisengefäßen
1934	Erster Direktumrichter auf $16^2/_3$ Hz zur Versorgung der Wiesentalbahn
1936	Erste Stromrichterlokomotive der Höllentalbahn (50 Hz, 20 kV). Enthielt gesteuerte Hg-Dampfgleichrichter, die bis 1960 erfolgreich in Betrieb waren.
1939	Excitron
1944	Kontaktumformer erlangen zeitweilig Bedeutung
1952	Erste HGÜ-Übertragung
1953	Leistungsdiode
1958	Entwicklung des Thyristors durch General Electric
1960	Erste Wechselrichter auf Thyristor-Basis
1970	Wasserkühlung bei Thyristoren
1974	Einführung der Neutrondotierung von Telefunken
1975	Rückwärtsleitender Thyristor
1977	Betrieb der Hochspannungs-Gleichstrom-Übertragungsanlage zwischen Dänemark und Norwegen 500 kV/1000 A
1978	Erste abschaltbare Thyristoren (GTO) in der Fertigung

Sektorsteuerung: Bei der Anschnittsteuerung wird immer der hintere Teilabschnitt der Sinuskurve belastend ausgetastet. Dies führt zu einer induktiven Belastung des Netzes. Die Sektorsteuerung tut das Gegenteil (Belastung des Vorderteils) und kompensiert so die Blindlast.

HGÜ = Hochspannungsgleichstromübertragung: Wechselstrom hat den Vorteil, leicht transformierbar zu sein. 50 Hz wurden dabei in Europa und 60 Hz in den USA/Kanada als wirtschaftliches Optimum festgelegt. Höhere Frequenzen bedeuten kleinere Transformatoren, aber höhere Verluste. Bei der Übertragung über große Entfernungen treten zusätzliche Verluste durch Induktivitäten und Kapazitäten der Leitungen auf. Sie lassen sich bei Gleichstrom voll vermeiden. So stellt sich bei Entfernungen über etwa 800 km und Leistungen über 500 MW die HGÜ als wirtschaftlicher heraus. Bei Seekabeln kann schon bei kleineren Entfernungen und kleineren Leistungen die HGÜ vorteilhaft sein. Vorteilhaft ist die HGÜ weiter für die Verkopplung frequenzasynchroner Netze (USA — Kanada). Typische Anlagen weisen folgende Beispiele auf:

Kasachstan, UdSSR: 2414 km $\pm$ 750 kV bis zu 6000 MW

Pacific Interfie, USA: 1320 km $\pm$ 400 kV bis zu 1440 MW

Skagerrak, Dänemark — Norwegen: 110 km Freileitung, 130 km Seekabel $\pm$ 500 kV bis zu 500 MW.

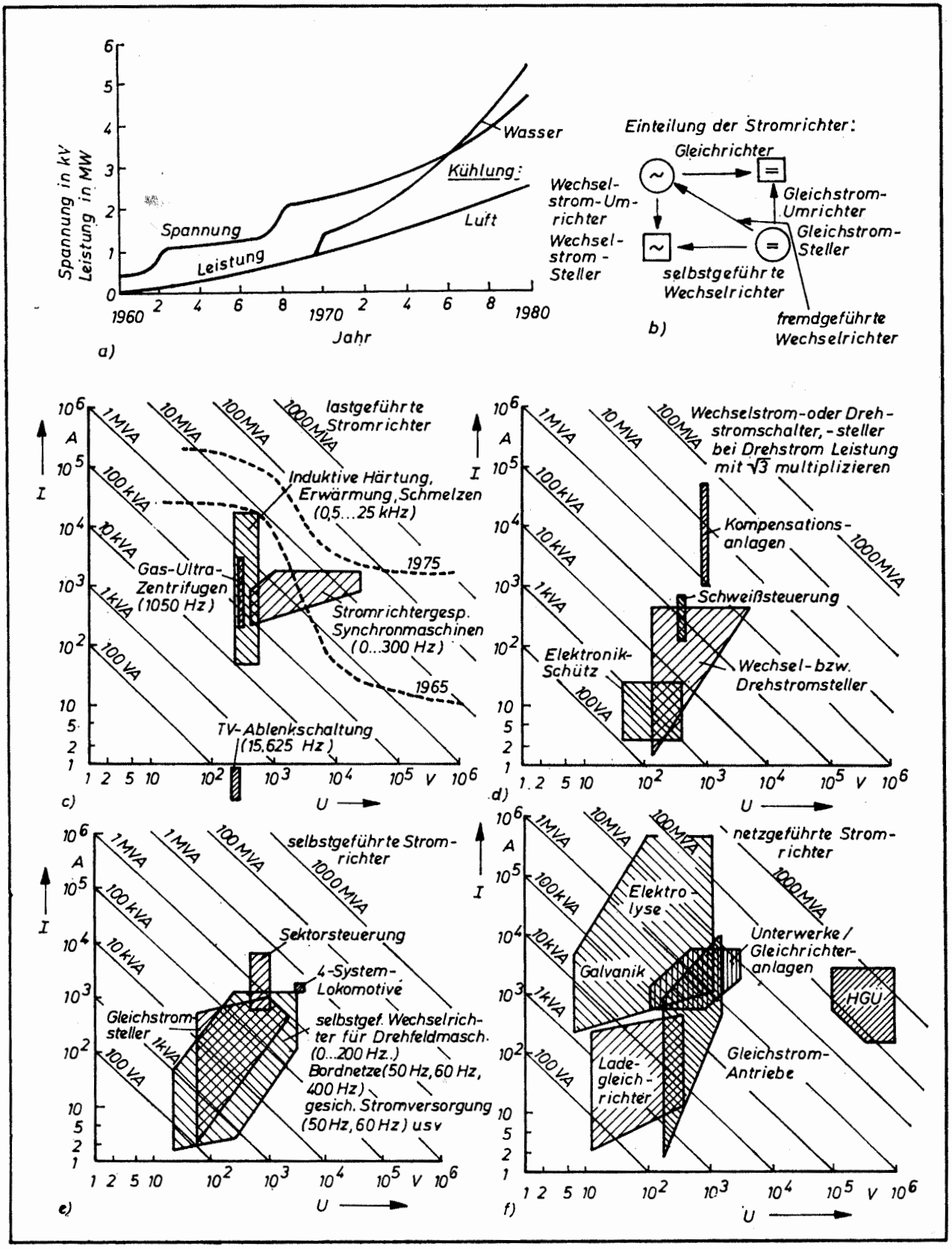


Bild 6.4.1. Zur Leistungselektronik

a) Leistungsgrenzen der Thyristoren

b) Einteilung der Stromrichter aus der Sicht der Quellen

c) bis f) Arbeitsbereiche verschiedener Typen von Stromrichtern

USV = Unterbrechungsfreie Stromversorgung: Bezüglich verschiedener Computersysteme, z. B. in Überwachungssystemen, werden hohe Forderungen an die Betriebssicherheit gestellt. Hier darf z. T. nicht einmal eine Halbschwingung ausfallen. Im wesentlichen haben sich hierfür als Reserve selbstgeführte Wechselrichter bewährt.

6.4.2. Vierschichtdiode als Grundlage

Es gibt eine *Vielzahl* von Halbleiterbauelementen der Leistungselektronik. Die meisten leiten sich *von der Vierschichtdiode*, auch Thyristordiode oder SUS (= silicon unilateral switch) genannt, ab. Sie sei daher als erste anhand von Bild 6.4.2 behandelt. Die Schichtenfolge ist *pnpn* mit den drei *pn*-Übergängen I bis III. Das genaue Dotierungs-

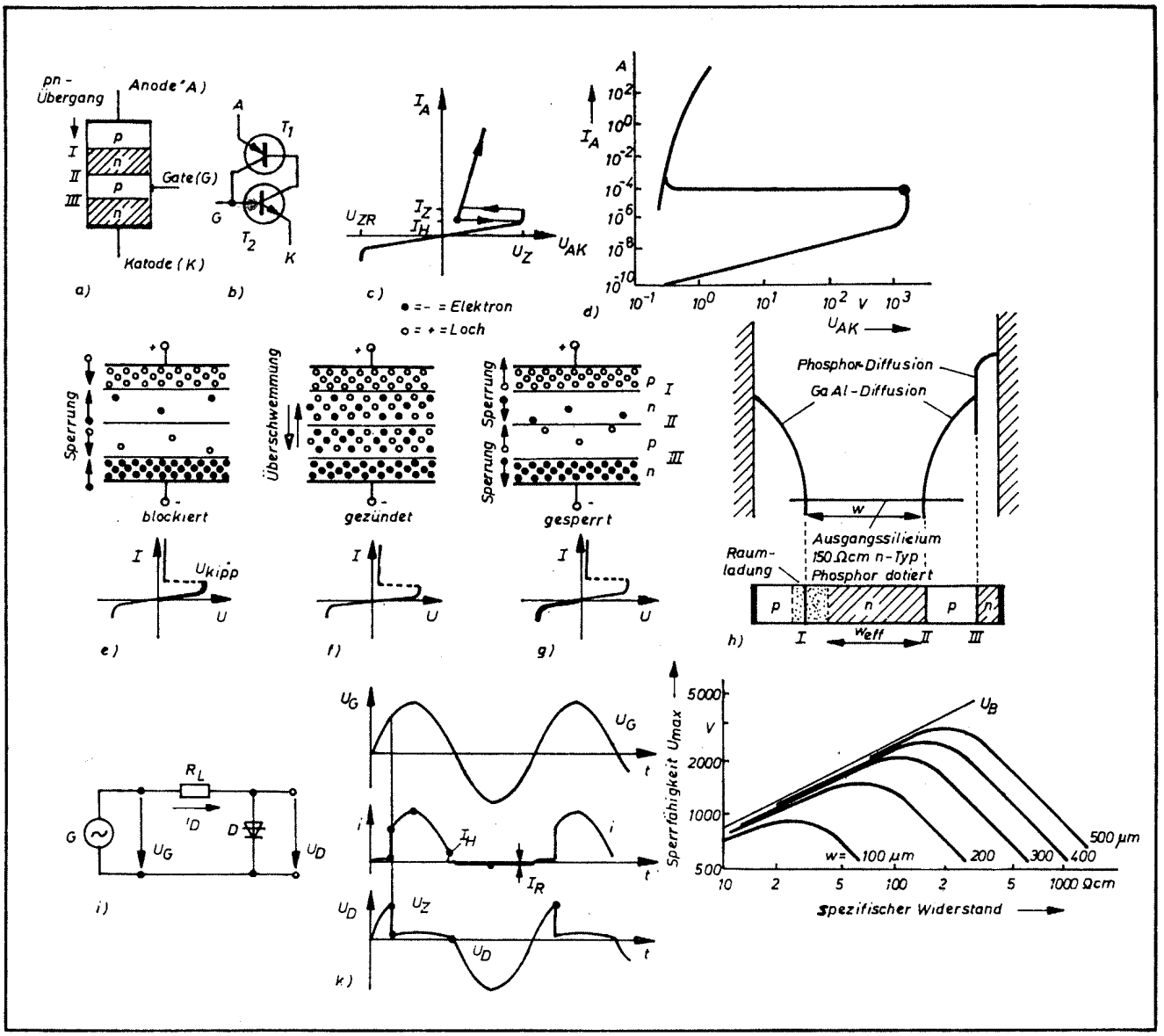


Bild 6.4.2. Aufbau und Kennlinien einer Vierschichtdiode (ohne G = Gate) bzw. eines Thyristors

a) Schichtfolge
b) Ersatzschaltung aus zwei Transistoren
c) Betriebskennlinie
d) Logarithmisch dargestellter positiver Kennlinienbereich
e) bis g) Verteilung der Ladungsträger in den drei Schichten: e) blockiert, f) gezündet, g) gesperrt
h) Dotierungsprofil einer Vierschichtdiode bzw. eines Thyristors und Definition von Weite und effektiver Weite
i) Schaltung und Zeitverläufe bei Verwendung einer Vierschichtdiode im Wechselstromkreis mit Lastwiderstand
j) Sperrspannung als Funktion der Dotierung und Weite W des inneren n-Gebietes (vgl. h))

profil mit etwa geometrisch richtigen Abmessungen zeigt Bild 6.4.2g). Insbesondere die innere *n*-Zone ist schwach dotiert. Der äußere *p-Anschluß heißt Anode.* An ihm liegt im Normalbetrieb eine positive Spannung. Der äußere *n-Anschluß heißt Katode.* Beide Begriffe sind aus der Röhrentechnik abgeleitet.

Zunächst wird der *Sperrbetrieb* der Vierschichtdiode betrachtet. An der Anode liegt dann eine negative Spannung, der *pn*-Übergang II, ist leitend, und die beiden anderen (I und III) sind gesperrt. Die Ladungsträgerkonzentration zeigt Teilbild g). Der Sperrstrom steigt entsprechend dem üblichen Exponentialgesetz nur wenig an, und zwar bis jene Spannung U_{ZR} erreicht ist, bei der für beide *pn*-Übergänge der *Avalanche-Durchbruch* eintritt. Dann erfolgt ein steiler Stromanstieg. Diese *Durchbruchspannung* liegt heute bei einigen kV.

Liegt eine *positive Spannung an der Anode*, so sind zunächst die *pn*-Übergänge I und III leitend, und II ist gesperrt. Für den *pn*-Übergang II gilt daher eine ähnliche Kennlinie wie im Sperrbetrieb (vgl. Teilbild e)). Die *Diode ist blockiert.* Bei U_Z wird der *Avalanche-Durchbruch* des *pn*-Überganges II erreicht. Jetzt tritt eine Änderung gegenüber dem Sperrbetrieb auf. Der Avalanche-Strom von *pn*-Übergang II bildet in den beiden in Durchlaßrichtung betriebenen *pn*-Übergängen I und III verstärkt Ladungsträger. Sie wirken dann wie ein Emitter bei Transistoren. Die von ihnen neu erzeugten Ladungsträger diffudieren in den „Kollektorübergang" II und lassen daher die Spannung an ihm zusammenbrechen. Die Diode ist gezündet (Teilbild f)). Dies erfolgt mit dem waagerechten Übergang in Bild 6.4.2b). Er geht bis zur kleinen Rest- bzw. Kniespannung U_R von rund 0,7 V. Der *Strom steigt* von hier ab entsprechend einer in Durchlaßrichtung gepolten Diode *steil* (exponentiell) *an.* Die Spannung wächst bei hohen Strömen auf etwa 1,5 V. Auf dieser *Kennlinie sind zwei typische Ströme* zu unterscheiden:

I_Z: *Zündstrom.* Bei ihm erfolgt der spontane Übergang von der Sperrkennlinie zur Durchlaßkennlinie. Die zugehörige Spannung ist die Zündspannung U_Z. Die typische Einschaltzeit liegt bei 1 µs.

I_H: *Haltestrom.* Ist die Vierschichtdiode im Durchlaßbetrieb, so läßt sich der Strom bis auf diesen Wert absenken. Wird er unterschritten, so kann nicht mehr der Avalanche-Mechanismus im *pn*-Übergang II aufrechterhalten werden, und die Vierschichtdiode schaltet spontan in die Sperrkennlinie zurück. Die typische Ausschaltzeit beträgt ca. 10 µs.
Der Haltestrom ist stark von der Temperatur abhängig. Beträgt er z. B. bei Zimmertemperatur 8 mA, so steigt er für −50 °C auf 17 mA und fällt für +120 °C auf 4 mA ab.

Es ist somit offensichtlich, daß die Vierschichtdiode *aus der gezündeten Durchlaßcharakteristik nur durch Unterschreiten des Haltestroms* herauskommt. Dieser Fakt ist besonders wichtig beim Thyristor.

Das Verhalten der Vierschichtdiode in einer Wechselstromschaltung mit Last zeigt Bild 6.4.2i).

In Bild 6.4.2d) ist der typische Spannungs-Stromverlauf des I. Quadranten der Vierschichtdiode vergrößert und mit logarithmischer Teilung der Achsen wiedergegeben. Aus der Ladungsträgerverteilung im gezündeten Zustand (Teilbild f)) wird deutlich, daß bei Stromfluß die Vierschichtdiode von einer normalen Diode mit *psn*-Struktur formal nicht zu unterscheiden ist. Daraus wird auch deutlich, daß die *Leitfähigkeit* von *der inneren n-Schicht* und die Spannungsbelastung der Diode von deren Weite begrenzt wird (Teilbild j)). Vergleiche hierzu auch Bild 3.2.8 und 3.4.4. Für das *niedrig dotierte Gebiet* besitzt die Dotierung durch *Neutronenstrahlung* mit der

Umwandlung

$$n + {}^{30}Si \dashrightarrow \gamma + {}^{31}Si \xrightarrow[2{,}6\,h]{} {}^{31}P + \beta^-$$

große Bedeutung.

Das geschilderte Verhalten der Vierschichtdiode läßt sich vollkommen analog mittels der Transistorersatzschaltung von Bild 6.4.2b) beschreiben. Wenn an A eine positive Spannung liegt, sind die Kollektorstrecken beider Transistoren gesperrt, die Emitterdioden in Durchlaßrichtung geschaltet. Sobald bei einem der beiden Transistoren die Kollektordurchbruchspannung erreicht wird, entsteht im anderen in der Emitterdiode ein Strom, der dessen Kollektorstrecke entsprechend öffnet. Dieser Strom fließt wiederum in die Emitterdiode des ersten und öffnet ihn weiter. So schaukelt sich der Prozeß in kürzester Zeit bis zum vollständigen Öffnen auf.

6.4.3. Thyristor

Das Wort Thyristor leitet sich aus *Thyr*atron und Trans*istor* ab. Das Bauelement besitzt nämlich Eigenschaften des nunmehr veralteten Thyratrons (spezielle Gasentladungsröhre mit Verstärkereigenschaften) und des Transistors, insbesondere bezüglich des Materials Silicium. Deswegen heißt der Thyristor im Angloamerikanischen auch SCR (= silicon controlled rectifier). Von der Vierschichtdiode unterscheidet er sich formal nur durch das herausgeführte Gate (Bild 6.4.2a)). *Mittels dieses Gates kann der Thyristor vom blockierten in den gezündeten Zustand überführt werden.* Die Katoden-Anordenspannung muß dabei größer als etwa 10 V und kleiner als U_Z sein. Wird dem Gate dann ein positiver Strom I_G zugeführt, so erfolgt die *Zündung*, sobald etwa

$$I_G \geqq I_Z \tag{1}$$

gilt. Der Gatestrom erzeugt dann so viele Ladungsträger, daß der Durchbruch des sperrenden *pn*-Übergangs II erfolgt. Hierfür genügt sogar ein nur kurzer Stromimpuls, der den Zündvorgang einleitet. Es muß nur ein Zustand erreicht werden, wo der fließende Strom größer als der Haltestrom I_H ist. Bleibt ein gewisser Gatestrom bestehen, so genügt sogar ein kleiner *Haltestrom* I_{H_0}. Es gilt etwa

$$I_{H_0} \geqq I_H - I_G\,. \tag{2}$$

Die verschiedenen Zündpunkte mit einem Gatestrom und bei verschiedenen Betriebsspannungen zeigt Bild 6.4.3a). Bezüglich des Zustromes ist die Eingangskennlinie des Gates zu beachten (Bild 6.4.3b)). Leider wirkt sich ein Gatestrom erheblich auf den

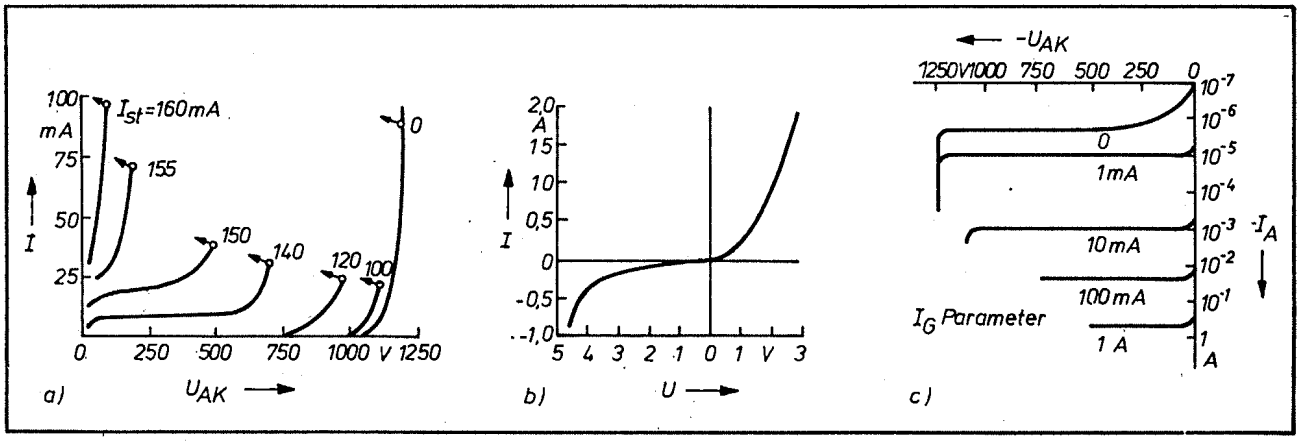


Bild 6.4.3. Thyristorkennlinien

a) Blockierungslinien und Zündpunkte bei verschiedenen Gateströmen

b) Gatekennlinie

c) Sperrströme bei verschiedenen Gateströmen

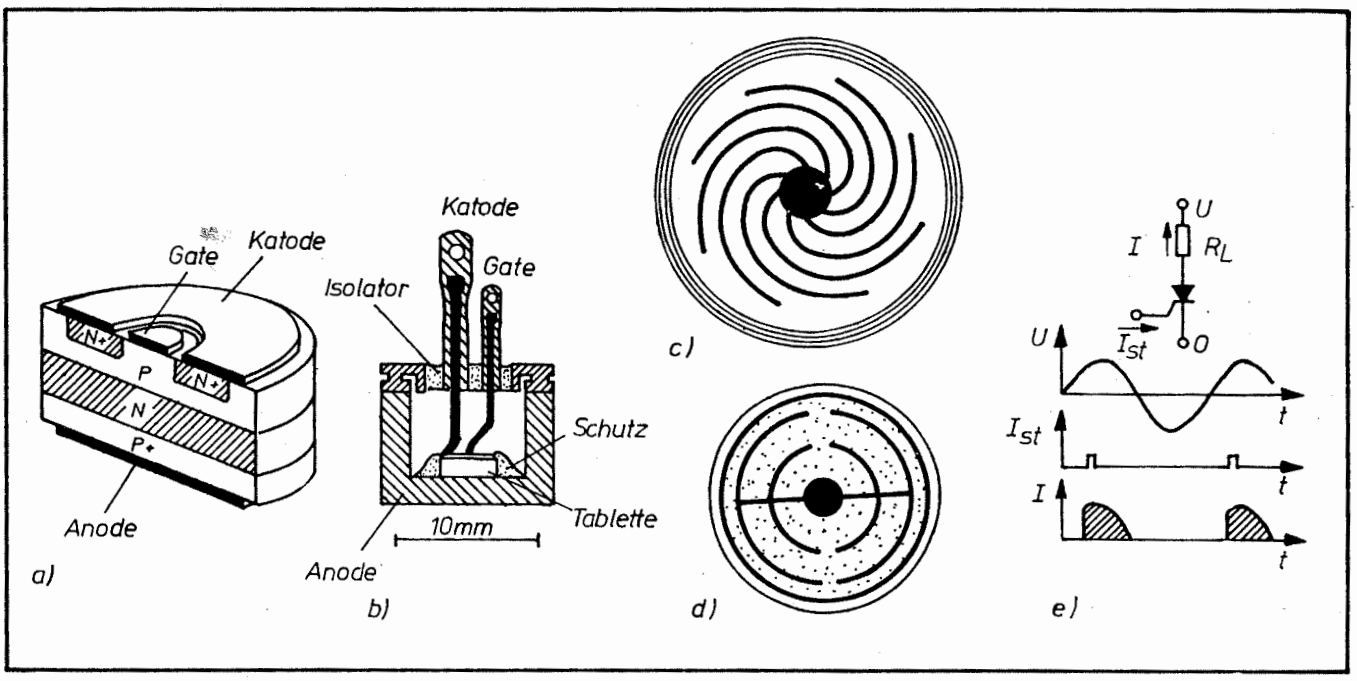


Bild 6.4.4. Aufbau und Verhalten von Thyristoren

a) Tablettenschnitt
b) Einbau einer Tablette in ein Press-fit-Gehäuse
c) radiales Streifengate
d) Gate mit Fingerstruktur
e) Ströme und Spannungen beim Thyristor

Sperrstrom, d. h. bei umgekehrt gepoltem Thyristor, aus, wie aus Bild 6.4.3c) ersichtlich. Deshalb ist es notwendig, den Gatestrom möglichst während der Umpolung auszuschalten. Den Aufbau eines Thyristors zeigt Bild 6.4.4a), den Einbau ins Gehäuse Bild 6.4.4b). Infolge der hohen Stromdichten im Transistor ist die *Kontaktierung zum Gehäuse sehr sorgfältig*, und zwar sowohl thermisch als auch mechanisch, bezüglich der *thermischen Wechselbeanspruchung* zu gestalten. Bezüglich der hohen Spannungen ist den Kanten besondere Aufmerksamkeit zu schenken. Deshalb die Schutzschicht in Teilbild b). Häufig werden auch die Kanten der Thyristortablette abgeschrägt.

Der *Zündimpuls pflanzt sich* vom Gate aus *relativ langsam* mit ca. 50 m/s *fort*. Dies macht es für die großen Thyristoren und für hohe Schaltgeschwindigkeiten erforderlich, die Zündung über die gesamte Tablette verteilt einzuleiten. So entstanden *verzweigte Strukturen* für das Gate, wie sie beispielhaft Bild 6.4.4c) und d) zeigen. Hierdurch können auch bei größten Leistungen Einschaltzeiten um 1 µs erreicht werden.

Soll der Thyristor ausgeschaltet werden, so muß kurzfristig der Strom unter den Haltestrom gesenkt werden. Dies erfolgt günstig, wenn eine Wechselspannung verwendet wird. Das Schema zeigt Bild 6.4.4e). Mit dem Zündimpuls wird der Thyristor gezündet, im Nulldurchgang schaltet er sich aus und geht in den blockierten bzw. gesperrten Zustand über. Durch die zeitliche Lage des Zündumpulses zur Wechselspannung kann die Stromfläche und damit die Leistung an R_L gesteuert werden.

Das *Ausschalten des Thyristors* kann auch durch *kurzfristiges Absenken des Stromes*, z.B. als Folge einer Gegenspannung oder Unterbrechung, erfolgen. Die Ausschaltzeit muß aber so lang sein, daß alle Ladungsträger ausgeräumt sind. Um diese Zeit zu verkürzen, wird der Thyristor oft mit Gold $1:10^6$ dotiert. Die *Freiwerdezeiten* liegen bei 5 bis 50 µs. Weiter läßt sich die Freiwerdezeit durch *negative Gateimpulse* verkürzen. Dies führt zum GATT bzw. GATO (= gate assisted turn off Thyristor). Solche Thyristoren können bei ca. 10 kHz betrieben werden.

Die hergestellten Thyristoren besitzen Flächen von 1 bis ca. 1500 mm² und damit max. Ströme von 0,5 bis ca. 1000 A. Die Stromdichten liegen bei einigen hundert A/cm².

Es gibt neben den statischen *Grenzen* eines Thyristors zwei wichtige *dynamische*:

- Die *maximale Spannungsanstiegsgeschwindigkeit* du/dt ist im blockierten Zustand zu beachten. Wird sie überschritten, so zündet der Thyristor auch ohne Zündimpuls. Infolge der Aufladung der Sperrschichtkapazität C bewirkt sie nämlich einen inneren Strom der Größe $C \cdot du/dt$, der den Zündstrom ersetzen kann. Eine praktische Grenze für du/dt liegt bei 20 V/µs.
- Die *maximale Stromanstiegsgeschwindigkeit* gilt für den Zündvorgang. Sie führt nämlich an den verschiedenen inneren Widerständen des Thyristors zu zusätzlichen Verlustleistungen, die den Thyristor überlasten können. Mittlere Werte liegen hier bei 20 A/µs.

6.4.4. Familie der Thyristorbauelemente

Bild 6.4.5 gibt einen Überblick zu den verschiedenen Thyristorbauelementen. Sowohl bei den Thyristordioden (Vierschichtdioden) als auch bei den Thyristoren (Thyristor-

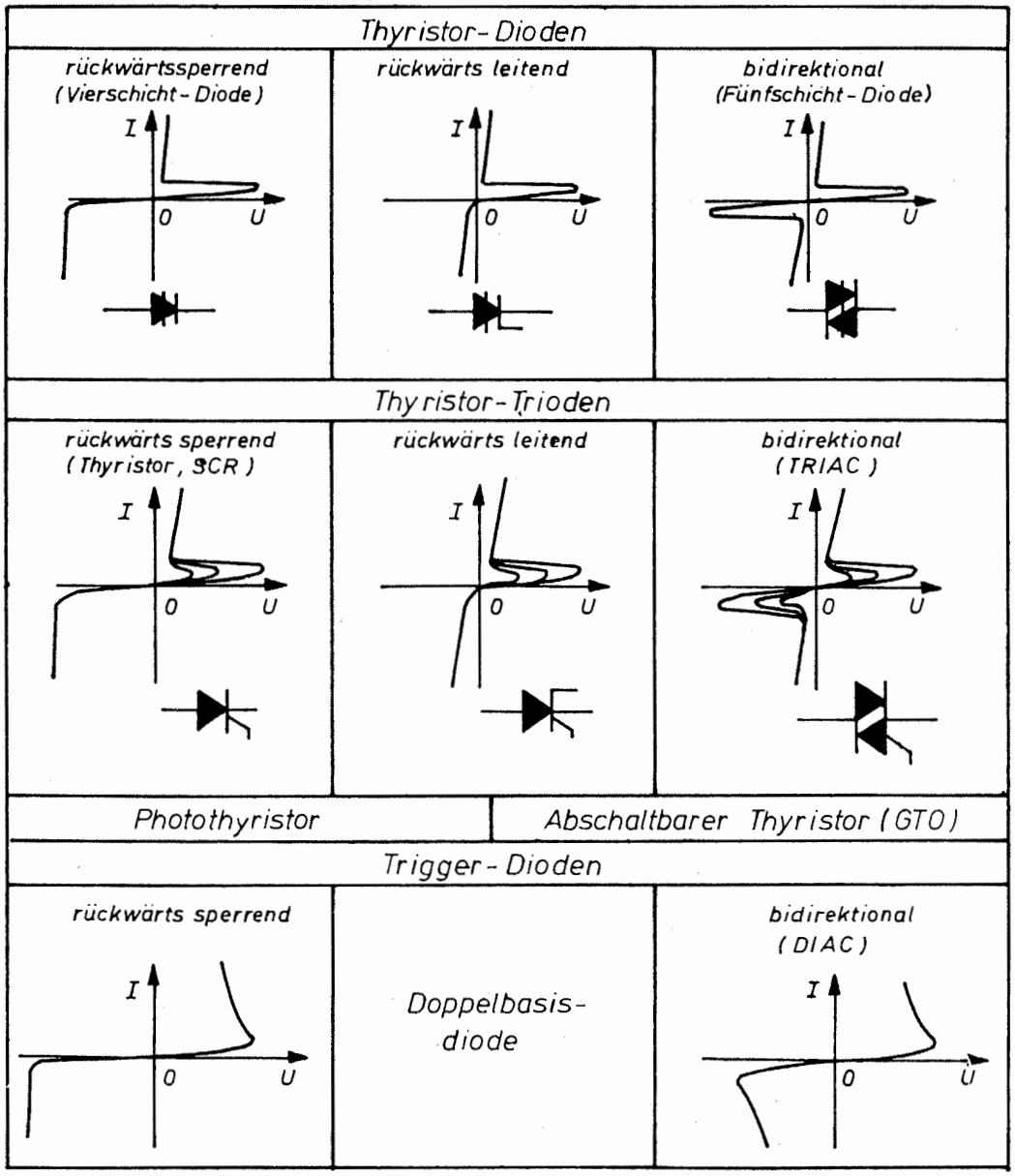


Bild 6.4.5. Schematischer Überblick zu Thyristorbauelementen und verwandten Bauelementen

trioden) kommen noch je zwei Varianten hinzu, also existieren drei Typen:

- rückwärtssperrend (bisher besprochener Typ)
- rückwärtsleitend
- bidirektional.

Die *rückwärtsleitende* Vierschichtdiode und der rückwärtsleitende Thyristor (RLT) entstehen, wenn auf der Thyristortablette zusätzlich eine Diode integriert wird. In der Regel erfolgt dies auf einem äußeren Radialsegment, wobei das Flächenverhältnis Thyristor:Diode bei etwa 3:1 bis 1:1 liegt. Sie haben Vorteile bei einigen spezifischen Schaltaufgaben und besitzen Freiwerdezeiten von etwa 3 µs. Daher entfällt bei ihnen der Schaltungsaufwand für GATO.

Die *bidirektionalen* Typen entstehen formal, wenn zwei rückwärtssperrende Typen antiparallel zusammengeschaltet werden. Erfolgt dies in einem Bauelement, so führt es zur Fünfschichtdiode SBS (= silicon bilateral switch) bzw. zum TRIAC (= triode alternating current), auch Zweiwegthyristor oder bidirektional schaltender Thyristor genannt. Den entsprechenden Aufbau der Fünfschichtdiode zeigt Bild 6.4.6a) und b). Daraus geht hervor, daß drei Schichten, nämlich *npn*, beiden Teildioden gemeinsam sind und nur die zusätzlichen *p*-Gebiete am Ober- bzw. Unterteil der Diode jeweils zur Hälfte eindiffundiert werden müssen.

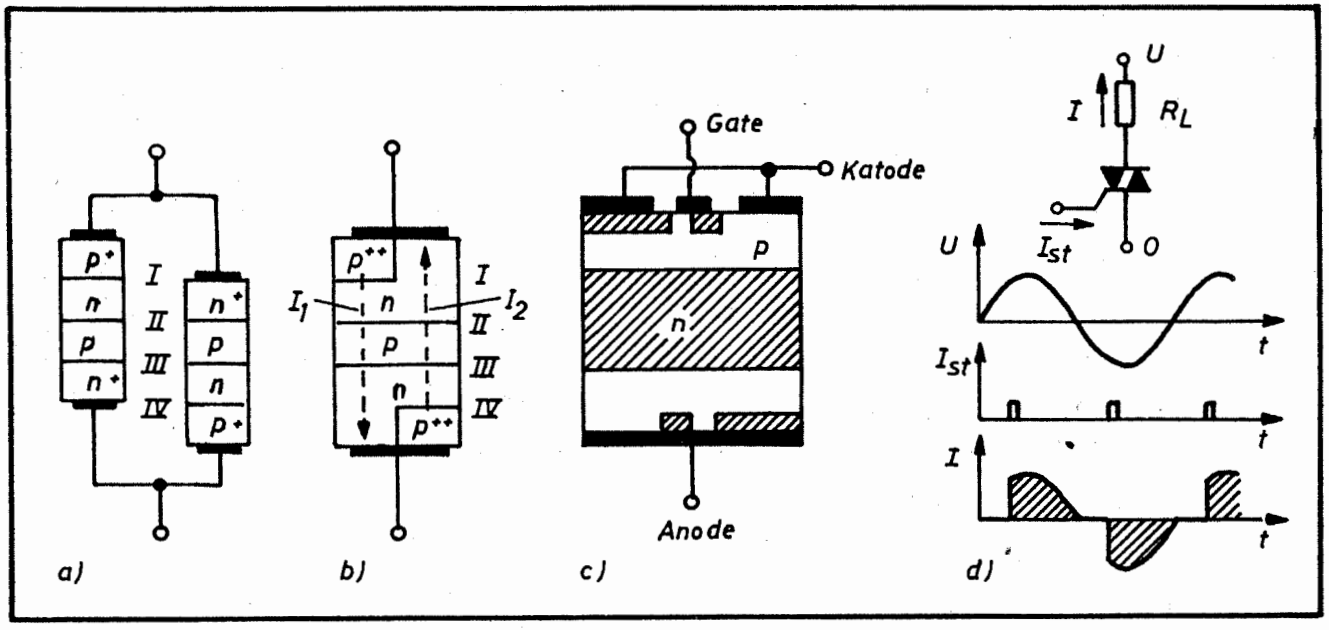
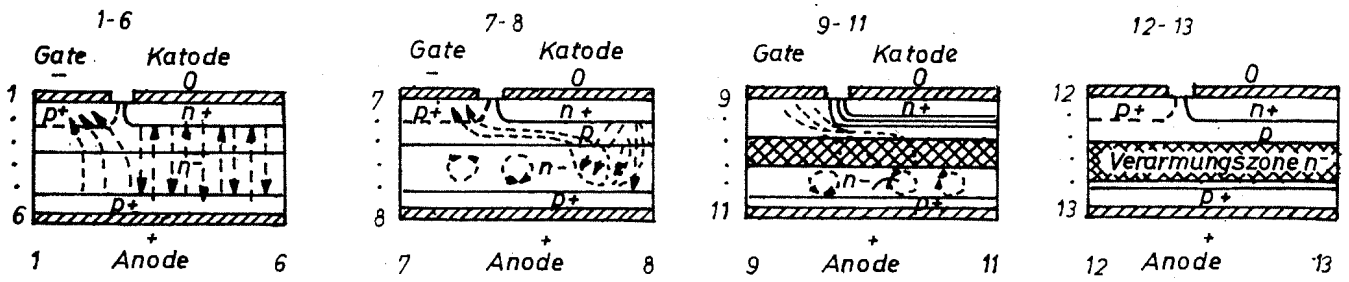


Bild 6.4.6. Zum Aufbau des TRIAC

a) Zwei antiparallel geschaltete Vierschichtdioden und
b) ihre Zusammenfassung zur Fünfschichtdiode
c) Aufbau eines TRIAC mit zusätzlich eingefügten *n*-Gebieten zur Zündung in beiden Richtungen mit beliebiger Polarität
d) Schaltung des TRIAC und zugehörige Ströme und Spannungen. Die Zündimpulse können für eine oder beide Halbwellen auch negativ sein

Beim TRIAC muß zusätzlich ein gemeinsamer Gateanschluß realisiert werden. Dies ist nur über Umwege möglich. Vorteilhaft für verschiedene Schaltungen ist es, wenn die Zündung des TRIAC sowohl mit positiven als auch negativen Impulsen (je nach Wunsch) möglich ist. So entsteht ein Gateanschluß mit zusätzlichem *pn*-Übergang (N_G), also mit Hilfstransistor gemäß Bild 6.4.6c). Die verschiedenen Betriebszustände des TRIAC sind kompliziert. Es sei deshalb auf [T 3, Teil 2; M 25] verwiesen.

Das Prinzip der Phasenanschnittsteuerung im Vergleich zum normalen Thyristor (Bild 6.4.4e)) zeigt Bild 6.4.6d). Die hier positiv gezeichneten Zündimpulse können ebensogut für eine Halbwelle oder für beide negativ sein.



1. Negative Vorspannung an das Gate angelegt.
2. Löcher von der Anode werden über das Gate abgeführt.
3. Injektion von Elektronen hört in der dem Gate benachbarten Zone auf.
4. Nachdem die Zonen in der Umgebung des Gates an Ladungsträgern verarmt sind, werden Löcher aus entfernteren Zonen abgeführt.
5. In nicht mehr injizierten Zonen sinkt die Ladungsträgerdichte durch Rekombination.
6. Immer größer werdende Zonen des n-Emitters hören auf, Elektronen zu injizieren.
7. Löcher müssen von den leitenden Zonen aus immer größerere Entfernung in Querrichtung durch die p-Basis zurücklegen, um an den Gateanschluß zu gelangen.
8. Der Anodenstrom wird in den vom Gateanschluß am weitesten entfernten Zonen zu Fäden hoher Stromdichte zusammengedrängt. Dies ist die kritischste Phase des Abschaltens über das Gate, da örtlich sehr schnell ansteigende Temperaturen den Kristall zerstören können, wenn die Stromfäden nicht schnell unterbrochen werden. Ein schnelleres Unterbrechen der Stromfäden ist durch eine höhere Gatesperrspannung möglich, aber die höchstmögliche Spannung ist die Gate-Katoden-Durchbruchsspannung.
9. Sind die letzten Stromfäden unterbrochen, so hört die Elektroneninjektion auf, und an der Gate-Katoden-Sperrschicht bildet sich eine Verarmungszone aus.
10. Die Verarmungszone beginnt sich auf die in Durchlaßrichtung blockierende Grenzschicht auszudehnen.
11. Der Katodenstrom setzt aus; aber dadurch, daß Träger aus dem Plasma der n-Basis in die Verarmungszone hineindiffundieren, fließt weiterhin ein Anoden-Gate-Strom. Dieser sinkt exponentiell, da die Trägerkonzentration durch Rekombination abnimmt.
12. Mit steigender Impedanz des Bauelementes nähert sich die Anodenspannung ihrem stationären Wert.
13. Die negative Gatespannung kann entfernt werden, wenn die Trägerkonzentration auf einen Wert abgesunken ist, der unter dem für das Fortbestehen des Rückkopplungsmechanismus erforderlichen liegt, und die Anodenspannung nahe ihrem stationären Wert ist.

Bild 6.4.7. Vereinfachte Zusammenfassung der Vorgänge bei einem abschaltbaren Thyristor

In Sonderfällen ist es wünschenswert, den gezündeten Thyristor ohne Abschaltung der Anodenspannung wieder löschen zu können. Dies führt zu *abschaltbaren Thyristoren* (Bild 6.4.7), auch GTO(= gate turn off) bzw. früher Transwitch genannt. Theoretisch und für kleine Leistungen bestanden hier schon länger Möglichkeiten. Erst in den letzten Jahren ist dies auch für hohe Leistungen gelungen. Die Zusammenhänge sind relativ kompliziert, deshalb der Literaturverweis auf [G 5]. Zur Löschung werden Spannungen von ca. 5 bis 10 V und Ströme vom 0,2- bis 1fachen des Betriebsstromes für etwa 0,1 µs benötigt. Dies erfordert spezielle Ansteuerschaltungen. Noch wesentlicher als beim normalen Thyristor ist eine stark verzweigte Struktur des Gates.

Als letztes eigentliches Thyristorbauelement sei der *lichtzündbare* Thyristor oder *Fotothyristor* genannt. An der Stelle des Gates ist er lichtdurchlässig aufgebaut, und vom Licht werden Ladungsträgerpaare erzeugt, die den Zündvorgang einleiten. Zur Erreichung einer hohen Empfindlichkeit sind Sondermaßnahmen notwendig, damit alle erzeugten Trägerpaare wirksam werden. Es muß eine bestimmte Elektroden- und Fensteranordnung gewählt werden (Bild 6.4.8a) und b)).

6.4.5. Sonderbauelemente für Thyristoren

Für das Zünden von Thyristoren sind Impulsspannungen von einigen V und Ströme von einigen hundert mA für Impulsdauern von einigen µs notwendig. Diese Impulse können mit einfachen Schaltungen auf vielfältige Weise erzeugt werden, u. a. mittels

- Vierschicht- bzw. Fünfschichtdioden (6.4.2. u. 6.4.5.)
- Lawinentransistoren (5.7.5.)
- Unipolartransistoren.

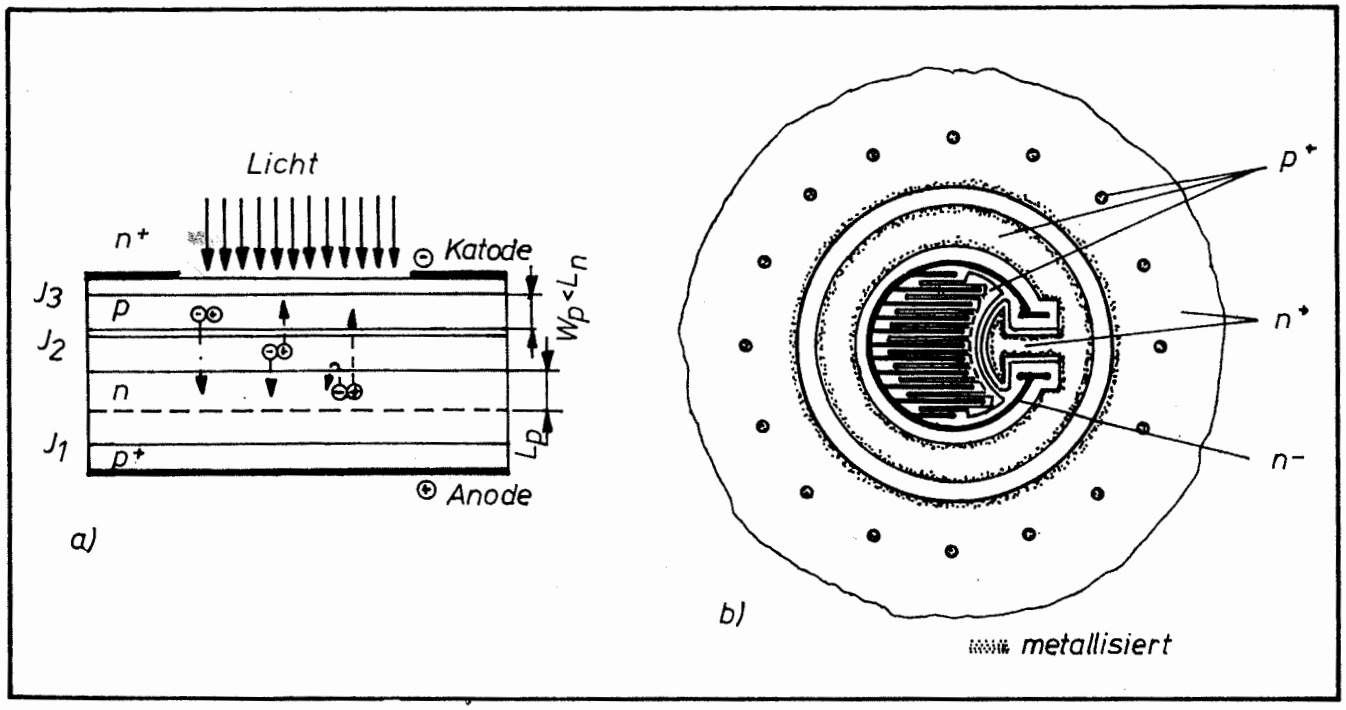


Bild 6.4.8. Zum Fototransistor

a) Prinzip der Lichtzündung

b) Beispiel einer Ausführung des Fensters zur Erhöhung der Lichtempfindlichkeit

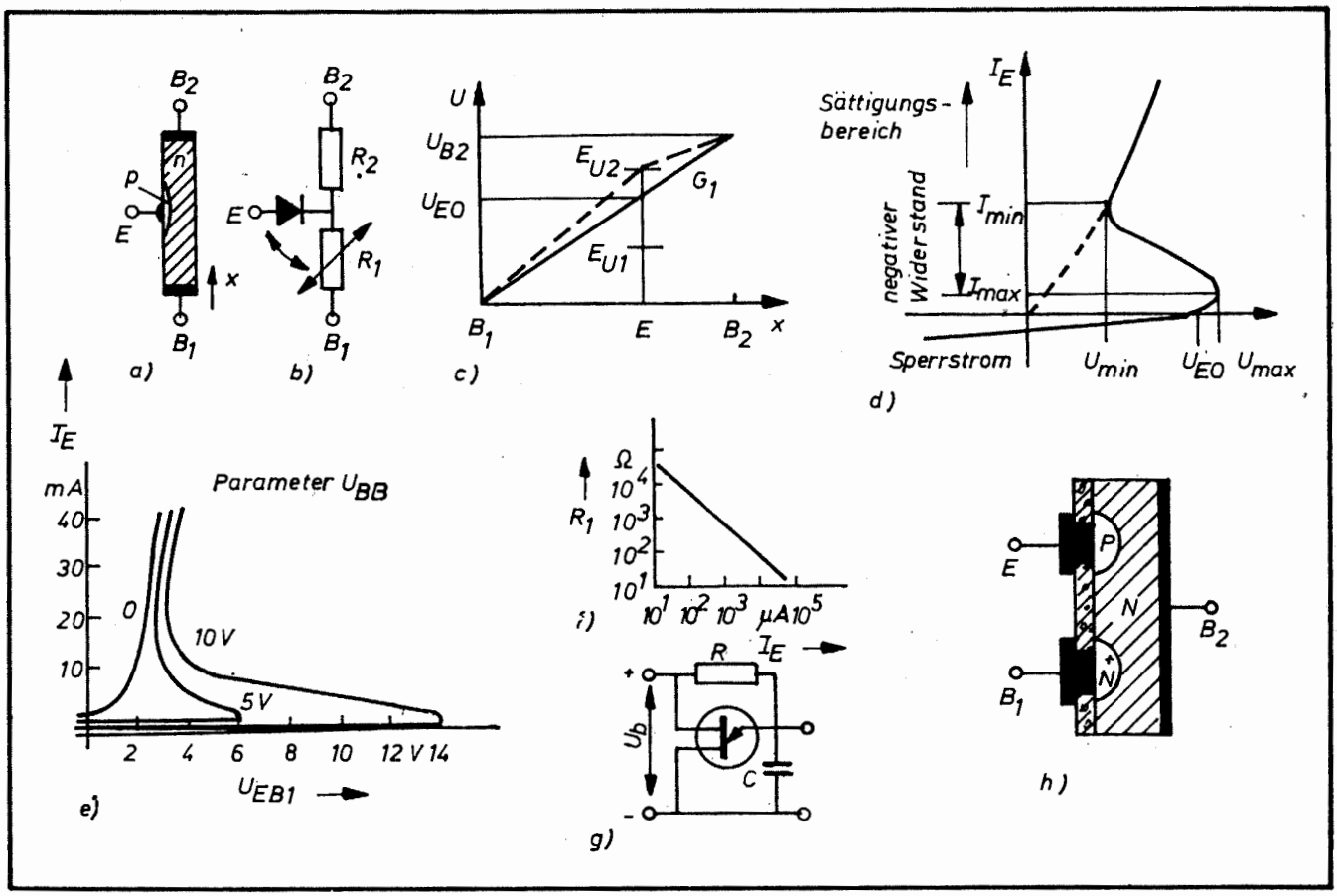


Bild 6.4.9. a) Prinzipieller Aufbau, b) Ersatzschaltung des Unipolartransistors; c) Konstruktion der Kennlinien (d); e) ihre Abhängigkeit von der Spannung U_{B2}; f) Widerstandsverlauf für $U_E > U_{E0}$; g) Schaltung zur Erzeugung von Sägezahnschwingungen; h) Planar-Unipolartransistor

Hier ist also nur noch der letzte Typ zu behandeln, für den auch die Namen *Unijunction-Transistor* oder *Doppelbasisdiode* verwendet werden (Bild 6.4.9a)). Gegenüber einem FET (leider auch Unipolartransistor genannt) sind das n-Material hochohmiger, die Emitterelektrode nur kleinflächig aufgesetzt und das Restvolumen hinter dem Emitter relativ groß. Obwohl er ohnehin anders betrieben wird, wäre dadurch die für den FET typische Kanaleinschnürung auch nicht möglich. Die beiden Elektroden B_1 und B_2 sind genau so wie beim Sperrschicht-FET sperrschichtfrei kontaktiert.

Die Wirkung des Unipolartransistors kann durch die Ersatzschaltung von Bild 6.4.9b) beschrieben werden, wobei der Doppelpfeil die Wirkung der Emitterdiode auf die Größe des Widerstandes R_1 ausdrücken soll. Zur Erklärung dieser Wechselwirkung diene Teilbild c). Die Spannungsverteilung im n-Material bei offenem Emitter zeigt die Gerade G_1. Dabei ist vorausgesetzt, daß an B_2 die gegenüber B_1 positive Spannung U_{B2} angelegt ist. An der Emitterdiode, genauer am pn-Übergang, entsteht die Spannung U_{E0}. Ihr Wert hängt nur von der Lage des Emitters in bezug auf B_1 und B_2 ab. Je nach Herstellung und Anwendung ist die Lage unterschiedlich.

Solange am Emitter ein negativeres Potential als U_{E0} angelegt ist (z.B. U_{U1} im Teilbild c)), ist der pn-Übergang gesperrt. Es fließt der sehr geringe Sperrstrom entsprechend Bild 6.4.9d). Der lineare Spannungsabfall von B_1 und B_2 bleibt also erhalten.

Sobald die Emitterspannung den Wert U_{E0} überschreitet, wird der pn-Übergang leitend, und es fließt ein Durchlaßstrom. Dabei wird die untere Strecke $E - B_1$ von Defektelektronen aus dem Emitter überschwemmt, und deshalb nimmt ihr Widerstand ab. Der Potentialverlauf längs des Unipolartransistors verändert sich gemäß der Lage E_{U2} in Bild 6.4.9c).

Die Spannung zwischen Emitter und Basis 1 durchläuft mit zunehmendem Emitterstrom sehr bald ein Maximum U_{max}, um dann auf das Minimum U_{min} abzusinken (Bild 6.4.9d)). Dies entspricht einer Kennlinie mit einem differentiellen, negativen Widerstand. Das Verhältnis $U_{max}:U_{min}$ liegt bei etwa 5. Beim Minimum geht der Verlauf in den einer normalen Diode über (gestrichelt gezeichnet). Wie dieser Übergang erfolgt, wird besonders mit Bild 6.4.9e) deutlich. Hier wird der Einfluß der Spannung U_{EB2} gezeigt.

Die Abhängigkeit des Widerstandes R_1 vom Durchlaßstrom zeigt Bild 6.4.9f). Mit echt guter Näherung gilt also

$$I_E R_1 \approx 1\,\mathrm{V}\,.$$

Es ist eine Widerstandsänderung von mehr als $1:10^3$ erreichbar.

Aus den Kennlinien von Bild 6.4.9d) und e) geht leicht der Zusammenhang mit den Triggerdioden gemäß Bild 6.4.6 hervor. Eine Antiparallelschaltung von zwei Unipolartransistoren führt dann auch zur Kennlinie der DIAC (= diode alternating current).

Eine *spezielle Anwendung* des Unipolartransistors zeigt Bild 6.4.9g). Sie erzeugt auf einfachste Weise Kippschwingungen. Zur Erklärung ihrer Funktion sei ein entladener Kondensator angenommen. Es fließt dann nur der sehr kleine Sperrstrom durch den Unipolartransistor. Über den Widerstand R wird der Kondensator mit der Zeitkonstante RC aufgeladen. Dies geschieht so lange, bis der Emitter das Potential U_{max} erreicht. Dann nimmt der Widerstand der Strecke Emitter — Bais 2 plötzlich ab und entlädt den Kondensator auf U_{min}. Hier wird die Emitterstrecke wieder hochohmig, und der genannte Vorgang wiederholt sich periodisch. So entsteht die Sägezahnspannung zwischen U_{max} und U_{min} mit der Frequenz, die durch R und C bestimmt ist.

Bild 6.4.9h) zeigt den planaren Aufbau eines Unipolartransistors.

6.4.6. Ausgewählte Schaltungen

Die Leistungselektronik ist nicht das eigentliche Anliegen dieses Buches. Deshalb werden hier nur wenige spezifische Schaltungen beschrieben, die die Funktion der behandelten Bauelemente etwas vertiefen. Sie sind in Bild 6.4.10 zusammengestellt. Teilbild a) zeigt einen extrem einfachen *Sägezahngenerator* mit einer Vierschichtdiode. Die Betriebsspannung U_b hat hier lediglich oberhalb der Zündspannung zu liegen. R und C bestimmt dann die Frequenz der Kippspannung.

Teilbild b) zeigt einen sehr einfachen SCHMITT-*Trigger* (vgl. 7.3.8.3.), wie er mit einem kleinen, abschaltbaren Thyristor realisiert werden kann. Mittels U_T kann das Tastverhältnis der Ausgangsrechteckimpulse eingestellt werden.

Teilbild c) zeigt eine einfache Schaltung, wie mit einem *Thyristor* eine Gleichspannung *durch zwei Kontakte ein-* ($S1$) *und ausgeschaltet* ($S2$) werden kann. Es sei angenommen, der Thyristor befinde sich im blockierten Zustand. Durch kurzzeitiges Schließen von $S1$ fließt über den Widerstand R_1 der Zündstrom, und der Thyristor wird leitend. Dabei wird auch der Kondensator C_1 über R_2 aufgeladen. Wird dann

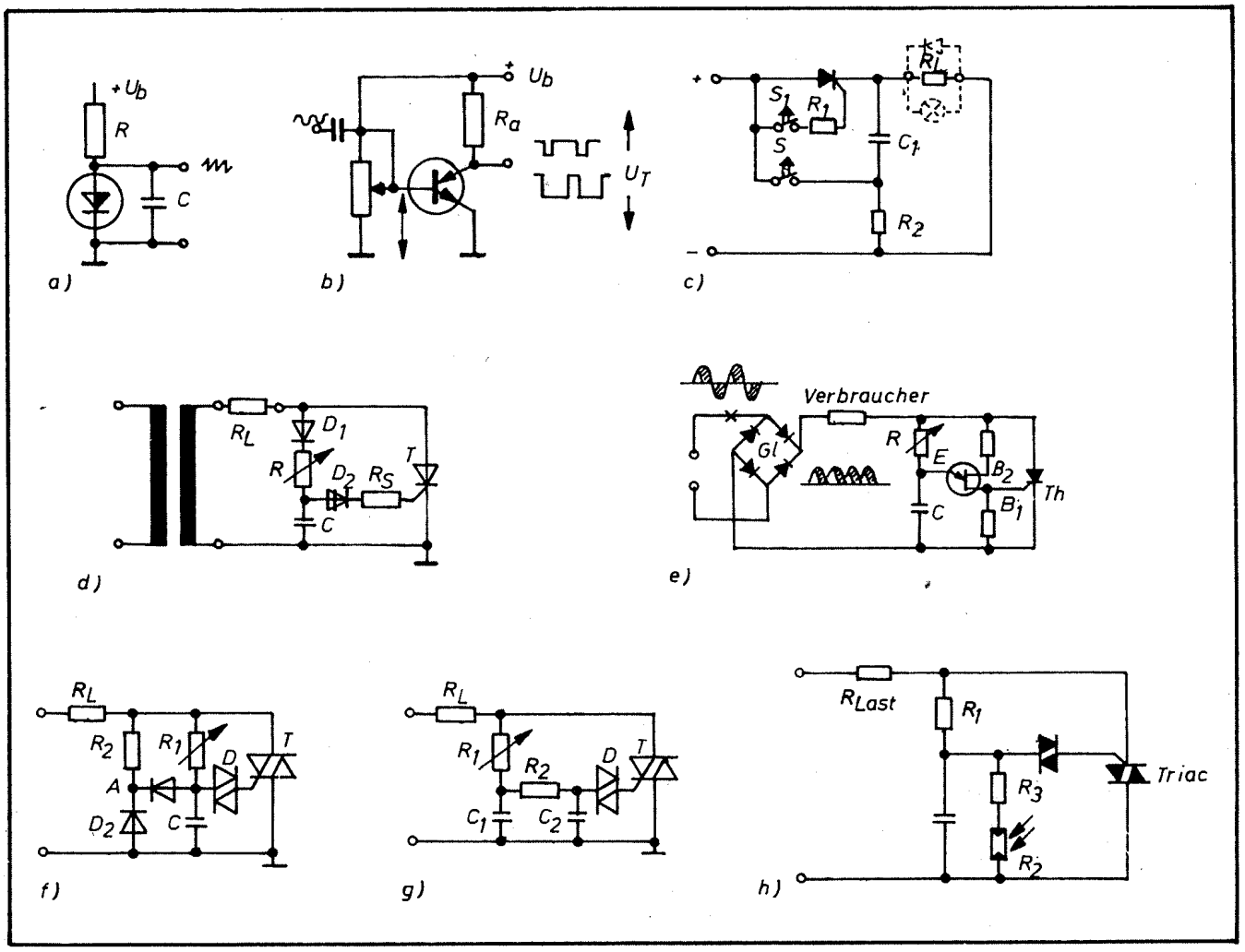


Bild 6.4.10. Zur Anwendung von Thyristoren

a) Vierschichtdiode als Sägezahngenerator
b) Doppelbasisdiode als SCHMITT-Trigger
c) Zündschaltung mit Thyristor, $S1$ zündet, und $S2$ löscht
d) Phasenanschnittsteuerung mit Triggerdiode $D2$ und Thyristor T
e) Doppelbasisdiode zur Zündung des Thyristors T, der in einer GRAETZ-Brücke liegt, um beide Halbwellen zu nutzen
f) und g) Zwei Schaltungen mit TRIAC = T und DIAC = D mit symmetrischem Schaltverhalten, das durch RC-Glieder erzwungen ist
h) Lichtgesteuerte TRIAC-Schaltung

der Schalter S_2 betätigt, so wird der Thyristor kurzfristig gesperrt und geht in den blockierten Zustand über. Da hierbei über S_2 ein beträchtlicher Strom fließt, kann es sinnvoll sein, hierfür eine analoge Thyristorschaltung einzusetzen. Die notwendige Größe für C_1 berechnet sich gemäß

$$C_1 \approx 1{,}5 t_F I_{max}/U_b;$$

darin bedeuten t_F die Freiwerdezeit, I_{max} den maximalen Laststrom und U_b die Betriebsspannung.

Teilbild d) zeigt eine thyristorgesteuerte *Einweg-Phasenanschnittschaltung*. Die Diode $D1$ sorgt dafür, daß nur der positiven Halbwelle über die Vierschichtdiode ein Zündimpuls zum Thyristor gelingt. Mit dem einstellbaren Widerstand R kann der Zeitpunkt des Zündpunktes gewählt werden. So steht an dem Lastwiderstand R_L ein mit R einstellbarer Ausschnitt aus der positiven Halbwelle zur Verfügung (vgl. Bild 6.4.4e)). Statt des Widerstandes R können auch ein Fotowiderstand oder ein Heißleiter eingebaut werden. Dann kann die Schaltung zur *Lichtsteuerung bzw. zur Temperaturregelung* verwendet werden.

Legt man die Schaltung von Teilbild d) in den Brückenzweig eines GRAETZ-*Gleichrichters*, so lassen sich beide Halbwellen für die Steuerung ausnutzen. Das entsprechende Prinzipschaltbild zeigt Teilbild e), diesmal allerdings mit einem Unijunction-Transistor.

Eine wesentliche Vereinfachung der Schaltungen von Teilbild d) entsteht, wenn **DIAC** und **TRIAC** verwendet werden. Hierbei entsteht jedoch im einfachsten Fall eine Einstellhysterese. Sie läßt sich mittels Dioden $D1$ und $D2$ in Teilbild f) bzw. durch die Zeitkonstante R_2C_2 in Teilbild g) vermeiden.

Eine einfache Lichtsteuerung zeigt schließlich noch Teilbild h). Ist der Fotowiderstand $R2$ stark beleuchtet, so kann der TRIAC nicht zünden. Je weniger Licht auf ihn fällt, desto mehr Strom fließt über Phasenanschnittsteuerung durch R_L.

6.4.7. Vergleich mit anderen Leistungsbauelementen

Für das Schalten und Verstärken von Leistungen stehen heute vor allem drei Bauelementtypen zur Verfügung:

- Leistungs-MOSFET (5.5.)
- Bipolare Leistungstransistoren (5.6. und 5.7.)
- Thyristoren (hier behandelt).

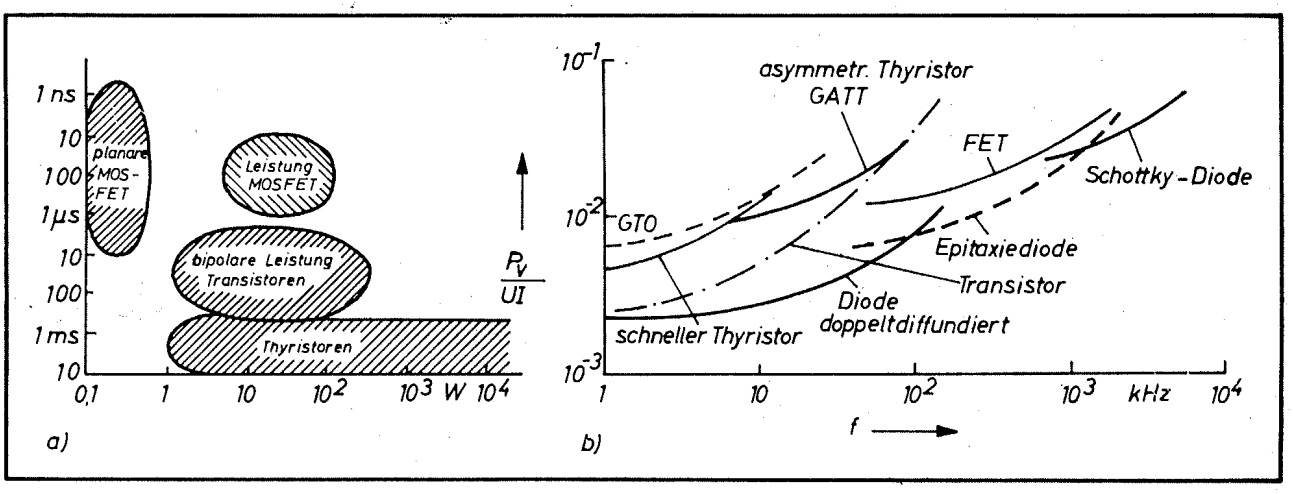


Bild 6.4.11. Zum Vergleich von Leistungsbauelementen

a) In der Zeit-Leistungs-Ebene

b) Als Funktion der Verlustleistung von der Frequenz

Sie besitzen unterschiedliche Eigenschaften, die u. a. aus Bild 6.4.11a) hervorgehen, also den Leistungs- und Frequenzbereich betreffen. Eine Ergänzung hierzu gibt Teilbild b), in dem für die verschiedenen Schaltbauelemente ausgewiesen ist, wie mit der Frequenz der relative Verlustanteil im Bauelement zunimmt.

Während mit beiden Transistortypen die Ströme im Prinzip analog geändert werden können, besteht beim Thyristor die entsprechende Möglichkeit nur durch Verändern von Impulslängen. Dieser Nachteil wird durch eine hohe Robustheit für viele Anwendungen ausgeglichen. So sind Thyristoren kurzzeitig bis zu 20fach überlastbar. Auf gleiche Leistung bezogen, benötigen sie die kleinste Chipfläche (FET die größte).

6.5. Sensoren

6.5.1. Überblick

Bauelemente oder Bausteine, die nichtelektrische Größen in elektrische Signale umwandeln, heißen heute *Sensoren* (engl. auch transducer). Dieser Begriff ist relativ neu. In etwa begann dieses Gebiet um 1950 mit dem Begriff *„elektronische Messung nichtelektronischer Größen"*. Die Sensoren hießen damals *Wandler* oder *Fühler* und später *Meßwertaufnehmer* bzw. *Meßgrößenumformer*.

Die Bedeutung der Sensoren hat in den letzten zwanzig Jahren ständig zugenommen. Dies bewirkten nach der Meßtechnik zunächst die Automatisierung, dann die Mikrorechentechnik und jetzt die Roboter. Für alle diese Anwendungsfälle und viele andere stellen sie die messende, kontrollierende, prüfende, kurz sensorische Verbindung der Elektronik zur Umwelt her, ganz ähnlich wie es die Sinnesorgane für den Menschen tun. Die Angaben über Produktionsvolumen (sowohl relativ als auch absolut) schwanken von Autor zu Autor ganz erheblich. Dies könnte daran liegen, daß der Begriff Sensor nicht eindeutig abgegrenzt ist; das betrifft sowohl seine Funktion als auch sein Anwendungsgebiet. In vielen Einschätzungen wird angegeben, daß etwa 80% aller Sensoren in drei großen Bereichen angewendet werden: Autoelektronik (45%), Konsumgüterelektronik (25%) und Meß-, Steuerungs- und Regelungstechnik (15%). In den USA betrug 1980 der Produktionswert „aller" Sensoren rund 300 Millionen Dollar. Die *Hauptmeßgrößen der Sensoren* betreffen u. a.:

> Temperatur, Druck, Position (Bewegung), Gewicht (Kraft, Moment, Druck), Durchfluß (Füllstand, Volumen), Feuchte, Stückzahl, Drehzahl, Winkel (Winkelgeschwindigkeit), Lichtgrößen (Strahlungsgrößen), Schwingungen (Schwingungsaufnehmer, Mikrophone, Tonabnehmer), Stoffgrößen (Masse, Dichte, pH-Wert, Viskosität, Zusammensetzung), magnetische Größen, . . .

Allein diese relativ ungeordnete Zusammenstellung zeigt Probleme der Sensoren: ihre Vielfalt, die Schwierigkeit ihrer Systematik und Abgrenzung: Sind z.B. Bildaufnahmeröhren Sensoren? Die Schwierigkeit wird noch größer, wenn man die *elektrischen Abbildgrößen* aufzuzählen versucht:

> Spannung, Strom, Frequenz, Widerstand, Impulszahl, Zeitabstände, Kapazität, Induktivität, . . .

Der *Mensch* mit seinen fünf *Sinnesleistungen* Sehen (10^7 bit/s), Hören ($5 \cdot 10^4$ bit/s), Riechen (50 bit/s), Schmecken (10 bit/s) und Fühlen ($2 \cdot 10^4$ bit/s) scheint dagegen viel übersichtlicher die Umwelt wahrzunehmen. Allerdings sind seine Sinnesorgane auch viel komplexer aufgebaut. Das Auge besitzt z.B. allein 10^8 Sinneszellen, und von ihm führen $2 \cdot 10^6$ Nervenbahnen weiter. Ja, hier lassen sich sogar die obigen absoluten Zahlen bezüglich der Leistungsfähigkeit angeben [V 11]. Dabei sind unter Fühlen die Sinneszellenleistung für Druck und Berührung ($2 \cdot 10^5$ bit/s), für Wärme und Kälte ($2 \cdot 10^3$ bit/s) und für Schmerz (100 bit/s) zusammengefaßt und nicht das emotionale Fühlen verstanden. Dem Menschen angepaßt ist daher die folgende Einteilung von Sensoren:

- *visuell*: punktförmig, zeilenförmig, als Matrix
- *taktil*: Taster, Dehnungsmeßstreifen usw.
- *auditiv*: Mikrofone, Schwingungsaufnehmer usw.

Diese bisherige, kurze Betrachtung zeigt, daß z. Z. der *systematische Stand der Sensoren* trotz ihrer sehr großen praktischen Bedeutung noch *nicht ausreichend geklärt* ist. Sowohl dieser Fakt als auch der breite Umfang und die große Vielfalt macht es notwendig, hier eine radikale Einschränkung vorzunehmen. So werden im folgenden *ausgewählte, typische* Sensoren auf der Basis von *Halbleitermaterialien* berücksichtigt. Zusätzlich muß auf jene Sensoren verwiesen werden, die bereits in den Abschnitten *Optoelektronik* (6.1.) sowie *Mechano-* und *Akustoelektronik* (6.3.) behandelt wurden. *Temperatursensoren* sind u. a. im Sinne der Temperaturabhängigkeit bei Halbleitermaterialien (3.1.), *pn*-Übergängen (3.2.) und bei den Heiß- bzw. Kaltleitern (5.1.3., 5.1.4.) möglich. Weiter muß auf den Abschnitt 7.7. verwiesen werden, in dem ähnliche Zusammenhänge aus der Sicht der *Meßtechnik* dargestellt werden. In [R 12, Bd. 5] sind darüber wesentliche Halbleitersensoren betrachtet. Ausgewählte geschichtliche Daten faßt Tab. 6.5.0 zusammen.

Hier seien noch zwei allgemeingültige Gesichtspunkte angefügt. Bild 6.5.1 zeigt den Zusammenhang einiger Begriffe und technischer *Bauelemente in der Umgebung des Sensors*. Es muß also zunächst einmal der nichtelektrische und der *elektrische Raum* unterschieden werden. Im *nicht-*

Tabelle 6.5.0. Auswahl zur Geschichte von Sensoren und Tieftemperaturelektronik
(Vgl. Tabelle 6.1.0 zu optischen und Tabelle 6.3.1 zu mechanischen Sensoren)

ca. 1800	OHM entdeckt, daß Bismut seinen Widerstand in Abhängigkeit vom Magnetfeld ändert
1821	SEEBECK beobachtet Effekt für Thermoelemente
1856	LORD KELVIN führt Prinzip des Dehnungsmeßstreifens ein
1879	HALL entdeckt die Querspannung stromdurchflossener Leiter
1908	Verflüssigung von Helium
1911	KAMERLINGH — ONNES entdecken Supraleitung
1933	MEISSNER-Effekt; kritisches Magnetfeld bei Supraleitern
1941	Oberflächenwellensonde von FÖRSTER und STRAMBKE entwickelt
1952	WELKER entdeckt magnetfeldabhängige Halbleiter
1957	BSC-Theorie von BARDEEN, COOPER und SCHRIEFFER; COOPER-Elektronen
1961	Nachweis der Flußquanten durch DOLL und NÄBAUER
1962	Magnetdiode (madistor)
1962	Tunneleffekte von COOPER-Paaren durch JOSEPHSON entdeckt
1965	Feldplatte wird produziert
ca. 1970	Halbleiter-Dehnungsmeßstreifen
1972	WIEGAND-Sensor
1975	Anisotroper magnetoresistiver Effekt
1975	Piezotasten
ca. 1980	Folientasten

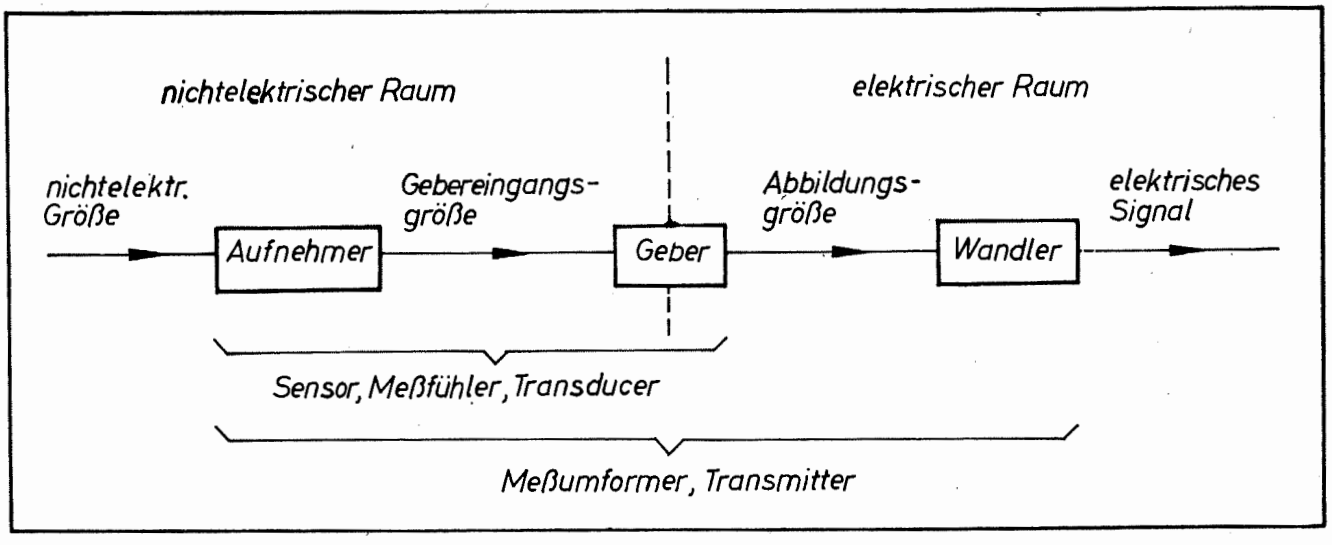


Bild 6.5.1. Begriffe und Signalflüsse in der Umgebung eines Sensors

elektrischen Raum existiert jene Größe, die kontrolliert, geprüft oder gemessen werden soll. Sie wird über einen Aufnehmer so verändert, daß sie dem Geber zugeführt werden kann. Der Aufnehmer ist bei einem Mikrofon z.B. die Membran. Dabei wird Schalldruck aufgenommen und als mechanische Bewegung dem Geber, z.B. einem Piezokristall zugeführt. Der Geber transformiert die nichtelektrische in einer elektrische Größe. Im gewissen Sinne stellt er also den Kern der Sensoren dar, während der Aufnehmer eine spezifische Zusatzeinrichtung ist.

Im elektrischen Raum wird in vielen Fällen das primäre Signal des Gebers, die Abbildungsgröße, noch einmal elektrisch gewandelt. Dies kann vielerlei Gründe haben, z.B. beim Mikrofon die Linearisierung des Frequenzganges. Der Geber kann aber auch nur Widerstandsänderung erfahren, z.B. bei einem Dehnungsmeßstreifen. Diese müssen dann in einem Wandler, z.B. einer Meßbrücke, in Spannungsänderungen umgewandelt werden.

Bezüglich der *Geber* lassen sich *zwei Gruppen,*

aktive und passive Geber,

unterscheiden. Die aktiven Geber benutzen die Energie der Eingangsgröße und wandeln sie in eine Spannung bzw. einen Strom um, der im Prinzip unmittelbar weiter gewandelt werden kann. Die passiven Geber benötigen dagegen eine elektrische (pneumatische, fluidische oder sonstige) Hilfsenergie, um die elektrische Abbildgröße entstehen zu lassen. Die Zusammenhänge zeigt, mit Beispielen belegt, Tab. 6.5.1.

Tabelle 6.5.1. Systematik der Gebereffekte

	Physikalischer Effekt	Anwendungsbeispiel
durch unmittelbare Energieumwandlung aus der Meßgröße (aktiv) → Erzeugung eines elektrischen Signals	thermoelektrische Effekte	Thermoelement
	piezoelektrische Effekte	Quarz-Druckaufnehmer
	fotoelektrische Effekte	Fotoelement
	elektromagnetische Induktion	Tachogenerator, Mikrofon
	chemoelektrische Effekte	Sauerstoffmeßzelle
mittels Parameter-Steuerung durch die Meßgröße und Auswertung über elektrische Hilfsenergie (passiv) → Erzeugung eines elektrischen Signals	Kapazität C, $C = \varepsilon \cdot f$ (Fläche, Abstand)	Feuchtefühler, Füllstandsensor
	Induktivität L, $L = \mu \cdot f$ (Windung, Geometrie, Kopplung)	Wegaufnehmer
	Widerstand R, $R = \varrho \cdot f$ (Länge, Querschnitt)	Platintemperaturfühler Pt 100, Dehnungsmeßstreifen
	Frequenz (Temperatur, Schichtdicke)	Quarz-TK, Schichtdicke

Im folgenden wird eine sehr kleine Auswahl von Sensoren behandelt: Etwas ausführlicher die magnetischen, da sie in engerem Zusammenhang mit der Elektronik stehen, dann die Dehnungsmeßstreifen, weil hier der Halbleitertyp große Bedeutung erlangt hat, und schließlich die Tasten, da sie durch die Mikroelektronik große Bedeutung erlangt haben.

6.5.2. Magnetische Sensoren

Messungen magnetischer Größen haben in der Elektronik eine besonders große Bedeutung. Es können folgende Prinzipien bzw. Bauelemente unterschieden werden:

HALL-Effekt
Feldplatte (Widerstandsänderungen)
Magnetdioden
WIEGAND-Effekt
Oberwellensonden.

Letztere nutzen die Nichtlinearität der Hysteresekurve aus und sind bei hohem elektronischen Aufwand besonders zur Messung extrem kleiner Felder geeignet [V 10, Bd. I].

6.5.2.1. HALL-Effekt

Frei bewegliche Elektronen und ebenso Löcher werden durch ein Magnetfeld abgelenkt. Ihre ursprüngliche Bewegung, die Richtung des Magnetfeldes und die Ablenkung stehen jeweils senkrecht aufeinander. Sie spannen also ein Dreibein auf (Bild 6.5.2a)). Im Festkörper wird die Ablenkung durch die ständigen Stöße mit

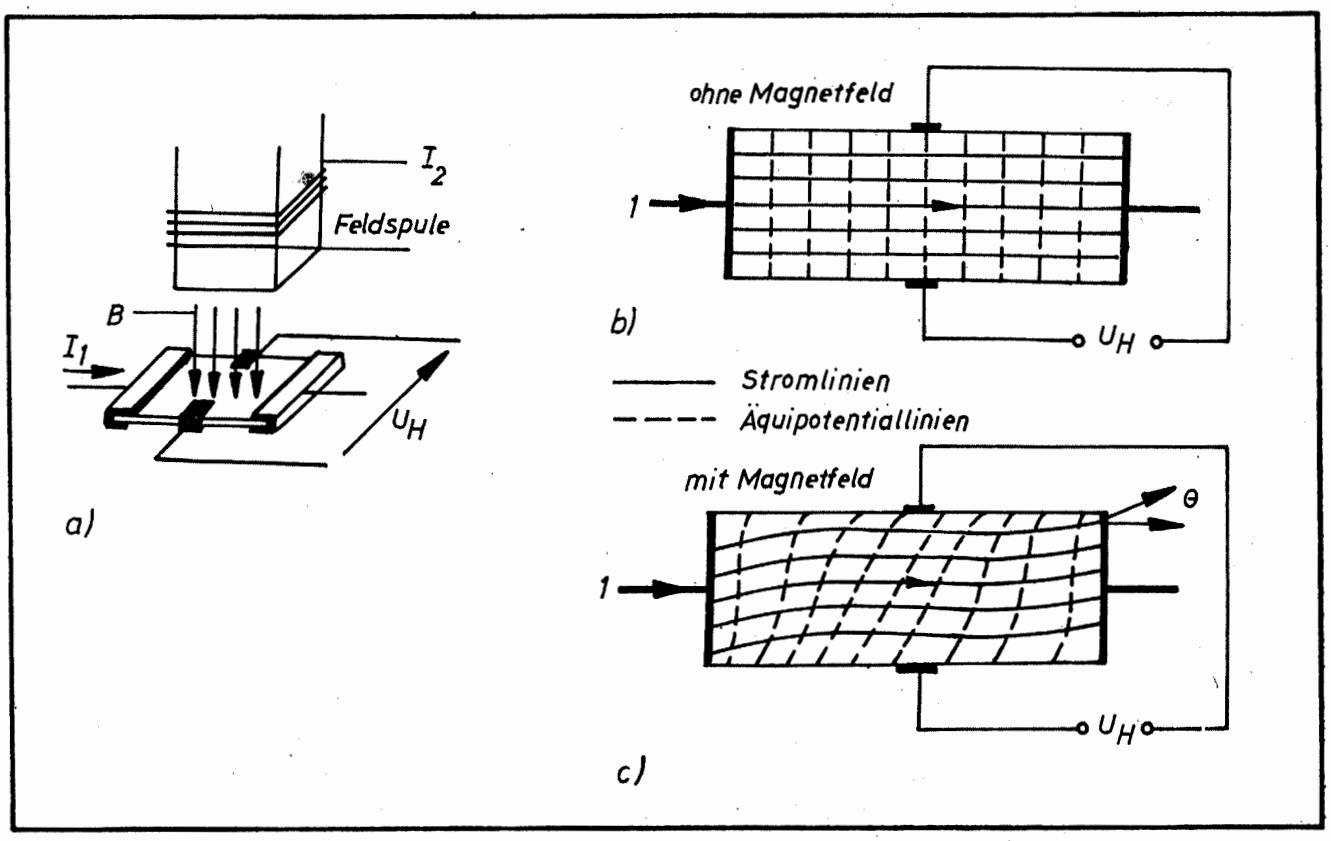


Bild 6.5.2. HALL-Sensoren

a) Es stehen jeweils senkrecht aufeinander: Magnetfeld, HALL-Steuerstrom und HALL-Meßspannung — Strom- und Äquipotentiallinien
b) ohne, c) mit Magnetfeld

anderen Ladungsträgern und Atomkernen vermindert. Es bildet sich deshalb ein sogenannter HALL-*Winkel* Θ aus. Um diesen Winkel weichen die Elektronenbahnen infolge des Magnetfeldes von den elektrischen Feldlinien ab. Diese Abweichung bewirkt den HALL-Effekt. Der Winkel Θ hängt unter anderem von der Zeit zwischen zwei Stößen und dem Weg, den ein Elektron zwischen diesen Stößen zurücklegt, ab. Hierfür ist hauptsächlich die in Abschnitt 3.1.6. eingeführte *Beweglichkeit* entscheidend.

Den Verlauf der Stromlinien mit und ohne Magnetfeld zeigt prinzipiell Bild 6.5.2b) und c). Hier sind ebenfalls die Äquipotentiallinien eingetragen. Unter Magnetfeldeinfluß verlaufen sie in der Mitte des Materials schräg. Folglich kann an zwei gegenüberstehenden Punkten eine Spannung, die sogenannte HALL-*Spannung* abgenommen

werden. Ihre Größe ist ersichtlich proportional dem Strom und dem HALL-Winkel. Der HALL-Winkel seinerseits ist proportional der magnetischen Induktion B. Deshalb gilt

$$U_H = R_H I_1 B . \tag{1}$$

Darin bedeutet R_H die HALL-Konstante. Mit einem Korrekturfaktor $1 \leqq r \leqq 1{,}93$ gilt für einen eigenleitenden Halbleiter

$$R_H = r \frac{1}{e_0 n_i} \frac{\mu^+ - \mu^-}{\mu^+ + \mu^-} . \tag{2}$$

Darin ist e_0 die Elektronenladung und n_i die Störstellenkonzentration. Typische Daten für einige Materialien enthält Tab. 6.5.2. Unter üblichen Verhältnissen liegen die HALL-Winkel bei Metallen unter einem Grad, während für InAs und InSb Werte um 80 Grad erreicht werden. Wegen seiner guten Integrierbarkeit hat anderseits vor allem Si als HALL-Material in integrierten Sensoren Bedeutung erlangt.

Tabelle 6.5.2. Materialwerte für HALL-Sensoren

Werkstoff	μ^- cm²/Vs	μ^+ cm²/Vs	$\Delta\varrho/\varrho$ bei 1 T	R_H cm³/As	ϱ Ωcm	TK von R_H %/K
Cu	40	—	—	$-4 \cdot 10^{-5}$	$1{,}6 \cdot 10^{-6}$	
Zn	—	7	—	$5 \cdot 10^{-5}$	$4{,}8 \cdot 10^{-6}$	
Bi	5000	—	0,45	10	0,002	0,4
Se	—	1	—	—	10^6	
Si	1900	350	—	10^8	63600	$<0{,}01$
Ge	3900	1800	0,3	10^5	46	0,3
InAs	27000	200	4,5	100	0,025	0,07
InSb	77000	700	23,6	470	0,007	1,5
$InAs_{0,8}P_{0,2}$	10000			1000	0,02	0,04

Neben der Magnetfeldmessung kann die HALL-*Sonde* auch für die *Multiplikation* günstig eingesetzt werden. Wenn wie in Bild 6.5.2a) das Magnetfeld durch einen Strom I_2 erzeugt wird und dem proportional ist, gilt nämlich

$$U_H \sim I_1 I_2 . \tag{3}$$

6.5.2.2. Feldplatte

Gemäß Bild 6.5.2c) erreicht die HALL-Spannung einen Grenzwert, wenn die Länge des HALL-Elementes groß im Vergleich zu seiner Breite ist. Dann verlaufen die Stromlinien in der Mitte der Platte parallel zu ihren Kanten, und die Äquipotentiallinien nehmen eine Neigung an, die dem HALL-Winkel entspricht. Wird die Platte dagegen kürzer gewählt, so können sich nur noch die geneigten Stromlinien ausbilden. Der Stromweg wird entsprechend dem HALL-Winkel verlängert. Die Verlängerung des Stromweges und die Größe der HALL-Spannung sind also offensichtlich zueinander reziprok.

Die Verlängerung des Stromweges in einem Material bedeutet eine Widerstandsänderung. Die Werte der *Widerstandsänderung* in Abhängigkeit vom Magnetfeld und von der Geometrie des Materials zeigt Bild 6.5.3a). Besonders große Widerstandsänderungen werden für kurze, aber breite Proben erreicht. Ein langes schmales Element zeigt dagegen praktisch keine Widerstandsänderung. Eine besonders große Widerstandsänderung ohne jede HALL-Spannung zeigt die runde Scheibe (CORBINO-

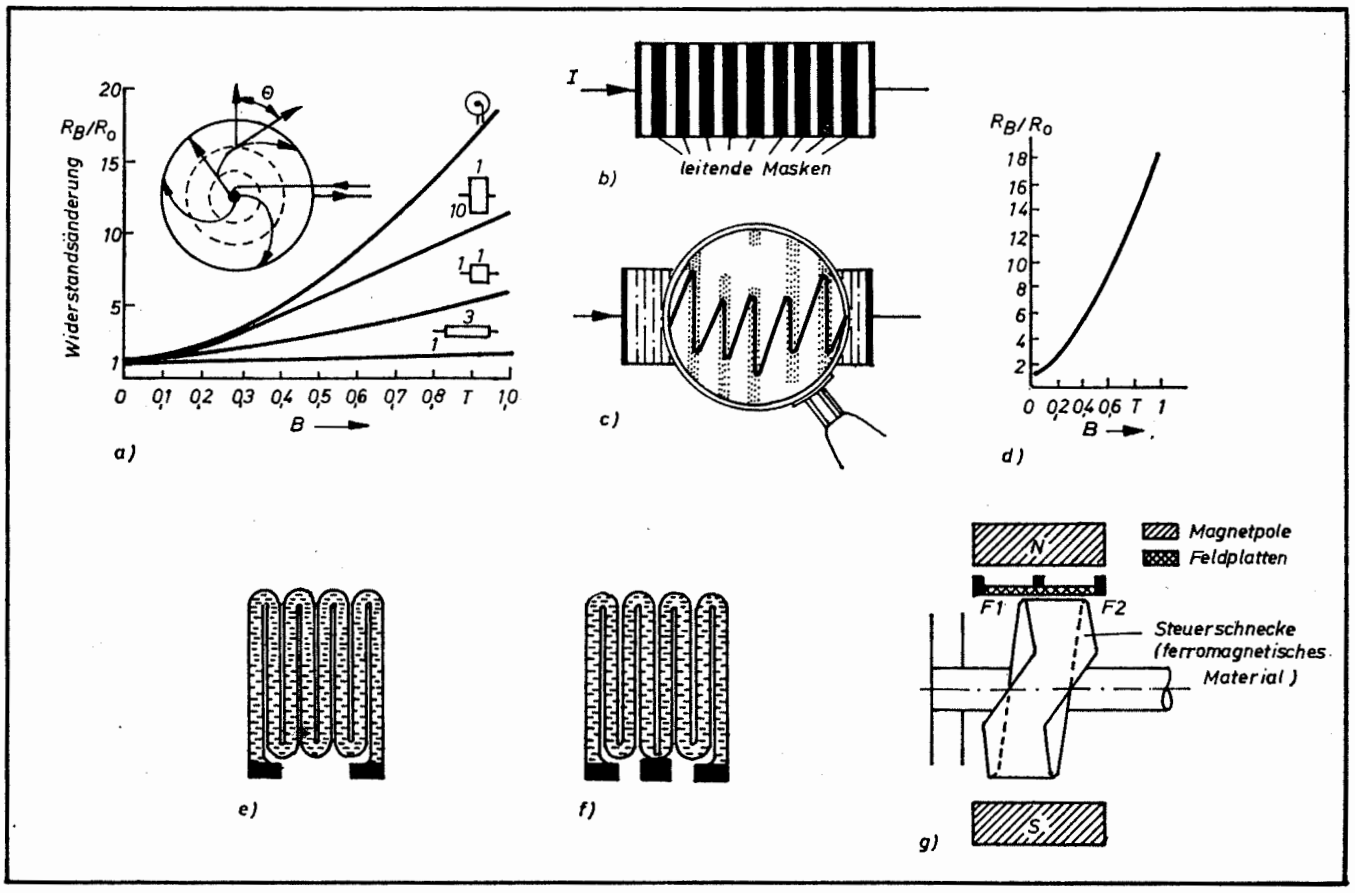


Bild 6.5.3. Zur Feldplatte

a) Widerstandsänderung bei HALL-Materialien verschiedener Geometrie. Die Zahlen an den Rechtecken geben das Längen-Breiten-Verhältnis der Proben an. Für die CORBINO-Scheibe sind zusätzlich vergrößert die Stromflußlinien mit den HALL-Winkel θ angegeben
b) Schema einer alten, bedampften Feldplatte
c) Stromlinien in der Feldplatte auf Basis von InSb-NiSb-Kristalliten
d) Kennlinie einer Feldplatte
e) Feldplatte mit möglichst langem Weg
f) Feldplatte für Differenzmessungen
g) Konstruktiver Aufbau eines Potentiometers mit einer Feldplatte nach f)

Scheibe). Sie ist jedoch wegen der hohen Stromdichte beim Mittelkontakt praktisch nicht anwendbar. *Optimal magnetfeldabhängige Widerstände* entstehen also durch eine *Reihenschaltung vieler kurzer breiter Streifen.* Zuerst wurde ihre Herstellung durch Bedampfung mit gut leitenden Masken entsprechend Bild 6.5.3b) (Rasterplatte) versucht. Seit einiger Zeit ist diese Aufgabe jedoch technologisch einfacher gelöst. In einem InSb-NiSb-Eutektikum lassen sich kleine, *gut leitende NiSb-Nadeln* erzeugen. Ihre Abmessungen liegen bei 50 μm Länge und 1 μm Durchmesser. Sie lassen sich ferner definiert ausrichten. Die Leitfähigkeit beträgt etwa $7 \cdot 10^4\ (\Omega m)^{-1}$. Das Grundmaterial InSb hat dagegen eine Leitfähigkeit von $200\ (\Omega m)^{-1}$. Den Stromfluß durch derartige Platten zeigt schematisch Bild 6.4.3c) mit Magnetfeld. Der Stromweg durch das relativ schlechtleitende Material wird also durch die Wirkung des Magnetfeldes erheblich verlängert. Insgesamt ergibt sich so eine Kennlinie wie in Bild 6.5.3d). Bild 6.5.3e) zeigt den Aufbau einer Feldplatte mit langem Weg für die Elektronen zur Erreichung einer hohen Empfindlichkeit. Teilbild f) zeigt eine doppelt ausgeführte Feldplatte. Sie kann in einem *kontaktfreien Potentiometer* entsprechend der Konstruktion von Teilbild g) verwendet werden.

In den letzten Jahren haben dünne Schichten aus magnetischen Materialien, wie Ni-, Co- und Fe-Legierungen, im Sinne der Feldplatte Bedeutung erlangt. Für sie hat sich jedoch der Begriff *magnetoresistiver Effekt* durchgesetzt. Die Material-

konstante $\Delta\varrho/\varrho$ liegt hier nur bei 1 bis 5%, kann aber relativ gut in Brückenschaltung genutzt werden. Vor vielen Jahren war der gleiche Effekt in Bi die meistbenutzte Methode zur Magnetfeldmessung.

6.5.2.3. Magnetdioden

Den prinzipiellen Aufbau einer Magnetdiode zeigt Bild 6.5.4a). Sie besteht aus einem Ge-Kristall, und zwischen dem p- und n-Gebiet befindet sich ein langes i-Gebiet, welches einseitig ein Rekombinationsgebiet RZ enthält. Es entsteht durch Einbau von Fremdatomen, z.B. Gold. Die Diode wird in Durchlaßrichtung betrieben. Infolge des Magnetfeldes werden sowohl die Löcher als auch die Elektronen nach einer Seite also z.B. zur Rekombinationsschicht, abgelenkt. Dadurch rekombinieren sie, und der Diodenwiderstand nimmt zu. Bei Umpolung des Feldes ist dies im Gegensatz zur Feldplatte, wie Teilbild b) zeigt, bei der Diode nicht der Fall. Um einen symmetrischen Verlauf zu bekommen, können nun zwei Dioden in Reihe geschaltet werden. Sie müssen aber an den entgegengesetzten Seiten das Rekombinationszentrum besitzen. So entsteht die Kennlinie von Bild 6.5.4d). Sie zeigt auch zugleich den großen Vorteil der Magnetdioden, nämlich ihre hohe Ausgangsspannung.

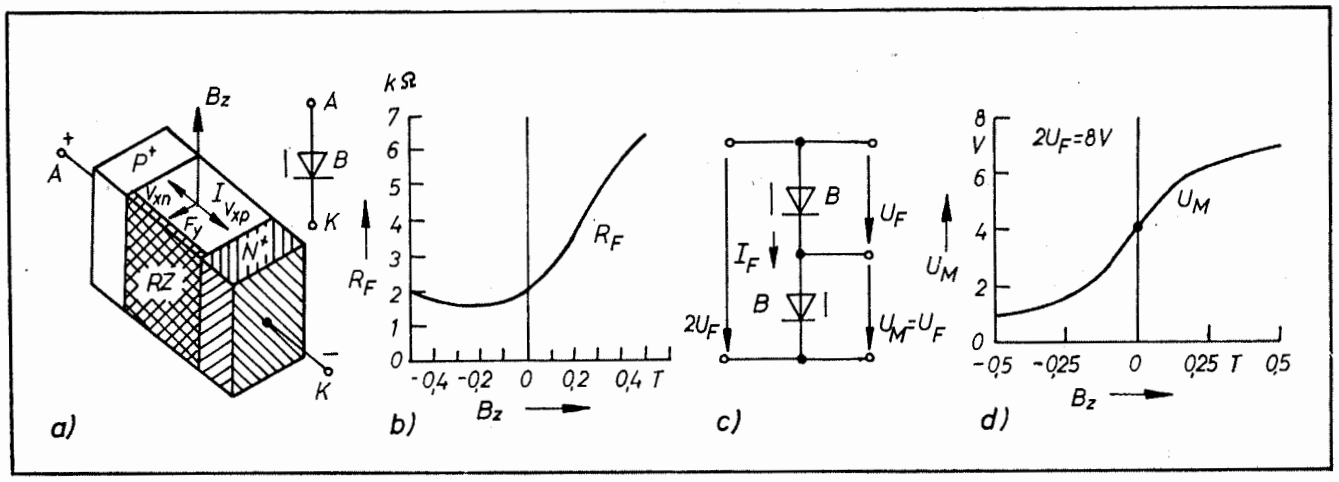


Bild 6.5.4. Magnetdiode

a) Schematischer Aufbau und Schaltbild
b) Kennlinie
c) Reihenschaltung von zwei Magnetdioden zu einem Bauelement
d) Kennlinie einer kommerziellen Magnetdoppeldiode AHY10 bei 10 V und 10 mA Betriebswerten

6.5.2.4. WIEGAND-Sensor

Zur Herstellung eines WIEGAND-Sensors benutzt man einen ferromagnetischen Draht von etwa 0,3 mm Durchmesser, der durch abwechselnd gegenläufiges Verdrillen verformt und anschließend etwas gedehnt wird. Danach folgt eine Wärmebehandlung zur Stabilisierung der Eigenschaften. Durch das Verdrillen wird der äußere Drahtbereich stärker verformt als der Kern. Man erhält eine mechanisch und magnetisch härtere Schale bei einem weichen Kern. Im Draht entsteht eine ausgeprägte Verformungstextur in axialer Richtung, wobei sich die Magnetisierung der WEISSschen Bezirke vorzugsweise parallel zur Drahtachse ausrichtet. Der Draht wird mit einer Spule von ca. 1000 Windungen umgeben. Die äußeren Schichten schalten bei etwa 40 A/cm und die inneren bei ca. 10 A/cm (Bild 6.5.5a)). Für die Anwendung werden die äußeren Schichten in einer Richtung festgehalten, so daß nur die inneren unter Einfluß eines entsprechenden Feldes (< 40 A/cm) schlagartig umschalten und in der

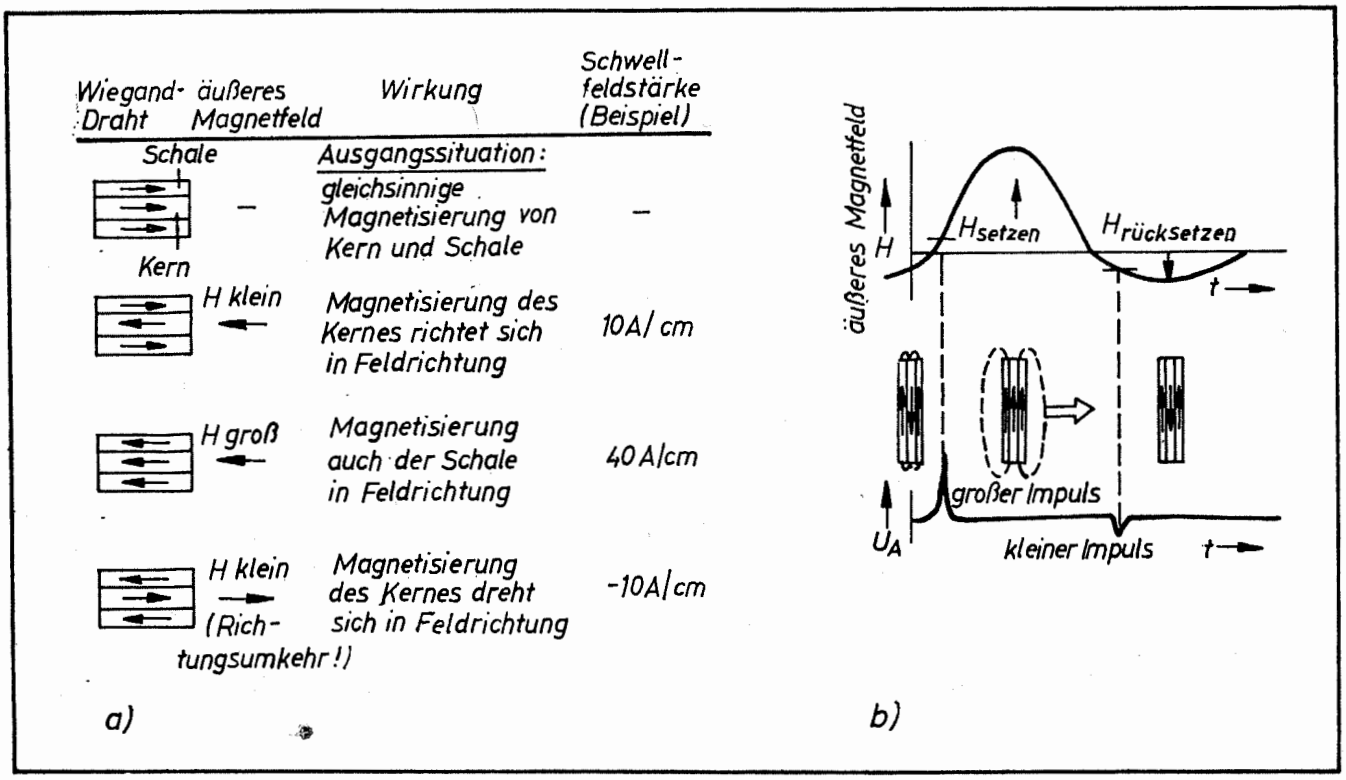


Bild 6.5.5. WIEGAND-Sensor

a) Seine verschiedenen magnetischen Zustände
b) Verlauf von magnetischer Feldstärke (oben), Zustände des Sensors (Mitte) und Impulse an den Spulen (unten)

Spule einen großen Spannungsimpuls induzieren (Bild 6.5.5b)). Die Impulse besitzen eine Dauer von ca. 20 µs und eine Amplitude von ca. 5 V. Der Sensor arbeitet von ca. —150 bis 200 °C und bis zu 10 kHz.

6.5.3. Dehnungsmeßstreifen (DMS)

Wird ein Draht in seiner Längsrichtung belastet, so wird er länger und dünner. Dies hat schon rein mechanisch zur Folge, daß sein Widerstand zunimmt. Zusätzliche Effekte können noch durch die Verlagerung der Kristallbereiche auftreten. Dieser zweite Effekt, der piezoresistive, ist besonders stark bei Silicium ausgeprägt und beruht darauf, daß sich die Bandstruktur ändert. Insgesamt läßt sich für die Widerstandsänderung also schreiben

$$\mathrm{d}R = \frac{\varrho}{A}\,\mathrm{d}l - \frac{\varrho l}{A^2}\,\mathrm{d}A + \frac{l}{A}\,\mathrm{d}\varrho\,, \tag{4}$$

wobei A die Fläche, l die Länge und ϱ den spezifischen Widerstand bedeuten. Für die *relative Änderung* gilt

$$\frac{\mathrm{d}R}{R} = \frac{\mathrm{d}l}{l}\left(1 - \frac{l}{A}\,\frac{\mathrm{d}A}{\mathrm{d}l} + \frac{l}{\varrho}\,\frac{\mathrm{d}\varrho}{\mathrm{d}l}\right). \tag{5}$$

Der Klammerausdruck wird als k-Faktor bezeichnet. Er gibt an, wie stark eine relative Längenänderung in eine relative Widerstandsänderung übersetzt wird. Beispielwerte sind:

Material	k-Faktor
Platin	+6,10
Chrom-Nickel	+2,63
Konstantan	+2,15
Manganin	+0,50
Nickel	−12,10
Silicium	±100 ... 210

Negative Werte besitzen nur Nickel, Bismut und n-leitendes Silicium. Mit einer Kombination von n-Si und p-Si können also k-Faktoren von 200 bis 240 erreicht werden. Die Konstanz des k-Faktors ist am besten bei Konstantan, sowohl bezüglich der Temperatur als auch der Zeit und Belastungen.

Den prinzipiellen Aufbau von Dehnungsmeßstreifen zeigt Bild 6.5.6, Teilbilder a) und b) für Drähte und c) für Halbleiter. Dehnungsmeßstreifen werden u. a. für

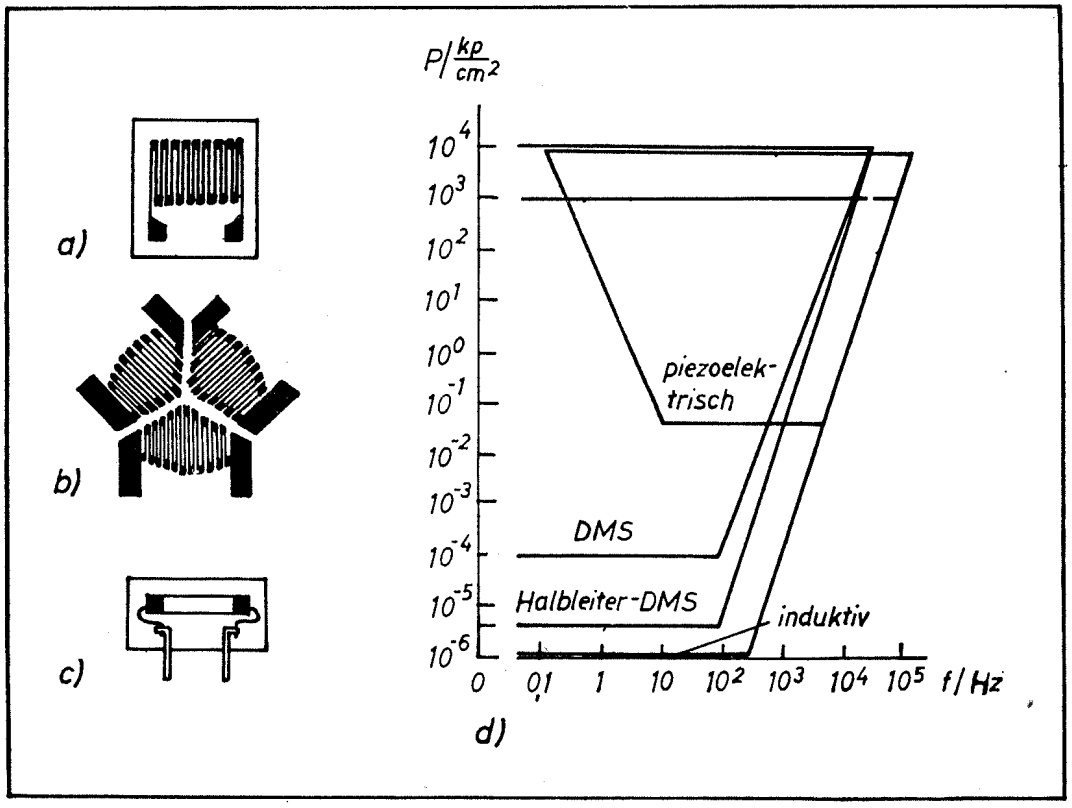


Bild 6.5.6. Dehnungsmeßstreifen (DMS)

a) Typischer Aufbau eines einfachen Gebers. Auf einer dünnen, elastischen Folie sind dünne Drähte, die an den Enden verstärkt verbunden sind

b) Ein DMS, der Messungen in allen Richtungen zuläßt

c) Halbleiter-DMS

d) Anwendungsbereiche für Druckmessungen verschiedener Geber

Messungen der Kraft, des Druckes und des Weges eingesetzt. Den Vergleich mit piezoelektrischen und induktiven Gebern zeigt Teilbild 6.5.6d). Als Meßschaltung für die DMS wird eine WHEATSTONE-Brücke eingesetzt. Zur Temperaturkompensation wird oft ein gleichartiger, aber mechanisch nicht belasteter DMS verwendet.

6.5.4. Tasten

Da die Eingabe per Sprache in elektronische Geräte auf längere Sicht nicht befriedigend und hinreichend einfach zu lösen ist, bleiben die Tasten die *wichtigsten*, fast

einzigen *Eingabemöglichkeiten* für den Menschen. So hat sich mit der Entwicklung der Mikroelektronik eine große und vielfältige Entwicklung ergeben. Ursprünglich wurde von Schaltern ausgegangen, die über einen Mechanismus betätigt wurden. Hierbei treten, wie bei Relais, *Kontaktprellungen* im Bereich von mehreren μs auf. Dies wird heute durch spezielle Schaltungen bzw. über Programme abgefangen. Weiter weisen Kontakte eine begrenzte Anzahl von Schaltspielen auf. So wurden *andere Mechanismen*, z.B. auf kapazitiver Basis mittels des HALL-*Effektes* usw., untersucht. Eine gewisse Bedeutung konnte der *piezoelektrische Quereffekt* (vgl. Bild 6.3.7, ganz rechts) erlangen. Der Sensor hat die Abmessungen von ca. 8 mm Breite, 0,2 mm Dicke und 20 mm Höhe. Betätigt wird er mit Druck in Richtung der Höhe, während die Kontakte an den 20 × 8 mm² großen Breitseiten liegen. Nach den letzten Entwicklungen dürfte aber für die meisten Tastaturen die *Folientastatur* (Bild 6.5.7) die größte Anwendungsbreite (ca. 80% wird geschätzt) gewinnen.

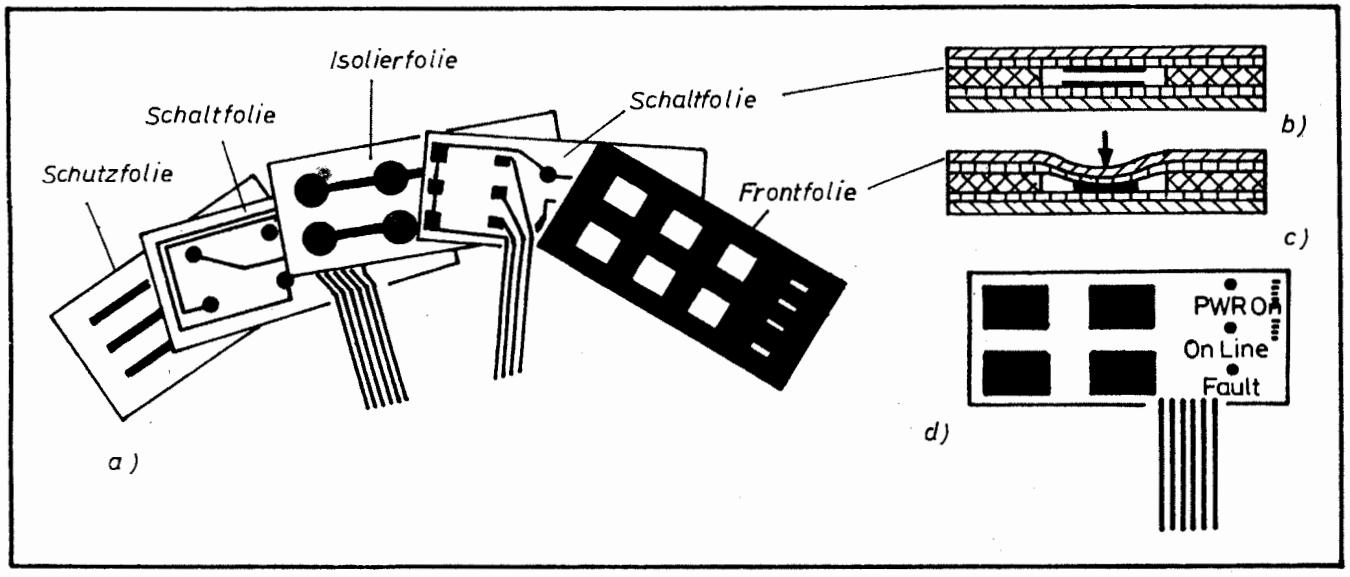


Bild 6.5.7. Aufbau einer Folientastatur

a) Fünf Folienelemente
b) Zusammenbau der Folien, für einen Kontakt gezeichnet
c) Kontaktgabe bei Betätigungsdruck
d) Äußeres Aussehen der Folientastatur mit einem Anzeigefeld (rechts)

Auf zwei dünne Kunststoffolien aus Polyester oder Polykarbonat werden Leiterbahnen und Kontaktflächen mittels eines leitenden Mediums aufgebracht. Für die Leiterbahnen und Kontakte kommt die Dickschichttechnik (9.4.4.) zum Einsatz. Diese beiden Folien werden so zusammenmontiert, daß sie durch eine Zwischenmaske getrennt sind und nach oben und unten dicht durch je eine Deckfolie abgeschlossen werden. Mittels Druck wird die obere Folie verformt und der Kontakt geschlossen. Der auf diese Art zusammenmontierte Membrantaster ist nur zum Schalten von Steuerströmen im mA-Bereich geeignet. Diese Impulse müssen durch eine nachgeschaltete Steuerelektronik an die gewünschte Schaltanwendung adaptiert werden. Im Vergleich zu traditionellen Schaltungen legt die Membranschaltung eine Strecke von 0,1 bis 0,15 mm zurück. Die dünnen Materialien, die für den Aufbau verwendet werden, ermöglichen die Aktivierung der Schaltung durch minimalen Fingerdruck, wobei der Druck durch geeignete Wahl der Frontfolie und Zwischenmaske variiert werden kann. Es hat den Anschein, als ob dieses Prinzip für hochwertige Tastaturen kaum Anwendung findet. Hierfür dürften andere Prinzipien noch weiterentwickelt werden.

6.6. Mikrowellenbauelemente

Die Mikrowellenbauelemente sind nicht Inhalt dieses Buches. Sie haben einen hohen Stand erreicht. Dies wurde bezüglich des Materials GaAs und bezüglich der FET kurz erwähnt. Darüber hinaus sind eine Vielzahl von aktiven Mikrowellendioden entstanden. Es seien hier nur das GUNN-Element, die Lawinenlaufzeitdioden (z.B. READ-Diode, Trapattbetrieb), die Barittdioden und die Tunneldioden erwähnt.

Einen Überblick hierzu geben u. a. [H 13, H 17, P 11]. Hier sollen nur zwei Bauelemente ausgewählt werden: die Tunnel- und Backwarddioden.

6.6.1. Tunneldioden

Eine Tunneldiode, auch ESAKI-Diode genannt, besitzt bis auf zwei spezielle technische Größen einen normalen *pn*-Übergang. Einmal ist der *pn*-Übergang extrem abrupt und erfolgt innerhalb von weniger als 0,01 μm. Zum andern ist die Dotierung der Materialien größer als $2{,}5 \cdot 10^{19}$ Störstellen/cm³. Bei normalen Dioden enthält das *p*-Material etwa 10^{17} und das *n*-Material ungefähr 10^{15} Störstellen/cm³. Infolge der hohen Dotierung wandert das FERMI-Niveau gemäß Bild 3.1.8c) ins Leitungs- bzw. Valenzband. Zusätzlich sind durch die hohe Dichte die Akzeptoren- und Donatorenniveaus zu Bändern verbreitert. So entsteht das Bändermodell der Tunneldiode, wie es Bild 6.6.1a) zeigt (vgl. mit einer normalen Diode, Bild 3.2.1h). Das Leitungsband

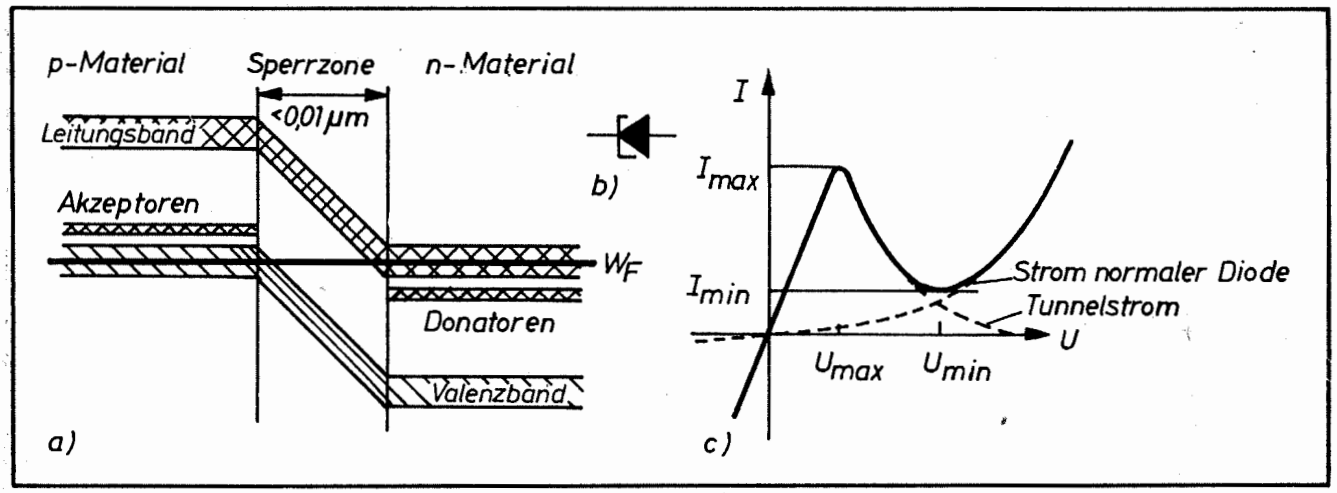


Bild 6.6.1. a) Bändermodell für eine Tunneldiode, b) ihr Schaltungssymbol und c) ihr Strom-Spannungsverlauf

des *n*-Materials liegt also auf demselben Energieniveau wie das Valenzband des *p*-Materials. Beide Bänder sind nur durch die sehr schmale Sperrzone getrennt. Deshalb können Ladungsträger den Potentialwall der Sperrzone quantenmechanisch ohne Energieverlust durchdringen. Dieser Vorgang heißt *Tunneleffekt*. Dadurch können auch bei sehr kleinen äußeren Spannungen Ladungsträger die Sperrzone durchtunneln und somit relativ große Ströme hervorrufen. Dies führt zum steilen linearen Anstieg der Strom-Spannungs-Kennlinie aus dem Ursprung gemäß Bild 6.6.1c). Der Tunneleffekt wird geringer, wenn eine äußere Spannung die Bänder gegeneinander verschiebt. Die abnehmende Überlappung der beiden Bänder führt dazu, daß der Tunnelstrom nach seinem Maximum auf Null absinkt. Dies zeigt die gestrichelte Linie in Bild 6.6.1c). Wird die Tunneldiode immer weiter in Durchlaßrichtung vorgespannt, so liegen die Bänder ähnlich wie in Bild 3.2.1h). Es erscheint also der normale Durchlaßstrom einer Diode (linke gestrichelte Linie in Bild 6.6.1c)).

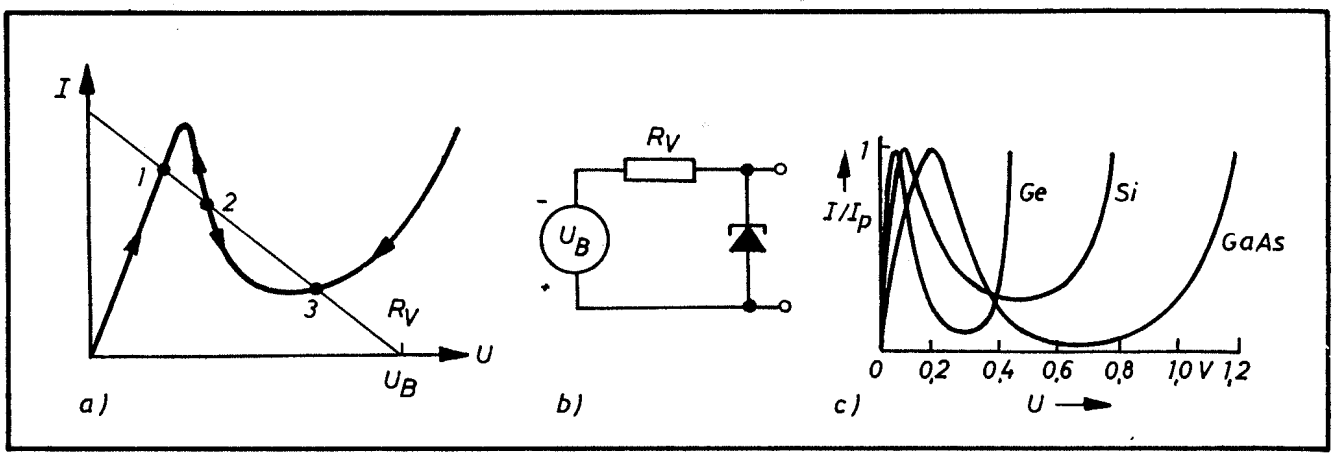


Bild 6.6.2. Zum bistabilen Verhalten einer Tunneldiode (a), b) zugehörige Schaltung und c) Vergleich der Kennlinien für die drei Typen von Tunneldioden

Insgesamt besitzt also die Tunneldiode einen Kennlinienteil, wo mit steigender Spannung der Strom abnimmt. Dies bedeutet einen differentiellen negativen Widerstand. Dieser Bereich existiert zwischen der Spannung U_{max} und U_{min}. Unabhängig von dem verwendeten Material liegt die Maximumspannung bei etwa 50 mV und die Minimumspannung zwischen 300 und 500 mV (vgl. Bild 6.6.2c)). Das Verhältnis von I_{max} zu I_{min} ist stark materialabhängig. Es beträgt für Si etwa 6, für Ge ungefähr 10 und für GaAs ungefähr 60. Der maximale Strom hängt stark vom Aufbau der Diode ab. Er kann weniger als ein mA betragen und bis zu einigen Ampere gewählt werden.

Die Anwendung der Tunneldiode erfolgt hauptsächlich infolge ihres differentiellen negativen Widerstandes. Er kann vor allem zur Entdämpfung und Schwingungserzeugung im Sinne von Abschnitt 7.3.7. herangezogen werden. Da der Tunneleffekt sich schneller als in Nanosekunden vollzieht, ist die statische Kennlinie bis in den GHz-Bereich hinein gültig. Infolge der geringen Dicke der Raumladungszone des *pn*-Überganges existieren allerdings hohe spezifische Kapazitäten in der Größenordnung von $\mu F/cm^2$. Die Fläche des *pn*-Überganges wird deshalb stets kleiner als 0,01 mm^2 gehalten.

Eine zweite wichtige Anwendung der Tunneldiode erfolgt in der Digitaltechnik. Sie wird dazu über einen Vorwiderstand R_V aus einer Gleichquelle U_B betrieben (Bild 6.6.2). Die Widerstandsgerade schneidet unter bestimmten Bedingungen dreimal die Kennlinie der Tunneldiode. Lediglich die Arbeitspunkte 1 und 3 sind dabei stabil. Kleine Änderungen des Widerstandes, der Kennlinie oder der Betriebsspannung führen immer wieder zu diesen Arbeitspunkten zurück. Beim Punkt 2 erfolgt dagegen sofort ein Übergang entweder zum Punkt 1 oder Punkt 3. In dieser Schaltung ist die Tunneldiode ein bistabiler Speicher oder Schalter mit extrem kurzer Umschaltzeit.

6.6.2. Backwarddioden

Durch eine spezielle Technologie (genaue Dotierung) ist es möglich, den Höcker der Tunneldiode zu unterdrücken. Dann entsteht die sogenannte Backward- (Rückwärts-) Diode. Ihre Kennlinie zeigt Bild 6.6.3a). Sie dient zur Gleichrichtung kleinster Spannungen. Ihre Kennlinie ist nämlich, wie Bild 6.6.3b) im Vergleich zu anderen Dioden zeigt, besonders stark gekrümmt. Im Vergleich zu einer normalen Diode ist sie dabei entgegengesetzt gepolt. Darauf geht ihr Name zurück. Auch bei ihr ist der Tunneleffekt vorhanden. Deshalb erfolgt ihr Einsatz meist bei höchsten Frequenzen bzw. bei kleinsten Spannungen. Infolge der hohen Dotierung sind bei der Tunnel- und Backwarddiode Minoritätsträger praktisch nicht vorhanden.

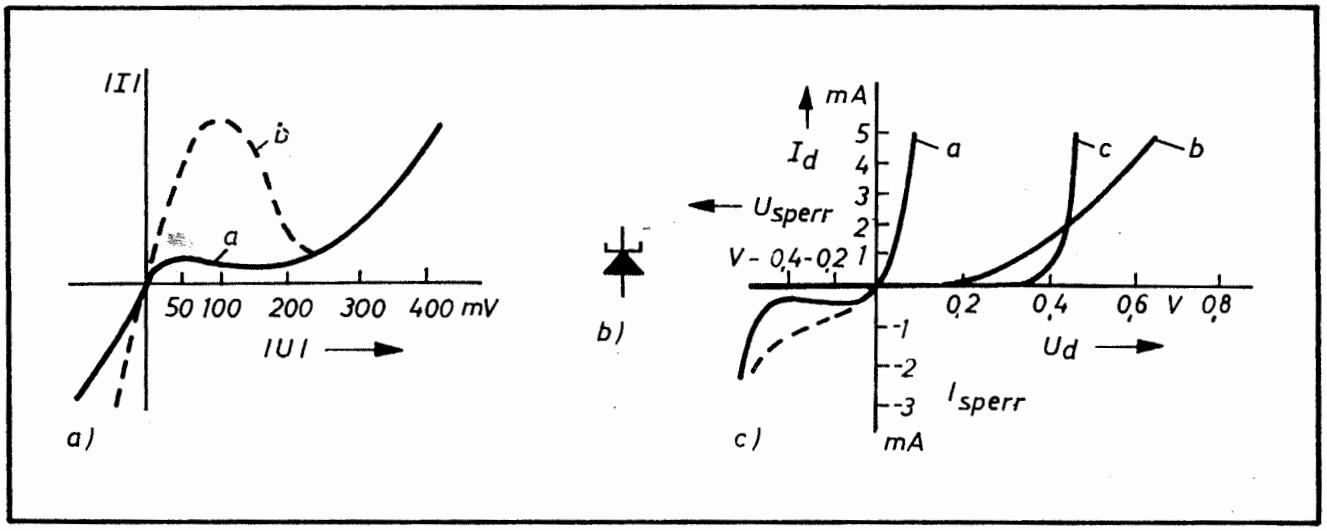


Bild 6.6.3. a) Kennlinie der Backwarddiode (———) und der Tunneldiode (— — —) zum Vergleich
b) Schaltsymbol
c) Vergleich der Kennlinien einer Rückwärtsdiode (*a*) mit und ohne Parallelwiderstand, einer Germaniumdiode (*b*) und einer Siliciumdiode (*c*)

6.7. Tieftemperaturelektronik

Die *Halbleiterelektronik strebt* in verschiedenen Daten *physikalischen Grenzen entgegen.* Dies betrifft u. a. minimale Störungen, Schaltgeschwindigkeiten und Leistungen je Bit. Hier bringt die Tieftemperaturelektronik bessere Parameter. Dies ist einer der Gründe, weshalb an ihrer Erforschung und künftigen Anwendung intensiv gearbeitet wird. Entscheidend für ihren Einsatz ist dabei die Möglichkeit, *tiefe Temperaturen* im Thermostaten einzustellen. Dies erfolgt mit *verflüssigten Gasen.* Die Regelung erfolgt dabei u. a. durch den jeweiligen Gasdruck im Bereich von ca. 7 bis 200 kPa (50 bis 1500 Torr) und führt zu den *folgenden Temperaturen*:

Kältemittel	Temperatur /K
Ethylen	134 ... 182
Methan	92 ... 120
Sauerstoff	70,5 ... 97
Stickstoff	63,5 ... 84
Neon	24,6 ... 30
Wasserstoff	14,1 ... 23
Helium	2,5 ... 5,0

Der hierfür notwendige *technische Aufwand* muß also durch Vorteile aufgewogen werden. Die wesentlichen Effekte bei den tiefen Temperaturen stehen im Zusammenhang mit der *Supraleitung*, und für sie sind die COOPER-*Paare* aus Elektronen entscheidend. Für sie gilt nicht die FERMI-Verteilung, wie sie im Abschnitt 3.1.4. eingeführt wurde. Anschaulich kann ein COOPER-Paar gemäß Bild 6.7.1 erklärt werden. Zwei Kugeln befinden sich auf einer gespannten Gummifolie und drücken sich dementsprechend etwas ein. Werden die Kugeln auf einen relativ kleinen Abstand genähert, so laufen sie automatisch zusammen und bilden mit dem Energiegewinn ΔW eine Einheit.

Der Durchmesser von COOPER-Paaren liegt bei ca. 10^{-7} ... 10^{-6} m, also dem 100- bis 1000fachen eines Atomdurchmessers. Das bedeutet, daß in der Regel *viele* COOPER-

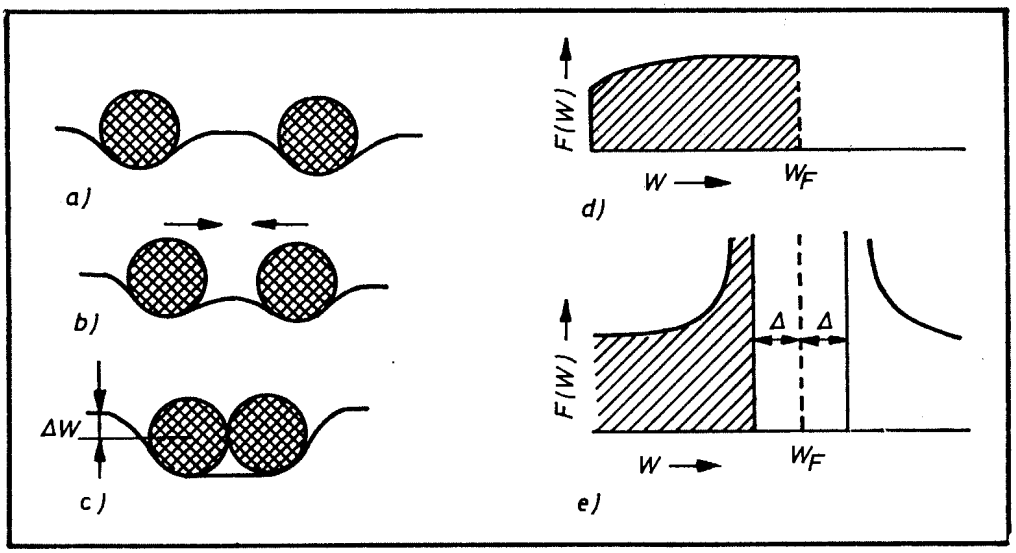


Bild 6.7.1. Zum COOPER-Paar: Modell mit zwei Kugeln auf gespannter Gummimembran

a) Mit großem Abstand als stabile Lage
b) Bei Annäherung wird die Lage instabil, um dann
c) ein zusammengefaßtes Paar zu bilden
d) FERMI-Verteilung von Fermionen (Elektronen) am absoluten Nullpunkt
e) Auftreten des Gap bei Supraleitung mit COOPER-Paaren am überhöhten linken Ast

Paare eine Einheit, d. h. ein einziges makroskopisches, quantenmechanisches System bilden. Dieses kann *nicht mehr mit dem Gitter in Wechselwirkung* treten. So können sich die COOPER-Paare ohne Widerstand bewegen. Erst wenn die COOPER-Paare eine Energie aufnehmen, die größer als das ΔW ist, werden sie aufgebrochen, und das Material wird wieder normalleitend.

Damit läßt sich in der Umgebung von $T = 0$ die *Elektronenverteilung* beim Leiter und Supraleiter gemäß Bild 6.7.1d) und e) beschreiben. Alle COOPER-Paare liegen $\Delta W/2$ vor der FERMI-Kante. Da für sie das PAULI-Prinzip nicht gilt, können viele gleichartige „Elektronen" existieren. Es entsteht der Gap (Spalt) um die FERMI-Kante. Bei höheren Temperaturen $T > 0$ können, solange noch COOPER-Paare existieren, Elektronen und COOPER-Paare nur Terme unterhalb des rechts vom FERMI-Niveau in Teilbild e) gezeichneten Verlaufs annehmen.

Energie an COOPER-*Paare* kann *durch ein äußeres Magnetfeld* bzw. durch *erhöhte Temperatur* zugeführt werden. So gibt es eine *kritische Temperatur* T_c, unterhalb

Tabelle 6.7.1. Kritische Temperaturen T_c (Temperaturen, bei denen die Stoffe vom normalleitenden in den supraleitenden Zustand übergehen)

Element		T_c/K	Element		T_c/K
Wolfram	W	0,01	Thorium	Th	1,37
Iridium	Ir	0,14	Palladium	Pa	1,40
Hafnium	Hf	0,16	Rhenium	Re	1,7
Titanium	Ti	0,39	Thallium	Tl	2,37
Cadmium	Cd	0,52	Indium	In	3,41
Zirconium	Zr	0,55	Zinn	Sn	3,72
Osmium	Os	0,66	Quecksilber	Hg	4,15
Zink	Zn	0,85	Tantal	Ta	4,48
Molybdän	Mo	0,92	Vanadium	V	5,03
Gallium	Ga	1,09	Blei	Pb	7,19
			Niobium	Nb	9,1
			Technetium	Tc	11,2

Verbindung	T_c/K
UCo	1,70
InSb	1,9
Nb_6Sn_5	2,07
Ti_2Co	3,44
La_3In	10,4
Nb_3An	11,5
MoN	12,0
NbN	16,0
V_3Ga	16,5
V_3Si	17,1
Nb_3Al	17,5
Nb_3Sn	18,05
Nb_2Sb	18,00
$Nb_3(Al_{0,8}\ Ge_{0,2})$	20,8

der verschiedene Materialien supraleitend werden. Werte sind in Tab. 6.7.1 zusammengestellt. Supraleitung wurde bisher nicht nachgewiesen bei Li, Na, K, Rb, Cs, Cu, Ag, Au, Pt sowie Ca, Ba, Fe, Co, Ni. Die Kenntnisse über Materialien mit unterschiedlichem T_c im Verlaufe der Zeit zeigt Bild 6.7.2a).

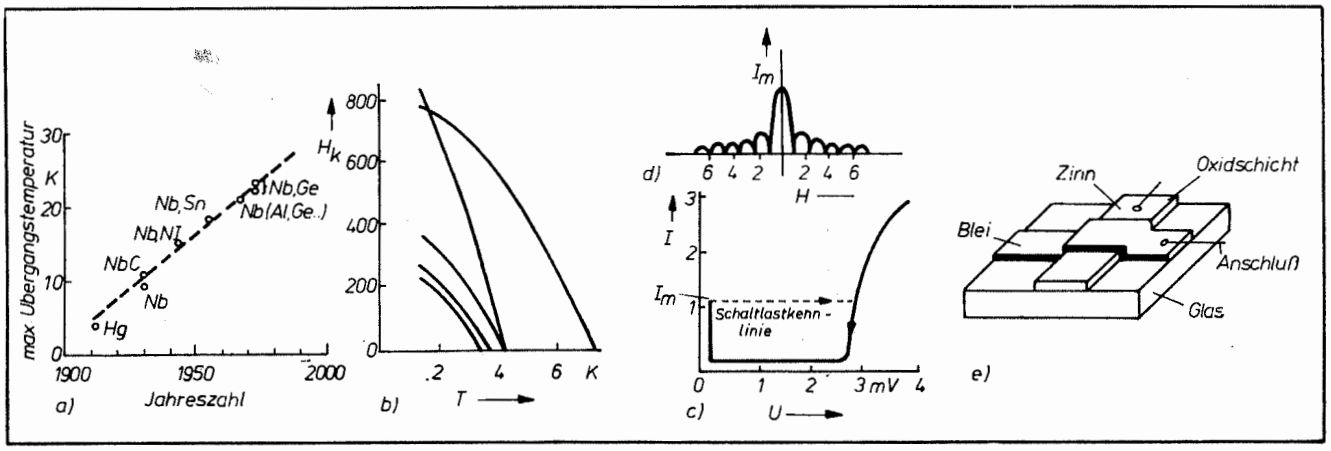


Bild 6.7.2. Eigenschaften von Supraleitern

a) Geschichtliche Entwicklung bez. der höchsten Temperatur, bei der Supraleitung auftritt
b) Grenzkennlinien bei Supraleitern bez. Magnetfeld und Temperatur (H_k in A/cm)
c) JOSEPHSON-Tunneleffekt, linker Ast Gleichstrom, rechter Ast Wechselstromkennlinie
d) Mögliche maximale Gleichströme bezüglich der Schaltkennlinie als Funktion des umschlossenen Flusses
e) Prinzipieller Aufbau eines JOSEPHSON-Tunnelelements

Zwischen der kritischen Temperatur und dem Gap besteht die Beziehung

$$\Delta W = 3{,}5kT_c\,. \tag{1}$$

Die Energie liegt also in der Größenordnung von 10^{-4} bis 10^{-3} eV.

Den Einfluß des *Magnetfeldes* zeigt Bild 6.7.2b). Für den hier nicht interessierenden Fall der Supraleiter 2. Art gibt es auch Materialien, die bei sehr hohen Feldern, z.B. über 10 T, noch supraleitend sind. Sie werden u. a. für starke Magnete verwendet.

Im Laufe der Zeit sind *viele supraleitende Bauelemente erprobt* worden. Sie haben aber kaum praktische Bedeutung erlangen können. Die *größte Aussicht* hat z. Z. der JOSEPHSON-*Tunneleffekt*. Hierbei werden zwei Supraleiter durch einen sehr dünnen (ca. 10^{-9} m) Isolator getrennt. JOSEPHSON wies 1962 theoretisch nach, daß über diesen Isolator COOPER-Paare tunneln können. Das bedeutet, daß der Isolator dann ebenfalls supraleitend erscheint, und ohne Spannungsabfall kann ein Gleichstrom zwischen den beiden Supraleitern fließen (Bild 6.7.2c)). Dies ist aber nur bis zu einer maximalen Größe I_m möglich. Wird der Strom weiter erhöht, so tritt ein Wechselstrom-Tunneleffekt mit der rechten Kennlinie auf. Der maximale Strom hängt gemäß Bild 6.7.2d) auch von dem einwirkenden Magnetfeld ab. Die Nullstellen des Stromes hängen mit den Flußquanten

$$\Phi = \frac{h}{2e} \approx 2{,}0678 \cdot 10^{-15}\ \mathrm{Vs} \tag{2}$$

zusammen. Es können auf der Verbindungsfläche zwischen den Leitern nur ganzzahlige Vielfache hiervon wirksam werden.

Ersichtlich kann also ein so einfaches JOSEPHSON-Tunnelelement die Zustände der beiden Kennlinien speichern (das ist 1 Bit). Diese Möglichkeit und Varianten davon können u. a. zu hochleistungsfähigen Speichern und Datenverarbeitungstechniken führen.

Den typischen Aufbau eines JOSEPHSON-Tunnelelementes zeigt Bild 6.7.2e).

7. Analoge Schaltungen

Die Definition von analog im Vergleich zu digital erfolgt zu Anfang von Kapitel 8. Hier wird die intuitive Interpretation im Sinne von kontinuierlich, stetig, nicht-digital vorausgesetzt, wobei auch die angrenzenden Gebiete mit dem Übergang zu impulsförmigen Signalen einbezogen werden. Gemeint sind die entsprechenden Modulationen und Schaltungen, also insbesondere z. B. der D-Verstärker und die Multivibratorfamilie. Zunächst werden dabei die allgemeinen Prinzipien von Rückkopplung und Modulation behandelt. Erst dann folgen die speziellen Schaltungen im Sinne der Verstärkung von Signalen, der Endstufen, wo die Leistung im Vordergrund steht, der Spannungs- und Stromstabilisation und weiterer Prinzipien. Den Abschluß bildet die Meßtechnik.

7.1. Rückkopplung

Von Rückkopplung (engl. feed-back, etwa zurückspeisen) wird dann gesprochen, wenn von einem Verstärker Anteile seiner Ausgangsgrößen erneut zum Eingang zurückgeleitet werden. Etwas allgemeiner ist der in der Regelungstechnik gebräuchlich gewordene Ausdruck *Rückführung*. Die allgemeine Struktur der Rückkopplung zeigt Bild 7.1.1a). Der innere Verstärker V ist der Ausgangspunkt. Für ihn besteht mit der Eingangsgröße x_i und der Ausgangsgröße x_a der Zusammenhang

$$x_a = V \cdot x_i \,, \tag{1a}$$

wobei die Größen x_a und x_i Spannungen und/oder Ströme darstellen. Für die Rückkopplung wird nun die Ausgangsgröße x_a mittels eines Vierpols K in die rückzuführende Größe x_r umgewandelt. Der Vierpol K ist ein beliebiges passives Netzwerk aus R, C und L, welches u. a. durch Frequenz- und Phasengang gekennzeichnet ist.
Die Zusammenschaltung von Verstärker V und Vierpol K zu einem neuen Verstärker V^+ erfolgt entsprechend Bild 7.1.1a) am Eingang. Die Eingangsgröße x_e wird mit der Rückkopplungsgröße x_r zur Eingangsgröße des inneren Verstärkers x_i zusammengefaßt:

$$x_i = x_e + x_r \,. \tag{1b}$$

Für den im Bild besonders hervorgehobenen, rückgekoppelten Verstärker gilt analog zu Gl. (1a) dann

$$x_a = V^+ \cdot x_e \,. \tag{1c}$$

Für die entsprechenden Parameter gilt folglich der funktionelle Zusammenhang

$$V^+ = f(V; K) \,. \tag{1d}$$

Die Größen x können in unterschiedlicher Weise Ströme bzw. Spannungen sein. So ergeben sich bezüglich des Eingangs die beiden Möglichkeiten der Teilbilder b) und c)

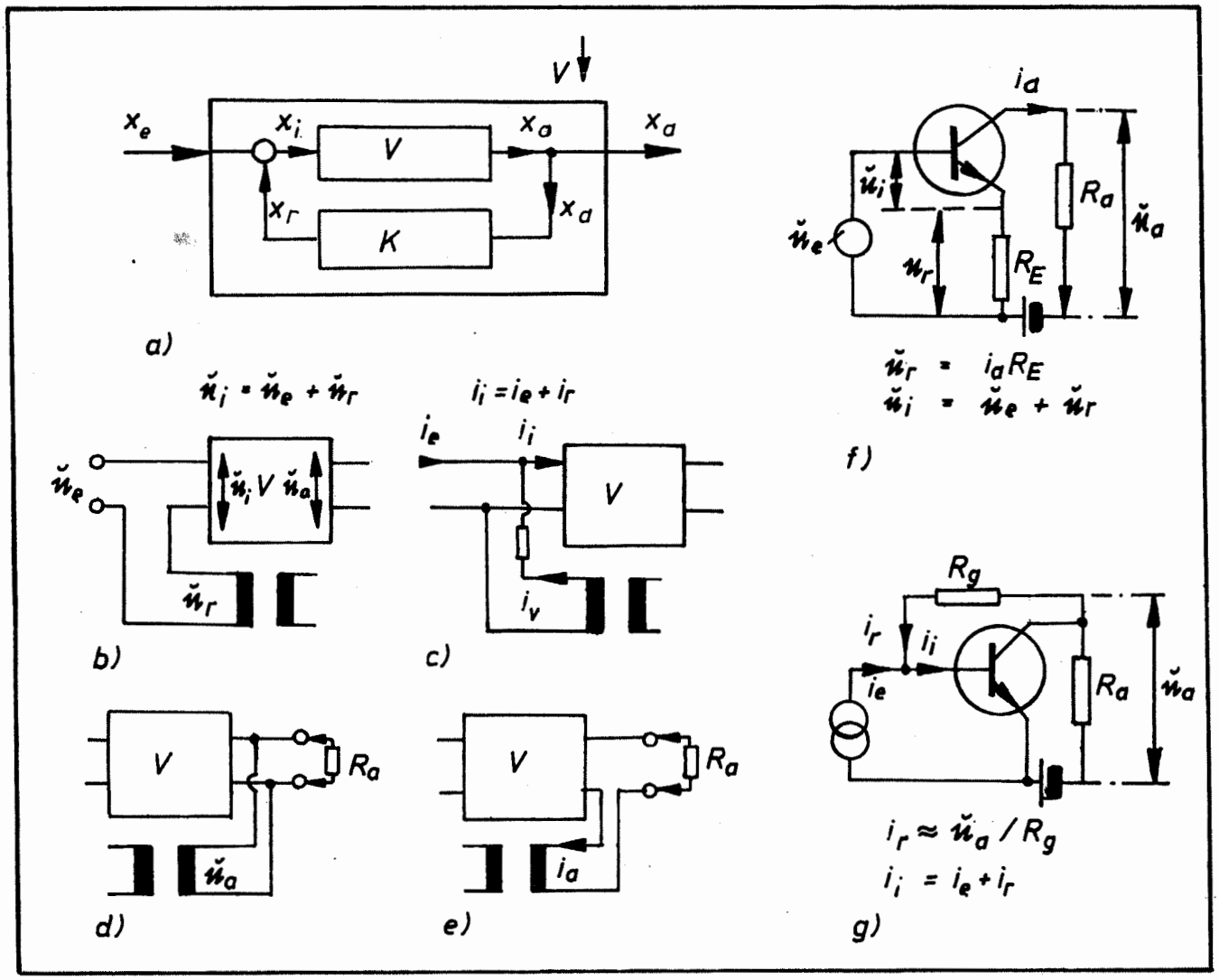


Bild 7.1.1. Prinzip der Rückkopplung

a) Allgemeines Schema aus innerem Verstärker V, passivem Rückkopplungsvierpol K und rückgekoppeltem Verstärker V^+
b) Spannungszusammenschaltung am Eingang
c) Stromzusammenschaltung am Eingang
d) Spannungsauskopplung am Ausgang
e) Stromauskopplung am Ausgang
f) Gegenkopplung mittels R_E als Beispiel einer Serien-Serien-Rückkopplung
g) Gegenkopplung mittels R_g als Beispiel einer Parallel-Parallel-Rückkopplung

und bezüglich des Ausgangs die beiden Varianten d) und e) und damit in deren Kombination die *vier Rückkopplungsprinzipien*:

	Eingang	Ausgang
Serien-Parallel-Schaltung	U	U
Serien-Serien-Schaltung	U	I
Parallel-Parallel-Schaltung	I	U
Parallel-Serien-Schaltung	I	I

Zwei Beispiele hierzu zeigt Bild 7.1.1. Die Schaltung f) verwendet den Ausgangsstrom über R_E und führt ihn in guter Näherung als Spannungsgröße an den Eingang zurück. Sie stellt also eine Serien-Serien-Rückkopplung dar. Bild g) verwendet die Ausgangsspannung und führt sie als Eingangsstrom zurück, also eine Parallel-Parallel-Schaltung. In beiden Fällen sind dies jedoch wegen der Eingangsleistung am Bipolartransistor nur Näherungen der obigen, ideal gemeinten Einteilungen. Ob der Eingang bezüglich der Rückkopplung mit Strom bzw. mit Spannung betrieben wird, ist vor

allem eine Konsequenz des inneren Verstärkers. Bipolartransistoren sind z.B. besser linear bezüglich des Stromes, Feldeffekttransistoren sind primär spannungsgesteuert. Der Unterschied, ob Strom oder Spannung am Ausgang verwendet wird, ist erst dann entscheidend, wenn R_a veränderbar bzw. nichtlinear ist. Der innere und der rückgekoppelte Verstärker kann durch *verschiedene Parameter beschrieben* werden. Hierfür sind u. a. für Bipolartransistoren die h-Parameter geeignet. Besonders übersichtlich werden jedoch die Verhältnisse, wenn die Parameter gemäß Abschnitt 3.3.7. Steilheit S, Durchgriff D und Innenwiderstand R_i verwendet werden. Dabei erhalten die Parameter des inneren Verstärkers eine zusätzliche Kennzeichnung. Beim rückgekoppelten Verstärker wird ein hochgestelltes + verwendet.

Größere praktische Bedeutung als den bisherigen Einteilungen kommt den Relationen bezüglich x_r und x_i zu. Hierdurch ergibt sich die *Dreiteilung* bezüglich des übergeordneten Begriffs *Rückkopplung*:

Gegenkopplung: Sie liegt vor, wenn x_r und x_i um 180° phasenverschoben sind. Die rückgeführte Größe also der Eingangsspannung x_e entgegenwirkt.
Mitkopplung: Sie liegt vor, wenn x_r und x_i gleichphasig schwingen. Die rückgeführte Größe vergrößert also die Wirkung der Eingangsspannung x_e.
Blindkopplung: Sie liegt vor, wenn x_r und x_i weder gleich- noch gegenphasig schwingen, insbesondere also, wenn zwischen ihnen eine Phasenverschiebung in der Umgebung von 90° besteht. Die Blindkopplung steht im Zusammenhang mit Induktivitäten und Kapazitäten (Blindelementen).

7.1.1. Phasenreine Rückkopplung

Phasenreine Rückkopplungen sind die Gegen- und Mitkopplung. Sie können gemeinsam behandelt werden und nur durch ein Vorzeichen + bzw. − unterschieden werden.

7.1.1.1. Kennwerte

Sowohl für den inneren als auch für den rückgekoppelten Verstärker werden die Kenngrößen des Abschnitts 3.3.7. und die Betrachtungen des Abschnitts 5.5.1. verwendet. Dies bedeutet, daß der Eingangsstrom des inneren Verstärkers vernachlässigbar ist. Verallgemeinert ohne inhaltliche Veränderung gelten dann diese Kenngrößen für jeden beliebigen (linear angenommenen) Verstärker unabhängig von der Stufenzahl und der Phasendrehung zwischen Ein- und Ausgang, wobei dann allerdings die Steilheit S und der Durchgriff D eventuell negatives Vorzeichen besitzen können. Zur formalen *Vereinfachung* werden im folgenden jedoch *immer positive Werte* angenommen. Gegebenenfalls kann diese Phasenumkehr in den noch einzuführenden Faktoren k und l berücksichtigt werden. Dann gelten die BARKHAUSEN-Beziehung

$$SDR_i = 1\,, \tag{2a}$$

die Steuergleichung

$$i_a = S(u_i + Du_a)\,, \tag{2b}$$

der Zusammenhang

$$u_a = -i_a R_a \tag{2c}$$

und schließlich

$$V = -\frac{u_a}{u_i} = \frac{1}{D}\,\frac{R_a}{R_a + R_i} = S\,\frac{R_a \cdot R_i}{R_a + R_i}\,. \tag{2d}$$

Außerdem sei die Stabilität als Abweichung ε der Verstärkung von ihrem Normalwert V_0 eingeführt. Diese Größe ε kann u. a. durch Temperatureinflüsse, Schwankungen der Betriebsspannung, Alterungseffekte und ähnliches bewirkt werden. Es gilt also

$$V = V_0(1 + \varepsilon) . \tag{2e}$$

Schließlich seien noch die Klirrkoeffizienten k_n gemäß Abschnitt 1.4.8. eingeführt.

Für die folgenden Berechnungen wird nun eine Struktur zugrunde gelegt, wie sie Bild 7.1.2 zeigt. Die Rückkopplungsspannung u_r kann über den Spannungsteiler P aus der Ausgangsspannung abgeleitet werden. Insbesondere dann, wenn ein Verstärker

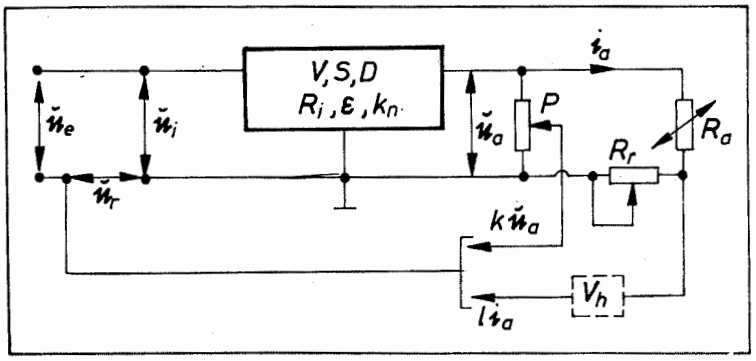


Bild 7.1.2. Anordnung zur Berechnung des rückgekoppelten Verstärkers

mit unterschiedlichen Abschlußwiderständen R_a betrieben wird, gibt es noch die Möglichkeit, sie aus dem Ausgangsstrom i_a abzuleiten. Hierzu dient der Widerstand R_r mit $R_r \ll R_a$. Ist der Spannungsabfall für die Rückkopplung nicht groß genug, so kann ein Hilfsverstärker V_h eingeschaltet werden. Die *Spannungsrückkopplung* erfolgt also über P mit dem Wert ku_a und die *Stromrückkopplung* über R_r mit dem Wert li_a. Die Größe l hat dabei die Einheit eines Widerstandes gemäß $R_r \cdot V_h$. Insgesamt gilt somit für die Rückkopplungsspannung

$$\boxed{u_r = -ku_a + li_a} . \tag{3}$$

Die Faktoren l und k sind dabei bereits im Vorzeichen so festgelegt, daß sie Auskunft über Gegen- oder Mitkopplung geben. Es gilt

$$\begin{Bmatrix}\operatorname{sign} k\\ \operatorname{sign} l\end{Bmatrix} = \begin{Bmatrix}+\\ -\end{Bmatrix} \text{ bei } \begin{Bmatrix}\text{Mitkopplung}\\ \text{Gegenkopplung}\end{Bmatrix}.$$

Wird Gl. (3) in Gl. (2b) eingesetzt, so folgt

$$i_a\left(\frac{1}{S} - l\right) = u_e + u_a(D - k) . \tag{4}$$

Hieraus folgen sofort zwei *Kennwerte des rückgekoppelten Verstärkers* mittels Vergleich mit (2b), nämlich

$$\boxed{S^+ = \frac{S}{1 - lS}} \tag{5}$$

und

$$\boxed{D^+ = D - k} . \tag{6}$$

Stromrückkopplung ändert also die Steilheit und nicht den Durchgriff. Bei Gegenkopplung verkleinert sich die Steilheit, bei Mittkopplung wird sie erhöht. Für

$$S = 1/l \tag{7}$$

entsteht eine unendlich große Steilheit. Dies bedeutet noch nicht unbedingt eine Instabilität des rückgekoppelten Verstärkers. Gemäß Abschnitt 5.5.2. wird nämlich durch den Arbeitswiderstand die Arbeitssteilheit kleiner als die statische Steilheit. Erst wenn diese gegen Unendlich geht, wird der Verstärker instabil, z.B. durch Eigenschwingungen.

Die Spannungsrückkopplung ändert den Durchgriff des Verstärkers. Eine Gegenkopplung macht ihn größer, eine Mitkopplung kleiner. Bei

$$D = k \tag{8}$$

wird er Null. Der Verstärker arbeitet dann vollkommen rückwirkungsfrei.

Entsprechend der BARKHAUSEN-Beziehung (2a) folgt für den Innenwiderstand:

$$R_i^+ = \frac{1}{S^+ D^+} = R_i \frac{1 - lS}{1 - k/D}. \tag{9}$$

Bei einer Dimensionierung gemäß Gl. (8) strebt der Innenwiderstand gegen Unendlich. Ein derartig dimensionierter Verstärker stellt also eine *ideale Stromquelle* dar. Der Anodenstrom des Verstärkers ist damit unabhängig von dem jeweils angeschlossenen Arbeitswiderstand. Grenzen treten erst dann auf, wenn die Ströme oder Spannungen die Grenzen des Kennlinienfeldes überschreiten. Genau werden diese Grenzen im Abschnitt 7.4. behandelt.

Bei der Dimensionierung nach Gl. (7) wird der *Innenwiderstand* des Verstärkers zu *Null*. Seine Ausgangsspannung ist also unabhängig von der Last. Damit besitzt der Verstärker die Eigenschaften einer idealen Spannungsquelle. Auch dies gilt nur so lange, wie nicht die Grenzkennlinien überschritten werden.

Eine Anwendung des Innenwiderstandes Null besteht in der Resonanzunterdrückung von schwingfähigen Gebilden, z.B. Lautsprechern. Diese Zusammenhänge wurden bereits im Abschnitt 2.2.3. erläutert. Wird die Strommitkopplung stärker als nach Gl. (7) ausgelegt, so tritt ein negativer Innenwiderstand auf. Seine Eigenschaften sind in Abschnitt 7.1.1.3. behandelt.

7.1.1.2. Änderungen im Kennlinienfeld

Die bisher diskutierten Zusammenhänge des rückgekoppelten Verstärkers lassen sich im Kennlinienfeld veranschaulichen. Für Bild 7.1.3 wurden dazu wenige Kennlinien eines FET und ein Arbeitspunkt ausgewählt. Alle Betrachtungen gelten nur in der Umgebung dieses Arbeitspunktes.

Die Stromrückkopplung beeinflußt die Steilheit des Verstärkers. Dies bedeutet, daß sie die Eingangskennlinie (links in Bild 7.1.3) verdreht. Der Durchgriff beeinflußt dagegen den Abstand der benachbarten Eingangskennlinien. Bei $k = 0$ fallen alle Kennlinien zusammen, für $k > D$ vertauschen sie ihre Lage (rechts bzw. links zur Bezugskennlinie).

Im Ausgangskennlinienfeld wird die Neigung der Kennlinien entsprechend Gl. (9) geändert. Sie können steiler und flacher verlaufen. In den Sonderfällen von Gl. (7) und (8) verlaufen sie genau waagerecht bzw. senkrecht.

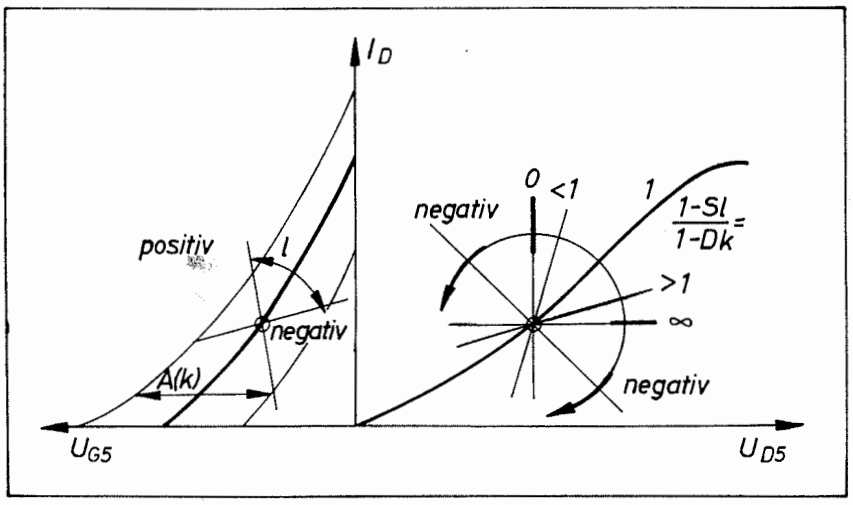


Bild 7.1.3. Zur Wirkung der Rückkopplung in den Kennlinienfeldern. Die Stromrückkopplung verändert die Steilheit der Eingangskennlinie, die Spannungsrückkopplung verändert den Abstand A der Kennlinien. Er wird größer für $k < 0$; er wird kleiner für $D > k > 0$, er verschwindet für $D = k$, und die Lage der Kennlinien rechts und links vertauscht sich für $k > D$, wobei mit wachsendem k der invertierte Abstand ständig zunimmt. Spannungs- und Stromrückkopplungen zusammen beeinflussen gemeinsam die Steilheit der Kennlinie im Ausgangskennlinienfeld (rechts)

7.1.1.3. Negativer Innenwiderstand

Gemäß Bild 7.1.3 bzw. Gl. (9) lassen sich durch eine Rückkopplung auch negative Innenwiderstände erzeugen. Sie besitzen Eigenschaften, die bisher noch nicht behandelt waren. Deshalb ist hierauf näher einzugehen. Zunächst gelten die Betrachtungen auch nur in der Umgebung des Arbeitspunktes.

In der statischen Kennlinie ist ein negativer Widerstand dadurch gekennzeichnet, daß sie eine negative Steigung besitzt. Dies bedeutet, daß der Anodenstrom fällt, wenn die Anodenspannung der Röhre zunimmt. Strom und Spannung verändern sich also genau gegenphasig (180° Phasenverschiebung). Damit erfüllt der negative Innenwiderstand die in Bild 2.2.1 fehlende linke Halbachse. Beim SMITH-Diagramm liegen seine Werte außerhalb des Einheitskreises.

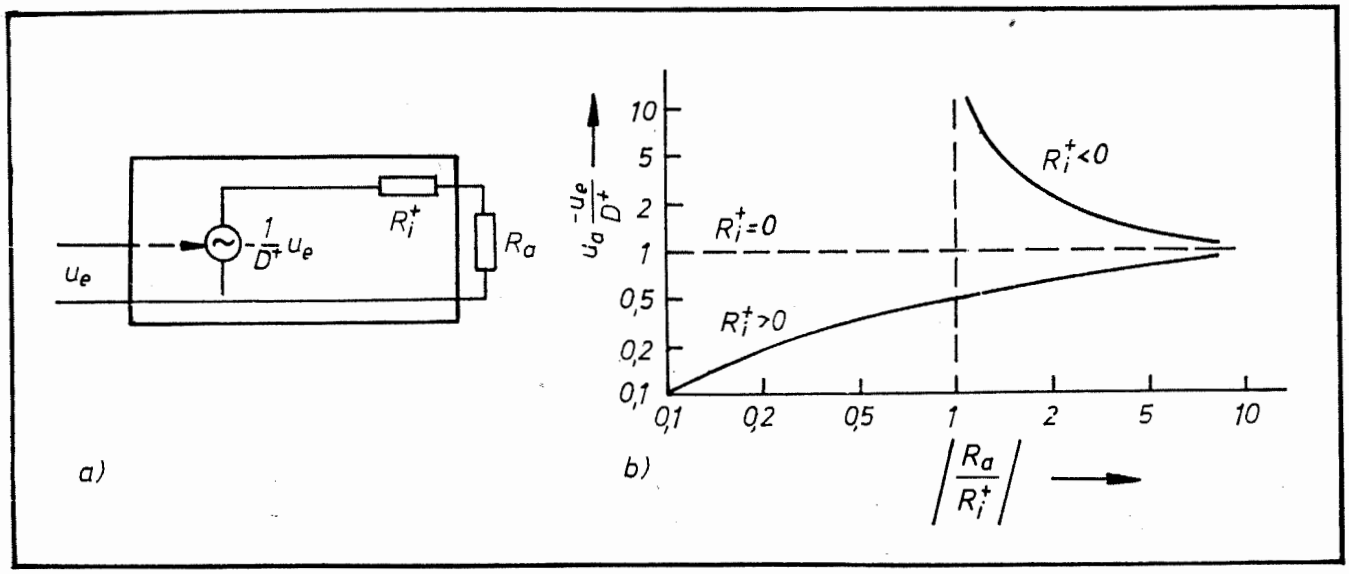


Bild 7.1.4. Zur Wirkung des negativen Innenwiderstandes

a) Spannungsersatzschaltung eines Verstärkers

b) Änderung der relativen Ausgangsspannung in Abhängigkeit vom Arbeitswiderstand bei negativem bzw. positivem Innenwiderstand

Eine weitere Erklärung des negativen Innenwiderstandes läßt sich mittels des Spannungs-Ersatzschaltbildes des FET von Bild 5.5.2b) geben. In abgewandelter Form gilt dann Bild 7.1.4a). Die Spannung am Außenwiderstand beträgt

$$u_a = -\frac{u_e}{D^+} \cdot \frac{R_a}{R_i^+ + R_a}. \tag{10}$$

Es werde nun die Eingangsspannung u_e konstant gehalten. Dann ist auch die innere Spannung des Ersatzschaltbildes konstant. Interessant ist der Zusammenhang $u_a = f(R_a)$, wobei R_i als Parameter auftritt. Normiert zeigt diesen Zusammenhang Bild 7.1.4b). Für positive Innenwiderstände fällt die Ausgangsspannung mit kleiner werdendem R_a ab. Für einen Innenwiderstand von Null ist die Spannung unabhängig von R_a und gleich der inneren Spannung der Ersatzschaltung. Dieser Verlauf ist in Bild 7.1.4b) gestrichelt gezeichnet. Bei negativen Innenwiderständen steigt sie an und erreicht schließlich bei $-R_i^+ = R_a$ den Wert Unendlich. Beim negativen Innenwiderstand ist die Ausgangsspannung an R_a also stets größer als die innere Spannung des Ersatzschaltbildes. Dies ist dadurch erklärbar, daß der negative Innenwiderstand einen Teil der Verluste des Arbeitswiderstandes kompensiert. Der negative Innenwiderstand ist also ein aktiver, Energie erzeugender Widerstand. Dies gilt sogar dann, wenn an dem Verstärker keine Eingangsspannung liegt. Der Verstärker wird dann zu einem aktiven Zweipol ohne innere Spannungsquelle. Er ist lediglich durch den Widerstand R_i^+ gemäß Gl. (9) gekennzeichnet.

In der letzten Zeit hat der negative Innenwiderstand als sogenannter Zweipolverstärker größere Bedeutung erlangt. Seine Anwendung sei am Beispiel einer Leitung, z. B. Telefonleitung erklärt. Sie besitzt Verluste, die durch eine Kettenschaltung aus Längs- und Querwiderständen gemäß Bild 7.1.5 darstellbar sind. Ein negativer Widerstand parallel zu den Querwiderständen gestattet, teilweise oder ganz diese Verluste zu beseitigen. Vereinfacht sei dies für einen Spannungsteiler gemäß Bild 7.1.5b) berechnet. Hierzu werde der Innenwiderstand des Spannungsteilers

$$R_{is} = \frac{R_1 \cdot R_2}{R_1 + R_2} \tag{11}$$

eingeführt. Mit ihm folgt für die Ausgangsspannung nach kurzer Rechnung

$$u_2 = u_1 \frac{R_2}{R_1 + R_2} \cdot \frac{1}{1 + \frac{R_{is}}{R_i}}. \tag{12}$$

Sobald

$$R_i^+ = -R_2 \tag{13}$$

ist, verschwindet die Dämpfung des Spannungsteilers. Die Verluste von R_2 werden dann gerade aufgehoben.

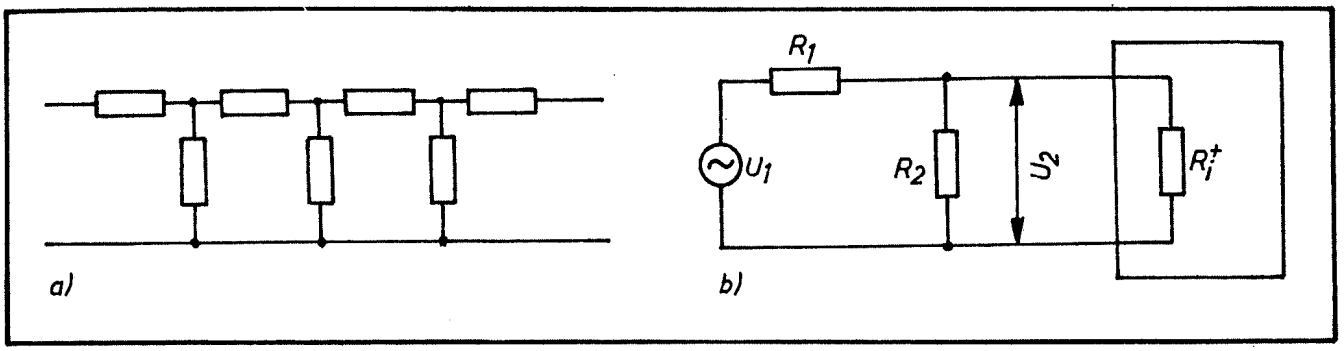


Bild 7.1.5. Zur Eigenschaft des Zweipolverstärkers mit negativem Innenwiderstand

a) Vereinfachte Darstellung einer Leitung als Kettenschaltung

b) Einwirkung des Zweipolverstärkers zur Entdämpfung der Leitung

Über Leitungen werden normalerweise Signale (Spannungen) in beiden Richtungen übertragen. Bei der Verwendung der üblichen Vierpolverstärker ist es notwendig, für jede Richtung einen getrennten Verstärker einzufügen. Außerdem sind Kunstschaltungen notwendig, damit sich die Verstärker nicht im Sinne einer gegenseitigen Mitkopplung erregen. Der Zweipolverstärker mit dem negativen Innenwiderstand bringt den weiteren Vorteil, daß seine Entdämpfung unabhängig von der Richtung der Signalübertragung wirkt.

7.1.1.4. Verstärkung und Stabilität

Die Größe der Verstärkung des rückgekoppelten Verstärkers kann am besten aus den Gl. (2d) und (3) berechnet werden. Es gilt

$$-\frac{u_a}{V} = u_i = u_e - k u_a + l i_a \,. \tag{14}$$

Gemäß der Definition von Gl. (2c) und (2d) folgt

$$V^+ = \frac{1}{\frac{1}{V} - k - \frac{l}{R_a}} \,. \tag{15}$$

Eine Mitkopplung erhöht also unabhängig davon, ob sie vom Ausgangsstrom oder von der Ausgangsspannung abgeleitet wird, die Verstärkung; eine Gegenkopplung senkt sie. Es sei angenommen, daß infolge verschiedener Störungen die Verstärkung um einen relativen, kleinen Betrag ε schwanke. Dann gilt Gl. (2c), und es wird die Verstärkung des rückgekoppelten Verstärkers V^+ um einen anderen Betrag ε^+ schwanken:

$$V^+ = V_0^+(1 + \varepsilon^+) \,. \tag{16}$$

Es sei ε klein gegen 1 bzw. V_0/V_0^+. Wegen

$$\frac{1}{1+\varepsilon} \approx 1 - \varepsilon \tag{17}$$

folgt dann aus Gl. (15) nach kurzer Rechnung

$$\boxed{\frac{V^+}{\varepsilon^+} = \frac{V}{\varepsilon}} \,. \tag{18}$$

Wird also mittels Gegenkopplung die Verstärkung gesenkt, so erhöht sich um den gleichen Faktor die relative Stabilität des Verstärkers und umgekehrt. Mittels einer Gegenkopplung lassen sich also sehr konstante Verstärkungsfaktoren erzeugen. Dieses Ergebnis ist auch direkt aus Gl. (15) zu erkennen. Liegen sehr große Verstärkungen $V_0 \to \infty$ sowie eine starke Spannungsgegenkopplung und keine Stromrückkopplung vor, so ist V in Gl. (16) nur noch eine Korrekturgröße. Mit guter Näherung gilt dann

$$V^+ \approx -\frac{1}{k} \,. \tag{19}$$

Die Verstärkung wird praktisch ausschließlich durch den Spannungsteiler in Bild 7.1.2 festgelegt. Ein Spannungsteiler läßt sich aber extrem konstant aufbauen.

7.1.1.5. Nichtlinearitäten

In Analogie zu Abschnitt 1.4.7., (1.4.14), werde der nichtlineare Zusammenhang $u_a = f(u_i)$ durch eine Potenzreihenentwicklung genähert:

$$u_a = -V_1 u_i + V_2 u_i^2 + V_3 u_i^3 + \ldots \tag{20}$$

Für die Berechnung ist eine Auflösung dieser Gleichung nach u_i (Umkehrung) notwendig:

$$u_i = -\frac{u_a}{V_1} + \frac{V_2}{V_1^3} u_a^2 - \frac{2V_2 + V_1 V_3}{V_1^5} u_a^3 + \dots \tag{21}$$

Da die folgenden Betrachtungen für einen konstanten Arbeitswiderstand durchgeführt werden sollen, ist es nicht notwendig, zwischen Strom- und Spannungsrückkopplung zu unterscheiden. Statt Gl. (1b) und (3) gilt dann mit dem einzigen Rückkopplungsfaktor m

$$u_i = u_e - m u_a \,. \tag{22}$$

Wird dieser Zusammenhang in Gl. (21) eingesetzt und werden die dann entstehenden Gleichungen $u_e = f(u_a)$ wiederum umgekehrt, so folgt

$$u_a = \frac{1}{1 - mV_1}\left\{-V_1 u_e + \frac{V_2}{(1 - mV_1)^2} u_e^2 + \left(\frac{V_3}{(1 - mV_1)^3} - \frac{2mV_2^2}{(1 - mV_1)^4}\right) u_e^3 + \dots\right\}. \tag{23}$$

Der Ausdruck vor der geschweiften Klammer gibt die bekannte Verstärkungsänderung durch die Rückkopplung an (vgl. (15)). Der Ausdruck in der geschweiften Klammer ist eine Gl. (20) analoge Potenzreihenentwicklung. Durch Koeffizientenvergleich ergibt sich ein Zusammenhang mit den Klirrkoeffizienten des rückkopplungsfreien und des rückgekoppelten Verstärkers. Gemäß den Klirrkoeffizienten von Abschnitt 1.4.7. bzw. Gl. (1.4.14) folgt

$$\frac{k_2^+}{k_2} = \frac{1}{(1 - mV_1)^2}\,, \tag{24a}$$

$$\frac{k_3^+}{k_3} = \frac{1}{(1 - mV_1)^3} - \frac{2mV_2^2}{V_3(1 - mV_1)^4} \approx \frac{1}{(1 - mV_1)^3}\,. \tag{24b}$$

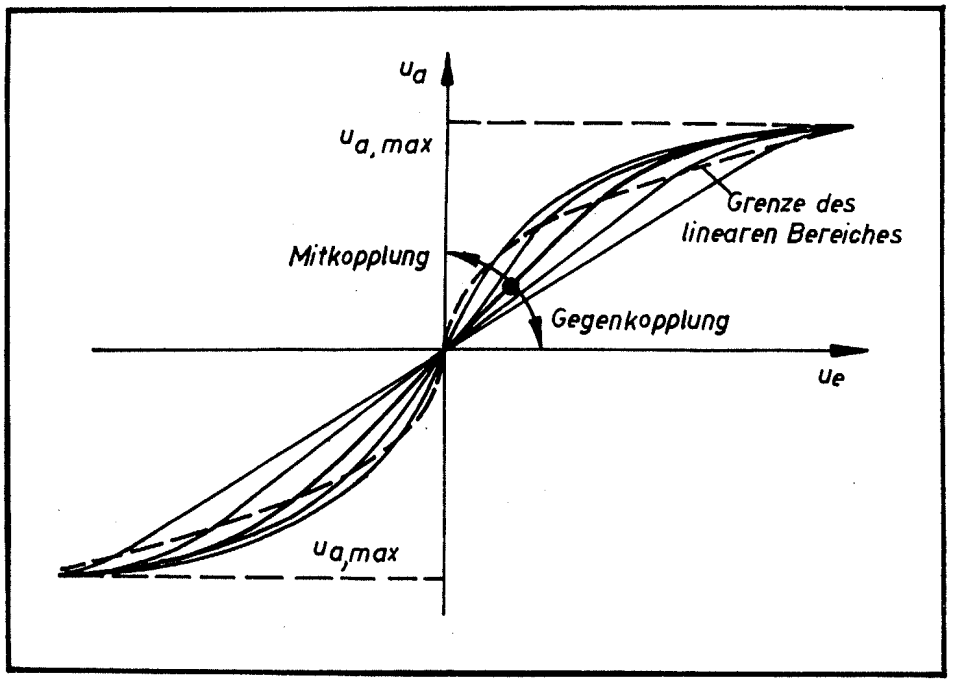


Bild 7.1.6. Zur Kennlinienlinearisierung durch Gegenkopplung sowie die noch stärker werdende Krümmung bei Mitkopplung. Die Linearitätsgrenze ist gestrichelt gezeichnet. Hieraus ist ersichtlich, wie stark die Rückkopplung den Linearitätsbereich beeinflußt

Die Klirrkoeffizienten und damit auch der gesamte Klirrfaktor nehmen also bei Gegenkopplung (m negativ) erheblich stärker ab, als der Verstärkungsfaktor gesenkt wird. Dies bedeutet, daß eine Gegenkopplung die Kennlinien eines Verstärkers stark linearisiert. Die Tangenten im Arbeitspunkt gemäß Bild 7.1.3 gelten also bei Gegenkopplung auch noch in relativ weiter Entfernung vom Arbeitspunkt. Bei extrem starker Gegenkopplung läßt sich sogar erreichen, daß die Kennlinie bis zu den Grenzen des möglichen Aussteuerungsbereiches (vgl. Abschnitt 7.4.) praktisch linear bleibt und dann unstetig umknickt. Bei einer Mitkopplung nimmt die Kennlinienkrümmung schneller als der Verstärkungsgewinn zu. Bild 7.1.6 zeigt den prinzipiellen Verlauf.

7.1.2. Frequenzabhängige Rückkopplungen

7.1.2.1. Frequenzselektive Verstärker

Bisher waren sowohl der Verstärker als auch die Rückkopplung phasenfrei und frequenzunabhängig angenommen. Jetzt werde für die Rückkopplung ein Frequenz- und Phasengang zugelassen. Als besonders übersichtlich erweisen sich dann Betrachtungen mit Hilfe der Ortskurven. Wegen der einfachen Kreisform sei als Beispiel das WIEN-Glied betrachtet. Mit ihm ergibt sich eine Anordnung, wie sie Bild 7.1.7a) zeigt. Zunächst sei ein Verstärker angenommen, bei dem u_i und u_a gegenphasig schwingen. Dann gilt die Konstruktion von Bild 7.1.7d). Aus ihm lassen sich die Frequenzgänge rechts daneben konstruieren bzw. berechnen. V_0 sei die Verstärkung des rückkopplungsfreien Verstärkers. Für große V_0 ist u_i klein, und für $V_0 \to \infty$ gilt näherungsweise $\boldsymbol{u}_e \approx -\boldsymbol{u}_r$. Die Verstärkung ist dann also reziprok zur Ausgangsspannung des RC-Gliedes. Sie ist am geringsten bei der Frequenz $\Omega = 1$ und zeigt damit das *Verhalten eines Serienkreises*. Je geringer die Verstärkung ist, desto länger wird $\boldsymbol{u}_i$ und desto weniger macht sich der Einfluß des WIEN-Gliedes bemerkbar.

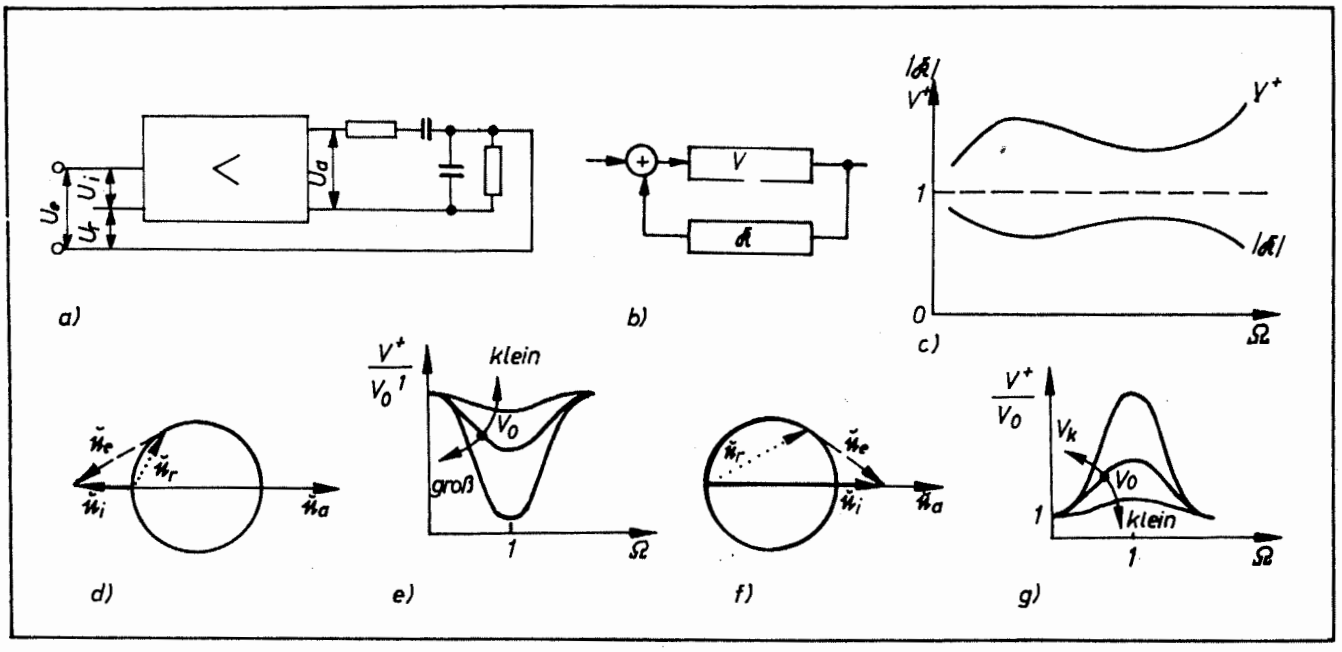


Bild 7.1.7. Zur frequenzabhängigen Rückkopplung

a) Phasenreiner und frequenzganglinearer Verstärker, der über ein WIEN-Glied rückgekoppelt ist
b) Phasenreiner und frequenzganglinearer Verstärker, der über einen komplexen Vierpol $\boldsymbol{K}$ rückgekoppelt ist
c) Bei hinreichend großer Verstärkung in b) sind Betrag von Verstärkung V^+ und $\boldsymbol{K}$ reziprok (Gl. (26))
d) Ortskurve und e) Frequenzgänge bei einem einstufigen Verstärker, $\boldsymbol{u}_i$ und $\boldsymbol{u}_a$ gegenphasig
f) Ortskurve und g) Frequenzgang, wenn der Verstärker in a) keine Phasendrehung besitzt

Die Verstärkungskurve tendiert zu einem relativ flachen Frequenzgang. Mittels der Verstärkung lassen sich unterschiedliche Frequenzgänge einstellen.

Für einen Verstärker, der zwischen dem Ein- und Ausgang keine Phasendrehung besitzt, gilt Bild 7.1.7f). Bei einer relativ kleinen Verstärkung V_0 ist die Eingangsspannung u_i größer als der Durchmesser der Ortskurve. Für die Frequenz $\Omega = 1$ liegt eine Mitkopplung vor, wodurch die Verstärkung V^+ größer als V_0 ist. Die Verstärkung wächst hier ersichtlich an, je mehr sich die Größe der Spannung u_i dem Durchmesser des Kreises der Ortskurve nähert. Wird die Verstärkung noch größer, so treten Fälle auf, die erst im Abschnitt 7.1.2.2. bzw. 7.1.3. behandelt werden.

Der Durchmesser des Kreises war in Gl. (2.3.8) für alle *RC*-Glieder als Größe v_k eingeführt. Die weiteren Betrachtungen erfolgen in dem Verstärkungsbereich, wo

$$u_i > \frac{u_a}{v_k} \tag{25}$$

gilt. Den zugehörigen Frequenzgang des Verstärkers in Abhängigkeit von der Verstärkung V_0 zeigt Bild 7.1.7g). Je stärker die relative Rückkopplung wird, desto mehr selektiv wird die Schaltung.

Sowohl beim rück- wie beim gegengekoppelten, frequenzselektiven Verstärker ist ersichtlich, daß die Bandbreite sich reziprok mit der Verstärkung ändert. Insbesondere ist es interessant, daß die mitgekoppelte Schaltung so wirkt, als ob ein Parallelschwingkreis einstellbarer Güte vorhanden ist. Im Sinne von Gl. (5.5.17) folgt auch umgekehrt, daß durch die Gegenkopplung eine Erweiterung des Frequenzbereiches von Verstärkern möglich ist. Dies kommt so zustande, daß der Gegenkopplungsgrad in der Nähe der Grenzfrequenzen abnimmt und auf diese Weise die Verstärkung den Werten in der Frequenzmitte genähert wird. Auf diese Weise lassen sich vorteilhaft breite Frequenzbänder mit relativ stromarmen Verstärkeranordnungen aufbauen. Insgesamt gilt dies auch für den frequenzselektiven Verstärker von Abschnitt 5.5.6. Nähere Betrachtungen zeigen aber auch die Gültigkeit der Gl. (18) bezüglich der Stabilität.

Die Kurven in Bild 7.1.7 zeigen, daß der Frequenzgang eines Verstärkers stark mit dem des Rückkopplungsgliedes zusammenhängt. Durch bestimmte frequenzabhängige Rückkopplungsglieder lassen sich gewünschte Frequenzgänge realisieren. Dieses Prinzip wird vielfältig bei Frequenzgangentzerrungen, u. a. in Magnetbandgeräten, Tonabnehmerverstärkern und bei Tonblenden verwendet. Vorteile der frequenzabhängigen Rückkopplungsschaltungen bestehen vor allem darin, daß mit ihnen ein geringeres Verstärkerrauschen, ein kleinerer Klirrfaktor und eine bessere Stabilität erreicht werden kann. Der Nachteil derartiger Schaltungen liegt vor allem in ihrem unübersichtlicheren Phasengang.

Entsprechende Zusammenhänge zeigt Bild 7.1.7b). $\boldsymbol{K}$ stellt darin das komplexe Übertragungsmaß des Rückkopplungszweipols dar. Die Verstärkung V des Verstärkers sei reell mit einem Phasenwinkel π, so daß Gegenkopplung auftritt. Dann gilt Gl. (15), nur mit dem Unterschied, daß $l = 0$ ist und alle Größen bis auf V komplex sind, also

$$\boldsymbol{V}^+ = \frac{1}{\frac{1}{V} + \boldsymbol{K}} \approx \frac{1}{\boldsymbol{K}}. \tag{26}$$

Da in den meisten Fällen V sehr groß gewählt wird, gilt annähernd, daß $\boldsymbol{K}$ und $\boldsymbol{V}^+$ reziprok zueinander sind. Für den Betrag demonstriert dies Bild 7.1.7c).

7.1.2.2. Schwingungserzeugung

Die Verstärkung in Bild 7.1.7f) sei jetzt so gewählt, daß die Spannung u_i etwa dieselbe Länge wie der Durchmesser des Kreises habe. Die Differenz zwischen beiden sei durch eine kleine Spannung u_ε mit der entgegengesetzten Richtung wie u_i gegeben. Es gelte also

$$u_i - u_\varepsilon = \frac{u_a}{v_k}. \tag{27}$$

Bei der Frequenz $\Omega = 1$ ist nur diese kleine Spannung u_ε notwendig, um die Ausgangsspannung u_a zu erzeugen. Für jede andere Frequenz ist eine viel größere Spannung erforderlich. Die erforderliche Spannung u_ε sei so klein, daß sie bereits durch die *Rauschspannung* der Widerstände und Transistoren im Verstärker vorhanden ist. Dann wird dieser Verstärker ohne zusätzliche Eingangsspannung genau diese Frequenz erzeugen. Die statistischen Amplitudenschwankungen des Rauschens lassen sich leicht automatisch unterdrücken. Bestehen bleibt jedoch die statistische Frequenzschwankung. Infolgedessen besitzen alle *Oszillatoren* eine gewisse *Frequenzunsicherheit*. Die Häufigkeitsverteilung der erzeugten Frequenz entspricht dabei einer Kurve von Bild 7.1.7g). Die mögliche Schwankung der erzeugten Frequenz wird um so geringer, je geringer das Rauschen des Verstärkers im Vergleich zu seiner Ausgangsspannung und je besser die Selektivität der Rückkopplung sind. Aus der Weitab-Selektion der Ortskurve und mit den Kennlinienkrümmungen läßt sich auch der Oberwellenanteil von Generatoren berechnen. Entsprechend dieser Darstellung ist also das Rauschen der eigentliche Ausgangspunkt für die Erzeugung von Spannungen durch Generatoren.

Die Schwingungserzeugung kann allgemein für Verstärker $\boldsymbol{v}$ und Rückkopplungsglieder $\boldsymbol{K}$ berechnet werden, die frequenz- und phasenabhängig sind. Ihre Größen werden analog Bild 7.1.1a) bezeichnet. Entsprechend Gl. (27) entsteht genau dann eine Frequenz, für die

$$\boxed{\boldsymbol{Kv} = 1 - \varepsilon} \tag{28}$$

erfüllt ist. Dabei bedeutet ε wiederum die möglichst kleine Größe des relativen Rauschanteils der Schaltung. Es kann grundsätzlich nur eine Frequenz erzeugt werden, für die $\boldsymbol{Kv}$ reell und positiv ist. Außerdem muß für dieses Produkt der Betrag ganz nahe bei Eins liegen. Ist er kleiner, so entsteht ein selektiver Verstärker. Ist er größer, so ergibt sich der Fall, wie er im Abschnitt 7.1.3. behandelt wird.

7.1.2.3. Grenzen der Gegenkopplung

Bisher wurden betont phasenreine Verstärker betrachtet. Jeder Verstärker hat obere und untere Grenzfrequenzen. Zur Vereinfachung sei ein *RC*-Verstärker betrachtet, wie er im Abschnitt 5.5.3. behandelt wurde. Dieser Fall gilt generell, denn zum Arbeitswiderstand liegen in jedem Fall Schaltkapazitäten parallel, außerdem aber auch die Ein- und Ausgangskapazitäten des jeweiligen aktiven Bauelementes, sei es nun FET oder Bipolartransistor. Bei mittleren Frequenzen ist die Phasendrehung Null bei den Grenzfrequenzen 45°, und maximal kann sie den Wert in der Nähe von 90° erreichen. Bei zwei Verstärkerstufen bleibt der Winkel also normalerweise noch unter 180°. Dieser Winkel würde die in einem mittleren Frequenzbereich mit $\varphi = 0$ vorhandene Gegenkopplung in eine Mitkopplung verwandeln. Da aber beim zweistufigen Verstärker für 180° die Verstärkung gegen Null geht, kann bei ihm normalerweise noch keine Instabilität auftreten. Wesentlich ungünstiger sind die Verhältnisse beim dreistufigen Verstärker. Zu 180° gehören bei ihm noch beträchtliche Verstärkungen. Es sei zunächst vereinfachend angenommen, daß alle drei Stufen gleichartige Eigenschaften besitzen, dann sind je Stufe 60° notwendig, und das jeweilige *RC*-Glied dämpft nur auf 0,5, also alle drei auf 1/8. Hierbei kann die bei mittleren Frequenzen vorhandene Gegenkopplung zur Mitkopplung mit Schwingungserzeugung übergehen. Die genauere Abschätzung dieser Zusammenhänge erfolgt am Beispiel des Verstärkers von Bild 7.1.8a).

Im mittleren Frequenzbereich ist er gemäß dem Spannungsteilerverhältnis R_1/R_2 gegengekoppelt. Sein Verhalten bei höheren Frequenzen wird durch die Grenz-

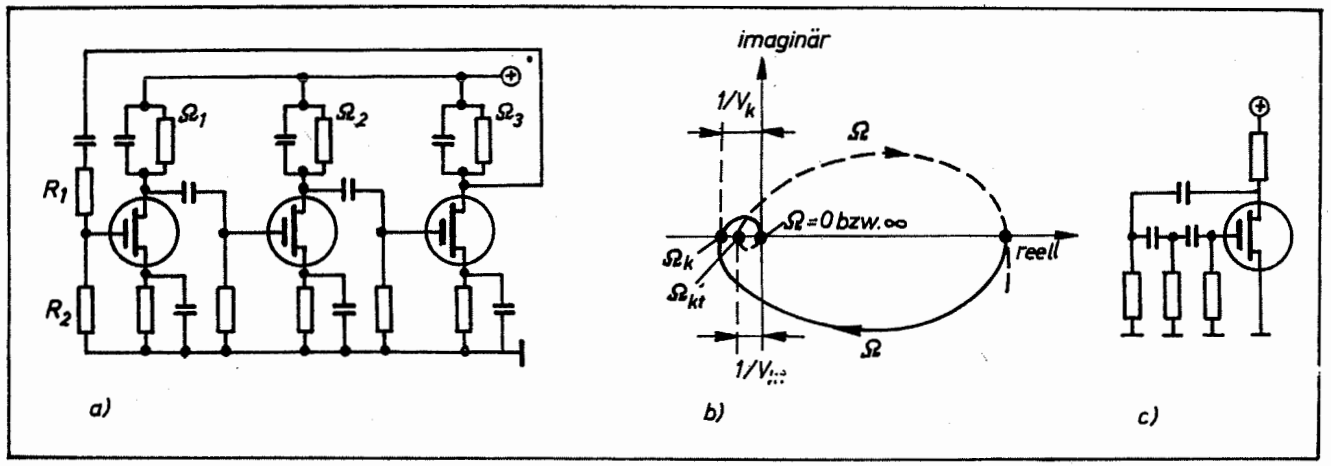


Bild 7.1.8. Zur Bestimmung der Grenzen der Gegenkopplung beim dreistufigen Verstärker
a) Schaltung
b) Ortskurve des Frequenz- und Phasenganges
(ausgezogen: der drei Tiefpaßfglieder, gestrichelt: der drei Hochpaßglieder)
c) *RC*-Generator nach diesem Prinzip

frequenzen Ω_1 bis Ω_3 der *RC*-Glieder festgelegt. Sie allein führen zu einer Ortskurve, wie sie Bild 7.1.8b) zeigt:

$$\frac{1}{K} = 1 - (\Omega_1\Omega_2 + \Omega_1\Omega_3 + \Omega_2\Omega_3)$$

$$+j(\Omega_1 + \Omega_2 + \Omega_3 - \Omega_1\Omega_2\Omega_3)\,. \tag{29}$$

Diese Ortskurve schneidet die negative reelle Achse. Für die zugehörige Frequenz Ω_k besteht Schwingungsneigung. Ist der Abstand vom Ursprung $1/v_k$, so ergibt sich für die maximal zulässige Gegenkopplung

$$k = \frac{u_r}{u_a} < \frac{1}{v_k}. \tag{30}$$

Eine stärker ausgelegte Gegenkopplung würde zwar im mittleren Frequenzbereich die Verstärkung weiter senken, aber gleichzeitig für die Frequenz Ω_k den Verstärker zum Generator machen.

Allgemein kann aus diesen Betrachtungen gefolgert werden, daß eine *Gegenkopplung über mehr als zwei Verstärkerstufen*, genauer 2 *RC*-Glieder oder einen Schwingkreis, zu *Instabilitäten* führen kann. Die Auswirkung der Instabilitäten zeigt sich an den Grenzen des Übertragungsbereiches. Ist eine Rückkopplung über drei Glieder notwendig, so kann durch eine *günstige Wahl* der einzelnen *Grenzfrequenzen* ein gewisser Vorteil für die maximal zulässige Gegenkopplung erreicht werden. Sind alle *RC*-Glieder auf eine Frequenz abgestimmt, also $\Omega_1 = \Omega_2 = \Omega_3$, so folgt die kritische Verstärkung zu $v_k = -8$. Wird dagegen ein Glied um den Faktor A verstimmt, d. h.

$$\Omega_1 = \Omega_2 = A\Omega_3\,, \tag{31}$$

so gilt

$$-v_k = 4 + 2A + \frac{2}{A}. \tag{32}$$

Je mehr der Faktor A von Eins abweicht, um so größer ist der Gewinn bezüglich einer maximalen Gegenkopplung.

Genau wie bei den hohen Frequenzen besteht beim Verstärker von Bild 7.1.8a) auch für die tiefen Frequenzen die Gefahr der Selbsterregung. Dies bewirken die drei *RC*-

Glieder zwischen den Stufen. Die zugehörige Ortskurve ist gestrichelt im Bild 7.1.8b) eingezeichnet. Die zugehörige *kritische Frequenz* ist Ω_{kt} und die kritische Verstärkung v_{kt}.

Die *kritische Rückkopplung* über drei und mehr *RC*-Glieder kann auch bewußt zur Schwingungserzeugung benutzt werden. Auf diesem Prinzip beruht der Phasenschiebergenerator des Bildes 7.1.8c).

7.1.3. Pendelrückkopplung

Zunächst sei noch einmal zur Einleitung dieses Kapitels zurückgekehrt. Im Bild 7.1.9 ist eine Einteilung der möglichen Rückkopplungsfälle zusammengestellt. Sie ist auf die eingangsseitigen Spannungen des Bildes 7.1.2 bezogen. Alle Fälle bis auf den einen von d) sind bisher besprochen. Hier ist die Rückkopplungsspannung größer als u_i, und zugleich zeigen beide in dieselbe Richtung. Alle anderen Arten der Rückkopplung konnten statisch, d. h. im eingeschwungenen Zustand betrachtet werden. Dieser Spezialfall erfordert eine *dynamische Behandlung*. Wesentlich dabei ist, daß es eine

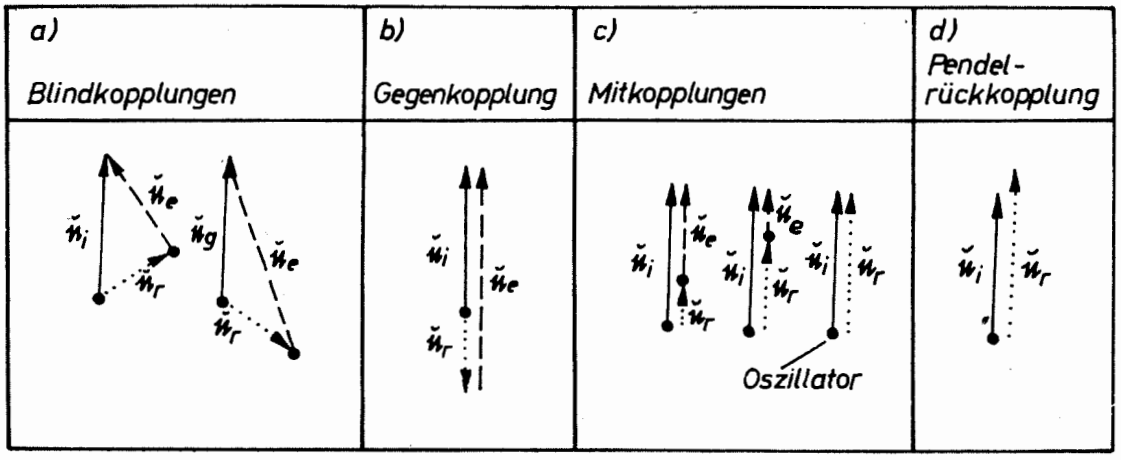


Bild 7.1.9. Phasenlage und Amplitude der rückgekoppelten Spannung bestimmen das Verhältnis von neuer Eingangsspannung u_e und Eingangsspannung ohne Rückkopplung u_i

bestimmte Zeit dauert, bis eine Eingangsschwingung u_i die Ausgangsschwingung erzeugt. Erst nach dieser *Verzögerung* steht auch die Rückkopplungsschwingung zur Verfügung. Für einen stark selektiven Resonanzverstärker (Bild 7.1.7 oder 7.1.8) beträgt diese Verzögerungszeit ziemlich genau $1/f$. Weiter ist das Verhältnis von $u_r : u_i$ bedeutsam. Es heißt im allgemeinen *Schleifenverstärkung*:

$$V_s = \frac{u_r}{u_i} = \frac{V_0}{v_k} = \boldsymbol{Kv} \,. \tag{33}$$

V_0 ist der Verstärkungsfaktor des Verstärkers in Bild 7.1.7, v_k bedeutet die kritische Verstärkung eines *RC*-Gliedes. $\boldsymbol{v}$ sowie $\boldsymbol{K}$ beziehen sich auf Bild 7.1.7b).

Zur Teit $t = 0$ werde am Eingang die Spannung u_i eingeschaltet. Zur Zeit $1/f$ liegt dann am Ausgang die Spannung $V_s \cdot u_i$. Sie wird zur Eingangsspannung addiert, und so entsteht die Eingangsspannung $(1 + V_s)\, u_i$. Zur Zeit $2/f$ wird daraus durch Verstärkung, erneute Rückkopplung und Addition eine entsprechend höhere Eingangsspannung. Mit jedem Schwingungsvorgang wächst folglich die Ausgangsspannung des Verstärkers exponentiell an. Nach einer Zeit T gilt

$$u_a \approx u_i \cdot e^{V_s/T} \,. \tag{34}$$

Durch eine hinreichend lange Zeit T ist es also möglich, jede beliebig große Verstärkung zu erreichen. In die Größe der Ausgangsspannung gehen Änderungen der Eingangsspannung nur während der ersten 2 bis 3 Perioden ein. Danach ist die Ausgangsspannung lediglich noch von der Schleifenverstärkung bei der Schwingungsfrequenz und der Zeit beeinflußbar. Für eine definierte Verstärkung ist es daher erforderlich, die Rückkopplung nach einer bestimmten Zeit T zu unterbrechen. Geschieht diese Unterbrechung nicht, so wird die Ausgangsspannung bis an die Grenzkennlinien (vgl. Abschnitt 7.4.) anwachsen. Diese Grenzen bestimmen dann eine gegebene Maximalamplitude. Infolge der dort vorhandenen starken Kennlinienkrümmungen wird sich diese Maximalamplitude so einstellen, daß im Mittel Gl. (28) erfüllt ist.

Definierte Verstärkungen für Signaländerungen können für den soeben besprochenen Fall nur erreicht werden, wenn die Rückkopplung für eine günstig gewählte Zeitdauer T periodisch ein- und ausgeschaltet wird. Diese Verstärkung berechnet sich zu

$$\boxed{V_p = e^{V_s f T}} \,. \tag{35}$$

Derartige Schaltungen mit periodischer Einschaltung der Rückkopplung werden als Pendelrückkopplung bezeichnet. Ein Schaltprinzip und die Wirkungsweise seien mit Bild 7.1.10 erklärt. Das Ein- und Aussetzen der Rückkopplung wird durch die rechteckförmige Spannung U_p der Frequenz f_p bewirkt. Diese Frequenz muß klein im Vergleich zur Übertragungsfrequenz f_{HF} und groß gegenüber der zu verstärkenden Schwingungsänderung f_{NF} sein.

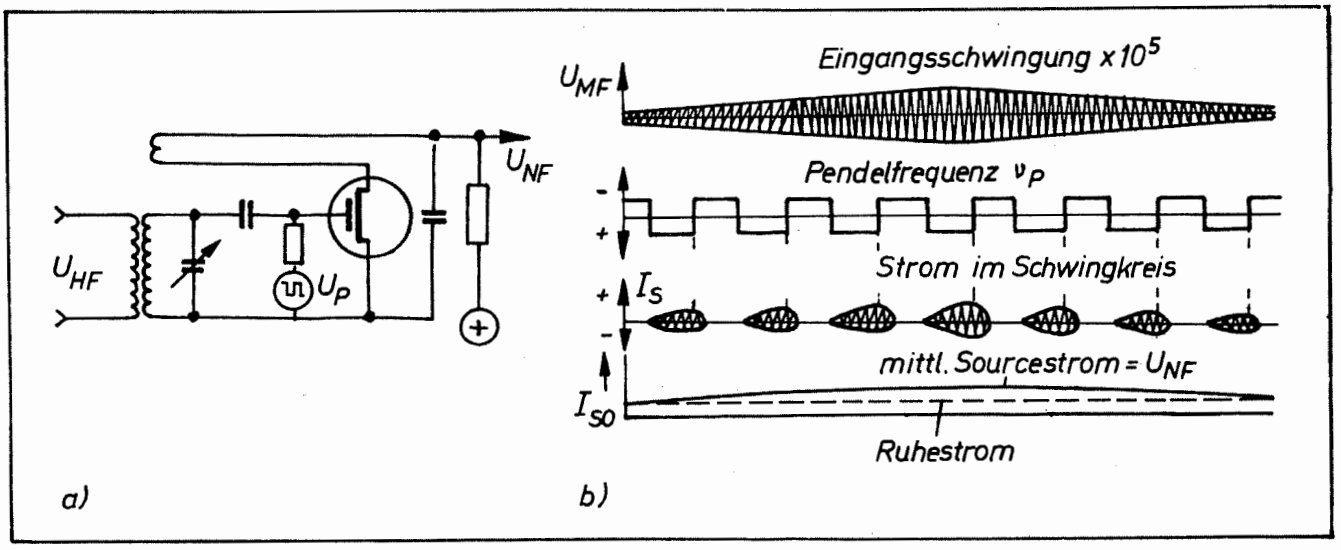


Bild 7.1.10. Schaltung für die Pendelrückkopplung und zugehöriger Schwingungsverlauf

Mit der Pendelrückkopplung kann mit einer einzigen Stufe das geringste Signal, auch Rauschen bis zu Spannungen von einigen Volt verstärkt werden. Unter günstigen Verhältnissen sind Verstärkungsfaktoren um 10^6 erreichbar. Derartige Schaltungen hatten größere Bedeutung zur Anfangszeit des Rundfunks und erneut zu Beginn der UKW-Technik. Zuweilen werden sie auch noch in der Meßtechnik mit Erfolg angewendet.

7.1.4. Elektronisch erzeugte Bauelemente

Bisher besaßen aktive Schaltungen dreimal Eigenschaften, die denen passiver Bauelemente entsprachen. Im Abschnitt 5.5.3. wurde die Wirkung der Gate-Drain-Kapazität um den Faktor $1 + V$ verstärkt. Im Abschnitt 7.1.1.3. gelang es mit der Rückkopplung, einen negativen Widerstand mit

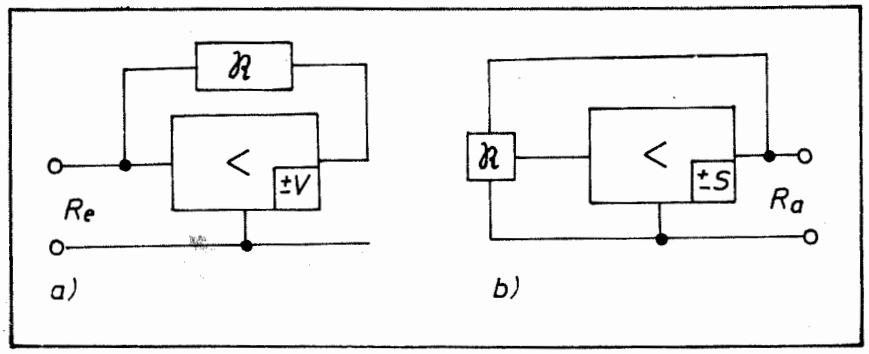


Bild 7.1.11. Anordnungen zur Erzeugung elektronischer Widerstände
a) am Eingang, b) am Ausgang

seinen speziellen Eigenschaften herzustellen. Im Abschnitt 7.1.2.1. wurde gezeigt, wie mittels *RC*-Gliedern ein Resonanzverstärker aufzubauen ist, der Eigenschaften eines Schwingkreisverstärkers besitzt. In der letzten Zeit sind vielfältige derartige Schaltungen entstanden. Deshalb soll hier eine systematische Betrachtung erfolgen. Dabei ist gemäß Bild 7.1.11 zwischen zwei Arten von elektronischen Bauelementen zu unterscheiden: Sie entstehen entweder am Eingang oder am Ausgang des Verstärkers.

7.1.4.1. Elektronische Eingangswiderstände

Ein elektronischer Eingangswiderstand entsteht gemäß dem Ersatzschaltbild 7.1.11a). Zur Vereinfachung sei angenommen, daß der Eingangswiderstand des rückkopplungsfreien Verstärkers unendlich groß und der Ausgangswiderstand Null seien. Für exakte Berechnungen sind beide Werte zusätzlich zu berücksichtigen. Der Verstärker selbst sei phasenrein. Die Phasendrehung zwischen seinem Ein- und Ausgang betrage je nach der Stufenzahl 0 oder 180°. Beide Fälle werden durch das Vorzeichen der Verstärkung erfaßt. Die Wirkung des überbrückenden Widerstandes ist ähnlich wie in Abschnitt 5.5.3. Am Eingang des Verstärkers befinde sich die Spannung u_e. Dann steht am Ausgang des Verstärkers die Spannung $\pm V u_e$. Beide Spannungen treiben durch den Widerstand $\boldsymbol{R}$ den Strom

$$i_R = \frac{(1 \pm V)\, u_e}{\boldsymbol{R}}. \tag{36}$$

Von der Eingangsseite her wird dieser Strom lediglich durch die Eingangsspannung aufgebracht. Damit liegt für sie der *wirksame Eingangswiderstand*

$$\boxed{\boldsymbol{R}_e = \boldsymbol{R}\frac{1}{1 \pm V}} \tag{37}$$

vor. Wird der Verstärker mit einer Induktivität bzw. Kapazität überbrückt, so treten an seinem Eingang folgende Widerstände auf:

$$R_L = \mathrm{j}\omega \frac{L}{1 \pm V} \tag{38a}$$

bzw.

$$R_C = \frac{1}{\mathrm{j}\omega C(1 \pm V)}. \tag{38b}$$

Eine gegenphasige Verstärkung verkleinert also alle Werte um den Faktor $1 + V$. Dies gilt sowohl für reine ohmsche wie auch für komplexe Widerstände. Der Frequenz-

gang der Widerstände bleibt hiervon unbeeinflußt. Folglich werden ohmsche Widerstände und Induktivitäten verkleinert, aber Kondensatoren entsprechend vergrößert. Insbesondere bei der Kapazität spricht man dann von MILLER-*Kapazität*. Ihre Anwendung bei Integratoren führt zum MILLER-*Integrator* (vgl. 7.5.6.).

Für einen Verstärker ohne Phasendrehung muß zwischen Verstärkungen, die kleiner oder größer als 1 sind, unterschieden werden. Bei V kleiner als Eins werden alle realen Widerstände und Induktivitäten vergrößert und alle Kapazitäten verkleinert. Bei V genau gleich Eins sind Eingangs- und Ausgangsspannung genau gleich groß. Dies bedeutet, daß durch den Widerstand kein Strom fließen kann. Er wird also unendlich groß bzw. seine Kapazität gleich Null. Schaltungen, die mit einem V betrieben werden, das nur wenig kleiner als Eins ist, heißen oft *Bootstrap-Schaltungen* (Bootstrap (engl.) = Schleife am hinteren Teil des Schuhs, zur Erleichterung des Anziehens). Mit ihnen wird der Eingangswiderstand gemäß den obigen Betrachtungen möglichst groß gemacht. Dieser Effekt hat für viele Anwendungen große Bedeutung. Er wird u. a. an Beispielen im Abschnitt 7.3.2. behandelt.

Wird V größer als 1, so tritt im Nenner von Gl. (37) und (38) ein Vorzeichenwechsel auf. Dadurch entstehen *negative Widerstände, Induktivitäten und Kapazitäten*. Die negativen Widerstände sind bereits im Abschnitt 7.1.1.3. behandelt worden. Negative Induktivitäten und Kapazitäten sind in ihrer Wirkung ähnlich. Sie erzeugen jedoch keinen neuen Punkt in der Ortskurve für Widerstände (Bild 2.2.1). Sie sind aber imstande, normale Kapazitäten und Induktivitäten zu kompensieren. Beim Schwingkreis in Abschnitt 2.2.2. wurde gezeigt, daß eine Induktivität oder Kapazität sich ganz oder teilweise kompensieren kann. Der Grad ihrer *Kompensation* ist von der Frequenz abhängig. Eine genaue Frequenzkompensation erfolgt bei der Resonanzfrequenz. So entsteht für den Schwingkreis eine Ortskurve, welche bei der Resonanzfrequenz die reelle Achse schneidet. Negative Blindelemente kompensieren im Gegensatz hierzu die positiven vollständig und unabhängig von der Frequenz. Die Zusammenhänge bezüglich des komplexen und reellen Anteils der verschiedenen Blindelemente zeigt Tab. 7.1.1. Die Kompensation einer Induktivität bzw. Kapazität durch ihren negativen Wert wird vielfach auch als Neutralisation bezeichnet. So wird durch eine entsprechende Verstärkerschaltung ähnlich wie in Bild 7.1.11a) z.B. die Gate-Drain-Kapazität eines FET in ihrer Wirkung beseitigt.

Werden *elektronische Bauelemente* mittels eines Verstärkers erzeugt, so sind sie zugleich relativ *leicht veränderbar*. Ihr Wert hängt nämlich von der Verstärkung ab. Mit einer Veränderung der Verstärkung ist eine Änderung ihrer Größe möglich.

7.1.4.2. Elektronische Ausgangswiderstände

Die Schaltung zur Erzeugung elektronischer Ausgangswiderstände zeigt Bild 7.1.11b). Wiederum sei ein unendlich großer Eingangswiderstand vorausgesetzt. Auch der Innenwiderstand sei diesmal unendlich groß. Damit wird der Verstärker am besten durch seine Steilheit S gekennzeichnet. Die Phasenlage sei wiederum im Vorzeichen enthalten. Aufgrund der BARKHAUSEN-Formel muß der Durchgriff des Verstärkers gleich Null sein. Die Steuergleichung nimmt damit die Form

$$i_a = \pm S \cdot u \tag{39}$$

an. Das Rückkopplungsglied sei durch sein Übertragungsmaß $\boldsymbol{K}$ gekennzeichnet. Daher gilt

$$u_i = \boldsymbol{K} u_a \,. \tag{40}$$

Für den Ausgangswiderstand ergibt sich somit

$$\boxed{\boldsymbol{R}_a = \frac{u_a}{i_a} = \frac{1}{\pm \boldsymbol{K} S}} \tag{41}$$

Als Rückkopplungsglied werde ein Spannungsteiler benutzt. (Vgl. Bild 2.3.1 bis 2.3.5.) Für das Übertragungsmaß gilt allgemein

$$\boldsymbol{K} = \frac{1}{a + \mathrm{j}b}. \tag{42}$$

Wegen Gl. (41) folgt deshalb der Widerstand

$$\boldsymbol{R}_a = \frac{a + \mathrm{j}b}{\pm S}. \tag{43}$$

Er stellt also eine Reihenschaltung aus einem ohmschen und einem Blindwiderstand dar. Dabei geht der ohmsche Widerstand auf den reellen Anteil a und der imaginäre Widerstand auf den Anteil b zurück. Die Steilheit $\pm S$ versieht das Übertragungsmaß des Spannungsteilers mit der Dimension eines Widerstandes. Ist b sehr groß gegen a, so entstehen praktisch reine Blindwiderstände. Definitionsgemäß ist dies gerade für *Differenzier- bzw. Integrierglieder* gemäß Bild 2.3.2 erfüllt. Bild 7.1.12 zeigt einige Beispiele von Rückkopplungsvierpolen und die sich damit ergebenden reellen und Blindwiderstände. Mittels des Vorzeichens der Steilheit wird das Vorzeichen dieser Widerstände und mittels des Betrages der Steilheit ihre Größe festgelegt. Die Widerstände sind also über die Steilheit steuerbar. Wird dieses Verhaltens bewußt ausgenutzt, so tragen derartige Schaltungen die Bezeichnung *Reaktanzstufen*.

Der Vorteil der ausgangsseitigen Erzeugung von Blindwiderständen liegt vor allem darin, daß relativ große Ströme, Spannungen bzw. Leistungen verarbeitet werden können. An der Eingangsseite sind dagegen Blindelemente zu erzeugen, deren Werte erheblich größer sind. In diesem Fall muß jedoch die Ansteuerspannung so klein gehalten werden, daß die Ausgangsseite des Verstärkers nicht übersteuert wird.

7.1.4.3. Impedanzkonverter, Gyrator

Insbesondere Bild 7.1.12 und Tab. 7.1.1 zeigen, daß es mit elektronischen Mitteln möglich ist, aus einer Impedanz eine andere zu erzeugen, z. B. aus einer Kapazität eine Induktivität. Dies hat große praktische Bedeutung, weil Induktivitäten meist sehr unpraktische Bauelemente sind. Das betrifft u. a. ihre teure Herstellung, ihr oft großes Volumen und ihre Verluste. So wurde schon in den vierziger Jahren versucht, das Prinzip der Impedanzkonvertierung zu realisieren (H. Feigs: Funk und Ton, 1949, S. 459ff.). Zunächst entstanden dabei aber nur die formalen *Ansätze aus der Vierpoltheorie*. Heute werden Positiv- und Negativ-Impedanzkonverter (PIC bzw. NIC) unterschieden. Bei den NIC gibt es noch die auf Spannung (UNIC) bzw. Strom (INIC) bezogenen. Der PIC wird meist Gyrator genannt. Dieser Begriff leitet sich vom griechischen gyro = Kreise, Drehbewegung ab und deutet so auf die bei ihm wirkende Phasendrehung hin.

In der Vierpolschreibweise gilt für diese drei Typen

$$\text{INIC:} \quad \begin{pmatrix} U_1 \\ I_1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} U_2 \\ I_2 \end{pmatrix} \tag{44}$$

$$\text{UNIC:} \quad \begin{pmatrix} U_1 \\ I_1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} U_2 \\ I_2 \end{pmatrix} \tag{45}$$

$$\text{PIC, Gyrator:} \quad \begin{pmatrix} I_1 \\ I_2 \end{pmatrix} = \begin{pmatrix} 0 & 1/R \\ 1/R & 0 \end{pmatrix} \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \tag{46}$$

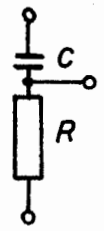
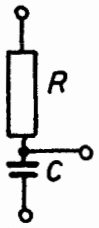
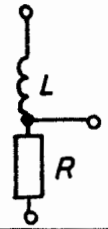
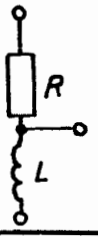
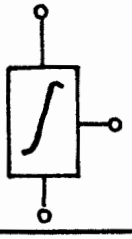
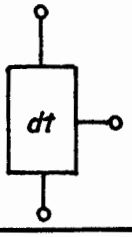
Schaltung						
$\frac{1}{\mathcal{R}}$	$1+\frac{1}{j\omega RC}$	$1+j\omega RC$	$1+j\omega\frac{L}{R}$	$1+\frac{R}{j\omega L}$	$j\omega X$	$\frac{1}{j\omega X}$
Realteil von $\mathcal{R}_a$	$\frac{1}{\pm S}$	$\frac{1}{\pm S}$	$\frac{1}{\pm S}$	$\frac{1}{\pm S}$	0	0
Blindanteil von $\mathcal{R}_a$	$C^{+}_{-}=\pm SRC$	$L^{+}_{-}=\mp\frac{RC}{S}$	$L^{+}_{-}=\mp\frac{L}{SR}$	$C^{+}_{-}=\pm\frac{SL}{R}$	$L^{+}_{-}=\mp\frac{X}{S}$	$C^{+}_{-}=\pm XS$

Bild 7.1.12. Schaltungen von Spannungsteilern für die Anordnung von Bild 7.1.11 b) zur elektronischen Erzeugung von Blindelementen und die dabei entstehenden Daten

Tabelle 7.1.1. Eigenschaften der verschiedenen Blindelemente

Bezeichnung	Widerstand	Phase	Frequenzgang
Induktivität	$j\omega L$	j	ω
Kapazität	$\frac{1}{j\omega C}$	$-j$	$\frac{1}{\omega}$
Negative Induktivität	$-j\omega X$	$-j$	ω
Negative Kapazität	$\frac{1}{-j\omega X}$	j	$\frac{1}{\omega}$

Für den Eingangswiderstand des Gyrators folgt hierzu

$$R_1 = \frac{U_1}{I_1} = \frac{I_2}{U_2}. \tag{47}$$

Liegt an den Ausgangsklemmen des Gyrators eine Kapazität

$$R_2 = 1/j\omega C\,, \tag{48}$$

so erscheint am Eingang eine Induktivität von der Größe

$$R_1 = j\omega C\,.$$

Dies zeigt symbolisch Bild 7.1.13. (49)

Bei den NIC sind die Verhältnisse leichter zu verstehen. Durch sie wird ein ohmscher Widerstand zum negativen Widerstand (vgl. 7.1.1.3.), eine Induktivität zur negativen Induktivität, eine Kapazität zur negativen Kapazität (vgl. Tab. 7.1.1.).

Formal und praktisch läßt sich aus zwei NIC ein Gyrator aufbauen. Hierauf und auf praktikable Schaltungen wird im Abschnitt 7.5.8. eingegangen.

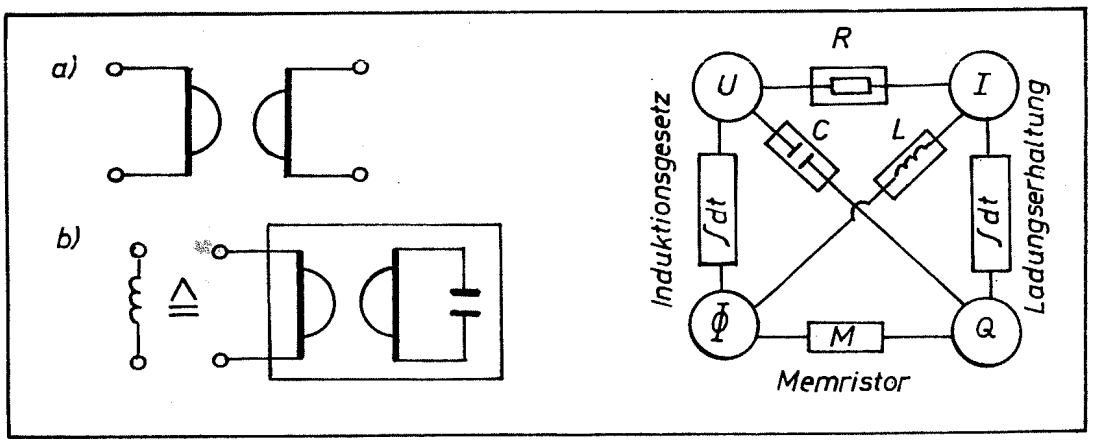


Bild 7.1.13. a) Schaltbild des Gyrators und b) die Erzeugung einer Induktivität mit ihm aus einer Kapazität

Bild 7.1.14. Schematischer Zusammenhang der Bauelemente aus den Größen Strom I, Spannung U, Ladung Q und magnetischer Fluß Φ sowie das fehlende Bauelement Memristor und die Gesetze der Induktion und Ladungserhaltung

7.1.4.4. Vollständigkeit der passiven Bauelemente

Im Abschnitt 1.5.1. und 2.2.1. sind die passiven Bauelemente R, C und L aus den vier Hauptgrößen der Elektrizität Spannung U, Strom I, Ladung Q und magnetischer Fluß Φ abgeleitet. Rein formal müssen zwischen diesen vier Größen sechs Beziehungen bestehen. Den entsprechenden Zusammenhang zeigt Bild 7.1.14. Da zwei Beziehungen durch das Induktionsgesetz

$$U = -n \frac{\mathrm{d}\Phi}{\mathrm{d}t} \tag{50}$$

bzw. Ladungserhaltungsgesetz

$$I = \mathrm{d}Q/\mathrm{d}t \tag{51}$$

realisiert sind, fehlt nur das mit M (von Memristor) bezeichnete Bauelement. Es sind keine auf Material bezogenen Gesetze bekannt, die dieses Bauelement realisieren ließen. Es besteht jedoch wieder die Möglichkeit, es elektronisch zu erzeugen.

Eine andere Betrachtung bezüglich der Vollständigkeit der Bauelemente geht davon aus, daß die Reihe

$$\frac{1}{\mathrm{j}\omega C};\ R;\ \mathrm{j}\omega L$$

formal nach beiden Seiten fortgesetzt wird, also

$$\ldots - \frac{1}{\mathrm{j}\omega^3 X_3};\ - \frac{1}{\omega^2 X_2};\ \frac{1}{\mathrm{j}\omega C};\ R;\ \mathrm{j}\omega L;\ -\omega^2 Y_2;\ -\mathrm{j}\omega^3 Y_3; \ldots$$

Dann wird von Bauelementen höherer Ordnung gesprochen. Auch sie lassen sich elektronisch erzeugen.

7.1.5. Zusammenfassung

Die vorangegangenen Betrachtungen zeigen eine große Variabilität der Rückkopplung. Deshalb ist es günstig, einen Überblick zu den verschiedenen Möglichkeiten zu erhalten. Dies zeigt Bild 7.1.15. Nicht berücksichtigt sind hierin jene Anwendungen, die erst im Abschnitt 7.5.7. bei den Operationsverstärkern besprochen werden, wie z. B.

- idealer Gleichrichter
- logarithmischer Verstärker
- Multiplizierschaltungen
- Aktivfilter (7.5.8.).

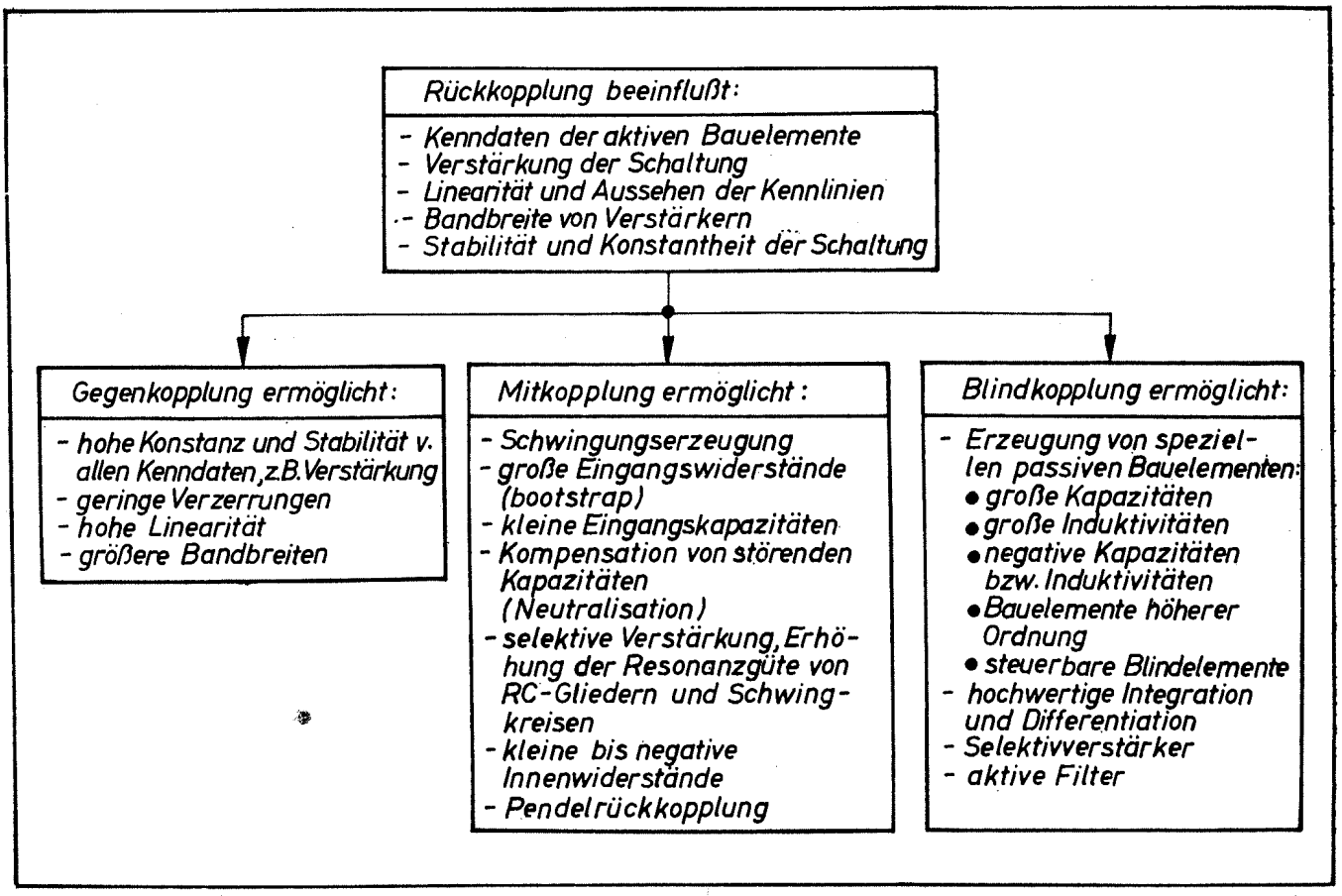


Bild 7.1.15. Überblick zu den Eigenschaften und Möglichkeiten der Gegenkopplung. Für weitere Anwendungen vgl. Abschnitt 7.3.0.

7.2. Modulationen

Ursprünglich wurden die Modulationen für die Nachrichtenübertragung entwickelt, die nicht Inhalt dieses Buches ist. Die Modulationen haben aber im Laufe der Zeit Anwendungen in vielen anderen Gebieten der Signalverarbeitung erlangt. Dies ist der Grund, hier einige Grundlagen zu behandeln. Ausführlicher ist auf sie u. a. in [F 12, F 20, H 26, H 41] eingegangen.

Die Vielzahl der heute verwendeten Modulationen ist schwer zu klassifizieren. Dabei ist auch zu beachten, daß verschiedene Modulationen mit diskreten bzw. digitalen Signalen arbeiten. Obwohl eine Klassifikation dieser Begriffe erst im Abschnitt 8.1. gegeben wird, sei hier bereits mittels Tab. 7.2.1 der Versuch einer groben Einteilung der Modulationen unternommen. Dabei ist zu beachten, daß Modulationen, die diskrete oder digitale Werte betreffen, meist Kodierungen (vgl. 8.1.4.) genannt werden.

Modulation wird hier als *übergeordneter Begriff* verwendet. Generell sind dann *zwei typische Eigenschaften* zu unterscheiden:

— die Eigenschaften des zu übertragenden Signals,
— die Eigenschaften des dazu vorhandenen Kanals.

Damit das Signal übertragen werden kann, muß es so umgewandelt werden, daß es den Eigenschaften des Kanals entspricht. Dieser Vorgang heißt *Modulation*. Nach der Übertragung muß das Signal wieder möglichst unverändert in seinen Originalzustand zurückgebracht werden. Dieser Vorgang heißt *Demodulation*. Den Gesamtzusammenhang stellt Bild 7.2.1 dar. Die Berechnungsunterlagen zu solchen Systemen

Tabelle 7.2.1. Einteilung verschiedener Modulationen

Zeit	Signal			
	kontinuierlich (stetig)	quantisiert (diskret)	digital	
			parallel	sequentiell
Kontinuierlich (stetig)	Analogtechnik AM FM	Quantisierungstechnik ohne Taktung (in Ausnahmefällen auch PCM)	Kombinationsschaltungen	asynchroner Automat
Quantisiert (diskret, getaktet)	Abtasttechnik Zeitmultiplex	getaktete Quantisierungstechnik Δ-Modulation	Folgeschaltungen	getakteter Automat
		PAM PFM PPM ≙ PZM PDM ≙ PLM ≙ PVM	Pulscodemodulation (PCM) Kodierungen	Speichermodulationen wie NRZ, RZ, Biphase

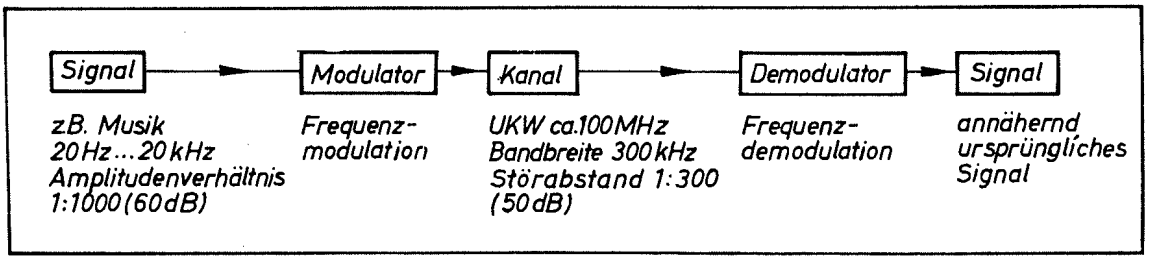


Bild 7.2.1. Zur Erklärung und Notwendigkeit der Modulation
Oben: Ablaufschema, unten: Beispiel

sind ingenieurtechnischer Art. Grundlagen, die im Sinne des Wirkungsgrades solcher Ketten zu betrachten sind, liefert die Informationstheorie.

7.2.1. Stetige Modulationen

Zur Vereinfachung der folgenden Betrachtungen werde ein sinusförmiges Signal

$$u = u_s \cos \omega t \tag{1}$$

angenommen. Darin bedeuten

u_s Amplitude,
ω Signalfrequenz,
t Zeit.

Entsprechend den Betrachtungen des Abschnittes 1.4.6. lassen sich alle möglichen Signale aus einer Summe von Sinus- und Cosinusschwingungen zusammensetzen.

Das vereinfachte Signal von Gl. (1) soll nun mittels einer hochfrequenten Schwingung

$$u = u_t \cos (\Omega t + \varphi) \tag{2}$$

übertragen werden. Darin bedeuten

u_t Trägeramplitude,
Ω Trägerfrequenz
φ Trägerphase für $t = 0$.

Hierzu können drei Wege unterschieden werden:

— Änderung der Trägeramplitude = *Amplitudenmodulation* (AM),
— Änderung der Trägerfrequenz = *Frequenzmodulation* (FM),
— Änderung der Phase = *Phasenmodulation* (PM).

Da Gl. (2) auch in der Exponentialform geschrieben werden kann (vgl. 1.4.3.),

$$u = u_t \, e^{j(\Omega t + \varphi)} , \tag{2a}$$

werden die Frequenz- und Phasenmodulation zuweilen auch unter dem Begriff *Exponentialmodulationen* zusammengefaßt.

Im allgemeinen erfolgen die Änderungen der drei Parameter einzeln, also nicht gemischt und proportional dem Signal. Ausnahmen werden hier nicht behandelt.

7.2.1.1. Amplitudenmodulation

Bei der Amplitudenmodulation (AM) wird die Trägeramplitude durch das Signal proportional beeinflußt,

$$u_t = u_{t_0}(1 + m \cos \omega t) , \tag{3}$$

wobei m der Modulationsgrad ist. Er hängt von u_s ab, muß aber kleiner als 1 bleiben. u_{t_0} ist die Ruheamplitude, und u_t kann damit zwischen 0 und $2u_{t_0}$ liegen. Wird nun (3) in (2) eingesetzt, so folgt

$$u = u_{t_0}(1 + m \cos \omega t) \cdot \cos (\Omega t + \varphi) . \tag{4}$$

Dies ist die relativ unübersichtliche Zeitdarstellung der AM. Die zu ihr gehörende Spektraldarstellung kann mittels des Additionstheorems

$$2 \cos \alpha \cdot \cos \beta = \cos (\alpha + \beta) + \cos (\alpha - \beta) \tag{5}$$

erhalten werden:

$$\frac{u}{u_{t_0}} = \cos (\Omega t + \varphi) + \frac{m}{2} \cos [(\Omega + \omega) t + \varphi] + \frac{m}{2} \cos [(\Omega - \omega) t + \varphi] . \tag{6}$$

Es existieren also der Träger Ω und die beiden Seitenfrequenzen $\Omega \pm \omega$. Wenn ω einen gewissen Frequenzbereich von ω_1 bis ω_2 gemäß Bild 7.2.2 ausfüllt, entstehen die beiden Seitenbänder. Dies hat dazu geführt, daß als Abarten der Amplitudenmodulation spezielle *Einseitenbandmodulationen* (SSM (engl.) = single sideband modulation) entstanden sind. Sie unterscheiden sich dadurch, ob der Träger vorhanden oder unterdrückt ist bzw. teilweise übertragen wird. Eine Einseitenbandmodulation mit teilweise unterdrücktem Träger und teilweise übertragenem zweiten Seitenband wird bei der Fernsehtechnik angewendet. In den Abschnitten 5.4.7. bis 5.4.9. wurde beispielhaft gezeigt, wie mittels Dioden eine Amplitudenmodulation erzeugt und demoduliert werden kann.

7.2.1.2. Frequenz- und Phasenmodulation

Die Frequenz- und Phasenmodulation haben gemeinsam, daß die zu beeinflussende Größe Argument einer trigonometrischen bzw. der Exponentialfunktion ist (Gl. (2) bzw. (2a)). Sie lassen sich gemeinsam behandeln, wenn das Argument als Winkel gemäß

$$\varphi(t) = \Omega t + \varphi \tag{7}$$

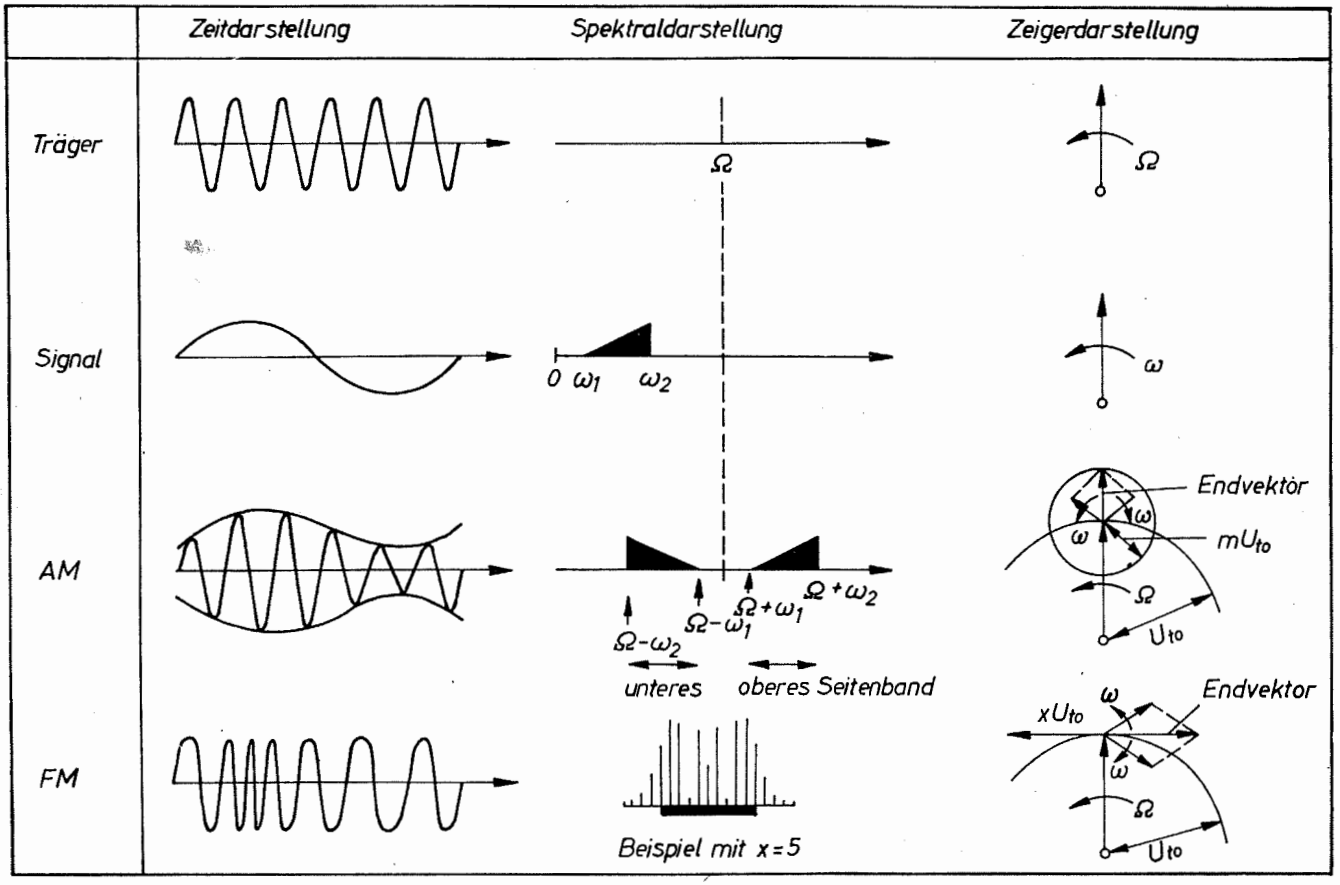


Bild 7.2.2. Amplituden- und Frequenzmodulation

Links: Zeitdarstellung für eine Sinuswelle als Signal
Mitte: Spektraldarstellung für ein Signal mit dreieckförmigem Spektrum zwischen ω_1 und ω_2
rechts: Zeigerdarstellung

aufgefaßt wird. Bei der Frequenzmodulation ist dann die Trägerfrequenz durch eine Momentanfrequenz

$$\Omega_h = \Omega + \Delta\Omega \cos \omega t \tag{8}$$

zu ersetzen und bei der Phasenmodulation die Phase durch

$$\varphi = \varphi_0 + \Delta\varphi \cos \omega t\,. \tag{9}$$

Darin bedeuten $\Delta\Omega$ bzw. $\Delta\varphi$ den Frequenz- bzw. Phasenhub. Infolge der zusammenfassenden Schreibweise von Gl. (7) kann auch Gl. (8) als Phasenmodulation geschrieben werden:

$$\varphi(t) = \int_0^t \Omega_h \, dt\,. \tag{10}$$

Frequenz- (FM) und Phasenmodulation (PM) können also durch *Integration* bzw. *Differentiation* des Signals ineinander übergeführt werden:

$$\text{Integration} + \text{PM} = \text{FM}\,,$$

$$\text{Differentation} + \text{FM} = \text{PM}\,.$$

Deshalb soll im folgenden nur die praktisch wichtigere FM behandelt werden. Sie lautet in Zeitdarstellung

$$\frac{u}{u_t} = \cos[\Omega t + (\Delta\Omega \cdot \cos \omega + t)\varphi]\,. \tag{11a}$$

In der Schreibweise der PM gilt wegen des eben behandelten Zusammenhanges die äquivalente Gleichung

$$\frac{u}{u_t} = \cos(\Omega t + x \sin \omega t + \varphi) . \tag{11b}$$

Der *Modulationsindex* lautet

$$x = \frac{\Delta\Omega}{\omega} . \tag{12}$$

Im Gegensatz zur Amplitudenmodulation ist zumindest theoretisch x nicht nach unten und oben beschränkt. Gewisse praktische Bedeutung hat der *relative Hub*

$$\Delta\Omega_r = \frac{\Delta\Omega}{\Omega} . \tag{13}$$

Ist schon die Zeitfunktion der FM, Gl. (11), wenig anschaulich (Bild 7.2.2), so gilt das noch weitaus mehr für das Spektrum. Da als Argument einer trigonometrischen Funktion wiederum eine trigonometrische Funktion auftritt, folgen für das Spektrum BESSEL-Funktionen. Auf die Ableitung sei hier verzichtet. Einen Überblick zu einigen Spektren gibt Bild 7.2.3. Daraus folgt, daß die *Bandbreite B* der FM beträchtlich ist.

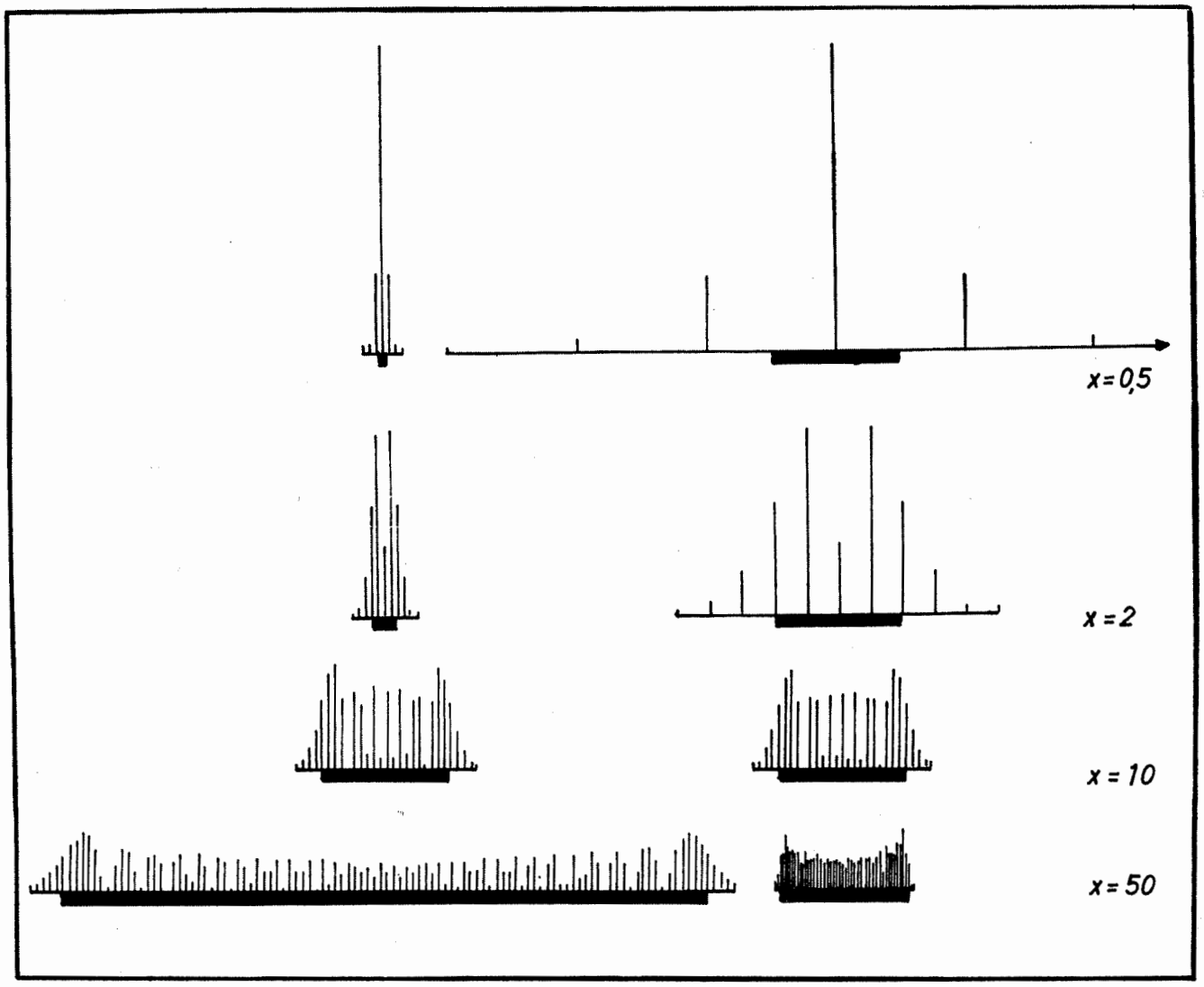


Bild 7.2.3. Frequenzmodulierte Spektren in Abhängigkeit vom Modulationsindex, links bei konstanter Modulationsfrequenz ω_s, rechts bei konstantem Hub $\Delta\omega$. Der Abstand zwischen den Spektrallinien entspricht der Modulationsfrequenz.

Ihre genaue Berechnung ist kompliziert. Infolge einer Näherung für die BESSEL-Funktionen gilt etwa

$$B \approx 2(\Delta\Omega + \omega) . \tag{14}$$

Für die AM gilt dagegen nur

$$B = 2\omega . \tag{15}$$

Die Frequenzmodulation hat dafür *Vorteile* bezüglich der *Störsicherheit* und der *Dynamik*.

7.2.2. Pulsmodulationen

Vor allem im Zusammenhang mit der Fernmeßtechnik entstand das Problem, *Meßwerte vieler Meßstellen* an eine zentrale Stelle möglichst über eine Leitung zu übertragen. Dies ist dann besonders effektiv zu gestalten, wenn die einzelnen Meßstellen zeitlich nacheinander abgetastet werden. Im Ergebnis dieses Prinzips, das *Zeitmultiplex* genannt wird, entsteht eine Impulsfolge (Bild 7.2.4).

Die Rückgewinnung der einzelnen Meßwerte erfolgt mittels des zweiten synchron umlaufenden Schalters. Durch den Kennwertkontakt wird bei jeder Umdrehung einmal die Synchronität kontrolliert. Im Sinne der einleitenden Betrachtungen dieses Hauptabschnittes können die beiden Schalter als Modulator bzw. Demodulator aufgefaßt werden, und da über den Kanal (die Leitung) Impulse übertragen werden, liegt eine Impulsmodulation vor. Sie wird verkürzt als Pulsmodulation bezeichnet, weil dann Ähnlichkeit zum englischen pulse besteht und die Abkürzungen gleichlautend werden. Mit derartigen Impulsen können auch zeitliche Signale wie Sprache, Musik usw. übertragen werden, sie müssen dazu nur in zeitlich definierter Folge ab-

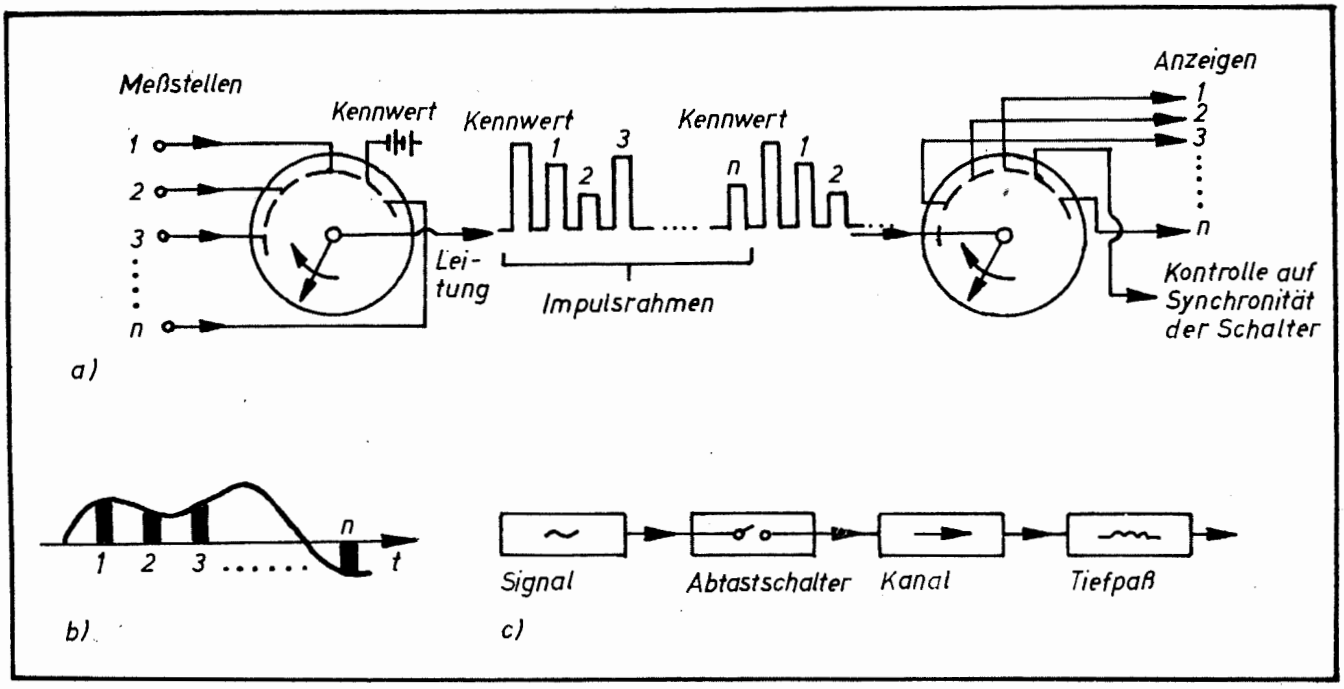


Bild 7.2.4. Prinzip von Zeitmultiplex als Übertragung von Meßwerten vieler Meßstellen über eine Leitung und Grundlage der Impulsmodulationen

a) Gewinnung und Rückgewinnung einer Impulsfolge aus Meßwerten
b) Entstehen der Abtastwerte eines Signals
c) Systemaufbau einer Impulsamplitudenmodulation mit Abtastschalter sowie Zurückgewinnung der ungefähr kontinuierlichen Schwingung durch einen passend gewählten Tiefpaß

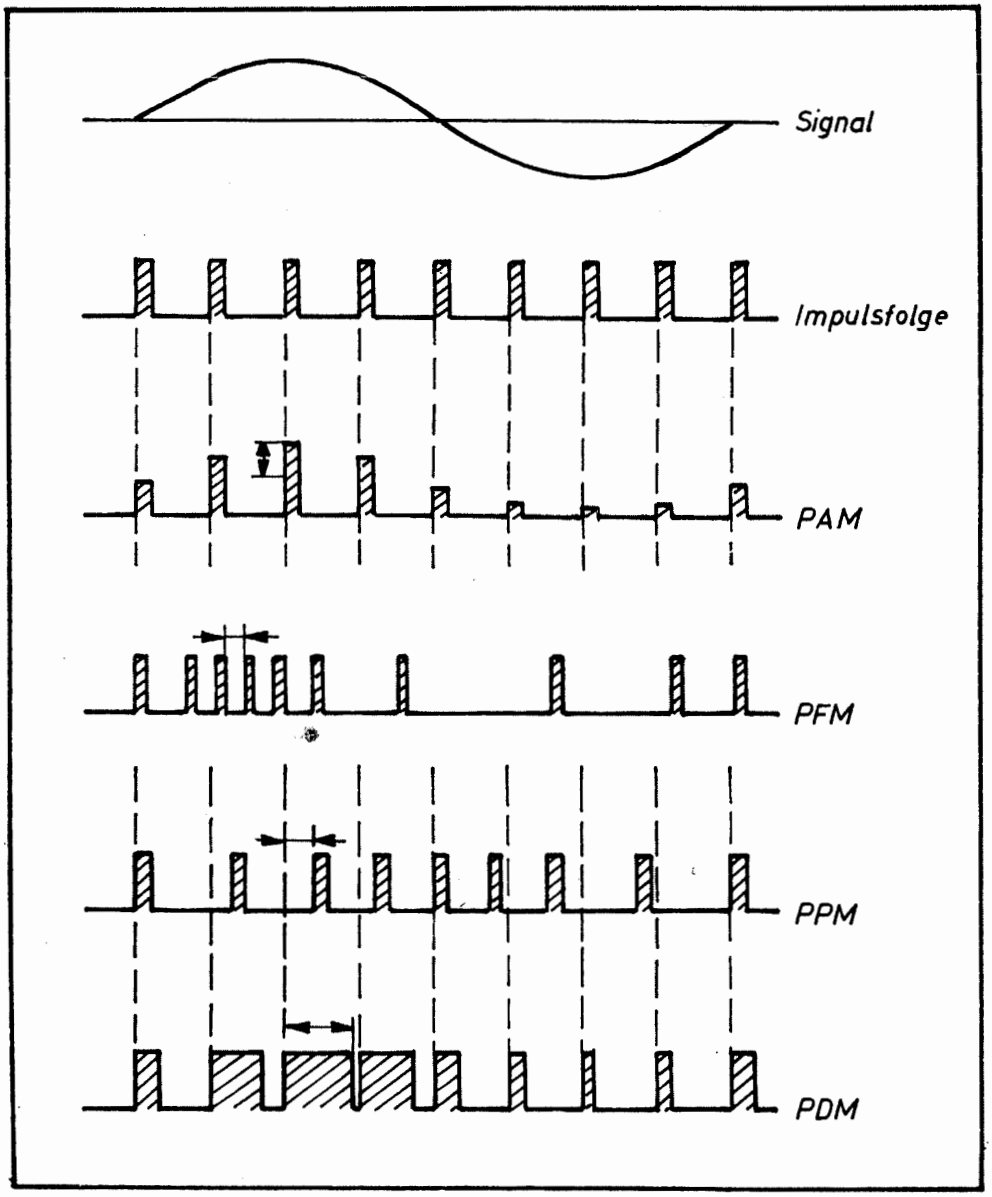


Bild 7.2.5. Die vier wichtigsten Impulsmodulationen. Die Abschnitte der positiven Impulse sind zur besseren Übersicht schraffiert. Dem Signalwert entspricht die durch einen Doppelpfeil gekennzeichnete Impulsgröße. Die Lage des Taktes ist über das gesamte Bild gestrichelt gekennzeichnet

getastet werden. Zunächst bietet dieses Prinzip keine sichtbaren Vorteile, außer dem des Zeitmultiplexes bei ebenfalls mehreren Signalen. Es brauchen nur entsprechend viele Signale zur Übertragung zeitlich ineinander geschachtelt zu werden. So, wie bei den stetigen Modulationen Amplitude, Frequenz oder Phase durch das Signal beeinflußt werden können, bestehen auch bei der Pulsmodulation *mehrere Möglichkeiten*. Sie zeigt Bild 7.2.5. Es gibt also die drei den stetigen Modulationen entsprechenden Formen: *Pulsamplituden-* (PAM), *Pulsfrequenz-* (PFM) und *Pulsphasenmodulation* (PPM), auch *Pulszeitmodulation* (PZM) genannt. Lediglich die PFM ist nicht zum Zeitmultiplex geeignet. Dafür gibt es aber noch eine vierte Form, die *Pulsdauermodulation* PDM. Sie wird im Deutschen zuweilen auch als Pulslängen- (PLM) oder Pulsverhältnismodulation (PVM) (Verhältnis der positiven und negativen Pulszeit) bezeichnet.

Eine Besonderheit der Pulsmodulationen liegt darin, daß sie den Signalwert nicht mehr kontinuierlich übertragen, sondern nur noch zu ausgewählten Zeiten, nämlich den *Abtastzeiten*. Sie bestimmt der *Abtasttakt*. Die zugehörige Taktperiode T_p muß in

einem bestimmten Verhältnis zur höchsten Frequenz des Signals f_0 stehen, wenn das Signal einigermaßen fehlerfrei rekonstruiert werden soll:

$$T_P \leqq \frac{1}{2f_0}. \tag{16}$$

Diese Aussage entspricht dem *Samplingtheorem* der Informationstheorie.

Eine weitere Besonderheit der Pulsmodulationen, insbesondere der PAM und PDM, jedoch nicht der PPM, besteht in ihrer *sehr einfachen Demodulation*. Es genügt nämlich ein Tiefpaß, der f_0 durchläßt, aber $1/T_P$ und alle Oberwellen sperrt. Ein solcher Tiefpaß entspricht einer Mittelwertsbildung, und die Rückgewinnung des Signals ist so leicht aus den schraffierten Flächen von Bild 7.2.5 zu erkennen. Gerade diese sehr einfache Demodulation bringt für viele Anwendungen Vorteile.

7.2.3. Quantisierte Signale

Bei den besprochenen Impulsmodulationen wird das kontinuierliche Signal in ein zeitlich diskretes verwandelt. Daneben besteht auch die Möglichkeit, bezüglich der Amplitudenwerte diskrete Signale zu erzeugen. Dies ist in Bild 7.2.6 gezeigt. Es sollen die acht möglichen Signalwerte (Sollwerte) existieren. *Zu jedem Sollwert gehört ein Toleranzbereich des Signalwertes.* Werden seine Grenzen, die *Schwellen*, überschritten, so erfolgt eine Umschaltung auf den neuen Sollwert. Der Vorteil derartiger quantisierter Signale besteht darin, daß sie völlig störfrei gegenüber Fremdstörungen zu halten sind. Die Fremdstörungen auf das quantisierte Signal müssen nur so klein gehalten werden, daß sie innerhalb des Toleranzbereiches bleiben. Dann kann nämlich das durch sie gestörte Signal wiederum durch einen *Quantisierer* genau auf die Sollwerte gebracht werden. In eine lange Übertragungsstrecke sind also zur *Störbefreiung* im jeweils richtigen Abstand nur Quantisierer einzuschalten. Der Nachteil der quantisierten Signale sind die systembedingten Störungen infolge der Differenz zwischen Signal- und Sollwert. Durch hinreichend viele Stufen (Sollwerte) kann dieses *Quantisierungsgeräusch* sehr klein gewählt werden. Ein weiterer Vorteil ergibt sich dadurch, daß die Sollwerte nicht äquidistant zu sein brauchen, sondern verschiedenen For-

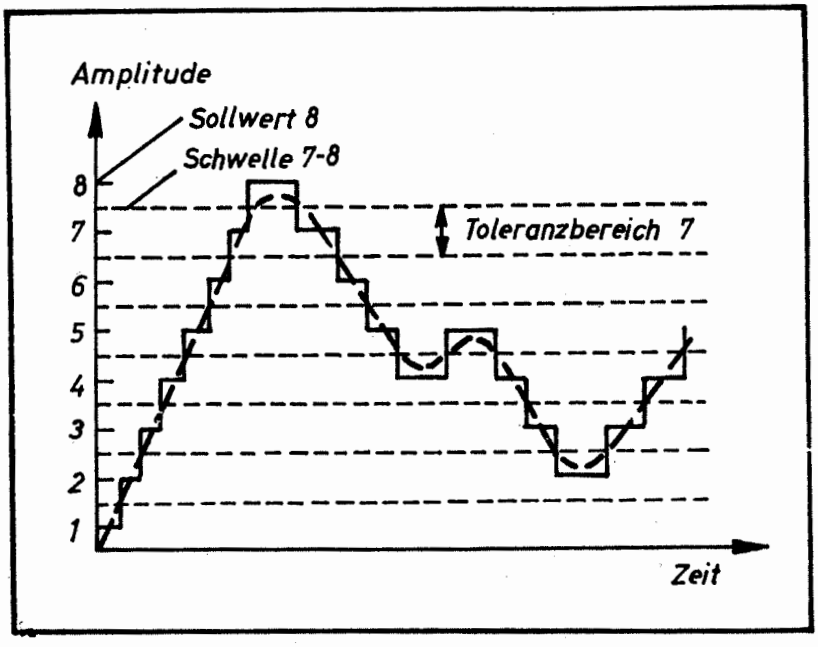


Bild 7.2.6. Erzeugung einer amplitudenquantisierten Schwingung (ausgezogen) aus einem kontinuierlichen Signal (gestrichelt)

derungen angepaßt werden können. Dadurch können zuweilen selbst bei wenig Stufen relativ geringe Auswirkungen des Quantisierungsgeräusches erreicht werden.

Mit quantisierten Signalen sind vielfältige Techniken realisierbar. Zunächst können sie mit allen bisher behandelten Modulationen übertragen werden. Quantisierte Signale können aber auch zu verschiedenen Pulscodemodulationen PCM führen. Solche PCM ergeben sich, wenn z.B. zum Zwecke der Geheimhaltung die Sollwerte irgendwie vertauscht werden. Eine andere PCM liegt vor, wenn die Sollwerte digitalen Signalen (s. 8.1.) zugeordnet werden. Eine weitere Variante ist die Deltamodulation (Δ-Modulation), bei der nur die Zu- bzw. Abnahme in den Stufenschritten übertragen wird. Sie ist vielfach sehr ökonomisch bezüglich der notwendigen Übertragungsbandbreite.

7.3. Einfache Analogschaltungen

7.3.0. Allgemeiner Überblick

Die Grundvoraussetzungen für diesen Abschnitt sind bezüglich der FET in den Abschnitten 3.3. und 5.5. sowie für den Bipolartransistor in 3.4. und 5.6. geschaffen worden. Hier werden zunächst weitere Schaltungsvarianten besprochen.
Sie können in zwei Gruppen geteilt werden:

- Schaltungen aus diskreten Halbleitern
- Grundschaltungen für analoge Schaltkreise.

Damit soll erreicht werden, daß die Grundlagen für die wichtigsten analogen Funktionsgruppen und integrierte Schaltkreise für die folgenden Abschnitte bereitstehen. Die Klassifizierung dieser analogen Funktionsgruppen ist seit jeher ein Problem, und zwar ganz deutlich im Gegensatz zur digitalen Technik (Kap. 8.). Sinnvoll dürfte neben anderen die folgende Einteilung sein:

- *Verstärkerschaltungen*
  - Vorverstärker für niedrige bis mittlere Frequenzen
  - Vorverstärker für hohe Frequenzen
  - Breitbandverstärker (u. a. Kettenverstärker, 5.5.7.)
  - Schmalbandverstärker (u. a. aktive Filter, 7.5.8.)
  - Operationsverstärker (7.5.)
  - Leistungsverstärker (7.4.)
- *Oszillatorschaltungen* (7.3.)
  - für Sinussignale
  - für Rechteck- und Sägezahnsignale
  - Multivibratoren
  - mit Frequenzsteuerung (u. a. U/f-Wandler, VCO)
- *Stromversorgungsschaltungen* (7.6.)
  - Stromquellen
  - Spannungsquellen (u. a. 5.4.5.)
  - elektronische Stabilisationsschaltungen
  - Schaltregler
- *frequenzumsetzende Schaltungen*
  - Frequenzteiler
  - Frequenzvervielfacher
  - Frequenzmischung (u. a. 5.4.9.)
  - Modulations- und Demodulationsschaltungen (u. a. 5.4.7. und 7.2.)

- *Gleichrichterschaltungen*
  - — Signalgleichrichter (u. a. 5.4.4.)
  - — Leistungsgleichrichter (u. a. 6.4.)
  - — phasenempfindliche Gleichrichter (u. a. 5.4.9.)
- *Meß- und Sonderschaltungen* (7.7.)
  - — Schaltungen für mathematische Funktionen, u. a. Addition, Subtraktion, Multiplikation, Division, Logarithmierung, Exponentialkennlinien, Integration, Differentation
  - — Oszillographenschaltungen
  - — Schaltungen zur Amplitudenbegrenzung
  - — Schaltungen zur Phasenselektion, u. a. PLL
  - — Korrelationselektronik
  - — Eimerkettenschaltungen
  - — analoge Magnetbandspeicher.

Diese Aufzählung muß immer unvollständig sein, denn stets werden neue spezielle Schaltungen entsprechend den steigenden Forderungen entworfen werden. Weiter wären zu berücksichtigen

- *Schaltungen zwischen analoger und digitaler Technik*
  - — ADL und DA-Wandler
  - — Impulsschaltungen
  - — Abtast- und Halteschaltungen
  - — Impulsmodulation und -demodulation.

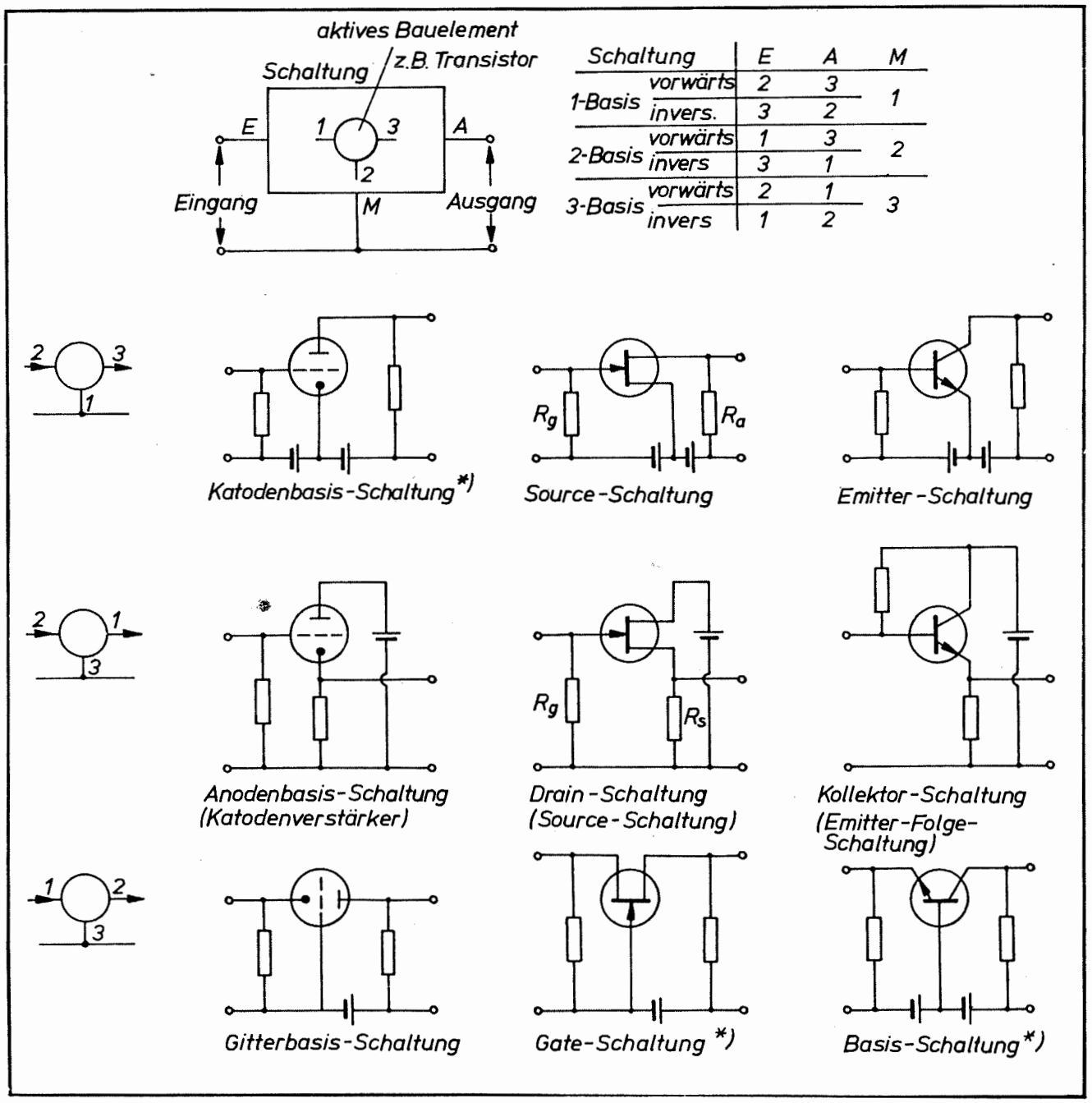
7.3.1. Einstufige Schaltungen

7.3.1.1. Einstufige Verstärker

Im Abschnitt 5.6.1. wurden ohne wesentliche Begründung die drei möglichen Schaltungen eines Bipolartransistors eingeführt. Im Grunde handelt es sich hierbei um ein universelles Prinzip, nämlich die Zuordnung der drei Anschlüsse eines beliebigen Dreipols zu den drei Anschlüssen einer Schaltung. Dieses Prinzip wurde für *RC*-Glieder bereits im Abschnitt 2.3.4. und 2.3.5. behandelt. Hier soll es unter Bezug auf die Rückkopplung (7.1.1.) für die aktiven Bauelemente behandelt werden. Dabei wird ausnahmsweise auch die Röhre, um die Allgemeingültigkeit hervorzuheben, einbezogen (Bild 7.3.1). *Ein Dreipol* kann folglich *in sechs verschiedenen Schaltungen betrieben werden.* Je zwei davon sind zueinander invers. Bei den aktiven Bauelementen ist jene in Verstärkerrichtung bevorzugt und wird, wenn die Betonung notwendig ist, als vorwärts gerichtet bezeichnet. Die entgegengesetzte heißt immer *invers* bzw. invers betrieben. Sie hat in Sonderfällen Bedeutung. Dies ist in Bild 7.3.1 durch ein * gekennzeichnet.

Die Berechnung dieser Schaltungen soll hier nur am *Beispiel des FET* betrachtet werden. Sie gilt ähnlich für Bipolartransistoren, ist bei ihnen infolge der starken Rückwirkungen jedoch etwas unübersichtlicher. Daten für sie enthält ohnehin schon Tab. 5.6.1 und 5.6.4. Zur Verallgemeinerung werden sofort die Source- und die Drainschaltung gemäß Bild 7.3.2a) gemeinsam behandelt. Sie besitzt zwei Ausgänge, die durch Aufteilung des Arbeitswiderstandes auf den Sourcewiderstand R_S und den Drainwiderstand R_D entstehen. Zunächst sei die Ausgangsspannung u_1 betrachtet. Für sie bewirkt der *Sourcewiderstand eine Stromgegenkopplung* mit

$$l = -R_S; \qquad k = 0\,. \tag{1}$$



Schaltung		E	A	M
1-Basis	vorwärts	2	3	1
	invers.	3	2	
2-Basis	vorwärts	1	3	2
	invers	3	1	
3-Basis	vorwärts	2	1	3
	invers	1	2	

Bild 7.3.1. Mögliche Dreipolschaltungen von aktiven Bauelementen. Zur Demonstration der Universalität dieser Betrachtung wurde hier ausnahmsweise die Röhre einbezogen Oben im Bild wird schematisch gezeigt, welche Kombinationen für einen Dreipol in der Zuordnung zu den drei äußeren Anschlüssen bestehen. Von den sechs möglichen Schaltungen werden in der Regel nur die drei vorwärts betriebenen verwendet. Für Sonderanwendungen kommen auch die invers betriebenen zum Einsatz. Dies ist bei den neun Schaltungen mit einem * bezeichnet

Gemäß den Gl. (7.1.5) bis (7.1.9) gilt daher für die mit dem Index D bezeichneten neuen Werte

$$S_D = \frac{S}{1 + SR_S}, \tag{2a}$$

$$D_D = D, \tag{2b}$$

$$R_{i,D} = R_i(1 + SR_S), \tag{2c}$$

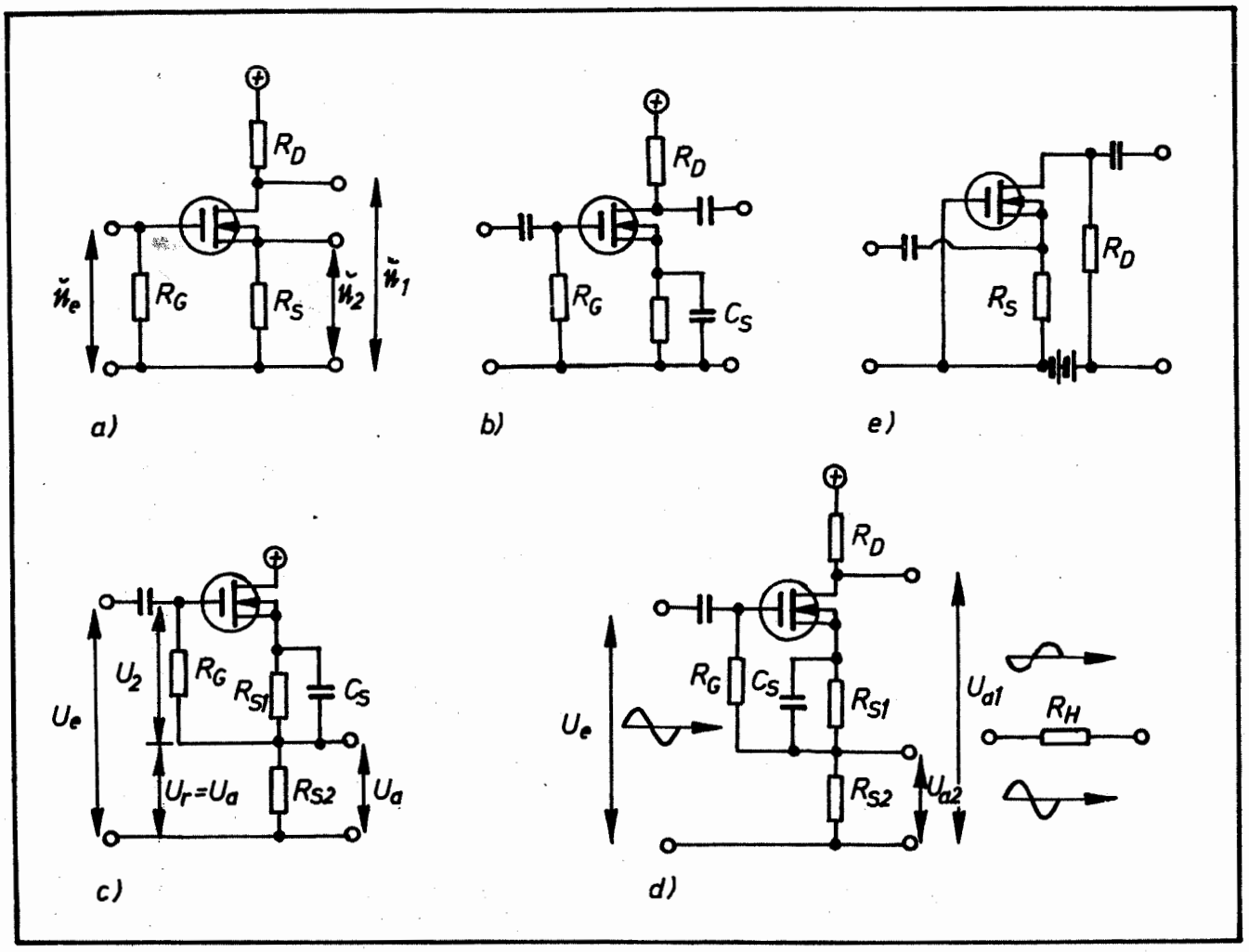


Bild 7.3.2. Schaltungsvarianten eines einstufigen Verstärkers

a) Grundschaltung zur Berechnung
b) Sourceschaltung mit kapazitiv überbrücktem Sourcewiderstand
c) Sourcefolger mit kleiner Ausgangs- und großer Eingangsimpedanz
d) Schaltung mit zwei gegenphasigen Ausgangsspannungen
e) Gateschaltung

und gemäß Gl. (7.1.15) bzw. (5.5.5a) und (5.5.5b) folgt für die neue Verstärkung

$$V_D = \frac{1}{D} \frac{R_D}{R_a + R_i(1 + SR_S)} \tag{3a}$$

$$= \frac{R_D}{R_S + D(R_D + R_i)} . \tag{3b}$$

Insbesondere für kleine D und große R_S gilt die Näherung

$$V_D \approx R_D/R_S . \tag{3c}$$

Für den *zweiten Ausgang* mit u_2 liegt dagegen eine *Spannungsgegenkopplung* vor. Es wird die volle, am Anfang vorhandene Spannung zurückgeführt:

$$l = 0; \qquad k = -1 . \tag{4}$$

Für diesen Ausgang gelten folglich die mit dem Index S gekennzeichneten Kennwerte

$$S_S = S , \tag{5a}$$

$$D_S = 1 + D , \tag{5b}$$

$$R_{i,S} = \frac{1}{S} \frac{1}{1 + D} . \tag{5c}$$

Für die Verstärkung gilt folglich

$$V_S = \frac{1}{1 + \frac{1}{SR_S} + D}. \tag{6}$$

Sie ist also stets etwas kleiner als Eins, kommt aber für $SR_S \gg 1$ und $D \ll 1$ ganz in die Nähe von Eins.

Auffällig ist, daß in die Kenngrößen für den zweiten Ausgang nicht direkt der Sourcewiderstand eingeht. In der Praxis geschieht dies immer über den Arbeitspunkt und damit über die statischen Kennwerte ohne Gegenkopplung. Bei gleicher Betriebsspannung und fester Gate-Source-Spannung (nicht in der Schaltung von Bild 7.3.2a) gegeben) bestimmt nämlich die Summe $R_S + R_D$ den Arbeitspunkt.

Weiter ist auffällig, daß der *Innenwiderstand der zweiten Spannungsquelle sehr klein* ist, etwa $1/S$, also z.B. bei 5 mA/V ca. 200 Ω. Dies bedeutet, er ist meist viel kleiner als der Sourcewiderstand, der hierbei durchaus einige kΩ betragen kann. Dies bedeutet weiter, daß für eine wechselstrommäßige Überbrückung des Sourcewiderstandes — um die Gegenkopplung für den Ausgang 1 zu vermeiden — nicht R_S als Bezugswert, sondern etwa $1/S$ zu wählen ist. Für die entsprechende Grenzfrequenz gilt folglich (Bild 7.3.2b)

$$f_g \approx \frac{1}{2\pi} \frac{S}{C_S}. \tag{7}$$

Hieraus kann bei vorgegebener Grenzfrequenz auch näherungsweise das notwendige C_S bestimmt werden. Damit sind Eigenschaften für die einfache Sourceschaltung mit durch C überbrücktem Sourcewiderstand gegeben. Läßt man nun umgekehrt R_D gegen Null gehen und teilt R_S in zwei Teile, den für die Gatevorspannung R_{S1} und den für den Ausgang entscheidenden Teil R_{S2} auf, so folgt der bekannte *Sourcefolger* gemäß Bild 7.3.2c). Er besitzt den niedrigen Ausgangswiderstand von etwa $1/S$, und sein Eingangswiderstand R_E wird nach dem *Bootstrap-Prinzip* aus R_G vergrößert. Es gilt mit Gl. (6) etwa

$$R_E \approx R_G \frac{1}{1 - V_S} = R_G \left(1 + \frac{1}{\frac{1}{SR_S} + D} \right) \tag{7a}$$

$$\approx R_G/D\,. \tag{7b}$$

Die Schaltung hat also eine Verstärkung von maximal Eins. Der Vorteil liegt in der *Impedanzwandlung*. Sie wird oft als Meßkopf verwendet. Dann belastet sie die Meßstelle sehr gering und gibt an ihrem Ausgang dieselbe Spannung mit der kleinen Impedanz so weiter, daß sie über längere Kabel geführt werden kann.

Eine Anwendung der allgemeinen Schaltung von Bild 7.3.2a) zeigt Teilbild d). Hier werden wirklich beide Ausgänge benutzt; wenn

$$R_{S2} = R_D \tag{8}$$

gilt, treten an ihnen gleich große, aber entgegengesetzte Spannungen auf. Diese Schaltung hat u. a. für spezielle *Phasenschieberschaltungen* (z.B. Bild 2.3.11 oben) und für die Ansteuerung von Gegentaktstufen (7.4.3.) Bedeutung. Mit einem bestimmten, evtl. veränderlichen Verhältnis von $R_D : R_{S2}$ können auch alle *Brückenschaltungen*, z.B. für die *RC*-Glieder (2.3.5.), durch sie günstig realisiert werden. Gegebenenfalls ist darauf zu achten, daß der Ausgang 2 niederohmiger ist. Dies kann durch einen Reihenwiderstand R_H weitgehend kompensiert werden.

Schließlich bleibt noch die *Gateschaltung* gemäß Bild 7.3.2e) zu besprechen. Sie folgt zwangsläufig ebenfalls aus den Betrachtungen zur Schaltung von Bild 7.3.2a). Der Eingang ist bei ihr durch den kleinen Innenwiderstand gemäß Gl. (5c) gegeben, während für den Ausgang die Beziehungen (2) gelten. Lediglich für die Verstärkung addiert sich aufgrund der Phasenbeziehungen noch u_e zur Spannung u_a. Deshalb gilt

$$V_G = 1 + V_D \,. \tag{9}$$

Da aber V_D in Gl. (3) meist groß gegen Eins ist, hat die Gateschaltung in etwa die Verstärkung V_D. Insgesamt ist sie damit bezüglich der Verstärkereigenschaften relativ ungünstig. Sie hat jedoch für *hohe Frequenzen* den großen Vorteil, daß Eingangs- und Ausgangskreis durch das auf Masse liegende Gate *gut entkoppelt* sind. Die sonst oft störenden Rückwirkungen fehlen bei dieser Schaltung fast vollständig.

7.3.1.2. Konstant-Stromquellen

Ein *aktives Bauelement*, z.B. Transistor, kann *nicht nur als Verstärker* verwendet werden. Von den vielen anderen Möglichkeiten seien im folgenden vor allem die Strom- und Spannungsquelle, die elektronische Siebung und einige Anwendungen bei Nichtlinearität beschrieben. Prinzipien des steuerbaren Widerstandes sind bereits im Zusammenhang mit Bild 5.1.2 behandelt. Das *Grundprinzip* von Spannungs- bzw. Stromquellen ist bereits im Zusammenhang mit Bild 5.4.8 eingeführt worden. Es kommt also darauf an, daß die entsprechende Schaltung im Arbeitspunkt ein möglichst großes Verhältnis bezüglich des Gleich- und Wechselstromwiderstandes besitzt. Dies ist mehr oder weniger gut bei den Feldeffekt- und Bipolartransistoren im Sättigungsbereich erfüllt. Allgemein gilt mit Bild 7.3.3a)

$$I = \frac{U}{R_V + R_L} = \frac{U}{R_V} \cdot \frac{1}{1 + \frac{1}{n}}, \tag{10}$$

wobei n durch das Widerstandsverhältnis gegeben ist:

$$n = R_V / R_L \,. \tag{11}$$

Werden die (mit kleinen Buchstaben bezeichneten) Änderungen eingeführt, so gilt

$$\frac{i}{I} = \frac{u}{U} + \frac{r_L}{R_L} \cdot \frac{1}{1+n} + \frac{r_V}{R_V} \cdot \frac{n}{1+n}. \tag{12}$$

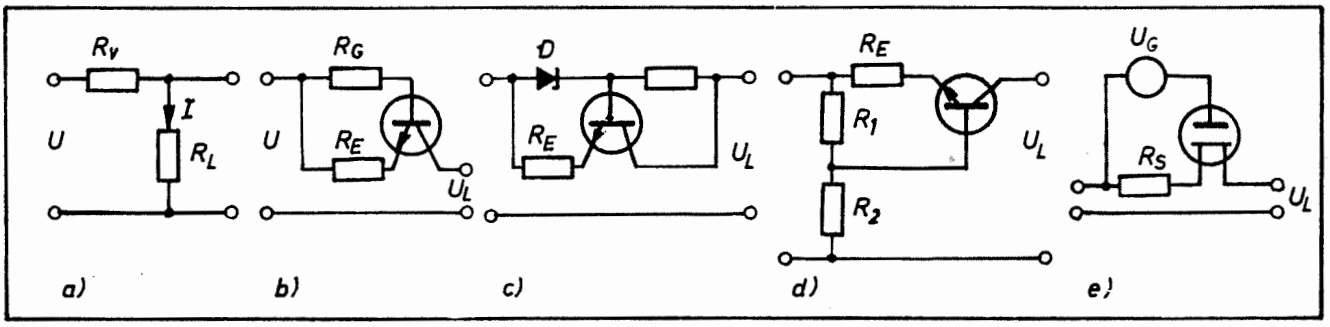


Bild 7.3.3. Schaltungen für einfache Konstant-Stromquellen

a) Grundprinzip für Berechnungen
b) Grundschaltung mit Bipolartransistor
c) Schaltung mit Anwendung einer Z-Diode D zur Einstellung eines günstigen Arbeitspunktes
d) Einstellung des Arbeitspunktes über Spannungsteiler $R_1 : R_2$
e) Grundprinzip mit einem FET. Die Spannungsquelle U_G dient zur Einstellung des Arbeitspunktes

Es gelten also folgende *Stabilitätsbeziehungen*:

- Die relativen Stromschwankungen setzen sich aus drei Termen zusammen.
- Die relativen Spannungsschwankungen werden voll 1:1 übertragen.
- Die relativen Schwankungen der Last werden um den Faktor $1/(1+n)$ gedämpft. Für ein großes Verhältnis n werden sie also sehr klein.
- Die relativen Schwankungen des Vorwiderstandes werden insbesondere für große n nur unwesentlich reduziert.

Hier können noch Temperaturänderungen aller Größen hinzugefügt oder aber auch einbezogen werden.

Im Prinzip kann bei einer Konstant-Stromquelle durch sehr großen Widerstand R_V der Einfluß von Laständerungen fast vollständig unterdrückt werden. Gerade deshalb sind Konstant-Stromquellen relativ einfach zu realisieren. Die entsprechenden Schaltungen zeigt Bild 7.3.3b) bis e). Die Berechnung dieser Schaltungen kann mit den Methoden der Gegenkopplung (7.1.1.) erfolgen. Eine Betrachtung der absoluten Änderungen zwischen Strom und Spannung führt wieder zum Begriff der *Steilheit*

$$S^+ = \frac{\mathrm{d}I}{\mathrm{d}U}. \tag{13}$$

Dementsprechend ist auch der *Innenwiderstand* R_i^+ einführbar, wenn auf die Spannung am Lastwiderstand Bezug genommen wird:

$$R_i^+ = \frac{\mathrm{d}U_L}{\mathrm{d}I}. \tag{14}$$

Für die Steilheit beim Bipolartransistor gemäß Bild 7.3.3b) gilt

$$S^+ = \frac{1}{R_E} \cdot \frac{h_{21} - h_{22}R_E}{\left(h_{22} + \frac{1}{R_E}\right)(h_{11} + R_G) + (1 - h_{12})(1 + h_{21})}. \tag{13a}$$

Mit den meist gültigen Näherungen

$$R_G \ll h_{11}; \qquad h_{12} \ll 1; \qquad h_{21} \gg 1 \tag{15}$$

folgt

$$S^+ \approx \frac{1}{R_E} \cdot \frac{1}{1 + \frac{h_{11}}{h_{21}}\left(h_{22} + \frac{1}{R_E}\right)}. \tag{13b}$$

Für den Innenwiderstand folgt

$$R_i^+ = \frac{(1 + h_{21}) \cdot (1 - h_{12}) + (h_{11} + R_G) \cdot (1/R_E + h_{22})}{h_{22} \cdot \left[1 + \frac{1}{R_E} \cdot \left(h_{11} + R_G - \frac{h_{12}h_{21}}{h_{22}}\right)\right]}, \tag{16}$$

und mit den Näherungen (14) gilt

$$R_i^+ \approx \frac{R_E h_{21}}{h_{22} h_{11}} \frac{1 + \frac{h_{11}}{h_{21}} \cdot \left(\frac{1}{R_E} + h_{22}\right)}{1 + \frac{1}{h_{11}} \cdot \left(R_E - \frac{h_{12}h_{21}}{h_{22}}\right)}. \tag{17}$$

Zwei mögliche Abwandlungen der Schaltung von Bild 7.3.3b) zeigt Bild 7.3.3c) und d). Sie zeichnen sich vor allem durch unterschiedliche Festlegung des Arbeitspunktes

und geändertes R_G aus. Für einen Feldeffekttransistor gilt Bild 7.3.3e). Hier ist es günstig, mit den y-Parametern zu rechnen, wobei bezüglich der Kenndaten des *FET folgende Näherungen* gelten:

$$y_{11} = 1/R_{\text{eing}} \to 0; \qquad y_{12} \to 0; \qquad y_{21} \approx S; \qquad y_{22} \approx 1/R_i\,. \tag{18}$$

Für die Steilheit folgt dann

$$S^+ = \frac{y_{21}\left(y_{12} + \dfrac{1}{R_S}\right) - y_{22}y_{11}}{y_{11} + y_{12} + y_{21} + y_{22} + \dfrac{1}{R_S}}$$

$$\approx \frac{1}{R_S}\,\frac{y_{21}}{y_{21} + y_{22} + \dfrac{1}{R_S}} \tag{19}$$

und für den Innenwiderstand

$$R_i^+ = \frac{1}{y_{22}}\,\frac{\dfrac{1}{R_S} + y_{11} + y_{12} + y_{21} + y_{22}}{\dfrac{1}{R_S} + y_{11} - \dfrac{y_{12}y_{21}}{y_{22}}}$$

$$\approx \frac{1}{y_{22}}\left[1 + R_S(y_{21} + y_{22})\right]$$

$$\approx (1 + SR_S)/y_{22}\,. \tag{20}$$

Die Schaltungen mit dem Bipolartransistor und dem FET weisen also in den Näherungen große Ähnlichkeiten auf. Die Steilheit, d. h. die Auswirkung von Eingangsspannungsänderungen wird vor allem durch $1/R_E$ bzw. $1/R_S$ bestimmt, während der Innenwiderstand in erster Linie durch den Innenwiderstand des Bauelements im Arbeitspunkt gegeben ist.

7.3.1.3. Konstant-Spannungsquellen und Siebung

Bei Konstant-Spannungsquellen interessieren wiederum zwei Kenngrößen, der *Stabilisationsfaktor*

$$A = 1/D^+ = \frac{\mathrm{d}U}{\mathrm{d}U_L} \tag{21}$$

und der *Innenwiderstand* gemäß Gl. (14). Der Stabilisationsfaktor drückt aus, wie stark sich Änderungen der Eingangsspannung U auf die stabilisierte Ausgangsspannung U_L auswirken. Der Innenwiderstand weist wiederum aus, wie stark Änderungen der Last die Ausgangsspannung ändern. Die beiden Grundschaltungen zeigt Bild 7.3.4a) bis c). In erster Näherung gilt bez. der Kenndaten

$$A = 1 + 1/D\,, \tag{22}$$

$$\boxed{R_i^+ = \frac{1}{S(1 + D)}}\,. \tag{23}$$

Da die Quelle den Innenwiderstand R_{iQ} besitzt (Bild 7.3.4d)), wird die an ihm durch Laständerungen bewirkte Spannungsänderung mit A geschwächt, folglich gilt für den

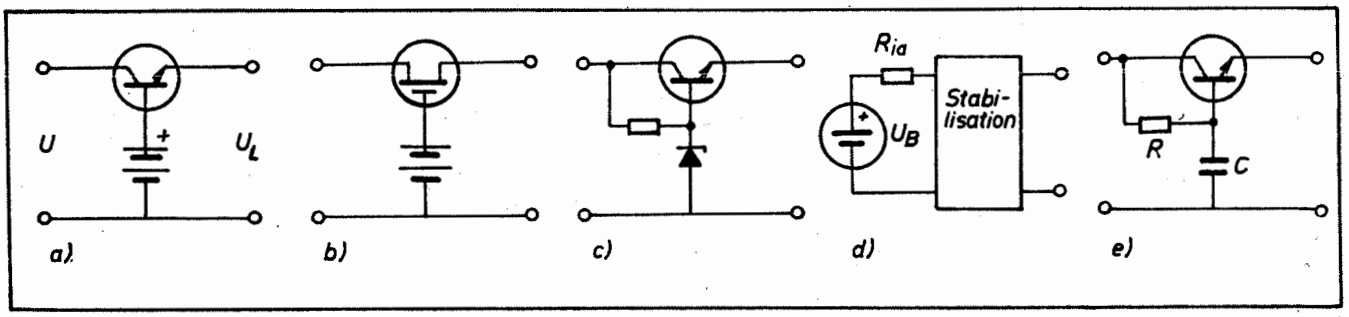


Bild 7.3.4. Einfache Schaltungen für Konstant-Spannungsquellen
a) Grundschaltung für Bipolartransistor und
b) für FET
c) Wie in a), jedoch mit Z-Diode
d) Zur Berücksichtigung des Innenwiderstandes der Quelle R_{iQ}
e) Einfache Schaltung, die ähnlich einer Drossel wirkt

wirksamen Innenwiderstand der Stabilisatorschaltung

$$R_i^{++} = R_i^{+} + \frac{R_{iQ}}{A} = \frac{1}{S(1+D)}\left(1 + \frac{R_{iQ}}{R_i}\right). \tag{24}$$

Eine weiter vereinfachte Schaltung gemäß Bild 7.3.4e) kann zur *Siebung ohne Stabilisierung* herangezogen werden. Durch den Emitterwiderstand R ist für einen günstigen Arbeitspunkt im Sättigungsbereich des Transistors bei eventuell möglichst kleinem Spannungsabfall zu sorgen. Der Kondensator reduziert die am Eingang vorhandenen Wechselspannungen an der Basis. Die Schaltung wirkt daher wie eine Drossel mit Wechselstromwiderstand R_i, und zwar für alle Frequenzen, die höher als die Grenzfrequenz des RC-Gliedes sind.

7.3.1.4. Einstellbare Z-Diode und Umkehrung von NTC-Widerständen

Wird eine Si-Diode in Durchlaßrichtung betrieben, so liegt die Begrenzerspannung bei ca. 0,6 bis 0,8 V. Mit Z-Dioden sind Werte von ca. 2 bis 3 V aufwärts zu erreichen. Eine einstellbare Spannung im Zwischenbereich ist mit einer Schaltung gemäß Bild 7.3.5a) erreichbar. Die zugehörigen Kennlinien zeigt ebenfalls das Bild. Für die Spannung U_D gilt

$$U_D = U_{BE}\left(1 + \frac{R_1}{R_2}\right) + \frac{R_2}{\beta + 1} I_{SS} \cdot \exp\left(\frac{e_0 U_{BE}}{kT}\right), \tag{25a}$$

während für den Strom gilt

$$I_D = \frac{U_{BE}}{R_2} + I_{SS} \cdot \exp\left(\frac{e_0 U_{BE}}{kT}\right). \tag{25b}$$

Darin bedeuten U_{BE} die Basis-Emitter-Spannung, e_0 die Elementarladung, T die absolute Temperatur, k die BOLTZMANN-Konstante, I_{SS} den Sättigungsstrom in Sperrichtung und β die Stromverstärkung in Emitterschaltung. Die beiden obigen Gleichungen besitzen keine analytische Lösung. Sie muß mit numerischen oder grafischen Methoden gewonnen werden. Für den Temperaturkoeffizienten gilt noch

$$\frac{dU_D}{dT} \approx \frac{U_{BE}}{T}\left(1 - \frac{R_1}{R_2}\right). \tag{26}$$

Mittels einfacher Methoden lassen sich Kennlinien durch einfache Schaltungen invertieren. Dies demonstriert Bild 7.3.5b) und c) am Beispiel eines NTC-Widerstandes. So entsteht ein PTC mit relativ linearem Temperaturgang (im Gegensatz zu den PTC

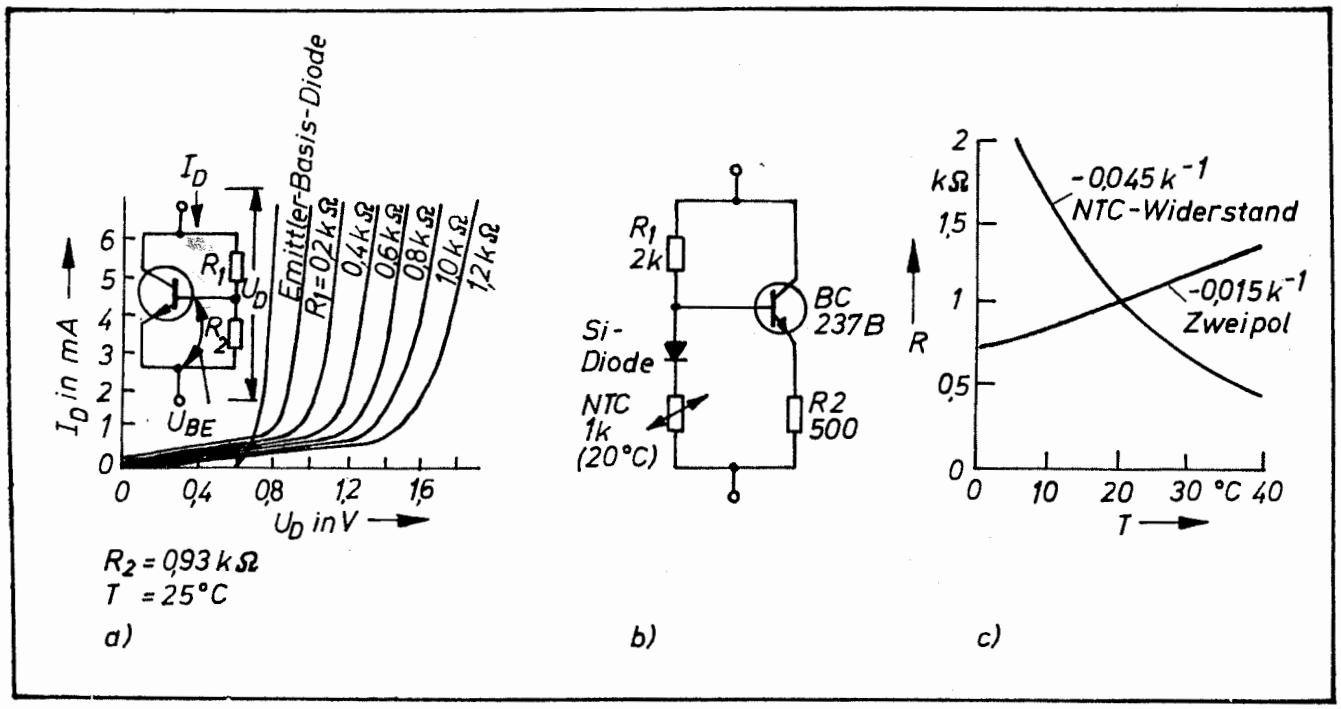


Bild 7.3.5. a) Schaltung einer einstellbaren Diodenkennlinie. Für die abfallende Spannung gilt näherungsweise $U_D \approx U_{BE} \cdot R_1/R_2$. Deshalb heißt diese Schaltung auch Spannungsmultiplikation b) Umkehrung eines NTC. An den Klemmen der Schaltung erscheint ein PTC mit der Kennlinie c). Der um $^1/_3$ verringerte Temperaturkoeffizient folgt aus den Werten von R_1 und R_2

im Abschnitt 5.1.3.). Die Schaltung hat einige Besonderheiten. Am oberen Anschluß muß mindestens eine positive Spannung von etwa 1,5 V liegen. Freiheitsgrade in der Dimensionierung sind jedoch durch die Widerstände $R1$ und $R2$ gegeben:

- Festlegung des Widerstandes des Zweipols bei der Arbeitstemperatur,
- Verkleinerung des Temperaturkoeffizienten um den Faktor $q < 1$.

7.3.2. Mehrstufige Gleichstromverstärker

Im allgemeinen bereitet es Schwierigkeiten, wenn mehrere Verstärkerstufen gleichstrommäßig gekoppelt werden sollen. Dabei müssen vor allem zwei Punkte beachtet werden:

- Die Stufen müssen im Potential angepaßt werden.
- Es muß eine Gleichstromdrift z.B. durch Temperatur- oder Betriebsspannungsänderungen vermieden werden.

Hier sei vor allem der erste Punkt beachtet. Bei FET besteht dann die Möglichkeit, die Gatepotentiale so zu legen, daß eine direkte Kopplung möglich ist. Diesen Fall zeigt Bild 7.3.6a). Bei Bipolartransistoren ergeben sich eventuell vorteilhafte Möglichkeiten mit der Kombination von *npn*- und *pnp*-Transistor gemäß Bild 7.3.6b).

Für kleine Aussteuerungen ist es auch möglich, mehrere Transistoren direkt in Reihe zu schalten. Dies hat insbesondere für integrierte Schaltungen Bedeutung. Einer der ersten integrierten Verstärker benutzt dieses Prinzip. Die Schaltung und das Layout zeigt Bild 7.3.6c) und d).

7.3.2.1. Darlington-Transistoren

Der Darlington-Transistor ist nach seinem Erfinder benannt. Hierbei werden meist zwei, selten drei Transistoren so zusammengeschaltet, daß eine große Stromverstär kung entsteht und die Leistung der Einzeltransistoren gestuft ist. Darlington-

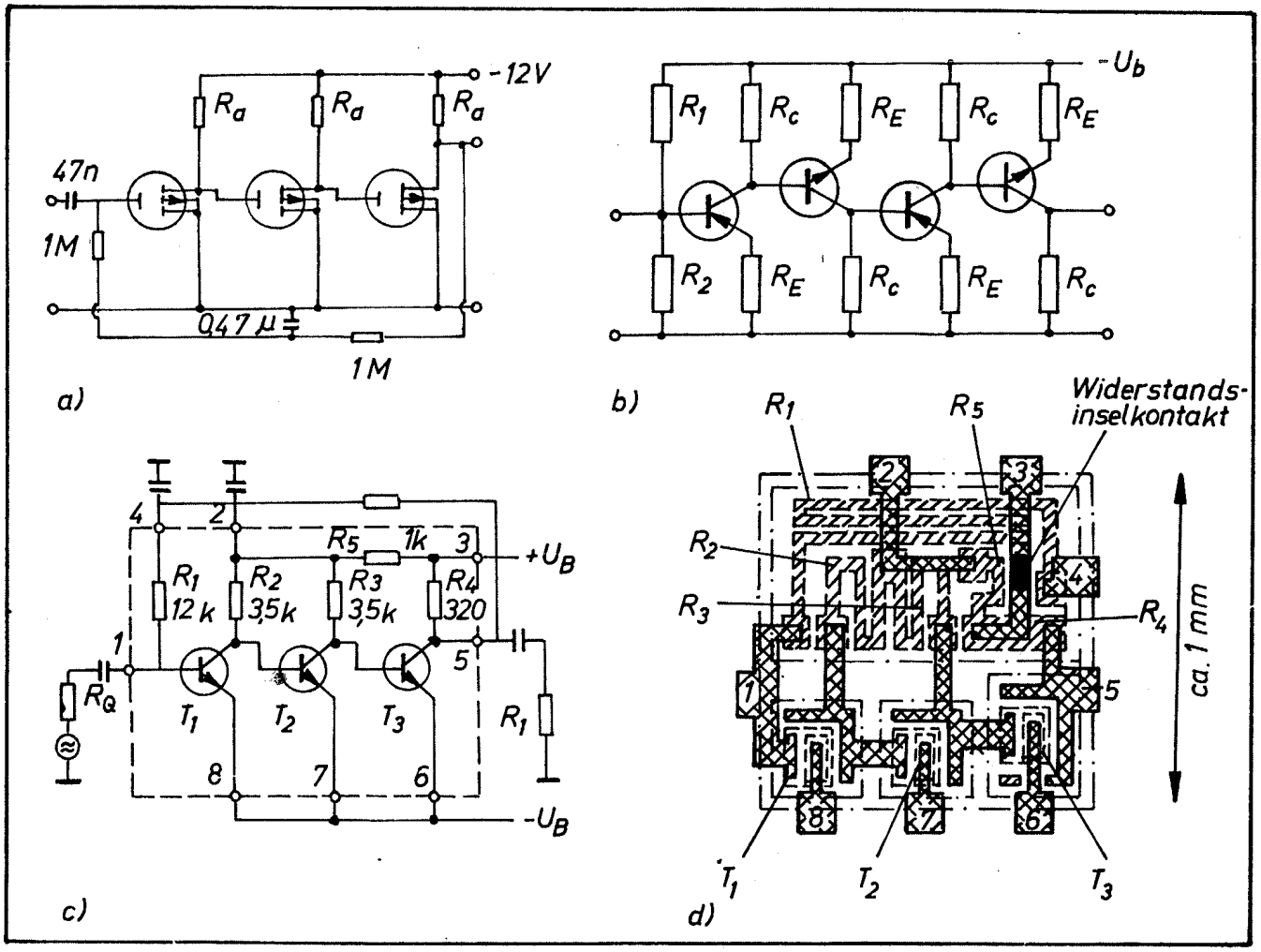


Bild 7.3.6. Beispiele von gleichstromgekoppelten Verstärkerstufen

a) Mit FET der Hochvolt-MOS-Serie. Der dreistufige Verstärker ermöglicht je nach den R_a von 5 bis 60 kΩ Verstärkungsfaktoren von 200 bis 2000. Die beiden 1 MΩ-Widerstände stabilisieren die Arbeitspunkte, und der Kondensator 0,47 µF hebt die Gegenkopplung für Frequenzen ab etwa 100 Hz auf

b) Mit *pnp*- und *npn*-Transistoren. Die Widerstände R_E dienen der Gegenkopplung und Stabilisierung der Arbeitspunkte, die Widerstände R_C der Verstärkung

c) Schaltung eines der ersten integrierten, analogen Schaltkreise mit äußerer Beschaltung. Diese Schaltung gibt es in mehreren Varianten, die sich durch Zusammenlegen der Emitter bzw. Auftrennen der einzelnen Stufen unterscheiden. Dadurch sind vielfältige Anwendungen möglich, z.B. [S29, S60]

d) Layout zur Schaltung c). Es ist deutlich der große Flächenverbrauch für die Widerstände zu erkennen

Transistoren können aus diskreten Bauelementen aufgebaut werden. In der Regel werden jedoch integrierte Bauelemente darunter verstanden. Die eigentliche *integrierte Schaltung* wird allgemein *zu den diskreten Bauelementen gezählt.* Hierfür sprechen folgende Fakten:

- Es sind nur die drei üblichen Anschlüsse eines Transistors vorhanden.
- Bei der Fertigung sind meist keine zusätzlichen Technologieschritte, wie bei den üblichen Schaltkreisen, notwendig.

Den Grundaufbau von DARLINGTON-Transistoren zeigt Bild 7.3.7a) für zwei *npn*-Transistoren und b) für je einen *pnp*- und *npn*-Transistor. Im Prinzip sind auch die beiden anderen Kombinationen (zwei *npn*- bzw. je ein *npn*- und *pnp*-Transistor) möglich. Bei zwei gleichen Transistoren ist der zusammengefaßte Transistor vom gleichen Typ, bei zwei unterschiedlichen bestimmt der erste Transistor den Typ des zusammengefaßten. Für die *Daten des zusammengefaßten Transistors* gelten die

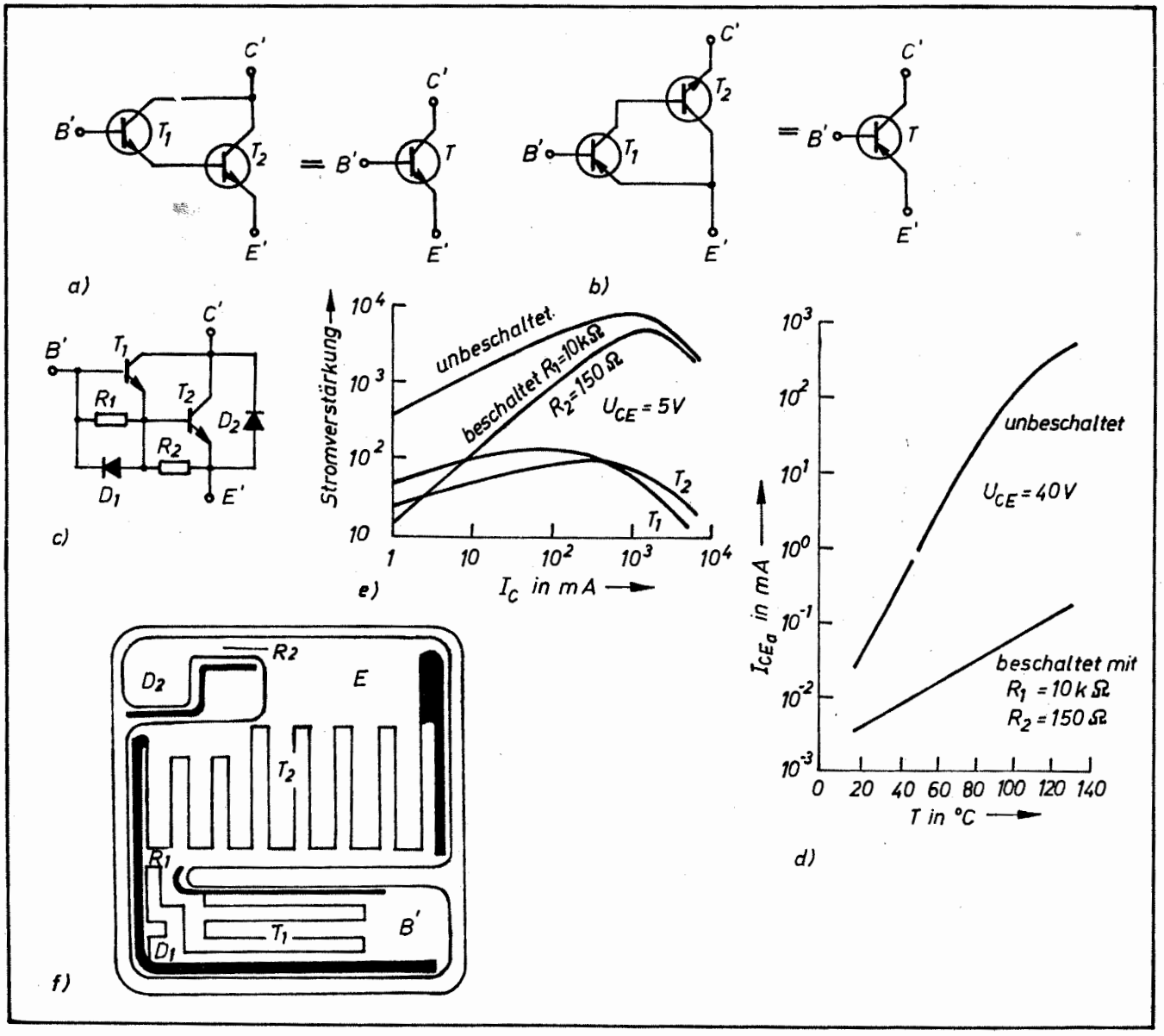


Bild 7.3.7. DARLINGTON-Transistoren

a) Zwei *npn*-Transistoren und deren Ersatztransistor
b) Ersatztransistor für je einen *pnp*- und *npn*-Transistor
c) Standardschaltung heutiger Leistungs-DARLINGTON-Transistoren. D_1 erhöht die Schaltgeschwindigkeit, und D_2 begrenzt die Umkehrspannung bei induktiver Last. Die Widerstände R_1 und R_2 bewirken bessere Arbeitspunkte
d) Sperrströme als Funktion der Temperatur
e) Stromverstärkung als Funktion des Kollektorstromes
f) Layout eines modernen Leistungs-DARLINGTON-Transistors für große Sperrspannungen und Lastströme

Näherungen

$$\left.\begin{aligned} h_{11}{}^{+} &\approx {}^{1}h_{11} + {}^{2}h_{21}\,{}^{2}h_{11} \approx 2\,{}^{1}h_{11}\,, \\ h_{12}{}^{+} &\approx {}^{1}h_{22}\,{}^{2}h_{11}\,, \\ h_{21}{}^{+} &\approx {}^{1}h_{21}\,{}^{2}h_{21}\,, \\ h_{22}{}^{+} &\approx {}^{2}h_{22} + {}^{2}h_{21}\,{}^{1}h_{22} \approx {}^{2}h_{22}\,. \end{aligned}\right\} \qquad (27)$$

Diese Betrachtungen gelten für die Emitterschaltung. Für die beiden anderen Schaltungen des DARLINGTON-Transistors — die jedoch so gut wie nicht verwendet werden — sind erst nach der Zusammenfassung der Emitterdaten die Umrechnungen vorzunehmen.

Der *Hauptvorteil* der DARLINGTON-Transistoren liegt darin, mit wenig Aufwand große Leistungen bei kleinen Steuerströmen zu verstärken. Dieser Vorteil ist mit einer Reihe von *Nachteilen* verknüpft (z.B. [S 27]):

- Die *Sättigungsspannung* ist größer als bei diskreten Bauelementen.

Dies ist dadurch bedingt, daß eine Basis-Emitter-Spannung und eine Kollektor-Emitter-Spannung in Reihe liegen.

- Der *Sperrstrom* ist größer und stärker temperaturabhängig.

Dies ist durch die nicht gleichzeitig mögliche ideale Arbeitspunkteinstellung für beide Transistoren bedingt. Dem störend überlagert ist noch die Wärmekopplung vom Leistungstransistor auf den Vorstufentransistor.

Beide Nachteile sind durch *zusätzliche Beschaltung* mit Widerständen und Dioden gemäß Bild 7.3.7c) zu verringern (Teilbild d)). Dies hat jedoch, wie Teilbild e) zeigt, wieder geringere Stromverstärkung zur Folge. Dennoch ist die Schaltung von Teilbild c) der heutige Standardaufbau von DARLINGTON-Transistoren, wozu Teilbild f) ein Layout eines Leistungs-DARLINGTON-Transistors für Sperrspannungen um 500 V und Ströme um 20 A zeigt. Die Stromverstärkung liegt dabei aber nur um 50. Aus diesen Daten und dem Vergleich mit den Leistungs-FET wird deutlich, daß die Bedeutung der DARLINGTON-Transistoren wahrscheinlich zurückgeht (vgl. Bild 5.7.21h)).

7.3.2.2. Kaskodeverstärker

Bei der Kaskodeschaltung werden Transistoren „übereinander" (Kaskade) angeordnet. Sie kann mit Bipolartransistoren (Bild 7.3.8a)), mit FET (Teilbild b)) oder auch gemischt ausgeführt werden. Der untere Transistor arbeitet in Emitter- bzw. Sourceschaltung, der obere in Basis- bzw. Gateschaltung. Der *Hauptvorteil* der Kaskodeschaltung liegt *bei hohen Frequenzen* in zwei Fakten:

- sehr gute Entkopplung von Ein- und Ausgang
- sehr kleine MILLER-Kapazität.

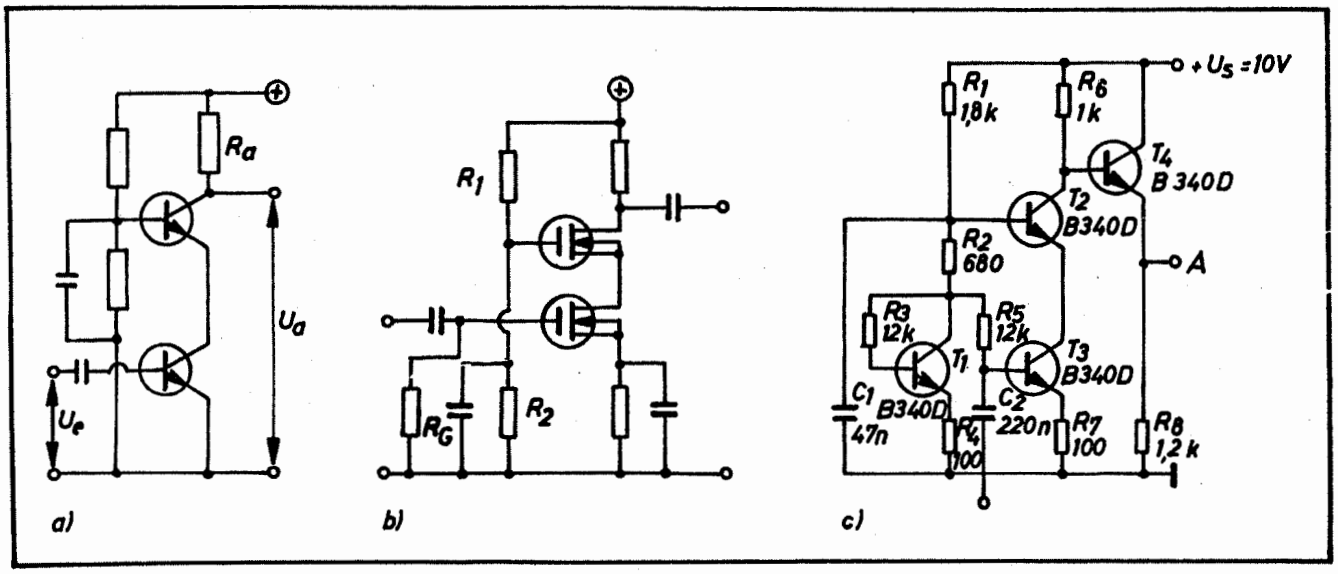


Bild 7.3.8. Kaskodeschaltungen

a) mit bipolaren Transistoren, b) mit FET

c) mit einem Transistorarray B340D, wobei T_2 und T_3 die eigentliche Kaskode, T_4 einen Emitterfolger darstellen und T_1 zur Arbeitspunkteinstellung dient

Die Entkopplung erfolgt vor allem in der oberen Stufe dadurch, daß das Gate bzw. die Basis wechselstrommäßig auf Masse liegt. Deshalb ist der Kondensator notwendig. Die kleine MILLER-Kapazität entsteht, weil die untere Stufe auf dem kleinen Eingangswiderstand der oberen Stufe arbeitet. Dadurch liegt ihre Verstärkung bei etwa Eins, und die Gate-Drain- bzw. Emitter-Kollektor-Kapazität wird bezüglich des Eingangs nicht erhöht (5.5.3. und 7.1.4.). Die *Spannungsverstärkung* der Kaskodeschaltung wird also vor allem durch die obere Stufe bewirkt.

Die Kaskodeschaltung kann auch durch einen *Ersatztransistor*, ähnlich wie die DARLINGTON-Schaltung, beschrieben werden. In erster Näherung gilt dann, daß der Eingangswiderstand dem des unteren Transistors und die Stromverstärkung der des oberen Transistors entspricht. Für den Innenwiderstand gilt bei Bipolartransistoren

$$R_i \approx \frac{{}^{u}h_{21}}{{}^{o}h_{22}\,{}^{u}h_{22}\,{}^{u}h_{11}}. \tag{28}$$

Eine sehr effektive, dimensionierte Kaskodeschaltung mit einem Transistorarray zeigt Bild 7.3.8. Die eigentliche Kaskode besteht aus den Transistoren T_2 und T_3. Der Transistor T_4 sorgt als Kollektorschaltung für einen sehr niederohmigen Ausgang. Die Widerstände R_1 bis R_4 und der Transistor T_1 erzeugen die notwendigen Arbeitspunkte für T_2 und T_3. Die Spannungsverstärkung ist in erster Näherung durch das Verhältnis $R_6 : R_7$ (hier 10fach) gegeben. Die Bandbreite dieser Schaltung liegt bei ca. 6 MHz.

7.3.2.3. WHITE-Folger und ähnliche Schaltungen

Bei der Kollektor- bzw. Drainschaltung wäre es günstig, wenn der Emitter bzw. Sourcewiderstand für Gleichstrom sehr klein und für Wechselstrom sehr groß wäre. Dann könnte nämlich der statische Arbeitspunkt mit dem Gleichstromwert auf günstige Parameter, z. B. große Steilheit, eingestellt und trotzdem die Gegenkopplung voll wirksam werden. Außerdem wären große, unverzerrte Ausgangsspannungen möglich. Dieses Prinzip wird im WHITE-Folger dadurch realisiert, daß der *Widerstand durch eine Wechselstromquelle*, z. B. mit einem zweiten Transistor, realisiert wird. Bild 7.3.9a) zeigt dies in einer kombinierten Schaltung, wobei der FET als Verstärker und der bipolare Transistor wie eine Stromquelle wirken. Die WHITE-Schaltung ist in jeder Kombination aus bipolaren Transistoren und FET zu realisieren.

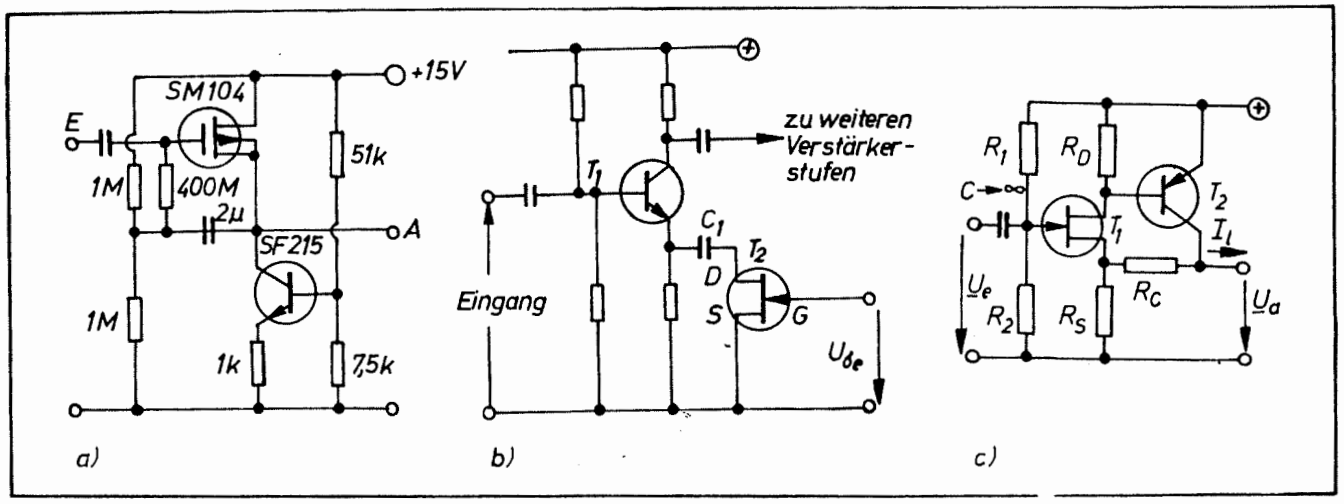


Bild 7.3.9. a) WHITE-Folger aus FET als Verstärker und bipolarer Transistor als Stromquelle
b) Über den Innenwiderstand des FET einstellbare Verstärkerregelung
c) Source- und Emitterschaltung, die für $R_C \to 0$ in einen fast idealen Sourcefolger übergeht

Eine andere Anwendung des Wechselstromwiderstandes eines Transistors (hier FET) zeigt Bild 7.3.9b). Hier überbrückt der FET mehr oder weniger den Emitterwiderstand des bipolaren Transistors. Dadurch ist eine einfache *Verstärkungsregelung* mit U_{be} möglich.

Bild 7.3.9c) zeigt eine Kombination von Source- und Emitterschaltung. Sie realisiert bei hohem Eingangswiderstand eine durch das Verhältnis $(R_S + R_C):R_S$ bestimmte Verstärkung. Für $R_C \to 0$ geht sie in einen fast idealen Sourcefolger (Drainschaltung) über. Ihre Verstärkung weicht extrem wenig von 1 ab,

$$V = 1 - \frac{1}{1 + (1 + h_{21})\, SR_S}, \tag{29}$$

während der Innenwiderstand sehr klein ist:

$$R_i \approx \frac{1}{h_{21} S}. \tag{30}$$

7.3.2.4. Schaltungen mit großer Verstärkung

Infolge der Sättigungscharakteristik von FET und Bipolartransistoren ist ein ohmscher Arbeitswiderstand immer klein gegen den Innenwiderstand. Dies führt entsprechend der Ersatzschaltung in Bild 5.5.2c) bzw. 5.6.1d) dazu, daß die maximal mögliche Verstärkung von z. B. $1/D$ auch nicht annähernd erreicht wird. Wesentlich höhere Verstärkungsfaktoren lassen sich mit einer *Stromquelle über ein aktives Bauelement* erreichen. Dies zeigt Bild 7.3.10a), wobei der bipolare Transistor den *dynamischen Widerstand* darstellt.

Eine andere Möglichkeit bietet die *Bootstrap-Schaltung* gemäß Bild 7.3.10b). T_2 ist hier ein Emitterfolger, und die Rückkopplung über C_1 bewirkt eine beträchtliche Vergrößerung von R_1. Gleichzeitig wird infolge des Emitterfolgers an A ein kleiner Innenwiderstand erzeugt.

Eine weitere Möglichkeit zeigt Bild 7.3.10c). Hier ist T_1 eine Stromquelle und T_3 ein Sourcefolger. Die Gleichstromgegenkopplung R_4, R_1 stabilisiert den Arbeitspunkt der Schaltung.

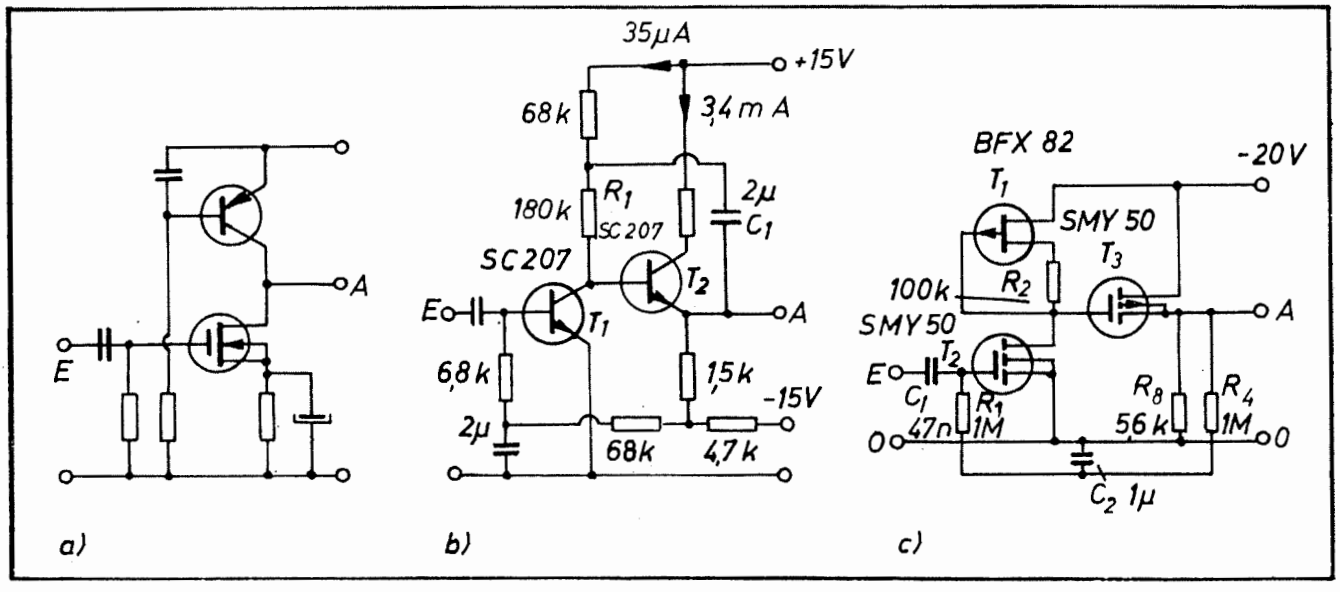


Bild 7.3.10. Schaltungen mit großer Verstärkung

a) mit Bipolartransistor als Stromquelle,

b) mit Bootstrap-Schaltung für den Arbeitswiderstand (Verstärkung etwa 2000fach),

c) mit T_1 als Stromquelle und T_3 als Sourcefolger

7.3.2.5. Stabiler Innenwiderstand Null

Für den Innenwiderstand einer rückgekoppelten Schaltung gilt Gl. (7.1.9). Mittels Strommitkopplung kann er zu Null gemacht werden. Dies bedeutet aber, wie dort gezeigt wurde, noch nicht Instabilität. Die Stabilität wird nämlich durch das Verhältnis aus der Verstärkung ohne und mit Rückkopplung (Gl. (7.1.18)) bestimmt. Dies beides wird in einer Schaltung gemäß Bild 7.3.11 ausgenutzt. Mittels R_k erfolgt bei ihr eine Spannungsgegenkopplung und mittels R_{SS} eine Strommitkopplung. Beide zusammen bestimmen die Verstärkung gemäß

$$V^+ \approx R_k/R_{S1} ,$$

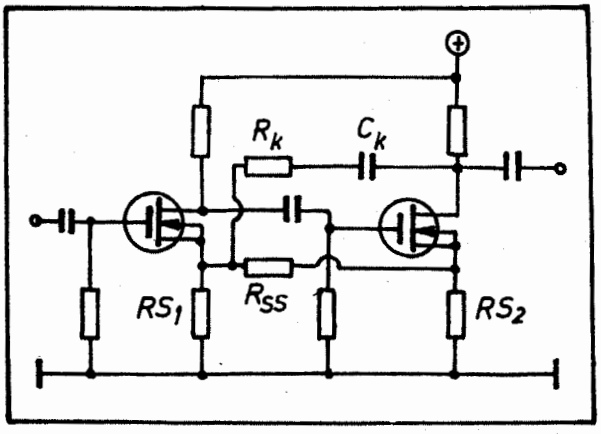


Bild 7.3.11. Verstärker mit extrem kleinem und stabilem Innenwiderstand. Die Verstärkung wird im wesentlichen durch das Verhältnis R_k/R_{S1} bestimmt. Der Innenwiderstand läßt sich mittels R_{SS} auf positive und negative Werte um null Ohm abgleichen

sie kann also in einem weiten Bereich und damit auch kleiner als V eingestellt werden. Der Innenwiderstand Null wird allein durch R_{SS} bestimmt. Er wird Null für etwa

$$R_{SS} = R_{S1}R_{S2}V_1S_2 ,$$

wobei V_1 die Verstärkung der ersten Stufe bei Abtrennung von R_k und R_{SS} ist. Bei günstiger Einstellung der Widerstände R_k und R_{SS} bietet die Schaltung Möglichkeiten, die über andere von zweistufigen Verstärkern hinausgehen. So kann z.B. in die Ausgangsleitung, auch wenn sie nicht abgeschirmt ist, keine Störung eindringen (Innenwiderstand Null!).

7.3.3. Differenzverstärker

7.3.3.1. Einführung

Ein idealer Differenzverstärker kann gemäß Bild 7.3.12a) beschrieben werden. Er hat einen invertierenden (mit — im Dreieck gekennzeichnet) und einen nichtinvertierenden (mit + bezeichnet) Eingang. Für die *Eingangsspannung zwischen beiden Eingängen* gilt

$$u_e = u_{e1} - u_{e2} . \tag{31}$$

Diese Eingangsspannung wird zum Ausgang hin mit der Verstärkung V_D (von Differenz) verstärkt:

$$u_a = V_D(u_{e1} - u_{e2}) . \tag{32}$$

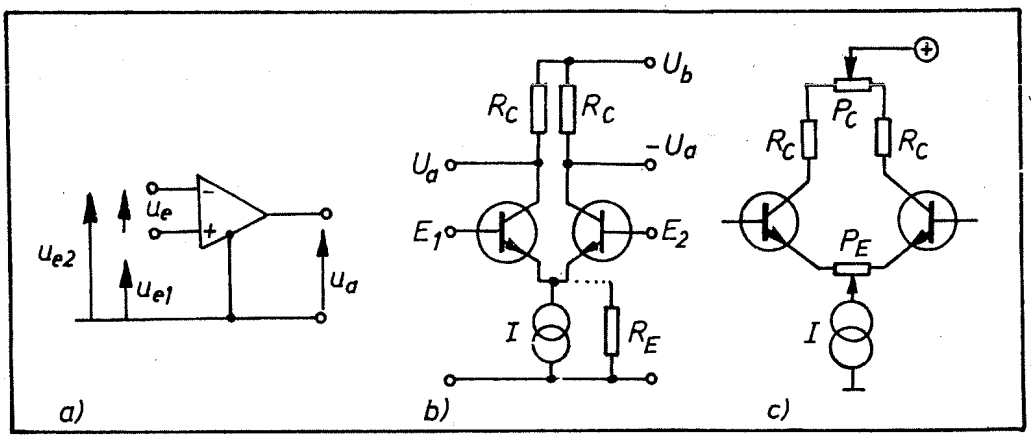


Bild 7.3.12. Zum Differenzverstärker

a) Prinzipanordnung

b) Grundschaltung zur Berechnung der wichtigsten Daten

c) Prinzip des Offsetabgleichs mit P_E und des Verstärkerabgleichs mit P_C.

Es geht daraus hervor, daß für

$$u_G = u_{e1} = u_{e2}\,, \tag{33}$$

also *Gleichtaktansteuerung*, keine Ausgangsspannung entsteht. Dies ist jedoch nur theoretisch richtig. In der Praxis wird auch bei Gleichtaktansteuerung eine gewisse, wenn auch sehr kleine Gleichtaktverstärkung V_G vorhanden sein:

$$u_a = V_G \cdot u_G\,. \tag{34}$$

Das Verhältnis der beiden Verstärkungen wird *Gleichtaktunterdrückung* oder **CMRR** (common mode rejection ratio) genannt:

$$G = V_D/V_G\,. \tag{35}$$

Zuweilen wird auch die Abkürzung CMR benutzt, sie wird dann meist in dB angegeben

$$G_{dB} = 20 \log G \tag{36}$$

7.3.3.2. Grundschaltung

Für den Aufbau eines Differenzverstärkers gibt es viele Möglichkeiten. Fast alle heutigen Schaltungen leiten sich jedoch aus dem *Prinzip* von Bild 7.3.12b) ab, wobei angenommen sei, daß die beiden Transistoren und Kollektorwiderstände genau gleich seien. Die Stromquelle I besitze einen Widerstand R_E oder sei durch den entsprechenden ohmschen Widerstand ersetzt. Werden dann die Eingänge E_1 und E_2 genau gegenphasig erregt, so teilt sich der Strom dadurch unterschiedlich auf die beiden Transistoren auf, wobei das Emitterpotential gleich bleibt. Dies bedeutet, daß in diesem Fall R_E keine Gegenkopplung bewirkt, beide Stufen verstärken wie die normale Emitterschaltung mit R_C als Arbeitswiderstand. Deshalb folgt die *Differenz-* (*Gegentakt-*) *Verstärkung* näherungsweise zu

$$V_D \approx -\frac{1}{2}\frac{h_{21}}{h_{11}} R_C\,. \tag{37}$$

Der Faktor 1/2 folgt, weil in Gl. (32) die Verstärkung auf die Differenzspannung gemäß Gl. (31) bezogen ist. Werden beide Eingänge mit der Gleichtaktspannung gemäß Gl. (33) betrieben, so ändert sich in fast gleichem Maße das Emitterpotential der beiden Transistoren. Infolge des endlichen Innenwiderstandes R_E der Stromquelle

folgt dadurch auch eine Stromänderung, die an den R_C eine Spannungsänderung hervorruft. Sie ist zwar ebenfalls an beiden R_C gleichphasig, wird jedoch voll wirksam, da meist nur einer der Ausgänge weiter verwendet wird. Damit folgt für die *Gleichtaktverstärkung*

$$V_G \approx \frac{h_{21}}{h_{11}} R_E , \tag{37a}$$

und folglich beträgt die Gleichtaktunterdrückung

$$G \approx R_C/2R_E . \tag{38}$$

Der Ausgangswiderstand an jedem der beiden Kollektoren beträgt

$$R_i \approx h_{11}/h_{21} . \tag{39}$$

Für die Eingangswiderstände müssen die beiden Betriebsarten unterschieden werden. Der Differenzeingangswiderstand besitzt — da keine Gegenkopplung auftritt — den doppelten Wert einer Stufe:

$$R_{eD} \approx 2h_{11} . \tag{40}$$

Bei Gleichtaktansteuerung wird dagegen die Gegenkopplung über R_E voll wirksam, und von einem Eingang nach Masse beträgt der Eingangswiderstand

$$R_{eG} \approx h_{21} R_E . \tag{41}$$

Bisher war vorausgesetzt, daß die beiden Transistoren identisch sind. Dies ist in der Praxis jedoch nie exakt erfüllt, nicht einmal dann, wenn sie auf dem gleichen Chip unmittelbar nebeneinander integriert sind. Die Differenz bewirkt, daß die exakte Aufteilung des Stromes I für etwas unterschiedliche Basis-Emitter-Spannungen erfolgt. Die zum Abgleich notwendige Differenzspannung heißt *Offsetspannung*, oft auch nur Offset. Ihre Kompensation erfolgt am besten mittels eines Potentiometers P_E gemäß Bild 7.3.12c). Zuweilen wird auch der von außen über eine Basis notwendige zusätzliche Strom als *Offsetstrom* angegeben.

Weiter kann in der Schaltung eine Differenz zwischen den verschiedenen h-Parametern der Transistoren bzw. den beiden Kollektorwiderständen auftreten. Dies führt zu unterschiedlich großen Verstärkungen an den beiden Ausgängen. Diese Differenz kann mittels eines Potentiometers P_C (s. Bild 7.3.12c)) abgeglichen werden.

Eine weitere Unsymmetrie kann durch Temperatureinflüsse auftreten. Sie wirken vor allem als *Offsetdrift*.

Es ist zu beachten, daß der Differenzverstärker von Bild 7.3.12b) bei einseitiger Ansteuerung eine Gegentaktausgangsspannung erzeugt, die gegebenenfalls mittels P_E in Bild c) symmetriert werden kann. Die eine Basis ist zu diesem Zweck auf Masse zu legen. Der Differenzverstärker stellt dadurch auch eine wertvolle Ergänzung zu der Schaltung in Bild 7.3.2d) dar.

7.3.3.3. Gleichphasiger Ausgang

Die übliche Differenzstufe gemäß Bild 7.3.12b) hat im Gegensatz zu Teilbild a) einen *Gegentaktausgang*. Im Prinzip — und so wurde bisher gerechnet — genügt es, nur einen Ausgang als asymmetrischen Ausgang zu verwenden. Dies bringt aber Nachteile für das Ausgangssignal. In ihm sind dann enthalten

- Änderungen der Versorgungsspannung
- noch vorhandene Gleichtaktsignale
- Driftspannungen der Folgestufen.

Gelingt es dagegen, nur die genau gegenphasigen Signale weiter zu verstärken, so werden bessere Ergebnisse erreicht. Die entsprechenden Schaltungen werden zuweilen phasenaddierende Schaltungen genannt und sind u. a. in [H 24] ausführlich behandelt. Hier sind in Bild 7.3.13 einige Beispiele zusammengestellt. In Teilbild a) wirkt am Transistor T_3 nur die Differenz zwischen den beiden Ausgängen des Differenzverstärkers. Die Diode D_1 gleicht dabei das Emitter-Basis-Potential von T_3 zum Teil sogar mit dem Temperatureinfluß aus. Der Nachteil ist eine relativ hohe Belastung und eine kleine Verstärkung. Eine Trennstufe mit T_4 in Kollektorschaltung, wie es Teilbild b) zeigt, vermeidet einige Nachteile.

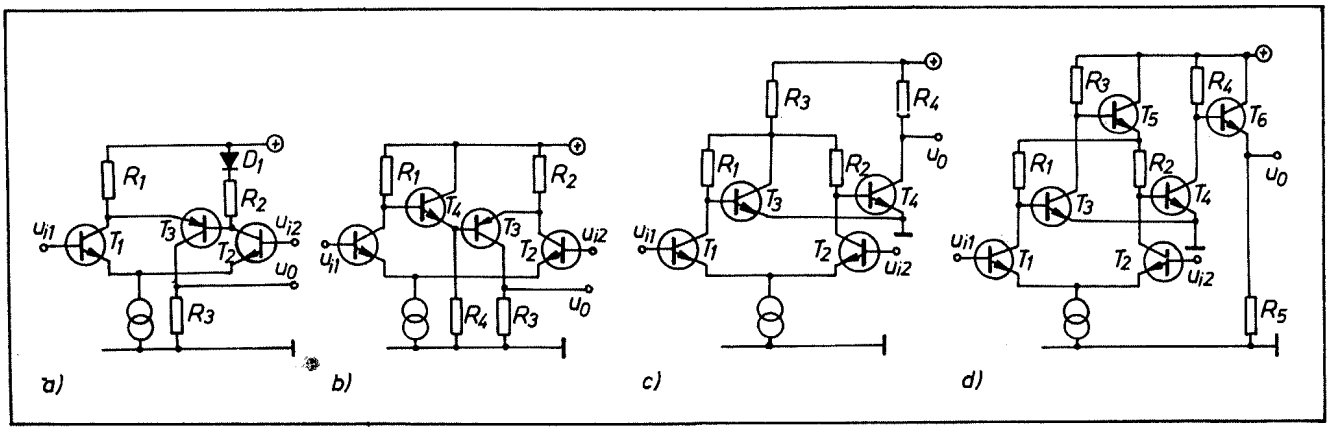


Bild 7.3.13. Phasenaddierende Schaltungen

a) Die Spannung zwischen Emitter und Basis von T_3 wird weiter verstärkt. D_1 bewirkt Kompensation von U_{EB} von T

b) Verbesserung von Schaltung a) durch Einfügen der Kollektorstufe T_4

c) Einfügen eines Differenzverstärkers aus T_3 und T_4

d) Weitere Verbesserungen des Prinzips von c) mittels T_5 und T_6

Eine andere Möglichkeit besteht darin, auch für den Übergang zur Asymmetrie das Prinzip eines Differenzverstärkers zu verwenden. Ein entsprechendes Beispiel zeigt Teilbild c). Mit dem Einsatz von zwei weiteren Transistoren läßt sich diese Schaltung noch weiter verbessern (s. Teilbild d)).

Ein weiteres Prinzip beruht auf dem im Abschnitt 7.3.4.2. behandelten Stromspiegel. Es ist in Bild 7.3.14e) dargestellt und beruht darauf, daß die Ströme durch T_3 und T_4 gleich groß sind. Abweichungen davon bewirken über die großen Innenwiderstände große Spannungsänderungen.

7.3.3.4. Schaltungsvarianten

Das Prinzip des Differenzverstärkers von Bild 7.3.12b) kann neben den bipolaren Transistoren auch auf FET angewendet werden. Dadurch ergeben sich vor allem Änderungen der erreichbaren Daten, aber nicht in der Funktion. Bedeutsam sind u. a. die unterschiedlichen Offsetwerte:

	Bipolartransistor	SGFET	MOSFET
$\Delta U/\Delta T$	0,5 ... 10 µV/K	10 µV/K	5...100 µV/K
$\Delta I/\Delta T$	1 pA/K	1 ... 10 pA/K	$<$ 1 pA/K

Weiter können die einzelnen Transistoren auch durch DARLINGTON- bzw. Kaskodestufen ersetzt werden. Hierfür zeigen Bild 7.3.14a) und b) Beispiele. Bild 7.3.14c) zeigt, wie mit Dioden die Eingangsspannung zwischen den beiden Eingängen begrenzt werden kann. Teilbild d) nutzt das Prinzip der sehr hohen Verstärkung aus (s. 7.3.2.4.). Dabei ist sogar ein relativ einfacher, wenn auch nicht sehr effektiver asymmetrischer Übergang möglich.

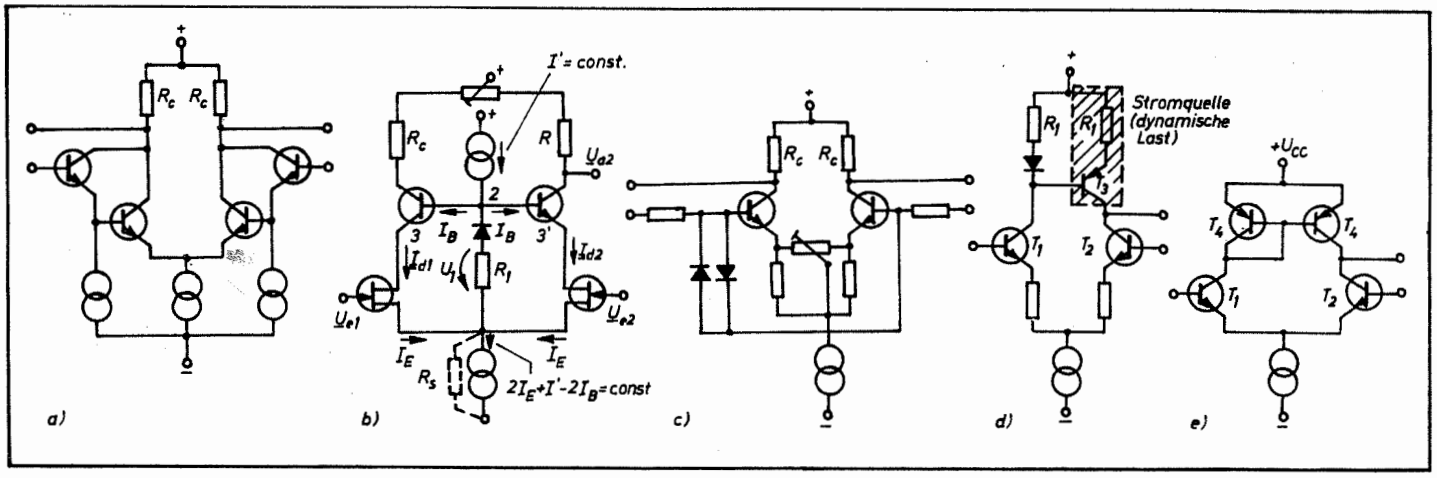


Bild 7.3.14. Varianten von Differenzverstärkern

a) Mit DARLINGTON-Stufe
b) Mit FET-Eingang und gemischter DARLINGTON-Stufe
c) Mit Stromgegenkopplung und Diodenbegrenzung der Eingangsspannung
d) Mit dynamischem Arbeitswiderstand über Stromquelle
e) Mit Stromspiegel im Kollektorkreis

7.3.3.5. NORTHON-Verstärker

Alle bisher behandelten Differenzverstärker sind durch zwei Effekte gekennzeichnet. Sie sind spannungsgesteuert und benötigen eine zusätzliche Schaltung zum asymmetrischen Ausgang. Der NORTHON-*Verstärker* ist dagegen

- gesteuert durch die Differenz von *Eingangsströmen* und
- hat unmittelbar einen *asymmetrischen Ausgang*.

Dies bringt den weiteren Vorteil mit sich, daß

- die *Versorgungsspannung nur eine Polarität* benötigt.

Die Funktionsweise wird wesentlich durch den im Abschnitt 7.3.4.2. zu behandelnden *Stromspiegel* bestimmt. In Bild 7.3.15 sind dies die Transistoren T_1 und T_2. Der Eingangsstrom I_{ip} bewirkt an dem als Diode geschalteten Transistor T_1 den zugehörigen Spannungsabfall U_{EB}. Dieser wirkt zugleich am Transistor T_2 und läßt durch ihn den gleichen Strom fließen. Wird dieser Strom als I_{in} eingekoppelt, so gelangt an die

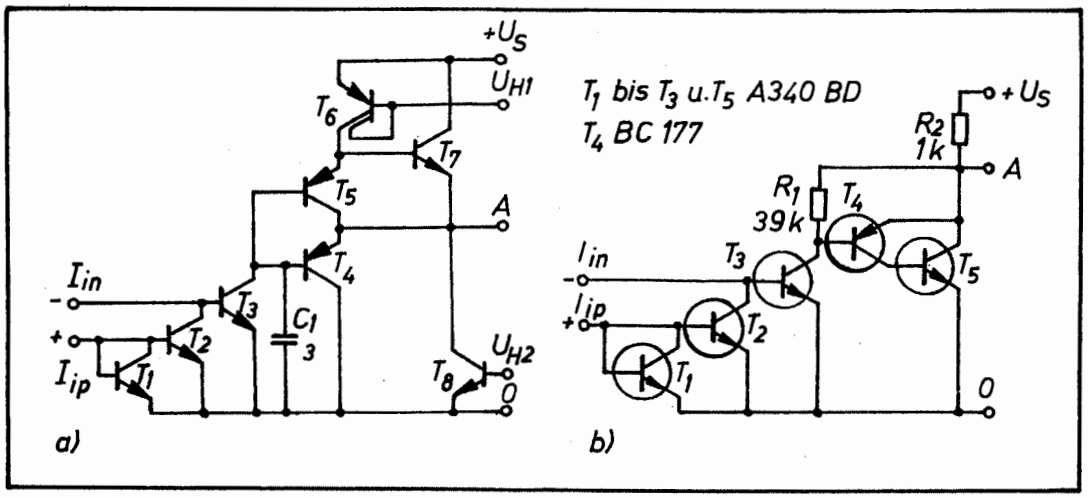


Bild 7.3.15. NORTHON-Verstärker

a) Vereinfachte Innenschaltung des Schaltkreises LM 3900
b) Diskreter Aufbau mit Transistorarray B340D

Basis von T_3 kein Strom. Weicht I_{in} jedoch hiervon ab, so muß die Basis von T_3 die Differenz aufnehmen und kann sie weiter verstärken. Die weiteren Transistoren bewirken eine Stabilisierung der Schaltung und eine ausreichende Verstärkung. So entstehen mit der Gesamtschaltung Daten, wie sie Tab. 7.3.1 zeigt. Daraus geht ein weiterer Vorteil hervor, nämlich daß nur eine Betriebsspannung gegen Masse notwendig ist, aber auch ein Betrieb mit symmetrischer Betriebsspannung möglich ist. Weiter sind die Betriebsdaten in einem weiten Bereich von der Größe der Betriebsspannung unabhängig.

In Teilbild b) ist noch ein diskret aufgebauter NORTHON-Verstärker mit Daten aus Tab. 7.3.2 gezeigt.

Tabelle 7.3.1. Übersichtsdaten eines Verstärkers des Bausteins LM 3900

Speisespannungsbereich	+4 ... +36 V oder ±2 ... ±18 V
Leerlaufverstärkung bei $R_1 = 10$ kΩ	70 dB
Bandbreite-Verstärkungsprodukt	2,5 MHz
Eingangsstrom	30 nA
Anstiegsgeschwindigkeit	0,5 V/μs
Maximaler Ausgangsstrom	30 mA
Ausgangswiderstand	8 kΩ

Tabelle 7.3.2. Übersichtsdaten des in Bild 7.3.15 b) dargestellten Verstärkers

Speisespannungsbereich	+3 ... +15 V oder (±) 2,5 ... (±) 7,5 V
Leerlaufverstärkung bei $R_I = 10$ kΩ	65 dB
Eingangsstrom	150 nA
Bandbreite-Verstärkungsprodukt	4 MHz

7.3.4. Grundschaltungen für IC

7.3.4.1. Prinzipien der analogen IC

Integrierte Schaltungen (integrated circuits = IC) sind gegenüber den Schaltungen aus diskreten Bauelementen durch mehrere *Besonderheiten* gekennzeichnet:

- Die Grundprinzipien sind trotz der folgenden Fakten im wesentlichen gleich denen der diskreten Schaltungen.
- Integrierte Schaltungen werden vollständig in einem geschlossenen technologischen Prozeß (meist Epitaxie-Planartechnik) hergestellt. Dies bedeutet u. a., daß alle Anschlüsse von der Oberfläche des Chips erfolgen müssen. Beispiele für Transistoren enthält hierzu z.B. der Vergleich von Bild 5.6.7 mit 5.6.8. Ein auffälliges Beispiel dazu ist der *pnp*-Lateraltransistor.
- Der Preis für die integrierte Schaltung wird vor allem durch die benötigte Si-Oberfläche bestimmt. Hierdurch müssen alle Kapazitäten > 50 pF und große Widerstände (> 20 kΩ) vermieden werden bzw. durch Dioden, Transistoren und deren Zusammenschaltungen realisiert werden.
- Infolge der Kleinheit der Chips (wenige mm²) ist die Wärmeabfuhr beschränkt. Deshalb werden leistungsarme Schaltungen angestrebt.
- Weiter sind durch die Kleinheit alle Strukturen thermisch verkoppelt. Daraus entstehen Vorteile bei Stabilisierungsschaltungen, Nachteile bei Leistungsstufen.

- Die Toleranzen zwischen verschiedenen Chargen sind groß, auf einem Chip dagegen klein. Dies läßt gute Transistorpaare erreichen, bringt aber Toleranzen zwischen verschiedenen vollständigen Schaltungen. Insbesondere streuen die Emitterstromdichte von *npn*-Transistoren, die Stromverstärkung und U_{BE} von Transistoren auf einem Chip nur sehr wenig.
- Induktivitäten und große Kapazitäten müssen als externe Bauelemente angefügt werden.
- Soweit wie möglich wird die Gleichstromkopplung zwischen Verstärkerstufen verwendet. Daraus ergeben sich Probleme des Drifts, der Notwendigkeit von Stufen zum Potentialversatz.
- Im Gegensatz zu digitalen integrierten Schaltungen ist bei den analogen infolge der äußeren Beschaltung meist eine gute Kenntnis ihrer inneren Schaltungsstruktur notwendig.

Bedeutsam für die praktische Realisierung von analogen Schaltkreisen ist, daß sie aus einer *relativ kleinen Anzahl von Grundfunktionen*, die zum größten Teil schon besprochen sind, aufgebaut werden. Dies sind vor allem

- Widerstände (5.1.)
- Kondensatoren (5.2.)
- Dioden (5.4.)
- FET (5.5.8.)
- Bipolartransistoren (5.6.5.)
- Stromquellen (7.3.1.2.; #)
- Spannungsquellen (7.3.1.3.;#)
- Potentialversatzstufen (7.3.2.; #)
- DARLINGTON-Stufen (7.3.2.1.)
- Kaskadestufen (7.3.2.2.)
- dynamischer Arbeitswiderstand (7.1.4.1., 7.3.2.1., 7.3.3.3.)
- Endstufen (7.4.).

Ein # bedeutet, daß hierzu im folgenden noch Ergänzungen gegeben werden. Die Endstufen werden erst im Abschnitt 7.4. und dann gleich zusammenfassend behandelt.

7.3.4.2. Stromspiegel

Der Stromspiegel ist eine spezielle *Kontaktstromquelle* der integrierten Technik. Sie ist nur mit weitgehend gleichen Transistoren effektiv zu gestalten. Das Grundprinzip zeigt Bild 7.3.16a). Für den Basisstrom der beiden gleichen Transistoren T_1 und T_2 gilt

$$I_B = \frac{I_S}{h_{21} + 1} [\exp (U_{BE}/U_T) - 1] . \tag{42}$$

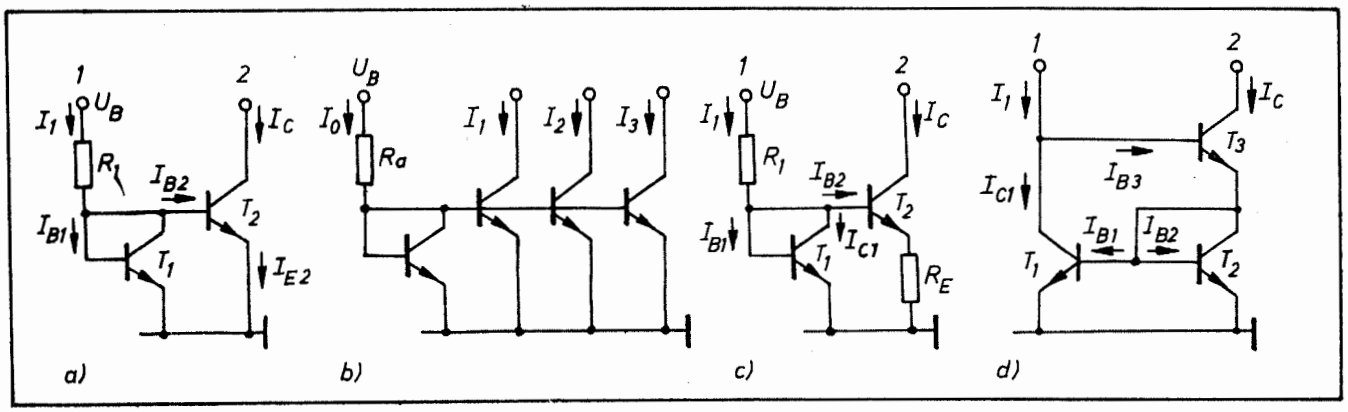


Bild 7.3.16. Zum Stromspiegel

a) Grundprinzip
b) Strombank bzw. Mehrfachstromspiegel
c) Stromspiegel für kleine Ströme
d) Stromspiegel mit besonders hohem Innenwiderstand

Darin bedeuten I_S den Emitter-Sättigungsstrom, h_{21} die Stromverstärkung, U_{BE} die Basis-Emitter-Spannung und U_T die Temperaturspannung e_0/kT. Für den Strom I_1 gilt

$$I_1 = (h_{21} + 1)\, I_{B1} + I_{B2} \tag{43}$$

und für I_C

$$I_C = h_{21} I_{B2}\,. \tag{44}$$

Die Ströme I_{B2} und I_{B1} verhalten sich infolge konstanter Emitterstromdichte wie die Emitterflächen A_E der beiden Transistoren. Dies gilt auch für die entsprechenden Sättigungsströme. Für das Verhältnis beider Ströme, das Spiegelverhältnis, folgt damit

$$\frac{I_C}{I_1} = \frac{h_{21} I_{S2}}{(h_{21} + 1)\, I_{S1} + I_{S2}} \approx \frac{A_{E2}}{A_{E1}}\,. \tag{45}$$

Die Näherung wird besonders gut für große Stromverstärkungen erfüllt und bedeutet, daß das Verhältnis beider Ströme sehr gut über die Emitterflächen, also geometrisch, festzulegen ist. Für gleich große Flächen wird also ein eingespeister Strom I_1 als I_C gespiegelt am zweiten Anschluß erscheinen. Dabei bleibt infolge der Sättigungscharakteristik des Transistors die Spannung am Kollektor von T_2 fast ohne Einfluß. Mit dem Flächenverhältnis können auch andere Stromwerte erzeugt werden, sofern das Verhältnis nicht zu stark von Eins abweicht (ca. 10:1 bis 1:10). Weiter läßt sich auf dieser Basis eine *Strombank*, auch Mehrfachstromquelle, gemäß Bild 7.3.16b) aufbauen. Der Strom I_0 bestimmt dann den Wert der Ströme I_1, I_2, I_3 usw.

Für größere Stromverhältnisse sind geringe *Abwandlungen* des Stromspiegels notwendig. Den wichtigsten Fall zeigt Bild 7.3.16c). Mit dem eingefügten Widerstand R_E wird die wirksame Basis-Emitter-Spannung von T_2 verkleinert:

$$U_{EB2} = U_{EB1} - I_{E2} R_E\,. \tag{46}$$

Mit den Näherungen $h_{21} > 1$ und $\exp(U_{BE}/U_T) > 1$ läßt sich dann analog zu oben ableiten

$$\frac{I_1}{I_C} \approx \frac{A_{E1}}{A_{E2}} \exp(I_C R_E / U_T)\,. \tag{47}$$

Soll z.B. ein Strom von $I_C = 10\,\mu A$ erzeugt werden und wird $A_{E1}/A_{E2} = 10$ gewählt, so folgt für einen Widerstand von $R_E = 5\,k\Omega$ $I_1/I_C = 10\,e^2 \approx 74$, also ein Strom I_1 von 0,74 mA. Dieser Strom läßt sich aus einer konstanten Spannung von 5 Volt mit einem R_1 von ebenfalls 5 kΩ erzeugen.

Die Schaltung von Bild 7.3.16d) erzeugt einen besonders hohen Innenwiderstand des Stromspiegels. Er wird um die Stromverstärkung des Transistors T_3 nochmals vergrößert. Beträgt er bei den anderen Schaltungen $1/h_{22}$, so gilt hier

$$R_i \approx \frac{1 + h_{21}}{h_{22}}\,, \tag{48}$$

Eine andere Möglichkeit der Verbesserung von Eigenschaften von Stromspiegeln besteht darin, statt der einfachen Transistoren DARLINGTON-Transistoren zu verwenden. Dies ist dann besonders wichtig, wenn die Stromverstärkung, wie z.B. bei (Lateral-)*pnp*-Transistoren, nicht sehr groß ist.

7.3.4.3. Potentialverschiebestufen

In der integrierten Technik müssen wegen *fehlender Koppelkondensatoren* die einzelnen Verstärkerstufen gleichstrommäßig gekoppelt werden. Hierbei summieren sich im

allgemeinen die Arbeitspunktpotentiale auf, und es müssen spezielle Schaltungen verwendet werden, um die Potentiale wieder auf den günstigen Wert zurückzuschieben. Auf diese Probleme wurde u. a. schon bei den DARLINGTON-Transistoren (7.3.2.1.) und generell in Abschnitt 7.3.2. eingegangen. Dort wurde im Zusammenhang mit Bild 7.3.6b) auch auf eine Möglichkeit verwiesen, nämlich die Verwendung *komplementärer Transistoren*, die jedoch in der Strukturierung und bezüglich der schlechten Eigenschaften der *pnp*-Transistoren problematisch ist. Deshalb wird von diesem Prinzip relativ selten Gebrauch gemacht. Es wurden vielmehr die speziellen Potentialverschiebeschaltungen entwickelt. Sie beruhen meist auf *zwei Prinzipien*:

- direkte Verlagerung mit *Spannungsquellen*,
- Verwendung von *Spannungsteilern* mit unterschiedlichen Widerständen für Gleich- und Wechselstrom.

Das Grundprinzip der ersten Variante zeigt Bild 7.3.17a). Als *Spannungsquellen* kommen vor allem

- Dioden in Durchlaßrichtung,
- Z-Dioden,
- einstellbare Spannungen, z.B. gemäß Bild 7.3.5a),

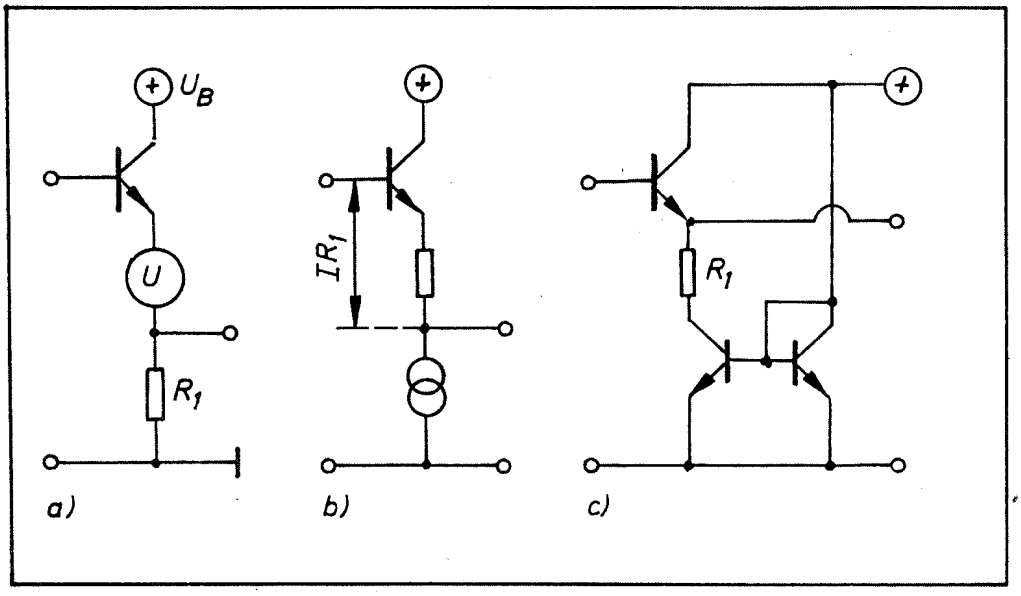


Bild 7.3.17. Potentialverschiebestufen

a) mit Spannungsquelle,
b) mit Stromquelle
c) Beispiel mit Stromspiegel

in Betracht. Bei diesen Schaltungen wirkt sich immer die Substratkapazität der Dioden bzw. des Transistors nachteilig aus. Der Wechselspannungsverlust ist etwa durch $R_{iu}:R_1$ gegeben. Der Innenwiderstand R_{iu} der Spannungsquelle sollte also klein sein.

Das *zweite Prinzip* zeigt Bild 7.3.17b). Mit der Wahl der Stromquelle und des Widerstandes R_1 kann die Potentialverschiebung in weiten Grenzen gewählt werden. Der Wechselspannungsverlust beträgt hier $R_{ii}/(R_1 + R_{ii})$. Der Stromquelleninnenwiderstand R_{ii} sollte also möglichst groß sein. Das Beispiel einer derartigen Stromverschiebeschaltung mit Stromspiegel zeigt Bild 7.3.17c).

7.3.4.4. Spannungsquellen

Noch mehr als bei den Potentialversatzschaltungen ist eine gegenüber Alterungen, Temperaturänderungen und Betriebsspannungsschwankungen stabilisierte Gleichspannungsquelle bei Normalen notwendig, also z.B. bei integrierten Schaltungen zur Strom- bzw. Spannungsstabilisation (7.6.2.) bzw. AD-Wandlern (9.2.). Hier seien nur einige Fakten bezüglich der *Temperaturstabilisation* behandelt.

Als Grundlage von stabilisierten Quellen dienen die *Basis-Emitter-Strecken* normaler Transistoren. Hiervon kann verwendet werden

- die Durchlaßrichtung mit ca. 0,7 V und einem Temperaturkoeffizienten von ca. −2 mV/K,
- die Sperrichtung mit einer Durchbruchspannung von 5,5 bis 7 V je nach Dotierung und einem Temperaturkoeffizienten von ca. +3 mV/K.

Dies legt es nahe, beide Methoden in *Reihenschaltung* gemäß Bild 7.3.18a) zu verwenden. Dann entsteht ein Temperaturkoeffizient von näherungsweise Null. Er ist vor allem deswegen nicht genau Null, weil für die Durchbruchspannung zwei Effekte mit unterschiedlichen Temperaturkoeffizienten zusammenwirken (3.2.6.). Dadurch hängt der Restwert von der gerade erreichten Diffusion ab.

Anwendungsbeispiele dieses Prinzips aus zwei integrierten Schaltungen zeigen die Teilbilder 7.3.18b) und c). In Teilbild b) stellen die Transistoren T_1 bis T_3 einen Stromspiegel gemäß Bild 7.3.16d) dar. Der Transistor T_4 dient zum getrennten Auskoppeln der zwei Spannungen U_{A1} und U_{A2}. Der Widerstand R_1 ist extrem hochohmig. Der durch ihn fließende Strom kann daher vernachlässigt werden. Er sorgt nur beim Einschalten für einen Startstrom des Stromspiegels. Den eigentlichen Strom für den Stromspiegel erzeugt R_2.

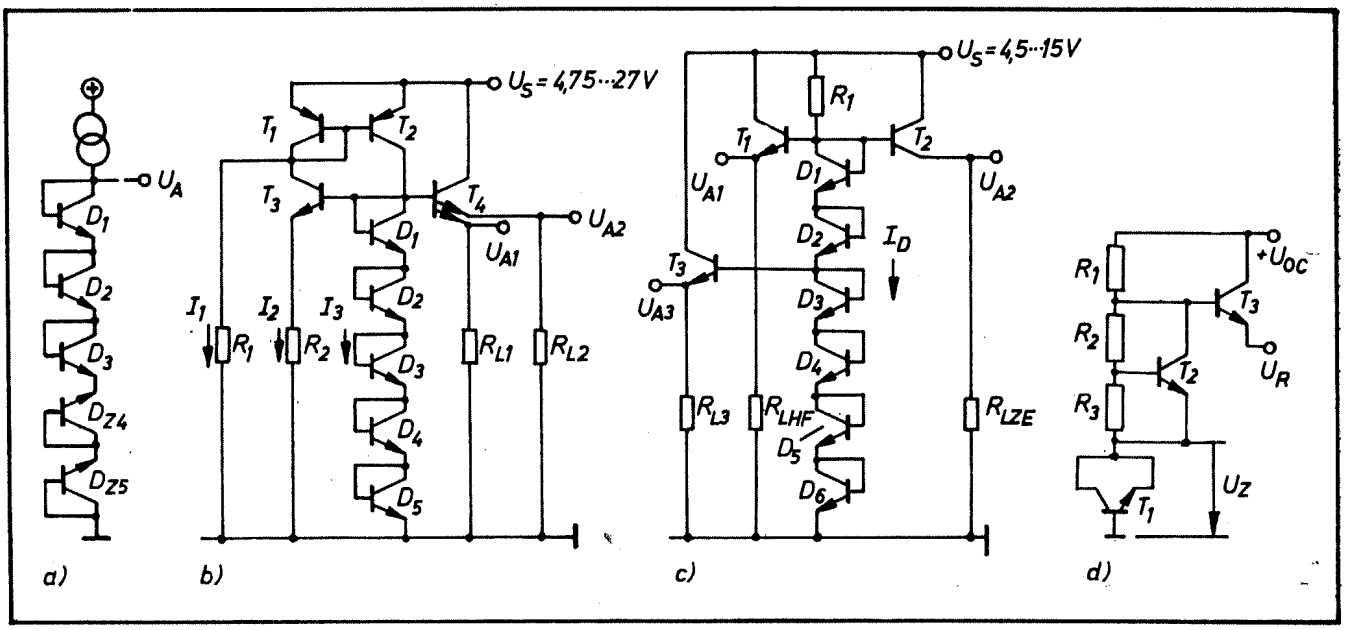


Bild 7.3.18. Spezielle Spannungsquellen

a) Über eine Stromquelle versorgte Reihenschaltung von drei in Durchlaß- und zwei in Sperrichtung betriebenen, als
b) Dioden geschalteten Transistoren
c) Anwendung des Prinzipschaltbildes von Teilbild a) in der Spannungsstabilisatorschaltung des A301D
d) Vereinfachte Spannungsquelle aus sechs als Dioden geschalteten Transistoren in Durchlaßrichtung mit Auskopplung von drei Spannungen beim A244D
Anwendung von einer „Z-Diode" T_1 und der Spannungsschaltung von Bild 7.3.5a) für Transistor T_2 und Auskopplung durch T_3

Die Schaltung von Bild 7.3.18c) verwendet keinen Stromspiegel, sondern nur den Widerstand R_1. Außerdem sind hier alle Dioden D_1 bis D_6 in Durchlaßrichtung betrieben. Die Nutzspannung liegt also bei $5\times$ bzw. $3\times U_{BE}$, ca. 3,5 und 2 Volt, und wird über die Transistoren T_1 bis T_3 ausgekoppelt.

Ein weiteres Prinzip mit Kompensation des Temperaturkoeffizienten zeigt Bild 7.3.18d). Der in Durchbruch als Diode betriebene T_1 hat einen Temperaturkoeffizienten von wiederum $+3$ mV/K. Der Transistor T_2 wird in der Schaltung gemäß Bild 7.3.5a) betrieben. Mit den Widerständen R_3 und R_2 wird nicht nur die Spannung, sondern auch der Temperaturkoeffizient geändert. Er ist einschließlich des Auskopplungstransistors T_3 ca. Null, wenn gilt $R_2 \approx 1{,}5R_3$. Mit einer aufwendigeren Schaltung, deren Grundlage von WIDLOR stammt, läßt sich u. a. mit einem speziellen Stromspiegel und einem Operationsverstärker eine quasi-temperaturunabhängige Spannung erzeugen. Da sie dann genau dem Bandabstand für $T \to 0$ entspricht, heißt sie auch *Bandabstandsreferenz*. Sie wird im Abschnitt 7.6.4.2. behandelt.

7.3.5. Weitere Stromquellen

Einfache Stromquellen zeigte Bild 7.3.3, Stromquellen, die aus dem Stromspiegel abgeleitet sind, Bild 7.3.16. Darüber hinaus sind noch einige Stromquellen entwickelt worden. Die wichtigsten faßt Bild 7.3.19 zusammen. Eine Kaskadierung der Stromquelle von Bild 7.3.3c) zeigt Bild 7.3.19a). Sie besitzt einen sehr hohen Innenwiderstand und einen großen Anwendungsbereich, nämlich bis zu Spannungen von mehr als hundert Volt. Der minimale Spannungsabfall liegt bei etwa $2U_Z$. Werden statt der Z-Diode je zwei normale Dioden in Flußrichtung verwendet, läßt er sich auf etwa 2 V reduzieren. Die analoge Schaltung mit FET zeigt Bild 7.3.19b). Es stellt eine Verdopplung der Schaltung von Bild 7.3.3e) dar.

Eine völlig andere Stromquelle zeigt Bild 7.3.19c). Hier wird die Reihenschaltung einer Z-Diode und einer Diode in Durchlaßrichtung zur Stabilisierung der Emitter-

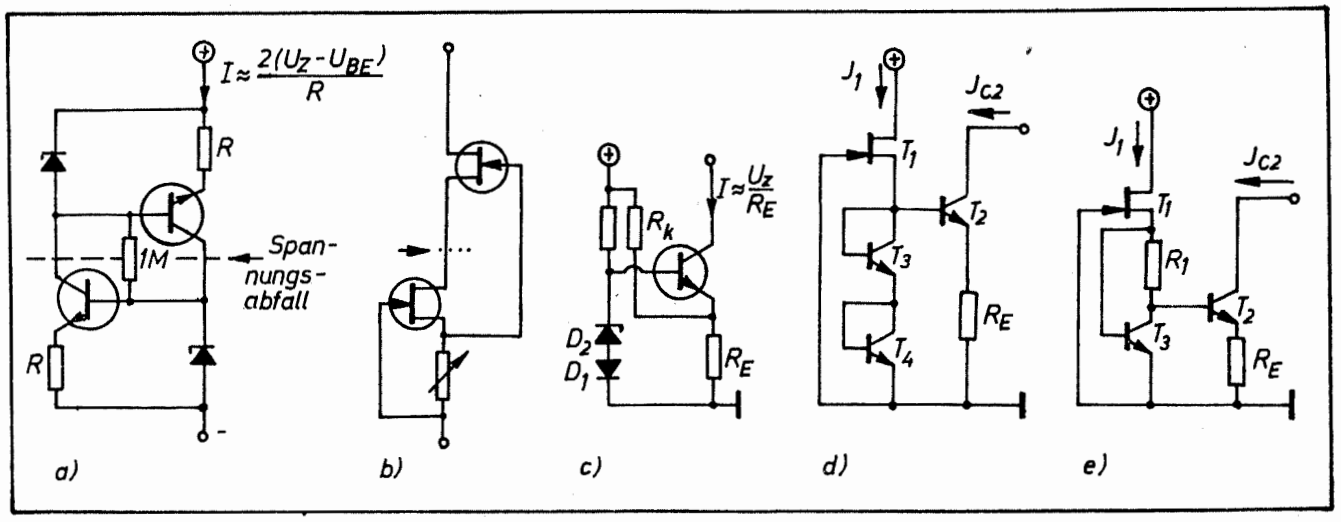


Bild 7.3.19. Weitere Stromquellen

a) Reihenschaltung von zwei Schaltungen gemäß Bild 7.3.3c), wodurch ein sehr hoher Innenwiderstand entsteht
b) Ähnliche Schaltung wie in a), jedoch mit FET
c) Stromquelle über eine Konstantspannungsquelle aus den beiden Dioden
d) Anwendung des FET als Stromquelle für die Spannungsquelle aus den beiden, als Dioden geschalteten Transistoren T_3 und T_4, die den Transistor T_2 als Konstantstromquelle steuern
e) Prinzip wie in d), mit dem Transistor T_3 als über R_1 einstellbarer Stromquelle

Basis-Spannung eines Transistors verwendet. Der erreichbare Strom ist dann näherungsweise durch U_Z und R_E gegeben.

Schließlich sind in Bild 7.3.19d) und e) noch zwei Stromquellen gezeigt, welche den hohen dynamischen Innenwiderstand eines FET nutzen, um damit durch die Dioden bzw. einen als Spannungsquelle geschalteten Transistor einen relativ konstanten Strom zu treiben, der seinerseits eine recht stabile Spannung der Spannungsquelle realisiert. Bei den Schaltungen gemäß Teilbild c) bis e) handelt es sich also um eine Kaskadierung von Strom- und Spannungsquellen zur Erzeugung konstanter Ströme.

7.3.6. Transmissionsgatter, Multiplexer, Sample-hold

In verschiedenen Anwendungen müssen *Analogsignale gemultiplext* werden. Sie müssen von einer oder mehreren Quellen auf einen oder mehrere Ausgänge umgeschaltet werden. Für diese Aufgabe sind FET aus mehreren Gründen *besonders geeignet*:

- Im Ein- (On-) Zustand gehen ihre Kennlinien durch Null, der Widerstand ist relativ niederohmig, und es können positive und negative Spannungen verzerrungsfrei übertragen werden (vgl. Bild 5.1.2).
- Im Aus- (Off-) Zustand sind sie in beiden Richtungen sehr hochohmig und besitzen eine kleine Drain-Source-Kapazität, sperren also die anliegenden Spannungen sehr gut.

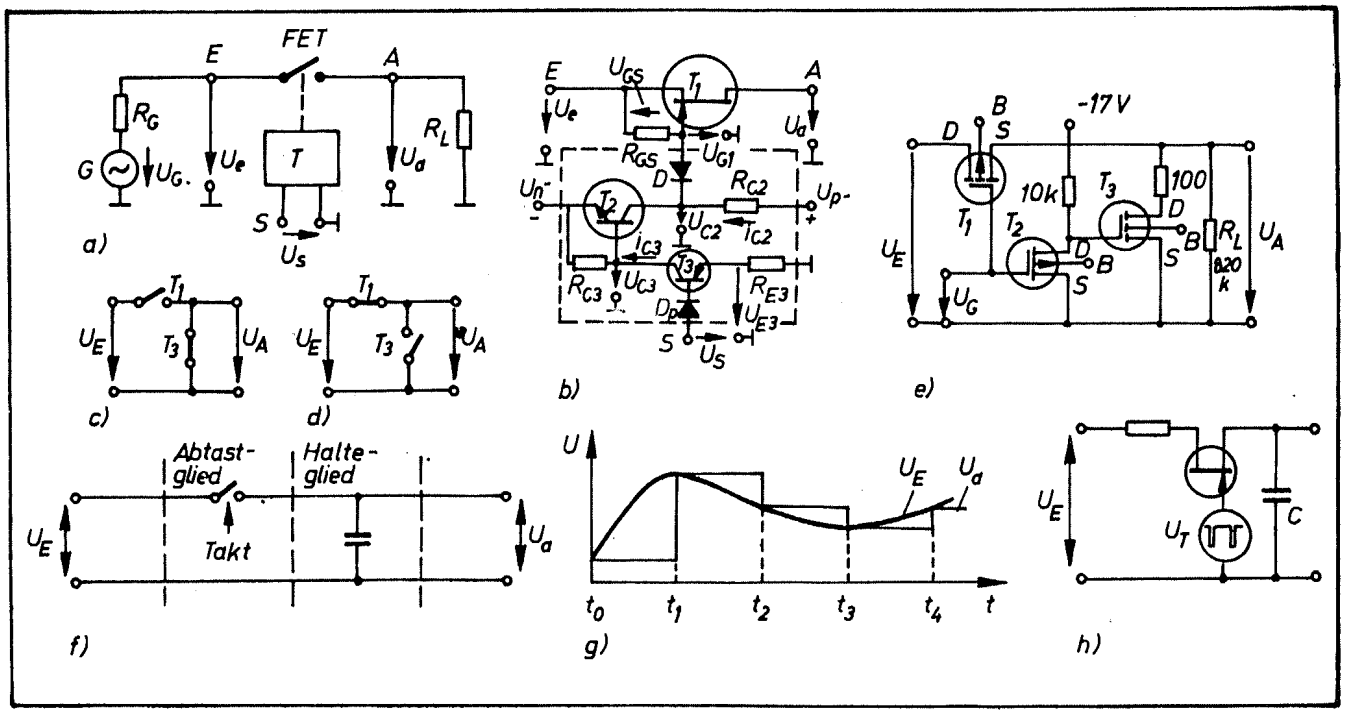


Bild 7.3.20. FET als Multiplexer und Sample-hold

a) Prinzipaufbau eines einzelnen Gatters
b) Aufbau eines Gatters mit FET und Ansteuerschaltung T
c) Ersatzschaltung eines Serien-Parallel-Schalters im Off-Zustand
d) Dito im On-Zustand
e) Prinzip eines Serien-Parallel-Schalters mit den beiden FET T_1 und T_3, während T_2 als Umkehrstufe den FET T_3 entgegengesetzt zu T_1 schaltet
f) Prinzip von sample (Abtastglied) und hold (Halteglied)
g) Spannungsverlauf am Eingang U_E und Ausgang U_A eines idealen Sample-hold
h) Prinzip des Sample-hold mit FET. Der Kondensator wird meist als MILLER-Kapazität in einer folgenden Verstärkerschaltung realisiert

Um diese beiden Zustände stabil zu erreichen, muß zu dem FET eine passende *Ansteuerschaltung* (T in Bild 7.3.20a)) vorgesehen werden. Beispielhaft zeigt eine solche Schaltung Bild 7.3.20b).

Andere Multiplexschaltungen werden z.B. mit *Diodenbrücken*, ähnlich dem Ringmodulator in Bild 5.4.13d), realisiert. Ω kann dann z.B. als Eingangsspannung und ω als Steuerspannung mittels einer Ansteuerschaltung realisiert werden. Einen besonders hohen Störabstand erreicht man mit zwei FET, die als Parallel-Serienschalter arbeiten. Dieses Prinzip ist im Bild 7.3.20c) und d) dargestellt. In der FET-realisierten Schaltung von Teilbild e) ist T_1 der Serien- und T_3 der Reihenschalter, während T_2 das letzte Glied der Ansteuerschaltung darstellt.

Für verschiedene Anwendungen ist es wichtig, aus einem kontinuierlichen Signal zu definierten Taktzeiten t_0, t_i bis t_n Proben zu entnehmen und während der Zwischenzeit zu speichern. Diese Problematik wird später noch beim AD-Wandler (9.2.) genauer begründet. Sie ist im Prinzip aber bereits bei verschiedenen Modulationen, z.B. im Zusammenhang mit den Bildern 7.2.4 bis 7.2.6, zu erkennen. Das Prinzip einer solchen Schaltung zeigt Bild 7.3.20f). Zu den Taktzeiten wird der Schalter (das *Abtastglied*, engl.: sample = Probe) geschlossen, und der jeweilige am Eingang liegende Spannungswert lädt den Kondensator auf diesen Wert. Dieser hält (engl.: hold) den abgetasteten Wert zwischen den Abtastungen. Aus der in Bild 7.3.20g) dick gezeichneten Linie der Eingangsspannung U_E entsteht dann die dünn gezeichnete Stufenkurve der Ausgangsspannung U_A. Den konkreten Aufbau mit einem FET zeigt Bild 7.3.20h), wobei der Kondensator meist in einem nachfolgenden Verstärker als MILLER-Kapazität (7.1.4.1.) aufgebaut wird. Dadurch wird die sonst hohe Belastung des Haltegliedes reduziert, so daß Verluste durch den Restwiderstand des eingeschalteten FET weitgehend vermieden werden.

7.3.7. Oszillatorschaltungen

Wie im Abschnitt 7.1.2.2. gezeigt wurde, können mit *Verstärkerschaltungen* unter zwei Bedingungen *Schwingungen erzeugt* werden:

- Die Verstärkung muß mindestens so groß wie die Dämpfung des Rückkopplungsvierpols sein.
- Es muß eine Frequenz bzw. einen Frequenzbereich geben, in dem die Phase der rückgekoppelten Spannung phasenrichtig erfolgt. In der Ortskurve muß der Punkt -1 richtig umlaufen werden.

Das Prinzip der Schwingungserzeugung — jedoch nicht ihre konkreten Schaltungen — würden vor allem im Zusammenhang mit dem Bild 7.1.7 für den *RC*-Generator und bezüglich Bild 7.1.8 für den Phasenschieber deutlich.

Eine andere *Variante der Schwingungserzeugung* beruht auf einem

- negativen Innenwiderstand.

Hierbei erfolgt eine Entdämpfung gemäß Abschnitt 7.1.1.3. Eine Schwingungserzeugung entsteht dabei, wenn der negative Innenwiderstand größer als die äußeren Verlustwiderstände sind. Formal kann dieses Prinzip der Schwingungserzeugung analog einem rückgekoppelten Verstärker behandelt werden.

Für die praktische Ausführung sind vor allem *drei Typen von Oszillatoren* zu unterscheiden:

- Sinusoszillatoren
- Rechteckgeneratoren
- Dreieck- bzw. Sägezahngeneratoren.

Ergänzend zu den Rechteckgeneratoren existieren noch verschiedene

- Impulsschaltungen.

Ein Teil von ihnen bildet die

- Multivibratorfamilie.

7.3.7.1. Sinusoszillatoren

Für die Sinusoszillatoren ist die *Einhaltung* einer definierten *Schwingungsamplitude* entscheidend. Dies kann auf zwei Arten erfolgen:

- Sättigungsbegrenzung
- Amplitudenregelung.

Bei der *Sättigungsbegrenzung* wird ausgenutzt, daß alle Verstärker (und negativen Innenwiderstände) eine *natürliche Amplitudengrenze* besitzen. Sie wird deutlich aus zwei Fakten:

- den nichtlinearen Kennlinien zwischen Ausgangs- und Eingangsspannung sowie deren Beeinflussung durch Rückkopplung (z.B. s. Bild 7.1.6),
- der Abhängigkeit der Verstärkung, insbesondere der Steilheit beim FET (Bilder 3.3.8a) und 5.5.12e)) und der Stromverstärkung beim Bipolartransistor (Bilder 3.4.6c), 5.6.3c) u. 5.7.4c), d)).

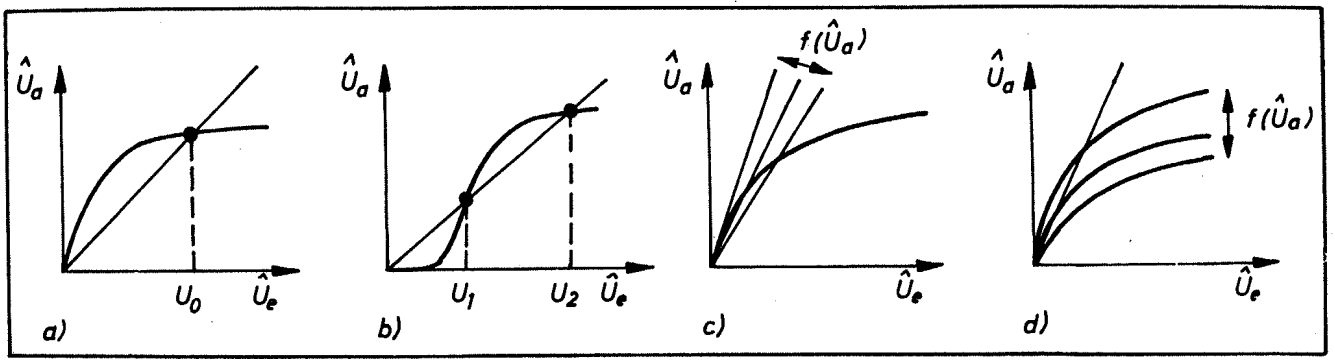


Bild 7.3.21. Zur Amplitudeneinstellung bei Oszillatoren. Die Geraden gelten für den Rückkopplungsvierpol, die gekrümmten Kennlinien für den Verstärker

a) Normales Aussehen bei Sättigungsbegrenzung. Für $u_e < u_0$ erfolgt das Einschwingen auf die Amplitude u_0. Falls $u_e > u_0$ wird, ist die Verstärkung zu klein, und die Amplitude sinkt wieder auf u_0

b) Verstärkerkennlinie mit für Oszillatoren ungünstigem Verlauf. Unterhalb von u_1 kann keine Schwingungsanfachung auftreten

c) Aussehen der Kennlinie bei künstlicher Amplitudenbegrenzung mittels Dämpfungssteuerung des Rückkopplungsvierpols. Es lassen sich so kleine Amplituden erzeugen, daß die Verstärkerkennlinie noch als linear betrachtet werden kann

d) Ähnliches Verhalten wie bei c), jedoch durch Steuerung der Verstärkung bewirkt

Im allgemeinen kann der Rückkopplungsvierpol dagegen als linear angenommen werden. Dadurch ergibt sich für die Eingangs- und Ausgangsamplituden ein Diagramm, wie es Bild 7.3.21a) zeigt. Für Amplituden, die kleiner als u_0 sind, ist dabei $kv > 1$, und die Amplitude der Schwingung wächst an. Für Amplituden größer als u_0 gilt dagegen $kv < 1$, und die Amplitude fällt ab. Es stellt sich also infolge der Nichtlinearität die Amplitude u_0 bzw. der zugehörige Ausgangswert automatisch ein.

Der *Nachteil dieses Prinzips* liegt einmal darin, daß die Verstärkerkennlinien bis in den nichtlinearen Teil ausgesteuert werden müssen. Es entstehen starke *Oberwellenanteile*, die nur durch eine hochselektive Rückkopplungsschaltung reduziert werden können. Deshalb ist dieses Prinzip praktisch nur ab Schwingkreisgüten größer 10 verwendbar.

Ein zweiter Nachteil dieses Prinzips kann eventuell in *schlechtem Anschwingen* in Erscheinung treten. Dies sei am Beispiel von Bild 7.3.21b) erklärt. Für Amplituden kleiner als u_1 gilt $kv > 1$. Die Schaltung kann, solange keine größere Steuerspannung, z.B. durch Rauschen oder andere Einflüsse, auftritt, nicht schwingen. Erst wenn u_1 überschritten wird, gilt $kv < 1$. Von u_1 an erfolgt die Amplitudeneinstellung wie bei Teilbild a). Eine Kennlinie wie in Bild 7.3.21b) kann nur dann auftreten, wenn der Arbeitspunkt so gewählt wurde, daß bei ihm nicht die größte Steilheit bzw. Verstärkung vorliegt. Da aber bei Oszillatorschaltungen für Schwingamplituden Null oft kein Wert bewußt günstig festgelegt wird, treten solche ungünstigen Fälle durchaus auf.

Die beiden oben genannten Nachteile sind mit einer *Amplitudenregelung* zu vermeiden. Hierbei wird eine Regelspannung abgeleitet, die der Amplitude der Oszillatorspannung proportional ist. Mit dieser Spannung kann, z.B. über einen FET-gesteuerten Spannungsteiler, die Dämpfung (Neigung in Bild 7.3.21c)) des Rückkopplungsvierpols verstellt werden. Es kann aber auch der Verstärkungsfaktor verändert werden, z.B. mittels einer Schaltung analog zu Bild 7.3.9b). Dieses Prinzip zeigt Bild 7.3.21d).

Mit beiden Varianten kann einmal erreicht werden, daß die Schwingamplitude so klein bleibt, daß nur der lineare Teil der Verstärkerkennlinien benutzt wird, also nur

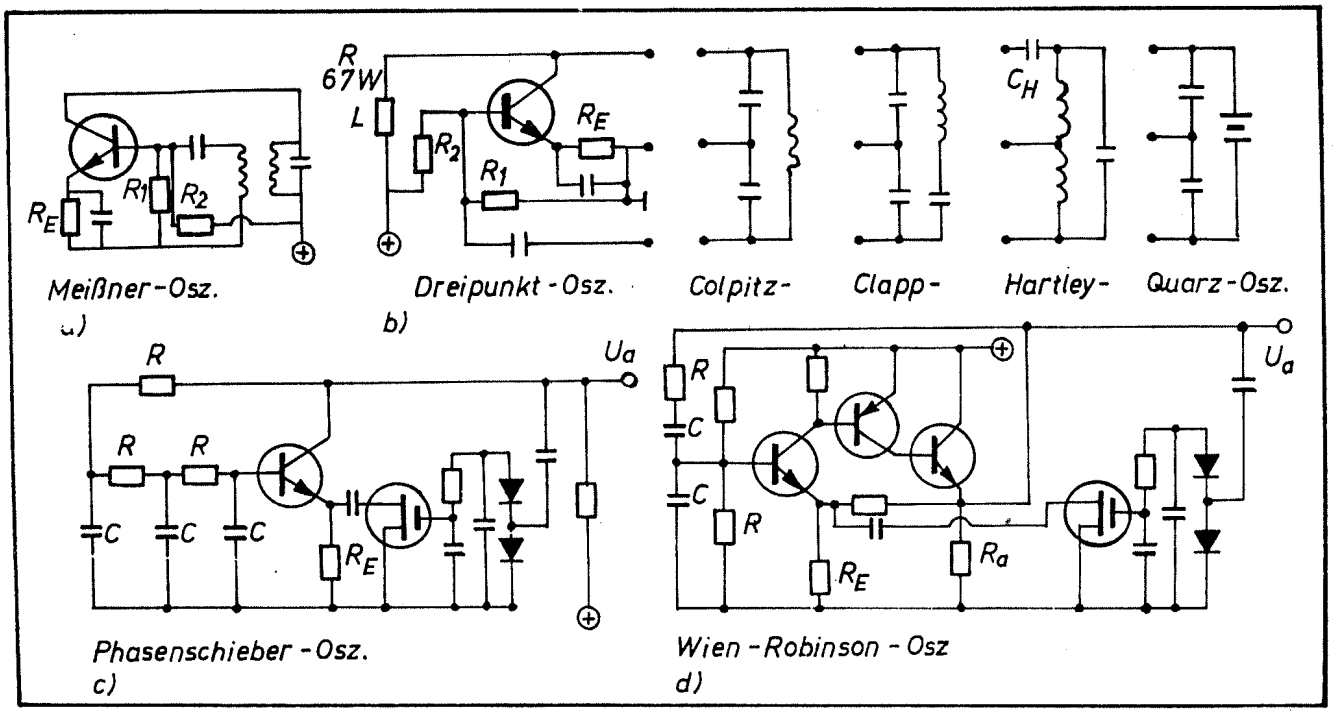


Bild 7.3.22. Beispielschaltungen für Sinusoszillatoren, die Namen beziehen sich auf Entwickler die erstmals diese Schaltung verwendeten. Die Widerstände R_1 und R_2 sowie R_E legen einen günstigen Arbeitspunkt fest

a) MEISSNER-Schaltung, gekennzeichnet durch die induktive Rückkopplung

b) Prinzip der allgemeinen Dreipunktschaltung mit den Varianten
Colpitz: Spannungsteilung durch zwei C, die zugleich die Kapazität des Parallelschwingkreises bilden;
Clapp: Spannungsteilung mit zwei C, jedoch dazu parallel liegendem Serienkreis;
Hartley: mit induktiver Spannungsteilung
Das C_H verhindert Gleichstromkurzschluß. Bei dieser Schaltung wird der Mittelabgriff der Spule oft auf die Betriebsspannung gelegt, dann sind die Spannungen anders zu trennen
Quarz mit kapazitiver Spannungsteilung

c) Phasenschieberoszillator und

d) WIEN-ROBINSON-Oszillator. Beide verwenden einen FET parallel zum Emitterwiderstand R_E und stellen dadurch die Verstärkung gemäß Bild 7.3.21d) auf den richtigen Wert ein. Die mit R und C bezeichneten Bauelemente bestimmen die Frequenz der Oszillatoren

sehr geringe Obertonkomponenten auftreten. In diesem Fall lassen sich auch oberwellenarme Sinusschwingungen mit *RC*-Gliedern erreichen. Zum anderen gestattet dieses Prinzip, relativ leicht ein gutes Anschwingen und eine konstante Schwingamplitude zu realisieren.

Im Laufe der Entwicklung sind eine Vielzahl von verschieden aufgebauten Oszillatorschaltungen entstanden. Einen Überblick hierzu gibt Bild 7.3.22.

7.3.7.2. Rechteckgeneratoren

Beim Rechteckgenerator wird genau das Gegenteil wie vom Sinusgenerator gefordert, nämlich die *Ausbildung möglichst vieler Oberwellen*, und zwar so, daß eine möglichst gute Rechteckspannung entsteht. Er darf daher weder eine indirekte Amplitudenbegrenzung noch selektive Bauelemente besitzen. Es muß also eine möglichst breitbandige Rückkopplung und eine natürliche Amplitudenbegrenzung vorgesehen werden. So entsteht die Schaltung, wie sie Bild 7.3.23a) zeigt. In die Schaltung sind die Schwingungsformen an verschiedenen Punkten eingezeichnet. Dies hat den Vorteil einer leichteren Diskussion der Funktion dieser Schaltung. Da in ihr nur sehr bedingt selektive Bauelemente auftreten, bestimmen vielmehr Zeitkonstanten das Verhalten. Sie entstehen durch die *Umladung der Kondensatoren* C_1 und C_2. Diese Umladung wird durch die Widerstände R_{a1}, R_{iT1}, R_{eT2}, R_2 bzw. R_{a2}, R_{iT2}, R_{eT1}, R_1 bewirkt.

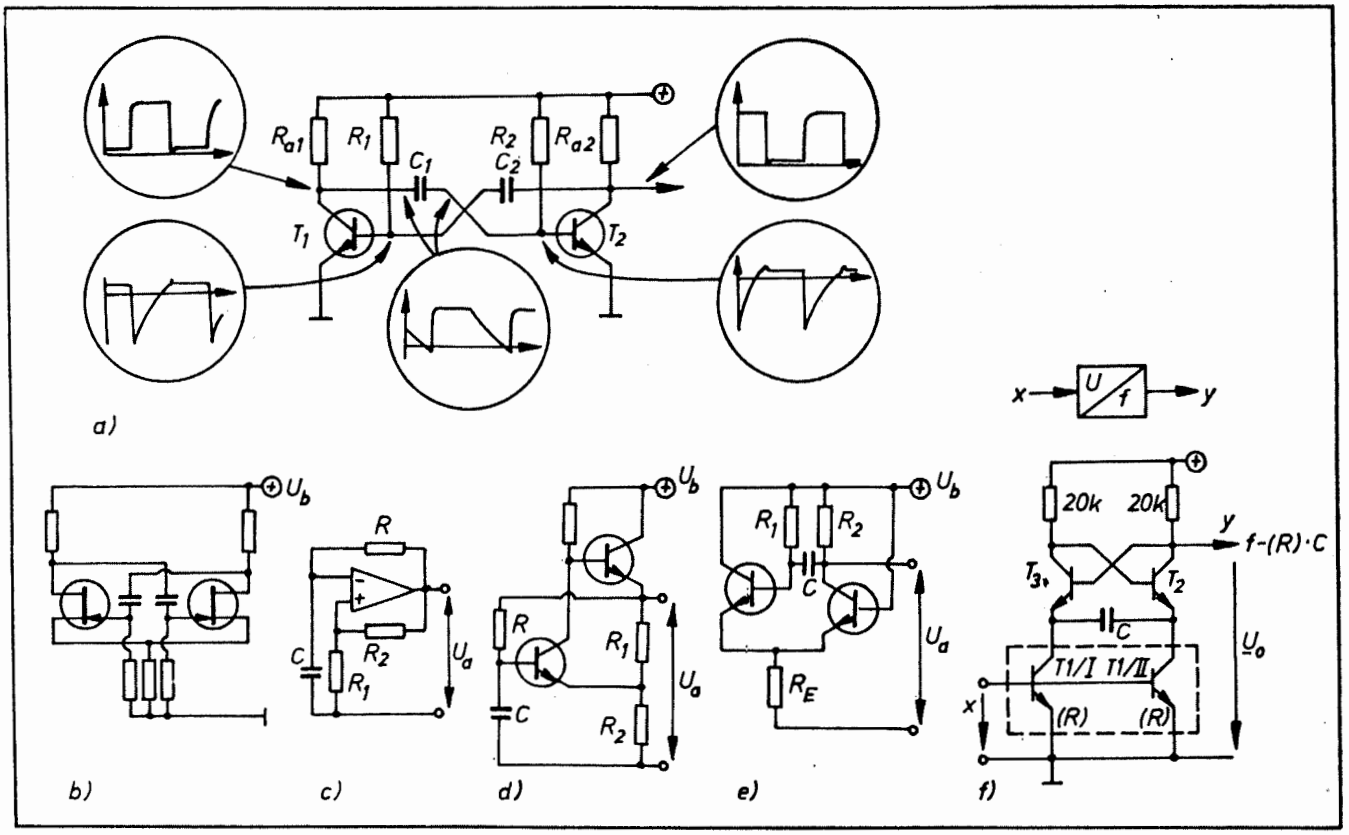


Bild 7.3.23. Multivibratoren bzw. Rechteckgeneratoren

a) Prinzipschaltung mit angezeigten Spannungsverläufen
b) Prinzipschaltung für FET
c) Schaltung mit einem Differenzverstärker (Operationsverstärker), die nur eine Kapazität C benötigt und symmetrische Rechteckimpulse erzeugt
d) Ähnliche Schaltung wie in c), jedoch mit zwei Transistoren
e) Schaltung in Emitterkopplung mit sehr unsymmetrischen Impulsen. Die Zeitkonstante bez. R_E ist extrem kurz. Die lange Impulszeit und damit die Frequenz wird durch C und die zugehörigen Ladewiderstände bestimmt
f) Schaltung eines VCO (voltage controlled oszillator) bzw. Rechteckgenerators, dessen Frequenz linear von der Spannung U_{ref} zur Steuerung der beiden Stromquellen abhängt

Es sei angenommen, daß T_1 gerade gesperrt und T_2 leitend sei. Die Sperrung von T_1 ist dann erfolgt, weil der Kollektor von T_2 auf eine negativere Spannung überging und dadurch, über C_2 gekoppelt, die Basis von T_1 negative Werte angenommen hat. Der Kondensator C_2 entlädt sich jetzt exponentiell über R_1 und R_{a2}, und zwar solange, bis ein geringer Basis-Emitter-Strom fließt. Er ruft einen mit h_{21} verstärkten Kollektorstrom in T_1 hervor. Dadurch entsteht ein Spannungsabfall an R_{a1}, der über C_1 die Basis von T_2 auf negative Werte zieht. Der Basisstrom von T_2 wird kleiner, um h_{21} der Kollektorstrom; damit erhöht sich das Kollektorpotential von T_2 und über C_2 wird der Basisstrom von T_1 weiter erhöht. Infolge der sehr starken Mitkopplung geht fast schlagartig T_1 in den leitenden und T_2 in den gesperrten Zustand über. Die Folge ist, daß die Kollektorpotentiale beider Transistoren fast idealen, rechteckförmigen Verlauf zeigen, die Basispotentiale dagegen einen z. T. exponentiellen Verlauf. Mit der Zeitkonstante dieses exponentiellen Verlaufs hängen die Schaltzeiten in beiden Richtungen zusammen. Sie können durch unterschiedliche Wahl von C_1 und C_2 unterschiedlich lang gemacht werden, so daß dann unsymmetrische Rechteckspannungen entstehen.

Es gibt eine Vielzahl von *Varianten* der Rechteckgeneratoren. Sie unterscheiden sich vor allem durch die Rückkopplungszweige und die verwendeten Transistoren. Einige Beispiele zeigt Bild 7.3.23b) bis e). Teilbild b) ist die zu a) analoge Schaltung mit FET, c) verwendet einen Differenzverstärker. Infolge der Mitkopplung über R_1 und R_2 kann sein Ausgang nur positive oder negative Werte annehmen. Die nächste Umschaltung erfolgt dann, wenn über $R\,C$ soweit umgeladen ist, daß wieder der Strom in der entgegengesetzten Richtung fließen kann. Da dies für beide Polungen gleich schnell erfolgt, gilt für beide Zeitkonstanten

$$T_1 = T_2 = RC \ln\left(1 + \frac{2R_1}{R_2}\right). \tag{49}$$

Ein ganz analoges Prinzip, das ebenfalls nur mit einem Kondensator auskommt, zeigt Bild 7.3.23d). Das Teilbild e) enthält gleichfalls nur einen Kondensator. Hier sind die Verhältnisse jedoch anders. Eine der Zeitkonstanten wird nämlich mittels der Emitterkopplung durch R_E bewirkt. Dazu gehört eine Zeitkonstante, die durch Schalt- und innere Kapazitäten gebildet wird und sehr gering ist. Die andere Zeitkonstante ist dagegen durch C und die zugehörigen Ladewiderstände bestimmt. Sie läßt sich relativ groß gestalten, so daß diese Schaltung eine extrem unsymmetrische Rechteckschwingung erzeugt.

Die *Rechteckgeneratoren haben Eigenschaften*, die sie ganz wesentlich von den Sinusoszillatoren unterscheiden:

- Es werden Rechteckschwingungen erzeugt, die sich leicht auch mit unterschiedlichen Zeitkonstanten für beide „Halbwellen" realisieren lassen.
- Durch Anlegen von äußeren Impulsen (die z.B. auf Basis bzw. Gate einwirken) sind sie leicht mit diesen Impulsen zu synchronisieren.
- Die Zeitkonstanten lassen sich ändern, wenn die Entladeströme verändert werden.

Der zweite Fakt hat große Bedeutung für die *Synchronisation* von Schwingungen und wird für viele Anwendungen genutzt. Er beruht darauf, daß der jeweils gesperrte Transistor vorfristig leitend gemacht wird. Je nach der Amplitude dieser Impulse sind auch relativ leicht bestimmte Frequenzverhältnisse, z.B. bis 1:10, bei Synchronisation einzuhalten.

Der letzte Effekt wird bei *Spannungs-Frequenzwandlern* genutzt. Diese Schaltungen heißen VCO (voltage controlled oszillator), ein Beispiel zeigt Bild 7.3.23f). Es ist über

die gesteuerten Stromquellen in einem Frequenzbereich bis 1:1000 frequenzlinear, wobei nur Fehler von etwa $1^0/_{00}$ auftreten. VCO-Schaltungen (Abschnitt 7.2.1.2.) können auch als Modulatoren für Frequenzmodulation oder Pulsfrequenzmodulation (Bild 7.2.5, PFM) aufgefaßt werden.

7.3.7.3. Sägezahn- bzw. Dreieckgeneratoren

Der Zusammenhang von Sägezahn- und Dreieckspannungen wurde bereits in den Bildern 1.5.2 und 1.5.4 gezeigt. Prinzipiell können diese zeitlinearen Spannungen *auf zwei Wegen* erzeugt werden:

- durch Integration von Rechteckschwingungen,
- durch Auf- bzw. Entladung von Kondensatoren mit konstantem Strom.

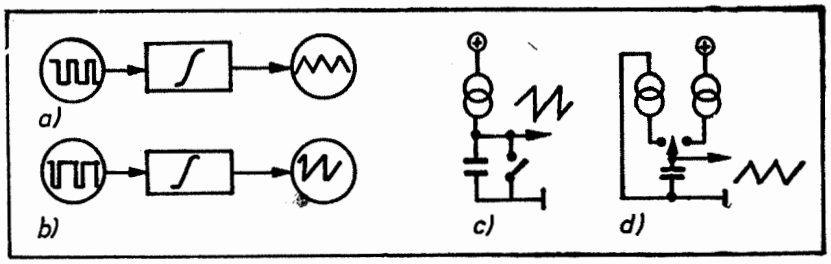


Bild 7.3.24. Dreieck- und Sägezahngeneratoren

a) Dreieck durch Integration aus Rechteckschwingungen
b) Sägezahn aus stark unsymmetrischen Rechteckschwingungen
c) Sägezahnschwingung durch Aufladung eines Kondensators mit Konstantstrom
d) Dreieckschwingung durch Aufladung und Entladung eines Kondensators aus Umschaltung zwischen zwei Stromquellen

Die entsprechenden Prinzipien zeigt Bild 7.3.24. Der Schalter bzw. Umschalter ist von äußeren Quellen zu steuern. Eine Möglichkeit hierzu bietet wieder der Multivibrator. Für die Stromquellen sind alle bisher behandelten Prinzipien anwendbar (u. a. 7.3.1.3., 7.3.4.2. und 7.3.5.).

Mittels des Stromspiegels ist relativ leicht über Steuerung des ersten Stromes ein VCO dann zu realisieren, wenn der Schalter in Bild 7.3.24c) nur durch die jeweils erreichte Spannung (Schwellenwert) ausgelöst wird. Die Frequenz ist dann etwa reziprok zum ersten Strom.

7.3.8. Impulsschaltungen

7.3.8.1. Überblick

Eine Zusammenstellung wichtiger Impulsschaltungen zeigt Bild 7.3.25. Sie zeichnen sich dadurch aus, daß an ihrem *Ausgang* im wesentlichen *nur zwei Signalwerte* angenommen werden können. Ihre Eingangssignale können analoger oder impulsförmiger Art sein bzw. beim astabilen Multivibrator entfallen. Die ersten drei Schaltungen können vor allem durch ihre *Kennlinie* unterschieden werden.

Ist sie eine ideale Rampenfunktion, so liegt der *Komparator* vor; weist sie einen Übergangsbereich Δu auf, so liegt ein *Begrenzer* vor, und besitzt sie schließlich eine Hysterese, so ist die entsprechende Schaltung ein SCHMITT-*Trigger*.

Bei den impulsgesteuerten Schaltungen ist der monostabile und bistabile Multivibrator zu unterscheiden. Der *monostabile Multivibrator* gibt nach jedem Eingangsimpuls einen Ausgangsrechteckimpuls der konstanten Dauer T ab. Er heißt daher

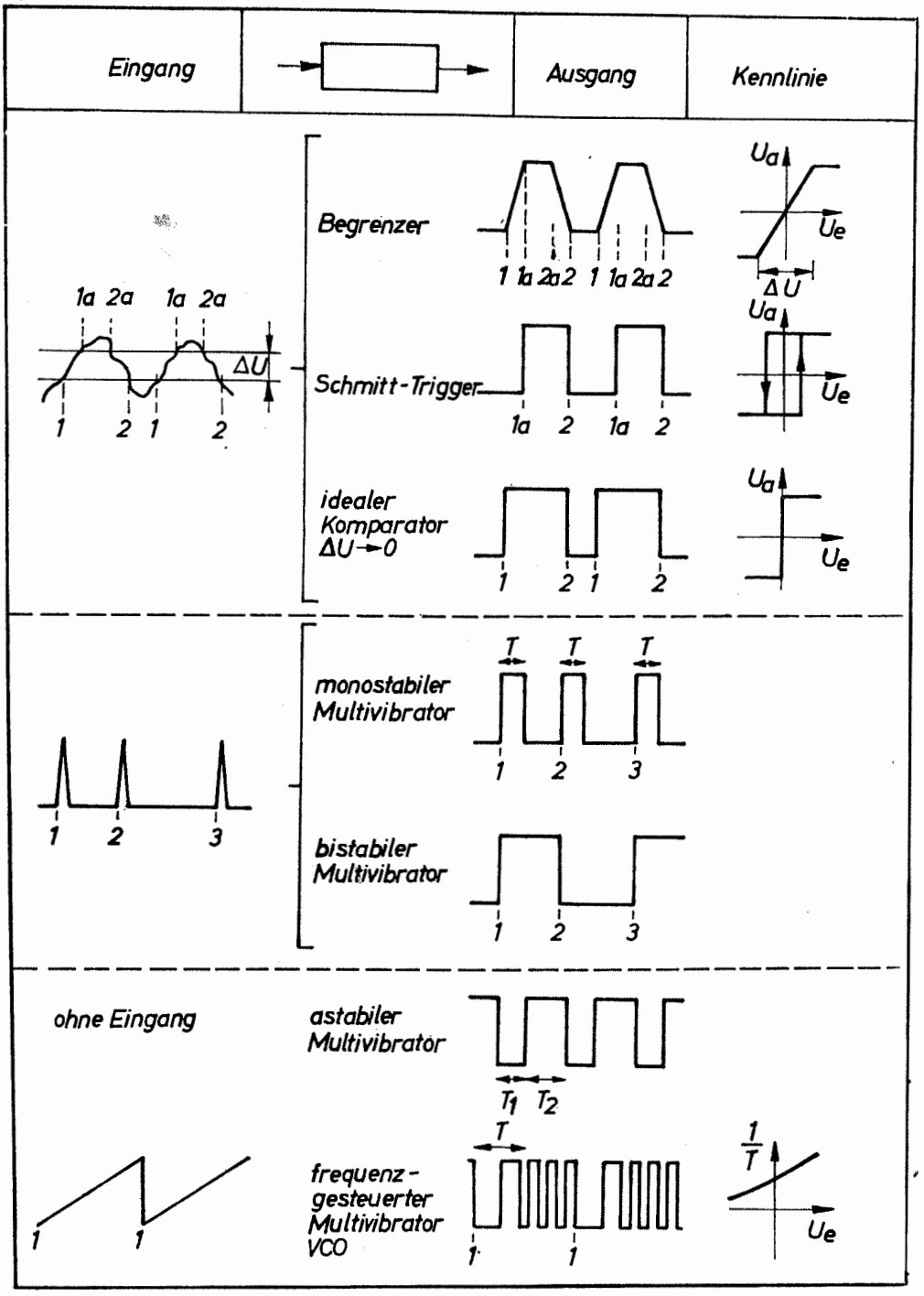


Bild 7.3.25. Überblick zu wichtigen Impulsschaltungen

auch *Univibrator*. Die Zeitdauer T ist durch Dimensionierung der Schaltung in weiten Grenzen wählbar. Der *bistabile Multivibrator* schaltet nach jedem Eingangsimpuls in den anderen Zustand um. Er heißt deshalb auch *Frequenzteiler*, *Flipflop*, *Scale of two*.

Der *astabile Multivibrator* ist ohne Eingangssignal. Er gibt fortlaufend Rechteckimpulse ab. Die Zeitdauern der beiden Impulszustände sind durch die Dimensionierung der Schaltung festgelegt.

Die beiden letzten Schaltungen wurden bereits im Abschnitt 7.3.7.2. behandelt. Der mono- bzw. bistabile Multivibrator unterscheidet sich von dem Rechteckgenerator nur durch die Ausführung der beiden möglichen Mitkopplungen. Ist nur eine für Gleichstrom ausgelegt, so entsteht ein Monoflop = monostabiler Multivibrator. Sind beide für Gleichstrom ausgelegt, so entsteht der bistabile Multivibrator bzw. Flip-

flop. Da seine Eigenschaften wesentlich für die Digitaltechnik sind, wird er dort behandelt (z.B. 8.5.2.).

7.3.8.2. Monostabiler Multivibrator

Schaltungen von monostabilen Multivibratoren zeigt Bild 7.3.26. Insbesondere am Teilbild a) wird der Unterschied zu den astabilen deutlich. Es existiert eine Gleichspannungskopplung $R_1 R_2$. Im Ruhezustand ist der Transistor T_1 leitend. Seine Basis erhält über R_B den notwendigen Basisstrom. Sein Kollektor liegt auf so tiefem Potential, daß über $R_1 R_2$ der Transistor T_2 gesperrt ist. Trifft nun als u_E ein negativer Impuls ein, so wird T_1 gesperrt, sein Kollektorpotential wird so positiv, daß über R_1 und R_2 T_2 leitend wird. Dieser Vorgang wird noch durch ein passend gewähltes C verstärkt.

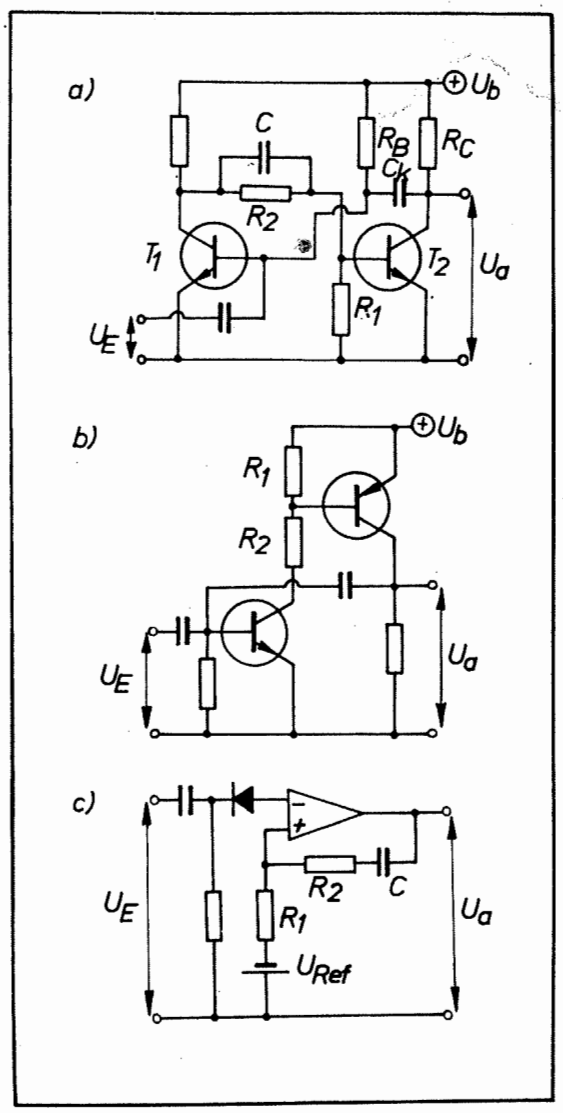


Bild 7.3.26. Schaltungsbeispiele für monostabile Multivibratoren (Uni- bzw. Monoflops)

Da T_2 leitend wird, sinkt sein Kollektorpotential, und über C_k wird so T_1 noch weiter gesperrt, selbst über die Zeit des Impulses bei u_E hinaus. Er wirkt nur auslösend für diesen Vorgang. Nach Ablauf dieses Umschaltvorganges wird C_k über R_C und R_B mit parallelgeschaltetem Basis-Emitter-Widerstand entladen. Sobald dabei der Transistor T_1 wieder etwas leitend wird, stellt sich verstärkt über die Mitkopplung der Schaltung der Ruhezustand ein. Die *Impulszeit* des Monoimpulses beträgt also

$$T \approx C_k \left(R_C + \frac{R_B \cdot R_{BE}}{R_B + R_{BE}} \right). \tag{50}$$

Sehr lange Zeiten können realisiert werden, wenn für T_1 ein selbstleitender FET gewählt wird. Die Auslösung der Schaltung kann in beiden Fällen auch durch positive Impulse an der Basis von T_2 erfolgen. Eine Schaltung des monostabilen Multivibrators mit komplimentären Transistoren zeigt Bild 7.3.26b). Bei ihm sind im Ruhezustand beide Transistoren gesperrt. Er nimmt deshalb nur während der Impulszeit Strom auf. Ausgelöst wird er durch positive Eingangsimpulse.

Einen monostabilen Multivibrator auf der Basis eines Operationsverstärkers zeigt noch Teilbild c). Infolge der negativen Referenzspannung U_{Ref} befindet sich am Ausgang im Ruhezustand die negative Sättigungsspannung U^-. Mit einem größeren negativen Eingangsimpuls kann der Operationsverstärker zur positiven Sättigungsspannung U^+ umgeschaltet werden. (Die Diode verhindert, daß positive Impulse zum Eingang des Operationsverstärkers gelangen.) Dabei wird über C, R_2 und R_1 dieses Umschalten verstärkt und beständig gemacht. Erst wenn sich C hinreichend weit entladen hat, tritt wieder der Ruhezustand ein. Die *Dauer des Monoimpulses* beträgt etwa

$$T = C(R_1 + R_2) \ln \frac{(U^+ - U^-)}{|U_{\mathrm{Ref}}|} \cdot \frac{R_1}{R_1 + R_2}. \tag{51}$$

7.3.8.3. Begrenzer, Komparator, SCHMITT-Trigger

Begrenzer wurden bereits im Abschnitt 5.4.6. im Zusammenhang mit Bild 5.4.9 behandelt. Sie lassen sich jedoch weitaus effektiver mit Verstärkern realisieren, die bis an ihre Aussteuerungsgrenzen betrieben werden. Dann erfolgt, wie in Bild 7.3.25, ein relativ steiler Übergang zwischen den beiden Grenzzuständen. Besonders günstig ist hierfür ein Differenz- bzw. Operationsverstärker. Mittels der Bezugs- bzw. Referenzspannung U_{Ref} kann dann nämlich der Schaltmittelwert in relativ weiten Grenzen gewählt werden (Bild 7.3.27a)). Mit Erhöhung der Verstärkung wird der Schaltbereich ΔU, in dem zwischen den beiden Ausgangspegeln übergegangen wird, immer kleiner. Erst bei unendlich großer Verstärkung wird er theoretisch Null und liefert den idealen Komparator. Praktisch ist dies nicht realisierbar, da dann bereits das

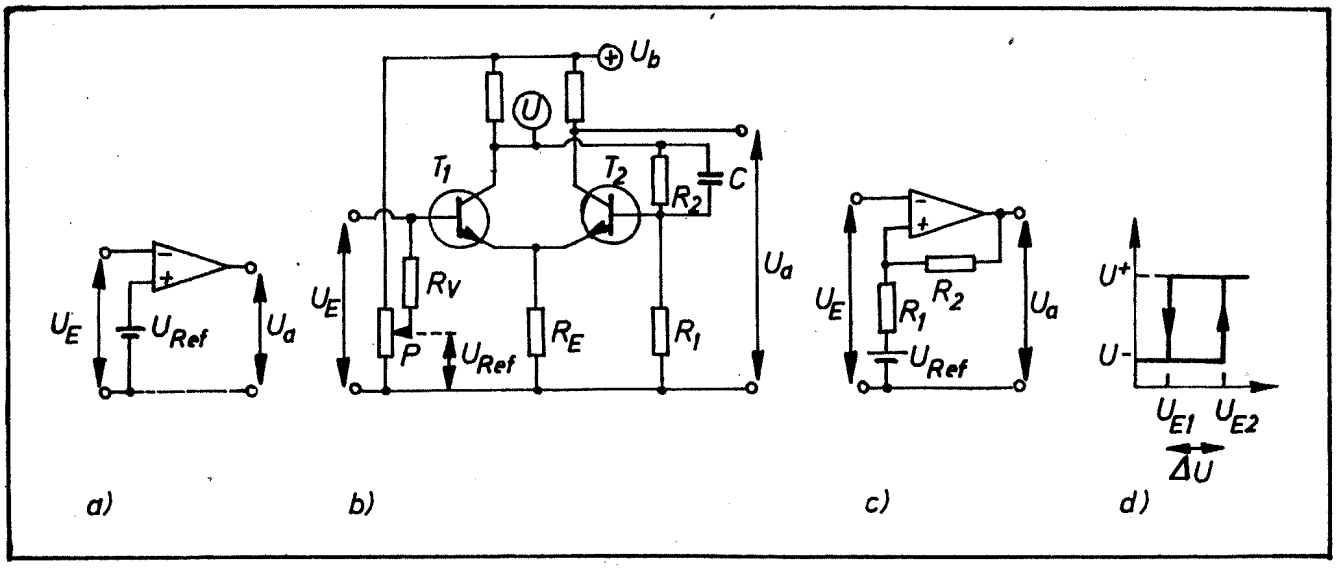


Bild 7.3.27. Signalverarbeitende Impulsschaltungen

a) Begrenzer
b) SCHMITT-Trigger mit emittergekoppelten Transistoren
c) SCHMITT-Trigger mit rückgekoppeltem Differenzverstärker
d) Kennlinie von SCHMITT-Triggern

Eingangsrauschen ein ständiges Umschalten am Ausgang bewirken würde. Einen Ausweg bildet hierbei der SCHMITT-*Trigger*. Er besitzt trotz eines *Schaltspiels* ΔU extrem steile Übergänge. Dies wird durch eine Gleichstromrückkopplung erreicht, die es nur zuläßt, daß einer der beiden Zustände angenommen wird. Dieses bedeutet eine Hysterese gemäß Bild 7.3.27d). Eine hierfür geeignete Schaltung zeigt Bild 7.3.27b). Je nach der Größe des Basisstroms an T_1 ist dieser oder der andere Transistor leitend. Beide Zustände sind jeder für sich stabil. Um von einem Zustand in den anderen zu kommen, muß der Basisstrom so stark geändert werden, daß die Stabilität des vorhandenen Zustandes instabil wird. Die kleinste Tendenz hierzu erzwingt über die dann starke Emitterkopplung das schlagartige Umklappen in den anderen Zustand. Dies ist noch deutlicher an der Schaltung mit dem Differenzverstärker von Bild 7.3.27c) zu erkennen. Der Spannungsteiler $R_1 R_2$ wirkt immer stabilisierend für den jeweils vorhandenen Zustand. Es muß dann am invertierenden Eingang eine Spannung angelegt werden, die gerade die positiv über $R_1 R_2$ zurückgekoppelte Spannung überwiegt. Das Vorzeichen dieser Spannung ist je nach dem gerade vorhandenen Zustand unterschiedlich. Die Schaltspannung ΔU ist also der doppelte Wert der mitgekoppelten Spannung vom Ausgang. Sind U^+ und U^- die Grenzwerte, so gilt für die beiden *Schaltspannungen*

$$\left.\begin{aligned} U_{E1} &= U_{Ref} + (U^+ - U_{Ref}) \frac{R_1}{R_1 + R_2}, \\ U_{E2} &= U_{Ref} + (U^- - U_{Ref}) \frac{R_1}{R_1 + R_2}, \end{aligned}\right\} \tag{52}$$

und daraus folgt

$$\Delta U = \frac{R_1}{R_1 + R_2} (U^+ - U^-) . \tag{53}$$

Damit wirklich eine ausreichende Mitkopplung vorliegt, muß für die *Verstärkung* gelten

$$V > \frac{R_1 + R_2}{R_1} . \tag{54}$$

Je größer die Verstärkung ist, desto kleiner kann die Schaltschwelle ΔU eingestellt werden.

7.3.8.4. Pulslängenmodulator

Mittels eines *Komparators* läßt sich leicht ein Pulslängenmodulator aufbauen. Als *Referenzquelle* wird dazu eine sehr gut lineare *Sägezahnspannung* verwendet. So entsteht die Anordnung von Bild 7.3.28a). Die Konstruktion zur Bildung der Rechteckimpulse zeigt Teilbild b). Sobald die Sägezahnspannung positiver als die Eingangsspannung U_E ist, liegt am Ausgang des Operationsverstärkers die positive Sättigungsspannung, sonst die negative. Die Impulslänge T_2 ist daher linear von U_E abhängig. Mit einer Konstanten K gilt also

$$T_2 = K(U_E + U_0) . \tag{55}$$

Die konstante Spannung U_0 ist notwendig, damit auch bei negativen U_E eine positive Impulszeit T_2 verbleibt. Der mittlere Gleichstromanteil der Rechteckfolge beträgt andererseits

$$U_m = \frac{T_2}{T_1 + T_2} \hat{U} . \tag{56}$$

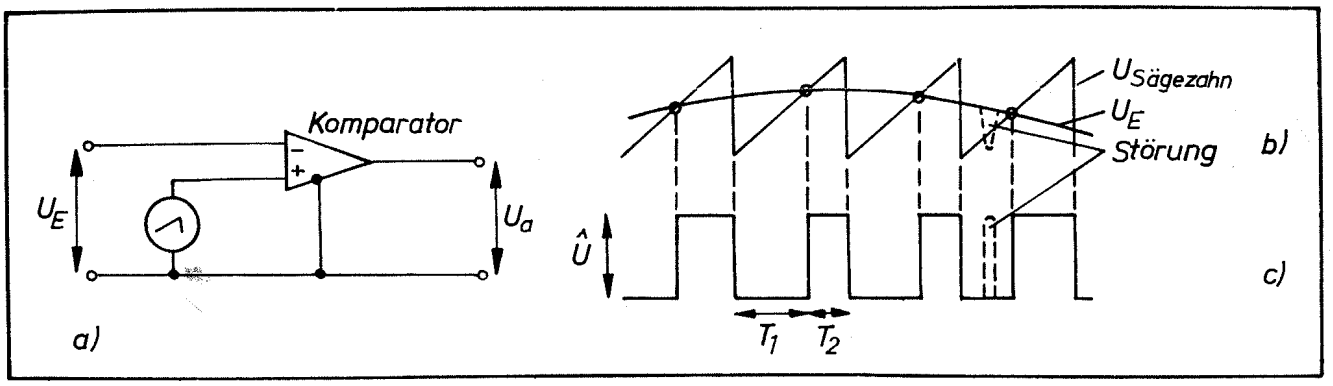


Bild 7.3.28. Prinzip des Pulslängenmodulators

a) Schaltung aus einem Komparator und sehr gut linearem Sägezahngenerator
b) Überlagerung der Sägezahnspannung mit der Eingangsspannung. Die Schnittpunkte schalten den Komparator um und
c) legen so die Impulsfolge fest

Deshalb kann diese Impulsfolge einfach durch einen Tiefpaß demoduliert werden. Es wird aus dem Bild offensichtlich, daß eine Störspannung das Verhalten des Pulslängenmodulators erheblich stören kann. Deshalb wird vor derartigen Modulatoren eine Sample-hold-Schaltung gemäß Abschnitt 7.3.6. vorgeschaltet.

7.4. Ausgangsstufen und Leistungsverstärker

7.4.1. Allgemeine Grundlagen

Die Begriffe Ausgangsstufe und Leistungsverstärker sind nicht eindeutig definiert. Hier geht es um *zwei Arten von Schaltungen*, die z. T. mit sehr *ähnlichen Prinzipien* realisiert werden:

- *Ausgangsstufen* sind vor allem in analogen Schaltkreisen vorhanden. Sie sind deshalb so ausgelegt, daß eine gewisse maximale Leistung, Spannung bzw. Strom abgegeben werden, ohne daß störende Rückwirkungen auf die Funktion der Schaltung auftreten.
- *Leistungsverstärker* stellen in erster Linie gesteuerte Energiewandler dar, die mit Hilfe einer Steuerspannung (bzw. eines Steuerstromes oder Leistung) eine vorhandene Gleichleistung möglichst verlustarm in eine der steuernden Größe äquivalente Wechselleistung umwandeln

Im letzten Teil erscheint zunächst Wechselleistung einengend zu sein, denn es gibt auch Schaltungen, die Gleichleistungen (und auch Wechselleistungen) in definierte, andere Gleichleistungen umwandeln. Dies sind *Stromversorgungsschaltungen*. Sie werden z. T. schon im Abschnitt 1.4.8., 5.4.3. bis 5.4.6. sowie 6.4. behandelt. Sie werden noch weiter im Abschnitt 7.6 behandelt werden.

7.4.1.1. Grenzen der Ausgangsarbeitsfläche

Eine wesentliche Grundlage aller Ausgangs- und Leistungsstufen ist das Ausgangskennlinienfeld eines Transistors. Es ist beispielhaft in Bild 7.4.1 gezeigt. Im Gegensatz zu den Darstellungen in Bild 5.5.1h) für den FET und 5.7.15a) für den Bipolartransistor sind hier lineare Strom- und Spannungsachsen gewählt. Sie sind für die folgenden Betrachtungen notwendig. Dieses Bild gilt generell für FET und Bipolartransistoren. Daher werden im folgenden die Indizes für Kollektor bzw. Drain weggelassen. Es ist *durch sechs Grenzen* gekennzeichnet:

— Der Kollektor- bzw. Drainstrom I darf einen maximalen Wert I_{max} nicht überschreiten.
— Der Kollektor- bzw. Drainstrom kann (bzw. sollte) nicht kleiner als Null, also nicht negativ werden.
— Die Betriebsspannung U (Kollektor- bzw. Source-Drain-Spannung) darf einen maximalen Wert U_{max} nicht überschreiten.

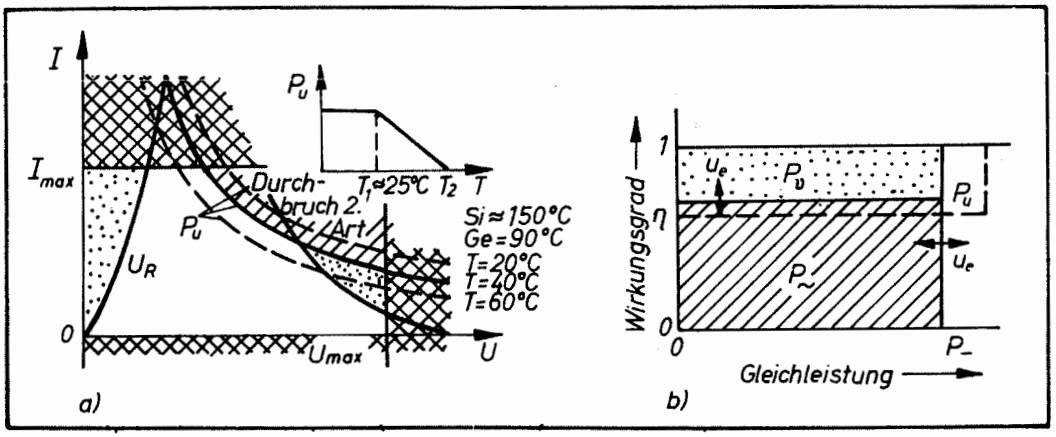


Bild 7.4.1. a) Ausgangsgrenzfläche für Transistoren mit der zugehörigen Deratingkurve (rechts oben) als Einfluß der Umgebungstemperatur auf die maximal umgesetzte Leistung P_u
b) Verteilung und Zusammenhang der Leistungen bei Leistungsendstufen. Es bedeuten P_- die zugeführte Gleichleistung, $P_\sim$ die daraus erzeugte Wechselleistung, P_v die in Wärme umgewandelte Verlustleistung und P_u die maximal vom Bauelement umsetzbare Wärmeleistung entsprechend der Hyperbel im Teilbild a). Da P_- und $P_\sim$ von der Aussteuerung durch u_e abhängen (Doppelpfeile), muß für P_u der ungünstigste Fall berücksichtigt werden

- Die Kollektor- bzw. Source-Drain-Spannung kann infolge von Verlusten nicht die Restspannung U_R unterschreiten. Sie ist bei Bipolartransistoren durch die Grenzen in Flußrichtung (vgl. z.B. Bild 5.6.4a) bzw. 5.7.3a)), beim FET durch den Spannungsabfall am Kanal, am Ein-Widerstand (vgl. z.B. Bild 5.5.1d)) begrenzt. Für die meisten Anwendungen sind diese Werte zumindest in erster Näherung vernachlässigbar klein. Deshalb wurde dieser Bereich in Bild 7.4.1 auch nur gepunktet. Dann wird $U_R = 0$ gewählt. Falls ein extrem kleiner Wert hier erforderlich ist, z.B. bei kleinen Betriebsspannungen, sind Ge-Transistoren am günstigsten.
- Die im Transistor umgesetzte Leistung $P_u = U \cdot I$ wird in Wärme umgewandelt. Dadurch ist diese Leistung entsprechend dem Aufbau des Transistors und der Wärmeableitung (s. 5.7.6.) begrenzt. Dadurch entsteht die Leistungshyperbel in Bild 7.4.1. Sie hängt gemäß der Deratingkurve (vgl. auch Bild 1.1.14) von der Umgebungstemperatur ab. Deshalb die drei Hyperbeln, einschließlich der stark ausgezogenen für 40 °C. Bei den Hyperbeln ist weiter zu beachten, daß sie für kurze Zeiten (z.B. durch Impulse) durchaus überschritten werden können. Die Grenzhyperbel gilt in diesem Sinne in erster Näherung im zeitlichen Mittel (vgl. Bild 5.5.1h) und 5.7.15a)). Deswegen erfolgt hier nur eine einfache Schraffur.
- Für Bipolartransistoren, und nur für diese, gibt es noch einen zweiten Durchbruch (s. 5.7.5.4), der das Ausgangsfeld zusätzlich begrenzt.

7.4.1.2. Bedeutung der Ausgangsfläche

Es erscheint zunächst nicht sinnvoll, die maximale Ausgangsarbeitsfläche allein zu betrachten. Es läßt sich jedoch schnell einsehen, daß der *Verlauf der Kennlinien* in dieser Fläche für eine Endstufe von untergeordneter Bedeutung ist. Die Verstärkung der Eingangssignale für die Endstufe kann im wesentlichen ohne Rücksicht auf Leistungsumwandlung erfolgen. Es werden nämlich die *Steuerleistungen vergleichsweise klein* sein. Weiter kann über die Endstufen, eventuell einschließlich der Vorstufen, eine starke Gegenkopplung realisiert werden. Hierdurch gelingt im Sinne von Bild 7.1.6. eine Linearisierung der Kennlinien bis an die Grenzen der Aussteuerung, und das sind im ungünstigsten Fall gerade die Grenzen der Ausgangsarbeitsfläche. Krümmungen der Kennlinien der Endstufen sind daher in erster Näherung zu vernachlässigen. Weiter können mit gemischter Strom-Spannungs-Gegenkopplung entsprechend Bild 7.1.3 die Steuerkennlinien in nahezu jede Richtung gedreht werden. Dies sind die Gründe, warum in den folgenden Betrachtungen nur ausnahmsweise die Eingangsspannungen bzw. -ströme für Endstufen Bedeutung besitzen.

7.4.1.3. Leistungsumsatz

In einer Leistungsendstufe wird eine *Gleichleistung* P_- teilweise in eine *Wechselleistung* $P_\sim$ umgesetzt. Die Differenz wird als *Verlustleistung* $P_v = P_- - P_\sim$ in Wärme umgewandelt und muß abgeführt werden. Sie darf daher höchstens gleich der *umsetzbaren Leistung* $P_u \geqq P_v$ der aktiven Bauelemente sein.
Im allgemeinen soll die Ausgangsleistung $P_\sim$ direkt linear (d. h. unverzerrt) der Eingangsspannung u_e folgen (Doppelpfeil an der Grenze zwischen P_v und $P_\sim$). Um dies optimal zu realisieren, wird in verschiedenen Schaltungen die Gleichleistung P_- ebenfalls als Funktion von u_e gestaltet. Dies ist in Bild 7.4.1b) durch den zweiten Doppelpfeil angedeutet. In vielen Fällen interessiert der *Wirkungsgrad* η der Umwandlung,

$$\eta = \frac{P_\sim}{P_-}. \tag{1}$$

Da die Steuerungen von $P_\sim$ und P_- nicht genau gleichartig erfolgen können, ist dieser Wirkungsgrad ebenfalls eine Funktion von u_e. Darüber hinaus ist er eine typische Funktion der jeweils realisierten Endstufe. Dies bedeutet, daß auch die Verlustleistung P_v eine Funktion von u_e ist. Deshalb muß die *maximal im Bauelement umsetzbare Leistung* P_u auf den ungünstigsten Fall bezogen werden. Dabei interessiert oft, wie groß P_u im Vergleich zu $P_\sim$ ist. Dies kann z. B. durch den kleinsten auftretenden Wirkungsgrad beschrieben werden. Allgemein gilt jedoch

$$r = P_\sim / P_u . \tag{2}$$

Endstufen verbrauchen in vielen Geräten die meiste Leistung. Dies wirkt sich sowohl auf die Ökonomie der Geräte als auch bezüglich des Aufwandes zur Ableitung der Wärme nachteilig aus. Deshalb geht das Bestreben dahin, zumindest η möglichst groß, im Idealfall 100% zu machen. Andererseits geht damit zumindest teilweise auch ein möglichst großes r einher, also der Einsatz kleinerer Bauelemente.

Die *Amplitudenstatistik* für Sprache und Musik weist aus, daß die maximale Leistung einer Endstufe in weniger als 1% der Zeit benötigt wird und daß in 90% der Zeit bestenfalls 20% der Leistung notwendig sind. Derartige Werte gelten für Verstärker in Sendern. Bei Rundfunkempfängern, insbesondere der HiFi-Klasse, sind heute die Reserven so groß, daß in 90% der Zeit nur noch 1% der maximalen Leistung auftritt und sie selbst meist nie benötigt wird, sondern als „stille Reserve" brach liegt. Diese „*Überdimensionierung*" ist eigentlich nicht zu verantworten. Deshalb sind im Laufe der Zeit viele Varianten entwickelt worden, die einen optimalen Wirkungsgrad auch bei verschiedenen Aussteuerungen ermöglichen. Bereits in den 30er Jahren wurden effektive Lösungen, insbesondere bei den Großsendern mit einigen hundert Kilowatt und bei transportablen, batteriebetriebenen Geräten entwickelt, und noch immer werden neue, interessante Vorschläge unterbreitet. Einige werden im folgenden analysiert, andere kurz erwähnt.

7.4.2. Endstufe mit nur einem Transistor

Die folgenden Betrachtungen werden an einem *npn*-Transistor durchgeführt. Sie gelten voll analog auch für einen *pnp*-Transistor bzw. einen SGFET oder MOSFET

7.4.2.1. Optimaler Arbeitswiderstand

In Bild 7.4.2a) ist die Schaltung einer Endstufe mit nur einem Transistor gezeigt. Es sei angenommen, daß der Transistor und die Betriebsspannung U_b fest gegeben seien.

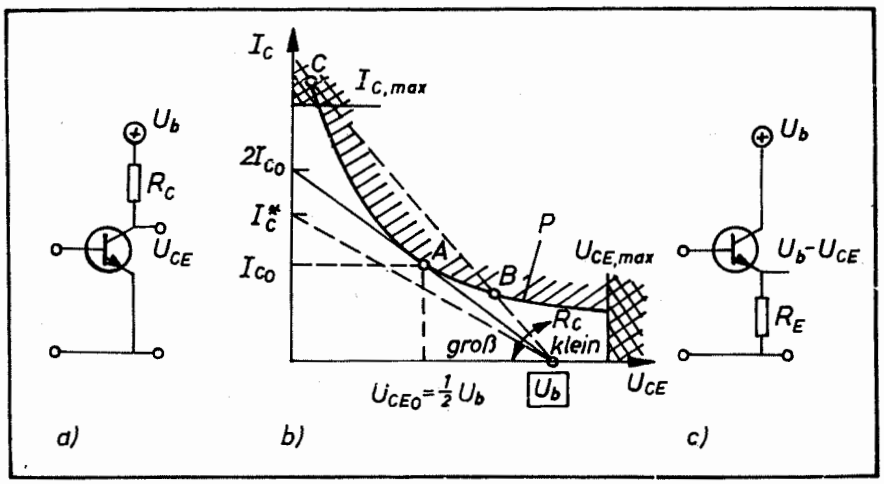


Bild 7.4.2. Zur Endstufe mit einem Transistor und ohmschen Arbeitswiderstand

a) Emitterschaltung
b) Betrachtungen im Ausgangsfeld bei fester Betriebsspannung U_b und unterschiedlichem Kollektorwiderstand R_C
c) Kollektorschaltung

Dann ist der Arbeitswiderstand R_C wählbar. Die entsprechende graphische Konstruktion zeigt Bild 7.4.2b). Da für die Spannung am Kollektor

$$U_{CE} = U_b - I_C R_C \tag{3}$$

gilt, folgen die Geraden, die von U_b ausgehen und bei

$$I_C^* = U_b/R_C \tag{4}$$

auf der Ordinate enden. Je nach der Größe von R_C verlaufen sie mehr oder weniger steil. Auf diesen Geraden liegen alle möglichen Arbeitspunkte der Endstufe. Dabei gibt es einen Widerstand, der die Leistungsparabel P_u im Punkt A berührt. Er ist dadurch ausgezeichnet, daß im Punkt A der Transistor die maximal zulässige Leistung aufnimmt und in allen anderen Punkten auf der Geraden darunter bleibt. Bei kleinerem R_C wird die Hyperbel geschnitten, z. B. in B und C, und für die meisten Arbeitspunkte wird der Transistor überlastet. Bei größerem R_C wird der Transistor an keiner Stelle bis zur Grenze ausgelastet. Der ausgezeichnete Widerstand werde daher R_{opt} genannt. Wegen der Berührung im Punkt A muß er die *Tangente an die Hyperbel* darstellen,

$$U_{CE} = \frac{P_u}{I_C}, \tag{5}$$

und für die Tangente gilt allgemein

$$R_{tan} = -\frac{dU_{CE}}{dI_C} = \frac{P_u}{I_C^2} = \frac{U_{CE}^2}{P_u}. \tag{6a}$$

Diese Gleichung gilt auch für den Punkt A und muß Gl. (3) erfüllen. Nach R_C aufgelöst, gibt sie

$$R_C = \frac{U_b - U_{CE}}{I_C}. \tag{6b}$$

Die Bedingung

$$R_{opt} = R_C = R_{tan} \tag{6c}$$

ergibt für den Punkt A

$$U_b = 2U_{CE_0}. \tag{7}$$

Dies bedeutet, daß die *Widerstandsgerade die Hyperbel gerade in dem Punkt berührt*, von dem aus eine Spannungsaussteuerung nach beiden Seiten gleich groß ist. Dasselbe gilt offensichtlich auch für den Strom. Der Punkt A ist also zugleich ein besonders günstiger Arbeitspunkt für maximale Spannungs- und Stromausnutzung bei der maximalen Aussteuerung. Daher gehören zum Punkt A die Größen U_{CE_0} und I_{C_0}.

Mit allen anderen Widerständen sind nur weniger günstige Bedingungen zu erreichen. Soll R_{opt} aus U_b und P_u bestimmt werden, so folgt

$$\boxed{R_{opt} = \frac{U_b^2}{4P_u}} \,. \tag{8}$$

In einigen Fällen, insbesondere bei kleinen Betriebsspannungen, kann $I_{C,max}$ die Grenze für R_C festlegen, dann gilt

$$R_{opt}^* = \frac{U_b}{I_{C,max}} \,. \tag{9}$$

Dieser Fall wird nicht weiter berücksichtigt.

7.4.2.2. Kollektor- und Emitterschaltung

Die bisherigen Betrachtungen zur Bestimmung des optimalen Arbeitswiderstandes erfolgten mittels der Emitterschaltung von Bild 7.4.2a). Sie gelten in analoger Weise unverändert auch für die Kollektorschaltung von Bild 7.4.2c). Die auf den Transistor bezogenen Leistungsbilanzen sind davon unberührt. Betroffen ist nur die Steuerspannung, selbst der notwendige Steuerstrom bleibt unverändert. Bei der Emitterschaltung genügen kleine Steuerspannungen, während die Kollektorschaltung infolge der vollständigen Spannungsgegenkopplung Steuerspannungen benötigt, die praktisch genauso groß wie die Ausgangsspannungen an R_E sind. Dieser erhebliche Nachteil wird aber zumindest teilweise durch die so bedingte direkte und sehr gute Linearisierung der Kennlinien ausgeglichen.

In Anbetracht dieser Aussagen bleiben Gl. (8) und alle im folgenden daraus abgeleiteten Leistungsbilanzen davon unberührt, ob eine Emitter- oder Kollektorschaltung für die Endstufe verwendet wird.

7.4.2.3. Leistungsbilanzen

Wird der Transistor ohne Aussteuerung betrieben, so benötigt die Schaltung die Leistung

$$P_- = U_b I_{C_0} \,. \tag{10}$$

Auf den Transistor selbst entfällt

$$P_V = U_{CE_0} I_{C_0} = \tfrac{1}{2} U_b I_{C_0} \,. \tag{11}$$

Wird der Transistor ausgesteuert, dann überlagern sich den Gleichwerten Wechselwerte. Bei erhöhtem Kollektorstrom ist entsprechend der Widerstandsgeraden die Kollektorspannung kleiner und umgekehrt. Es werde ein Aussteuerungsfaktor A mit

$$0 \leqq A \leqq 1 \tag{12}$$

und sinusförmiger Aussteuerung eingeführt. Allgemein kann er für Strom und Spannung unterschiedlich sein. Es werde jedoch vorausgesetzt, daß *gleichzeitig 100%ige Strom- und Spannungsaussteuerung* erfolgen. Dies ist bei R_{opt} erfüllt. Also gilt

$$U_{CE} = U_{CE_0}(1 + A \sin \omega t) \tag{13}$$

sowie

$$I_C = I_{C_0}(1 - A \sin \omega t) \,. \tag{14}$$

Für die *im Transistor umgesetzte Leistung* gilt dann

$$P_v = \frac{U_{CE_0} I_{C_0}}{T} \int_0^T (1 + A \sin \omega t)\,(1 - A \sin \omega t)\,\mathrm{d}t$$

$$= U_{CE_0} I_{C_0} \left(1 - \frac{A^2}{2}\right). \tag{15}$$

Der Wärmeumsatz im Transistor sinkt also mit der Aussteuerung.

Die *Wechselleistung* in Abhängigkeit von der Aussteuerung beträgt

$$P_\sim = \frac{U_{CE_0} I_{C_0}}{T} \int_0^T A^2 \sin^2 \omega t \,\mathrm{d}t = A^2 \frac{U_{CE_0} I_{C_0}}{2}. \tag{16}$$

Dies ist genau jener Anteil, der gemäß Gl. (15) beim Transistor als Wärmemenge weniger auftritt. Dieser Anteil wird also von Gleichleistung auf die Gesamtleistung gemäß Gl. (10) bezogen, dann gilt für den *Wirkungsgrad* gemäß Gl. (1)

$$\eta = \frac{A^2}{4} \leqq \frac{1}{4}. \tag{17}$$

7.4.2.4. Verschiedene Betriebsfälle

Der Wärmeumsatz im Transistor hängt gemäß den vorangegangenen Betrachtungen deutlich von der Aussteuerung A ab. Dies hat zur Folge, daß *drei Arten der Aussteuerung* zu unterscheiden sind:

o) Die Endstufe ist in der meisten Zeit nur wenig oder gar nicht ausgesteuert. Für diesen Fall ohne wesentliche Aussteuerung gilt $A \approx 0$.

s) Ständige und vollständige Aussteuerung. Es wird immer möglichst genau $A = 1$ eingehalten. Dieser Fall trifft auf Oszillatoren, Generatoren und Senderendstufen zu.

z) Von zweiteiliger Aussteuerung wird gesprochen, wenn sie sich je nach den Umständen und Gegebenheiten zwischen den beiden anderen Extremen bewegt. Dies trifft insbesondere bei Schallsignalen, also z. B. Musik und Sprache zu. Hier kann für die weiteren Betrachtungen ein mittlerer Wert für A gewählt werden.

Die Wahl des Arbeitspunktes gemäß Abschnitt 7.4.2.1. gilt exakt für den Fall ohne Aussteuerung. Bei *ständiger Aussteuerung* (insbesondere mit so hohen Frequenzen, daß dadurch noch keine thermischen Fluktuationen auftreten) könnte eine Widerstandsgerade gewählt werden, welche die Hyperbel schneidet und gemäß Gl. (15) mit $A = 1$ einen „Ruhearbeitspunkt" mit der doppelten Leistung gegenüber dem Fall ohne Aussteuerung besitzt. Wird weiterhin U_b konstant gehalten, so bedeutet dies den doppelten „Ruhestrom" und damit einen halb so großen optimalen Arbeitswiderstand. Bei dieser Dimensionierung wird mit dem gleichen Transistortyp doppelt so viel Wechselleistung erzeugt. Es besteht jedoch die Gefahr, daß bei Ausfall der Steuerleistung der Transistor im „Arbeitspunkt" doppelt überlastet und damit zerstört wird.

Gemäß Gl. (2) interessiert das Verhältnis aus erzeugbarer Wechselleistung und notwendiger Verlustleistung eines Transistors. Bei ständiger Aussteuerung für beide Leistungen A und je nach der *Aussteuerung* folgt aus Gl. (15) und (16)

$$r_s = \frac{A^2}{2 - A^2} \leqq 1. \tag{18}$$

Bei zeitweiliger Aussteuerung ist der ungünstige Fall für die Verlustleistung, also $A = 0$, in Gl. (15) zu verwenden, und es gilt,

$$r_z = A^2/2 \leqq 1/2 . \tag{19}$$

Zur Verallgemeinerung der vorangegangenen Betrachtungen sei die *Leistung im Arbeitspunkt*, der nicht auf der Leistungshyperbel zu liegen braucht, als

$$P_0 = U_{CE_0} I_{C_0} \tag{20}$$

eingeführt und alle anderen Leistungen darauf normiert. Die so *normierten Leistungen* werden klein geschrieben. Folglich gilt

$$p_v = 1 - A^2/2 , \tag{15a}$$

$$p_\sim = A^2/2 . \tag{16a}$$

Je nach der gegebenen Aussteuerung (auch mittleren) kann dann eine relative Verlustleistung und Wechselleistung gefunden werden. Die relative Verlustleistung gibt z.B. an, um welchen Faktor je nach Aussteuerung die Hyperbel unterhalb des Arbeitspunktes liegen darf.

Schließlich seien noch die *Strom- und Spannungseffektivwerte* berechnet. Hierbei ist zu beachten, daß bei sinusförmiger Aussteuerung nur $1/\sqrt{8}$ der Spitzenwerte gilt. Daher folgt

$$i_{C,\,eff} = \frac{A}{\sqrt{2}} I_{C_0} , \tag{21}$$

$$u_{CE,\,eff} = \frac{A}{\sqrt{2}} U_{CE_0} = \frac{A}{\sqrt{8}} U_b . \tag{22}$$

7.4.2.5. Induktiv überbrückter Arbeitswiderstand

Etwa die Hälfte der Gleichstromleistung wird bei der Schaltung in Bild 7.4.2 am Widerstand R_C umgesetzt. Dies läßt sich vermeiden, wenn er mit einer Induktivität überbrückt wird. Dann ist im Ruhearbeitspunkt die Kollektorspannung gleich der Betriebsspannung. Lediglich für Wechselströme, wo der induktive Widerstand wirksam wird, treten neue Bedingungen auf.

Zunächst sei vereinfachend angenommen, daß die Indultivität unendlich groß sei. Dann wirkt sie nur noch als Energiespeicher für die Gleichstromkomponente I_{C_0}. Wird der Strom erhöht, so sinkt U_{CE}, wird er erniedrigt, so wächst U_{CE} sogar über die Betriebsspannung hinaus. Dies wird meist anschaulich schwer einzusehen sein. Es kann wie folgt erklärt werden. Fließt durch eine Induktivität ein Strom I, so ist in ihr gemäß Tab. 1.3.1 die Energie I^2L gespeichert. Ändert sich der Strom, so muß die entsprechende Energiedifferenz frei werden. Sie erhöht über die LENZsche Regel, Gl. (1.5.9), die Spannung. Die Spannungserhöhung entsprechend der Arbeitsgeraden kann nur für eine bestimmte Zeit wirksam werden. Daraus folgt eine minimale Größe für die Induktivität. Es gilt mit einer unteren Grenzfrequenz f_u daher

$$L \geqq \frac{R_{opt}}{2\pi f_u} . \tag{23}$$

Es wurde gleich der optimale Arbeitswiderstand gewählt. Beim Gleichheitszeichen sind induktiver und ohmscher Widerstand gleich groß. Die Arbeitsgerade ist daher entsprechend Bild 5.5.6 zum Kreis geworden. Mit größerem L oder bei höherer Frequenz geht sie in eine Ellipse über, um schließlich wieder praktisch zur Geraden zu werden. Im Grenzfall ergeben sich Widerstandsgeraden, wie sie in Bild 7.4.3 a) gezeigt sind. Es gibt wiederum für einen auf der Hyperbel gewählten Arbeitspunkt einen optimalen Widerstand. Insgesamt sind sogar alle Betrachtungen des vorigen Abschnittes voll zu übernehmen, wenn

statt U_b jetzt $2U_b$

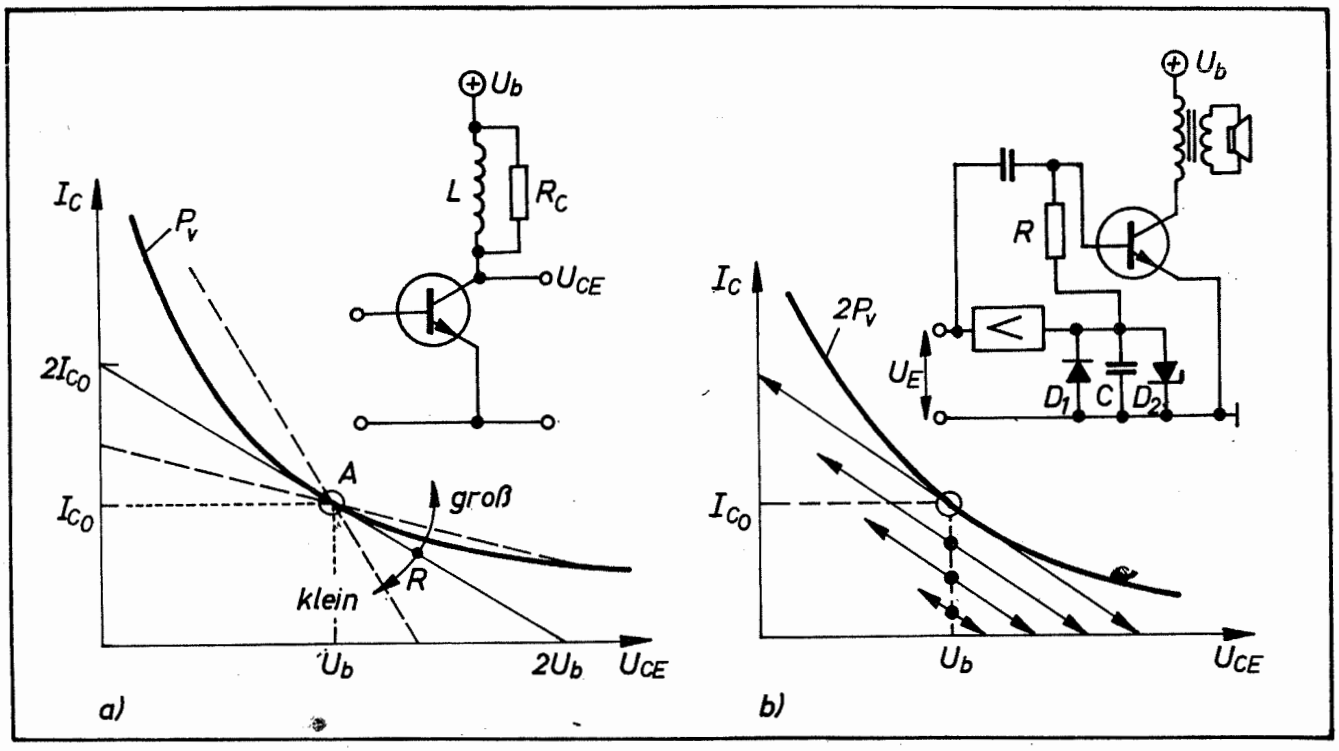


Bild 7.4.3. a) Endstufe mit einer Drossel, die den Arbeitswiderstand für Wechselstrom überbrückt b) Endstufe mit gleitendem Arbeitspunkt bez. des Stromes

gesetzt wird. Ergänzend sei erwähnt, daß wenn schon eine Drossel Anwendung findet, sie auch gleich als Transformator ausgelegt werden kann und so den Widerstand des Verbrauchers genau auf den R_{opt} transformiert.

7.4.2.6. Gleitende Arbeitspunktverlagerung

Die Arbeitsweise dieses Prinzips ergibt sich aus Bild 7.4.3 b). Da weder der Arbeitswiderstand noch die Betriebsspannung mit der Aussteuerung zu ändern sind, wird der Kollektorstrom passend verstellt. Dies erfolgt mittels Gleichrichtung der verstärkten Eingangsspannung durch D_1. Der Kondensator C hält Signalanteile fern. Die Verstellung des Arbeitspunktes erfolgt also relativ langsam. Die Z-Diode D_2 sorgt dafür, daß nicht die Leistungshyperbel überschritten wird, daß also maximal der Arbeitspunkt A eingenommen wird. Der Arbeitspunkt wird genau so eingestellt (Idealfall), daß immer eine 100%ige Stromdurchsteuerung vorliegt. Dies deuten die Doppelpfeile als Grenzen der Arbeitskennlinie mit der Neigung R_{opt} an. Da bei einer solchen Schaltung der Arbeitspunkt A nie als Ruhearbeitspunkt der Aussteuerung auftritt, kann entsprechend Gl. (15) die Hyperbel mit der doppelten Verlustleistung gewählt werden, deshalb ist auch R_{opt} nur halb so groß wie ohne Steuerung des Arbeitspunktes. Es werde wieder der Aussteuerungsfaktor A als Maß für die Eingangsspannung verwendet. Dann gilt für den Kollektorstrom

$$I_C = A I_{C_0}. \tag{24}$$

Da immer die Stromdurchsetzung 100% ist, folgt weiter

$$I_C = A I_{C_0}(1 - \sin \omega t)\,. \tag{25}$$

Gemäß Abschnitt 7.4.3.2. und mit Gl. (13) ergibt sich für die im Transistor erzeugte Wärmemenge

$$P_V = U_b I_{C_0} A(3 - 2{,}5A)\,. \tag{26}$$

Für die Wechselleistung gilt wieder Gl. (16), jedoch mit U_b statt U_{CE_0}, und Gl. (24) für I_{C_0}:

$$P_\sim = \tfrac{1}{2} A^2 U_b I_{C_0}\,. \tag{27}$$

Die insgesamt aufgenommene Gleichleistung beträgt

$$P_- = A U_b I_{C_0}\,. \tag{28}$$

Dadurch ist der Wirkungsgrad nicht mehr quadratisch, sondern nur noch linear von der Aussteuerung abhängig:

$$\eta = A/2\,. \tag{29}$$

7.4.3. Gegentaktendstufen

Mit dem durch die Drossel überbrückten Widerstand konnte jene Gleichleistung vermieden werden, die im Arbeitspunkt bereits am ohmschen Widerstand entsteht. Ansonsten wird der Wirkungsgrad hierdurch nicht verändert. Effektiver ist da schon der gleitende Arbeitspunkt. Er besitzt aber nur bei voller Aussteuerung auch die volle Spannungsaussteuerung. Es gibt jedoch bisher keinen Weg, Ruhestrom und Betriebsspannung gleichmäßig zu steuern. Aber selbst unter diesen zunächst ideal erscheinenden Bedingungen würde der Wirkungsgrad immer nur $^1/_2$ sein. Dies ist dadurch bedingt, daß von einem aktiven Bauelement ausgegangen wird, für das ein Arbeitspunkt mit Gleichleistung gewählt wird, und um diesen Arbeitspunkt herum die Aussteuerung (bestenfalls für Strom und Spannung zu je 100%) erfolgt. Eine völlig andere Möglichkeit bieten Gegentaktschaltungen. Hierbei übernimmt jeder Transistor genau die Verstärkung einer Halbwelle und benötigt daher keinen Ruhestrom.

Im Laufe der Technikgeschichte gab es zu dem heute üblichen Weg viele Varianten. Sie werden hier nicht behandelt, da die sogenannte „eisenlose Endstufe", wie sie bei ihrem Entstehen hieß, heute als einzige noch Bedeutung hat. Interessenten an der älteren Technik seien auf die früheren Auflagen dieses Buches verwiesen.

7.4.3.1. Prinzip; Gegentakt-A-Betrieb

Die Funktionsweise der heute üblichen Gegentaktstufen sei mittels Bild 7.4.4 erklärt. In Teilbild a) sind je eine *pnp- und npn-Emitterendstufe übereinander* dargestellt. Durch passende Wahl der Steuerspannungen ist es möglich, die beiden Stufen so anzusteuern, daß zu jedem Zeitpunkt die gleiche Spannung an den jeweiligen R_E abfällt. Dies bedeutet, daß der obere Transistor gerade dann einen höheren Strom führt, wenn durch den unteren weniger fließt. Infolge der unterschiedlichen Transistortypen ist dieser Fakt erfüllt, wenn die U_E sich gleichartig ändern.

Unter den genannten Bedingungen können gemäß Teilbild b) die *beiden Emitterwiderstände durch einen halb so großen* gemeinsamen ersetzt werden. An beiden dadurch in Reihe geschalteten Transistoren liegt damit stets die volle Betriebsspannung, und darüber hinaus fällt im Arbeitspunkt ohne Aussteuerung an dem gemeinsamen Emitterwiderstand keine Spannung ab. Die Bedingungen sind für ihn also bezüglich

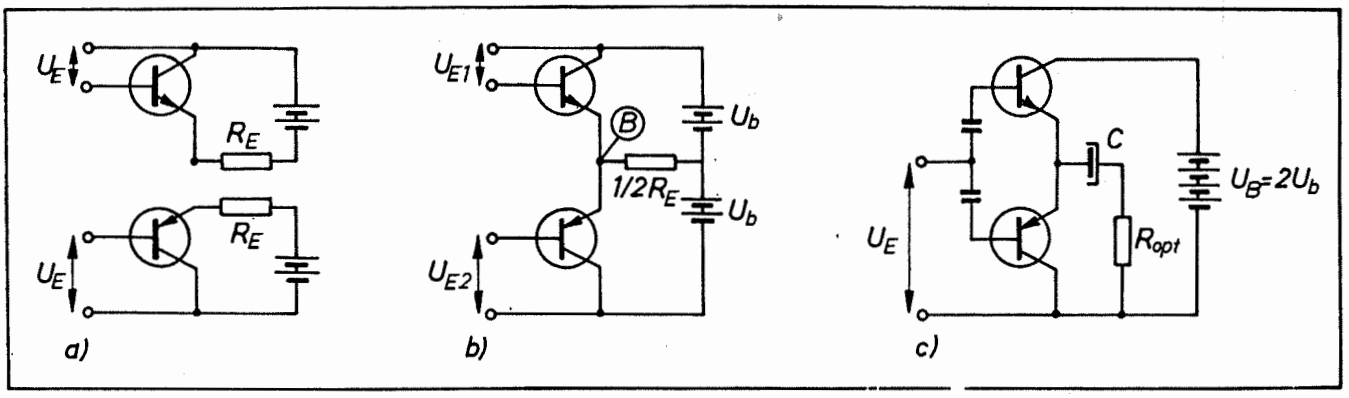


Bild 7.4.4. Prinzip der Gegentaktschaltung

a) Je eine *pnp*- und *npn*-Emitterendstufe

b) Zusammenfassung der beiden Stufen zu einer Schaltung

c) Zusammenfassung der Spannungsquellen und Steuerspannungen

Gleichstrom so wie jene bei der einfachen Endstufe, wo er mit Drossel überbrückt wurde. Der *Ruhestrom* geht voll *durch die beiden* in Reihe geschalteten *Transistoren*.

Bei Ansteuerung der beiden Transistoren wird für Gleichstrom der eine Transistor hoch- und der andere niederohmiger. Das Potential am Punkt B schwankt im Takt der Aussteuerung, und dadurch fließt ein entsprechender Wechselstrom durch den gemeinsamen Emitterwiderstand. Aus dieser Betrachtung wird die Bezeichnung Brücken-Gegentakt-Schaltung verständlich. Eine *andere Umschreibung* dieses Geschehens ist folgende. Der eine Transistor liefert einen zusätzlichen, erhöhten Strom durch den gemeinsamen Emitterwiderstand. Der andere zieht in diesem Fall weniger, und auch dieser Differenzstrom muß vom gemeinsamen Emitterwiderstand ausgeglichen werden. Da die Spannungsquelle, vom Widerstand aus gesehen, entgegengesetzt gepolt ist, verstärkt sich der Spannungsabfall am Widerstand nochmals. Der Widerstand übernimmt also den vollen Differenzstrom zwischen den beiden Transistoren. Hieraus ist die englische Bezeichnung für Gegentaktschaltungen, nämlich *push-pull* (schieben und ziehen), zu verstehen.

Mit Teilbild c) wird nur noch eine geringe Änderung der Schaltung durchgeführt. Die Ausgangswechselgrößen werden *über den Kondensator C ausgekoppelt*. Dadurch entfällt die Anzapfung der Spannungsquelle. Die beiden Basisanschlüsse werden ebenfalls wechselstrommäßig zusammengeschaltet, was allerdings Transistoren mit gleicher Stromverstärkung (Transistorpaare) erfordert. Bei der konkreten Ausführung der Schaltung ist weiter für die richtige Arbeitspunkteinstellung zu sorgen. Die *typischen Kennwerte* dieser als Gegentakt-A-Schaltung bezeichneten Betriebsart betragen mit der *Aussteuerung* A gemäß Gl. (13) und (14)

$$\left.\begin{aligned} R_{\mathrm{opt}} &= \frac{U_{\mathrm{B}}}{4I_{\mathrm{C_0}}}, \qquad \eta = A^2/2\,, \\ P_{\sim} &= \frac{A^2U_{\mathrm{B}}^2}{8R_{\mathrm{opt}}}, \qquad P_{\mathrm{v}} = \frac{U_{\mathrm{B}}^2(1 - A^2/2)}{4R_{\mathrm{opt}}}\,. \end{aligned}\right\} \tag{30}$$

7.4.3.2. Idealer B-Betrieb

Der wesentliche Verlustanteil der Gegentakt-A-Schaltung liegt in dem *Ruhestrom* $I_{\mathrm{C_0}}$, der ohne Aussteuerung beide Transistoren *nutzlos* durchfließt. Eine sehr vorteilhafte Weiterentwicklung der Gegentaktendstufen besteht darin, daß *jeder Transistor genau eine Halbwelle* zur Verstärkung übernimmt, während der andere zu dieser Zeit voll gesperrt ist. Dies bedeutet, daß ohne Aussteuerung der Ruhestrom beider Transistoren Null ist. Bei der positiven Halbwelle der Steuerspannung führt (in Bild 7.4.4b)) der obere Transistor den notwendigen Strom durch den gemeinsamen Emitterwiderstand, während der untere durch die positive Halbwelle nur noch mehr gesperrt wird. Für die negative Halbwelle liefert der untere Transistor die volle Leistung und der obere bleibt gesperrt. Für die weiteren Aussagen seien zur Vereinfachung zunächst ideale Bedingungen auch im Übernahmebereich vorausgesetzt. Die Abweichungen hiervon werden erst später berücksichtigt. Dann genügt es gemäß Bild 7.4.5a), die Verhältnisse jeweils nur für den Transistor zu betrachten, der gerade Strom führt, sich also in der aktiven Phase befindet. Hieraus sind leicht über beide Transistoren die Leistungsbedingungen für beide Halbwellen zu erhalten. Der Ruhearbeitspunkt für den Transistor liegt bei U_{b} und $I_{\mathrm{C}} = 0$. Da er keinen Strom im Arbeitspunkt führt, muß die *Widerstandsgerade* mit dem Arbeitswiderstand R durch den *maximalen Strom*

$$I_{\mathrm{Cm}} = U_{\mathrm{b}}/R \tag{31}$$

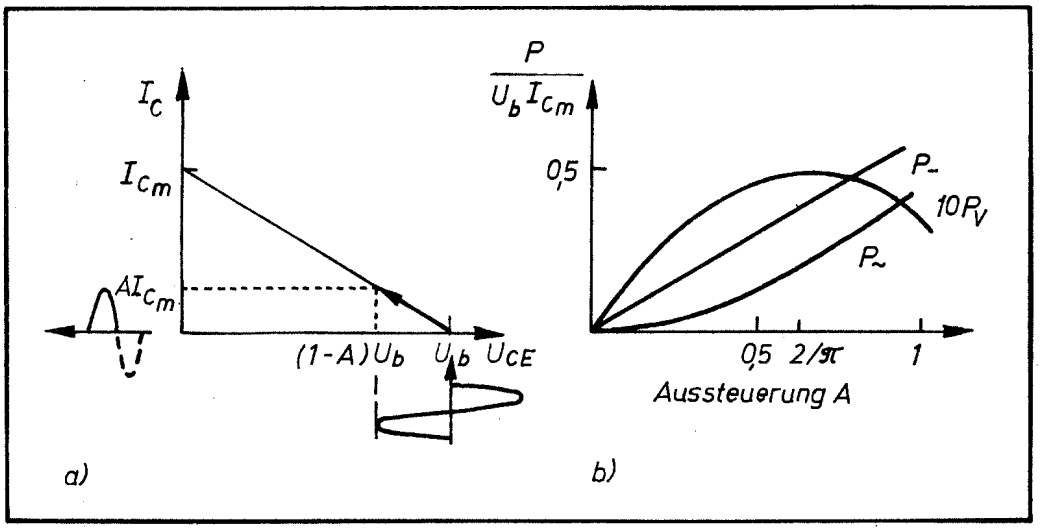


Bild 7.4.5. Zu den Eigenschaften einer Gegentakt-B-Endstufe

a) Zur Festlegung der Größen. Die Leistungshyperbel fehlt bewußt hier. Ihre Anordnung ergibt sich aus
b) dem Verlauf der drei Leistungen in Abhängigkeit von der Aussteuerung
Die in jedem Transistor erzeugte Wärmeleistung ist zehnfach überhöht

beschrieben werden. In Abhängigkeit von der Aussteuerung A wird mit beiden Transistoren dann folgende *Wechselleistung* erzeugt:

$$P_\sim = \frac{A^2}{2} U_b I_{Cm} . \tag{32}$$

Die *aufgenommene Gesamtleistung* ist so groß wie die Gleichleistung, die von einem Transistorzweig während der aktiven Phase aufgenommen wird:

$$P_- = \frac{2}{T} \int_0^{T/2} A U_b I_{Cm} \sin \omega t \, dt = \frac{2}{\pi} A U_b I_{Cm} . \tag{33}$$

Folglich gilt

$$\eta = \frac{A\pi}{4} \leqq 78.8\% . \tag{34}$$

Der Wirkungsgrad nimmt also wie bei gleitendem Arbeitspunkt nur linear mit der Aussteuerung ab, ist jedoch 1,57mal so groß.

Der Wärmeumsatz in den Transistoren kann auf zwei Wegen berechnet werden; einmal als Differenz von Gl. (33) und (34). Zum anderen kann er auch aus dem *Schwingungsverlauf am Transistor,*

$$P_v = \frac{1}{T} \int_0^{T/2} A U_b I_{Cm} (1 - A \sin \omega t) \sin \omega t \, dt = U_b I_{Cm} A \left(\frac{1}{\pi} - \frac{A}{4} \right), \tag{35}$$

bestimmt werden. Die Verlustleistung je Transistor ist also stark von der Aussteuerung abhängig und *durchläuft ein relatives Maximum* bei,

$$A = \frac{2}{\pi} \tag{36}$$

mit

$$P_{v,\,max} = \frac{U_b I_{Cm}}{\pi^2} . \tag{37}$$

Damit ist die bisherige Rechengröße I_{Cm} festgelegt:

$$I_{Cm,\,z} = \frac{\pi^2 P_u}{U_b}. \tag{38}$$

Der Index z wurde angefügt, weil dieser Wert für zeitweilige Aussteuerung gilt und wegen des Maximums von Gl. (35) der ungünstigste Fall gewählt wurde, nämlich daß durch Zufall oder Absicht eine dauernde Aussteuerung gemäß Gl. (36) erfolgt. Der zugehörige *optimale Arbeitswiderstand* kann gemäß

$$R_{opt,\,z} = \frac{2U_b}{I_{Cm,\,z}} = \frac{U_B}{I_{Cm,\,z}} \tag{39}$$

bestimmt werden. Der Faktor 2 folgt aus Bild 7.4.4e). Mit Gl. (38) und (39) gilt

$$R_{opt,\,z} = \frac{2U_b^2}{\pi^2 P_u}. \tag{40}$$

Nun kann r_z bestimmt werden. Aus Gl. (32) und (37) folgt

$$r_z = \frac{\pi^2 A^2}{2}. \tag{41}$$

In Bild 7.4.5b) ist zusammenfassend der Verlauf der drei Leistungen in Abhängigkeit von der Aussteuerung gezeigt. Daraus wird deutlich, daß auch im B-Betrieb bei ständiger Aussteuerung noch bessere Werte zu erhalten sind. Gemäß Gl. (35) beträgt dann die umgesetzte Leistung nur

$$P_{vs} = U_b I_{Cm}\left(\frac{1}{\pi} - \frac{1}{4}\right). \tag{42}$$

Daraus folgen ein größerer $I_{Cm,\,z}$ und ein noch kleinerer $R_{opt,\,z}$.

$$\left.\begin{aligned} I_{Cm,\,s} &= \frac{1}{\frac{1}{\pi} - \frac{1}{4}} \frac{P_u}{U_b} \approx \frac{1}{15} \frac{P_u}{U_b}, \\ R_{opt,\,s} &= 1\left(\frac{1}{\pi} - \frac{1}{4}\right)\frac{U_b^2}{P_u} \approx 0{,}137\,\frac{U_b^2}{P_u}, \\ r_s &= \frac{A}{2\left(\frac{1}{\pi} - \frac{A}{4}\right)} \leqq 7{,}32\,. \end{aligned}\right\} \tag{43}$$

Bei B-Stufen ist die Gleichstromaufnahme stark von der Aussteuerung abhängig. Deshalb ändert sich, wenn keine besonderen Maßnahmen getroffen werden (z.B. Stabilisierung), auch U_b mit der Aussteuerung. Dies hat Vor- und Nachteile. Bei Übertragung von Sprache und Musik liegt das Amplitudenspektrum meist nur bei relativ kleinen Werten. Deshalb können kurzzeitig auftretende Amplitudenspitzen von Elkos aufgefangen werden. Dies ist nicht möglich bei Entnahme einer *Dauerleistung*, dann sinkt U_b. Auf diese Weise können derartige Schaltungen so dimensioniert werden, als ob ständige Aussteuerung vorliegt, und z.T. sogar noch etwas darüber. Aus diesem Grunde wird die Endleistung solcher Stufen auch getrennt nach *Musik- und Sinusleistung* angegeben.

Da

$$\left(\frac{1}{\pi} - \frac{1}{4}\right) \approx \frac{1}{14{,}64} \approx 0{,}06831 \tag{43a}$$

beträgt, wird oft der Wert 1/15 für die Näherung angesetzt. Andererseits sind verschiedene Verluste z. B. durch die Restspannung am Transistor zu beachten, so daß auch der Wert von der zeitweiligen Aussteuerung als real angesehen werden kann.

7.4.3.3. Grafische Dimensionierung

Bei der Dimensionierung von Leistungsendstufen sind folgende *Parameter* interessant:

U_b^+ als Betriebsspannung je Transistor. Sie ist evtl. um die Restspannung gemäß Bild 7.4.1 zu reduzieren, deswegen das hochgestellte +.
I_{Cm} der maximal auftretende Strom
$P_\sim$ die maximal abgebbare Wechselleistung für beide Transistoren
P_u die Verlustleistung je Transistor
R_{opt} der optimale Arbeitswiderstand.

Die Verlustleistung je Transistor ist für viele Anwendungen besser durch

K den totalen Wärmewiderstand bis zur Umwelt

zu ersetzen. Dafür gilt für die Temperaturerhöhung (vgl. 5.7.6.)

$$\Delta T = K P_u . \tag{44}$$

Für Si kann etwa $\Delta T \approx 100$ K (für Ge nur 50 K) angesetzt werden. Unter Beachtung dieser Umstände gelten für den Gegentakt-B-Betrieb die grafischen Darstellungen von Bild 7.4.6.

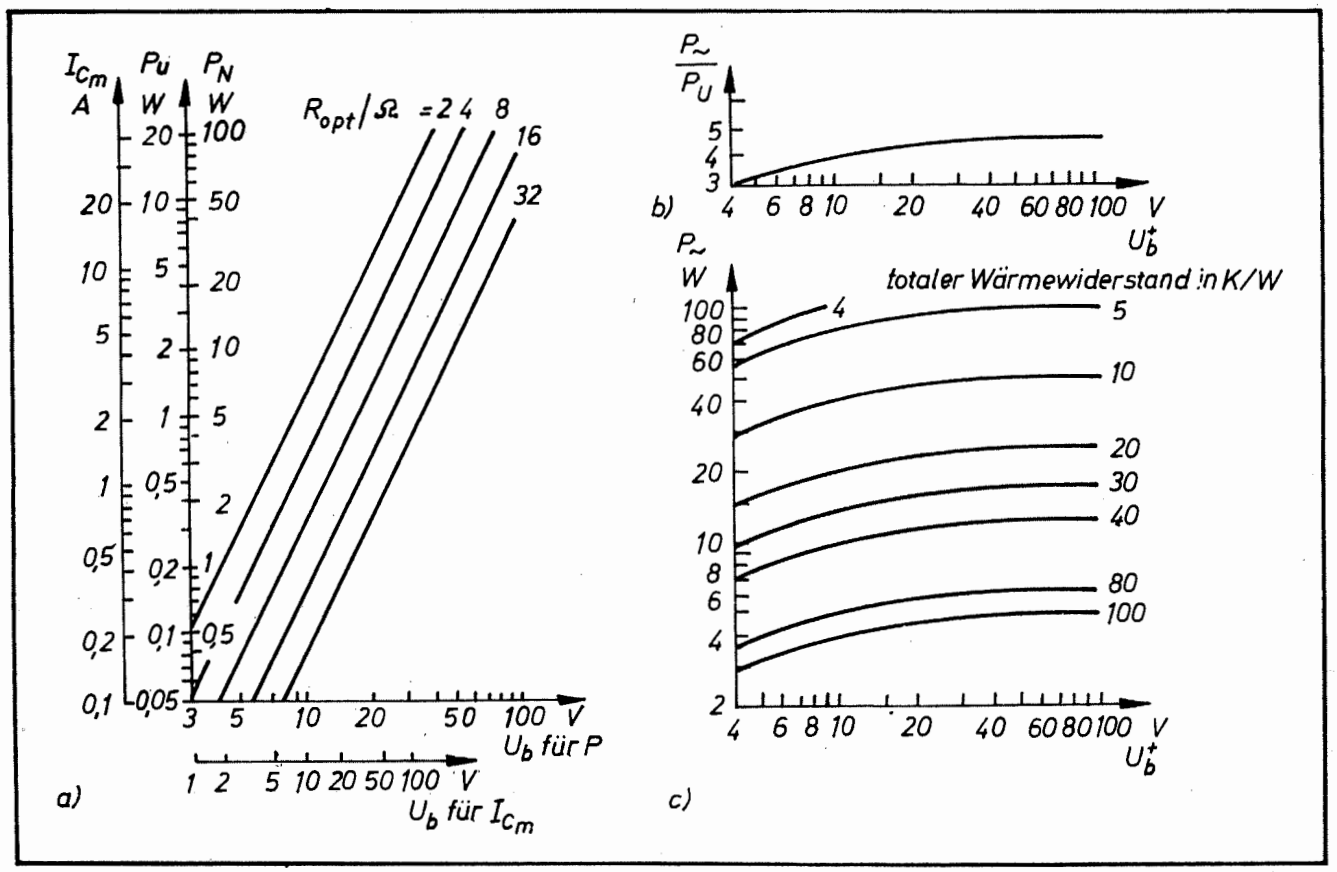


Bild 7.4.6. Grafisch dargestellte Zusammenhänge zur Gegentakt-B-Endstufe. U_b ist die Betriebsspannung und U_b^+ die um die Restspannung erhöhte

a) Einfluß des optimalen Arbeitswiderstandes
b) Verhältnis der erreichbaren Wechselleistung zur maximalen Verlustleistung je Transistor
c) Einfluß des totalen Wärmewiderstandes je Transistor auf die erreichbare Wechselleistung

7.4.3.4. Übernahmeverzerrungen; AB-Betrieb

Bisher wurden *ideal lineare Kennlinien* im Arbeitsfeld vorausgesetzt. Dies ist insbesondere aber an der Grenze für $I_C = 0$ nicht der Fall. Hier fließen u. a. Restströme und treten gerade die größten Nichtlinearitäten auf. Sie werden, wie z. B. die Bilder 7.1.3 und 7.1.6 zeigen, an den Grenzen der Aussteuerung sogar bei starker Gegenkopplung größer. Dadurch ist es für die Gegentaktendstufe notwendig, auch die *Eingangskennlinie einzubeziehen.* Dies zeigt Bild 7.4.7a). Daraus wird deutlich, daß gerade für kleine Signale — nämlich im sogenannten *Übernahmebereich* (des Stromes von einem Transistor zum anderen) — die größten Verzerrungen auftreten. Eine gewisse Kompensation dieses Effektes läßt sich erreichen, wenn ein definierter Ruhestrom beider Transistoren so eingeführt wird, daß sich im unteren Bereich die beiden Transistoren in der Leistung ergänzen und erst für große Ausgangsamplituden nur ein

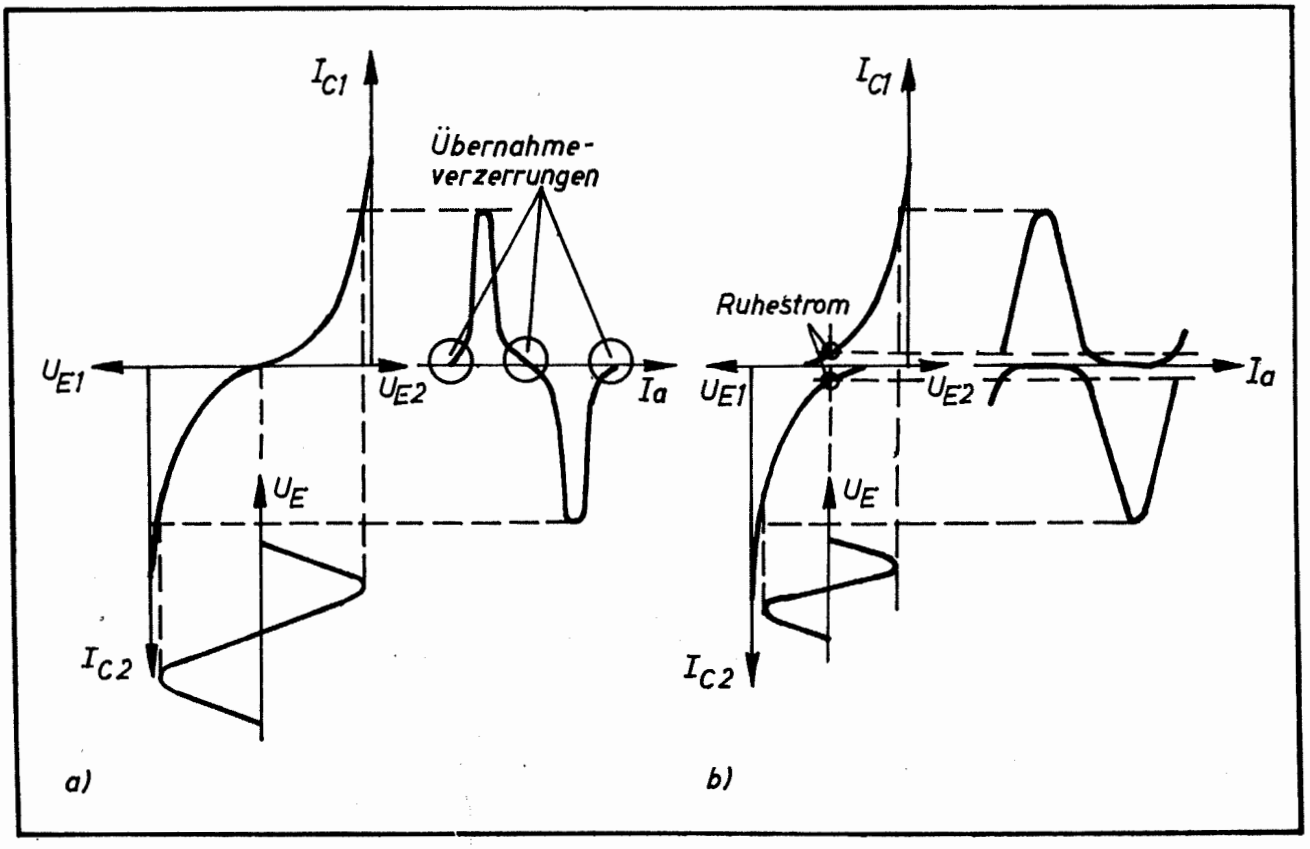


Bild 7.4.7. Zur Betrachtung a) der Übernahmeverzerrungen und b) des Ruhestromes bei Gegentakt-AB-Endstufen

Transistor allein wirksam wird. Ein derartiger Betrieb liegt bezüglich des Arbeitspunktes mit dem Ruhestrom I_{CR} zwischen dem A- und dem B-Betrieb und heißt daher AB-Betrieb. Bild 7.4.7b) zeigt, wie durch Addition der beiden stark nichtlinearen Teilströme ein relativ verzerrungsarmer Gesamtstrom entsteht. Es wird daraus aber auch verständlich, daß dieser *Ruhestrom* je nach dem gewünschten *Verzerrungsverlauf* bezüglich der Ausgangsleistung kritisch eingehalten werden muß. Insbesondere muß er gegen Schwankungen der Betriebsspannung, der Aussteuerung und der Temperatur stabilisiert werden.

7.4.3.5. Verzerrungen bei Gegentaktschaltungen

Bei gegebenem Arbeitspunkt, z. B. mit dem Ruhestrom I_{CR}, gelte für den Kollektorstrom die *Reihenentwicklung*

$$I_C = I_{CR}(1 + k_1 u_e + k_2 u_e^2 + k_3 u_e^3 + \ldots) . \tag{45}$$

Die Steuerspannungen (evtl. auch Ströme) sind für beide Transistoren gleich. Weiter seien beide Transistoren auf gleiche Konstanten k_n ausgesucht. Infolge der unterschiedlichen Typen *npn* und *pnp* wirkt auf beide Transistoren die Eingangsspannung mit entgegengesetztem Vorzeichen. Im *Arbeitswiderstand* tritt weiter nur der *Differenzstrom* auf. Für ihn gilt folglich

$$I_C = 2I_{CR}(k_1 u_e + k_3 u_e^3 + k_5 u_e^5 + \ldots) . \tag{46}$$

Es entfallen also sowohl der Gleichanteil als auch die geradzahligen Klirrkoeffizienten. Dies bedeutet, daß eine *Gegentaktschaltung* — gleich in welcher Betriebsart — *generell erheblich geringere Verzerrungen* als eine einzelne Endstufe besitzt. Dies ist ein erheblicher Vorteil aller Gegentaktstufen, der schon ohne Gegenkopplung auftritt.

7.4.3.6. Nichtsinusförmige Steuersignale

Alle bisherigen Ableitungen und Berechnungen erfolgten auf der Basis sinusförmiger Steuersignale gemäß Gl. (13) und (14). Dies hatte seinen Grund darin, daß für viele Anwendungen und Messungen die Sinussignale von grundlegender Bedeutung sind. Die dabei erhaltenen Ergebnisse werden aber z. T. sehr ungenau, wenn andere Signale, z. B. Dreieck- oder Rechtecksignale, auftreten. Dann sind die Integrale in den Gl. (15), (16), (33), (35) entsprechend anzusetzen. Das Ergebnis ist z. B. bei Rechteckimpulsen, daß nur die beiden Festpunkte der Aussteuerung auftreten. Im A-Betrieb werden dadurch sofort bessere Leistungsbilanzen erreicht. In dem einen Punkt ist nämlich die Spannung, im anderen der Strom klein. Die Zwischenwerte werden extrem schnell durchlaufen.

Bei dreieckförmigen Signalen tritt genau das Gegenteil auf, denn es werden, relativ gesehen, der Ruhearbeitspunkt länger und die Grenzwerte kürzer eingenommen. Diese beiden Fälle zeigen, daß die Leistungsbilanzen stark von der Kurvenform abhängen und damit bei fest vorgegebenen Kurvenformen speziell zu berechnen sind.

7.4.4. Praktischer Aufbau von Gegentaktendstufen

Folgende fünf *Problemkreise* sind bei schaltungstechnischen Realisierungen von Gegentaktendstufen neben den bisher behandelten Grundlagen allgemein von *Bedeutung*:

- Einstellung des stabilen und optimalen *Ruhestromes*;
- *Bootstrap-Schaltung* für Steuerstufe;
- Stabilisierung der *Mittenspannung* der Endstufe;
- Möglichkeiten bei *gleichartigen Endstufentransistoren*;
- Schutzschaltungen für Kurzschluß, thermische Überlastung, Gleichstromverschiebung und Einschaltverzögerung.

7.4.4.1. Einstellung des Ruhestromes

Werden die beiden Basisanschlüsse der Endstufentransistoren (s. z. B. in Bild 7.4.8) direkt verbunden, so fließt jeweils dann ein Basisstrom, wenn die Spannung U_{BE} des nicht aktiven Transistors überschritten wird. Dies läßt sich vermeiden, wenn zwischen beide Basisanschlüsse *zwei Dioden* und ein *einstellbarer Widerstand* P eingefügt wird. Über die beiden gleich großen Widerstände R_1 und R_2 kann dann mittels P der Ruhestrom gut eingestellt werden. Für einfache, insbesondere integrierte Schaltungen kann auch P entfallen. Die beiden Dioden sollten darüber hinaus in Wärmekontakt mit je einem Endtransistor sein. Dadurch läßt sich weitgehend das *„thermische Hochlaufen" des Ruhestromes* vermeiden. Es ist ein Kettenprozeß, der wie folgt auftritt: Die Endtransistoren seien durch zu hohen Ruhestrom oder durch vorherige Belastung erwärmt. Dies bedeutet, daß ihre Basis-Emitter-Spannung gesunken

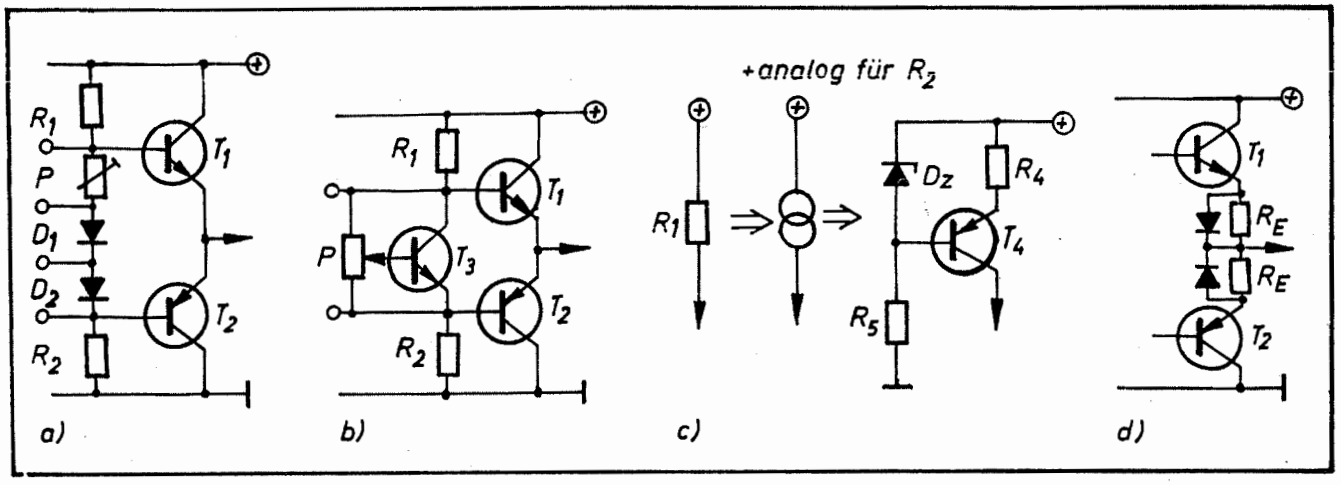


Bild 7.4.8. Teilschaltungen zur Einstellung und Stabilisierung des Ruhestromes. Die jeweils links befindlichen kleinen Kreise sind Anschlüsse für die Ansteuerung der Endstufe

a) mittels zweier Dioden und eines Regelwiderstandes P,
b) mittels einer einstellbaren Konstantspannungsquelle
c) Ersetzen des Widerstandes R_1 (bzw. analog R_2) durch eine Stromquelle mit der konkreten Realisierung durch T_4 und der ZENER-Diode D_Z
d) Zusätzliche Stabilisierung durch Emitterwiderstände, die evtl. mit Dioden überbrückt werden

ist und damit ein erhöhter Ruhestrom eintritt. Er erwärmt weiter die Transistoren und senkt erneut die Basis-Emitter-Spannung, was zu noch höherem Ruhestrom führt usw.

Eine weitere brauchbare Maßnahme hingegen ist das Einfügen von *Emitterwiderständen* gemäß Teilbild 7.4.8d). Sie wirken als Gegenkopplung und reduzieren so den Hochlaufeinfluß erheblich. Sie sollten deshalb möglichst groß sein. Anderseits fällt an ihnen aber auch eine Nutzwechselleistung ab. Deshalb sollten sie möglichst klein sein. Eine Überbrückung mit Kapazitäten kommt wegen der notwendigen extrem großen Werte in der Regel nicht in Betracht. Deshalb ist es günstig, sie einmal optimal zu bemessen und zum anderen für den aktiven Betrieb *mit Dioden zu überbrücken*. Durch Dioden treten zwar zusätzliche Verzerrungen auf, sie sind aber im Vergleich zu den anderen Verzerrungen meist sehr gering.

Für die Berechnung des Emitterwiderstandes wird von dem Temperaturkoeffizienten T_k der Basis-Emitter-Diode ausgegangen. Er liegt meist bei -2 bis $-2{,}5$ mV/K. Mit dem totalen thermischen Widerstand des Transistors R_{th} und der Betriebsspannung U_b gilt dann

$$R_E \approx T_k U_b R_{th} . \tag{47}$$

Für eine Leistungsendstufe (mit ca. 20 W) aus Transistoren mit 5 K/W und $U_b = 30$ V (je Transistor) folgt R_E zu 0,3 bis 0,4 Ω.

In der obigen Formel kann, wenn die Dioden D_1 und D_2 in thermischem Kontakt mit den Endstufentransistoren sind, für T_k die Differenz der Temperaturkoeffizienten der Dioden und Transistoren eingesetzt werden.

Eine andere Möglichkeit zur Einstellung des Ruhestromes zeigt Teilbild 7.4.8b). Hier wird die *einstellbare Konstantspannungsquelle* von Bild 7.3.5a) verwendet. Prinzipiell sind auch alle anderen Konstantspannungsquellen (7.3.1.3. und 7.3.4.4.) möglich. Ihr Temperaturkoeffizient sollte aber möglichst dem beider Basis-Emitter-Strecken der Endtransistoren entsprechen.

Eine Verbesserung besonders gegenüber Schwankungen der Betriebsspannung und bezüglich der Aussteuerung ist mit dem Ersatz der Widerstände R_1 und R_2 durch Stromquellen möglich. Dies zeigt Teilbild c) u. a. für eine bestimmte Ausführung der Stromquelle mit T_4. Alle anderen Varianten, wie sie u. a. in den Abschnitten 7.3.1.2., 7.3.4.3. und 7.3.5. behandelt sind, können ebenfalls eingesetzt werden.

7.4.4.2. Steuerstufen

Eine sehr einfache Steuerschaltung mit einem *npn*-Transistor zeigt Bild 7.4.9a). R_1 ist der Arbeitswiderstand, der wieder evtl. durch eine Stromquelle gemäß Bild 7.4.8c) ersetzt werden kann. Der Emitterwiderstand R_5 dient der Arbeitspunktstabilisierung und Gegenkopplung, wobei die Gegenkopplung für Wechselstrom auch durch C_5 vermieden werden kann.

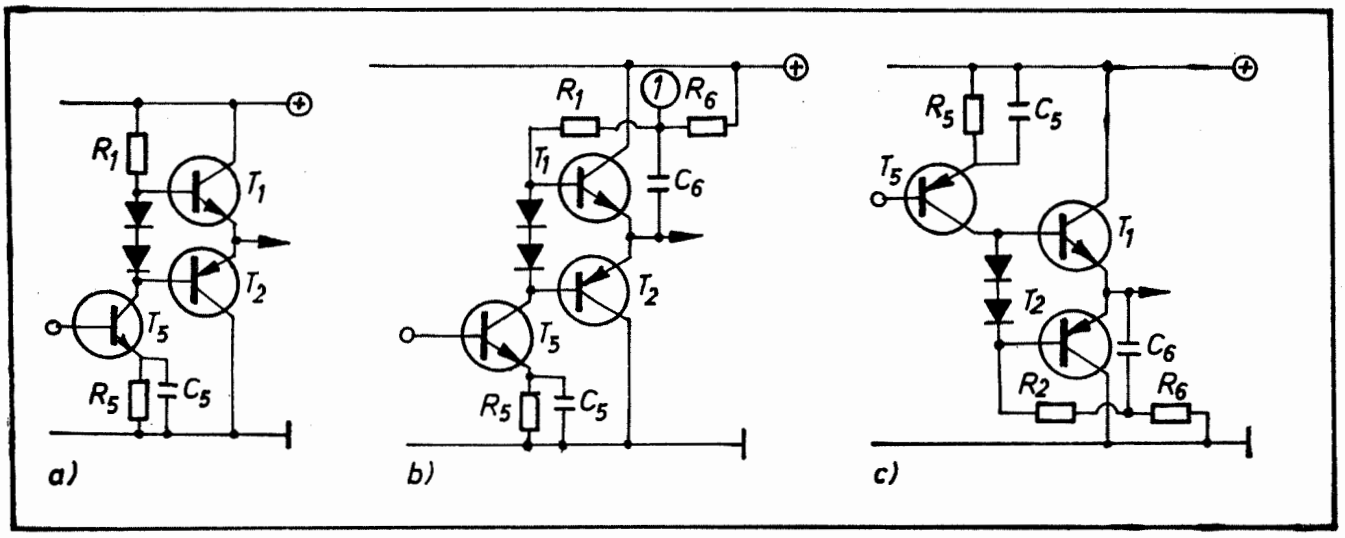


Bild 7.4.9. Steuerstufen für Gegentaktendstufen

a) Einfache Anordnung eines *npn*-Transistors T_5

b) Zusätzliche Bootstrap-Schaltung mit R_6, C_6

c) Steuerschaltung mit *pnp*-Transistor T_5 und Bootstrap-Schaltung R_6, C_6

Da die beiden Endtransistoren die volle Steuerspannung von praktisch Null bis zur Betriebsspannung benötigen, muß der Transistor T_5 sehr weit ausgesteuert werden. Dies hat erhebliche Nachteile, da sich bei diesen Spannungen bereits Verzerrungen zeigen können. Diese lassen sich beträchtlich reduzieren, wenn gemäß Bild 7.4.9b) eine *Bootstrap-Schaltung* aus R_6 und C_6 angewendet wird. So wird nämlich die Betriebsspannung für die Vorstufe am Ende von R_1, am Punkt ①, voll mit der Ausgangsspannung mitgeführt. Dadurch arbeitet T_5 immer in einem optimalen Arbeitspunkt.

Die zum Teilbild b) analoge Schaltung mit einem *pnp*-Steuertransistor zeigt Teilbild c).

7.4.4.3. Mittenspannung der Endstufe

Die insgesamt hohe Verstärkung von Endstufen mit ihren gleichstromgekoppelten Vorverstärkern läßt *Driftgrößen*, u. a. durch Temperatur- und Betriebsspannungsschwankungen, störend in Erscheinung treten. Betreffen sie die *Mittenspannung der Endstufe*, so treten auch Leistungsverluste und bei einer Gleichstromanschaltung der Last, z.B. gemäß Bild 7.4.4b), Gleichströme durch die Last auf. Aus diesen und weiteren Gründen ist es günstig, die Mittenspannung zu stabilisieren. Eine oft angewandte und besonders effektive Schaltung arbeitet gemäß Bild 7.4.10 mit einem *Differenzverstärker* aus T_6 und T_7, einschließlich der zugehörigen Konstantstromquelle aus T_8. Der Widerstand R_9 koppelt einen Stromanteil von der Mittenspannung auf den Differenzverstärker zurück. Die Mittenspannung ist durch die Widerstände R_7 und R_8 einstellbar. T_5 ist die Steuerstufe gemäß Bild 7.4.9, einschließlich der Bootstrap-

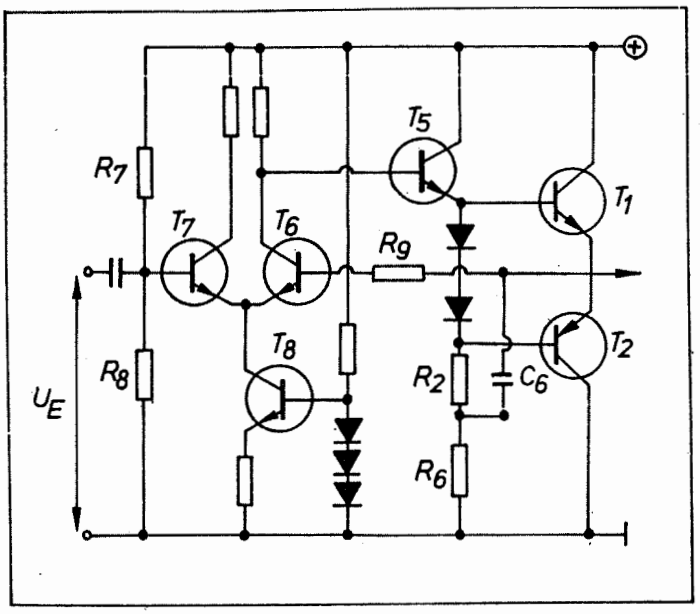


Bild 7.4.10. Schaltung einer Endstufe mit automatischer Spannungsmitteneinstellung. Der Differenzverstärker aus T_6 und T_7 sowie die Stromquelle aus T_8 wird über R_9 vom Mittenpotential angesteuert. T_5 ist die Steuerstufe entsprechend Bild 7.4.9c)

Schaltung aus R_6 und C_6. Die Emitterbauelemente R_5 und C_5 können hier entfallen, da die Stabilisierung über R_9 viel stärker wirkt.

Für den Differenzverstärker können alle u. a. im Abschnitt 7.3.3. behandelten Varianten verwendet werden.

7.4.4.4. Quasikomplementäre Endstufen

Alle bisherigen Endstufen benötigen zwei sehr ähnliche, komplementäre Endtransistoren. Sie sind schwierig herzustellen und darüber hinaus noch auszusuchen. Deshalb sind verschiedene Varianten mit *zwei gleichartigen Endtransistoren*, z.B. vom Typ *npn*, entwickelt worden. Zwei Beispiele hierzu zeigt Bild 7.4.11. Bei der Schaltung von Teilbild a) wird die Tatsache ausgenutzt, daß *komplementäre* DARLINGTON-*Transistoren* sich so verhalten, wie der erste, also der mit kleiner Leistung (vgl. 7.3.2.1.). Der Vorstufentransistor T_5 entspricht der Treiberstufe in Bild 7.4.8a).

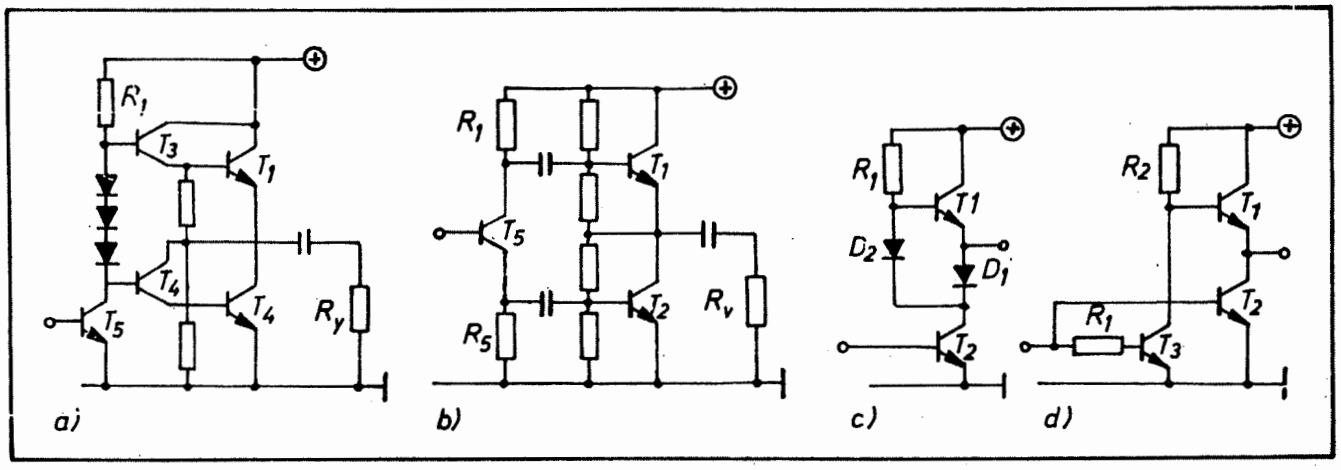


Bild 7.4.11. Quasikomplementäre Endstufen

a) mit Darlington-Transistoren ($T_3 = npn$, $T_4 = pnp$),
b) über eine Phasenumkehrstufe als Treiber,
c) mittels Steuerung über zwei Dioden
d) Phasenumkehr der Steuerung von T_1 über T_2
Die Schaltungen c) und d) haben vor allem für digitale Schaltungen Bedeutung

Deshalb kann hier auch die *Bootstrap-Schaltung* gemäß Bild 7.4.9b) zusätzlich genutzt werden. Im Gegensatz zu den dortigen zwei Dioden für den temperaturstabilen Ruhestrom müssen bei den DARLINGTON-Stufen infolge der zu kompensierenden Basis-Emitter-Spannung jetzt drei Dioden gewählt werden.

Ein zweites Prinzip für gleichartige *npn*-Endstufentransistoren zeigt Bild 7.4.11b). Die Steuerung des oberen Endstufentransistors erfolgt hier in der gewohnten Weise, wobei statt R_1 wiederum Bootstrap oder Stromquelle eingesetzt werden können. Für den unteren Transistor wird jedoch durch T_5 die Eingangsspannung nicht gegen-, sondern gleichphasig über R_5 bereitgestellt. Der Transistor wirkt bezüglich der beiden Endtransistoren wie eine Phasentrennschaltung, z.B. analog Bild 7.3.2a) und d). Ein Nachteil dieser Schaltung ist die komplizierte Ruhestromeinstellung von T_1 und T_2. Eine diesbezüglich geradezu trivial erscheinende Schaltung zeigt Bild 7.4.11c). Die beiden Dioden D_1 und D_2 sorgen dafür, daß beide Transistoren sowohl den gleichen Ruhestrom als auch die gegeneinander abgestimmte Aussteuerung erhalten. Hierbei treten jedoch Verzerrungen und Unsymmetrien in der Aussteuerung auf. Deshalb hat diese Schaltung vor allem *bei digitalen Schaltungen* Bedeutung. Das gleiche gilt für die Schaltung von Bild 7.4.11d), wobei T_2 direkt und T_1 mit Phasendrehung über T_3 angesteuert wird.

7.4.4.5. Schutzschaltungen

Folgende *vier Arten* von Schutzschaltungen haben bei Endstufen größere Bedeutung:

- Schutz vor *thermischer Überlastung*. Hierbei wird mit temperaturabhängigen Widerständen, Dioden oder Transistoren, die sich in thermischem Kontakt mit den Endtransistoren befinden, die Temperatur gemessen. Sofern ein kritischer Wert erreicht wird, kann die Steuerleistung reduziert oder gar die Endstufe abgeschaltet werden. Bei leistungsstarken, integrierten Endverstärkern befinden sich derartige komplette Schaltungen mit auf dem Chip und begrenzen so die Belastbarkeit des Schaltkreises (vgl. 7.4.6.).
- *Schutz für Verbraucher*, wie z.B. Lautsprecher, vor unzulässigen *Gleichspannungen*. Sie können entstehen, wenn eine Mittenpotentialverschiebung auftritt oder wenn bei Schaltungen gemäß Bild 7.4.4b) eine Hälfte der Spannungsquelle ausfällt.
- Die *Einschaltverzögerung* entlastet Endstufe und angeschlossene Verbraucher von den meist unregelmäßigen Schwankungen der Ausgangsspannung einer Endstufe kurz nach dem Einschalten. Sie wird in der Regel mit einer speziellen Schaltung und einem Relais realisiert.
- Eine *elektrische Überlastung* der Endtransistoren kann durch zu hohen Strom bzw. zu hohe Spannung oder Leistung entstehen. Diesbezügliche Schutzschaltungen werden im folgenden an den drei Beispielen des Bildes 7.4.12 besprochen.

In Bild 7.4.12a) ist, im Vergleich zu Bild 7.4.8d), zusätzlich in dem Emitterkreis des Endtransistors T_1 der Widerstand R_{13} eingeschaltet. Sobald die an ihm abfallende Spannung den Wert U_{BE} des zusätzlichen Transistors T_8 überschreitet, wird dieser leitend und übernimmt einen Teil des Steuerstromes, der für T_1 bestimmt ist. Die Folge davon ist, daß der *Strom* I_C auf etwa

$$I_{C,\max} = \frac{U_{BE,T8}}{R_{13}} \approx \frac{0{,}6\ \text{V}}{R_{13}} \tag{48}$$

begrenzt wird. Größere Ströme sind weder bei größerer Aussteuerung noch bei zufälligem Kurzschluß des Ausganges zu entnehmen. Eine solche Endstufe ist daher

kurzschlußsicher. Allerdings muß die Schaltung aus R_{13}, R_{14} und T_8 analog für den unteren Transistor T_2 (z.B. in Bild 7.4.8d)) eingefügt werden.

Eine andere Möglichkeit zur Strombegrenzung besteht bei Verwendung der *Stromquelle* von Bild 7.4.8c). Entsprechend Bild 7.4.12b) ist dann lediglich R_{15} in den Kollektorkreis des Transistors T_1 einzufügen und über eine Diode mit R_4 zu verbinden. Die Begrenzung tritt hier ein, wenn der Spannungsabfall an R_{15} größer als die *Z-Diodenspannung* U_Z ist. Dann wird ein Teil des für T_1 bestimmten Basisstromes über die Diode D und R_{15} abgeleitet. Für den maximalen Strom gilt hier

$$I_{C,\max} \approx U_Z/R_{15}\,. \tag{49}$$

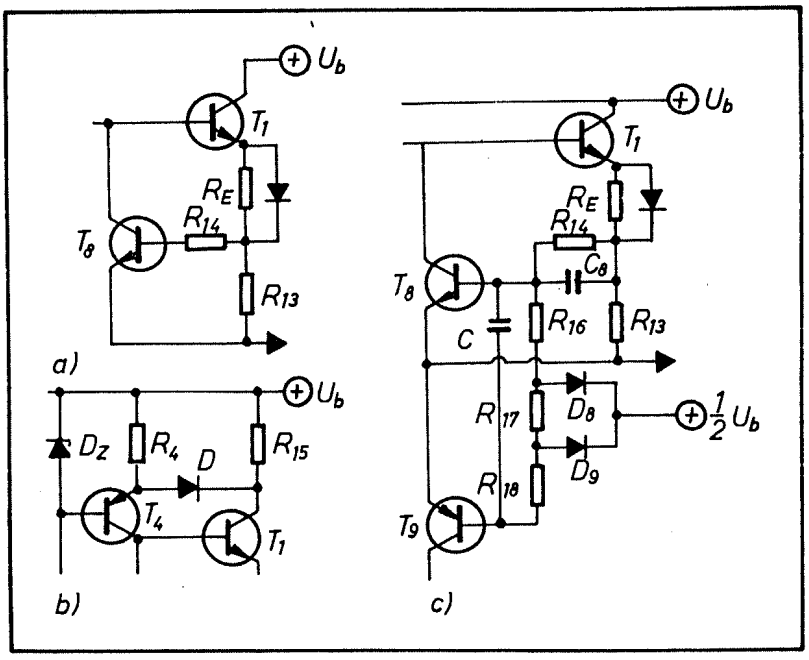


Bild 7.4.12. Schutzschaltung für zu hohen Strom bzw. zu hohe Leistung

a) Strombegrenzung über Spannungsabfall an R_{13}

b) Strombegrenzung über Spannungsabfall an R_{15}

c) Leistungsbegrenzung durch zusätzliche Addition eines Teils der Ausgangsspannung mittels der Dioden D_8 und D_9

Genaugenommen muß nicht nur der Strom durch den Transistor begrenzt werden, sondern auch die *Leistung*. In erster Näherung läßt sich dies mit der Schaltung von Bild 7.4.12c) realisieren. T_9 ist darin der T_8 entsprechende Transistor für den unteren Leistungstransistor T_2 gemäß den Bildern 7.4.8d) und 7.4.12a). Die Dioden D_8 und D_9 liegen auf dem festen halben Gleichpotential (bzw. in Bild 7.4.4b) auf der Mitte der angezapften Spannungsquelle). Durch die Schwankungen vom Ausgang (Pfeil) werden so über R_{16}, R_{17}, R_{18} spannungsproportionale Anteile bez. der Basis der Transistoren T_8 bzw. T_9 addiert, und die Strombegrenzung über R_{13} ist zusätzlich spannungsabhängig.

7.4.5. Ausgewählte integrierte Endstufen

7.4.5.1. A 211 D

Die wichtigsten Daten des 1-W-Verstärkerschaltkreises enthält Tab. 7.4.1. Er ist gemäß Bild 7.4.13a) in einem 14poligen DIL-Gehäuse untergebracht, bei dem die Anschlüsse 3 bis 5 und 10 bis 12 zur besseren Wärmeableitung zusammengefaßt sind. Die vollständige Innenschaltung des Schaltkreises zeigt Bild 7.4.13b). Sie besteht

Tabelle 7.4.1. Typische Kennwerte der A 211

Betriebsspannung	4,2 ... 15 V
Größte Betriebsspannung ohne Eingangssignal	18 V
Zulässige Verlustleistung bei 25 °C	1 W
Verlustleistung mit Kühlfläche	1,35 W
Betriebstemperatur	−10 ... +70 °C
Größter Ausgangsstrom	1 A
Typischer Ruhestrom	3,5 mA
Minimale offene Leerlaufverstärkung	60 dB
Empfohlene Verstärkung	48 dB
Typischer Lastwiderstand	4 bis 12 Ω
Eingangswiderstand ohne Gegenkopplung	390 kΩ
mit Gegenkopplung auf 48 dB	5 MΩ

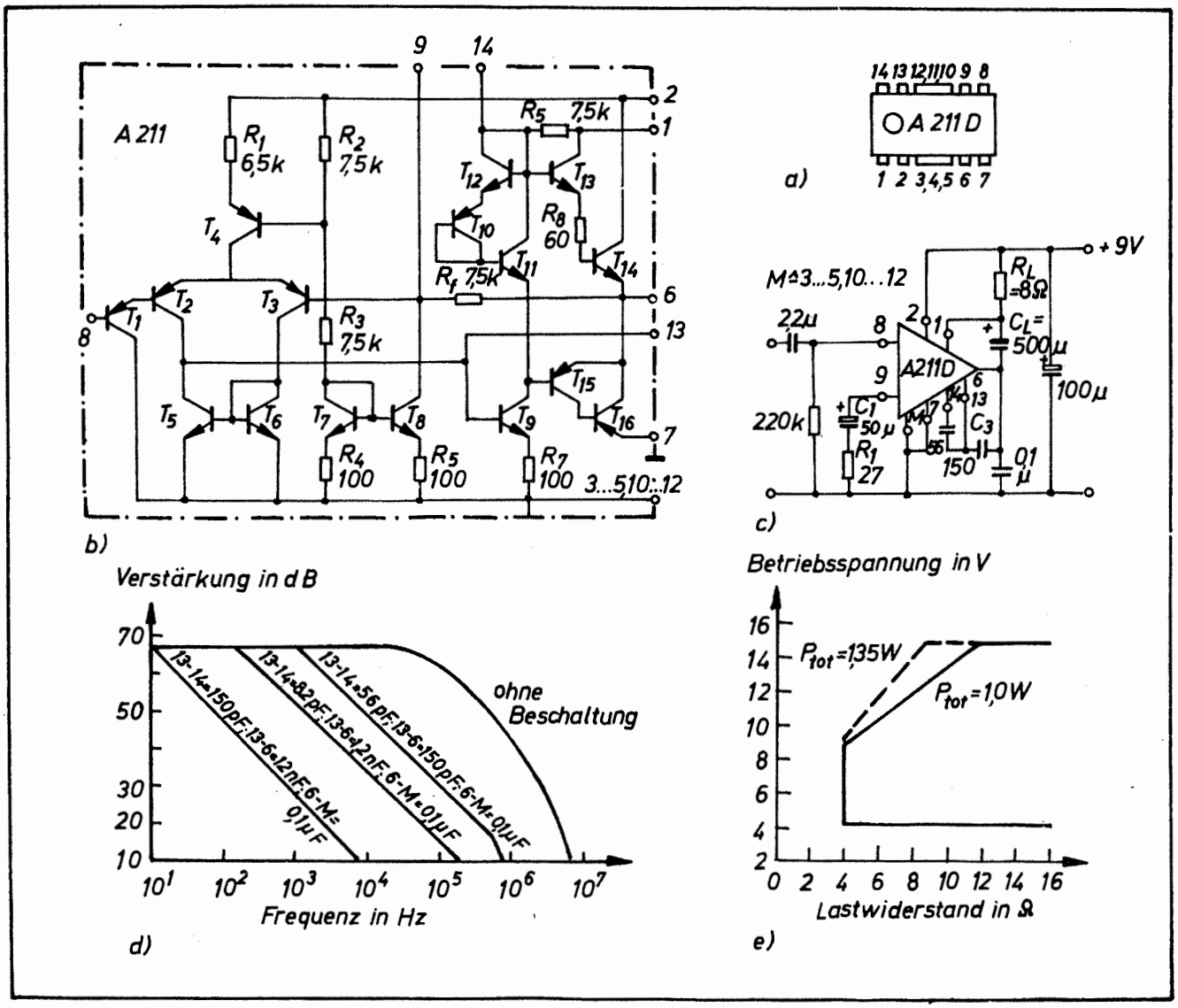


Bild 7.4.13. Zum integrierten Schaltkreis A 211 D (T_{16} ist vom Typ *npn*)

a) Sockelschaltung
b) Vollständige Innenschaltung
c) Meßschaltung
d) Frequenzgänge ohne Gegenkopplung bei verschiedenen Korrekturen
e) Belastungsgrenzen

aus drei Teilschaltungen:

— *Eingangsdifferenzstufe,*
— *quasikomplementäre Endstufe,*
— *Netzwerk zur Regelung der Ausgangsmittenspannung.*

Die *Eingangsstufe* besteht *wiederum aus drei Teilen*:

— eigentlicher Differenzverstärker aus T_2 und T_3, wobei T_2 noch T_1 als DARLINGTON-Transistor zur Erhöhung des Eingangswiderstands vorgeschaltet ist,
— Konstantstromquelle aus T_4,
— Stromspiegel aus T_5 und T_6. Er dient als hoher differentieller Arbeitswiderstand, damit die Schaltung eine große Verstärkung erreicht, und der Phasenaddition.

Die *quasikomplementäre Endstufe* beginnt mit der Treiberstufe T_9, die mit dem Emitterwiderstand R_7 gegengekoppelt ist. Das Diodennetzwerk T_{10}, T_{11}, T_{12} stockt die Treiberausgangsspannung für den oberen Ausgangstransistorzweig so auf, daß die Endstufe voll ausgesteuert werden kann. Es bestimmt außerdem den Ruhestrom der Endstufe. Während die untere Hälfte mit den in DARLINGTON-Schaltung betriebenen Transistoren T_{15}, T_{16} üblich ist, stellen die oberen beiden T_{13} und T_{14} einige Besonderheiten dar. Durch die Kombination *pnp*- und *npn*-Transistor wirken sie wie eine *pnp*-Stufe. Der Anschluß 1 dient ferner dazu, hier nach dem Bootstrap-Prinzip die Betriebsspannung über die Ausgangsspannung aufzustocken.

Die *Einstellung der Mittenspannung* erfolgt mittels des Stromspiegels T_7, T_8 und über die Widerstände R_f bzw. R_2, R_3. Zugleich führt R_f als Gegenkopplung zurück zum Differenzeingangsverstärker. Dadurch wird auch der Eingangswiderstand der Schaltung weiter erhöht.

Zur *Frequenzkompensation* können (vgl. 7.5.) jeweils zwischen 13 und 14, 13 und 6 sowie 6 und Masse Kondensatoren geschaltet werden. In der Meßschaltung 7.4.13c) sind die Werte so gewählt, daß die 56 pF und 0,1 µF zur Vermeidung von HF-Störschwingungen dienen, während die 150 pF (C_3) die obere Grenzfrequenz festlegen. Der Widerstand R_1 mit 27 Ω legt die Verstärkung mit dem inneren Widerstand R_f auf 48 dB fest.

Um die realen Frequenzgänge auch für andere Anwendungszwecke zu erkennen, sind im Bild 7.4.13d) die Frequenzgänge für einige Beschaltungen ohne Gegenkopplung dargestellt.

Bild 7.4.13e) zeigt schließlich die Leistungsgrenzen bezüglich Lastwiderstand und Betriebsspannung.

7.4.5.2. A 210 D bzw. K

Die wichtigsten Eigenschaften des Schaltkreises A 210 D bzw. K zeigt Tab. 7.4.2. Er wird in der Ausführung ohne Kühlkörper (D) gemäß Bild 7.4.14a) bzw. mit aufgepreßtem Kühlkörper (K) gemäß Bild 7.4.14b) gefertigt. Seine vollständige Innenschaltung zeigt Bild 7.4.14c) und die Standardbetriebsschaltung Teilbild d). Die Schaltung enthält die folgenden *fünf Funktionsblöcke*:

— Vorverstärker
— Zwischenverstärker
— Endverstärker
— Arbeitspunkteinstellung und Mittenregelung
— Wärmeschutzschaltung.

Der *Vorverstärker* beginnt mit den DARLINGTON-Transistoren T_1 und T_2, wodurch ein hoher Eingangswiderstand erreicht wird. T_2 ist dabei eine stromgegengekoppelte Emitterstufe (Anschluß 6 mit Standardbeschaltung 56 Ω, 500 µF in Reihe gegen

Tabelle 7.4.2. Kennwerte des A 210 D/K bei $T_a = 25\ °C - 5\ K$, $U_B = 15\ V$, $R_i = 50\ m\Omega$, $R_L = 4\ \Omega$, $f = 1\ kHz$

Betriebsspannung	U_B	4 bis 20 V
Eingangsgleichspannung	U_I	−3 bis +5 V
Eingangsstrom	I_I	2 mA
Ausgangsspitzenstrom	I_{OM}	2,5 A
Gesamtverlustleistung		
A 210 D	P_{tot}	1,3 W
A 210 K	P_{tot}	5 W
Gesamtstromaufnahme		
$U_B = 9\ V$, $U_I = 0$	I_{S0}	7,4 bis 10,0 mA
$U_B = 15\ V$, $U_I = 0$	I_{S0}	10,1 bis 14,0 mA
Mittenspannung	U_{00}	6,7 bis 8,3 V
Eingangsgleichstrom	I_{I0}	0,2 bis 0,5 µA
Offene Spannungsverstärkung bei $P_o = 1\ W$	V_{U0}	70,2 bis 73,5 dB
Geschlossene Spannungsverstärkg. bei $P_o = 2{,}5\ W$	V_U	34,0 bis 40,0 dB
Ausgangsleistung bei $k = 10\%$	P_o	5,45 bis 6,57 W
Eingangswechselspannung	U_i	30 bis 70 mV
Geräuschspannung	U_n	0,5 bis 1,2 mV
Obere Grenzfrequenz	f_o	31 bis 50 kHz

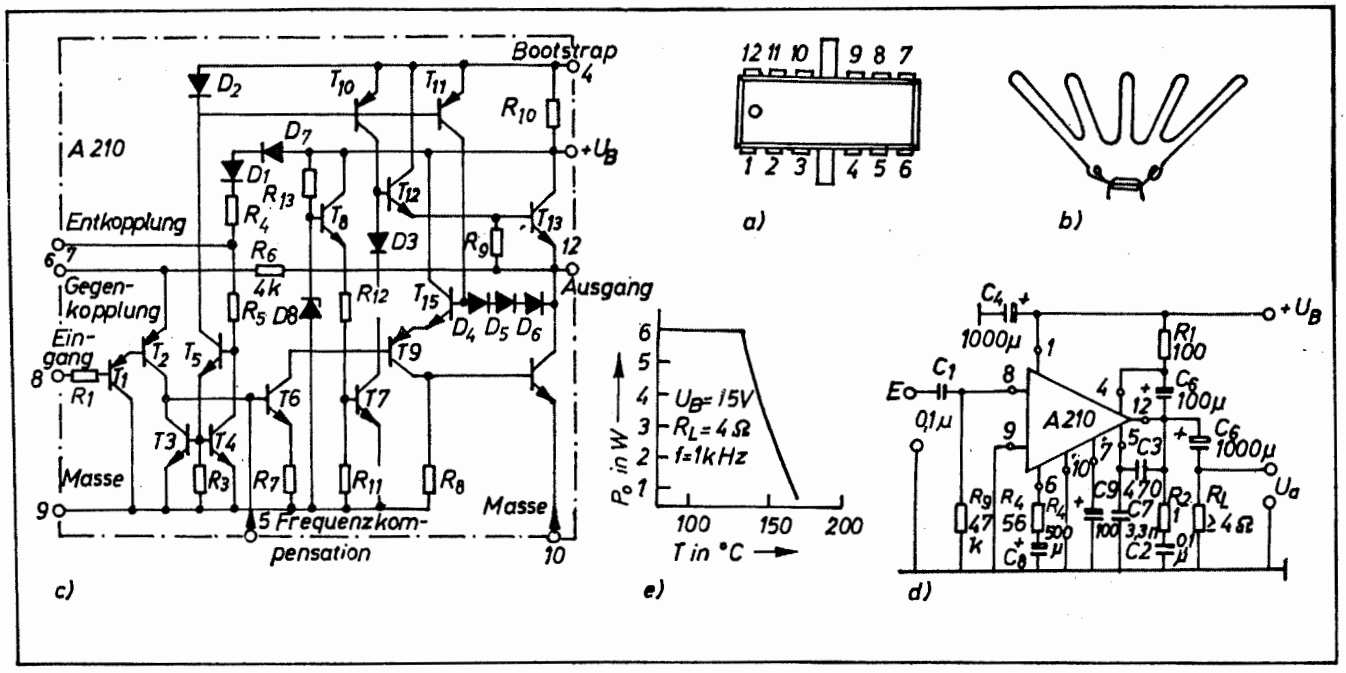


Bild 7.4.14. Schaltkreis A 210 D/K

a) Sockelschaltung ohne Kühlkörper (D) mit den beiden Kühlfahnen
b) Aussehen des aufgepreßten Kühlkörpers bei der Variante K
c) Vollständige Innenschaltung
d) Standardbetriebsschaltung
e) Wirkung der Wärmeschutzschaltung

Masse). Ihr Arbeitswiderstand wird durch eine Konstantstromquelle aus T_3 bis T_5 und R_2 bis R_5 gebildet.

Der *Zwischenverstärker* besteht aus T_6. Er ist über den Emitterwiderstand R_7 gegengekoppelt.

Die *Endstufe* enthält oben den *npn*-DARLINGTON-Transistor aus T_{12} und T_{13}, unten den *pnp*-DARLINGTON-Transistor aus T_9 und T_{14}. Der Ruhestrom — *Arbeitspunkt* — wird bezüglich des Übernahmeknicks einmal durch die Diode D_3, dann durch die Stromspiegelbank aus D_2, T_{10} und T_{11} versorgt und schließlich mit T_{11} und D_4 bis D_6 weiter stabilisiert. Die *Mittenstabilisierung* erfolgt über die Gegenkopplung mittels R_6. Für ihre Wirksamkeit muß die Gleichstromgegenkopplung möglichst groß gehalten werden. Deshalb ist C_8 in Bild 7.4.14d) notwendig.

Die *Wärmeschutzschaltung* wird in den Transistoren T_7 und T_8 realisiert. D_8 ist die Referenzspannung und stellt einen als Z-Diode geschalteten Transistor dar. Sein positiver Temperaturgang wird durch den negativen Temperaturgang von U_{BE} des Transistors T_7 kompensiert. Dadurch entsteht am Emitter von T_8 eine von der Temperatur unabhängige Spannung. Daraus wird mittels R_{11}, R_{12} an der Basis von T_7 eine konstante Spannung von ca. 0,4 Volt eingestellt. Bei zunehmender Chiptemperatur verringert sich U_{BE} des Transistors T_7, und wenn eine Schwellspannung überschritten wird, zieht er die Basis von T_9 gegen Masse. Die Wirkung dieser Wärmeabschaltung demonstriert Bild 7.4.14e) bezüglich der maximal erreichbaren Wechselleistung.

7.4.5.3. MDA 2010 bzw. MDA 2020

Die beiden tschechischen Leistungsschaltkreise können mit maximal ± 18 Volt (MDA 2010) bzw. ± 22 Volt (MDA 2020) betrieben werden und gestatten dann, Wechselleistungen von ca. 10 bzw. 20 Watt zu erzeugen. In beiden Fällen ist eine sehr effektive Kühlfläche vorzusehen, sonst wird die Temperaturbegrenzung früher wirksam. Das Prinzip der Schaltkreise zeigt Bild 7.4.15a). Für viele Anwendungen ist es nützlich, daß ein kompletter Differenzverstärker am Eingang existiert. Ihm folgt ein DARLINGTON-Zwischenverstärker. Den dann folgenden interessanten Teil zeigt Bild 7.4.15b) in vollständiger Schaltung. Die Treiberstufe T_1 steuert die beiden DARLINGTON-Endstufen aus T_2 und T_3 bzw. T_4 und T_5. Die *Temperaturbegrenzung* erzeugt mittels D_4 und T_6, T_7 eine konstante Emitterspannung von ca. 0,34 Volt. Überschreitet der Chip eine Temperatur von 150 °C, so werden beide Transistoren leitend und wirken auf die Begrenzungsschaltungen für beide Endstufen ein. Dies sei für den oberen Leistungstransistor erklärt: T_8 ist hier der Transistor, welcher der Endstufe den Steuerstrom entzieht. Die Leistungsbegrenzung beginnt mit einem an R_8 (Bonddraht ca. 10 mΩ) dem Strom proportionalen Spannungsabfall. Er wird über einen zweiten Bonddraht R_9 zurückgeführt. Die Spannungskomponente wird mittels R_5 eingekoppelt. Beide Komponenten wirken addiert auf T_9 ein. Sie addieren sich wiederum auf T_8 zur Temperaturbegrenzung und werden beim Überschreiten von Grenzbedingungen wirksam.

Den Aufbau eines Leistungsverstärkers mit *Netzteil* zeigt Bild 7.4.15c). Durch die Mittenanzapfung des Trafos werden gleichzeitig die positive und die negative Spannung erzeugt (± 17 V). Die Verstärkung wird durch die Gegenkopplungswiderstände R_1 und R_2 eingestellt. Es werden 30 dB erreicht, so daß der Verstärker mit 260 mV voll ausgesteuert werden kann. Die *Frequenzkompensation* erfolgt zur Stabilitätserhöhung mit C_4 und R_4, C_8. C_3, C_6 sowie VD 1 und VD 2 sind *Störschutzglieder*.

In Bild 7.4.15 wird schließlich noch gezeigt, wie durch Brückenschaltung von zwei Schaltkreisen die Leistung verdoppelt werden kann. Hierbei ist es besonders vorteilhaft, daß die Eingangsstufe ein echter Differenzverstärker ist. Soll dasselbe Prinzip z.B. bei A 210 oder A 211 angewendet werden, so ist für einen der Eingänge eine Phasenumkehrstufe notwendig.

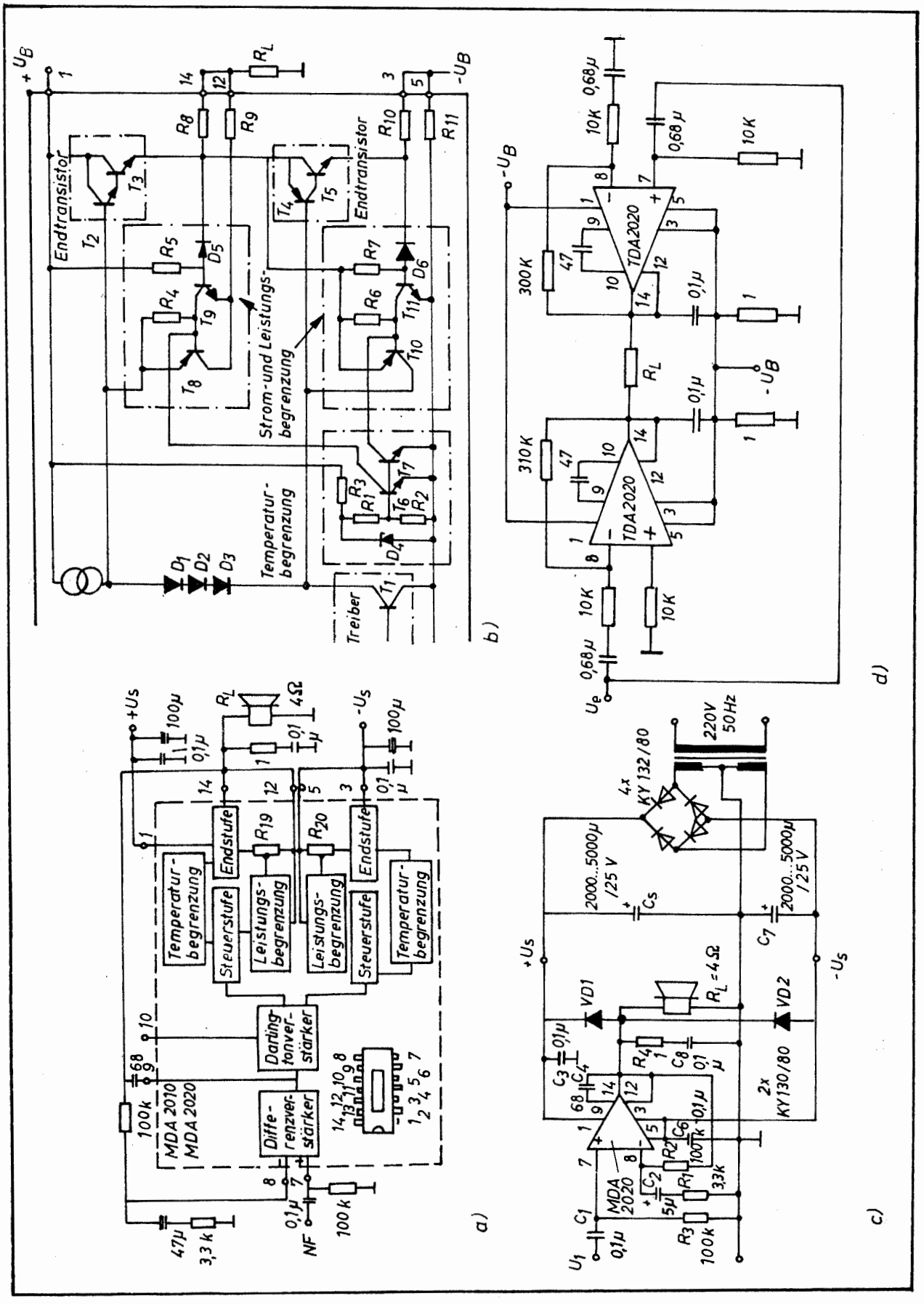


Bild 7.4.15. Schaltkreis MDA 2010/MDA 2020

a) Prinzipaufbau und Sockelschaltung
b) Teilschaltung aus Treiber, Endstufen und Schutzschaltungen
c) Aufbau eines 15-Watt-Verstärkers mit der MDA 2020
d) Aufbau eines 30-Watt-Verstärkers mit zwei MDA 2020

7.4.6. Weitere Prinzipien

Zwei Probleme aller üblichen Endstufen haben dazu geführt, ständig nach besseren Lösungen zu suchen:

- Der *Wirkungsgrad* erreicht selbst bei B-Betrieb maximal 78%.
- Die *Linearität* läßt in der Regel Wünsche offen.

Im folgenden sollen einige Varianten vorgestellt werden.

7.4.6.1. C-Endstufe

Der Übergang von der A- zur AB und zur B-Endstufe geht mit wachsendem Wirkungsgrad einher. Dies beruht darauf, daß die relative Zeit, in welcher ein Transistor Strom führt, also der Stromflußwinkel Θ (vgl. Bild 5.4.4c)), von 360° auf 180° abnimmt. Deshalb liegt der Gedanke nahe, ihn weiter zu senken. Dies bedeutet, daß ein Transistor nicht mehr während einer Halbwelle (wie im B-Betrieb) Strom führt, sondern eine kürzere Zeit. Von einer Eingangssinusschwingung treten dann nur kurze Impulse mit einem hohen Oberwellenanteil auf. An eine breitbandige, verzerrungsfreie Signalübertragung ist bei diesem Prinzip nicht mehr zu denken. Mit einem schmalbandigen Resonanzkreis können auf diese Weise jedoch für die eine Frequenz und schmale Seitenbänder hohe Leistungswirkungsgrade erreicht werden. Dieser C-Betrieb mit $\Theta < 180°$ hat daher nur für Sendeverstärker Bedeutung erreicht.

7.4.6.2. D-Verstärker

In der *Digitaltechnik* ist es üblich, einen Transistor nur in zwei Zuständen zu betreiben:

- *gesperrt*: stromlos mit voller Betriebsspannung U_b;
- *leitend*: maximaler Strom bei kleinstmöglichem Spannungsabfall, z.B. U_R in Bild 7.4.1a).

In beiden Zuständen entsteht im Transistor *nur sehr geringe Verlustleistung*. Mit speziellen Modulationen, z.B. der *Pulslängenmodulation* PDM (7.2.2., Bild 7.2.5 und 7.3.8), lassen sich dennoch analoge Signale übertragen. Das Prinzip im Vergleich zum C-Verstärker zeigt Bild 7.4.16a). Die Amplitudenbegrenzung, welche beim C-Verstärker zu den kleinen Stromflußwinkeln führt, ist durch den Modulator ersetzt. Die Verstärker werden in beiden Fällen möglichst nur in den beiden extrem leistungsarmen Zuständen leitend und gesperrt betrieben. Beim C-Verstärker wird die Grundwelle

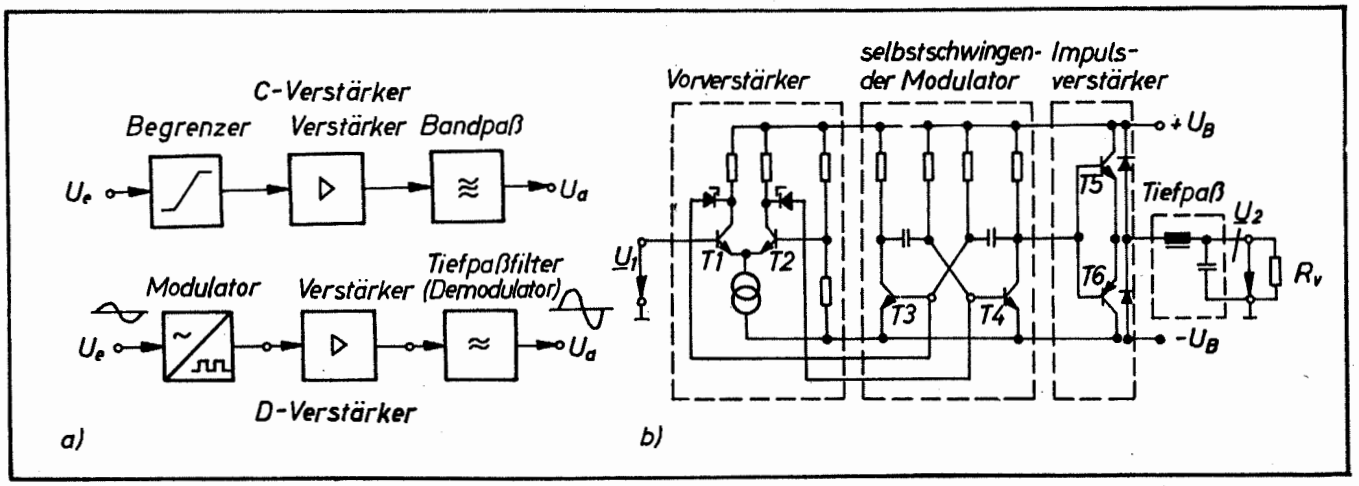


Bild 7.4.16. Zum D-Verstärker

a) Prinzipieller Aufbau eines D-Verstärkers im Vergleich zum C-Verstärker

b) Prinzipieller Aufbau eines integrierten D-Verstärkers

herausgesiebt, während der D-Verstärker durch einen Tiefpaß nur das mittlere Strom-Pausen-Verhältnis, also die Gleichkomponente des PDM-Signals, passieren läßt und so das PDM-Signal auf einfachste Weise demoduliert.

Für den *Wirkungsgrad* einer solchen Endstufe gilt in guter Näherung

$$\eta = 1 - U_b/U_R \tag{50}$$

Er erreicht also fast 100%. Bezüglich des Verhältnisses von erreichter Wechselleistung zur Verlustleistung der Transistoren werden so r-Werte von 10 bis 100 möglich, während im B-Betrieb nur ca. 5 (Gl. (41): $\pi^2/2$ bzw. (43): $< 7{,}3$) erreichbar sind.

Das erste Patent zum D-Verstärker stammt von BEDFORD aus dem Jahre 1930. Er machte seine Versuche mit Thyratons (thyristorähnlich wirkende Gasentladungsröhren). Wenn bis heute trotzdem der D-Verstärker keine größere Bedeutung erlangt hat, so liegt es an den *Problemen*, die mit ihm verknüpft sind:

- Der Schaltungsaufwand ist erheblich höher als bei anderen Endstufen.
- Die Taktfrequenz muß ca. 5- bis 7mal so hoch sein wie die höchste zu übertragende Frequenz (Abtasttheorem).
- Die Bandbreite muß ca. 200- bis 500mal so groß wie das höchste zu übertragende Signal sein. Nur dann sind in den Impulslängen hinreichend viele Abstufungen und damit ein guter Störabstand möglich.
- Durch die steilen Impulsflanken hoher Leistung besteht die Gefahr einer hohen Störstrahlung, die schwer zu reduzieren ist.

Der erste Punkt ist durch die Entwicklung der *integrierten Schaltkreistechnik* inzwischen praktisch bedeutungslos geworden. Der zweite und dritte Punkt hat durch die Entwicklung der *Leistungs-MOSFET* (s. Bild 6.4.11a)) wesentlich an Bedeutung verloren. Damit verbleibt heute im wesentlichen nur noch der letzte Punkt, und er dürfte beherrschbar werden. Es ist zu erwarten, daß der D-Verstärker in der Zukunft an Bedeutung gewinnen wird. Es gibt sogar schon Lösungen, die zu noch komplizierteren Modulationen, z.B. gemischten Pulsfrequenz- und Pulslängenmodulationen, übergehen, um weitere Verbesserungen zu erreichen. Mit dem Einsatz von D-Verstärkern werden vor allem *folgende Vorteile* auftreten:

- größter Wirkungsgrad und damit die Möglichkeit, in integrierten Schaltkreisen die heutige Grenzleistung von ca. 20 Watt zu überschreiten, bzw. die Möglichkeit, bei gleicher Leistung kleinere Baugruppen, kleine Netzteile bzw. kleinere Batterien verwenden zu können.
- An die Toleranzen der Bauelemente und ihre Linearität brauchen nur geringe Forderungen gestellt zu werden.
- Die Arbeitspunkteinstellung und -stabilisierung hat weitaus geringere Bedeutung.
- Die speziellen Verzerrungen von Transistorendstufen (7.4.6.4.) treten nicht auf. Es sind sehr gute Linearität und Dynamik erreichbar.
- Es besteht große Freiheit bezüglich des Lastwiderstandes und der Versorgungsspannung.

Den *Prinzipaufbau* eines D-Verstärkers zeigt Bild 7.4.16b). Der Vorverstärker ist Differenzverstärker mit den Transistoren T_1 und T_2. Die Z-Dioden bewirken eine Potentialverschiebung zum selbstschwingenden Modulator mit den beiden Transistoren T_3 und T_4. Es werden dabei die beiden Taktzeiten gegensinnig verstellt. Hinter dem Modulator befindet sich die Endstufe aus den Transistoren T_5 und T_6. Die beiden Dioden sind notwendig wegen der stark induktiven Komponente des Tiefpasses (Überschwingen). Für die Endstufe ist es wichtig, daß *zu keinem Zeitpunkt beide Transistoren gleichzeitig eingeschaltet* werden können. Dann würde nämlich ein Kurzschluß der Betriebsspannung auftreten. Hierfür sind eventuell Sondermaßnahmen notwendig.

7.4.6.3. Umschaltung zwischen zwei Endstufen

Das Grundprinzip dieser in der letzten Zeit in vielen Varianten beschriebenen Schaltung geht besonders deutlich aus Bild 7.4.17a) und b) hervor. Solange die Ausgangsleistung (-spannung) klein bleibt, werden nur die relativ kleinen Spannungsquellen V_1 und V_1' mit den Transistoren T_1 und T_1' benutzt. Sobald jedoch die verlangte Leistung die Grenze dieses Leistungsverstärkers überschreitet, werden die Spannungsquellen V_2 und V_2' mit den zusätzlichen Endtransistoren T_2 und T_2' eingeschaltet. Nach diesem Prinzip, das in sehr verschiedenen Schaltungen realisiert wird, läßt sich im Prinzip ein Wirkungsgrad erreichen, wie ihn Bild 7.4.17c) als DLD-Schaltung zeigt. Natürsich bleibt auch hierbei der Wirkungsgrad immer unter 78%, aber in einem großen Bereich ist er diesem Wert recht nahe. Das Problem aller dieser Schaltungen, welche u. a. die Namen E-Verstärker, G-Verstärker oder DLD (nach KEYWOOD: engl. Dynamic Linear Drive = dynamische Linearsteuerung) tragen, sind die Verzerrungen im Übergangsbereich zwischen den beiden Endstufen. Zuweilen werden hier Schaltzeiten im GHz-Bereich verlangt.

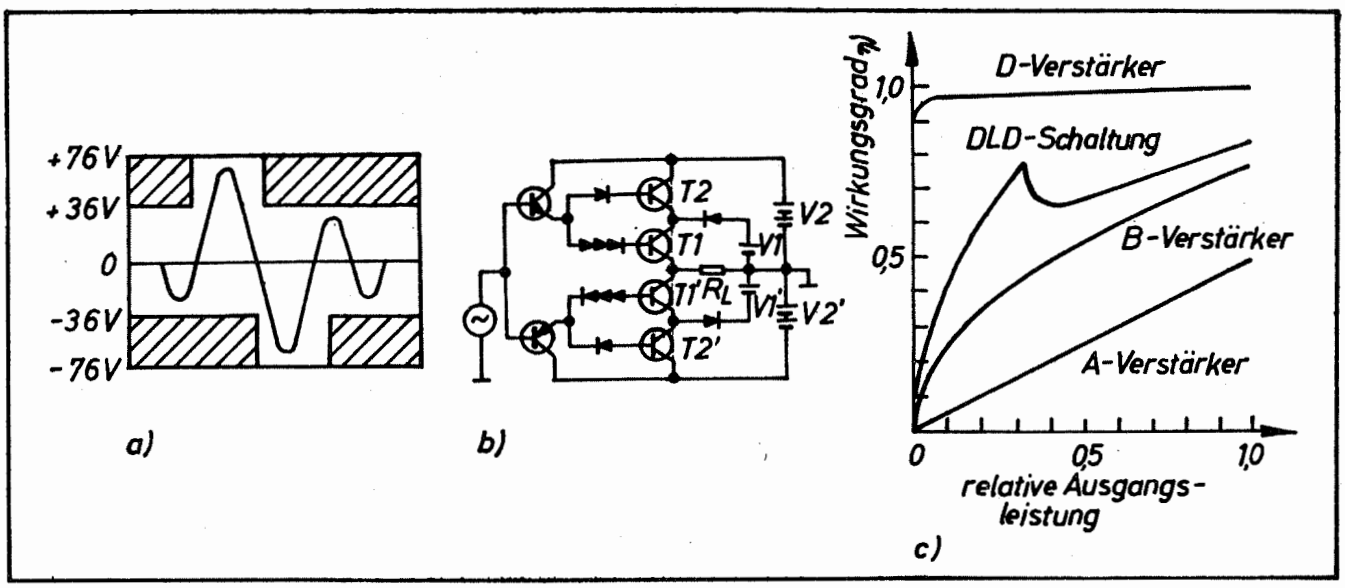


Bild 7.4.17. Zum DLD- bzw. E- oder G-Verstärker

a) Einschaltung der Spannungsquellen V_1 bzw. V_2 je nach der Aussteuerung

b) Prinzipschaltung der beiden Endstufen unterschiedlicher Leistung

c) Verlauf des Wirkungsgrades als Funktion der relativen Ausgangsleistung (auf maximale Leistung bezogen) bei verschiedenen Schaltungen

7.4.6.4. Spezielle Transistorverzerrungen

Immer wieder werden in der hochwertigen Elektroakustik Aussagen laut, daß bezüglich der *Klangqualität* der alte Röhrenverstärker dem modernen Transistorverstärker (*transistor sound*) deutlich hörbar überlegen ist. Unter anderem wurde die Entwicklung der Leistungs-MOSFET auch hierdurch deutlich stimuliert. Bisher konnte dieser subjektiv feststellbare Effekt meßtechnisch noch nicht eindeutig belegt werden. Auffällig ist jedoch, daß keine subjektive Unterscheidung bei Vorverstärkern möglich ist. Es kristallisieren sich jedoch *folgende Fakten* heraus:

- Der Übernahmeknick bringt zusätzliche Verzerrungen.
- Der Anteil höherer Oberwellen ist bei Transistorverstärkern höher.
- Transistorverstärker besitzen durch Kompromisse bei Arbeitspunktwahl, Stabilität und Bandbreite in der Sprungantwort zuweilen einen unregelmäßigen Verlauf.
- Durch die Frequenzkorrekturen treten bei kritischen Spannungssprüngen bereits in den Steuerstufen Nichtlinearitäten (Begrenzungen) auf.

Dies hat zur Folge, daß immer stärker auf *vollständige Symmetrie der beiden Leistungsendstufen und ihrer Vorstufen* geachtet wird. Ferner wird jetzt mehr Gewicht auf die Frequenzgangkorrekturen und damit auf die Stabilität gelegt. Doch alle durchgeführten Maßnahmen brachten bisher nur Teilerfolge, die gerade deshalb so schwer zu bewerten sind, weil noch objektive Meßmethoden

fehlen. Bedeutsam ist, daß der D-Verstärker von allen bisher angenommenen Effekten frei ist und, nach den wenigen bisher bekannten Aussagen von Hörexperten, die störenden Effekte nicht zeigen soll.

Eine spezielle Maßnahme zur Behebung der typischen Transistorverzerrungen sei hier wegen ihrer generell interessanten Schaltung noch beschrieben. Das Prinzip geht auf ein Patent aus dem Jahr 1928 zurück und wurde in der Zwischenzeit mehrfach modifiziert (Bild 7.4.18a)). Der Hauptverstärker besitzt Verzerrungen. Sie werden durch Vergleich mit der Eingangsspannung abge-

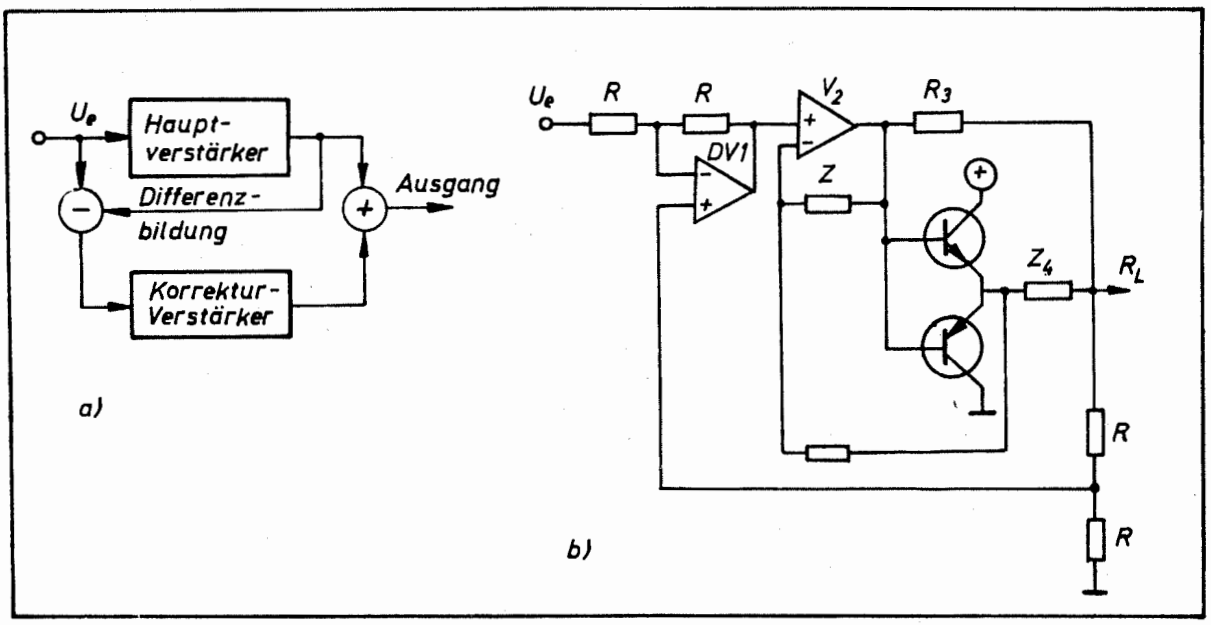


Bild 7.4.18. a) Prinzip der Vorwärtskopplung (feed forward error correction)
b) Prinzip der stromentlasteten Konstantstromquelle als verzerrungsarmer Endverstärker

trennt und einem *Korrekturverstärker* zugeführt. Dieser benötigt nur eine relativ kleine Leistung, dafür aber extrem hohe Linearität, denn seine Linearität (und Stabilität) bestimmt die Verzerrungseigenschaften des Systems aus beiden Verstärkern. Die Nachteile dieses Prinzips sind:

- Leistungen sind nur schwierig zu addieren.
- Es werden zwei Endstufen benötigt.

Die Weiterentwicklung dieser Grundidee führte zur stromentlasteten Konstantstromquelle gemäß Bild 7.4.18b). Im Differenzverstärker DV 1 wird der Verzerrungsanteil aus der Spannung am Lastwiderstand R_L herausgezogen und der Steuerung des A-Verstärkers kleiner Leistung V 2 hinzugefügt. Die Stromentlastung für die großen Amplituden besorgt die B-Stufe (ohne Ruhestrom aus den beiden Transistoren T_1 und T_2, wobei Z_4 in der Praxis eine Induktivität ist). Die Theorie und der Aufbau sowie die Eigenschaften eines nach diesem Prinzip realisierten 40-W-Verstärkers enthält die Zeitschrift Radio, Fernsehen, Elektronik 31 (1982), H. 11. Von Experten konnte kein Klangunterschied zu Röhrenverstärkern festgestellt werden.

7.5. Operationsverstärker

7.5.1. Einführung

7.5.1.1. Zur Geschichte

Der Operationsverstärker (OV) entstand im Zusammenhang mit der *Analogrechentechnik*. Dort war der Operationsverstärker die entscheidende Baugruppe, aus der in der Regel Integratoren aufgebaut und untereinander zur Simulation von Objekten mittels Differentialgleichungen verschaltet wurden. Die Anzahl der damals noch aus Röhren und später aus Transistoren aufgebauten

OV war ein wesentliches Kennzeichen für die Leistungsfähigkeit des jeweiligen Analogrechners. Inzwischen hat durch die rasante Entwicklung der Digitaltechnik der Analogrechner seine Bedeutung praktisch völlig eingebüßt. Insofern haben heutige OV nur noch Eigenschaften, die in etwa denen der historischen OV entsprechen. Sie sind — von extrem seltenen Ausnahmen abgesehen — voll integriert aufgebaut und dienen den verschiedensten Anwendungen.

Der integrierte OV stellte das erste universelle integrierte analoge Bauelement dar. Die Vielfalt seiner Möglichkeiten korreliert stark (auch inhaltlich) mit der *Universalität der Rückkopplung* (7.1), geht jedoch über die dort sonst üblichen Methoden hinaus, da beim OV auch vielfältige nichtlineare Betriebsweisen, einschließlich nichtlinearer Rückkopplungen, zur Anwendung kommen.

Der erste integrierte OV war der μA 702 von FAIRCHILD. Er wurde erstmalig 1963 produziert. Der erste erfolgreiche integrierte OV, der heute noch gewisse Bedeutung hat, war der 1966 von FAIRCHILD eingeführte Typ μA 709 (DDR A 109). Er wird deshalb ausführlich im Abschnitt 7.5.2. behandelt. Seitdem gibt es mehrere Generationen von universellen OV und eine Vielzahl von OV mit speziellen Eigenschaften. Hierauf wird im Abschnitt 7.5.4. eingegangen. Die Typenvielfalt ist heute so groß, daß dabei nur das Spektrum angedeutet werden kann.

7.5.1.2. Der ideale Operationsverstärker

Der ideale Operationsverstärker (OV) ist einmal eine Zielvorstellung für die Entwicklung konkreter OV. An ihm werden die Eigenschaften der realen OV gemessen. Er ist zum anderen aber auch ein Modell, mit dem leicht und meist näherungsweise Schaltungen mit OV verstanden bzw. berechnet werden können.

Der ideale OV besteht aus *drei Teilen*:

- *einem idealen Differenzeingang.* Dies bedeutet vor allem:
  - Gleichtaktverstärkung gleich Null,
  - unendlich großer Eingangswiderstand für Differenzsignale,
  - kein Offset, weder Strom noch Spannung,
  - keine Drift, weder von Temperatur noch Betriebsspannung oder Alterung,
  - kein Eigenrauschen.
- *einer idealen Endstufe.* Dies bedeutet vor allem:
  - Innenwiderstand gleich Null,
  - lineare Kennlinie bis an die Aussteuerungsgrenzen.
- *einem idealen Haupt- (Gesamt-) Verstärker.* Dies bedeutet vor allem:
  - unendlich große Verstärkung,
  - unendlich große Bandbreite ohne Frequenzgang,
  - keine Phasenfehler im gesamten Frequenzbereich.

Es ist sofort zu erkennen, daß einige dieser Daten gegen die physikalischen Gesetze verstoßen. Sie sind so zu interpretieren, daß hier der bestmögliche Kompromiß für eine gegebene Anwendungsklasse anzustreben ist.

7.5.2. A 109

7.5.2.1. Die Innenschaltung

Die Innenschaltung des A 109 zeigt Bild 7.5.1. Darin sind auch die Teilschaltungen deutlich hervorgehoben. Für den Betrieb des Operationsverstärkers (OV) A 109 ist zu beachten, daß er mit *zwei Spannungsquellen* von +15 V und −15 V gegenüber Masse betrieben werden muß. Ein Masseanschluß existiert für den Schaltkreis nur indirekt über die Mitte der Spannungsquelle (30 V).

Der *erste Differenzverstärker* mit T_1, T_2, R_1, R_2 enthält im Emitterkreis eine Stromquelle in Form eines *Stromspiegels* T_{10}, T_{11}. Durch den Widerstand R_{11} wird im Sinne der Schaltung von Bild 7.3.16c) der kleine Strom von 40 μA erzeugt. Dadurch wird

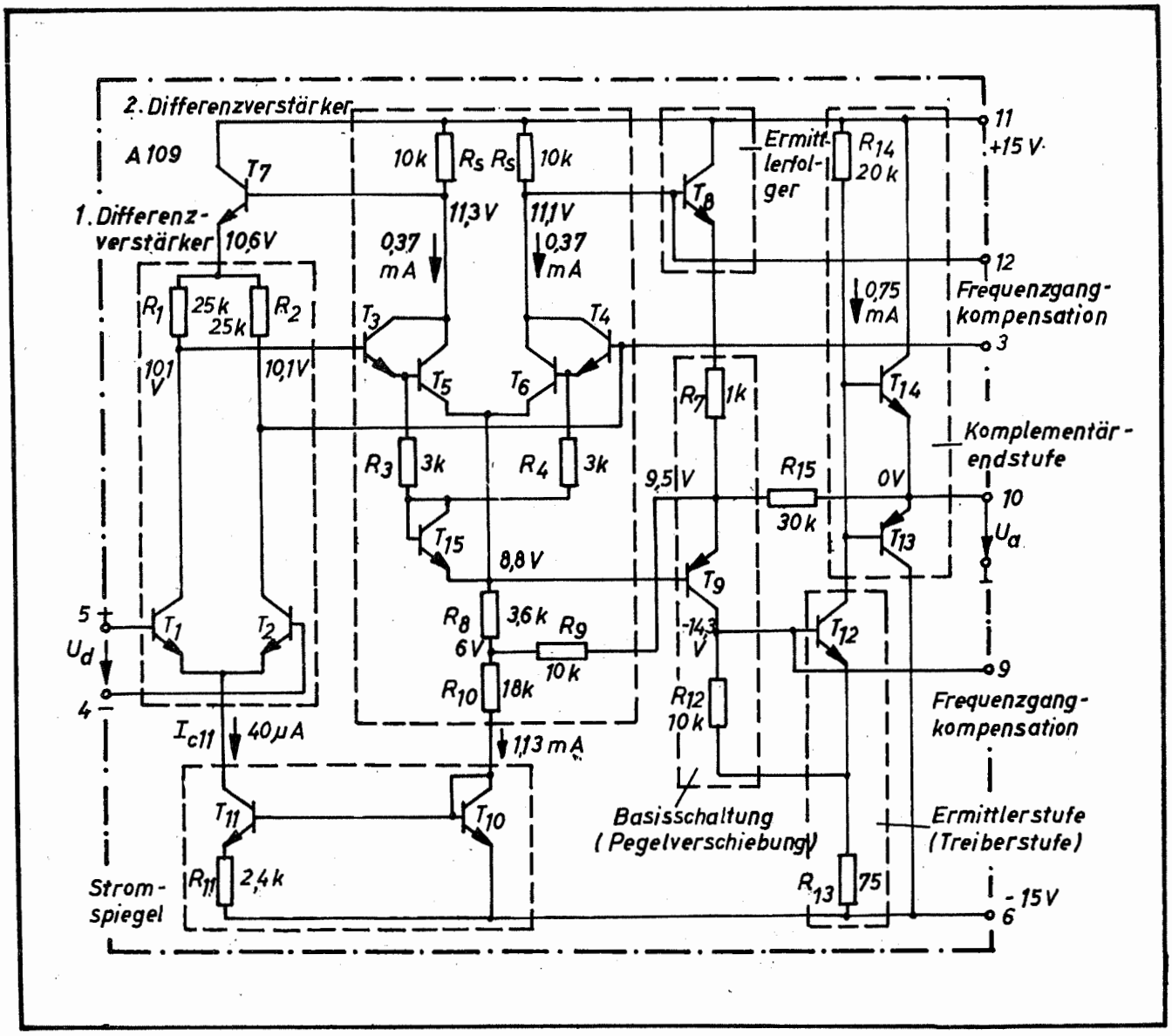


Bild 7.5.1. Innenschaltung des Operationsverstärkers A 109

ein großer Eingangswiderstand des OV erreicht. Die Verstärkung ist ca. 15fach. Die maximale Spannung zwischen beiden Eingängen wird auf etwa 2×7 V infolge der Durchbruchspannung beider Transistoren begrenzt. Eine gleichzeitige *Potentialverschiebung* beider Eingänge gegenüber Masse ist infolge der Stromquelle (T_{11}) in einem weiten Bereich (ca. ± 8 V) möglich.

Der *zweite Differenzverstärker* enthält zwei DARLINGTON-Stufen T_3, T_5 bzw. T_4, T_6. Hier wird *erhebliche Verstärkung* realisiert (ca. 100fach). Über den als Diode geschalteten Transistor T_{15} sind die DARLINGTON-Stufen *gut temperaturstabilisiert*. In Standardschaltungen zur Einstellung eines günstigen Arbeitspunktes wären die Widerstände R_3 bzw. R_4 direkt gegen den Emitter von T_5 und T_6 zu schalten. Sie müßten dann relativ groß sein (Fläche) und hätten vom Chip her sogar noch den falschen Temperaturkoeffizienten. Durch die Diode T_{15} wird die Emitter-Basis-Spannung von T_5 und T_6 voll im Temperaturgang nachgebildet, und an R_3, R_4 liegt die Spannung

$$U_{R4} = U_{BE6} - U_{BE15} = \frac{kT}{e_0} \ln \frac{I_{C6}}{I_{C15}}. \tag{1}$$

Die Transistoren T_7 bzw. T_8 bewirken den Übergang zur *unsymmetrischen* Spannung analog zur Schaltung in Bild 7.3.13d). Gleichzeitig stellt T_8 als Emitterfolger den Anfang der *Potentialverschiebeschaltung* T_9, R_7 (Emittergegenkopplung), R_{12} dar. Ihr

folgt die *Treiberstufe* mit T_{12}. Beide zusammen verstärken ca. 400- ... 800fach. Die Komplementärendstufe aus T_{13} und T_{14} arbeitet im B-Betrieb. Um die Übernahmeverzerrungen (Totzone etwa 1 V) zu reduzieren, ist die *Ausgangsstufe* einschließlich Treiberstufe über R_{15} gegengekoppelt. Die interne Schleifenverstärkung beträgt hier etwa 20. Folglich wird die Totzone auf 50 mV reduziert. Mit dem Widerstand R_9 wird eine gute *Nullpunktstabilität* des gesamten Verstärkers über den Stromspiegel erreicht.

7.5.2.2. Layout

Aus Gründen des Know-how, der Lizenzpolitik und aus verschiedenen anderen firmenstrategischen Gesichtspunkten werden Chipfotografien und Layouts bzw. Masken nur selten zur Publikation freigegeben. Da der Typ A 109 schon zur Tradition der OV gehört, sind hier die entsprechenden Unterlagen bekannt und z. T. in Bild 7.5.2 dargestellt. Dies ermöglicht, die prinzipielle Herstellung eines integrierten Schaltkreises etwas detaillierter anzugeben. Aufgrund von Weiterentwicklungen, die für die Herstellung eines jeden Schaltkreises gewisse Veränderungen bewirken, passen leider einige Details im Bild nicht exakt zueinander (z. B. R_{10} und das Fehlen des Testtransistors in Teilbild d)).

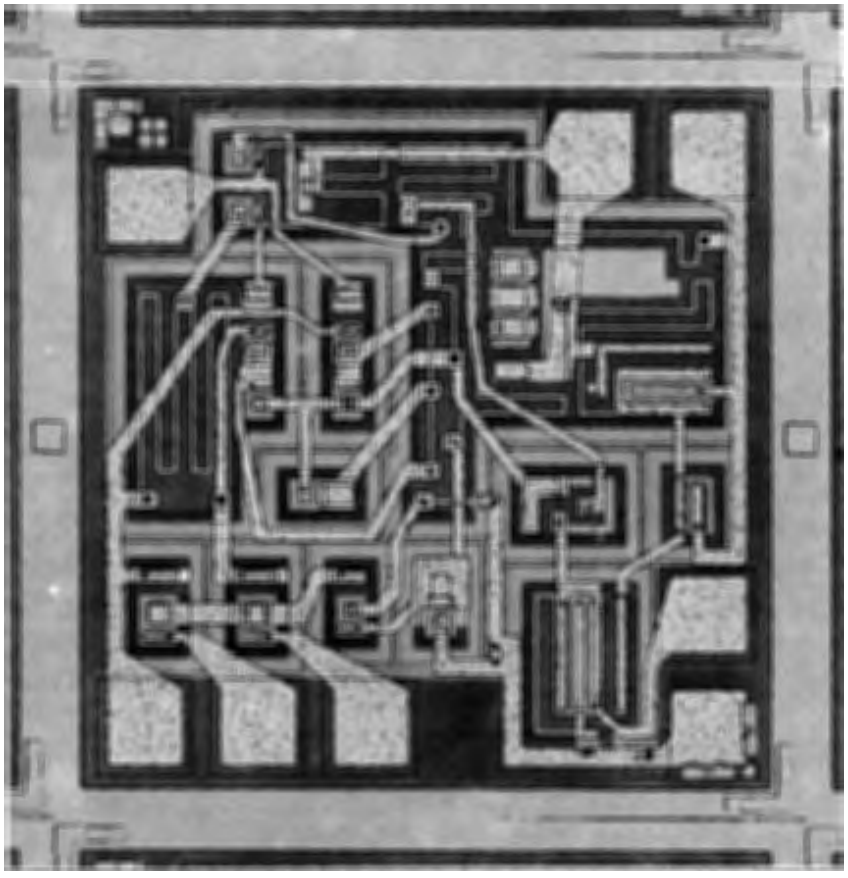
Teilbild a) zeigt die Chipfotografie des A 109, in der alle Strukturen (auch die tieferen) zu erkennen sind. Teilbild b) verdeutlicht dies schematisch und schafft die Zuordnung zu Bild 7.5.1. Als Beispiel wurden nur drei Masken ausgewählt. Vor und nach dem Aufwachsen der Epitaxieschicht werden mit der Maske von Teilbild c) die Isolierrahmen für die einzelnen Teilstrukturen eindiffundiert. Sie sind in Teilbild a) als dunkel umrandete parallele Linien von knapp 1 mm Breite zu erkennen. Teilbild d) zeigt die Diffusionsmaske für die Basis- und Widerstandsgebiete (hell in Bild a)), und Teilbild e) zeigt die Maske für die Emitterdiffusion und zur höheren Dotierung der Kollektoranschlußgebiete. Nicht dargestellt sind die Masken für das begrabene Gebiet, die Kontaktfenster und für die Leiterzüge. Die Leiterzüge sind ohnehin als helle Flächen und Linienzüge in Teilbild a) gut zu erkennen.

7.5.2.3. Statische Kenndaten und Eigenschaften

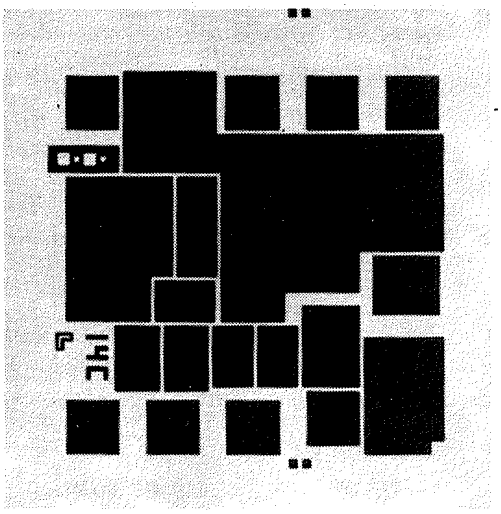
Wichtige Kenndaten des A 109 faßt Tab. 7.5.1 zusammen. Sie zeigen deutlich, wie weit die im Abschnitt 7.5.1.2. genannten Werte eines idealen OV genähert werden.

Tabelle 7.5.1. Elektrische Kennwerte des A 109
(für 25 °C — 5 K und $U_{B^+} = -U_{B^-} = 15$ V)

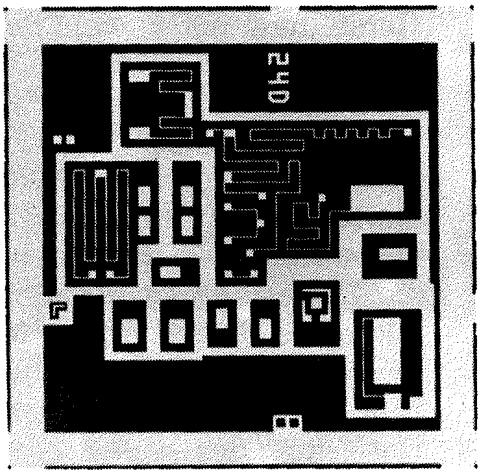
Eingangsoffsetspannung U_{I0}	$\leqq$ 7,5 mV
Betriebsspannungsunterdrückung SVR	$\leqq 2 \cdot 10^{-4}$
Eingangsruhestrom I_I	$\leqq$ 1,5 µA
Eingangsoffsetstrom I_{I0}	$\leqq$ 500 nA
Ausgangsspitzenspannung ($R_L = 2$ kΩ) U_A	$\geqq \pm$ 10 V
Ausgangsspitzenspannung ($R_L = 10$ kΩ) U_A	$\geqq \pm$ 12 V
Gleichtakt-Eingangsspannungsbereich U_I	$\geqq \pm$ 8 V
Gleichtaktunterdrückung CMR	$\geqq$ 65 dB
Großsignalverstärkung ($R_L = 2$ kΩ, $U_A = \pm 10$V) V_U	$\geqq 1{,}5 \cdot 10^4$
Eingangswiderstand R_E	$\geqq$ 50 kΩ
Gleichtakteingangswiderstand	$\approx 10^2$ ... 10^3 R_E
Leistungsaufnahme	$\leqq$ 200 mW
Ausgangswiderstand R_A	$\leqq$ 150 Ω
Spannungsdrift	$\leqq$ 3 µV/K



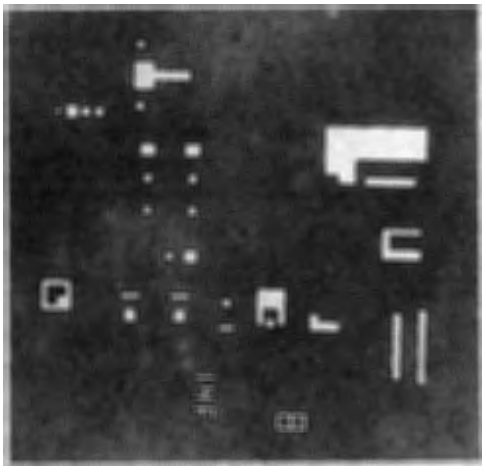
a)



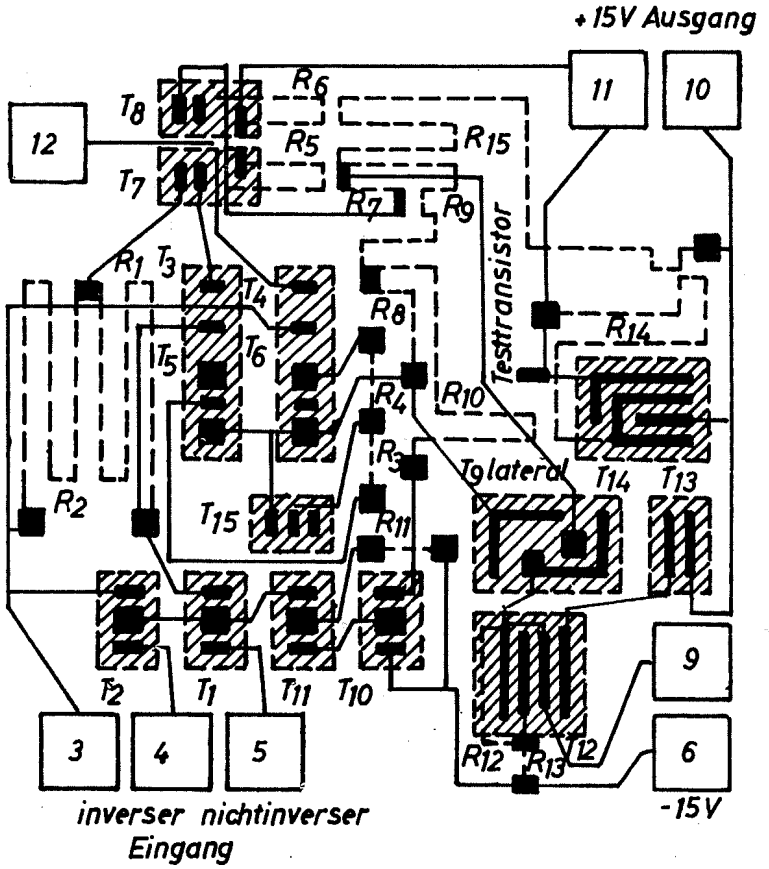
c)



d)



e)



b)

Bild7.5.2. Topologie des Operationsverstärkers A 109

a) Chipmikrofotografie
b) Zuordnung der Komponenten zum Schaltbild Bild 7.5.1
c) bis e) Ausgewählte Masken
c) Isolierrahmen für die Inseln
d) Basis- und Widerstandsdiffusion
e) Emitter- und Kollektorkontakte

Statische Kenndaten (ohne Frequenzgangeinfluß) ergänzt dazu Bild 7.5.3. Aus der Übertragungskennlinie gemäß Teilbild a) ist der gut lineare Verlauf bis kurz vor die Sättigungswerte zu erkennen. Der Temperatureinfluß ist im extremen Temperaturbereich gerade erkennbar. Die Abhängigkeit der Sättigungswerte sowie der maximalen Leerlaufverstärkung von der Betriebsspannung zeigt Teilbild 7.5.3b) und c). Den Einfluß des Lastwiderstandes auf die Sättigungswerte enthält Teilbild d). Der Temperatureinfluß auf den Eingangsruhestrom (e), den Eingangsoffsetstrom (f) und auf den Eingangswiderstand (g) zeigt, daß mit höherer Temperatur hierfür bessere Werte erreicht werden. Die Rauschzahl als Funktion des Quellenwiderstandes ergibt ein Optimum um etwa 10 kΩ und ist abhängig von der gewählten oberen Grenzfrequenz (Bild 7.5.3h)).

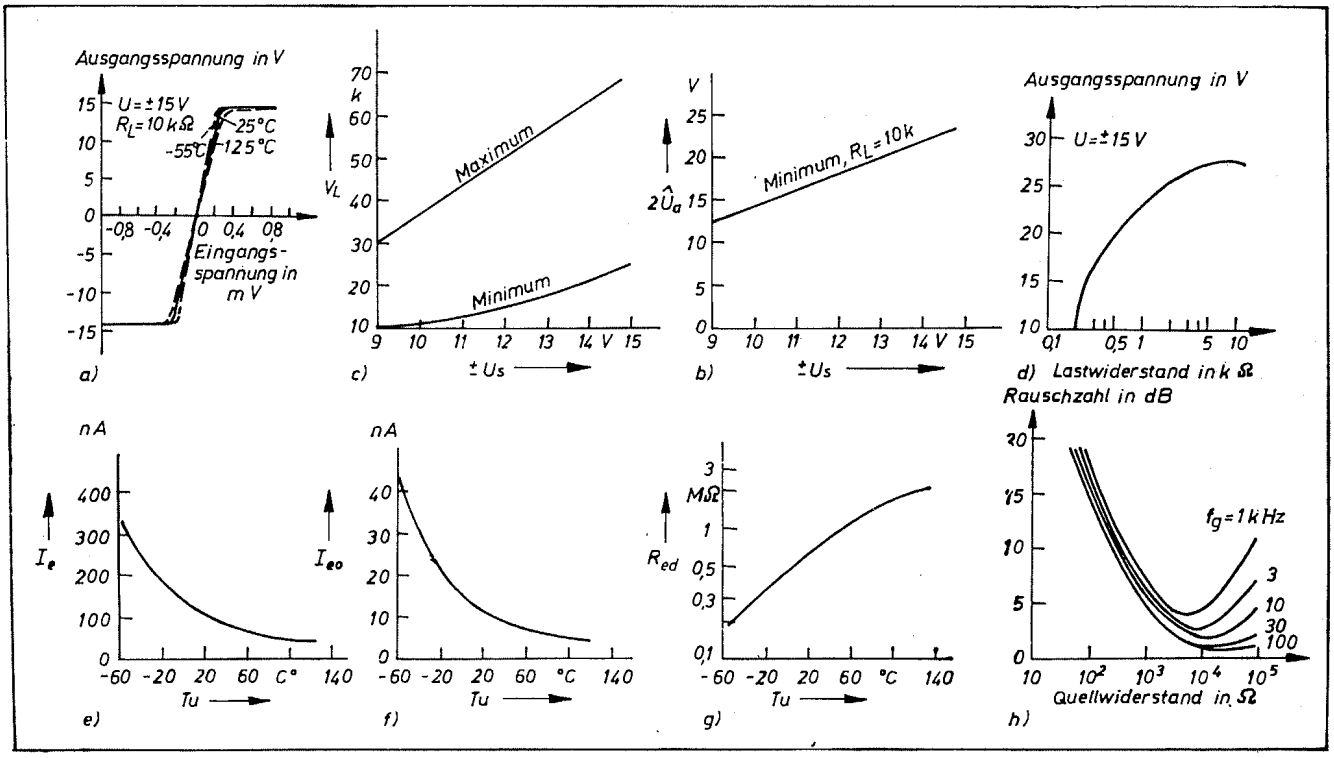


Bild 7.5.3. Wichtige statische Daten des Operationsverstärkers A 109

a) Übertragungskennlinie
b) Einfluß der Betriebsspannung auf die Differenz der beiden erreichbaren Sättigungsspannungen,
c) auf die Grenzen der Leerlaufverstärkung
d) Maximale Differenz der Sättigungsspannungen in Abhängigkeit vom Lastwiderstand
e) Temperatureinfluß auf den Eingangsstrom,
f) auf den Eingangsoffsetstrom,
g) auf den Eingangswiderstand zwischen den beiden Eingängen
h) Abhängigkeit der Rauschzahl vom Quellenwiderstand und der höchsten übertragenen Frequenz

7.5.3. Frequenzgang und -korrektur

Der Frequenzgang des offenen OV vom Typ A 109 ist durch *Tiefpaßfrequenzen* der drei Verstärkerstufen bestimmt. Sie liegen bei ca. 0,2; 1; 8 MHz. Die meisten OV können ähnlich beschrieben werden. Daher haben die folgenden Betrachtungen generelle Bedeutung, die rechte Kurve in Bild 7.5.4c) ist als beispielhaft aufzufassen. Damit sind bei Gegenkopplung des OV die Betrachtungen des Abschnitts 7.1.2.3., insbesondere des Bildes 7.1.8, gültig. Eine stabile Gegenkopplung ist nur möglich bis zu der Frequenz, wo der durch die Tiefpaßglieder verursachte Phasenwinkel kleiner als 180° bleibt. Dann geht nämlich die *Gegenkopplung in eine Mitkopplung über*, und

der Verstärker wird *instabil*. In der Regel wird eine *Phasensicherheit* von 45° bzw. 60° gegenüber den 180° angenommen, das bedeutet also Maximaldrehungen von 135° bzw. 120°. Bei den hier vorliegenden einfachen Schaltungen (ohne Kreuzglieder, Brücken usw.) ist eine direkte Verkopplung von Phasen- und Amplitudengang gegeben. Deshalb kann die obige Aussage auch so ausgedrückt werden: Gegenkopplung ist nur bis zu den Frequenzen möglich, wo der *Amplitudengang mit 20 dB/Dekade* abfällt (in Bild 7.5.4c) also maximal bis f_2). Dies bedeutet, daß maximal so stark gegengekoppelt werden darf, daß die Verstärkung mindestens 60 dB beträgt. Wird stärker gegengekoppelt, besteht die Gefahr zur Schwingneigung. Dieses Ergebnis ist unbefriedigend. Für viele Daten, wie Linearität der Kennlinien, Stabilität gegenüber Spannungsschwankungen usw., wäre es nützlich, stärker gegenkoppeln zu können. Bereits im Abschnitt 7.1.2.3. wurde auf eine *Lösung* dieses Problems hingewiesen: *Eine Grenzfrequenz ist dominant tief zu legen*. Diese Möglichkeit und daraus abgeleitete Varianten werden fälschlicherweise als Phasenkompensation bezeichnet. Tatsächlich kompensieren sie nicht den Phasengang, sondern verändern bzw. *korrigieren Frequenz- und Phasengang* so, daß eine stärkere Gegenkopplung möglich ist. Dabei werden zweimal je *zwei Fälle* unterschieden:

- äußere bzw. innere Beschaltung
- Beschaltung mit C bzw. R und C.

7.5.3.1. Äußere Beschaltung

Die äußere Beschaltung des OV setzt voraus, daß kein Eingriff in den Frequenz- und Phasengang des OV gemacht wird, sondern die Korrektur im Rückkopplungszweig gemäß Bild 7.5.4a) und b) erfolgt. Auf den Unterschied beider Schaltungen wird später einzugehen sein. Mit dem Potentiometer P kann die Spannungsgegenkopplung zwischen 0 und 1 eingestellt werden. Weiter sei R_2 zunächst Null, und betrachtet werde die Schaltung gemäß Teilbild a). Es erscheint unwahrscheinlich, daß die Erhöhung der Anzahl der Tiefpässe auf vier überhaupt Vorteile bringen kann. Dies sei erklärt. Die Grenzfrequenz des Tiefpasses aus dem zusätzlich eingefügten RC-Glied sei f_c. Er allein besitzt dann einen Amplitudengang gemäß Bild 7.5.4d). Von der Frequenz f_c an wächst seine Dämpfung ständig mit 20 dB/Dekade. Ohne Gegenkopplung wird sie (im logarithmischen Maßstab) zu der Dämpfung des OV addiert, und es entsteht Kurve ⓒ in Bild 7.5.4c). Damit ist der OV jetzt bis zu einer Verstärkung von 20 dB herab gegenzukoppeln. Er hat zwar einen weitaus „schlechteren" Leerlauffrequenzgang, besitzt dafür aber im stark gegengekoppelten Fall (zumindest bei den tiefen Frequenzen) eine erheblich linearere Kennlinie und Stabilität. Die maximal nutzbare Grenzfrequenz ist allerdings von ca. f_2 und f_1 reduziert worden.

Durch die Wahl von f_c können Daten erreicht werden, die gerade jede gewünschte Verstärkung mit Gegenkopplung optimal bez. Stabilität und Grenzfrequenz befriedigen.

7.5.3.2. Äußere Beschaltung mit R, C, R

Wird dem C in Bild 7.5.4a) ein R_2 in Reihe geschaltet, so besteht die Möglichkeit, die Wirkung dieser äußeren Beschaltung gerade dann unwirksam zu machen, wenn der OV selbst den Abfall von 20 dB/Dekade realisiert, also bei f_1. Die Grenzfrequenz f_h aus C und R_2 sollte gemäß Bild 7.5.4c) und d) möglichst genau f_1 des OV entsprechen. Für eine Verstärkung von wieder 20 dB mit Gegenkopplung entsteht dann die Kurve ⓑ in den beiden Teilbildern. Es tritt also nicht mehr etwa f_1 als Grenzfrequenz, sondern etwa f_t auf. Der Leerlauffrequenzgang beginnt nicht mehr bei f_c, sondern bei f_t abzufallen. Generell wird hierdurch also soviel an Bandbreite und anderen vorteilhaften

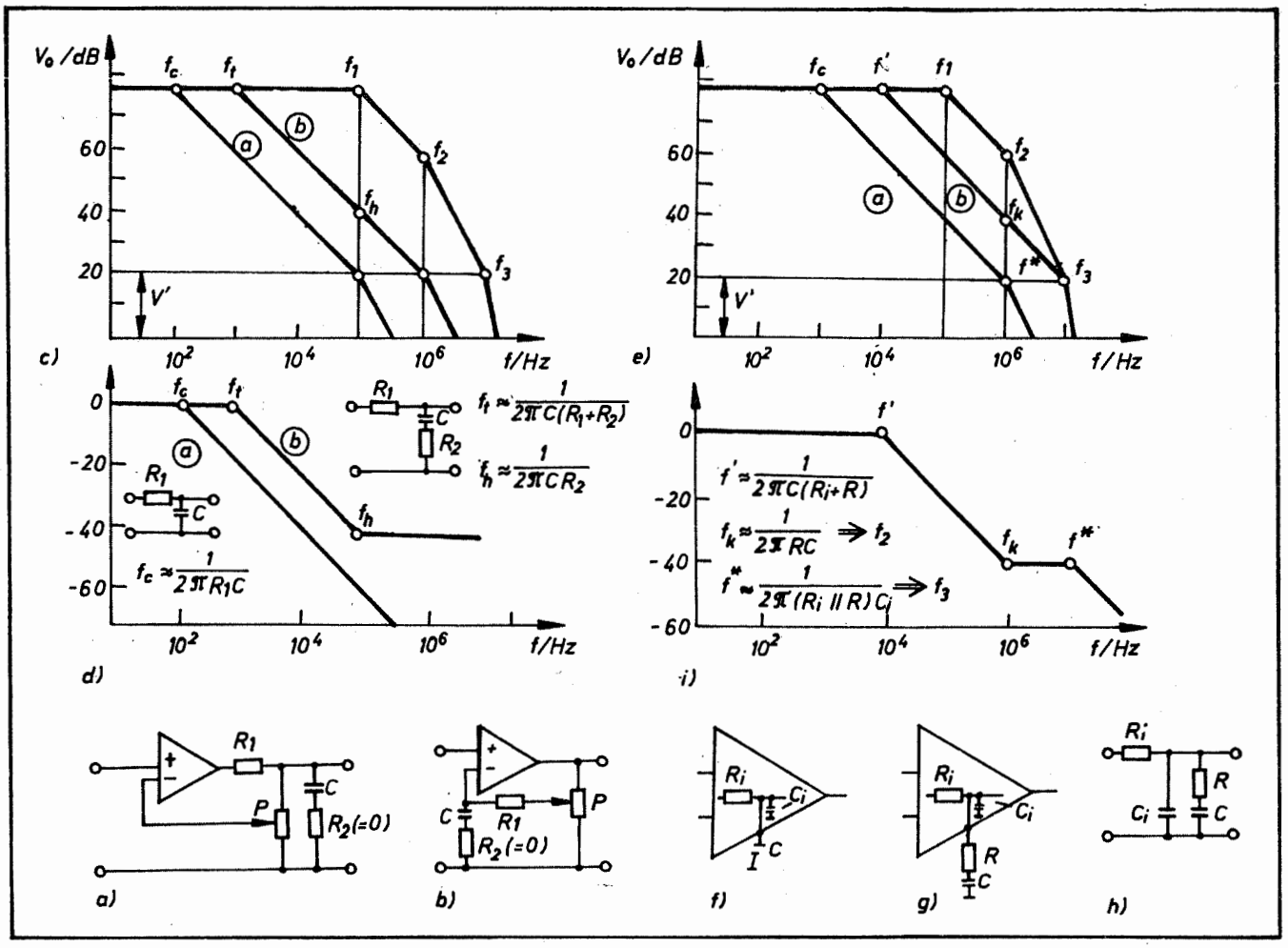


Bild 7.5.4. Zur Frequenzkorrektur bei Operationsverstärkern

a) Äußere Korrektur am Ausgang
b) Äußere Korrektur am Eingang
c) Frequenzgang eines OV mit den drei Tiefpaßfrequenzen f_1, f_2 und f_3 sowie Frequenzgang mit äußerer Beschaltung gemäß:
d) *RC*-Frequenzgang ⓐ und *RCR*-Frequenzgang ⓑ führen zu den korrigierten Frequenzgängen in Teilbild c),
e) Frequenzgänge bei innerer Beschaltung mit einer Kapazität ⓐ bzw. mit einer Reihenschaltung aus Kapazität und Widerstand ⓑ
f) Schema für die innere Beschaltung mit einer Kapazität,
g) mit einer Reihenschaltung aus *C* und *R*
h) Ersatzschaltbild für die innere *RC*-Schaltung
i) Frequenzgang des Korrekturgliedes von Teilbild h)

Eigenschaften gewonnen, wie das Verhältnis aus f_1 und f_2 beim unbeschalteten OV ist. Anders ausgedrückt, wird indirekt die Grenzfrequenz f_1 des OV nach f_t verschoben. Die dazu notwendige „Korrektur" bewirkt eben das auf f_t und $f_h = f_1$ abgeglichene R_1CR_2-Glied.

7.5.3.3. Innere Korrektur

Bei verschiedenen OV, insbesondere auch beim A 109, sind Anschlüsse herausgeführt, die es gestatten, die Grenzfrequenzen der Tiefpässe direkt zu ändern. Dadurch ist es nicht notwendig, den dreigliedrigen Tiefpaß auf einen viergliedrigen zu erhöhen, und die Korrektur der untersten Frequenz f_1 ist direkt mit einem parallelgeschalteten *C* möglich. So entsteht die Kurve ⓐ in Bild 7.5.4e). Es werden durch innere Beschaltung mit einem *C* bereits jene Werte erreicht, die ein äußeres *RCR*-Glied realisiert. Es gibt OV, in denen die Beschaltung bereits intern vorgenommen wird, z.B. der

μA 741 (7.5.4.1.). Die Grenzfrequenz bestimmt sich zu

$$f_c = \frac{1}{2\pi R_i(C + C_i)}. \tag{2}$$

Es ist naheliegend, daß mit einem R_iCR-Glied noch bessere Werte erreichbar sind. Das Prinzip ist jedoch erheblich anders. Dies läßt sich aus der Schaltung gemäß Bild 7.5.4g) und der zugehörigen Ersatzschaltung gemäß Teilbild h) erklären. Seinen prinzipiellen Frequenzgang zeigt Bild 7.5.4i). Zunächst wirkt nur das Zusatz-C über die Reihenschaltung aus $R_i + R$. Der zugehörige Frequenzgang mit 20 dB/Dekade beginnt bei f' und hört auf, sobald R und C von etwa gleicher Größe sind (Grenzfrequenz f_k). Ein neuer Abfall mit 20 dB/Dekade beginnt, wenn C_i den Wert der Parallelschaltung aus R_i und R erreicht, also bei der Frequenz f^*. Wird nun $f_k \approx f_2$ und $f^* \approx f_3$ gewählt, so läßt sich ein Frequenzgang mit 20 dB/Dekade von f' bis f_3 erreichen. Dies ist der bestmögliche Wert, der mit einem Korrekturglied erreicht werden kann. Mit zwei Korrekturgliedern bestehen weitere Möglichkeiten.

7.5.3.4. Sprungantwort

Für Einschwingverhalten bezüglich kleiner Impulse ist vor allem der Amplituden- und Phasengang an der oberen Frequenzgrenze wichtig. Für einfache, eingliedrige RC-Tiefpässe gilt, daß bei der Grenzfrequenz ein Phasenwinkel von 45° existiert. Bei mehrfachen Tiefpässen ist der Winkel größer, ebenso, wenn am Ende des Frequenzganges eine *Resonanzüberhöhung* existiert. Vereinfacht ist für das Einschwingverhalten entscheidend, wieweit der Phasenwinkel im Übertragungsbereich — also bis zur 6-dB-Grenze — vom Normalwert bei Frequenzmitte abweicht. Dieser Wert α_{krit} wird oft mit *Phasenrand* oder *Phasenspielraum* bezeichnet. Er wird für verschiedene Zwecke vorgegeben. Für einige Werte zeigt Bild 7.5.5a) das entsprechende Einschwingverhalten. Meist werden Werte zwischen 45° und 60° vorgegeben. Wie dies bezüglich einer äußeren Korrektur mit RCR-Gliedern optimal zu gestalten ist, zeigt beispielhaft Bild 7.5.5b). Es ist auffällig, daß durch die äußere Korrektur der Phasenwert des unkorrigierten Verstärkers nicht unterschritten werden kann. Es

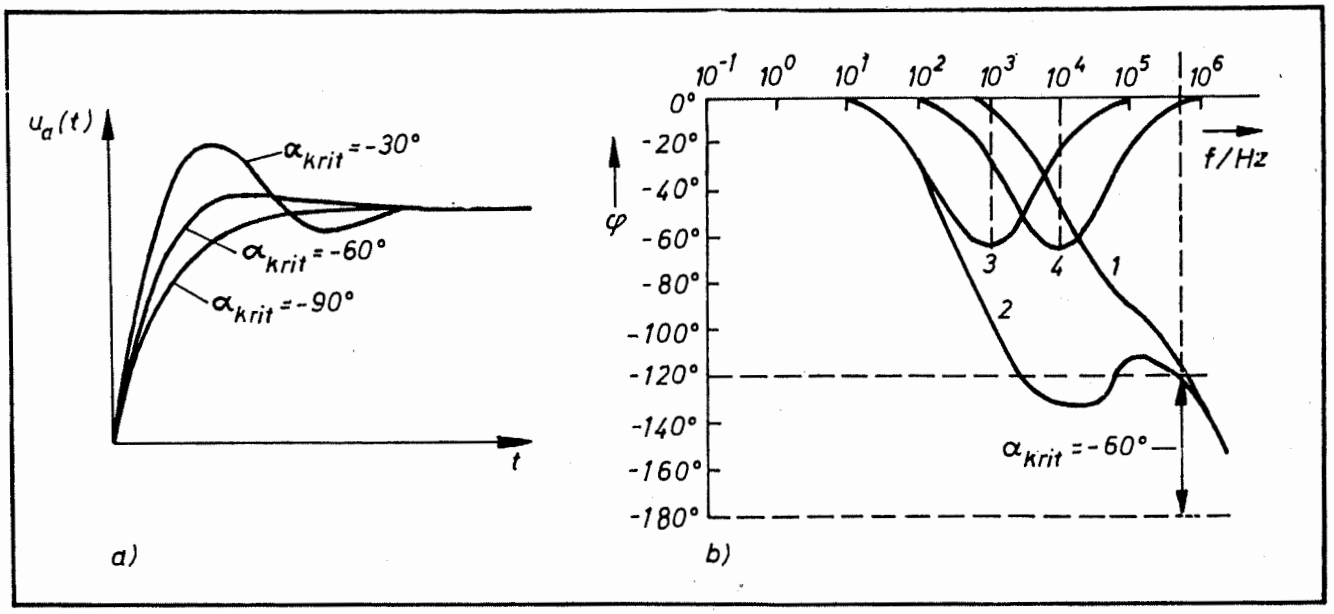


Bild 7.5.5. Zum Phasenrand

a) Impulsverhalten bei unterschiedlichem Phasenrand

b) Beispiel eines mit zwei RCR-Gliedern korrigierten Verstärkers: ① unkorrigiert, ② korrigiert, ③ und ④ Phasengang der beiden Korrekturglieder

wird aber der Amplitudengang so verändert, daß für die tieferen Frequenzen eine weitaus höhere Gegenkopplung mit allen ihren Vorteilen möglich ist (vgl. Bild 7.5.4c), Kurve ⓑ).

7.5.3.5. Optimale Anschaltung der Korrekturglieder

Die soeben durchgeführte Betrachtung gilt nur für kleine Eingangssignale. Bei *großen Eingangssignalen* treten neue Eigenschaften auf. Die Korrekturen — sowohl die inneren als auch die äußeren — bewirken nämlich bei den hohen Frequenzen zusätzliche Dämpfungen. (Das ist ja schließlich ihr Ziel!) Es sei angenommen, daß die Korrektur zwischen Treiber und Endstufe erfolge und, z.B. bei gegebener Frequenz, 20 dB betrage. Dann wird, um die Endstufe voll auszusteuern, vor dem Korrekturglied eine 10fache Spannung notwendig. Diese kann sie nicht aufbringen, und folglich wird sie bereits übersteuert, bevor die Endstufe voll ausgesteuert ist. Dies hat zur Folge, daß es eine *maximale Frequenz* gibt, für die gerade noch *Vollaussteuerung* der Endstufe möglich ist. Wird dagegen die Korrektur vor dem ersten Differenzverstärker vorgenommen, so kann dieser Fall nie eintreten. Dafür muß aber ein anderer Nachteil in Kauf genommen werden. Bei hohen Frequenzen werden die das Korrekturglied passierenden Signale so klein, daß die Gefahr besteht, daß sie im *Eigenrauschen* untergehen. Der Stör-Rausch-Abstand wird also wesentlich ungünstiger. Wird das Korrekturglied an irgendeiner anderen Stelle eingefügt, so überwiegt der eine oder andere Effekt. Dies bedeutet, daß es je nach dem Anwendungseffekt optimale Orte für das Anschalten der Korrekturglieder gibt.

7.5.3.6. Slew-Rate

Die Eingangsspannung werde schlagartig von der einen zur anderen Grenze durchgesteuert. Die einzelnen Stufen des OV folgen dann, so schnell es ihnen aufgrund ihrer Zeitkonstanten möglich ist bzw. wie sie von den Vorstufen eingesteuert werden. Am Ausgang der Endstufe entsteht so eine *maximale Änderungsgeschwindigkeit*, die in V/μs gemessen wird. Sie heißt allgemein Slew-Rate S_r und hängt stark von der jeweils gewählten Frequenzgangkorrektur und ihrer Lage im Verstärker ab. Der Zusammenhang mit der Frequenz f_m, bei der noch gerade die größte Aussteuerung möglich ist, ergibt sich aus der *größten Steilheit der Sinusschwingung*, nämlich bei ihrem Nulldurchgang, und dem maximalen Spitzen-Spitzen-Wert U_{ss} des OV:

$$S_r = 2\pi U_{ss} f_m \,. \tag{3}$$

7.5.3.7. Beispiel A 109

Für einen integrierten Operationsverstärker, wie z.B. den Typ A 109, liegen natürlich für die weiteren Anwendungen die optimalen Frequenzgangkorrekturen bereits in den Datenblättern vor. Sie hängen von der Verstärkung des gegengekoppelten Verstärkers ab und sind in Stufen zu 10 dB in Tab. 7.5.2 zusammengefaßt. Für Stufen im Abstand von 20 dB zeigt den Frequenzgang des offenen und geschlossenen (gegengekoppelten) OV Bild 7.5.6b) und c), während maximal erreichbare Spitzen-Spitzen-Spannung in Abhängigkeit von der Frequenz Teilbild d) zeigt.

Besonders starke Gegenkopplung liegt vor, wenn ein 1:1-Umkehrverstärker realisiert werden soll. Die Schaltung und die Daten mit innerer und äußerer Frequenzgangkorrektur zeigt Tab. 7.5.3 für unterschiedliche Korrekturen. Hieraus ist deutlich der zuvor (7.5.3.5.) diskutierte Einfluß auf Anstiegsgeschwindigkeit (Slew-Rate) und Rauschspannung zu erkennen; je größer die Slew-Rate, desto größer die Rauschspannung und umgekehrt.

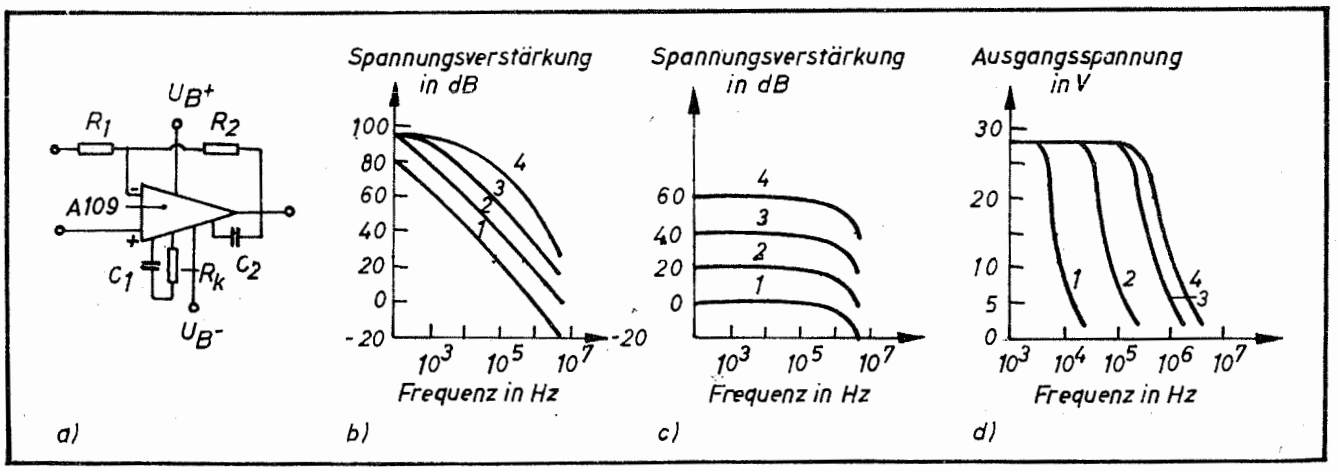


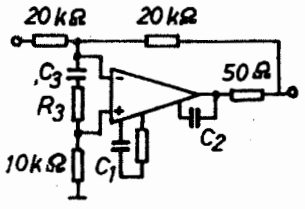
Bild 7.5.6. Eigenschaften des kompensierten OV vom Typ A 109. Die Werte 1 bis 4 beziehen sich auf Tab. 7.5.2

a) Anschalten der beiden Frequenzkorrekturen
b) Frequenzgang des offenen OV mit Korrekturen,
c) des gegengekoppelten OV
d) Maximal erreichbare Spitzen-Spitzen-Spannung am Ausgang

Tabelle 7.5.2. In etwa optimale Kompensationsglieder des A 109

Geschlossene Verstärkung in dB	C_1 in pF	R_k in kΩ	C_2 in pF	Kurve in Bild 7.5.6
0	5000	1,5	200	1
10	2700	1,5	100	—
20	500	1,5	20	2
30	270	1,5	10	—
40	100	1,5	3	3
50	27	1,5	3	—
60	10	0	3	4

Tabelle 7.5.3. Daten eines 1:1-Umkehrverstärkers gemäß nebenstehender Schaltung



Kompensationsart				
R_k in kΩ	1,5	0	1,5
C_1 in pF	5000	10	250
C_2 in pF	200	3	10
R_3 in Ω	—	39	1000
C_3 in µF	—	0,47	0,47
Bandbreite (3 dB) in kHz	500	500	500
Großsignalbandbreite in kHz	4,5	300	80
Anstiegsgeschwindigkeit in V/µs	0,3	22	6
Rauschspannung in mV	0,03	20	1,3

7.5.4. Ausgewählte Operationsverstärker

Zeitlich nach dem µA 709 kam bald (1968) der nächste bedeutende universelle OV, der µA 741, auf den Markt. Er hat auch noch heute beachtliche Bedeutung (7.5.4.1.). Neben gewissen Weiterentwicklungen bei universellen OV setzte von da ab immer stärker eine Entwicklung zu OV mit speziellen Eigenschaften ein, wobei andere Eigenschaften durchaus schlechter sein dürfen. Es ist schwierig, die verschiedenen Richtungen vollständig zu erfassen. Deshalb werden hier nur *einige Beispiele* gegeben:

- Universal — hohe Qualität
- Instrumentation — Offset $< 100\ \mu A$: Drift $< 1\ \mu V/K$
- FET-Eingang — extrem hoher Eingangswiderstand
- Hohe Geschwindigkeit — $S_r > 10\ V/\mu s$
- Hohe Spannung — $U_b > \pm 30\ V$
- Verstärker mit Eingangssteuerstrom
- Programmierbare — einstellbare Daten
- Maskenprogrammierbare

Tab. 7.5.4 gibt einen Überblick zu einigen typischen OV. Dabei sind die beiden FET-OV nach einer relativ neuen Technologie entwickelt worden, bei der auf einem Chip sowohl hochwertige SGFET als auch bipolare Transistoren realisiert werden können. So konnte für sie in etwa die gleiche geringe Drift wie bei bipolaren OV erreicht werden. Vor kurzem ist es auch gelungen, MOSFET mit Bipolartechnik zu verbinden. So entstanden BIMOS-OV. Die beiden Spezialverstärker wurden nach extrem geringer Drift bzw. extrem geringem Eingangsstrom ausgewählt.

Tabelle 7.5.4. Typische Daten integrierter Operationsverstärker ohne äußere Beschaltung bei $\pm$ 15 V Betriebsspannung

Parameter	Symbol	Standardverstärker		Spezialverstärker	
		µA 741 (bipolar)	LF 356 (FET)	µA 714A (bipolar)	3528 CM (FET)
Differenzverstärkung	V_D	10^5	10^5	$5 \cdot 10^5$	10^5
Gleichtaktunterdrückung	G	$3 \cdot 10^4$	10^5	10^5	10^4
3-dB-Bandbreite	f_g	10 Hz	50 Hz	2 Hz	7 Hz
Verstärkungs-Bandbreite-Produkt	f_T	1 MHz	5 MHz	1 MHz	700 kHz
Differenzeingangswiderstand	r_D	$10^6\ \Omega$	$10^{12}\ \Omega$	$6 \cdot 10^7\ \Omega$	$10^{13}\ \Omega$
Gleichtakteingangswiderstand	r_{Gl}	$10^9\ \Omega$	$10^{14}\ \Omega$	$2 \cdot 10^{11}\ \Omega$	$10^{15}\ \Omega$
Eingangsruhestrom	I_B	80 nA	30 pA	$\pm$ 0,7 nA	0,05 pA
Offsetspannung	U_0	1 mV	3 mV	10 µV	0,2 mV
Offsetspannungsdrift	$\Delta U_0/\Delta T$	6 µV/K	5 µV/K	0,2 µV/K	5 µV/K
Betriebsspannungsdurchgriff	$\Delta U_0/\Delta U_b$	15 µV/V	10 µV/V	3 µV/V	25 µV/V
Gleichtaktaussteuerbarkeit	$U_{Gl\,max}$	$\pm$13 V	+15 V, −12 V	$\pm$14 V	$\pm$12 V
Ausgangsaussteuerbarkeit	$U_{a\,max}$	$\pm$13 V	$\pm$13 V	$\pm$13 V	$\pm$12 V
Maximaler Ausgangsstrom	$I_{a\,max}$	$\pm$ 20 mA	$\pm$20 mA	$\pm$20 mA	$\pm$10 mA
Ausgangswiderstand	r_a	1 kΩ	50 Ω	60 Ω	1,5 kΩ
Betriebsstromaufnahme	I_b	1,7 mA	5 mA	2,5 mA	1 mA

7.5.4.1. µA 741/748/777

Die wesentlichen Besonderheiten des µA 741 gegenüber dem µA 709 sind interne Frequenzgangkorrektur und Ausgangsstrombegrenzung bei Kurzschluß. Seine Innenschaltung zeigt Bild 7.5.7. Der *Eingangsdifferenzverstärker* besteht aus Komplementärkaskadestufen, und zwar je einem *npn*-Transistor (T_1, T_2) mit hoher Stromverstärkung und je einem Lateral-*npn*-Transistor (T_3, T_4) mit geringer Stromverstärkung. Der

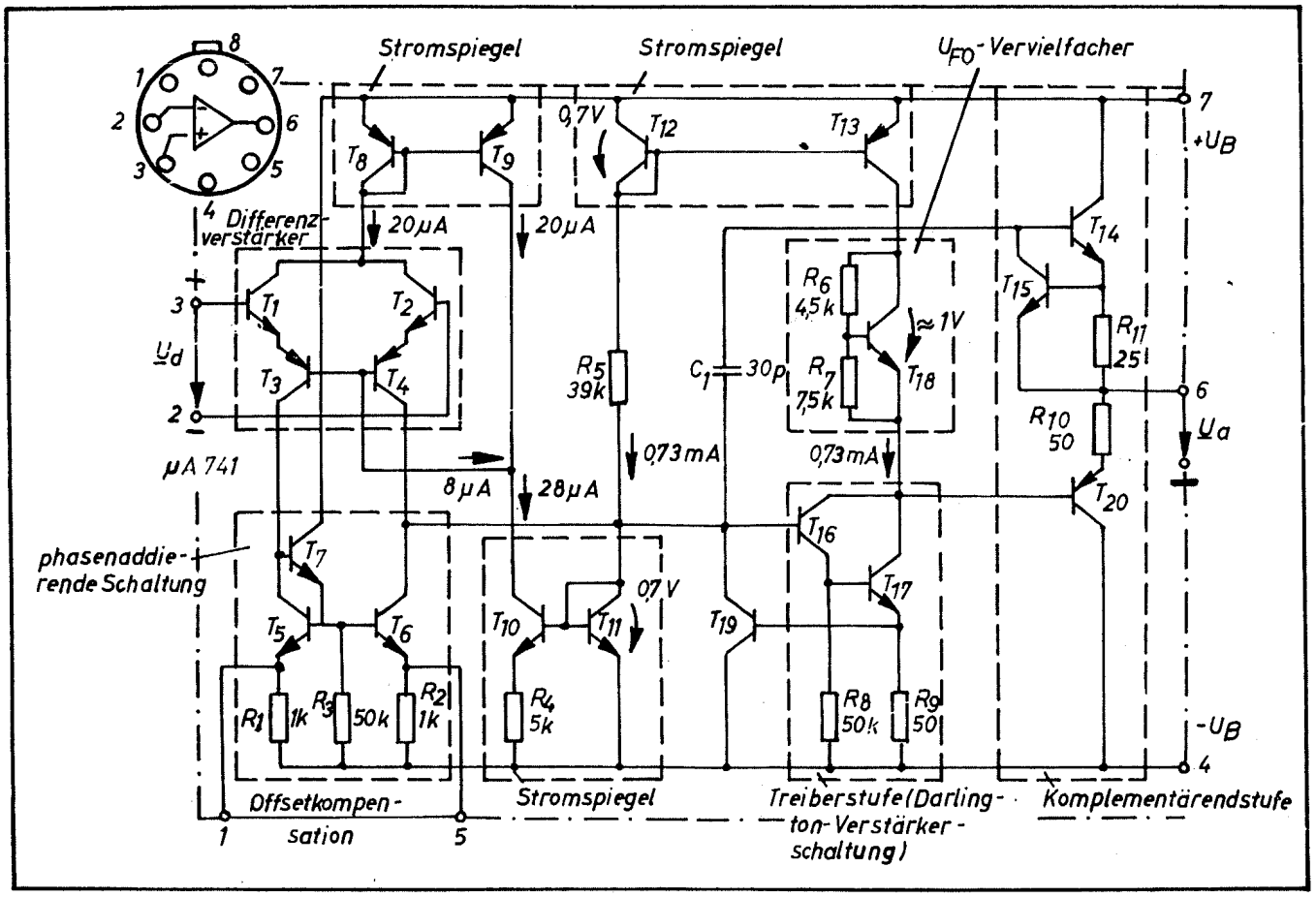


Bild 7.5.7. Innenschaltung des Operationsverstärkers µA 741 (T_{16} und T_{19} sind *npn*-Transistoren)

Emitterkreis enthält einen *Stromspiegel* (T_8, T_9). Die Verstärkung der Eingangsstufe ist etwa 1000fach.

Die *phasenaddierende Schaltung* ist wiederum ein Differenzverstärker (T_5 bis T_7). Durch sie gelangt der volle Kollektordifferenzstrom der Eingangsdifferenzstufe zur Basis von T_{16}. Hier beginnt die *Treiberstufe* als zweiter Verstärker (T_{16}, T_{17}), der aus einer DARLINGTON-*npn*-Stufe besteht. Diese besitzt infolge der Emitterwiderstände (R_8, R_9) einen sehr hohen Eingangswiderstand und mit dem Stromspiegel aus T_{12} und T_{13} einen aktiven Lastwiderstand, also ebenfalls eine große über 3000fache Leerlaufverstärkung.

Die *Komplementärendstufe* besteht aus T_{14}, T_{20}. Ihr Arbeitspunkt wird mittels der Konstantspannungsquelle aus T_{18} auf AB-Betrieb eingestellt. Der Ruhestrom beträgt 60 µA. Sie ist mit einer *Kurzschlußstrombegrenzung* (T_{15}, T_{19}, R_{10}, R_{11}) versehen. Wenn der Strom durch R_{11} 15 ... 20 mA übersteigt, wird T_{15} leitend und übernimmt einen Teil des für T_{14} bestimmten Basisstromes. Der untere Transistor wird mittels des Spannungsabfalls an R_{10} geschützt. Übersteigt er 15 ... 20 mA, so sinkt der Emitterwirkungsgrad von T_{20}. Der deshalb ansteigende Basisstrom fließt über R_9 und macht T_{19} leitend. Damit wird der Basisstrom von T_{16} reduziert. Die *interne Frequenzgangkorrektur* bewirkt sowohl eine integrierte Kapazität C_1 von 30 pF als auch eine erste Grenzfrequenz, die bei ca. 5 Hz liegt. Der Frequenzgang fällt von da an bis zur Verstärkung mit dem Faktor 1, d. h. bis 1 MHz, konstant mit 20 dB/Dekade ab. Dadurch ist der Verstärker für beliebige, aber phasenfreie Gegenkopplungen stabil. Die Bequemlichkeit wird allerdings durch *geringere Slew-Rate und Bandbreite* erkauft. Deshalb wurde später ein sonst identischer OV, jedoch ohne C_1, mit der Typenbezeichnung µA 748 herausgebracht. Eine geringfügige Weiterentwicklung bezüglich kleineren

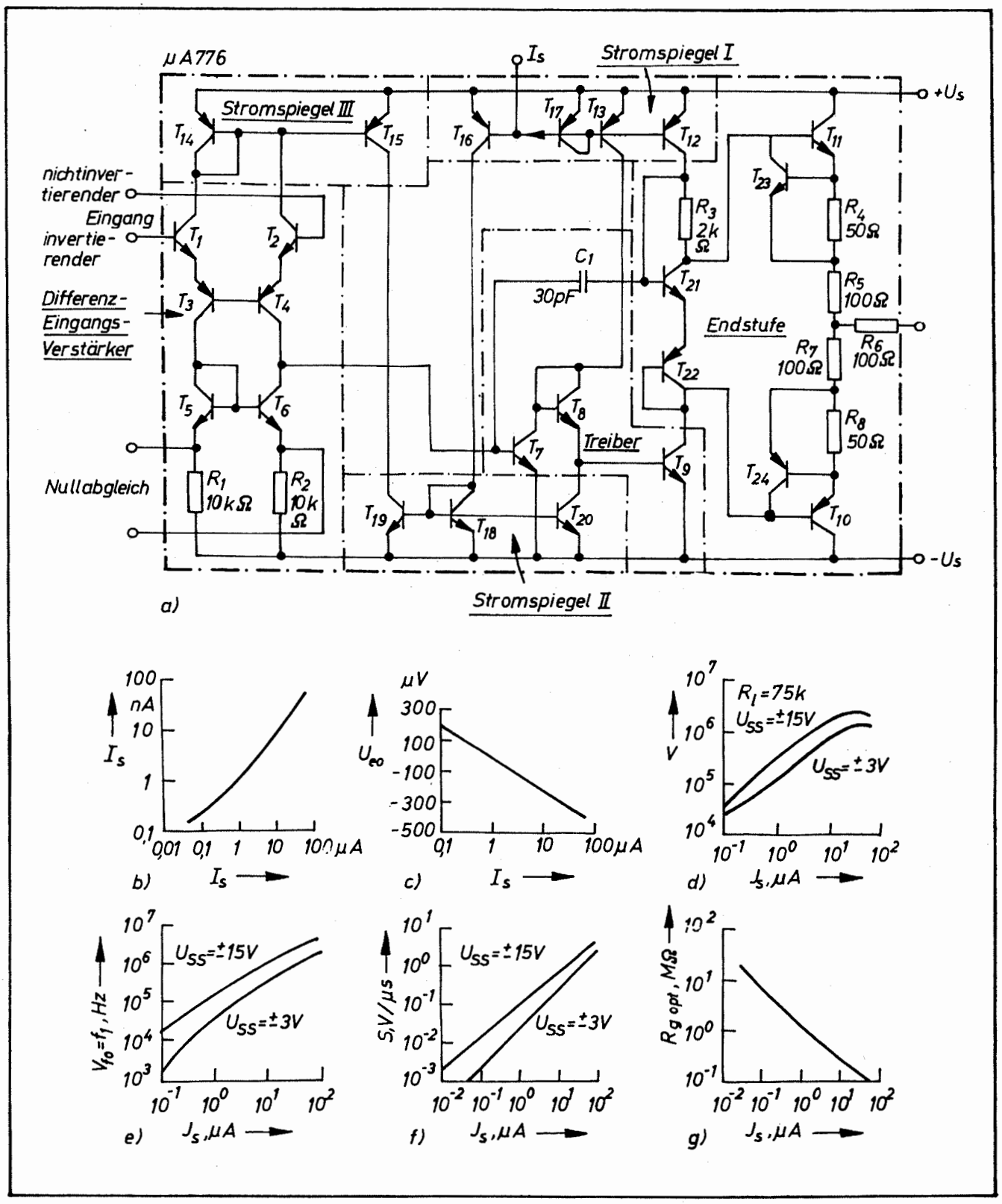


Bild 7.5.8. Programmierbarer Operationsverstärker μA 776

a) Innenschaltung
Abhängigkeiten von Daten vom Programmierstrom
b) Eingangsruhestrom
c) Offsetspannung
d) Leerlaufverstärkung
e) Grenzfrequenz mit der Verstärkung 1
f) Slew-Rate
g) Optimaler Generatorwiderstand

Eingangsstroms und geringerer Stromdrift ist der Typ μA 777. Äquivalent zum μA 741 sind K 140 УД 7 (SU), MAA 741 (ČSSR) und TA 72741 (UVR) und zum μA 748 MAA 748 (ČSSR) und TA 72748 (UVR).

7.5.4.2. Programmierbare Typen B 176 D, B 177 D und μA 776

Beim OV vom Typ μA 741 wird *der gesamte Ruhestrom mittels des Widerstandes R_5 festgelegt*. Ließe sich dieser Strom von außen beeinflussen, so könnten die meisten Daten des OV verändert werden. Dieser Gedanke führte zum programmierbaren OV vom Typ B 176 D, B 177 D bzw. μA 776, dessen Innenschaltung und einige programmierbare Daten Bild 7.5.8 zeigt. Gegenüber dem μA 741 wurden vor allem verändert

- Stromspiegel I zur Programmierung
- Strombegrenzung der Endstufe.

Die *Endstufe* ist jetzt voll symmetrisch gestaltet, und die Widerstände R_4 bzw. R_8 wirken mit den Transistoren T_{23} bzw. T_{24} genauso wie beim μA 741 (Bild 7.5.8) R_{11} und T_5. Der strombestimmende Widerstand R_5 vom μA 741 ist entfallen, dafür ist der Stromspiegel I beim μA 776 so gestaltet, daß er mit einem Strom I_S gegen Masse gesetzt wird. Der hierfür mögliche Strombereich liegt etwa zwischen 0,01 und 100 μA. Einige der sich damit stark ändernden Daten enthalten die Bilder 7.5.8 b) bis f). Kaum ändern sich dagegen z.B. die Gleichtaktunterdrückung (max. 85 bis 95 dB), die maximale Ausgangsspannung. Stärker ändern sich die Rauschspannung und der Rauschstrom des Verstärkers. Sie werden jedoch erheblich durch den optimalen Generatorwiderstand R_G beeinflußt. Äquivalent zum μA 776 ist der Typ K 140 УД 12 aus der SU.

7.5.3.4. Maskenprogrammierbare Typen

Folgende Operationsverstärker gehen auf ein *gemeinsames Layout* zurück: B 761 D (TAA 761), B 861 D (TAA 861), B 611 D (TCA 311), B 621 D (TCA 321) und B 631 D (TCA 331) (in Klammern äquivalente Siemens-Typen). Mit gleicher Innenschaltung gehört hierzu noch der Doppeloperationsverstärker B 2 761 D (TAA 2761). Die Grundschaltungen aller dieser Typen zeigt Bild 7.5.9a) Mitte, während links mögliche Eingangsänderungen und rechts mögliche Ausgangsänderungen gezeigt sind.

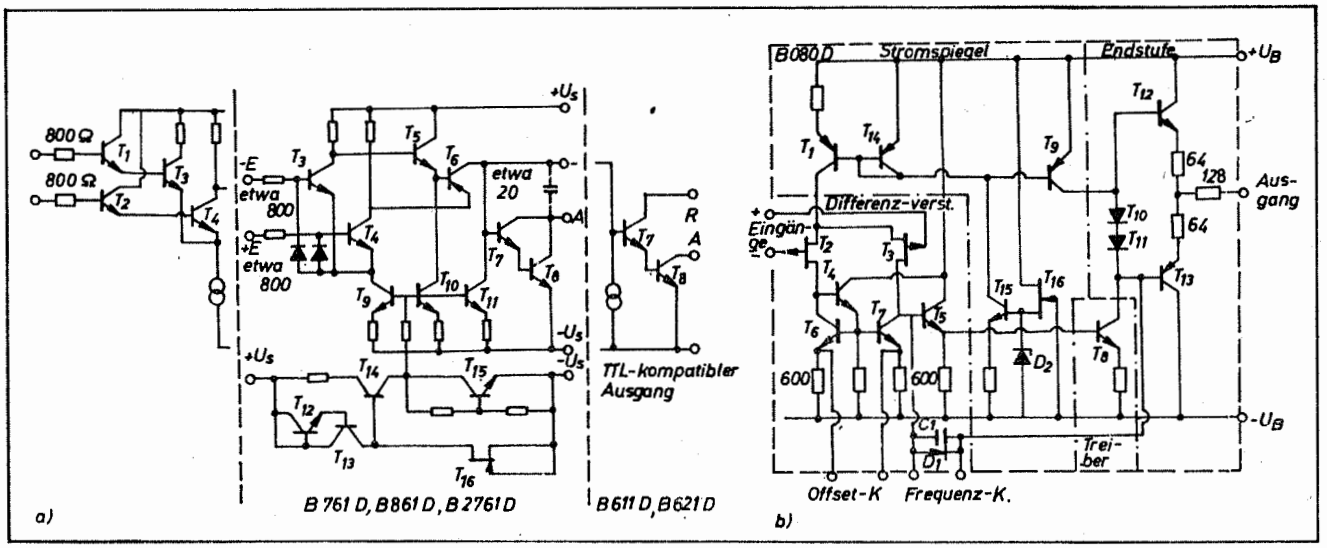


Bild 7.5.9. Schaltungsvarianten der Operationsverstärker um den Typ B 761 D (In a) ist T_{13} vom *npn*- und T_{13} vom *pnp*-Typ)

Die *Eingangsstufe* kann entweder mit zwei DARLINGTON-Transistoren (T_1 bis T_4) realisiert werden oder mit je zwei einfachen Transistoren (T_3, T_4), wobei dann die Transistoren T_1, T_2 als Schutzdioden fungieren. An den beiden Kollektorwiderständen wird das Ausgangssignal abgegriffen. Die Eingangsdifferenzstufe wird von der *Konstantstromquelle* T_9 (bis T_{11}) versorgt. Sie wird von der *Konstantspannungsquelle* T_{12} bis T_{16} mit einer temperaturstabilen Spannung von ca. 0,6 V versorgt. Die Transistoren T_5, T_6 bilden die *Phasenadditionsschaltung*, welche die Eintaktendstufe aus den Transistoren T_7 und T_8 steuert. Mit Ausnahme der Typen 611 und 621 ist die Endstufe als DARLINGTON-Transistor geschaltet. Es wird ein Arbeitswiderstand von etwa 2 kΩ benötigt. Seine Größe kann unter verschiedenen Gesichtspunkten ausgewählt werden, jedoch darf der maximale Strom von 70 mA nicht überschritten werden.

Alle OV außer 611, 621 und 2761 können von außen durch das gestrichelt angedeutete C mit ca. 3 bis 30 pF im *Frequenzgang korrigiert* werden. Der Typ 2761 enthält bereits intern 20 pF eingebaut und läßt daher keine weitere Korrektur zu. Die Typen 611 und 621 ermöglichen TTL-Kompatibilität, dann müssen A und R mit Abschlußwiderständen R_A und $R_R \approx 3$ bis 20mal R_A versehen werden. Einige Daten faßt Tab. 7.5.5 zusammen.

Tabelle 7.5.5. Daten der Operationsverstärker, die vom Typ B 761 D abgeleitet sind. Sie gelten für ± 15 V (861: ± 10 V) und $R_A = 2$ kΩ

Betriebsspannung	± 2 bis 15 V	861: ±1,5 bis 10 V; 761: ±1,5 bis 18 V
Ausgangsstrom	≦70 mA	
Eingangsoffset	≦ 6 mV	621: ≦ 7,5 mV 861: ≦10 mV 611 u. 631: ≦15 mV
	≦300 nA	611 u. 631: ≦25 nA
Eingangsstrom	≦ 1 µA	611 u. 631: ≦50 nA
Verstärkung	≦75 dB	761 u. 2761: ≦80 dB
Stromaufnahme	≦ 2,5 mA	861 u. 1761: ≦1,5 mA

7.5.4.4. BIFET-Operationsverstärker

Bild 7.5.9b) zeigt die Grundschaltung des B 080 D bzw. B 084 D (Äquivalenttypen TL 081 bis TL 084). Sie enthalten *vier Baugruppen*:

- Differenzverstärker mit Phasenaddition T_2 bis T_7
- Treiberstufe T_8
- Endstufe T_{13}, T_{14}
- Stromspiegel-Konstantstromquelle T_1, T_9, T_{14} bis T_{16}.

Aus den vorangegangenen Betrachtungen ist die Funktion der einzelnen Transistoren und deren Zusammenwirken mittels dieser Aufteilung leicht erkennbar. Daher erübrigt sich eine weitere Erklärung. Die Daten sind ähnlich denen des LF 356 in Tab. 7.5.4.

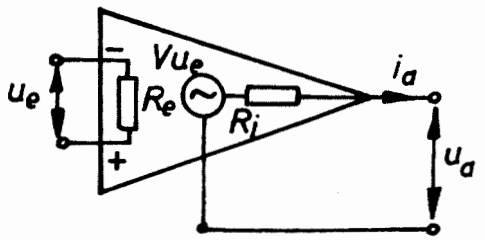
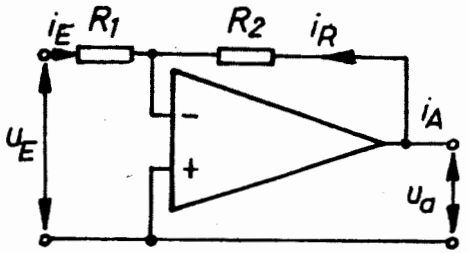
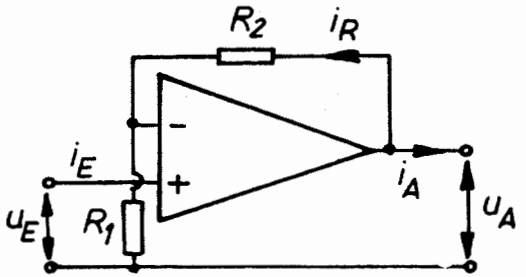
7.5.5. Die beiden Grundschaltungen des Operationsverstärkers

Die meisten Schaltungen mit OV lassen sich auf die beiden Grundschaltungen in Tab. 7.5.6,

Umkehrverstärker und *Elektrometerverstärker*,

zurückführen. Sie sind durch die beiden Widerstände R_1 und R_2 bestimmt, die allgemein komplex und nichtlinear sein können. Der durch sie gebildete Spannungs-

Tabelle 7.5.6. Prinzipschaltbild des Operationsverstärkers, seine Schaltung als Umkehr- und Elektrometerverstärker mit den zugehörigen wesentlichen Daten. Die Näherungen gelten bei $kV \gg 1$; $R_i \ll R_{2i}$; $R_e \gg R_1 + R_2$ und $R_i \ll R_e$. Es bedeutet $k = \dfrac{R_1}{R_1 + R_2}$

		Einsatzschaltung	Umkehrverstärker	Elektrometerverstärker
				
1	Ausgangs-Eingangs-Spannungsverhältnis	$V_U = \dfrac{u_A}{u_E} =$	$-\dfrac{R_2}{R_1}\dfrac{1}{1+\left(\dfrac{1}{k}+\dfrac{R_2}{R_e}\right)\dfrac{R_2+R_i}{V R_2 - R_i}} \approx -\dfrac{R_2}{R_1}$	$\dfrac{V + k\dfrac{R_i}{R_e}}{1 + kV + \dfrac{R_i}{R_1+R_2} + k\dfrac{R_2+R_i}{R_e}} \approx \dfrac{1}{k}$
2	Stromübersetzung	$V_I = \dfrac{i_R}{i_E} =$	$\dfrac{1}{1+\dfrac{1}{1+V}\left(1+\dfrac{R_i}{R_2}\right)} \approx 1$	$\dfrac{V + \dfrac{R_i+R_2}{R_e}}{1+\dfrac{R_i+R_2}{R_1}} \approx kV$
3	Eingangsstrom erzeugt Ausgangsspannung	$V_R = \dfrac{u_A}{i_E} =$	$\dfrac{-R_2}{1+\left(1+\dfrac{R_2}{R_e}\right)\left(\dfrac{R_2+R_i}{V R_2 - R_i}\right)} \approx -R_2$	$R_e \dfrac{V + k\dfrac{R_i}{R_e}}{1+k\dfrac{R_i}{R_1}} \approx R_e V$
4	Eingangsspannung erzeugt Strom durch R_2	$V_G = \dfrac{i_R}{u_E} =$	$-\dfrac{1}{R_i}\dfrac{1+V}{V+\dfrac{1}{k}+\dfrac{R_2+R_i}{R_e}+\dfrac{R_i}{R_1}} \approx -\dfrac{1}{R_1}$	$\dfrac{1}{R_1}\cdot\dfrac{1-\dfrac{R_1}{VR_e}}{1+\dfrac{1}{kV}+\dfrac{1}{V}\left(\dfrac{R_2+R_i}{R_e}+\dfrac{R_i}{R_1}\right)} \approx \dfrac{1}{R_1}$
5	Eingangswiderstand	$R_E = \dfrac{u_E}{i_E} =$	$R_1 + \dfrac{R_2+R_i}{V+1+\dfrac{R_2+R_i}{R_e}} \approx R_1 + \dfrac{R_2}{1+V} \approx R_1$	$R_e\left[1+\dfrac{k\left(V+\dfrac{R_2+R_i}{R_e}\right)}{1+\dfrac{R_i}{R_2+R_1}}\right] \approx R_e(1+kV)$
6	Ausgangswiderstand	$R_A = \dfrac{u_A}{i_A} =$	$\dfrac{R_i}{1+V\dfrac{R_1//R_e + R_i/V}{R_1//R_e + R_2}} \approx \dfrac{R_i}{1+kV}$	mit $R_1//R_e = \dfrac{R_1 \cdot R_e}{R_1+R_e}$

teiler,

$$k = \frac{R_1}{R_1 + R_2}, \tag{4}$$

bestimmt die Gegenkopplung des Verstärkers. Der Unterschied zwischen beiden besteht nur darin, *wie sie* dann *angesteuert werden.*

Für die Betrachtungen dieser Schaltungen wird ein vereinfachtes Modell des OV gemäß der in Tab. 7.5.6 gezeigten Ersatzschaltung angenommen. Die Größen dieses Modells, u. a. der Eingangswiderstand R_e und der Ausgangswiderstand R_a sowie die Ströme und Spannungen, erhalten kleingeschriebene Indizes. Entsprechende Größen der beiden gegengekoppelten Schaltungen erhalten dagegen große Indizes. Es können vor allem die folgenden *sechs Größen* bei beiden gegengekoppelten Verstärkern unterschieden werden. (Dabei ist zu beachten, daß *ausgangsseitig Leerlauf*, also $i_a = 0$, vorausgesetzt wird):

1 das Verhältnis von Aus- und Eingangsspannung V_U, welches auch als Spannungsverstärkung bezeichnet werden könnte,
2 die Stromübersetzung V_I, welche die Stromverstärkung bezüglich R_2 darstellt,
3 die von einem Eingangsstrom bewirkte Ausgangsspannungsänderung. Da sie die physikalische Einheit eines Widerstandes (Kernwiderstand) besitzt, erhält sie das Symbol V_R.
4 die von einer Eingangsspannung in R_2 bewirkte Stromänderung entspricht einem Leitwert V_G,
5 der Eingangswiderstand R_E,
6 der Ausgangswiderstand R_A.

Zu diesen Werten sind in der Tab. 7.5.6 zunächst die genauen Daten des Modells und dann vereinfachte Formeln angegeben. Die dafür notwendigen Voraussetzungen stehen in der Legende zur Tabelle.

Die vereinfachten Formeln und viele Betrachtungen zum OV werden durch ein *einfaches, anschauliches Modell* schnell wie folgt zusammengefaßt:

Infolge der sehr großen (unendlich großen) Verstärkung wird sich am Eingang eines gegengekoppelten OV immer ein solches Betriebsverhalten einstellen, daß die Spannung zwischen den beiden Differenzeingängen vernachlässigbar klein (Null) ist.

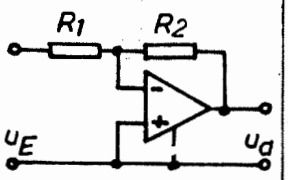
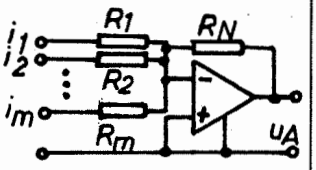
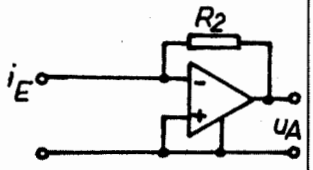
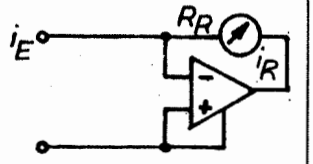
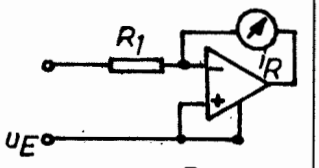
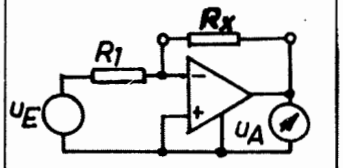
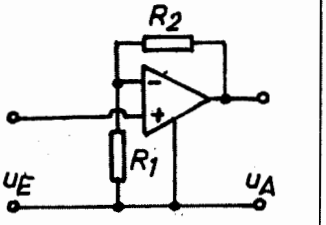
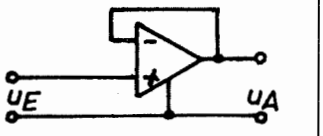
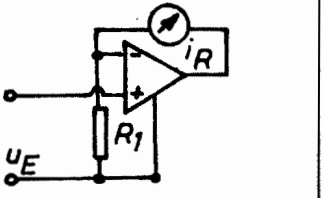
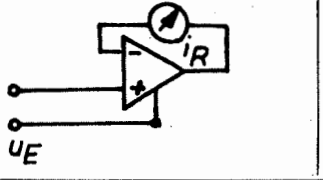
Der Unterschied zwischen den beiden Grundschaltungen besteht nun darin, daß beim *Umkehrverstärker* der invertierende Eingang scheinbar an Masse liegt. Die Eingangsspannung muß also voll am Widerstand R_1 abfallen. Dabei entsteht der Strom i_E, welcher ebenfalls durch R_2 fließen muß. So entsteht am Ausgang die Spannung

$$U_A \approx -\frac{R_2}{R_1} U_E \,. \tag{5}$$

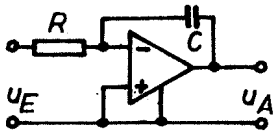
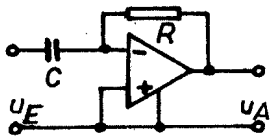
Hieraus sind alle weiteren Näherungswerte der Tab. 7.5.6 erhältlich. Lediglich für den Ausgangswiderstand ist eine andere Betrachtung notwendig. Er wird um die Kreisverstärkung von etwa $1 + kV$ verkleinert.

Beim *Elektrometerverstärker* wird der nicht invertierende und dazu unbeschaltete Eingang benutzt. Zu seiner Steuerung muß der invertierende Eingang so mitlaufen, daß wiederum die Potentialdifferenz zwischen beiden sehr klein (Null) ist. Dadurch kann die Verstärkung der Spannungsteilung aus R_1 und R_2 entsprechen. Durch das Mitlaufen des invertierenden Eingangs wird der Eingangswiderstand R_e des OV wie bei einer Bootstrap-Schaltung um die Kreisverstärkung von etwa $k \cdot V$ erhöht. So folgen wieder die weiteren Daten. Durch den sehr hohen Eingangswiderstand hat der Elektrometerverstärker seinen Namen. Als Elektrometer wurden früher Spannungsmesser bezeichnet, die so gut wie keinen Strom aufnahmen.

Tabelle 7.5.7. Beispiele linearer Schaltungen mit Operationsverstärkern mit $k = \dfrac{R_1}{R_1 + R_2}$ und Näherungsformeln

1	*Umkehrverstärker*		$\dfrac{u_A}{u_E} = -\dfrac{R_2}{R_1}$; $R_E = R_1$; $R_A = \dfrac{R_i}{1 + kV}$
2	Addierstufe		$R_{En} = R_n$ $u_A = -R_N \sum_{n=1}^{m} i_n$; $R_A = \dfrac{R_i}{1 + V \dfrac{\sum R_n}{R_N + \sum R_n}}$
3	Strom-Spannungs-Konverter		$u_A = -R_2 i_E$; $R_E = \dfrac{R_2 + R_i}{1 + V}$
4	Strommesser		$i_R = i_E$; $R_E = \dfrac{R_R + R_i}{1 + V}$
5	Spannungs-Strom-Konverter		$i_R = \dfrac{u_E}{R_1}$; $R_E = R_1$
6	Lineares Ohmmeter, einstellbare Spannungsquelle		$u_A = -\dfrac{u_E}{R_1} R_x$
7	*Elektrometerverstärker*		$\dfrac{u_A}{u_E} = k$; $R_E = R_e(1 + kV)$ $R_A = \dfrac{R_i}{1 + kV}$
8	Elektrometerimpedanzwandler (Spannungsfolger)		$u_A = u_E$; $R_E = R_e(1 + V)$ $R_A = \dfrac{R_i}{1 + V}$
9	Spannungs-Strom-Konverter		$i_R = \dfrac{1}{R_1} u_E$; $R_E = R_e(1 + V)$
10	Hochempfindlicher Spannungs-Strom-Konverter		$i_R = \dfrac{1}{R_e(1 + V)} u_E$; $R_E = R_e(1 + V)$

Fortsetzung Tabelle 7.5.7

11	*Integrator*		$u_A = -\frac{1}{RC}\int u_E \, dt; f_u = \frac{1}{2\pi VRC}$
12	*Differentiator*		$u_A = -RC\frac{du_E}{dt}$; $f_o = \frac{V}{2\pi RC}$

7.5.6. Einfache lineare Schaltungen

Die wichtigsten linearen Schaltungen faßt Tab. 7.5.7 zusammen. Sie sind in drei Gruppen geteilt, je nachdem, ob sie sich aus dem Umkehrverstärker (1 bis 6) oder aus dem Elektrometerverstärker (7 bis 10) ableiten bzw. zur Integration oder Differentation angewendet werden.

Die *Addierstufe* faßt alle Spannungen bzw. Ströme, die an den summierenden Eingängen 1 bis m stehen wie folgt zusammen:

$$u_A = -R_N \sum_{n=1}^{m} i_n = -R_N \sum_{n=1}^{m} \frac{U_n}{R_n}. \tag{6}$$

Dabei verhalten sich die einzelnen Eingänge zueinander rückwirkungsfrei, denn die Spannung zwischen den beiden Eingängen des OV ist praktisch Null.

Der *Strom-Spannungs-Konverter* (3) erzeugt aus einem Eingangsstrom bei sehr kleinem Spannungsabfall (gegen Null) eine Ausgangsspannung, die proportional dem Wert von R_2 ist. Nach diesem Prinzip läßt sich gemäß 4 ein *Strommesser* aufbauen, der einen extrem kleinen Eingangswiderstand besitzt.

Sollen *Spannungs-Strom-Konverter* aufgebaut werden, so bestehen gleich drei Möglichkeiten. In der Schaltung 5 bestimmt der Widerstand R_1 sowohl den Eingangswiderstand als auch den Faktor der Wandlung. Dies ermöglicht den Aufbau von Präzisionsschaltungen. Darüber hinaus besteht bei dieser Schaltung die zusätzliche Möglichkeit der Addition mehrerer Spannungsgrößen gemäß Schaltung 2. Die Schaltung 9 hat den gleichen Wandlungsgrad, jedoch mit entgegengesetzten Vorzeichen. Ihr Eingangswiderstand ist jedoch sehr groß, wenn auch weniger genau definiert. Die Schaltung 10 hat darüber hinaus auch einen großen, jedoch ebenfalls nicht sehr stabilen Wandlungsgrad.

Schaltung 6 stellt ein lineares *Ohmmeter* dar. Die Ausgangsspannung ist direkt proportional dem unbekannten Widerstand R_x, und mittels R_1 und u_e kann der zu messende Widerstandsbereich gewählt werden.

Mit der Schaltung 6 kann ohne Änderung auch eine über R_x *einstellbare Spannungsquelle* realisiert werden. Die Ausgangs- (Gleich- oder Wechsel-) Spannung wird durch u_e und R_1 sowie R_x bestimmt.

Schaltung 8 stellt einen *Spannungsfolger* mit sehr hohem Eingangs- und sehr kleinem Ausgangswiderstand dar. Es handelt sich also um eine Schaltung, die einem extrem verbesserten Emitter- bzw. Sourcefolger entspricht.

Beim *Integrator* (11) bzw. *Differentiator* (12) wird ein definierter Frequenzgang erzeugt, der direkt proportional $1/\omega$ bzw. ω ist und somit die Eigenschaften gemäß Abschnitt 2.3.1., Bild 2.3.2 erfüllt. Der Fehler solcher Schaltungen ist direkt proportional der Restspannung, die nicht an R (Integrator) bzw. C (Differentiator) ab-

fällt. Diese wird aber in den beiden Schaltungen 11 und 12 gerade sehr klein (bzw. Null). Insofern realisieren diese Schaltungen praktisch ideal diese Funktionen, und zwar solange, wie an den Eingangsgliedern R bzw. C eine hinreichend große Spannung abfällt und der OV wenig Phasenfehler besitzt. Die Wirkung der Schaltung 11 läßt sich auch so interpretieren, daß C zum Eingang hin auf V-mal vergrößert wird. Dies entspricht dem MILLER-*Effekt* (5.5.3., Bild 5.5.5). Beim Differentiator wird der Widerstand R um den Faktor V verkleinert.

Bei beiden Schaltungen lassen sich die Prinzipien der Integration bzw. Differentation durch mehrere R bzw. C bezüglich *mehrerer Eingangsgrößen* wieder gemäß der Schaltung 1 addierend zusammenfassen. Für die eine entsprechende Integrationsschaltung gilt dann z.B.

$$u_A = \sum_{n=1}^{m} -\frac{1}{R_n C} \int u_{En}(t)\, dt\,. \tag{7}$$

In Bild 7.1.1b) bis e) wurden *die vier möglichen Gegenkopplungsschaltungen* eingeführt. Mittels OV lassen sie sich in idealer Weise einzeln realisieren. Dies zeigt Bild 7.5.10. Dort sind nur die Grenzfälle für unendlich große Verstärkung des OV eingetragen. Mit Hilfe des Abschnitts 7.1.1. lassen sich aber relativ leicht alle Eigenschaften — falls nötig — direkt bestimmen.

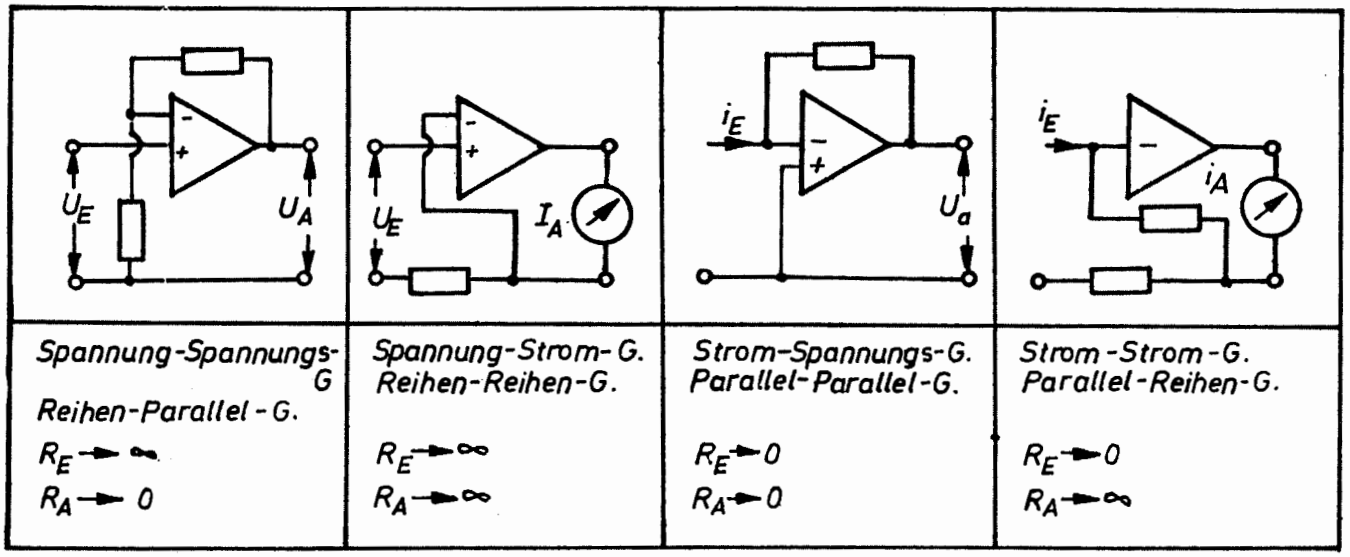


Bild 7.5.10. Die vier prinzipiellen Grundtypen der Rückkopplung, am Beispiel des Operationsverstärkers dargestellt. Die Tendenzpfeile gelten für unendlich große Verstärkung des OV

7.5.7. Nichtlineare Schaltungen

Nichtlineare Schaltungen sind in den meisten Fällen schwieriger als lineare zu realisieren. Erst die integrierte Schaltkreistechnik hat hier in den letzten Jahren erhebliche Fortschritte gebracht. Es können *zwei Gruppen von nichtlinearen Funktionen* unterschieden werden:

- *stetige*, die etwa einer Funktion $y = f(x)$ gehorchen;
- *unstetige*, die im Prinzip durch Schalter realisiert werden.

Die *stetigen Funktionen* werden heute vor allem nach *drei Prinzipien* realisiert:

- *logarithmischer* Kennlinienzusammenhang bei Dioden und Transistoren,
- *Proportionalität* zwischen *Emitterstrom* und *Steilheit* bei Bipolartransistoren,

- *Approximation* der Kennlinien durch lineare Teilabschnitte, zwischen denen weitgehend unstetig umgeschaltet wird.

Die *unstetigen Funktionen* betreffen im wesentlichen

- ideale Gleichrichtung, Begrenzung und Signumfunktion.

Darüber hinaus gehören zu den unstetigen Funktionseinheiten auch alle Baugruppen, die bereits in Abschnitt 7.3.7. und 7.3.8. im Sinne von

- *Rechteck-, Dreieck- und Sägezahngeneratoren* sowie als
- *Impulsschaltungen* (u. a. monostabiler Multivibrator, SCHMITT-Trigger)

behandelt wurden.

7.5.7.1. Logarithmische und exponentielle Verstärker

Bild 7.5.11 zeigt in der linken Spalte, wie mit einer nichtlinearen Kennlinie $i = f(u)$ ein entsprechender Zusammenhang zwischen Aus- und Eingangsspannung realisiert werden kann. Dies ist unmittelbar aus Tab. 7.5.6 bzw. 7.5.7 für den Umkehrverstärker ableitbar. Darunter befindet sich das Bild für die Realisierung der Umkehrfunktion $u = f^{-1}(i)$ mit dem gleichen nichtlinearen Glied. Es ist dazu nur in den Gegenkopplungszweig zu legen. In der Mitte des Bildes befinden sich die daraus abgeleiteten Schaltungen für den Logarithmus bzw. für die Exponentialfunktion. Sie gehen auf die exponentielle Kennlinie der Diode zurück (vgl. 3.2.2., Gl. (5)):

$$I = I_S(\exp(U_D/U_T) - 1) \quad \text{mit} \quad U_T = e_0/kT\,. \tag{8a}$$

Für Spannungen U, die mindestens 100 mV ($U_T \approx 25$ mV) groß sind, gilt die Näherung

$$I \approx I_S \exp(U_D/U_T) \tag{8b}$$

über mehrere Zehnerpotenzen des Stromes.

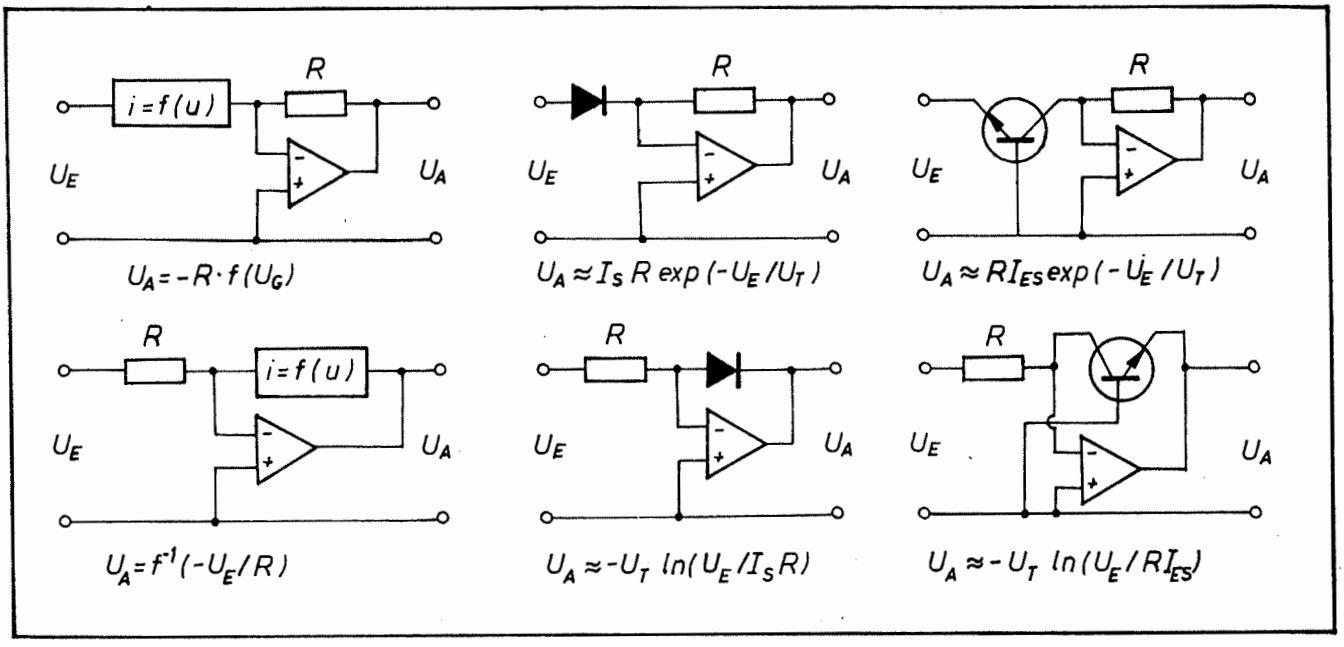


Bild 7.5.11. Zum Aufbau von Schaltungen mit Operationsverstärkern, die nichtlineare Kennlinien realisieren. Jeweils oben und unten die zueinandergehörenden Umkehrfunktionen. Links: Eine beliebige Funktion $i = f(u)$ und deren Umkehrung $u = f^{-1}(i)$. Mitte: Aus Dioden und rechts: aus Transistoren abgeleitete exponentielle (oben) und logarithmische (unten) Verstärker

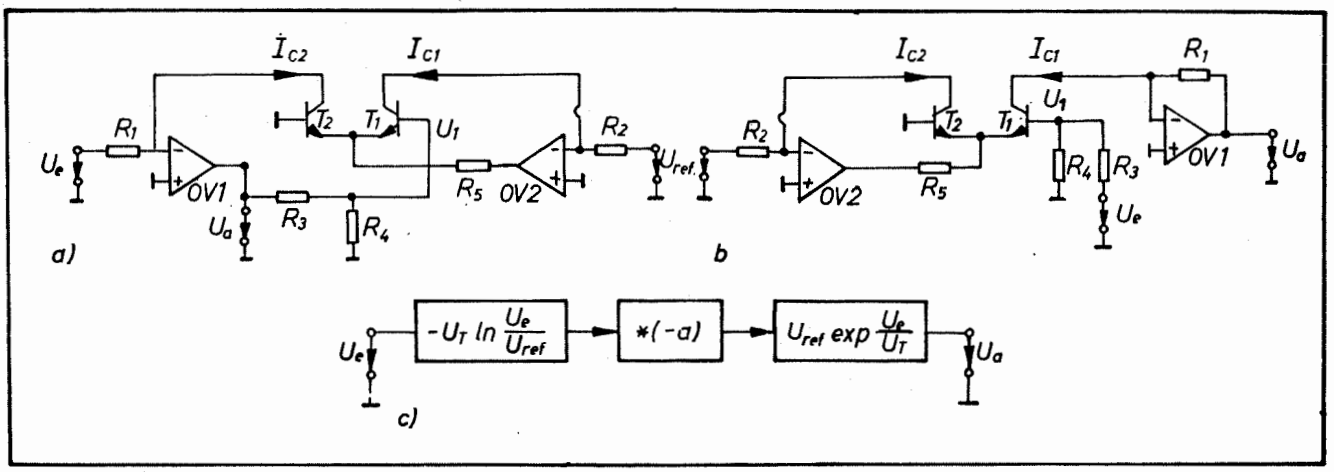


Bild 7.5.12. Aufbau von temperaturkompensierten Schaltungen

a) Logarithmierender Verstärker
b) Verstärker mit exponentieller Funktion
c) Prinzip eines Verstärkers, der die Potenzfunktion realisiert und aus den beiden anderen Verstärkern a) und b) besteht und einem dazwischen geschalteten Verstärker, der die Verstärkung a realisiert. Hierfür ist u. a. die Schaltung 6 aus Tab. 7.5.7 geeignet

Die analogen Gleichungen gelten auch für einen Bipolartransistor, dessen Kollektorspannung Null ist (3.4.2.). Dabei ist I_S durch I_{ES} und U_D durch U_{BE} zu ersetzen:

$$I_C \approx I_{ES} \exp (U_{EB}/U_T) . \tag{8c}$$

Für Si-Planartransistoren kann hierbei ein besonders großer Bereich für I_C überstrichen werden. I_{ES} liegt für Vorstufentransistoren bei 10^{-13} bis 10^{-14} A. Die entsprechenden Schaltungen zeigt der rechte Teil von Bild 7.5.11.

Die einfachen Schaltungen von Bild 7.5.11 besitzen den großen Nachteil, daß U_T stark von der Temperatur abhängt. Dies läßt sich korrigieren, wenn zwei solche Schaltungen aufgebaut werden, von denen eine gesteuert und die andere mit einer Bezugsspannung U_{ref} betrieben wird, und dann die Differenz der Ausgangsspannung von beiden genutzt wird. Solche Schaltungen existieren in vielen Varianten, u. a. in [H 45, K 42, P 1, S 41, T 4].

Bild 7.5.12 zeigt je eine Variante für den logarithmischen und den exponentiellen Verstärker. Es zeigt auch, wie aus beiden die Potenzfunktion zu realisieren ist.

7.5.7.2. Funktionsgeneratoren mit Beispiel für Sinus

Funktionen können durch *Geradenapproximation* genähert werden. Eine solche Approximation weicht von der üblichen mathematischen Approximation durch Polynome ab, ist dafür aber technisch gut realisierbar. Dies erfolgt, indem bei bestimmten, wählbaren Spannungen *mittels Dioden zusätzliche Widerstände* ein- bzw. ausgeschaltet werden. Ein Beispiel hiervon zeigt Bild 7.5.13a). Sobald an einem der Spannungsteiler die Spannung am Teilerpunkt einen solchen Wert erreicht, daß die entsprechende Diode durchschaltet, fließt ein zusätzlicher Strom zum invertierenden Eingang des OV. Dies bedeutet, daß mit steigender Eingangsspannung der Strom zum OV überproportional zunimmt. Die Kennlinie zeigt also generell einen nach oben gekrümmten Verlauf, wie es im Bild oben angedeutet ist. Die realisierte Kurvenform hängt dabei von den durch die Spannungsteiler bestimmten Umschaltpunkten ab. Unter anderem sind nach diesem Prinzip x^2, x^3, arcsin x und ähnliches zu realisieren, natürlich nur in einem Quadranten. Dabei ist generell von Vorteil, daß die Dioden nicht plötzlich schalten. So entsteht vorteilhaft ein allmählicher Übergang zwischen den Geradenstücken.

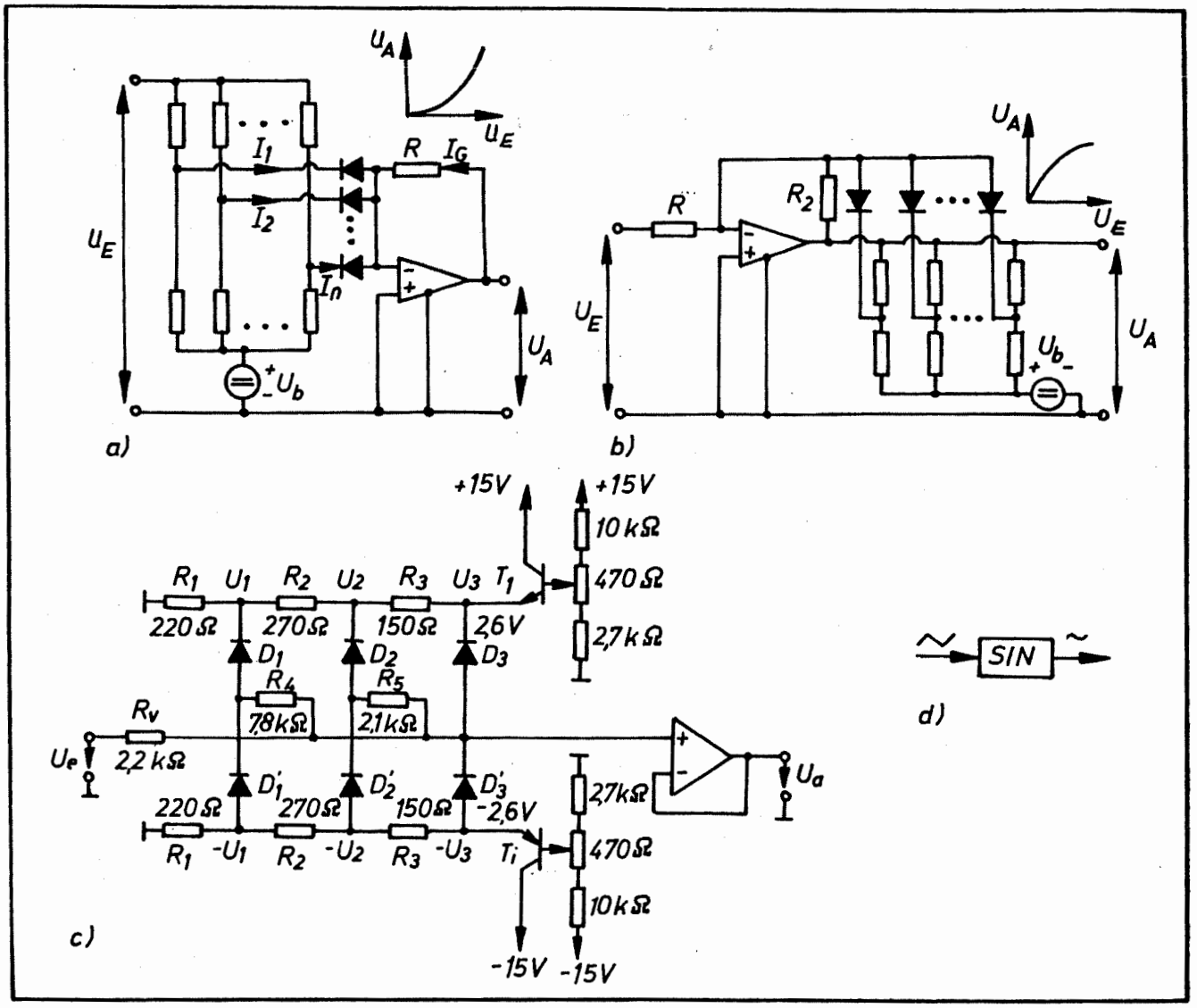


Bild 7.5.13. Funktionsgeneratoren mit Diodenteilern

a) Prinzip mit wachsendem Strom vor einem OV erzeugt steiler werdende Kurven
b) Dasselbe im Gegenkopplungszweig erzeugt flacher werdende Kurven
c) Für die Erzeugung einer Sinusfunktion dimensionierte Schaltung mit dem Spannungsteilerprinzip
d) Wirkung der Schaltung c) bei einem Eingangs-Dreiecksignal

Wird dasselbe Bauelement in den Gegenkopplungszweig geschaltet, so entsteht Bild 7.5.13b). Jetzt sind Funktionen zu realisieren, die nach unten gekrümmt sind, also z. B. $\sqrt{x}$ und $\sin x$.

Ein anderes Prinzip für Funktionsgeber zeigt Bild 7.5.13c). Hier ist ein Spannungsteiler vor den Operationsverstärker geschaltet. Zunächst gelangt die volle Eingangsspannung über R_V zum OV. Beim Erreichen der Spannung U_1 bzw. $-U_1$ wird R_4 belastend eingeschaltet und bewirkt eine Spannungsteilung mit R_V. Wird U_2 bzw. $-U_2$ erreicht, so schaltet sich auch noch R_5 dazu und bewirkt eine weitere Spannungsteilung. Die Spannung U_3 bzw. $-U_3$ wirkt schließlich als absolut begrenzend. Mit dem dargestellten Netzwerk wird in recht guter Approximation eine Sinus- bzw. Kosinuskurve realisiert (weitere Details in [T 4]). Deshalb erzeugt die Schaltung *aus einem symmetrischen Dreieck eine brauchbare Sinusschwingung*, wie es das Bild 7.5.13d) andeutet.

7.5.7.3. Multiplikatoren

Die Multiplikation ist wie die Addition oder Substraktion eine Funktion von zwei Eingangsgrößen. Sie ist damit allgemeinerer Natur als die vorangegangenen Funk-

tionen und hat darüber hinaus eine *Vielzahl von Anwendungen*. Es seien hier nur die Beispiele *Modulation, Demodulation, Mischung von Frequenzen* und Phasenregelkreis genannt. Sie erfordern jedoch nicht die Präzision der speziellen Kennlinien, wie sie bei einer möglichst (numerisch) exakten Multiplikation von analogen Signalen verlangt wird. Dabei sind je nach dem Vorzeichenbereich der beiden Eingangsgrößen U_1 und U_2 folgende Methoden zu unterscheiden:

$U_1 \geqq 0$	und	$U_2 \geqq 0$	Einquadrantenmultiplikation
U_1 beliebig	und	$U_2 \geqq 0$	Zweiquadrantenmultiplikation
U_1 beliebig	und	U_2 beliebig	Vierquadrantenmultiplikation

Am einfachsten ist die Einquadrantenmultiplikation zu realisieren, während allgemein die Vierquadrantenmultiplikation gewünscht ist.

Im Laufe der Entwicklung hatten bzw. haben die in Bild 7.5.14 aufgezeigten vier Prinzipien Bedeutung. Sie werden hier vereinfacht so betrachtet, als ob U_1 und U_2 einheitenfreie Größen im Sinne von U_1/U_{Bezug} sind.

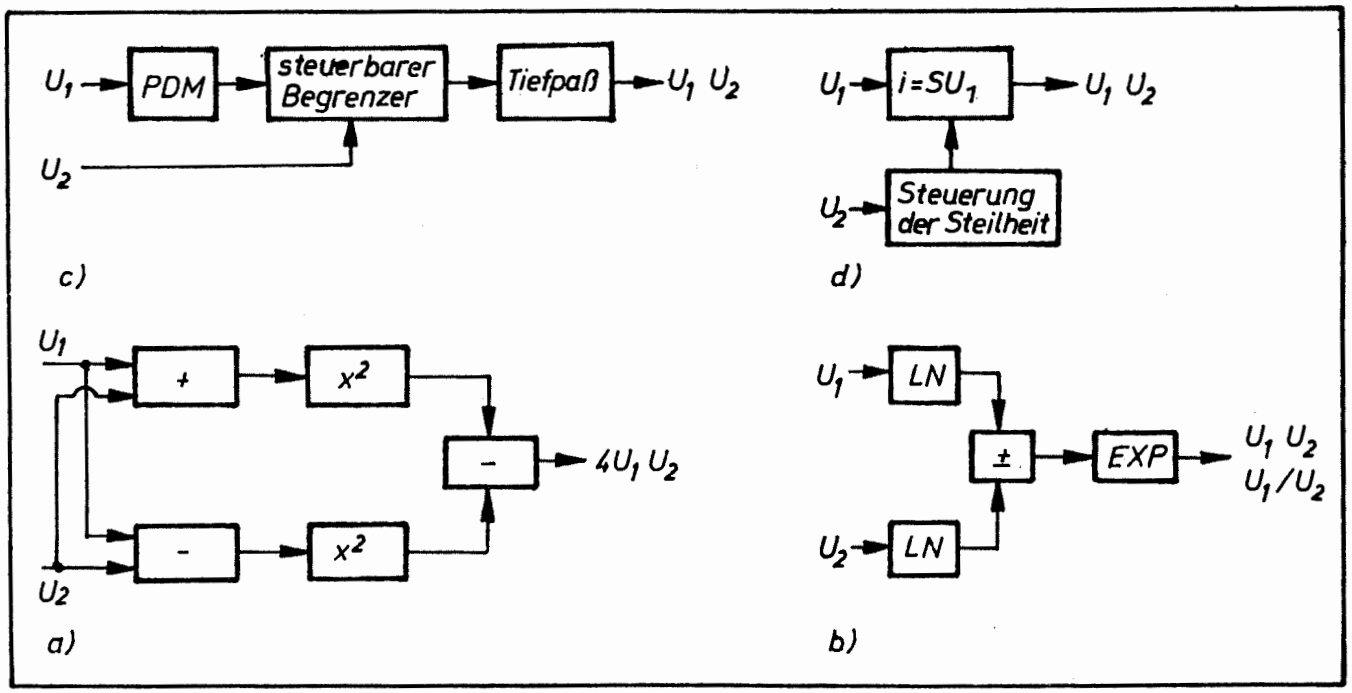


Bild 7.5.14. Vier Schaltungsprinzipien zur Multiplikation:

a) mittels $4U_1U_2 = (U_1 + U_2)^2 - (U_1 - U_2)^2$

b) mittels $U_1U_2 = \exp(\ln U_1 + \ln U_2)$ bzw. der analogen Beziehung für Division

c) Time-division-Prinzip mit Pulsdauermodulation und einem spannungsproportionalen Begrenzer

d) mittels Steilheitsmodulation

Die Methode in Teilbild a) beruht auf der *mathematischen Äquivalenz* von

$$4U_1U_2 = (U_1 + U_2)^2 - (U_1 - U_2)^2 . \tag{9}$$

Die Schwierigkeit ist hier in die beiden *Quadrierstufen* verlagert. Sie wurden früher mit Hilfe von speziell ausgesuchten CuO_2-Gleichrichtern sehr gut genähert. Die Methode hat heute kaum noch Bedeutung.

Eine andere *mathematische Äquivalenz*,

$$U_1U_2 = \exp(\ln U_1 + \ln U_2) , \tag{10}$$

nutzt das Schema von Teilbild b) aus. Es erreicht in der Regel Genauigkeiten um 0,1 bis 1%. Es hat darüber hinaus den Vorteil, daß damit auch leicht die Division gemäß

$$U_1/U_2 = \exp(\ln U_1 - \ln U_2) \tag{11}$$

realisiert werden kann. Es ist aber relativ aufwendig, u. a. wegen der schon im vorigen Abschnitt beschriebenen Probleme der Temperaturstabilisierung.

Das dritte Verfahren ist das *Time-division-Prinzip* (Teilbild c)). Es nutzt keine mathematische Relation, sondern die schon in den Abschnitten 7.2.2., 7.3.8.4. und 7.4.6.2. beschriebene Pulsdauer- bzw. Pulslängenmodulation aus. Sie wird z. B. mittels U_1 erzeugt. In einer zweiten Stufe wird mit einem steuerbaren Begrenzer der Spitzenwert proportional U_2 gemacht. Nach dem Tiefpaß steht dann das Produkt $U_1 U_2$ zur Verfügung. Problematisch sind hier die Taktfrequenz und die etwa 200- bis 500mal höhere Flankensteilheit der Frequenz des Signals (vgl. auch 9.1.3.).

Die letzte hier noch genauer zu beschreibende Methode (Teilbild d)) nutzt einen *Zusammenhang des Bipolartransistors bezüglich seiner Steilheit*

$$\frac{I_C}{U_{BE}} = S = I_C/U_T \tag{12}$$

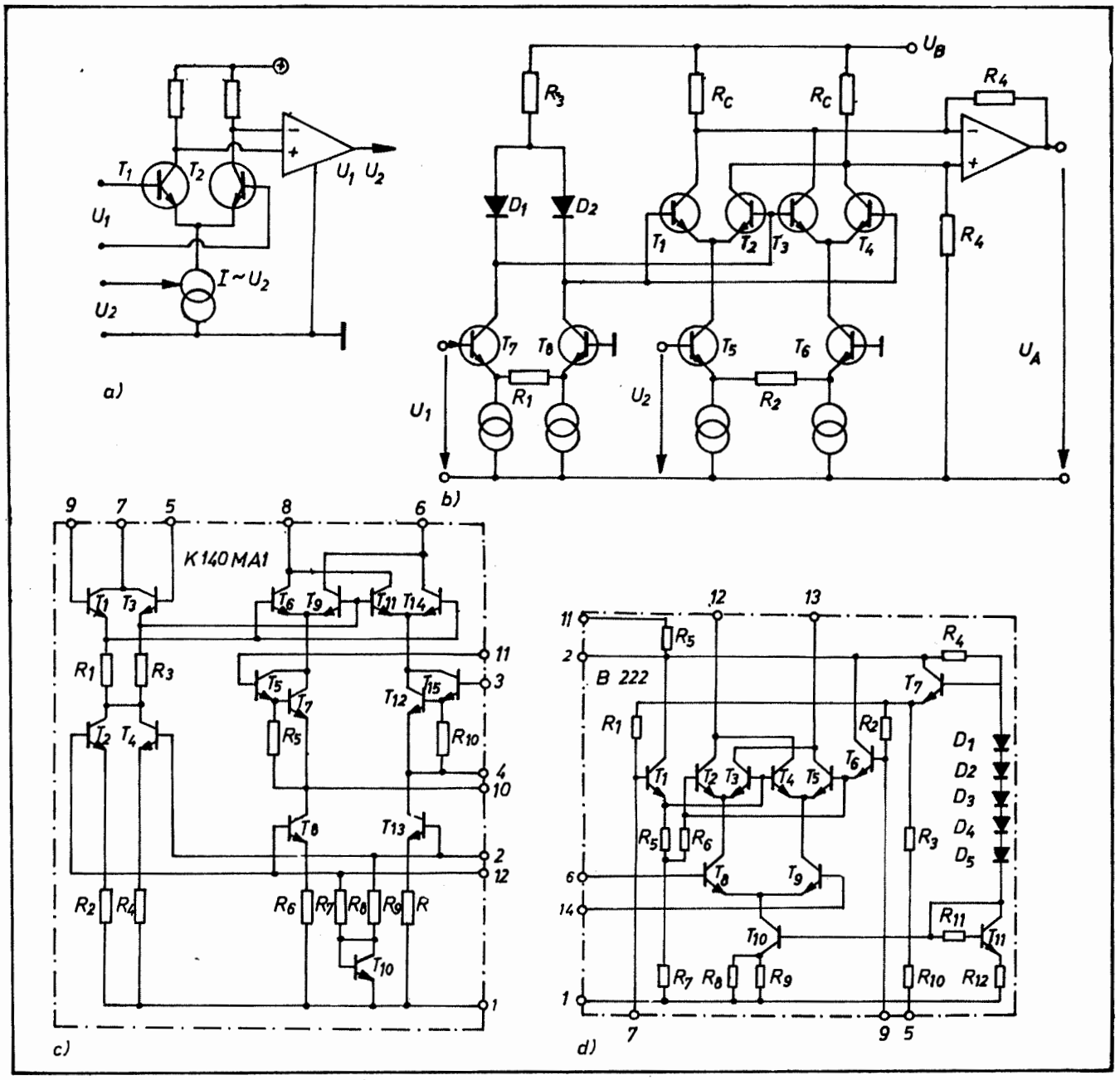


Bild 7.5.15. Multiplikatoren nach dem Prinzip der gesteuerten Steilheit

a) Prinzipschaltung für zwei Quadranten

b) Grundschaltung für vier Quadranten

c) Schaltung des Vierquadrantenmultiplikators vom Typ K 140 MA 1 (UdSSR)

d) vom Typ B 222 D (HFO) (T_{10} ist vom *npn*-Typ)

aus (vgl. Gl. (5.6.17)). Er ist mit guter Genauigkeit für nicht zu große Kollektorströme (I_C etwa kleiner 100 µA) über viele Dekaden erfüllt. Praktische Schaltungen sind u. a. in [H 24, B 2, K 42, R 4, S 41, T 4] ausführlicher behandelt. Das Grundprinzip geht aus Bild 7.5.15a) hervor. Der Differenzverstärker T_1T_2 wird mit der Spannung U_1 gesteuert. Dies ermöglicht ein Zweiquadrantenverfahren. Die Steilheitssteuerung des Verstärkers erfolgt über eine von U_2 gesteuerte Stromquelle. Die Erweiterung dieser Schaltung zur Vierquadrantenmethode unter Berücksichtigung von Temperaturstabilität führt zur Prinzipschaltung von Bild 7.5.15b). Zwei relativ einfache industrielle Schaltungen zeigt Bild 7.5.15c) und d).

Ist eine Multiplikatorschaltung vorhanden, so sind daraus leicht Quadrier-, Divisions- und Radizierschaltungen realisierbar. Hierzu sei auf die oben genannte Literatur verwiesen.

7.5.7.4. Gleichrichterschaltungen

Die einfachen Gleichrichterschaltungen aus den Abschnitten 5.4.4. und 5.4.5. bringen Probleme, wenn *kleine Spannungen* gleichgerichtet werden sollen, dann stört nämlich die Schwellspannung beträchtlich. Hier lag ja einer der Vorteile der im Abschnitt 6.6.2. behandelten Backwarddiode. Wesentlich günstigere Verhältnisse ergeben sich jedoch, wenn die Gleichrichtung mit einem OV verbunden wird, wie es z. B. Bild 7.5.16a) zeigt. Dann liegt die gesamte Gleichrichteranordnung im Gegenkopplungszweig. Der Eingangsstrom aus u_E/R wird, unabhängig vom Widerstand der Dioden, gleich groß in den Gegenkopplungszweig übertragen (vgl. Tab. 7.5.7, Nr. 5). Auf diese Weise

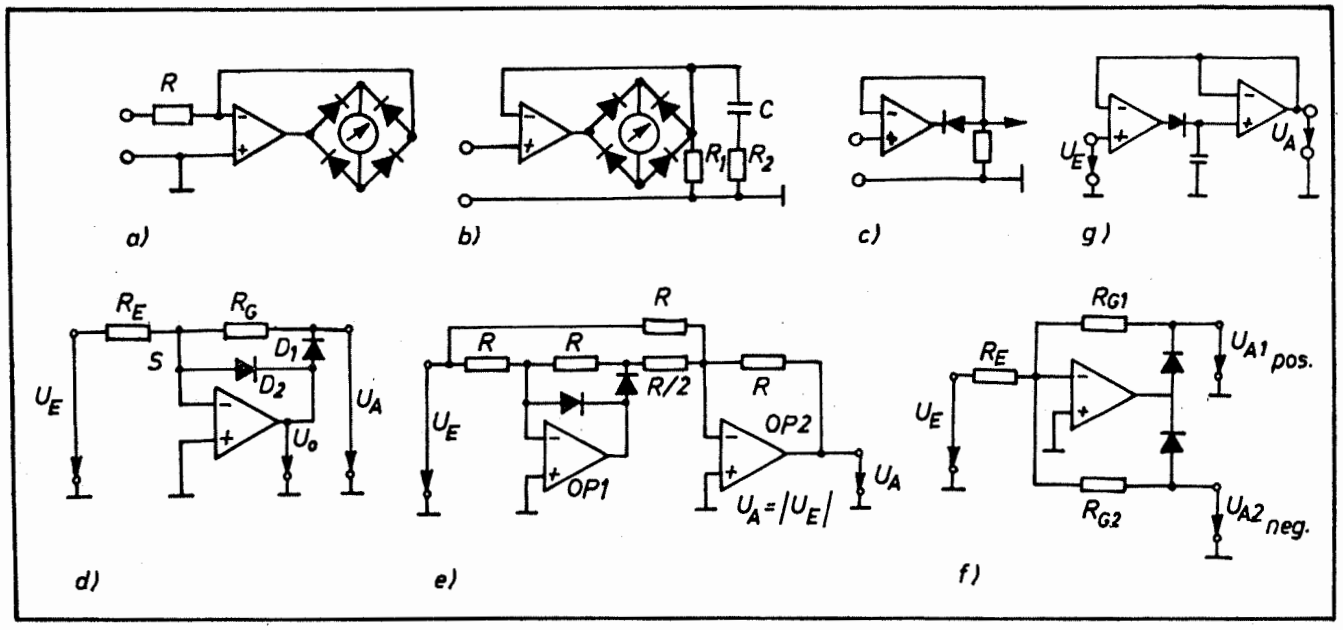


Bild 7.5.16. Gleichrichterschaltungen mit Operationsverstärkern

a) Ideal linearisierter GRAETZ-Gleichrichter

b) Ideal linearisierter GRAETZ-Gleichrichter, jedoch mit zusätzlicher Anzeige des Effektivwertes für Sinuswechselgrößen Dabei muß gelten

$$R_2 = \frac{R_1}{\pi/\sqrt{8} - 1} \approx 9{,}0317 R_1 .$$

c) Einfacher Einweggleichrichter mit geerdeter Ausgangsspannung

d) Verbesserter Einweggleichrichter

e) Aus d) abgeleiteter Zweiweggleichrichter durch Addition der halben Eingangsspannung

f) Schaltung eines Polaritätstrenners

g) Spitzenwertgleichrichter (vgl. S. 622)

erfolgt eine exakte und *hochlineare Zweiweggleichrichtung*. Anders ausgedrückt, liegt eine Betragsbildung vor. Das Vorzeichen der Eingangsspannung geht nicht mehr ein.

Wird die Schaltung 7.5.16 mit dem Elektrometerverstärker (Teilbild b)) realisiert, so bestehen sogar weitere Möglichkeiten. Für Gleichsignale bleibt C unwirksam, und es liegen die Verhältnisse von Tab. 7.5.7, Nr. 9, vor. Für Wechselstrom wird über C jedoch R_2 parallel zu R_1 geschaltet. Hierdurch kann die Wechselstromanzeige auf Effektivwerte korrigiert werden (vgl. Tab. 1.4.3).

Die bisher besprochenen beiden Schaltungen haben den großen Nachteil, daß die *gleichgerichtete Spannung nicht auf Nullpotential* liegt und daher schwierig weiter zu verarbeiten ist. Für die Einweggleichrichtung zeigt Teilbild c) eine sehr einfache Schaltung. Eine bessere Lösung zeigt Bild 7.5.16d). Für positives Eingangssignal wird u_0 negativ, die Diode D_2 leitend und damit u_A gleich Null, denn der Punkt S liegt virtuell auf Masse. Außerdem sperrt D_1. Für negative Eingangssignale ist D_2 gesperrt und D_1 leitend. Das Verhältnis $R_G:R_E$ bestimmt daher die Ausgangsspannung.

Aus der vorangegangenen Schaltung kann ein Zweiweggleichrichter gewonnen werden, wenn in halber Amplitude die Eingangsspannung addiert wird. So entsteht die Zweiweggleichrichterschaltung von Teilbild 7.5.16e).

In Bild 7.5.16f) ist eine Schaltung gezeigt, welche in geringer Änderung von Teilbild c) die beiden Halbwellen — genauer die Anteile mit positiven und negativen Vorzeichen — getrennten Ausgängen zuführt. Solche Schaltungen werden auch Polaritätstrenner genannt.

Im Zusammenhang mit Bild 1.4.10 und Tab. 1.4.3 wurde gezeigt, daß es bei *nicht sinusförmigen Wechselspannungen* sinnvoll sein kann, die *Spitzenwerte* zu messen. Dies leistet keine der bisher gezeigten Schaltungen. Das Problem wird deutlicher beim Rückblick auf Bild 5.4.4, insbesondere Teilbild c). Die Differenz gegenüber einer Spitzenspannung hängt u. a. davon ab, wie groß der Innenwiderstand der Quelle R_i und der Flußwiderstand einer Diode R_F ist. Sie hängt genauer von dem Verhältnis

$$X = (R_i + R_F)/R_L \tag{13}$$

ab, wobei R_L der Lastwiderstand (Eingangswiderstand des Folgeverstärkers/Meßinstruments mit Parallelwiderstand in Sperrichtung der Diode und Kondensatorverlusten) ist. Dieses Verhältnis bestimmt in erster Näherung, wie gut relativ kurze Impulse (Stromflußwinkel) bei einer Spitzenwertgleichrichtung erfaßt werden. So entsteht eine Schaltung gemäß Bild 7.5.16g), die diesbezüglich durch die beiden OV gute Werte aufweist.

7.5.7.5. Ringmodulatorschaltungen

Der *Ringmodulator* wurde im Abschnitt 5.4.9. einschließlich seiner vielen vorteilhaften Anwendungen eingeführt. Seine Nachteile waren das Problem der hohen *Symmetrie und die Transformatoren*. Deshalb entstanden im Laufe der Zeit mehrere Schaltungen mit äquivalenten Eigenschaften, jedoch unter weitgehender Vermeidung der Nachteile. Diese Schaltungen nutzen im wesentlichen die Schalterersatzschaltung von Bild 5.4.13f) aus. Der Schalter kann dann z.B. ein Differenzverstärker gemäß Bild 7.5.17a) sein, während ω einen Stromspiegel steuert. Die Weiterentwicklung dieses Prinzips führt wieder zur *Multiplikatorschaltung* von Bild 7.5.15.

Ein abweichendes Prinzip läßt sich mittels eines Operationsverstärkers und eines Schalttransistors realisieren. Seinen Grundbaustein zeigt Bild 7.5.17b). Der Operationsverstärker wird in Umkehrschaltung betrieben. Seine Verstärkung bei gesperrtem Transistor ist dann durch R_2/R_1 gegeben. Wird der Transistor geöffnet, so liegt parallel zu R_2 sein sehr kleiner Durchlaßwiderstand $R_D \leqq 10\ \Omega$. Bei richtiger Wahl von R_1 kann die Verstärkung sehr klein werden. Insgesamt wird also der Operationsverstärker

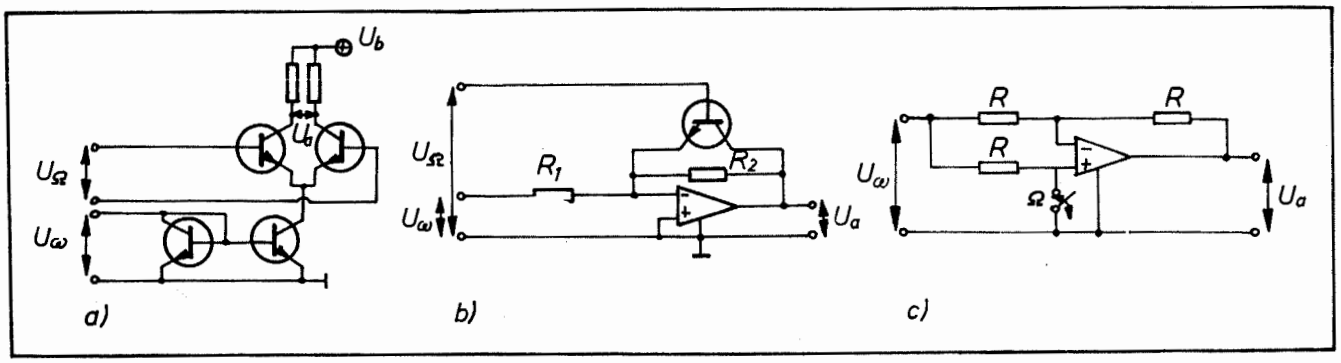


Bild 7.5.17. Ringmodulatorschaltungen

a) Auf der Basis von Differenzverstärkern

b) Schalten der Verstärkung eines Operationsverstärkers

c) Umschalten zwischen +1 und −1

auf- und zugeschaltet. Mit zwei solchen bezüglich u_Ω in Gegenphase betriebenen Stufen ist wiederum eine ringmodulatoräquivalente Schaltung aufzubauen. Für erdbezogene Spannungen sind ihnen eine Symmetriestufe für u_Ω vor- und ein Differenzverstärker für u_a nachzuschalten.

Ein weiteres Beispiel zeigt Bild 7.5.17c). Hier wird der Schalter im Ω-Takt geschlossen und geöffnet. Alle drei Widerstände sind gleich groß. Ist der Schalter geöffnet, so beträgt die Verstärkung der Schaltung −1. Wenn der Schalter geschlossen ist, so ist zu beachten, daß sich am invertierten Eingang immer etwa das Potential Null ausbildet. Wenn dann R klein gegen den Eingangswiderstand des Operationsverstärkers ist, beträgt seine Verstärkung also +1. Dies bedeutet, daß mit Öffnen und Schließen des Schalters das Vorzeichen der Verstärkung umgeschaltet wird. Damit liegt eine komplette ringmodulatoräquivalente Schaltung vor.

7.5.7.6. Begrenzer, Komparator und SCHMITT-Trigger

Begrenzer, Komparator und SCHMITT-Trigger sind bereits im Zusammenhang mit Bild 7.3.25 eingeführt. Sie unterscheiden sich durch folgende Haupteigenschaften:

Ein *Begrenzer* erzeugt an seinem Ausgang eine Spannung, die maximal zwischen zwei vorgegebenen Grenzwerten U_1 und U_2 liegt. Zwischen diesen beiden Grenzwerten besteht ein linearer Zusammenhang von Ein- und Ausgangsspannung. Begrenzer werden zuweilen auch als Clipper bezeichnet.

Ein idealer *Komparator* ist nur der beiden Werte U_1 und U_2 am Ausgang (und keiner Zwischenwerte) fähig. Die Umschaltung zwischen diesen beiden Werten erfolgt bei der zum Vergleich dienenden Spannung U_V. Ein Begrenzer mit sehr großer Verstärkung, also sehr steilem Übergang von U_1 nach U_2 und umgekehrt, ist bereits ein Komparator.

Ein SCHMITT-*Trigger* besitzt gegenüber einem Komparator eine Hysterese bezüglich der Eingangsspannung (vgl. Bild 7.3.23 bzw. 7.3.27d)). Er besitzt im Gegensatz zum Begrenzer keinen linearen Übergangsbereich, sondern echtes Schaltverhalten. Der Name SCHMITT-Trigger geht auf den Erfinder der entsprechenden Röhrenschaltung zurück.

Alle drei Schaltungen sind also geeignet, von einer analogen Eingangsgröße zu einer digitalen Ausgangsgröße überzugehen, wobei U_1 und U_2 dann die beiden logischen Pegel (vgl. 8.2.3.) sind. U_V ist der Vergleichswert, bei dem die Umschaltung erfolgt, und ΔU der Schaltwellenbereich. Die Anwendung der drei Schaltungen ist sehr vielfältig und nicht auf den Übergang zu digitalen Signalen begrenzt. Ein anderes Beispiel ist der in Bild 7.3.28 gezeigte Fall, wie mittels eines Komparators eine PDM

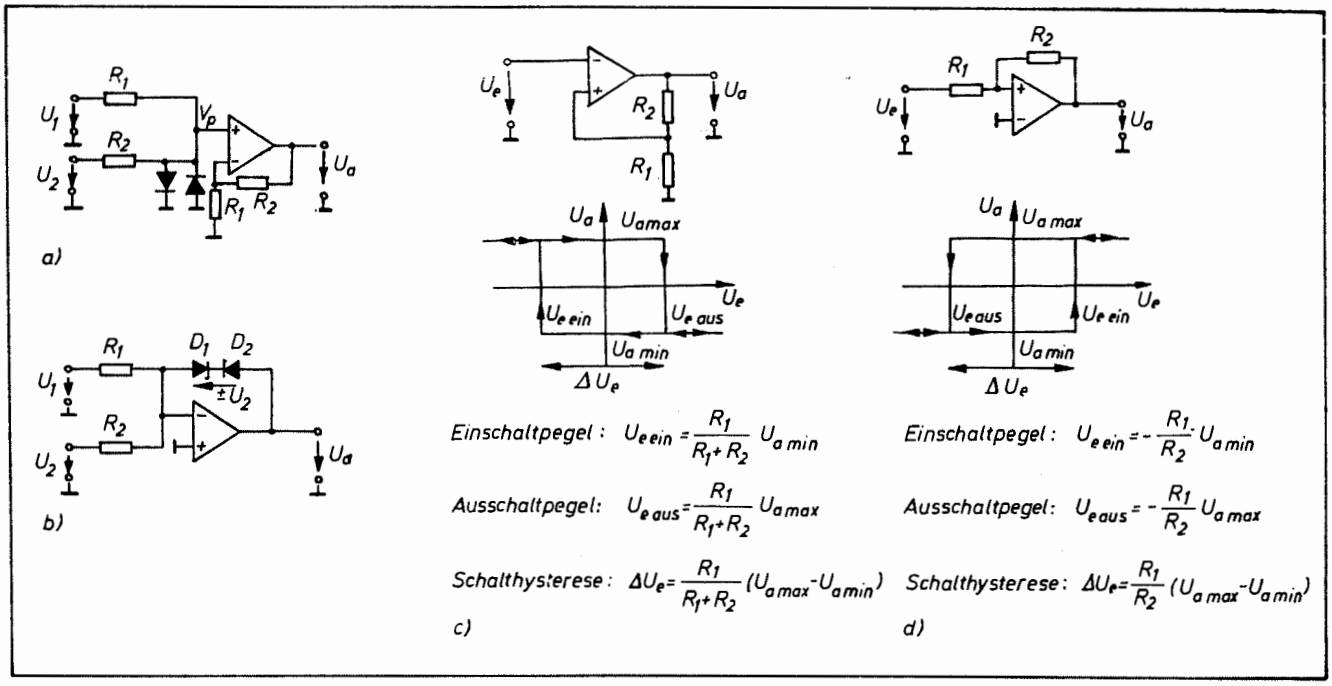


Bild 7.5.18. Schaltungen von Begrenzer, Komparator und SCHMITT-Trigger mit Operationsverstärkern

a) Begrenzer und Komparator (für $R_1 \to 0$ und $R_2 \to \infty$). Die Dioden begrenzen die mögliche Eingangsspannung auf $\pm U_S$ (Schwellenspannung), und die Ausgangsspannung beträgt $U_{min} \leqq -U_S R_1/R_2$ bzw. $U_S R_1/R_2 \leqq U_{max}$

b) Komparator für $U_1/R_1 = -U_2/R_2$ mit den Ausgangsspannungen $\pm U_Z$

c) SCHMITT-Trigger mit invertierendem Eingang

d) mit nicht invertierendem Eingang,

erzeugt wird. Bild 7.5.18 zeigt eine Auswahl für diese Schaltungen auf der Basis eines OV. Weitere Schaltungen enthalten die Bilder 7.3.27a) (Begrenzer) und 7.3.27c) (SCHMITT-Trigger).

Es gibt auch *spezielle OV*, die *als Komparator* besonders gut geeignet sind. Sie zeichnen sich durch die *folgenden Eigenschaften* gegenüber OV aus:

- Slew-Rate größer,
- Totzeit geringer.

Dafür wird *weniger beachtet*:

- Verstärkung (kann kleiner sein),
- Drift (kann größer sein).

7.5.7.7. Modulations- bzw. Chopperverstärker

Selbst die besten OV besitzen bei kritischen Anwendungen immer noch zu hohe Drift und Offset. Dadurch ist die Gleichspannungsverstärkung begrenzt. Ein Ausweg bietet hier der Modulationsverstärker, dessen Prinzip Bild 7.5.19a) zeigt. Die Gleichgröße wird in einem sehr sorgfältig zu bemessenden Modulator in eine modulierte Größe, z.B. AM, FM oder Pulsmodulation, umgesetzt. Die Größe ist dann auf unterschiedlichste Weise zu verstärken. Bei AM erfolgt die Verstärkung, indem der Träger reduziert und damit der *Modulationsindex* erhöht wird; bei FM dadurch, daß die *Trägerfrequenz zu tieferen Werten* verlagert (additiv verschoben) und so der relative Frequenzhub größer wird. Für Pulsmodulationen bestehen analoge Prinzipien. Danach wird das modulierte Signal rückgewandelt (demoduliert), und es entsteht ein entsprechend großes (verstärktes) Ausgangssignal. Ein Beispiel für dieses Prinzip wurde indirekt schon im Zusammenhang mit Bild 5.4.14 bezüglich der Lock-in-Schaltung

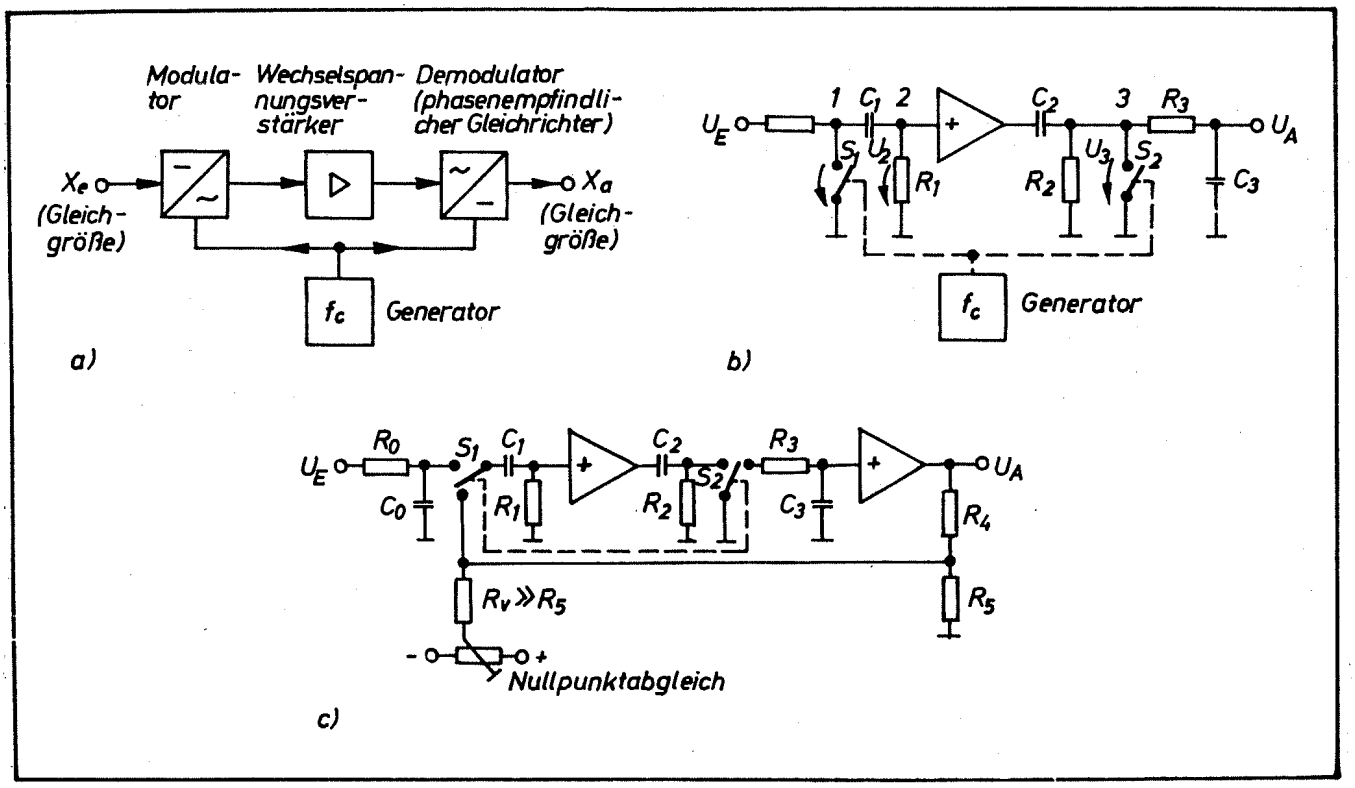


Bild 7.5.19. Chopper- bzw. Modulationsverstärker

a) Prinzip des Modulationsverstärkers
b) Prinzip des Chopper- bzw. Zerhackerverstärkers
c) Prinzip eines Chopperverstärkers mit Gegenkopplung

bzw. des *phasenempfindlichen Gleichrichters* erklärt. Ein fast gleichartiges Wirkprinzip liegt beim Chopper- bzw. *Zerhackerverstärker* vor. Hier werden am Ein- und Ausgang je ein Schalter (S_1 und S_2 in Bild 7.5.19b)) verwendet, die mit dem Generator im Takt der Frequenz f_c geschaltet werden. Infolgedessen werden die Kondensatoren C_1 und C_2 periodisch auf Masse geschaltet. Die Gleichspannung U_E am Punkt ① wird in eine Wechselspannung mit dem Spitzen-Spitzenwert U_E umgewandelt und verstärkt. Sie ist also driftfrei. Am Ausgang Punkt ③ liegt die verstärkte Wechselspannung vor. Der Schalter S_2 wird zu den Zeitpunkten, wo U_E auf Null geschaltet war, ebenfalls auf Null festgelegt. Hierfür ist der Ausdruck Klemmschaltung gebräuchlich, und so entsteht als Mittelwert zwischen der maximalen Ausgangsspannung und Null (geklemmter Wert) wieder die richtige, aber stark verstärkte Gleichspannung. Als Drift gehen hierbei nur Kontaktspannungen des Schalters S_1 ein. Um diese Spannungen klein zu halten, kann für den Schalter S_1 auch eine durch f_c gesteuerte Kapazität bzw. ein anstelle von C_1 liegender Schwingkondensator liegen. Daten derartiger Verstärker im Vergleich zu einem sehr guten OV mit Driftkorrektur faßt Tab. 7.5.8 zusammen.

Schließlich können alle Chopper- und Modulationsverstärker auch zur weiteren Stabilisierung ihrer Eigenschaften *gegengekoppelt* werden. So entsteht die Prinzipschaltung von Bild 7.5.19c).

7.5.7.8. Multivibratoren und Funktionsgeneratoren

Gemäß Bild 7.3.25 werden drei Arten von Multivibratoren unterschieden: mono-, bi- und astabile. Auch sie sind in idealer Weise mittels OV zu realisieren. Ein Beispiel des Monoflop zeigte Bild 7.3.26c), ein Beispiel des astabilen Multivibrators Bild 7.3.23d). Als *bistabiler Multivibrator* kann der SCHMITT-*Trigger* in allen seinen Varianten be-

Tabelle 7.5.8. Vergleich einiger Daten von Verstärkern für kleine Gleichgrößen

	Modulationsverstärker			Verstärker mit Driftkorrektur
	FET-Zerhacker	Varicap	Schwing-kondensator	
Eingangswiderstand	$10^7\,\Omega$	$10^{-10}\,\Omega$	$10^{16}\,\Omega$	10^7 ... $10^{12}\,\Omega$
Nachweisgrenze	0,01 ... 0,1 μV	10^{-14} A	10^{-16} A (10^{-17} A)	0,1 ... 1 μV
Temperaturdrift der Offsetspannung	10 ... 20 nV/K	10 μV/K	50 μV/K	$<$0,1μV/K
Langzeitdrift der Offsetspannung	$<$0,1 μV/24 h	30 μV/24 h	200 μV/24 h	$\approx$1 μ V/Monat, 2 μV/Jahr
Eingangsruhestrom, -offsetstrom	10^{-10} ... 10^{-11} A (Drift 1 pA/K)	10^{-14} A	10^{-16} A (10^{-17} A)	10^{-9} ... 10^{-12} A
Grenzfrequenz des Verstärkers (durch Modulator bedingt)	10^2 Hz	10^3 ... 10^6 Hz	10^1 ... 10^3 Hz	durch OV gegeben (z. B. 10^3 ... 10^5 Hz)
Anwendung zur Verstärkung von	kleinen Gleichspannungen ($\leqq$mV-Bereich)	kleinen Gleichströmen ($\leqq$nA ... pA-Bereich)	sehr kleinen Gleichströmen ($\leqq 10^{-10}$ ... 10^{-16} A)	kleinen Gleichspannungen ($\leqq$mV-Bereich; evtl. kleinen Gleichströmen)

trieben werden, wenn die Eingangsgleichspannung in die Mitte seiner Hysteresekurve gelegt wird und er durch positive bzw. negative Impulse umgeschaltet wird.

Unter *Funktionsgeneratoren* seien hier *Oszillatoren* verstanden, die einen bestimmten Verlauf der *Ausgangsspannung*, z.B. *Dreieck-, Sägezahn- oder Treppenfunktion*, erzeugen bzw. andere Funktionen, z.B. eine *Spannungs-Frequenz-Wandlung*, realisieren.

Sägezahnschaltungen sind bereits u. a. im Zusammenhang mit Bild 6.4.9 und 6.4.10 behandelt worden. Mit OV läßt sich ein Dreieckgenerator in einfacher Weise durch Reihenschaltung eines SCHMITT-Triggers und eines Integrators realisieren (Bild 7.5.20a)). Der SCHMITT-Trigger liefert in dem gerade festen Zustand eine konstante

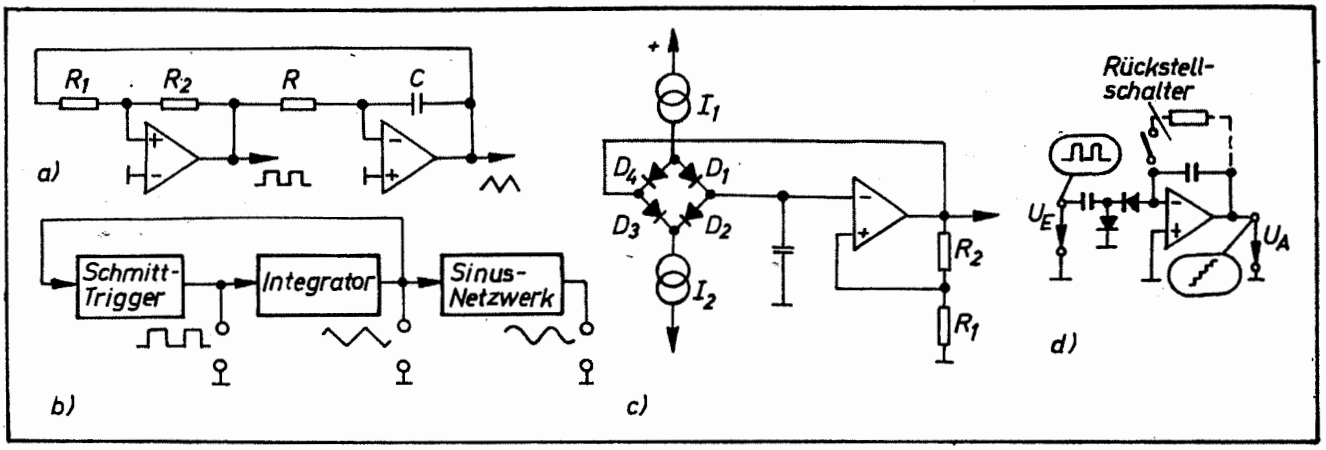


Bild 7.5.20. Beispiele von Funktionsgeneratoren

a) Die Reihenschaltung aus SCHMITT-Trigger und Integrator führt zum Rechteck-Dreieck-Generator

b) Mittels eines Sinusnetzwerkes gemäß Bild 7.5.13c) kann aus Teilbild a) ein Rechteck-Dreieck-Sinusgenerator erhalten werden

c) Sägezahngenerator, dessen Impulsverhältnis durch die Stromstärke der beiden Stromquellen eingestellt ist und für hohe Frequenzen geeignet ist

d) Wichtiger Teil eines Treppengenerators

Ausgangsspannung. Sie wird im Integrator integriert, bis am SCHMITT-Trigger die Schaltschwelle überschritten ist und er umkippt. Dann liegt am Integrator die entgegengesetzte Ausgangsspannung des SCHMITT-Triggers, und die integrierte Spannung läuft in die entgegengesetzte Richtung, bis erneut die Umschaltspannung des SCHMITT-Triggers erreicht ist, usw. Die Frequenz des *Dreieckgenerators* ist durch die Gleichung

$$f = \frac{1}{4RC} \cdot \frac{R_2}{R_1} \tag{14}$$

gegeben und ist voll symmetrisch.

Ein anderes Prinzip zeigt Bild 7.5.20c). Hier werden zwei Stromquellen I_1 und I_2 genutzt. Ihr Stromverhältnis bestimmt die Länge der Teilimpulse und damit die Unsymmetrie bzw. die Schrägen der beiden Sägezahnflanken. Die Umschaltung zwischen der ladenden Quelle I_1 und der entladenden I_2 erfolgt über die Diodenbrücke D_1 bis D_4.

Bild 7.5.20b) zeigt schließlich einen Funktionsgenerator, der auf einer Zusammenschaltung von Schaltungen aus Bild 7.5.20a) und 7.5.13c) beruht. Mit ihr können also *gleichzeitig rechteck-, dreieck- und sinusförmige Schwingungen* erzeugt werden.

Den wesentlichen Teil einer Schaltung, mit der eine Treppenfunktion erzeugt werden kann, zeigt Bild 7.5.20d). Die Eingangsrechteckimpulse werden über die Gleichrichter dem Integrator zugeführt und führen so zur Treppenfunktion. Sobald der Maximalwert erreicht ist, muß der Kondensator über den Rückstellschalter entladen werden. Prinzipien der *Spannung-Frequenz-Wandler* sind in den Schaltungen der Bilder 7.3.23 bis 7.3.25 beispielhaft behandelt.

7.5.8. Aktive Filter

In den Abschnitten 7.1.2. und 7.1.4. wurde gezeigt, wie durch Rückkopplung Frequenzselektion zu erreichen ist. Mittels der OV entwickelten sich auf diesem Prinzip die aktiven Filterschaltungen. Die entscheidenden Grundlagen hierzu brachten R. SALLEN und E. KEY 1955, also ca. 10 Jahre, bevor integrierte OV vorlagen. Inzwischen existieren so viele unterschiedliche Schaltprinzipien, daß ein Überblick kaum mehr möglich ist. Darüber hinaus haben die aktiven Filter durch eine Vielzahl andersartiger Filterprinzipien (vgl. 7.9.) nur noch bis ca. 100 kHz Bedeutung. Deshalb sollen *hier nur einige ausgewählte Varianten* vorgestellt werden.

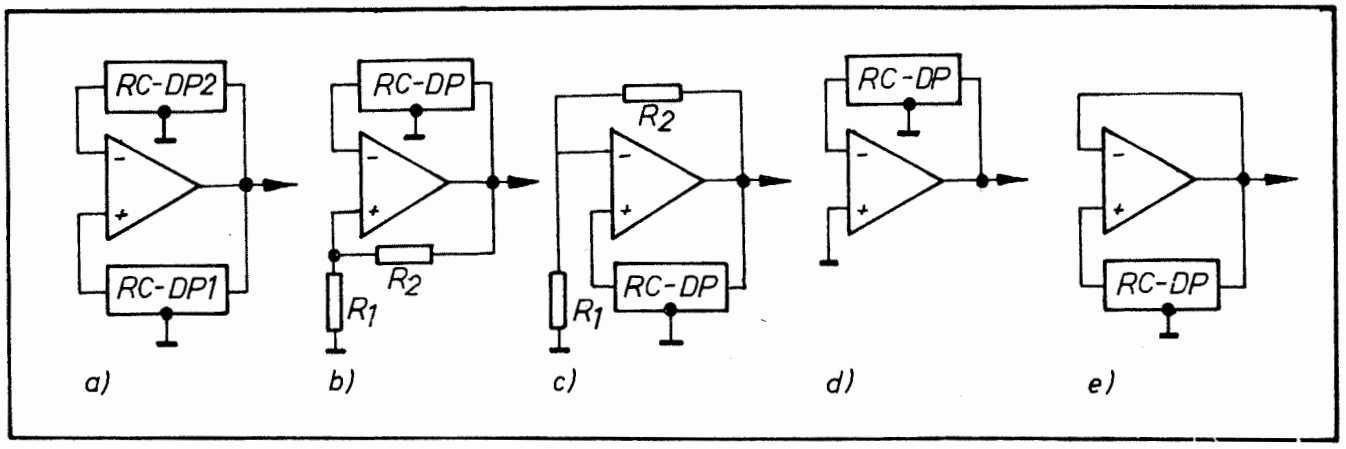


Bild 7.5.21. Wichtige Schaltungsvarianten eines OV für aktive Filter. (*RC*-DP bedeutet *RC*-Dreipol)

a) Allgemeiner Fall mit zwei *RC*-Dreipolen
b) Angefachte Inversionsschaltungen
c) Angefachte Einsverstärkerschaltungen
d) Inversionsschaltungen
e) Einsverstärkerschaltungen

In der Regel werden heute aktive Filterschaltungen so aufgebaut, daß mit je einem OV ein Filter 1. bzw. 2. Grades realisiert wird und diese Filter in Reihe geschaltet, d. h. kaskadiert, werden. Entsprechend Abschnitt 2.5. sind damit alle technisch realisierbaren Filter zu verwirklichen. In diesem Sinne werden hier nur aktive Filterschaltungen mit einem OV behandelt. Die entscheidenden Strukturen zeigt Bild 7.5.21, wobei zur Vereinfachung auf den Anschluß der Eingangsspannung zunächst verzichtet wurde. Die allgemeine Variante ergibt sich, wenn zu beiden Eingängen des OV *RC*-Dreipole (*RC*-DP) führen. Schaltungen dieser Art sind die oft komplizierten Brückenschaltungen, die hier nicht weiter behandelt werden. Sie sind u. a. in den generell weiterführenden Büchern [F 3, F 29, K 21 und T 4] enthalten. Wird ein Zweig von Teilbild a) durch einen ohmschen Spannungsteiler bzw. durch Kurzschluß ersetzt, so entstehen die anderen vier speziellen Schaltungen. Zur Einführung in die Problematik sei zunächst ein Beispiel für die Schaltung in Teilbild e) behandelt.

7.5.8.1. Einsverstärker

Die Schaltung eines Einsverstärkers als aktiven Tiefpaß zeigt Bild 7.5.22a). Durch die Wahl der Bauelemente R_1, R_2, C_1 und C_2 lassen sich die Eigenschaften des Schwingkreisspannungsteilers gemäß Abschnitt 2.3.2. realisieren. Formal kann dabei weiter auf den Reihenschwingkreis gemäß Abschnitt 2.2.3. zurückgegriffen werden. So ergeben sich Teilbild b) mit der Darstellung des PN-Plans (vgl. auch Bild 2.2.4 für den Reihenschwingkreis), Teilbild c) für den Amplituden- und d) für den Phasengang.

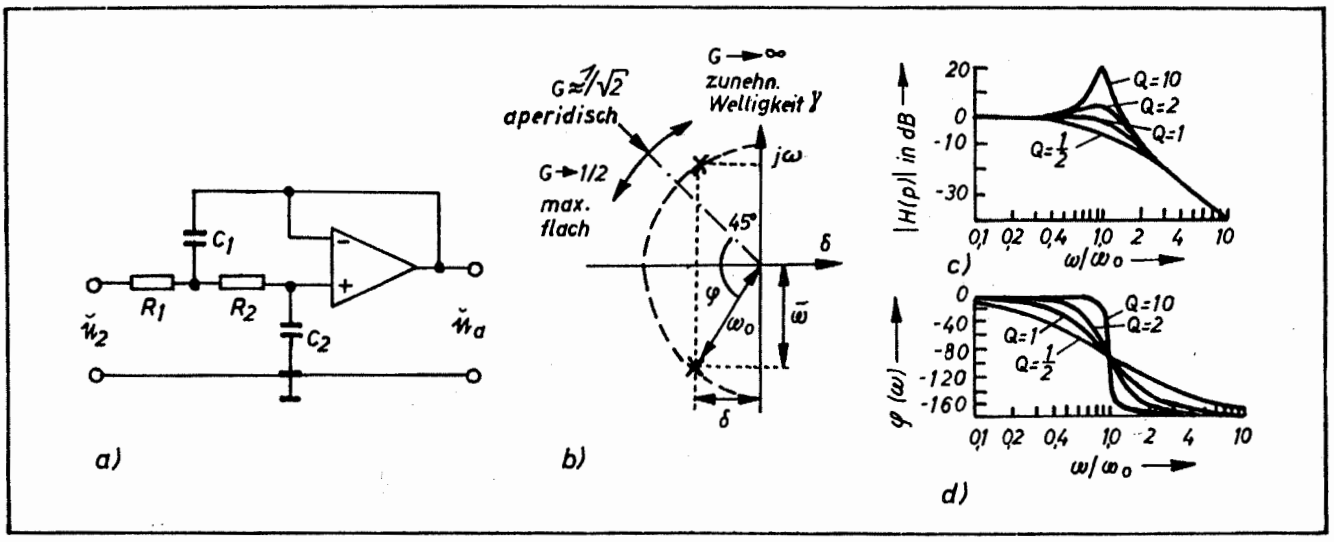


Bild 7.5.22. Aktiver Tiefpaß, aus dem Einsverstärker von Bild 7.5.21e) abgeleitet

a) Schaltung
b) Pol-Nullstellen-Plan
c) Amplitudengang für verschiedene Güten G
d) Phasengang für verschiedene Güten G

Dabei sind die *drei Fälle* zu unterscheiden:

$\frac{1}{2} \leqq G < 1/\sqrt{2}$: $\varphi < 45°$ überaperiodisch: max. flach: lineare Phase

$G = 1/\sqrt{2}$: $\varphi = 45°$ aperiodisch: Potenzfilter

$G < 1/\sqrt{2}$: $\varphi > 45°$ gedämpfte Schwingung: Welligkeit: TSCHEBYSCHEFF-Filter

Die Übertragungsfunktion der Schaltung kann mit der komplexen Frequenz $p = \mathrm{j}\omega$ — wenn der Quellenwiderstand von u_e Null ist — wie folgt geschrieben werden:

$$\frac{u_a}{u_e} = \frac{1}{p^2 R_1 R_2 C_1 C_2 + pC_2 \cdot (R_1 + R_2) + 1}. \tag{15a}$$

Für das Nennerpolynom ergeben sich die Lösungen als Pole

$$p_{1,2} = -\frac{R_1 + R_2}{2R_1R_2C_1} \pm \frac{\sqrt{C_2^2 \cdot (R_1 + R_2)^2 - 4R_1C_1R_2C_2}}{2R_1R_2C_1C_2}. \tag{15b}$$

Daraus folgen die Resonanzfrequenz

$$\omega_0 = \frac{1}{\sqrt{R_1C_1R_2C_2}}, \tag{16a}$$

die Güte

$$G = \frac{\sqrt{R_1R_2C_1C_2}}{C_2 \cdot (R_1 + R_2)} \tag{16b}$$

und die Dämpfung

$$\delta = -\frac{R_1 + R_2}{2R_1R_2C_1}. \tag{16c}$$

7.5.8.2. Angefachte Einsverstärkerschaltungen

Die *Verallgemeinerung* des vorangegangenen Beispiels kann *in zweierlei Richtungen* erfolgen, und zwar durch den Übergang zu

- Tief-, Hoch- und Bandpaß,
- Verstärkungen, die größer als Eins sind.

Die entsprechenden Fälle faßt Bild 7.5.23 zusammen. Im Vergleich zu Bild 7.5.21c) ist darin zur Vereinfachung der Gegenkopplungszweig aus R_1 und R_2 fortgelassen worden. Die sich durch ihn und ohne RC-Dreipol ergebende Leerlaufverstärkung wird mit V bezeichnet. *Für alle drei Schaltungen* läßt sich dann die *Übertragungsfunktion* wie folgt schreiben:

$$H(p) = \frac{z}{p^2 - 2\delta p + \omega_0^2}. \tag{17}$$

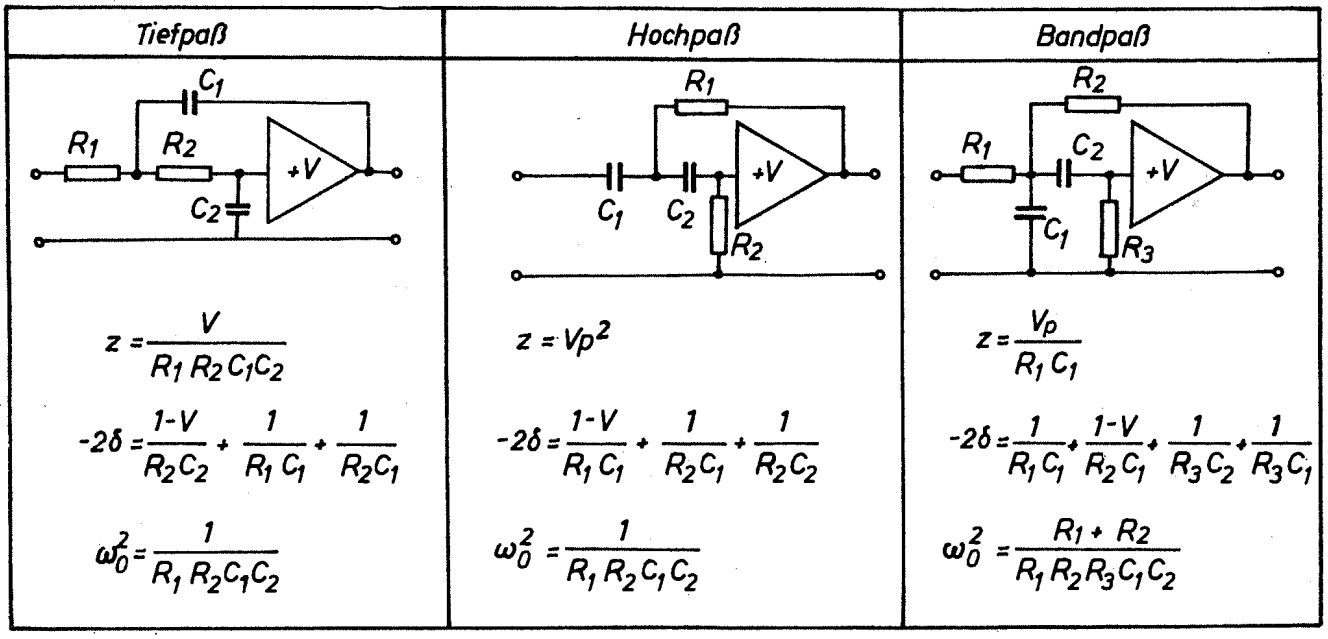


Bild 7.5.23. Zusammenstellung der Schaltungen und Hauptdaten der drei Grundtypen des angefachten Einsverstärkers. Im Vergleich zu Bild 7.5.21c) wurde der Zweig R_1, R_2 zur Vereinfachung weggelassen und die sich durch ihn ergebende Verstärkung V ohne RC-Dreipol in den OV hineingeschrieben

Die entsprechenden Kenngrößen sind in das Bild eingetragen. Der Tiefpaß hat also zusätzlich eine doppelte und der Bandpaß eine einfache Nullstelle im Ursprung. Die Güte der jeweiligen Schaltung berechnet sich gemäß

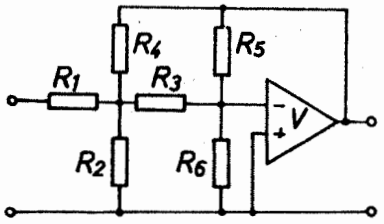
$$G = -\frac{\omega_0}{2\delta}. \tag{18}$$

7.5.8.3. Mit invertierendem Verstärker

Die allgemeine Grundschaltung des invertierenden Aktivfilters zeigt Tab. 7.5.9. Sie besitzt die *Übertragungsfunktion*

$$H(p) = \frac{-V}{1 + \boldsymbol{R}_1\left(\frac{1}{\boldsymbol{R}_2} + \frac{1}{\boldsymbol{R}_4}(1+V)\right) + \left(\boldsymbol{R}_1\left(1 + \frac{\boldsymbol{R}_3}{\boldsymbol{R}_2} + \frac{\boldsymbol{R}_3}{\boldsymbol{R}_4}\right) + \boldsymbol{R}_3\right)\cdot\left(\frac{1+V}{\boldsymbol{R}_5} + \frac{1}{\boldsymbol{R}_6}\right)}. \tag{19}$$

Tabelle 7.5.9. Schaltung und Berechnungsdaten im Sinne eines Ablaufschemas für aktive Filter in Inversionsschaltung. Die allgemeinen Impedanzen im Bild können durch Widerstände oder Kapazitäten realisiert werden. R_6 wird in den Beispielen unendlich gewählt.



TP	HP	BP
$R_1C_2R_3R_4C_5$	$C_1R_2C_3C_4R_5$	$R_1R_2C_3C_4R_5$
$C_5 = C$ $^1R_1 = 2\,\omega_9 CVG$ $C_2 = 8\,GC\,(V+1)$ $^1/R_3 = \omega_0 GC\,(V+1)$ $R_4 = VR_1$	$C_3 = C$ $C_1 = C$ $^1/R_2 = 2\,\omega_0 C\,(1 + {}^1/_2\,V)$ $C_4 = C/V$ $R_5 = \frac{(2V+1)\,G}{\omega_0 C}$	$C_3 = C$ $C_4 = C$ $^1/R_1 = V\omega C$ $^1/R_2 = (2G^2 - A)\omega CG$ $R_5 = \frac{2G}{\omega C}$

Dabei können die komplexen Widerstände z. T. durch ohmsche Widerstände, $\boldsymbol{R}_i = R_i$, und z. T. durch Kapazität, $\boldsymbol{R}_i = 1/pC_i$, ersetzt werden. Für drei Spezialfälle mit jeweils $\boldsymbol{R}_G \to \infty$ enthält Tab. 7.5.9 in einfacher Darstellung den Rechenweg. Es werden jeweils vorgegeben: V, G, ω_0 und eine bzw. zwei Kapazitäten. Dann können die dazugehörenden, weiteren berechnet werden. Aus der obigen Formel lassen sich bei Bedarf auch die allgemeinen Fälle entsprechend den vorangegangenen Beispielen bestimmen.

7.5.8.4. Allpässe und Breitbandphasenschieber

Im Zusammenhang mit Bild 2.3.11 (oberer Teil) ist ein *typischer Allpaß* beschrieben. Soll seine Ausgangsspannung auf Masse bezogen sein, so benötigt er jedoch eine Gegentakteingangsspannung, wie sie z. B. mit einer Schaltung gemäß Bild 7.3.2d)

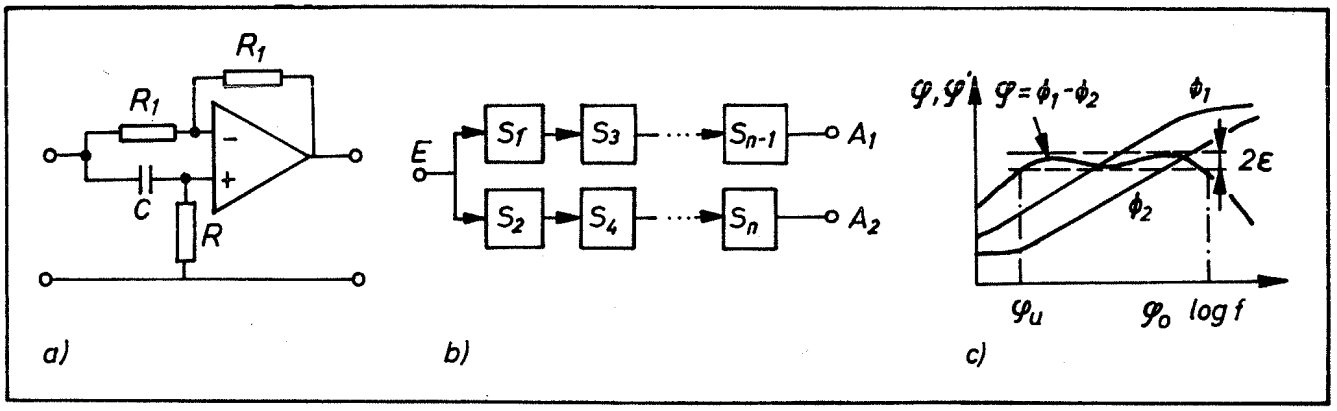


Bild 7.5.24. Allpaß und Breitbandphasenschieber

a) Allpaßschaltung eines Operationsverstärkers, die der Schaltung in Bild 2.3.11 entspricht
b) Zusammenschaltung von n Schaltungen gemäß a) mit gestaffelten Pol-Nullstellen-Frequenzen
c) Optimaler Frequenzgang einer Schaltung nach b) mit möglichst konstanter Phase zwischen den beiden Ausgängen

erzeugt werden kann. Der OV mit seinen beiden Eingängen ermöglicht auch hier eine effektive Lösung, wie sie als Beispiel Bild 7.5.24a) zeigt. Der *Phasengang* eines solchen Allpasses ist durch

$$\tan\varphi = \frac{2\Omega}{1-\Omega^2} \quad \text{mit} \quad \Omega = \omega RC \tag{20}$$

gegeben. Dieses Verhalten ermöglicht in einem relativ großen Frequenzbereich einen nahezu logarithmisch von der Frequenz abhängenden Phasengang. Werden mehrere solcher Allpässe in Reihe geschaltet, so kann mit mehreren Stufen der Frequenzbereich f_u bis f_o erweitert werden. Werden nun *zwei solcher Ketten* verwendet. so läßt sich gemäß Bild 7.5.24c) erreichen, daß *zwischen den beiden Ausgängen* A_1 und A_2 eine im Bereich f_u bis f_o bis auf den Fehler ε *konstante Winkeldifferenz* existiert. Die Approximation für ein solches System ermöglicht folgendes Formelsystem: Mit einer Hilfsgröße K bestehen zwischen den Grenzfrequenzen f_u und f_o, dem Fehler ε im Winkelmaß und der Winkeldifferenz φ bei n (gerade) Stufen die Zusammenhänge

$$K = \sqrt[n]{4 f_o/f_u}\,, \tag{21a}$$

$$\varepsilon = \frac{2\sin\varphi}{\cosh\left(\dfrac{\pi}{2\ln k}\right)}. \tag{21b}$$

Die Verhältnisse der einzelnen Pol-Nullstellen-Frequenzen der RC-Glieder berechnen sich dann zu

$$F_i = -\left(K^a + k_i + k_{N-i}\right), \tag{22a}$$

wobei für gerades i gilt

$$a = 2 - 2\varphi/\pi \quad \text{und} \quad k_i = \frac{K^{4-4\varphi/\pi} - 1}{K^{2i} - K^{2-2\varphi/\pi}} \tag{22b}$$

und für ungerades i

$$a = 2\varphi/\pi \quad \text{und} \quad k_i = \frac{K^{4\varphi/\pi} - 1}{K^{2i} - K^{2\varphi/\pi}}. \tag{22c}$$

Für die unterste Pol-Nullstellen-Frequenz gilt dann

$$f_1 = -\tfrac{1}{2} f_u (K^{1-\varphi/\pi} - K^{\varphi/\pi - 1}). \tag{23}$$

Alle weiteren Frequenzen können über die F_i berechnet werden.

7.6. Stabilisierte Stromversorgungen

7.6.1. Grundprinzipien

Alle elektronischen Einrichtungen und Geräte bedürfen einer Stromversorgung. Sie kann auf die *unterschiedlichste Weise* realisiert werden, z. B. aus

- dem allgemeinen Stromnetz
- Batterien und Akkumulatoren.

Im ersten Fall ist eine Gleichrichtung notwendig, wie sie im Abschnitt 5.4.4. behandelt wurde. Der zweite Fall wurde im Abschnitt 1.1.7. besprochen. In beiden Fällen besitzt die dabei entstehende Spannungsquelle *Mängel*, bedingt vor allem durch

- Schwankungen infolge von Rauschen, Restbrumm und Alterung
- den endlichen Innenwiderstand.

Für viele Anwendungen sind diese Mängel so störend, daß eine *Stabilisationsschaltung* notwendig ist. In anderen Fällen werden konstante Ströme benötigt. Dies führt zu speziellen Stabilisationsschaltungen, die im Abschnitt 7.6.5. behandelt werden.

Das Grundprinzip der Spannungsstabilisation zeigt Bild 7.6.1a). Sie ist zwischen Spannungsquelle und Verbraucher geschaltet und besteht aus

- Regelstrecke
- Regeleinrichtung.

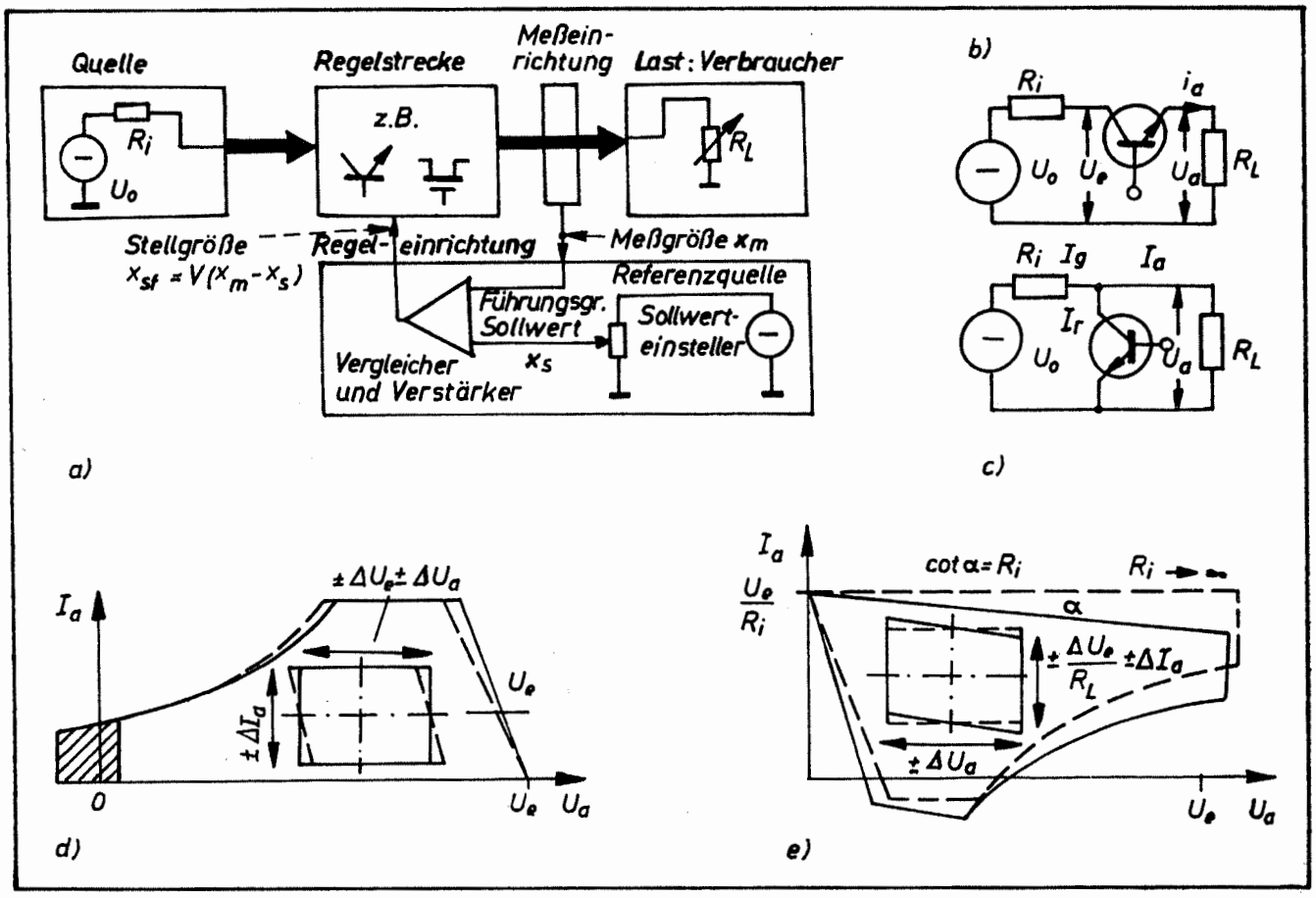


Bild 7.6.1. Grundprinzipien der Spannungsstabilisation

a) Prinzipielle Schaltung mit regelungstechnischen Blöcken
b) Stelltransistor bei Längsregelung und
c) bei Querregelung
d) Arbeitsfeld beim Längstransistor und
e) beim Quertransistor

Die gestrichelten Linien in d) und e) berücksichtigen den Innenwiderstand der Quelle. Bei e) ist darin auch der Vorwiderstand enthalten

Die *Regeleinrichtung* existiert in vielerlei Schaltungen. Sie wird daher in den folgenden Abschnitten behandelt.

An der *Regelstrecke* wird stets zusätzliche Energie verbraucht. Es kann hier also, ähnlich wie bei den Endstufen (7.4.), ein Wirkungsgrad betrachtet werden. Dementsprechend sind auch wieder *kontinuierlich wirkende* Regelstrecken und Regelstrecken mit *Schalterbetrieb* zu unterscheiden. Die Schaltregler werden erst im Abschnitt 7.6.6. behandelt. Für die kontinuierlich wirkende Regelstrecke sind *zwei Typen* möglich:

- Längsregler gemäß Teilbild b).
- Querregler gemäß Teilbild c).

Am *Längsregler* fällt immer eine bestimmte Spannung ab. Es gilt

$$U_a = U_e - U_r . \tag{1}$$

Der *Querregler* zweigt dagegen einen zusätzlichen Strom ab. Es gilt

$$I_a = I_g - I_r . \tag{2}$$

Beide Schaltungen haben Vor- und Nachteile. Es seien hier die jeweiligen *Vorteile* genannt.

Beim *Querregler* sind dies:

- Die Quelle wird immer mit dem konstanten Strom i_g belastet.
- Der Regeltransistor wird bei maximaler Ausgangslast minimal belastet. Er kann also nicht durch Kurzschluß überlastet werden.
- R_i der Quelle kann, ja, muß sogar relativ groß sein.

Beim *Längsregler* ist dagegen vorteilhaft:

- Bei kleiner Last wird auch wenig Energie verbraucht.
- Der Lastwiderstand kann vor zu hoher Belastung (Strom) leicht geschützt werden.

In den meisten Fällen überwiegen die Vorteile des *Längsreglers*. Er wird daher *am häufigsten angewendet*.

Die Leistungsverhältnisse am Regeltransistor lassen sich leicht gemäß den Betrachtungen zu Bild 7.4.1a) erfassen. Es sind dazu lediglich die Bereiche des *sicheren Arbeitsbereichs* so umzuordnen, daß sie auf Ausgangsstrom und -spannung bezogen sind. Dann entstehen die Bilder 7.6.1d) bzw. e). In die Fläche ist für beide Reglerfälle ein Rechteck bzw. ein Parallelogramm eingezeichnet, die mit Variation der Ausgangsspannung und des Ausgangsstromes entstehen.

Beim *Längsregler* ist zu erkennen, daß die *Gefahr* seiner Überlastung gerade für *kleine Ausgangsspannung* und *große Ausgangsströme* besteht. Je größer die Ausgangsspannung ist, desto mehr Strom kann bezüglich des Regeltransistors entnommen werden.

Beim *Querregler* besteht die *Gefahr* bei *großer Ausgangsspannung* und *kleinem Ausgangsstrom*. Dieser Fall ist aber ohnehin bei der Auslegung der gesamten Schaltung unmittelbar zu berücksichtigen. Mit R_i dreht sich die Arbeitsfläche um den Winkel α, und außerdem verlagert sich der Punkt U_e/R_i auf der Stromachse. Hierbei besteht die weitere Gefahr, daß bei relativ kleinem R_i keine kleinen Ausgangsspannungen einstellbar sind und so eine Überlastung der zu versorgenden Schaltung auftritt. Aus diesen und anderen Gründen wird für Querregler das R_i der Quelle oft bewußt vergrößert.

7.6.1.1. Kenngrößen

Für den Stabilisator gibt es drei Kenngrößen. Sie werden durch drei Störgrößen bestimmt:

Schwankt die *Eingangsspannung* U_e (oder die innere Spannung U_0), so ergibt sich als Folge immer eine wenn auch kleine Schwankung der Ausgangsspannung. Bezüglich der Relationen dieser Schwankungen gibt es unterschiedliche Definitionen. Hier sollen die (kleingeschriebenen) Schwankungen direkt verwendet werden. Dann ist der *Stabilisationsfaktor* als

$$\alpha = u_e/u_a \tag{3}$$

definiert. Er ist also reziprok zu der sonst bezüglich Ein- und Ausgangsspannung üblichen Definition. Er ist nämlich um so größer, je besser die Stabilität ist, je weniger sich also die Eingangsschwankungen am Ausgang bemerkbar machen. Ist V die Verstärkung der Meßgrößenänderung zur Stellgröße und D der Durchgriff des Regeltransistors beim Längstransistor (bei bipolaren $\mathrm{d}U_{CB}/\mathrm{d}U_{EB} \approx h_{12B}$), so gilt

$$\alpha \approx 1 + V/D\,. \tag{4}$$

Schwankt die *Last*, so ist der *Innenwiderstand* des Stabilisators bedeutsam. Er setzt sich aus zwei Anteilen zusammen:

$$R_{is} = R_{ie} + R_i/\alpha\,. \tag{5}$$

Der *elektronische Anteil* R_{ie} entsteht, wenn die Eingangsspannung konstant geschaltet wird. Die durch Laständerungen bedingten *Eingangsspannungen* werden ebenfalls um den Stabilisationsfaktor α verringert. Folglich erscheint der Innenwiderstand der Quelle um diesen Fehler kleiner. Ist S die Steilheit des Regeltransistors beim Längstransistor (bei bipolaren $\mathrm{d}I_C/\mathrm{d}U_{BE} \approx h_{21} = h_{21E}/h_{11E}$), so gilt

$$R_{ie} \approx 1/SV\,. \tag{6}$$

Die dritte Störung kann vom *Sollwert* kommen. Sie rührt u. a. von *Temperatureinflüssen*, *Rauschspannungen* und *Alterungen* her. Sie gehen statisch, d. h. für tiefe Frequenzen, voll als Änderung in die Ausgangsspannung ein. Dabei addieren sich sogar noch die Drift- und Rauschgrößen des Verstärkers. Bei *hohen Frequenzen* treten jedoch Abweichungen auf. Hier sinkt nämlich der Verstärkungsgrad des Verstärkers. Dies bedeutet, daß bezüglich des Frequenzganges der totale Frequenzgang des Verstärkers und des Stelltransistors $V_T(f)$ bis zur Ausgangsspannung wirkt. Da es sich hier um ein stark rückgekoppeltes System handelt, bei dem auch *Frequenzüberhöhungen* möglich sind, können sich bestimmte Störfrequenzen des Sollwertes und des Verstärkerrauschens erheblich in der Ausgangsspannung auswirken. Für die *Übertragungsgröße der Störgrößen* gilt folglich

$$R(f) = V_T(f)/V_{T0}\,, \tag{7}$$

wobei V_{T0} das Produkt aus V und der Spannungsverstärkung des Regeltransistors bei sehr niedrigen Frequenzen ist.

So, wie $R(f)$ frequenzabhängig ist, gilt dies natürlich auch für R_i und R_{ie} und damit auch für R_{is}. Da insgesamt ein stark gegengekoppeltes System vorliegt, sind Betrachtungen zur *Stabilität* analog zu denen im Abschnitt 7.1.2.3. und 5.3. notwendig. Vorteilhaft ist darüber hinaus ein *Kondensator am Ausgang* des Reglers, also parallel zur Last. Er darf jedoch nicht zu groß gewählt werden, weil dann die *Ladeströme* den Längstransistor zerstören können.

Für den Stabilisator mit Quertransistor ist der Stabilisationsfaktor auf u_0 zu beziehen, dann folgt

$$\alpha \approx \frac{R_i + R_{iT}}{R_{iT}}, \tag{4a}$$

wobei R_{iT} der Innenwiderstand des Transistors ist. Dementsprechend liegt für den Innenwiderstand des Stabilisators auch die Parallelschaltung von R_i und R_{ie} vor:

$$R_{is} = \frac{R_{ie} R_i}{R_{ie} + R_i}. \tag{5a}$$

Bezüglich $R(f)$ ergeben sich keine *Änderungen*.

7.6.1.2. Störwertaufschaltungen

Die Kenngrößen wiesen aus, daß generell bei der bisherigen Schaltung immer ein gewisser, wenn auch *meist sehr kleiner Rest der Störung* in der Ausgangsspannung wirksam wird. Das hier verwendete Prinzip der Regelung beruht ja gerade darauf, daß eine Regelabweichung $x_m - x_S$ existieren muß, damit überhaupt eine Stellgröße x_{st} (Bild 7.6.1a)) auftreten kann. Ein *hiervon abweichendes* Prinzip ist die *Störwertaufschaltung*. Sie wird aus der Eingangsspannung bzw. dem Ausgangsstrom abgeleitet und im Sinne einer Mitkopplung additiv dem Sollwert zugeschaltet. Dadurch gelingt es, soweit sich nicht Verstärkung, Steilheit und Durchgriff (also der Arbeitspunkt) ändern, diese Störungen zu kompensieren. Im Sinne der Rückkopplung (7.1.1.) wird so der Durchgriff und der Innenwiderstand Null sowie die Steilheit unendlich gemacht. Schaltungen hierzu werden im Abschnitt 7.6.2. behandelt. Sie ermöglichen *etwa eine zehnfache Verbesserung* von Kenndaten.

Die Störwertaufschaltung wurde früher einmal auch als Steuerung bezeichnet. Heute wird Steuerung auf beabsichtigte Einflüsse (keine Störungen) begrenzt. Ein Spannungsstabilisator kann also z. B. über eine bewußte Verstellung der Sollwerteinstellung realisiert werden. Dann wird auch von einer programmierbaren Spannungsquelle gesprochen.

7.6.2. Spannungsregler

Einige Schaltungen zur Erzeugung konstanter Spannungen wurden bereits in den Bildern 7.3.4, 7.3.5a) und 7.3.18 vorgestellt. Aus den Ergebnissen der letzten Abschnitte kann hier eine systematische Analyse der Schaltungen vorgenommen werden. Sie läßt sich für den Längstransistor auf der Basis von Bild 7.6.2 gewinnen.

Die *einfachste Schaltung* zeigt Teilbild a). Sie enthält nur den Regeltransistor, eine Z-Diode als Spannungsnormal und einen Widerstand, der den nötigen Basisstrom und Strom für die Z-Diode aufbringt. Die Kenndaten werden erheblich verbessert,

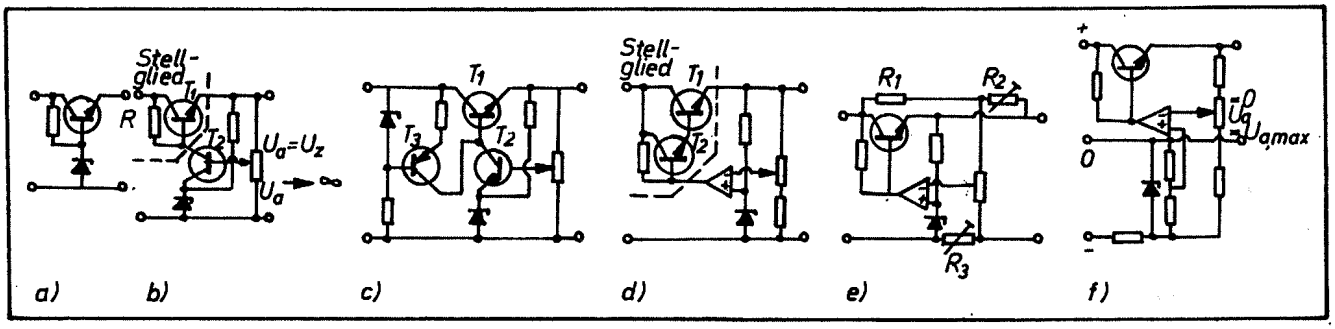


Bild 7.6.2. Grundschaltungen der Stabilisation mit Längstransistor als Stellglied

a) ohne Verstärker, b) mit T_2 als Verstärker,
c) mit zusätzlichem dynamischen Arbeitswiderstand mittels T_3,
d) mit DARLINGTON-Längstransistor T_1, T_2 und Differenz- bzw. Operationsverstärker,
e) mit Störwertaufschaltungen für $u_e(R_1, R_2)$ und $i_a(R_3)$,
f) für die Möglichkeit, die Ausgangsspannung Null zu erreichen

wenn ein *Verstärkertransistor* eingefügt wird (b). Gleichzeitig besteht dann auch eine bequeme Möglichkeit zur *Einstellung der Ausgangsspannung* über das Potentiometer. Da jetzt der Widerstand R_B sowohl der Arbeitswiderstand für T_2 ist als auch den Basisstrom für T_1 aufzubringen hat, lassen sich größere Verstärkungsfaktoren mit einem dynamischen Arbeitswiderstand durch den Transistor T_3 erreichen. Generell kann die Verstärkerstufe günstig durch einen *Differenz- bzw. Operationsverstärker* (7.3.3. bzw. 7.5.) ersetzt werden. So entsteht der rechte Teil von Bild c). Beim Stelltransistor ist dabei zugleich — was auch unabhängig in allen anderen Teilbildern erfolgen kann — eine DARLINGTON-*Variante* gewählt. Dies ist dann sinnvoll, wenn so große Ströme geregelt werden müssen, daß der Steuerstrom nicht direkt vom Verstärker aufgebracht werden kann.

Die Teilschaltung e) zeigt die beiden Störwertaufschaltungen. Mittels R_1, R_2 wird der Einfluß der Eingangsspannung reduziert. Deshalb liegt dieses Verhältnis bei etwa α. Mit R_3 wird der Innenwiderstand auf etwa Null eingestellt. Deshalb muß sein Wert bei etwa R_{is} liegen.

Bei allen Schaltungen mit Spannungsteiler für u_a-Einstellung ist darauf zu achten, daß auch das wirksame V um diesen Teilungsfaktor verkleinert wird.

Die Schaltung in Teilbild f) zeigt schließlich eine Möglichkeit, wie die *Ausgangsspannung bis auf Null* herab einstellbar ist. Gegenüber dem Nullpotential wird eine zusätzliche negative Spannung erzeugt und mittels der Z-Diode stabilisiert. Ein Teil dieser Spannung (in der Regel etwa $^1/_2$ bis $^1/_3$) dient als Vergleichsspannung, und die gesamte Spannung dient als Fußpunkt für den Spannungseinsteller.

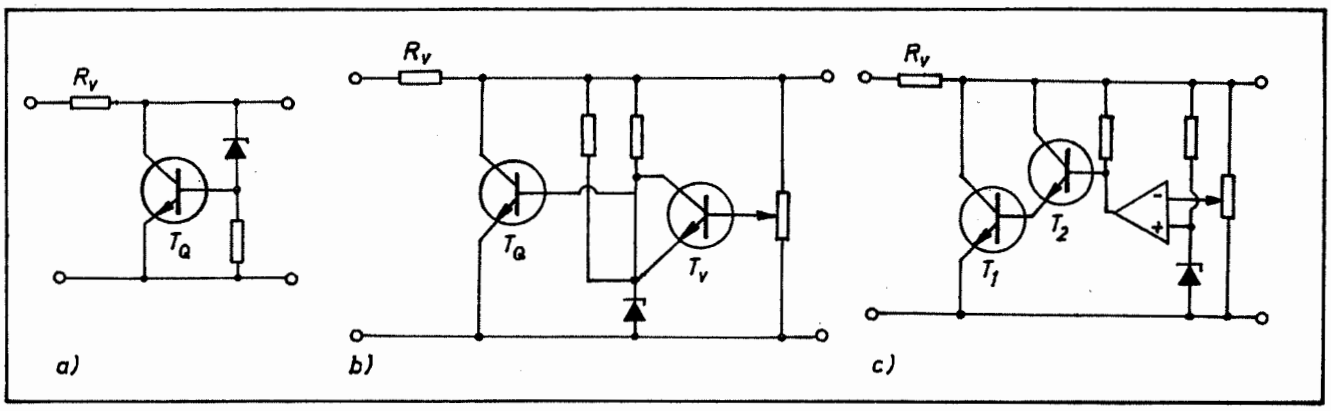


Bild 7.6.3. Grundschaltungen der Stabilisation mit Quertransistor
a) ohne, b) mit Verstärker T_V,
c) mit DARLINGTON-Quertransistor T_1, T_2 und Differenz- bzw. Operationsverstärker

In Bild 7.6.3 sind drei *Beispielschaltungen* für den *Querregler* dargestellt. Teilbild a) benutzt nur den Quertransistor T_Q. Das Normal für die Ausgangsspannung ist die Z-Diode. Der Vorwiderstand R_V dient zur Vergrößerung des Innenwiderstandes R_i der Eingangsquelle. Beim Teilbild b) findet zusätzlich ein Verstärker mittels Transistor T_V Anwendung. In Teilbild c) wird der Quertransistor durch ein DARLINGTON-Paar T_1, T_2 gebildet, und es findet darüber hinaus ein Differenz- bzw. Operationsverstärker Anwendung.

Insbesondere bei OV, aber auch bei vielen anderen Anwendungen (z.B. in der Digitaltechnik bei Speichern) werden positive und negative Spannungen benötigt, die in einem festen Verhältnis zueinander stehen. Besonders wichtig ist die *erdsymmetrische Spannung*. Sie kann zunächst gemäß Bild 7.6.4a) auf einfache Weise erzeugt werden. Mittels einer Schaltung gemäß Teilbild b) können daraus eine stabilisierte positive und negative Spannung abgeleitet werden. Mit P_1 ist die Größe der positiven Spannung im üblichen Sinne einstellbar. Mittels P_2 kann aus der positiven

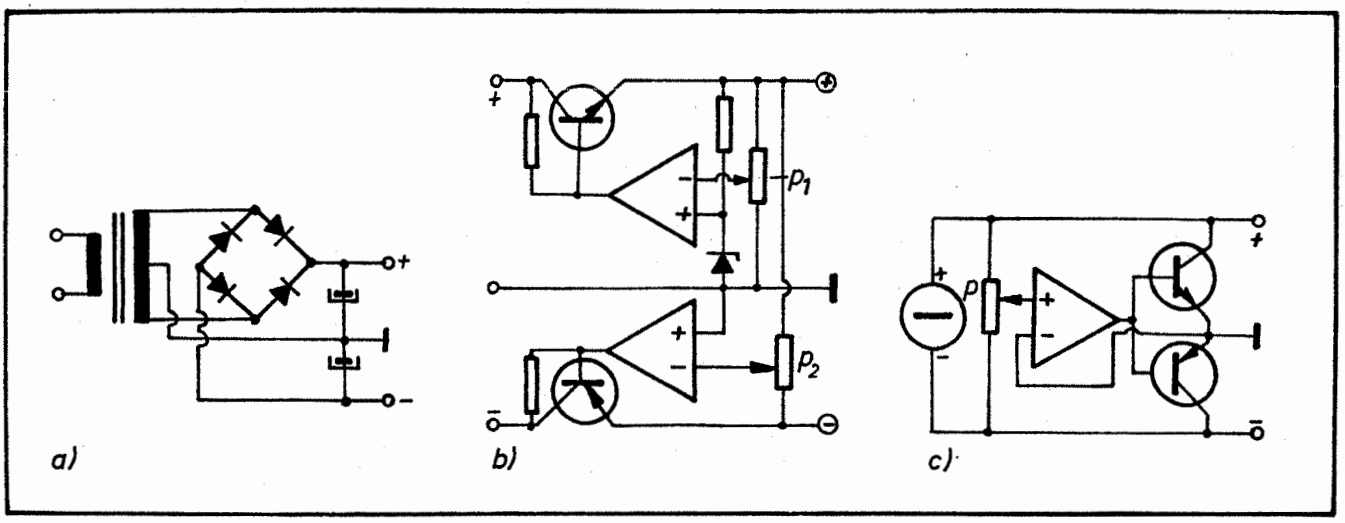


Bild 7.6.4. Schaltungen zur Erzeugung von positiven und negativen Spannungen mit festem bzw. einstellbarem Verhältnis

a) Gleichrichterschaltung zur Erzeugung von auf Masse bezogener positiver und negativer Spannung
b) Schaltung zur Doppelstabilisierung von positiver und negativer Spannung. P_1 stellt die Größe der positiven Spannung ein; P_2 das Verhältnis von positiver und negativer Spannung
c) Schaltung zur Bildung einer virtuellen und über P einstellbaren Masse bei einer erdfreien Eingangsspannung

Spannung die Größe der negativen abgeleitet werden. Wie sich auch die positive Spannung ändert, die negative Spannung wird sich im durch P_2 eingestellten Verhältnis (bei Mitteneinstellung 1:1) mit ändern. Von einer erdfreien Spannung kann mittels der Schaltung von Bild 7.6.4c) künstlich eine durch die Einstellung von P gegebene Mitte der Ausgangsspannung erreicht werden. Die beiden Leistungstransistoren übernehmen den Stromausgleich bei unterschiedlichen Lastströmen der beiden auf Masse bezogenen Ausgangsspannungen.

7.6.3. Schutzschaltungen

Insbesondere bei einstellbaren Spannungsstabilisatoren im *Labor* besteht die *Gefahr von Kurzschlüssen*. Sie würden zur Zerstörung des Stelltransistors führen. Feinsicherungen sind zum Schutz des Transistors zu träge. Bei doppeltem Nennstrom benötigen sie eine Schmelzzeit von ca. 30 Minuten, und noch bei 4fachem Strom sind 10 bis 300 ms notwendig, während die Zeitkonstante der Transistoren bei 5 ... 50 ms liegt. Deshalb wurden vielfältige Modifikationen von elektronischen Sicherungen entwickelt. Einige einfache Beispiele zeigt Bild 7.6.5. Übersteigt im Teilbild a) der Spannungsabfall am Widerstand R_1 die Schleusenspannung der Diode also um etwa 0,7 V, so wird die Diode leitend und beeinflußt den Transistor in Richtung zu einem größeren Spannungsabfall. Es entsteht so die gepunktete *Strombegrenzung* in Bild 7.6.5c). Dieses Prinzip läßt sich noch mit einem Transistor T_3 gemäß Teilbild b) verbessern. Der Widerstand R_2 dient zur Strombegrenzung des Basisstromes von T_3 und besitzt einen Wert zwischen 100 und 1000 Ohm. Mit dem Widetstand R_3 zusammen kann zusätzlich der Einfluß der Ausgangsspannung auf das Begrenzungsverhalten einbezogen werden. Dadurch ist es möglich, sich dem sicheren Arbeitsbereich von Bild 7.6.1d) weitgehend zu nähern. Im Normalbetrieb ist die Basis von T_3 negativ vorgespannt. Der Spannungsabfall an R_2 beträgt etwa $U_A R_2/R_3$. Leitend wird der Transistor T_3, sobald

$$U_{R2} \geqq U_{BES} \frac{R_2}{R_3} \tag{8}$$

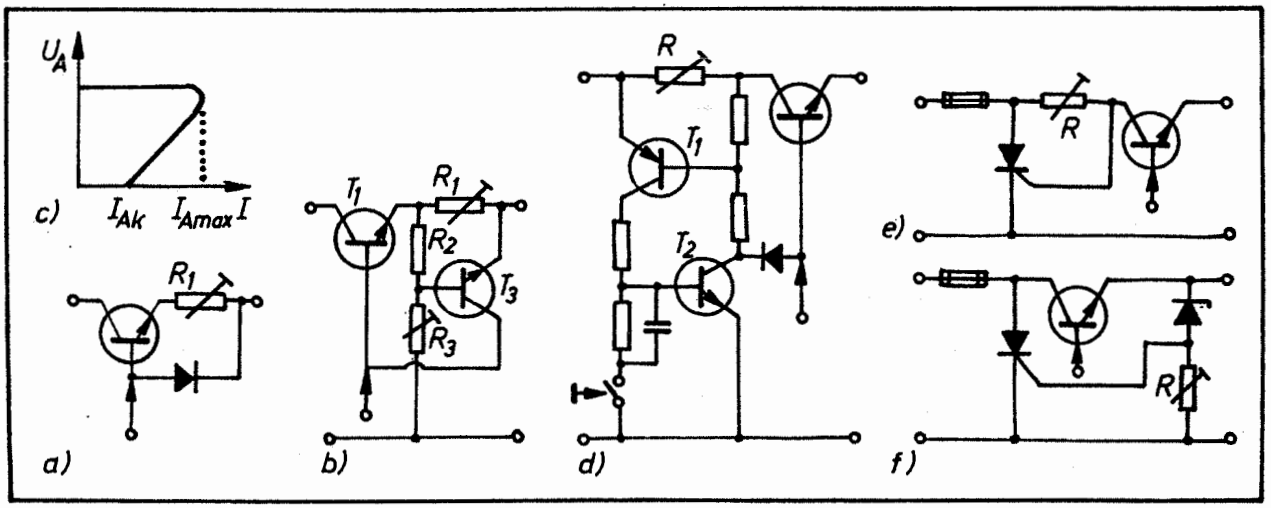


Bild 7.6.5. Schutzschaltungen

a) mit einer Diode,
b) mit Transistor T_3 und einstellbarem Einfluß von maximalem Strom (R_1) und Leistung am Transistor (R_3)
c) Kennlinien der Strombegrenzung: ausgezogen: Fold-back-Kennlinie für b); gepunktet: für a)
d) mit Flip-Flop am Eingang,
e) mit Thyristor bezüglich Strom durch R,
f) mit Thyristor bez. maximal zulässiger Ausgangsspannung

gilt, wobei U_{BES} die *Schleusenspannung* der Basis-Emitter-Strecke von T_3 ist. Insgesamt folgt so für den maximalen Strom gemäß Teilbild c)

$$I_{A,\,max} \approx \frac{U_{BES}}{R_1} + \frac{U_A R_2}{R_1 R_3}. \tag{9}$$

Bei Kurzschluß fällt an R_2 keine Spannung mehr ab, deshalb gilt für den Kurzschlußstrom

$$I_{AK} \approx \frac{U_{BES}}{R_1}. \tag{10}$$

Nachteilig bei diesen Schaltungen ist der Widerstand R_1. Er läßt sich mit entsprechend komplizierten Schaltungen vermeiden. Es ist auch möglich, diese elektronischen Sicherungen in die Siebkette vor den Stabilisator zu schalten, dann tritt R_1 nicht störend in Erscheinung. Auch gibt es Schaltungen, die nach einem Kurzschluß bzw. einer Überlastung nicht automatisch wieder in den normalen Betriebszustand zurückkehren, sondern erst wieder eingeschaltet werden müssen. Ein solches Beispiel zeigt Bild d). Die beiden komplementären Transistoren T_1 und T_2 bilden einen bistabilen Multivibrator im Sinne von Bild 6.4.2b). Sobald der Strom eine mittels R einstellbare Schwelle überschreitet, schaltet er, und der Regeltransistor ist gesperrt. Mittels der Taste kann wieder der Betriebszustand erreicht werden.

Ein anderes Prinzip von Schutzschaltungen wird vor allem bei fest eingebauten Stabilisatoren verwendet. Es arbeitet *mit Thyristoren*. In Teilbild e) zündet der Thyristor, sobald der Strom an R einen zu großen Spannungsabfall hervorruft. Dadurch brennt momentan die Sicherung durch.

Bei anderen Schaltungen, z. B. TTL-Schaltkreisen (8.2.4.) ist eine durch Störung bedingte *Überspannung gefährlich*. Die gesamte Baugruppe kann mit einem Überspannungsschutz gemäß Teilbild f) vor der Zerstörung geschützt werden. Sobald die Z-Spannung überschritten wird, entsteht ein Spannungsabfall an R. Übersteigt er einen Grenzwert, so zündet der Thyristor, und die Sicherung brennt durch.

7.6.4. Integrierte Schaltkreise

International existiert eine Vielzahl von Schaltkreisen für die Erzeugung stabilisierter Spannungen. Hier werden zwei typische Beispiele ausgewählt.

7.6.4.1. Spannungsregler MAA 723

Der integrierte Schaltkreis MAA 723 trägt auch die Bezeichnungen TRA 72400 in der UVR, IL 72723 in Jugoslawien, μA 723 bei Fairchild und L 123 bzw. TRV 1723 bei anderen Firmen. Er enthält, wie Bild 7.6.6 zeigt, *fünf Teilgruppen*:

- Stromspiegel zur komplexen Versorgung der weiteren Teilschaltungen,
- Erzeugung der temperaturstabilisierten Referenzspannung,
- Regeldifferenzverstärker,
- Stellglied,
- Hilfsmittel zum einfachen Aufbau von Schutzschaltungen.

Eine für die gesamte Funktion aller Baugruppen wichtige Einheit ist der mehrfache *Stromspiegel* aus den Transistoren T_2, T_3, T_7, T_8. Er wird versorgt von der durch die Z-Diode D_1 stabilisierten Spannung und dem Widerstand 15 kΩ. Dabei ist vorteilhaft, daß die Z-Diode und die Basis-Emitter-Strecke eines Transistors etwa je einen Temperaturkoeffizienten von 2,4 mVK^{-1}, aber mit entgegengesetztem Vorzeichen besitzen. Eine weitere Stabilisierung bewirkt die Versorgung durch die Konstantstromquelle mittels des FET T_1. Durch den FET können außerdem große integrierte Widerstände vermieden werden.

Die eigentliche *Referenzspannung* geht von der Z-Diode D_2 aus. Ihre typische Durchbruchspannung ist 6,2 V und der Vorstrom, mit dem sie in der Schaltung betrieben wird, 100 μA. Wegen der großen Bedeutung dieser Spannung für die Stabilität der vom Baustein erzeugten Spannung wurde hier ein hoher Aufwand mittels des Regelverstärkers T_4, T_5 und T_6 betrieben. T_4 und T_5 werden am besten als Stelltransistor in DARLINGTON-Schaltung aufgefaßt. Dann stellt T_6 den Regelverstärker für sie dar, der zudem noch durch die Konstantstromquelle T_3 einen hohen dynamischen Arbeitswiderstand, also große Verstärkung besitzt. Der integrierte MOS-Kondensator besorgt ein stabiles Verhalten dieses Verstärkers durch rechtzeitiges Senken des Frequenzganges um 20 dB/Dekade. Auch hier wirken wieder der Temperaturkoeffizient von D_2 und der Basis-Emitter-Strecke von T_6 entgegengesetzt, so daß eine hohe Temperaturstabilität erreicht wird. Ferner wird durch den Regelverstärker erreicht, daß die Referenzspannung mit einem sehr kleinen Innenwiderstand von etwa 100 Ω bereitsteht.

Der *Differenzverstärker* besteht aus den Transistoren T_{11} und T_{12}, die über T_{13} mit konstantem Strom gespeist werden. Das dazu notwendige Basispotential wird über den konstanten Strom aus T_7 abgeleitet und von T_{10} passend umgewandelt. T_9 stellt eine Rückführung hierzu dar und unterstützt außerdem die Wirkung des Differenzverstärkers wirksam. Die Verstärkung wird durch den großen dynamischen Widerstand der Stromquelle T_8 über 100fach erhöht. An dem Anschluß 9 kann u. a. eine Frequenzkompensation zur Erhöhung der Stabilität erfolgen. Dazu kann entweder ein Kondensator von etwa 100 pF zum Eingang von T_{12} geschaltet werden. Es kann aber auch eine Kapazität von ca. 1000 pF gegen Masse benutzt werden.

Das *Stellglied* besteht aus dem Leistungstransistor T_{15}, der von dem Transistor T_{14} meist in DARLINGTON-Schaltung angesteuert wird. T_{14} kann aber auch als Emitterfolger des Differenzverstärkers aufgefaßt werden.

Zum Aufbau der *Schutzschaltung* steht der Transistor T_{11} und beim DIL-Gehäuse die Z-Diode D_3 zur Verfügung.

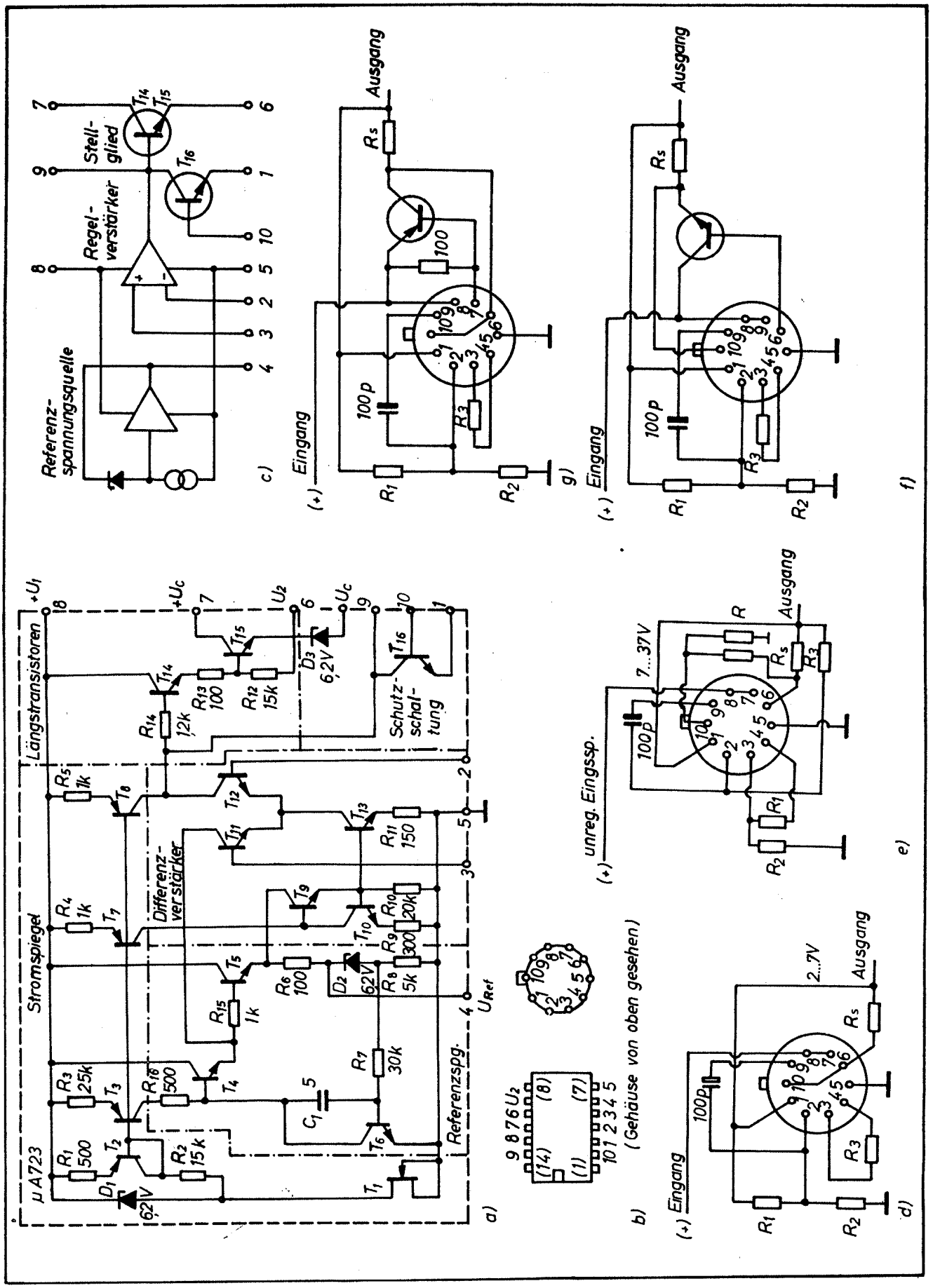


Bild 7.6.6. Zum MAA 723

a) Innenschaltung mit Teilfunktionen b) Sockelschaltungen c) vereinfachte Innenschaltung
d) Beschaltung für Ausgangsspannungen von 2 bis 7 V und eine senkrechte Strombegrenzung e) Beschaltung für 7 bis 37 V und mit Fold-back-Kennlinie der Strombegrenzung, f) für 2 bis 7 V und höhere Stromentnahme durch einen *npn*-Leistungstransistor g) Dito mit einem *pnp*-Leistungstransistor

Tabelle 7.6.1. Daten der integrierten Regelschaltung MAA 723

Eingangsspannungsbereich	9,5 ... 40 V
Stromaufnahme bei $U_e = 30$ V	2,3 ... 5 mA
Referenzspannung	6,8 ... 7,5 V
Grenzen des Spannungsabfalls am Stelltransistor	3 ... 40 V
Maximaler Laststrom	0,15 A
Maximale Verlustleistung	0,8 W
Stabilisierungsfaktor	$2 \cdot 10^2$
Temperaturkoeffizint der Referenzspannung	$2 \cdot 10^{-5}\,\mathrm{K}^{-1}$

Die wesentlichen Daten der Schaltung enthält Tab. 7.6.1. Einige Anwendungsbeispiele zeigen die Schaltungen 7.6.6d) bis g). Die Schaltung d) gestattet, über den Spannungsteiler aus R_1 und R_2 stabilisierte Spannungen im Bereich von etwa 2 bis 7 Volt zu erhalten. Dabei gilt

$$U_a = U_{ref} \frac{R_2}{R_1 + R_2}. \tag{11}$$

Die Schaltung e) ist für Spannungen von 7 bis 37 V geeignet. Für die Ausgangsspannung gilt hier

$$U_a = U_{ref} \frac{R_1 + R_2}{R_2}. \tag{12}$$

In beiden Fällen ist eine Stromentnahme bis ca. 150 mA möglich, sofern dabei nicht die Leistung von 0,8 W am Schaltkreis überschritten wird. In beiden Schaltungen sind Schutzschaltungen vorgesehen. In der Schaltung d) ist die scharfe Begrenzung gemäß dem gepunkteten Verlauf in Bild 7.6.5c) zur Anwendung gekommen. Für den maximalen Strom gilt etwa

$$I_{max} \approx 0{,}7\ \mathrm{V}/R_S\,. \tag{13}$$

In der Schaltung e) findet dagegen das Prinzip von Bild 7.6.5b) mit der ausgezogenen Kennlinie in Bild 7.6.5c) Anwendung. Für den Kurzschlußstrom I_{AK} und den maximalen Strom gelten jetzt

$$I_{AK} \approx 0{,}7\ \mathrm{V}/R_S\,, \tag{14a}$$

$$I_{max} = \frac{0{,}7\,\mathrm{V}}{R_S} + \frac{U_a R_5}{R_S R_6}. \tag{14b}$$

Der Widerstand R_3 soll den Einfluß des Offsetstromes und seines Temperaturkoeffizienten vom Differenzverstärker zur Referenzquelle reduzieren. Schließlich ist in f) und g) noch gezeigt, wie mittels außen angeschalteter Leistungstransistoren größere Ströme geregelt werden können. Die Schaltung mit dem zusätzlichen *pnp*-Transistor besitzt schlechtere Regeleigenschaften bezüglich Eingangsspannungsschwankungen und einen größeren Temperaturkoeffizienten bezüglich der Lastabhängigkeit. Sonst sind beide Schaltungen mit *pnp*- und *npn*-Transistor gleichwertig.

7.6.4.2. Festspannungsregler MA 78XX

Die Daten der integrierten Festspannungsregler MA 78XX enthält Tab. 7.6.2, die Innenschaltung Bild 7.6.7a) und ihr vereinfachtes Funktionsprinzip Teilbild b). Die verschiedenen *Ausgangsspannungen werden durch Maskenänderung* bezüglich des

Tabelle 7.6.2. Technische Daten der Festspannungsregler von TESLA

		MA 7805	MA 7812	MA 7815	MA 7824
Ausgangsspannung	V	4,8 ...5,2	11,5 ... 12,5	14,4 ... 15,6	23 ... 24
Max. Eingangsspannung	V	35	35	35	40
Ausgangswiderstand	mΩ	17	18	19	28
Ausgangsspitzenstrom	A	2,2	2,2	2,1	2,1
Ausgangskurzschlußstrom	A	0,75	0,35	0,23	0,15
Temperaturkoeffizient von U_a	mV/K	−1,1	−1,0	−1,0	−1,5

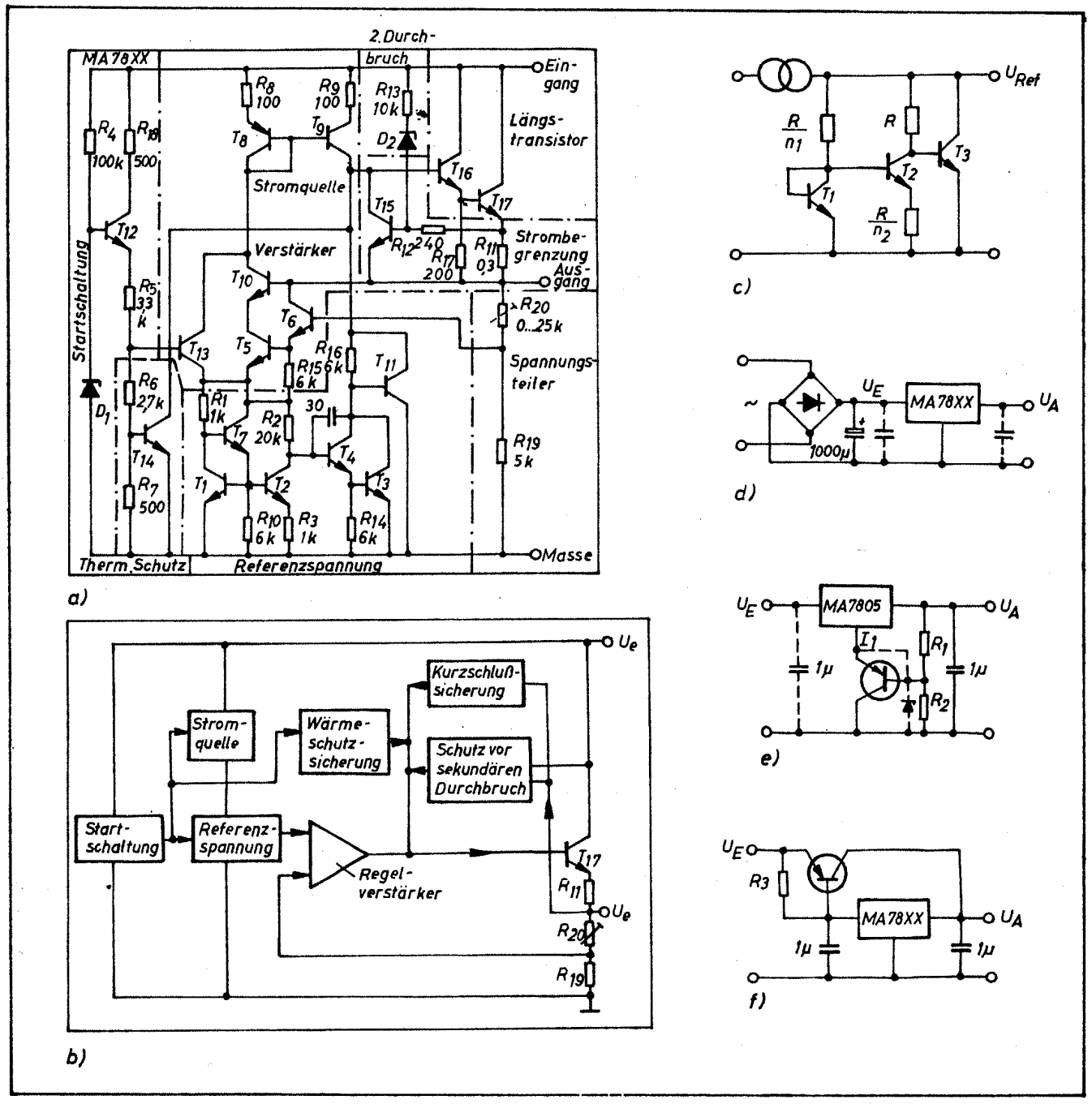


Bild 7.6.7. Zum MA 78XX

a) Innenschaltung (T_9 und T_{11} sind *pnp*-Transistoren, T_{13} ist *npn*-Transistor)
b) Prinzipschaltung
c) Bandabstandsreferenz
d) Betrieb der Schaltung
e) Erhöhung der Ausgangsspannung durch äußere Beschaltung
f) Erhöhung des maximalen Laststromes durch einen Leistungstransistor

Spannungsteilerwiderstandes R_{20} realisiert. Der Kern dieser Schaltung ist das *Spannungsnormal.* Es leitet sich aus der *Bandabstandsreferenz* (Band-gap-reference) gemäß Bild 7.6.7c) ab. Deshalb sei zunächst dessen Funktion beschrieben. Der als Diode betriebene Transistor T_1 führt zu einem Kollektorstrom

$$I_{C1} = \frac{U_{Ref} - U_{BE1}}{R/n_1}. \tag{15}$$

Für den Transistor T_2 gilt wegen seiner Stromsättigung, d. h. sehr kleinen Spannungsabfalls zwischen Kollektor und Basis,

$$I_{C2} \approx \frac{U_{Ref} - U_{BE2}}{R}. \tag{16}$$

Die beiden Ströme besitzen also das Verhältnis n_1. Damit kann der Spannungsabfall am Emitterwiderstand von T_2 berechnet werden:

$$U_{E2} = U_{BE1} - U_{BE2} = \frac{kT}{e_0} \ln n_1 . \tag{17}$$

Der Temperaturkoeffizient dieser Spannung ist positiv und beträgt

$$\frac{dU_{E2}}{dT} = \frac{1}{T} U_{E2} . \tag{18}$$

Dieser Temperaturkoeffizient liegt bei $n_1 = 10$ z. B. bei $+0{,}2$ mV/K. Er kann auf den zur Kompensation notwendigen Wert mit einem hinreichend großen Kollektorwiderstand von T_2 verstärkt werden. Dies bewirkt eine richtige Wahl von n_2, in dem Beispiel also 20. Damit beträgt die am Ausgang ohne Temperaturkoeffizient zur Verfügung stehende Spannung

$$U_{Ref} = U_{BE3} + n_2 U_{E2} \approx 1{,}2 \text{ V} . \tag{19}$$

Sie entspricht bei exaktem Abgleich genau dem Wert des Bandabstandes von Silicium mit 1,205 V. Daher der Name dieser und ähnlicher, daraus abgeleiteter Schaltungen [R 4, T 4].

In der Innenschaltung von MA 78XX sind in die Referenzspannungsquelle weitere Transistoren eingefügt. So entsteht eine *Referenzspannung von* $4 \times 1{,}2$ V.

Der *Differenzverstärker* besteht aus einem Teil der Referenzquelle und den Transistoren T_5, T_6, T_{10}, T_{13}, wobei T_8 als dynamischer Arbeitswiderstand und als Stromspiegel zur Übertragung an die DARLINGTON-Regeltransistorstufe (T_{16}, T_{17}) wirkt. Mit dem Widerstand R_{11} und dem Transistor T_{15} ist eine *Strombegrenzung* eingebaut. Der Widerstand R_{13} und die Z-Diode D_2 verhindern einen zweiten Durchbruch. Der Transistor T_{14} ist in gutem Wärmekontakt mit dem Leistungstransistor T_{17}. Seine Basis wird meist mit einer konstanten Spannung aus der *Startschaltung* versorgt. Sobald sich der Chip zu stark erhitzt, wirkt T_{14} ähnlich leistungsbegrenzend, wie in den Abschnitten 7.4.5.2. und 7.4.5.3. beschrieben. Durch diese *dreifache Schutzschaltung* ist der Spannungsregler *sehr robust.* Darüber hinaus ist sein Betrieb, wie Bild 7.6.7d) zeigt, sehr einfach. Es sind kaum externe Bauelemente erforderlich. Besonders wenn der Regler weit vom Ladeelko entfernt ist, sollte unmittelbar an seinen Eingang ein Kondensator von 1 bis 2 µF zur Vermeidung von Eigenschwingungen angeschaltet werden. Der Kondensator am Ausgang ist nicht notwendig, aber bezüglich Lastschwankungen vorteilhaft.

Prinzipiell ist es möglich, mit einem Festspannungsregler auch *höhere Spannungen* zu stabilisieren. Die dazu notwendigen Änderungen zeigt Bild 7.6.7e). Es können dabei u. a. drei Fälle unterschieden werden:

Nur mit R_1 und R_2 (ohne Transistor und Z-Diode) beträgt die neue Ausgangsspannung

$$U_A^+ \approx U_A + I_1 R_2 \left(1 + \frac{R_2}{R_1}\right), \tag{20}$$

wobei I_1 für $U_A = 5$ V etwa 30 mA beträgt. Die Widerstände sollten niederohmig, um ca. 100 Ω, sein.

Mit Transistor und ohne Z-Diode gilt dagegen

$$U_A^+ \approx \left(U_A + U_{BE} + \frac{I_1 R_2}{h_{21}}\right) \cdot \left(1 + \frac{R_2}{R_1}\right). \tag{21}$$

Wird die Z-Diode verwendet (ohne R_2 und Transistor), so ergibt sich

$$U_A^+ \approx U_A + U_Z \,. \tag{22}$$

Der *maximale Strom* kann auch hier durch einen äußeren *Leistungstransistor* gemäß Bild 7.6.7f) erhöht werden. Insgesamt sollten bei diesen Reglertypen jedoch nicht 5 A überschritten werden.

7.6.5. Stromregler

Konstantstromquellen sind schon in verschiedenen Abschnitten, u. a. in 7.3.1.2. und 7.3.5., behandelt worden. Darüber hinaus sei noch auf die Stromspiegel im Abschnitt 7.3.4.2. verwiesen. Hier sind spezielle Schaltungen zu behandeln, die sich aus der stabilisierten Spannungsquelle ableiten und insbesondere integrierte Bausteine verwenden.

Die wichtigsten Grundprinzipien faßt Bild 7.6.8 zusammen. Aus Teilbild a) ist erkennbar, wie die Wirkung der Last an R einen Spannungsabfall bewirkt, der mit der

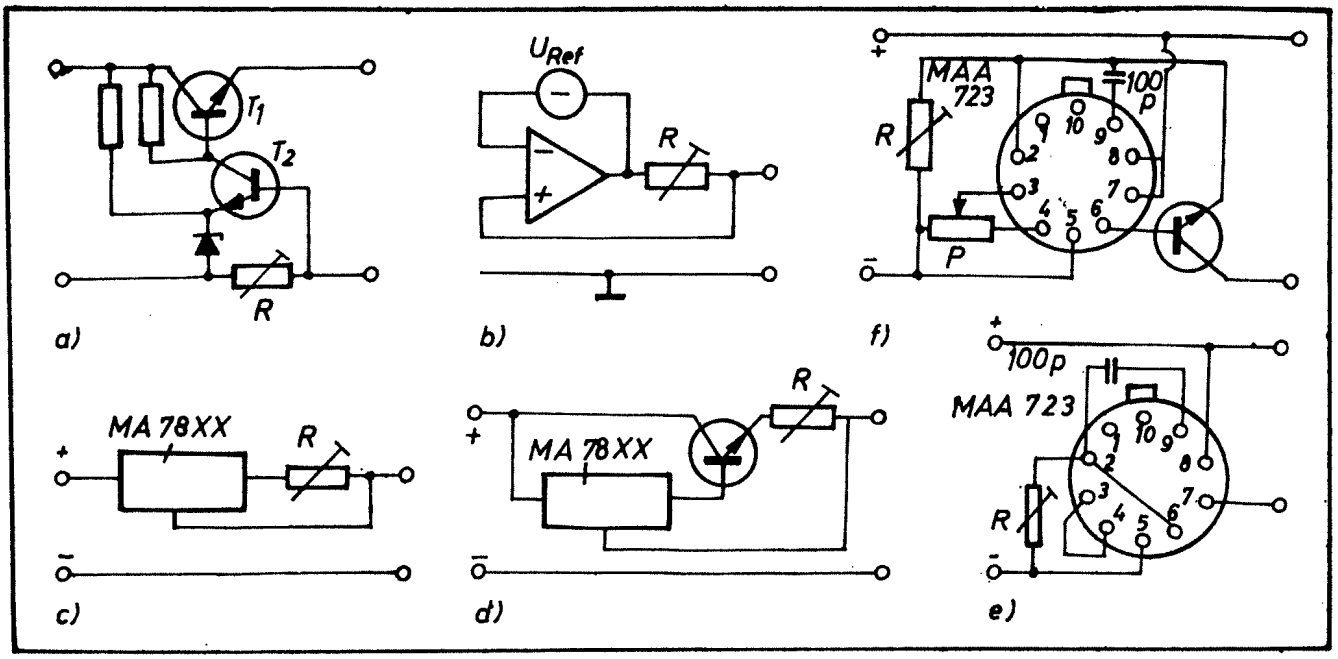


Bild 7.6.8. Beispiele von Stromreglern für veränderliche Last

a) Grundschaltung, aus der stabilisierten Spannungsquelle abgeleitet
b) Erklärung des Grundprinzips am Operationsverstärker
c) Anwendung des Schaltkreises MA 78XX und
d) mit zusätzlichem Leistungstransistor
e) Anwendung des Schaltkreises MAA 723 und
f) mit zusätzlichem Lasttransistor und über P einstellbarer Bezugsspannung (7 bis 37 V)

Referenzspannungsquelle (hier eine Z-Diode) verglichen wird, und dann über einen Verstärker T_2 der Längstransistor T_1 richtig gesteuert wird. Der sich dabei einstellende Strom beträgt

$$I_{\text{konstant}} = U_{\text{Ref}}/R \,. \tag{23}$$

Dieses Prinzip ist für alle weiteren Schaltungen verbindlich. Es ändern sich nur die Bausteine, die Spannungsquelle und Details der Schaltung.

In Teilbild b) ist dieses Prinzip bezüglich eines OV noch deutlicher zu erkennen, dafür ist aber der Zusammenhang mit der Stabilisation von Spannungen kaum mehr sichtbar. Dies wird jedoch wieder deutlicher bei der entsprechenden Anwendung der Schaltkreise MA 78XX. Die Bezugsspannung für Gl. (23) ist hierbei die vom Typ festgelegte Ausgangsspannung (Tab. 7.6.2). Genügt der maximale Strom von Schaltung c) nicht, so kann gemäß Teilbild d) ein zusätzlicher Leistungstransistor eingefügt werden.

Mit dem Schaltkreis MAA 723 bietet sich die Schaltung von Teilbild e) an. Die Bezugsspannung ist hier ca. 7 V. Mit der Schaltung f) können durch den Leistungstransistor wieder größere Ströme erreicht werden. Außerdem ermöglicht das Potentiometer P, größere Spannungswerte gemäß der Schaltung von Bild 7.6.6 einzustellen.

7.6.6. Geschaltete Stromversorgungen

7.6.6.1. Allgemeine Grundlagen

In den letzten Jahren sind mit der Weiterentwicklung der elektronischen Bauelemente, insbesondere durch die schnellen Leistungstransistoren sowohl der bipolaren als auch der MOSFET wesentliche Änderungen in der Stromversorgung eingeleitet worden. Hierbei war auch die ständige Miniaturisierung aller anderen Elektronikgruppen von wichtigem Einfluß. So entstanden einige neue Schaltprinzipien, die in Bild 7.6.9 zu systematisieren versucht wurden.

Teilbild a) zeigt die bisherige und in den vorangegangenen Abschnitten behandelte *Grundschaltung der Stabilisierung*. Sie besitzt den schon bekannten Nachteil, daß in der Stabilisierung viel Energie verbraucht wird. Der Verlust läßt sich durch einen geschalteten Regler (*Schaltregler*) reduzieren, der periodisch ein- und ausgeschaltet wird, wobei das Verhältnis von Ein- und Ausschaltzeit im Sinne einer PDM (vgl. 7.2.2., 7.3.8.4., 7.4.6.2.) genau die richtige Energiemenge weitergibt. Da ein Schaltregler auf diese Weise Impulse abgibt, wird er vor den Glättungsmitteln angeordnet. Sie beseitigen dann Brumm- und Impulsstörungen gleichermaßen. Dadurch wirkt der Schalttransistor zugleich auf eine Induktivität, die Energie speichern kann. So besteht hier prinzipiell die Möglichkeit, diesen Abschnitt der Schaltung auch wie Sperr- bzw. Durchlaßwandler (Teilbild f) bis h)) auszulegen. Weiter besteht die Möglichkeit, wie z. B. in Teilbild e), hinter dem Transistor einen Transformator für hohe Schaltfrequenzen (*Schaltnetzteil*) einzufügen, welcher dann die Geräte vom Netz trennen kann. Damit ist es möglich, wie in Teilbild c) den eigentlichen Netztrafo einzusparen und eine erhebliche Reduzierung von Volumen und Gewicht zu erreichen. Dies sei durch eine Näherungsformel für Leistungstransformatoren belegt. Sie gilt für eine magnetische Flußdichte von 1,2 T und eine Stromdichte von 3,5 A/mm². Bezüglich des Kernquerschnitts des magnetischen Materials A_{Fe}, der Leistung P und der Frequenz f gilt

$$A_{\text{Fe}} \approx 10\sqrt{P/f} \,. \tag{24}$$

Mit der Frequenz vieler Schaltnetzteile um 30 kHz sinkt also der Querschnitt auf etwa $^1/_{25}$. Mit MOSFET sind sogar 300 kHz erreichbar, dann liegt der Wert bei $^1/_{80}$. Bei ca. 30 kHz werden insgesamt Leistungsdichten von ca. 60 W/dm³ gegenüber 13 bei konventionellen Netzteilen erreicht. Darüber hinaus hat das Schaltnetzteil u. a. noch den Vorteil, daß es einen sehr großen Bereich der Eingangswechselspannung von z. B. 90 bis 300 V problemlos verarbeiten kann. Es wird eben nur das Pulsverhältnis über die Regelung umgesteuert. Allerdings ist für die Regelung am Ausgang

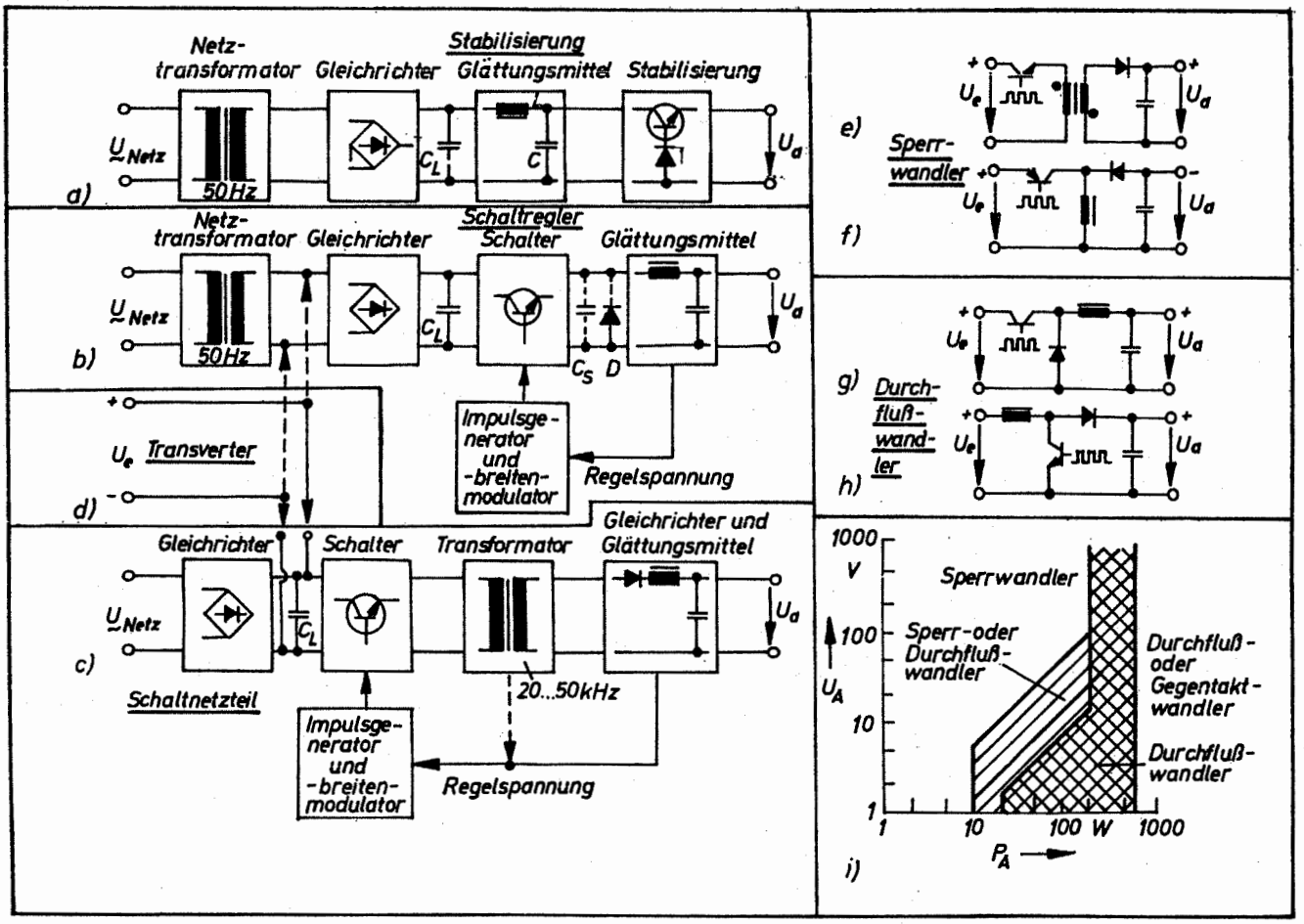


Bild 7.6.9. Prinzipien schaltender Netzteile

a) Klassische Stabilisierung zum Vergleich
b) Schaltregler mit PDM-gesteuertem Regeltransistor (Schalter)
c) Schaltnetzteil mit Potentialtrennung durch Ferrittransformator bei 20 bis 50 kHz
d) Transverter (Gleichspannungswandler), der entweder ein Teil des Schaltnetzteiles (c) bzw. e)) oder des Schaltreglers (b) bzw. f) bis g)) ist
e) und f) Zwei Varianten des Sperrwandlers mit positiver bzw. negativer Ausgangsspannung (die Punkte am Trafo bezeichnen den Wickelanfang)
g) und h) Zwei Varianten des Durchflußreglers
i) Optimale Einsatzgebiete der verschiedenen Varianten von Schaltreglern bzw. Schaltnetzteilen

ein Optokoppler zur Potentialtrennung notwendig. Eine andere Möglichkeit besteht darin, das Regelsignal aus dem Transformator abzuleiten.

Schließlich besteht noch die Möglichkeit, *Transverter* (Gleichspannungswandler) aufzubauen (Teilbild d)). Sie haben die Aufgabe, eine vorhandene Gleichspannung mit möglichst hohem Wirkungsgrad in eine andere umzuwandeln. Hierbei können die beiden Varianten des Bildes 7.6.9d) genutzt werden. Mit Trafo wird die Teilschaltung des Schaltnetzteils — und ohne Trafo die des Schaltreglers (gemäß f) bis h)) genutzt. Sie enthalten aber immer noch Induktivitäten. Es gibt aber auch Schaltungen ohne Induktivitäten. Sie werden im Abschnitt 7.6.6.6. behandelt.

7.6.6.2. Prinzip des Durchflußwandlers

Die *vier Grundprinzipien* für die Wandlung von Gleichspannungen mittels Schalter zeigt Bild 7.6.9e) bis h). Das Teilbild g) läßt sich am einfachsten aus der Schaltung der üblichen Stabilisierung von Teilbild a) ableiten. Der Längstransistor aus Teilbild a) wird zunächst als Schalter vor die Siebdrossel geschaltet. Wenn dann am Anfang der Siebkette die *Kapazität* C_S liegt, wird durch sie einfach das arithmetische Mittel der Einschaltzeit T_e zur Gesamtdauer der Periode T_g für die Ausgangs-

spannung verantwortlich:

$$U_a = U_e \frac{T_e}{T_g}. \tag{25}$$

Entfällt jedoch diese Kapazität, so ist beim Strom I_L durch die Drossel während der Einschaltzeit in der *Drossel die Energie*

$$W = I_L^2 \cdot L \tag{26}$$

gespeichert. Diese Energie wird bei Abschaltung des Transistors als großer Spannungsimpuls frei (vgl. Bild 5.7.13b)). Dies kann zur Zerstörung des Transistors führen, wenn die *zulässige Sperrspannung* überschritten wird. Deshalb ist die *Diode D* anstelle des Kondensators C_S notwendig. So entsteht die Schaltung von Teilbild g), und zugleich liefert auf diese Weise die Drossel in der Sperrphase die zuvor gespeicherte Energie nutzbar zum Ausgang. Deshalb wird die Funktion der Drossel in diesem Fall auch als *Speicherdrossel* bezeichnet. Da die Diode den Transistor in der Sperrphase vor Überspannung schützt, heißt sie auch Freilaufdiode. Die Wirkung der Ströme in den beiden Phasen — Ein und Aus — zeigt Bild 7.6.10 (links). Da der Hauptstrom zum Ausgangskreis in der Ein-Phase fließt, heißt diese Schaltung *Durchflußwandler*. Der maximale Strom in der Ein-Phase I_{Lm} ist natürlich größer als der gemittelte Ausgangsstrom I_a. Damit der Transistor nicht überdimensioniert werden muß, sollte als Faktor möglichst 1,2 nicht überschritten werden. Dies bedeutet eine minimale Induktivität L. Mit dem Lastwiderstand R_a gilt

$$L = \frac{R_a \cdot (1 - U_a/U_e)}{2f \cdot (I_{Lm}/I_a - 1)}, \tag{27}$$

wobei für die Spannungen wieder Gl. (25) erfüllt ist. Für die Ausgangswelligkeit folgt

$$\Delta U_a = \frac{U_a}{8LCf^2}\left(1 - \frac{U_a}{U_e}\right). \tag{28}$$

Da im Transistor nur kleine Verluste auftreten, ist der zeitlich gemittelte Strom durch den Transistor kleiner als der Ausgangsstrom. Mit guter Näherung gilt für ihn

$$\bar{I}_e = I_a U_a / U_e \,. \tag{29}$$

7.6.6.3. Prinzip des Sperrwandlers

Die Funktion der *Drossel als Energiespeicher* ermöglicht den Sperrwandlerbetrieb. In der Ein-Phase wird die Energie in der Drossel gespeichert und in der Aus-Phase zum Ausgang übertragen. Bild 7.6.10 (rechts) zeigt dieses Schema, einschließlich der fließenden Ströme. Schaltungsmäßig ist nur die Anordnung von Diode und Drossel vertauscht, d. h., der Stromfluß im Eingangs- und Ausgangskreis sind vertauscht. Daneben ist weiter auffällig, daß auch die Polarität beider Spannungen vertauscht

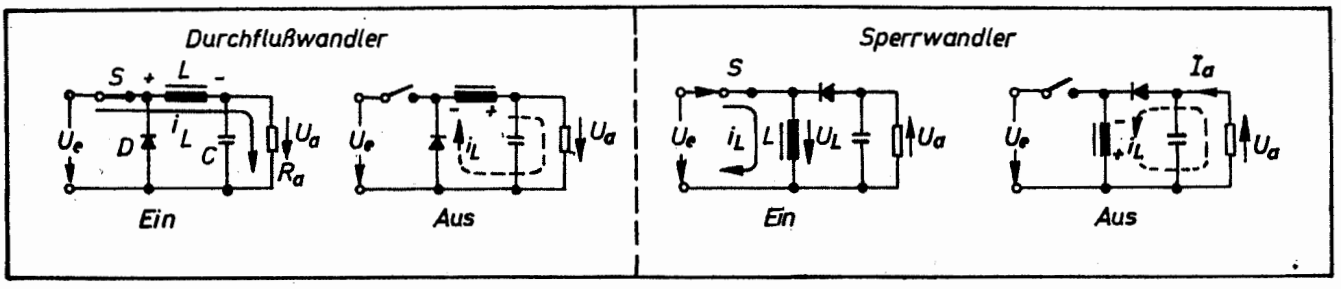


Bild 7.6.10. Ströme in der Ein- und Aus-Phase bei Durchfluß- und Sperrwandler

ist. Die Größe der Ausgangsspannung kann weiterhin über der der Eingangsspannung liegen. Mit der Ausschaltzeit T_a gilt

$$U_a = -U_e \frac{T_e}{T_a}, \tag{30}$$

wobei wieder ähnliche Beziehungen wie im vorigen Abschnitt für die Induktivität für die Welligkeit und den mittleren Eingangsstrom gelten.

7.6.6.4. Weitere Prinzipien

Neben den beiden behandelten Grundschaltungen existieren noch eine *Vielzahl von Schaltungsvarianten* mit den gleichen Prinzipien wie beim Durchfluß- bzw. Sperrwandler (z.B. Bilder 7.6.9h) und e)), aber auch andere, daraus abgeleitete Varianten. In jedem Fall kann der Impulsstrom durch einen zusätzlichen *Transformator* noch transformiert und darüber hinaus eine galvanische Trennung von Eingangs- und Ausgangskreis bewirkt werden. Lediglich der Sperrwandler nach Bild 7.6.9e) kommt mit nur einer Induktivität aus, die zugleich Transformator und Speicherdrossel ist. Dieser Sperrwandler ist besonders geeignet, wenn gleichzeitig mehrere Spannungen erzeugt werden sollen. Es sind dann nur zusätzliche Sekundärwicklungen und je eine Diode und Kapazität notwendig. Bei allen anderen Varianten ist für jede andere sekundäre Spannung darüber hinaus noch eine Speicherdrossel erforderlich.

Beim Durchflußwandler mit Transformator ist zusätzlich auf dem Trafo noch eine Wicklung mit Diode notwendig, welche die Gleichstromvormagnetisierung des Trafos in der Sperrphase aufhebt. Eine andere Methode besteht darin, im Gegentakt zwei Transistoren so zu schalten, daß im Primärkreis in beiden Phasen entgegengesetzt magnetisiert wird. Dies führt zu den *Gegentaktwandlern*. Schließlich ist es auch möglich, beide Phasen auf der Sekundärseite zu nutzen. Da hierbei in der Regel beide Halbwellenspannungen in Reihe geschaltet werden (DELON-Schaltung von Bild 5.4.6g)), wird dann auch vom *Summierwandler* gesprochen. Auch besteht die Möglichkeit, vom Transformator durch eine Rückkopplung die Schalttransistoren selbständig schwingen zu lassen. So entsteht der *selbstschwingende Transverter*. Für weitere Details sei auf [D 2, K 40, T 4] verwiesen.

7.6.6.5. Vor- und Nachteile

Die *Vorteile* der Schaltnetzteile (ohne Netztrafo) sind

- hoher energetischer *Wirkungsgrad* (kleiner Leistungstransistor),
- kleines *Volumen* und kleine *Masse* (kleiner Ferrittrafo infolge hoher Frequenz),
- großer *Eingangsspannungsbereich*,
- niedriger *Siebaufwand* infolge hoher Taktfrequenz,
- niedriger *Siebaufwand* auch im netzfrequenten Teil, da diese Welligkeit ausgeregelt wird.

Die *Nachteile* sind

- *komplizierte Schaltungstechnik* (heute durch spezielle Schaltkreise, vgl. 7.6.6.7., gelöst).,
- *spezielle Bauelemente*, u. a.
  - schnelle Hochvolttransistoren,
  - schaltfeste Elkos,
  - Spezialferrite für Transformatoren und Speicherdrosseln,
  - evtl. Optokoppler für galvanikfreie Impulsübertragung,
- *langsameres Reagieren* auf schnelle Lastwechsel (Grenze durch Taktfrequenz gegeben),
- *hoher Entstöraufwand*. Die Oberwellen der Schaltfrequenz besitzen erhebliche Energie und dürfen weder in das Netz noch in den Lastkreis gelangen (spezielle Filter an Ein- und Ausgang sowie evtl. Abschirmung).

- *besondere Schutzgütemaßnahmen,* da Schalttransistor und zugehörige Bauelemente galvanisch mit dem Netz verbunden sind (Kapselung und Vorsichtsmaßnahmen bei Reparatur).

Die Auswahl der Varianten Sperr- oder Durchflußwandler hängt vom Strom- und Spannungsbereich des Verbrauchers ab.

Der Sperrwandler ist die einfachste Schaltung. Seine Grenzen sind vor allem durch die notwendige feste Kopplung von Primär- und Sekundärwicklung des Trafos gegeben. Bei kleinen Leistungen ist eine Überdimensionierung des Trafos mit fester Kopplung möglich. Durch die Trennung von Transformator und Speicherdrossel ist der Durchflußwandler besser zu gestalten. Der höhere Aufwand ist vor allem bei großen Leistungen vorteilhaft. Im Übergangsgebiet und bei mehreren Ausgangsspannungen ist der Sperrwandler zumindest bei geringen Strömen, also höheren Ausgangsspannungen, vorteilhaft. So entsteht das Bild 7.6.9i) (vgl. auch Bild 6.4.1).

7.6.6.6. Eisenloser Gleichspannungswandler

Mit einem Schaltregler ist aufgrund der Induktivität gemäß Bild 7.6.10 eine höhere Ausgangsspannung möglich. Ein ganz anderes Prinzip läßt sich mit Verdopplungs- bzw. Vervielfachungsschaltungen (Bild 5.4.6) dann erreichen, wenn die Gleichspannung in eine Rechteckspannung umgeformt wird. So entsteht das Prinzip, welches Bild 7.6.11 zeigt. Ein frei schwingender Multivibrator steuert eine Gegentaktstufe an. Am Mittelpunkt entstehen so symmetrische Rechteckimpulse, die mit einer Verdopplungsschaltung zu negativem oder erhöhtem positiven Potential führen.

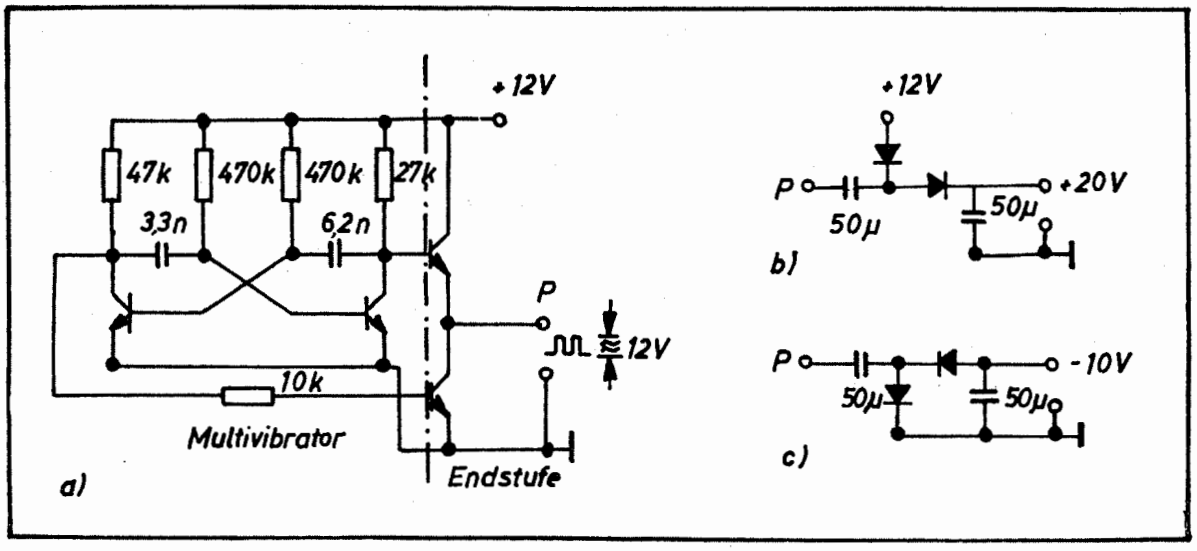


Bild 7.6.11. Eisenloser Spannungswandler

a) Multivibrator und Endstufe zur Erzeugung leistungsstarker Rechteckspannungen
b) Diodenschaltung zur Erhöhung der Spannung auf etwa das Doppelte
c) Diodenschaltung zur Erzeugung einer zusätzlichen negativen Spannung

7.6.6.7. Integrierte Ansteuerschaltkreise

Im Prinzip sind Schaltregler und Schaltnetzteile mit diskreten Bauteilen oder auch mit den üblichen Spannungsreglerschaltkreisen, wie z.B. MAA 723 oder MA 78XX, zu realisieren. Hierbei ist dann im einfachsten Fall über eine Spannungshysterese mit z. T. sich frei einstellender Frequenz zu schalten. Mit zusätzlichen Impulsschaltungen werden bessere Lösungen erreicht. Optimale Varianten ergeben sich jedoch mit *speziellen Ansteuerschaltkreisen,* wie z.B. B 260 (Vergleichstyp TDA 1060). Die Gesamtinnenschaltung ist dann jedoch bereits sehr komplex, so daß mit Bild 7.6.12 *nur ein Baugruppenbild* gezeigt sei.

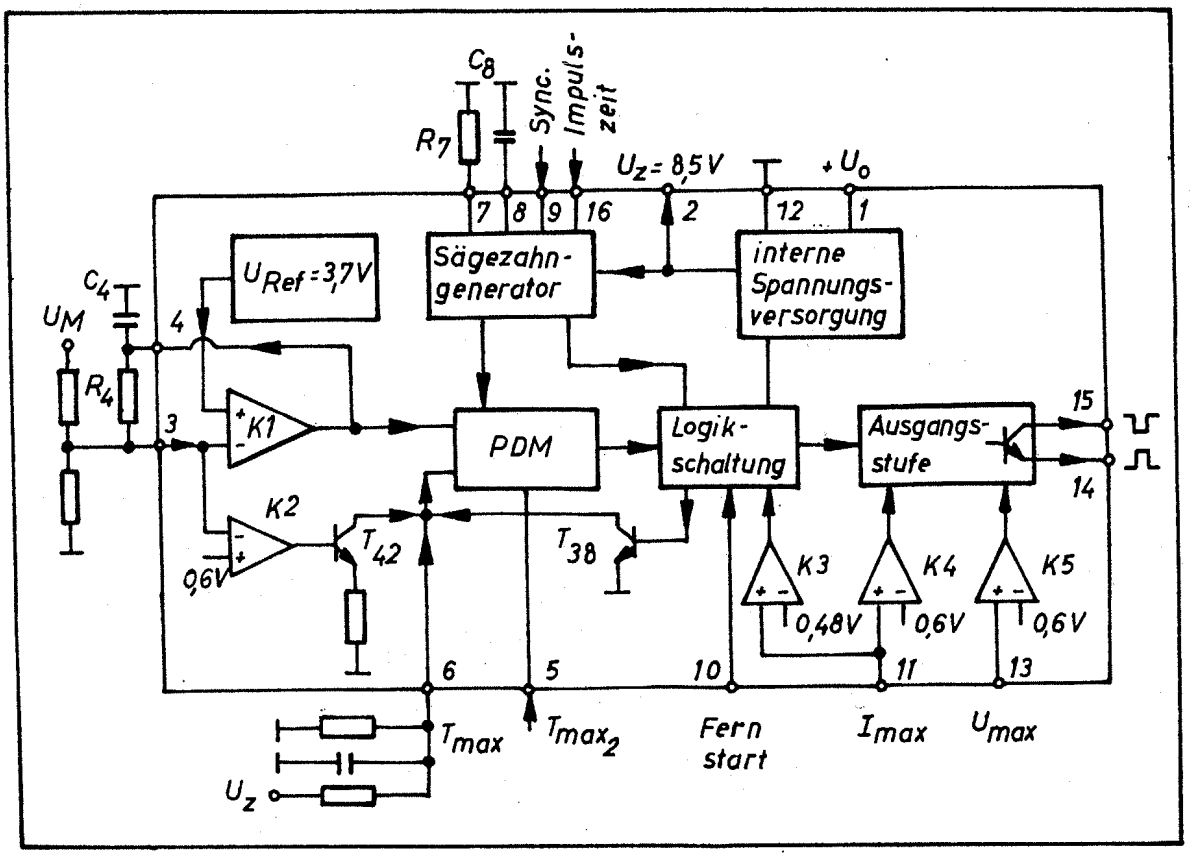


Bild 7.6.12. Blockschaltbild des Ansteuerschaltkreises für Schaltregler und Schaltnetzteile B 260

Versorgt wird der Schaltkreis mit einer positiven Spannung an Klemme 1. Sie soll im Bereich von 10,5 bis 18 V liegen. Durch eine interne Z-Diode (30 V) ist aber auch eine Versorgung per Strom (max. 30 mA) aus einer hohen Spannung möglich. Intern wird daraus eine *stabilisierte Spannung* von 8,5 V erzeugt, welche darüber hinaus mit max. 5 mA Belastung an 2 zur Verfügung steht. Die interne *Referenzspannung* von 3,7 V geht auf eine modifizierte *Bandabstandsquelle* (7.6.4.1.) zurück. Der *PDM-Modulator* entspricht dem Prinzip von Bild 7.3.28; wobei als Eingangsspannung die niedrigste von den Eingängen 4, 5 und 6 wirkt. Davon ist 4 zugleich der Ausgang des Komparators $K1$, wo aus U_M mit dem Spannungsteiler an 3 der Bezugswert zur Referenzspannung gültig ist.

Von den beiden anderen Anschlüssen werden *Schutzfunktionen* realisiert (siehe unten). Der *Sägezahngenerator* arbeitet nach dem Stromladeprinzip von Bild 7.3.24c). Seine Frequenz kann durch die Kapazität C_8 an 8 und der Strom mittels R_7 an 7 über Stromspiegel verstellt werden. Ferner kann er mit einer tieferen Frequenz an 9 auf eine *äußere Frequenz synchronisiert* werden. Über 16 ist eine weitere Beeinflussung möglich, die eine bestmögliche Auslastung der Transformatoren ermöglicht.

In der Ausgangsstufe steht ein angesteuerter Transistor zur Verfügung, über dessen Emitter bzw. Kollektor positive bzw. negative Impulse abgeleitet werden können. Zwischen PDM und Ausgangsstufe befindet sich noch eine Logikschaltung zur Auslösung weiterer Schutzfunktionen. So kann mittels 11 z.B. ein I_{max} aus einer Stromableitung hergestellt werden. Die auf 0,48 bzw. 0,6 V voreingestellten Komparatoren $K3$ und $K4$ *begrenzen dann das Impulsverhältnis* bzw. *schalten den Leistungstransistor ab.* Am Anschluß 13 kann dies unabhängig mittels des Komparators $K5$ für eine *maximale Spannung erfolgen.* Sowohl $K3$ als auch der Eingang 10 für *Fernsteuerung* (Ein — Aus) wirken über T_{38} auf den PDM ein. Sofern er eingeschaltet wird, ist 6 auf Masse geschaltet, und das Pulsverhältnis wird zu Null. Wenn er sich wieder öffnet, *läuft* über den an 6 liegenden Kondensator das *Pulsverhältnis langsam hoch.*

Es erfolgt also nicht ein plötzlicher, sondern ein „weicher" Start für den Leistungstransistor. Dies erfolgt auch, wenn die Ausgangsspannung des Reglers, aus der U_M abgeleitet ist, kurzgeschlossen wird. Dann wirkt der Komparator $K2$ über T_{42} genauso. Lediglich am Eingang 5 kann eine direkte Begrenzung des Pulsverhältnisses erfolgen. Mit R_4 kann die Verstärkung des Komparators $K1$ reduziert werden. C_4 erhöht die Stabilität.

7.7. Ausgewählte Gebiete zur Meßtechnik

7.7.1. Grundlagen des Messens

Vom Standpunkt der Technik ist das Messen von fundamentaler Bedeutung. Es schafft einmal reproduzierbare (meist quantifizierbare) Erkenntnisse über unsere Umwelt und trägt zur theoretisch vertieften Durchdringung der Welt und damit zu ihrer Beherrschung durch den Menschen bei. Zum anderen wird durch Messen aber auch wesentlich die Menge und Qualität von Gütern bestimmt. Hierdurch erlangt das Messen eine hohe ökonomische Bedeutung, und es wird verständlich, warum das Meßwesen durch die herrschende Klasse, durch den Staat, geregelt werden muß.

Für die Elektronik hat die Messung deshalb besondere Bedeutung erlangt, weil fast die gesamte moderne Meßtechnik sich elektronischer Methoden bedient. *Meßbar* sind *nicht die Objekte* selbst. Sie sind *nur zählbar. Meßbar* sind nur *Eigenschaften.* Sie müssen dabei zumindest zwei *Ausprägungsgrade,* z. B. das Geschlecht: also männlich und weiblich, besitzen. Existieren mehrere Ausprägungsgrade, so entsteht eine Skala, besonders einfach mit Rangfolge, wie bei der subjektiv feststellbaren Temperatur: eisig, frostig, kalt, kühl, lau, warm, heiß. Eine andere Möglichkeit besteht im *Vergleichen,* z. B. kälter bzw. wärmer. Der Übergang zu Zahlenwerten verlangt bereits mehr oder weniger objektive Vergleichswerte, z. B. im Sinne der Schulnoten oder bereits genauer im Sinne der Mohsschen Härteskala. Die größte physikalische Bedeutung haben die *Verhältnisskalen.* Sie führten erst zu den im System International (SI) festgelegten Bedingungen, die im weiteren vorausgesetzt seien [V 12].

7.7.1.1. Grundbegriffe

Einige Grundbegriffe der Meßtechnik seien mittels Bild 7.7.1a) erklärt. Bei dem Gegenstand wird hier die *Meßgröße* Länge mit der Schieblehre bestimmt. Die Skale (hier speziell mit zusätzlichem Nonius) dient der *Anzeige,* und an ihr ist der Zahlenwert von z. B. 12,4 ablesbar. Wird er mit der zugehörigen *Einheit* multipliziert, so entsteht der *Meßwert* von 12,4 mm. In einigen Fällen stellt der Meßwert das Ende der Messung dar, dann ist er zugleich das Meßergebnis. Im allgemeinen sind für das *Meßergebnis* noch weitere Messungen und Rechnungen notwendig. Ein anderes Meßergeb-

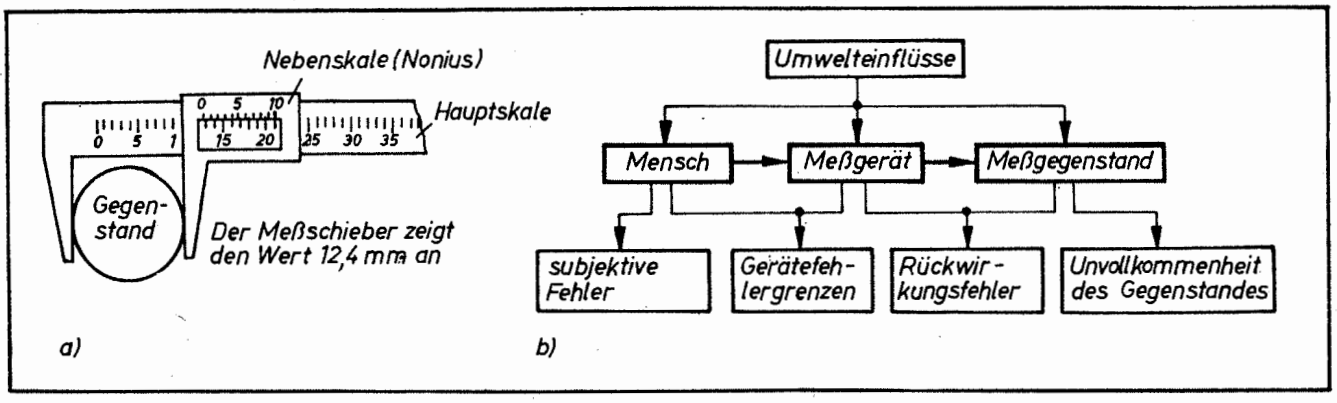


Bild 7.7.1. Prinzipien der Messung

a) Messung am Beispiel der Schieblehre

b) Messung als Wechselwirkung Mensch — Meßgerät — Meßgegenstand, mit Zuordnung einiger Fehlerprinzipien

nis für den in Bild 7.7.1a) gegebenen Gegenstand könnte sein Querschnitt, sein Volumen oder sein spezifisches Gewicht sein. Die Messung vollzieht sich — von den Ausnahmen der automatischen Meßtechnik abgesehen — gemäß dem Schema von Bild 7.7.1b). Der Mensch beobachtet ein Meßgerät, welches mit dem Objekt der Messung, dem Meßgegenstand, verkoppelt ist. Auf alle drei wirken störende Einflüsse der Umwelt, welche die Genauigkeit (Qualität) der Messung begrenzen. Die Abweichungen vom theoretisch richtigen Ergebnis heißen Fehler: Das Licht kann so dunkel sein, daß der Mensch nur sehr schwierig, und damit wahrscheinlich fehlerhaft, das Meßgerät ablesen kann. Die Temperatur, die Feuchte usw. können so extrem sein, daß das Meßgerät nur bedingt einwandfrei arbeitet. Der Meßgegenstand selbst kann unvollkommen sein, z.B. bei einer Kugel unrund. Aus diesen Gesichtspunkten lassen sich Fehlerbetrachtungen gewinnen.

Hier sollen lediglich die Fehlertypen behandelt werden:

Subjektive Fehler sind im wesentlichen Irrtümer des Menschen, z.B. eine falsche Ablesung.

In der Wechselwirkung von Mensch und Meßgerät liegen *Gerätefehlergrenzen.* Hierzu gehören z.B.: Wie gut ist die Skale des Gerätes geeicht? Wie sicher ist sie ablesbar? Durch die Digitaltechnik werden Fehler immer mehr eine Eigenschaft des Gerätes allein.

Die *Rückwirkungsfehler* sind prinzipieller Natur. Vom Meßgegenstand muß stets eine gewisse Energie zum Meßgerät übergehen, damit die Messung erfolgen kann. Dieser Energieentzug verfälscht aber Eigenschaften des Meßgegenstandes. Die physikalische Grenze ist durch die HEISENBERGsche Unschärferelation gegeben. Äquivalente Grenzen lassen sich aus der Informationstheorie ableiten.

Eine andere Fehlerklassifizierung bezieht sich auf die Art der Fehler: *Grobe Fehler* weichen sehr stark vom erwarteten Ergebnis bzw. vom Mittelwert ab. Im allgemeinen werden sie ausgeschieden. Hierbei ist jedoch Vorsicht angebracht. Schon öfter hat sich über vermeintliche grobe Fehler etwas unerwartetes Neues angekündigt. *Systematische Fehler* haben eine im Prinzip immer bestimmbare Größe und Richtung und können daher durch eine Korrektur zumindest teilweise berichtigt werden. *Zufällige Fehler* rühren von nicht beeinflußbaren und nicht erfaßbaren Veränderungen her. Sie besitzen bei jeder Einzelmessung einen anderen Wert. Deshalb schwankt der Meßwert bei wiederholten Messungen. Hierauf beruht die rechnerische Verbesserung eines Meßwertes aus vielen Messungen. Die zugehörige *Fehlerrechnung* ist in allen Standardwerken über Messen gründlich beschrieben und braucht deshalb hier nicht behandelt zu werden. Die *Meßunsicherheit* ergibt sich aus den berechneten Fehlern zuzüglich nur abschätzbarer systematischer Fehler.

7.7.1.2. Meßkette

Die Umwandlung der Meßgröße in den Meßwert, genauer Anzeige, kann unmittelbar, wie z.B. bei der Schieblehre und beim Drehspulinstrument, erfolgen. Dann wird von direkter Messung gesprochen. Sie kann aber auch mittelbar (indirekt) erfolgen, dann wird meist eine Aufteilung nach Bild 7.7.2 günstig sein. Hierbei ist vorausgesetzt, daß die Messungen auf elektrischem Wege erfolgen. Dies ist heute wegen der Flexibilität und Anpassungsfähigkeit der elektrischen und insbesondere der elektronischen Meßtechnik fast ausschließlich der Fall und gilt ganz besonders im Rahmen dieses Buches. Dann stellt der Meßwandler jenes Glied dar, das die eventuell nichtelektrische Größe in eine elektrische umformt. An Wandlern steht meist eine große Auswahl zur Verfügung (vgl. 6.3.4.). Auf sie kann daher hier nicht eingegangen werden, obwohl sie nicht selten für den Naturwissenschaftler von grundlegender Bedeutung sind. Es sei

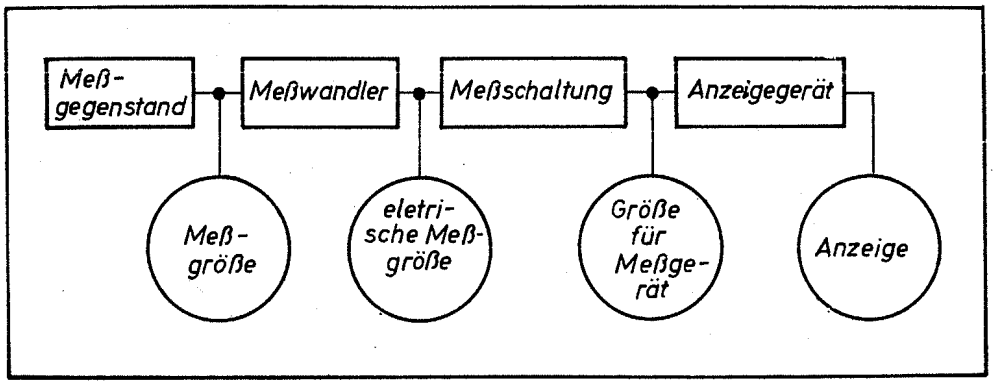


Bild 7.7.2. Meßkette als Beispiel für den oft komplexen Zusammenhang zwischen der gemessenen und der angezeigten Größe

aber auf die einschlägige Fachliteratur, z.B. [A 15, B 7, D 10, G 20, H 11, H 36, K 34, P 19 I, P 25, S 31, S 52 und T 5] hingewiesen.

Die vom Meßwandler erhaltene elektrische Größe ist meist noch nicht unmittelbar für das Anzeigegerät geeignet. Sie kann sowohl einen zu großen als auch einen zu kleinen Wert aufweisen, dann enthält die Meßschaltung Abschwächer oder Verstärker. Die Meßgröße kann aber auch an ein ungeeignetes Signal gebunden sein, z.B. Frequenz statt Spannung oder Phase statt Strom. Dann besorgt die Meßschaltung die entsprechende Umformung. Schließlich kann auch der Meßgrößenbereich nicht dem Anzeigebereich angepaßt sein, z.B. wenn die Frequenzschwankung eines Oszillators von etwa 10^{-5} angezeigt werden soll. Auf ausgewählte Meßschaltungen wird noch im folgenden eingegangen.

Die nach der Meßschaltung passend geformte Meßgröße wird dem Anzeigegerät zugeführt und dort abgelesen, registriert oder gespeichert. Wichtigste Anzeige- und Registriergeräte wurden bereits in dem Abschnitt 6.3.3. behandelt.

Für eine vollständige Meßanordnung bestehen verschiedene Möglichkeiten nach dem Umfang der bei einem Versuch erhaltenen Meßergebnisse. Dabei sollen die vier Qualitäten

- Einzelmessung,
- Meßreihe,
- Registrierung und
- Speicherung mit Reproduktion des elektrischen Signals

unterschieden werden. Ihr Verhältnis sei am Beispiel von Spannungsmessungen des Lichtnetzes erklärt:

Genügt es zu bestimmen, welchen Nennwert die Spannung eines unbekannten Netzes besitzt, so ist eine *Einzelmessung* mit dem Meßergebnis von vielleicht 217 V voll ausreichend. Es handelt sich dann um ein Netz von 220 V Nennspannung.

Soll dagegen die tageszeitliche Schwankung dieser Spannung bestimmt werden, so wird in der Regel eine mehrtägige *Meßreihe* im Asbtand von jeweils etwa 15 Minuten vollauf genügen. Die einzelnen Meßergebnisse werden abgelesen, aufgeschrieben und grafisch dargestellt.

Bei der Meßreihe ist der manuelle Aufwand des Ablesens, Aufschreibens und grafischen Darstellens erheblich. So entstanden *Registriergeräte*, die dieses automatisch erledigen. Die Aufnahme vieler Einzelmeßpunkte wird durch einen einzigen Registriervorgang ersetzt.

Entsprechend der Vielzahl von Möglichkeiten bei Meßreihen ist es zuweilen aufwendiger, die Registrierung vorzubereiten, als die Meßreihe durchzuführen. Deshalb hat auch heute noch die Meßreihe große Bedeutung, so zum Beispiel, wenn aus einer Vielzahl von Individuen, z.B. Halbleiterbauelementen, der typische Mittelwert und die zugehörige Schwankungsbreite bestimmt werden soll. In solchen Fällen kann dann die Auswertung der Meßergebnisse mittels der Rechentechnik zu erheblichen Vereinfachungen führen. Die Registrierung hat gegenüber der Meßreihe den vielleicht nicht zu unterschätzenden Vorteil, daß sie kontinuierlich bzw. in sehr dichter

Folge vorgenommen werden kann. Dadurch werden vielfach Einzelheiten sichtbar, die nicht erwartet wurden oder prinzipiell nicht mit einer Meßreihe zu erfassen sind. Die Registrierung kann nämlich auch mit sehr großen Geschwindigkeiten und damit großem zeitlichen Auflösungsvermögen erfolgen. So entsteht der große Vorteil der Dehnung des Zeitmaßstabes. Bei der Einzelmessung ist nur die Raffung möglich.

Ein weiterer Vorteil wird bei der elektromagnetischen *Speicherung von Meßwerten* gewonnen. Hierdurch kann nämlich die ursprüngliche Meßgröße auch nach dem Meßvorgang wieder als elektrisches Signal reproduziert werden. Die Hauptbedeutung hat hierbei der Magnetbandspeicher. Er ist so universell und anpassungsfähig, daß in dem letzten Jahrzehnt sein Einsatz sehr steil zunahm. Dabei spielte auch die Kosmosforschung eine wichtige Rolle. Bezüglich seiner vielfältigen Eigenschaften und Vorteile sei auf die entsprechende Fachliteratur, z. B. [V 10], sowie auf die neuen Möglichkeiten beim digitalen Oszillographen (7.7.4.7.) verwiesen.

7.7.1.3. Maßeinheiten

Der Übersichtlichkeit halber sei als das einfachste Messen das *Abzählen* erwähnt; es führt zu einer Zahl als Meßergebnis. Dabei ist aber bekannt, daß nur gleichartige Stückzahlen zusammengezählt werden dürfen, und es muß in der Regel der Begriff des Abgezählten der Zahl hinzugefügt werden: z. B. 57 Stück Transistoren. Etwas allgemeiner ist bereits die Angabe eines Preises, z. B. 7 Mark und 43 Pfennige. Die meisten Messungen beruhen auf einem direkten oder indirekten Vergleich mit einem vereinbarten *Normal*. Dieses Normal legt die Einheit, z. B. das Meter fest. Alle gemessenen Größen, kurz alle *Meßgrößen* bestehen daher aus einem *Zahlenwert*, z. B. 2,31 und der *Einheit*. Von allen Einheiten können Vielfache oder Teile gebildet werden. Hierzu stehen die genormten *Vorsätze* zur Verfügung. So entstehen Einheiten wie cm, km, mA, MΩ, THz usw. Eine Meßgröße kann auch infolge der geschichtlichen Entwicklung der Einheiten sehr verschieden angegeben werden. Dabei ist aber der „Wert" der Meßgröße immer derselbe, deshalb gilt z. B.

$$100\ \mathrm{km\ h^{-1}} \approx 27{,}8\ \mathrm{m\ s^{-1}} \approx 54\ \mathrm{sm\ h^{-1}} \approx 54\ \mathrm{kn}\,.$$

Nur unter dieser Voraussetzung können Gleichungen ohne zusätzliche Angabe der zu verwendenden Einheiten, also ganz allgemein, geschrieben werden. Es müssen in sie die Zahlenwerte aus einem zusammengehörigen Maßsystem eingesetzt werden. Auch für die Maßsysteme gibt es eine geschichtliche Entwicklung. Noch um die Jahrhundertwende bestanden viele Maßsysteme nebeneinander. Unter anderem wurde lange Zeit geglaubt, alle physikalischen Größen seien auf drei mechanische Grundgrößen, z. B. Länge, Masse und Zeit zurückführbar. Dabei entstanden u. a. Mehrdeutigkeiten sowie gebrochene Exponenten bei den zusammengesetzten Einheiten, insbesondere aber bei elektrischen und magnetischen Größen. Nach langer Forschung wurde der entscheidende Abschluß mit dem Beschluß der X. Generalkonferenz für Maße und Gewichte 1954 in Genf erreicht. Die DDR erhob als einer der ersten Staaten diesen Beschluß zum Gesetz, und zwar am 14. August 1958. Nach dieser Regelung wird von Grundeinheiten ausgegangen. Innerhalb der Mechanik werden davon drei benötigt. Festgelegt wurden das *Meter*, die *Sekunde* und das *Kilogramm*. Beim Übergang zum Elektromagnetismus wird eine zusätzliche Größe benötigt. Ausgewählt wurde das *Ampere*. Für jedes neue Gebiet ist jeweils eine weitere neue Grundgröße einzuführen. Für die Thermodynamik wurde die *Kelvin-Temperatur* und für die Lichttechnik die *Candela* festgelegt. So entsteht das System International (SI), wie es überall genauer beschrieben ist.

7.7.1.4. Logarithmische Maße

Viele elektrische Größen überstreichen einen Bereich von vielen Zehnerpotenzen. Dann ist die Angabe der Meßwerte mit einem logarithmischen Maß vorteilhaft. Mit

ihm werden oft nur Verhältnisse angegeben. Das wichtigste logarithmische Maß ist das *Dezibel* (dB) (zusammengesetzt aus dezi und BELL). Es war ursprünglich nur auf Leistungen bezogen. Entsprechend dem Vorsatz dezi gilt dann

$$b\,[\mathrm{dB}] = 10 \log \frac{P_1}{P_2}. \tag{1}$$

Weitaus wichtiger ist jedoch seine Anwendung auf Ströme und Spannungen geworden. Wegen der Beziehung

$$P = \frac{U^2}{R} = I^2 R \tag{2}$$

kann dies nur für einen gemeinsamen Widerstand bei beiden Leistungen gelten. Hiervon wurde im Laufe der Entwicklung aber abgesehen. Wenn dann das Quadrat vor den Logarithmus gezogen wird, gilt

$$b\,[\mathrm{dB}] = 20 \log \frac{U_1}{U_2} = 20 \log \frac{I_1}{I_2}. \tag{1a}$$

Für jedes Spannungs- bzw. Stromverhältnis kann so ein dB-Wert angegeben werden.

Zuweilen wird auch eine *absolute Dezibelskala* für Spannungen benutzt. Dann ist der *Bezugspegel* von 0 dB *auf* ca. 775 mV festgesetzt. Das ist jene Spannung, die an dem Widerstand von 600 Ω genau eine Leistung von 1 mW erzeugt.

Wird statt des BRIGGSschen Logarithmus in Gl. (1) und (3) der natürliche verwendet und der Faktor 10 fortgelassen, so entsteht ein anderes Dämpfungsmaß, das *Neper* (Np) (nach JOHN NAPIER):

$$b\,[\mathrm{Np}] = \ln \frac{P_1}{P_2}. \tag{1b}$$

Es unterscheidet sich von dem Dezibelmaß durch einen Faktor (20/ln 10):

$$1\ \mathrm{Np} \approx 8{,}69\ \mathrm{dB}\ . \tag{3}$$

Der zugehörige Bezugsstrom beträgt ca. 1,29 mA. In den USA existiert ein Bezugswert von 6 mW an 500 Ω.

In der Rundfunktechnik gibt es weiterhin eine Empfindlichkeitsangabe. Der Nullpegel ist hierbei auf dreierlei Art festgelegt:

- Amplitudenmodulierte Signale: 1 V (dB V)
- Amplitudenmodulierte Signale mit Ferrit- oder Stabantenne: 1 V/m (Feldstärke)
- Frequenzmodulierte Signale und Fernsehen: 1 mV bzw. 1 μV (dB μV) mit Angabe des Anpaßwiderstandes

7.7.1.5. Meßmethoden

Die Standardmethode des Messens besteht darin, daß die Anzeige auf ein Normal bzw. auf einen Bezugswert bezogen wird. Das Verhältnis bzw. die Differenz erscheint dann als Länge oder im digitalen Bereich als Zahl. Dieses Prinzip nennt man

Ausschlagmethode (Bild 7.7.3).

Daneben existieren noch

Differenzmethode
Kompensationsmethode
Brückenmethode.

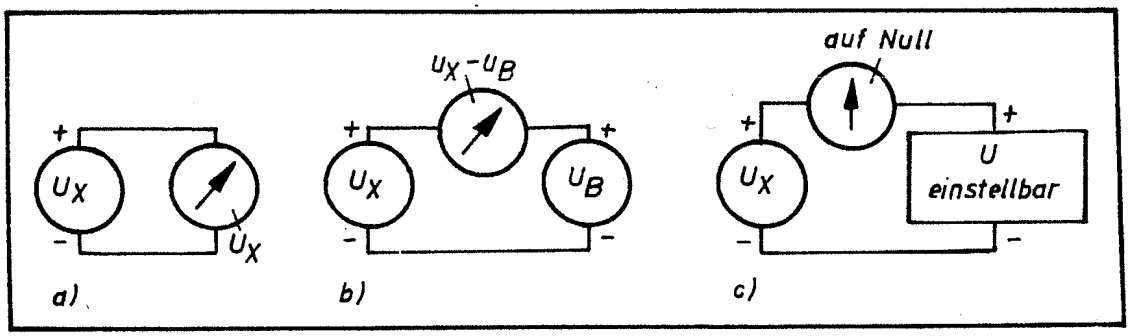


Bild 7.7.3. Die drei wichtigsten Meßmethoden

a) Ausschlagmethode: Die unbekannte Spannung wird unmittelbar zur Anzeige gebracht
b) Differenzmethode: Die unbekannte Spannung wird mit einer bekannten Bezugsspannung U_B verglichen und die Differenz zur Anzeige gebracht
c) Kompensationsmethode: Eine exakt einstellbare, bekannte Spannung wird solange verändert, bis sie mit der unbekannten übereinstimmt

Bei der *Differenzmethode* wird der Bezugswert möglichst in die Nähe des zu messenden Wertes gebracht, und nur die Differenz, die dann natürlich viel genauer meßbar ist, mittels Ausschlagmethode bestimmt.

Die *Kompensationsmethode* verändert den Bezugswert so lange, bis er mit dem zu messenden Wert übereinstimmt. Die Differenz wird also auf Null abgeglichen.

Weiter abgewandelt entsteht daraus die *Brückenmethode*, wie sie z. B. in den Bildern 2.3.8 bis 2.3.10 an Beispielen behandelt wurde. Sie stellt in anderer Betrachtung eine Meßschaltung dar und wird daher meist dann angewendet, wenn keine effektive, unmittelbare Messung nach den anderen Methoden möglich ist.

Von den bisher behandelten Methoden und Prinzipien weicht der

Oszillograph

erheblich ab. *Seine Funktion* ist in erster Linie, Schwingungsformen (-verläufe) und mehrdimensionale Funktionsabläufe *darzustellen*. Diese vermitteln einen guten Einblick in das ablaufende Geschehen und sagen in dieser Hinsicht meist sogar mehr aus als der Meßwert, aber Messen ist dies eben dennoch primär nicht, obwohl so auch Meßwerte zu erhalten sind (vgl. 7.7.4.).

7.7.1.6. Dem Messen verwandte Methoden

Im gewissen Sinne hätte auch hier der Oszillograph genannt werden können. Doch es gibt eine *Vielzahl* anderer *verwandter Methoden*, die u. a. in [H 11] und [H 36] vollständiger und genauer als hier behandelt sind.

Zählen ist primär eine digitale Meßtätigkeit, bei der nur die Anzahl von Stücken (und ähnlichem) in einer vorgegebenen Menge bzw. Zeit zu bestimmen ist (vgl. 8.5.4.).

Sortieren erfolgt, wenn aus einer Menge nach bestimmten Grenzmeßwerten Klassen gebildet werden sollen.

Eichen erfolgt, wenn ein Meßgerät auf die richtigen Bezugswerte eingestellt wird, damit es richtig mißt. Es wird in diesem Zusammenhang auch von *Kalibrieren* (*Justieren*) bzw. *Abgleichen* gesprochen. Darüber hinaus ist Eichen im engeren Sinne gebräuchlich für die amtliche Tätigkeit von Eichbehörden, in der DDR z. B. das Amt für Standardisierung und Meßwesen.

Prüfen ist eine spezielle Meßtätigkeit, bei der nur festgestellt wird, ob das Gerät, die Schaltung usw. innerhalb der vorgegebenen Toleranzen arbeitet. In diesem Sinne können subjektives (unmittelbar mit unseren Sinnesorganen erfolgendes) und objektives (mit Meßmitteln vorgenommenes) Prüfen unterschieden werden.

Speichern ist ein allgemeines Prinzip, mit dem Meßwerte so festgehalten werden, daß sie später wieder abrufbar sind.

Messen ist nicht Selbstzweck. Aus diesem Grunde werden zuweilen auch jene Bereiche mit einbezogen, die den Messungen folgen, wie z.B. das Verstellen bzw. Stellen von Größen (meist als ein Teil der Regelungstechnik aufgefaßt) oder das immer bedeutsamer werdende *Fernwirken* [F 6].

7.7.2. Einfache Meßschaltungen

7.7.2.1. Messung von Gleichwerten

Ein Drehspulinstrument erreicht bei einem bestimmten Strom I_m seinen *Vollausschlag*. Gleichzeitig besitzt es einen bestimmten *Innenwiderstand* R_{im}. Infolgedessen liegt beim Vollausschlag eine bestimmte Spannung U_m an der Drehspule:

$$U_m = I_m \cdot R_{im} . \tag{4}$$

Neben dem Stromverbrauch bei Vollausschlag I_m ist es üblich, den reziproken Wert als *Stromdämmung* gemäß

$$\boxed{D_i = \frac{1}{I_m} = \frac{R_{im}}{U_m}} \tag{4a}$$

in Ω/V anzugeben. Hierdurch erhalten Instrumente mit hoher *Empfindlichkeit* einen großen Wert der Stromdämmung zugeordnet, während ihr Strom für Vollausschlag ja gerade sehr klein ist. Heute sind Instrumente mit Werten zwischen 10000 und 100000 Ω/V, also 100 ... 10 μA durchaus üblich, während vor etwa 20 Jahren ein Wert von 1000 Ω/V schon als sehr gut angesehen wurde. Eine weitere Steigerung der Empfindlichkeit scheint in Anbetracht der Transistortechnik nicht mehr notwendig. Außerdem sind derartig elektrisch empfindliche Instrumente auch gegenüber mechanischen Beeinflussungen recht anfällig. Robuste Standardinstrumente benötigen etwa 1 mA an 100 Ω.

Entsprechend Gl. (4) ist das Drehspulinstrument auch zur *Spannungsmessung* anwendbar. Damit es für verschiedene Spannungen benutzt werden kann, sind mit dem Instrument Vorwiderstände R_v (Bild 7.7.4a)) in Reihe zu schalten. Es gilt dann

$$U_{m,R} = I_m(R_i + R_v) . \tag{5}$$

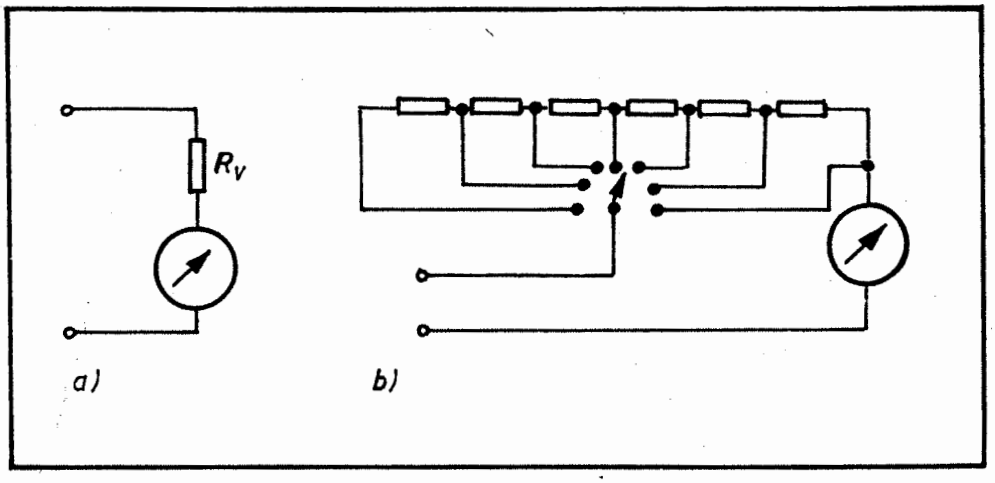


Bild 7.7.4. Zur Anwendung des Drehspulinstrumentes zur Spannungsmessung

a) Einschaltung des Vorwiderstandes

b) Umschaltung des Meßbereiches

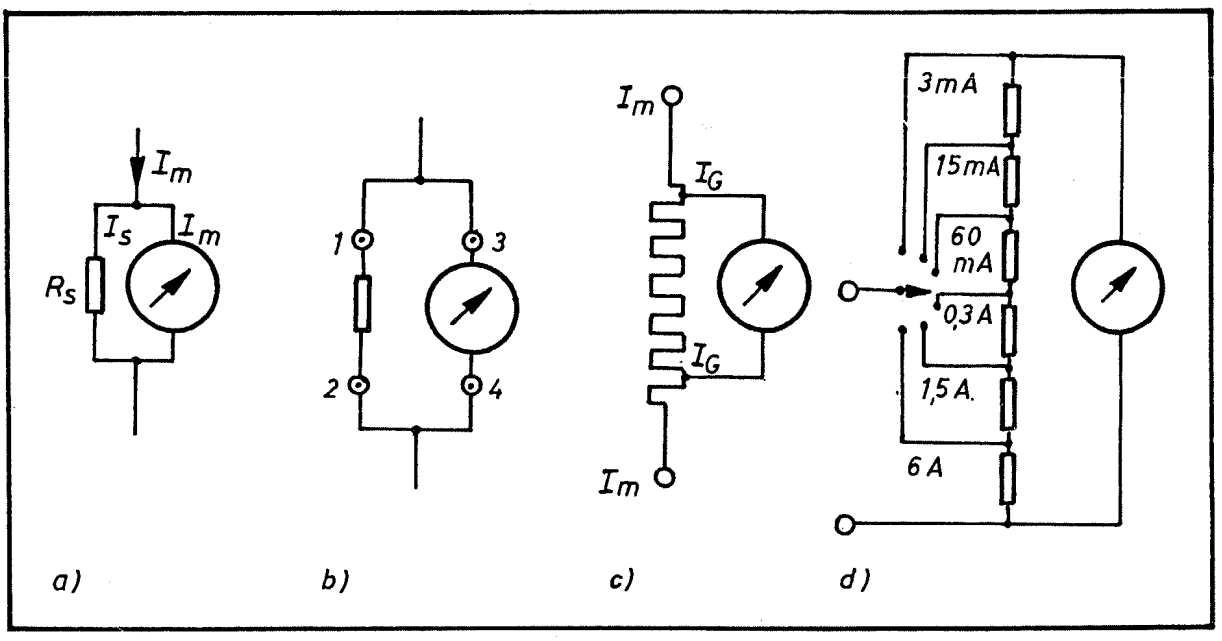


Bild 7.7.5. a) Änderung des Strommeßbereiches durch einen Shunt
b) Zur Bestimmung des Einflusses von Kontaktübergangswiderständen
c) Unterscheidung von Strommeßpfad I_m und Gerätemeßpfad I_G
d) Anordnung zur Umschaltung des Strommeßbereiches

Mit mehreren Vorwiderständen sind verschiedene Spannungsbereiche zu realisieren. Die Spannungsauswahl erfolgt z.B. über Bereichsschalter (Bild 7.7.4b)). Bei einer Messung ist es sinnvoll, mit dem *Bereich für die größte Spannung zu beginnen* und dann schrittweise bis in die Nähe des Vollausschlages herabzuschalten. So kann eine Überlastung des Meßgerätes vermieden werden. Ein Drehspulinstrument kann unmittelbar zur *Strommessung* benutzt werden, wenn der zu messende Strom innerhalb seines Meßbereiches liegt. Wesentlich kleinere Ströme zu messen, ist nur mittels eines zusätzlichen Verstärkers möglich. Für größere Ströme ergibt sich jedoch die Möglichkeit, das Meßinstrument durch *Parallelwiderstände* R_s (Shunts) gemäß Bild 7.7.5a) anzupassen. Für den Strom I_s durch den Shunt im Vergleich zum Strom I_m durch das Instrument mit dem Innenwiderstand R_{im} gilt

$$\frac{I_s}{I_m} = \frac{R_{im}}{R_s}. \tag{6}$$

Der Meßstrom bei Vollausschlag des geshunteten Instrumentes beträgt somit

$$I_{m,R} = I_m + I_s = I_m \left(1 + \frac{R_{im}}{R_s}\right). \tag{7}$$

Sind dagegen der erwünschte Strom bei Vollausschlag $I_{m,R}$ und die Eigenschaften des Drehspulinstrumentes, z.B. R_{im} und I_m, gegeben, so berechnet sich der notwendige Shunt zu

$$R_s = R_{im} \frac{I_s}{I_{m,R} - I_m}. \tag{8}$$

Dieser Widerstand kann, insbesondere wenn große Ströme gemessen werden sollen, sehr klein werden. Deshalb können (eventuell veränderliche) Kontaktwiderstände die Meßgenauigkeit erheblich beeinflussen oder sogar das Drehspulinstrument zerstören. Dies möge mittels Bild 7.7.5b) erklärt werden.

Wenn der *Übergangswiderstand einer der Kontaktstellen* 1 oder 2 größer wird, so entspricht das einem vergrößerten Shunt, und damit muß das Drehspulinstrument

einen größeren Strom aufnehmen und kann überlastet werden. Selbst sehr geringe Änderungen des Übergangswiderstandes verändern in dieser Schaltung erheblich die Meßgenauigkeit. Die Kontaktstellen 3 und 4 haben dagegen wegen des erheblich größeren Innenwiderstandes des Meßinstrumentes weitaus weniger Einfluß auf die Genauigkeit des Meßgerätes. Eine Überlastung des Instrumentes kann durch eine Änderung des Kontaktwiderstandes prinzipiell nicht auftreten.

Entsprechend den vorangegangenen Erkenntnissen ist die Anordnung des Shunts gemäß Bild 7.7.5c) gewählt. Sie unterscheidet zwei Anschlußstellen, einen *Strompfad* I_m und einen *Gerätepfad* I_G. In ähnlicher Art werden daher auch die Shunts bei Umschaltung betrieben (Bild 7.7.5d). Diese Schaltung hat dann jedoch den Nachteil, daß der Spannungsabfall bei den verschiedenen Strombereichen unterschiedlich ist. Er nimmt zu größeren Strombereichen hin zu.

7.7.2.2. Korrektur des Rückwirkungsfehlers

Sowohl bei der Spannungs- als auch bei der Strommessung treten infolge des Eigenverbrauchs des Instruments Meßfehler auf. Bei der Spannungsmessung kann der Meßpunkt durch die zu messende Ersatzspannungsquelle mit dem Innenwiderstand R_i ersetzt werden. Da der Innenwiderstand nicht bekannt ist, werden vorteilhaft zwei Messungen ausgeführt (Bild 7.7.6a)): eine in der üblichen Weise und eine zweite mittels eines Zusatzwiderstandes R_2. Beträgt dabei die Spannungsänderung

$$\Delta U = U_{m1} - U_{m2}\,, \tag{9a}$$

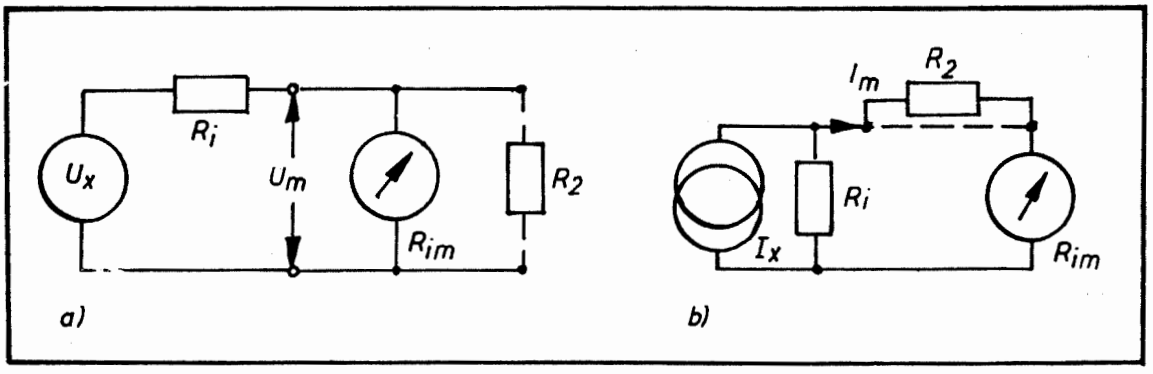


Bild 7.7.6. Zur Berücksichtigung des Verlustes am Meßinstrument

a) bei Spannungsmessung, b) bei Strommessung, mittels des zusätzlichen Widerstandes R_2 können über Gl. (9) bzw. (10) die wirklichen Werte berechnet werden

so gilt für den richtigen Spannungswert, ohne Belastung durch das Meßinstrument

$$U_x = \frac{U_{m1}}{1 - \dfrac{R_2}{R_{im}}\dfrac{\Delta U}{U_{m2}}} \tag{9b}$$

Besonders wenn R_2 groß gegen R_{im} gewählt wird, kann in Gl. (9a) $U_{m2} \approx U_{m1}$ gesetzt werden, und es ergibt sich eine verhältnismäßig einfache Auswertung mittels ΔU und U_{m1}. Bei der Strommessung muß der Stromkreis unterbrochen und das Instrument eingefügt werden. Es ist auch hier sinnvoll, für die zweite Messung einen kleinen Reihenwiderstand R_2 einzufügen. Analog zur Spannungsmessung gilt dann

$$I_x = I_1\left(1 + \frac{1}{\dfrac{R_2}{R_{im}}\dfrac{I_2}{\Delta I} - 1}\right) \tag{10}$$

7.7.2.3. Widerstandsmessung

Mit einer Schaltung gemäß Bild 7.7.7a) kann der Widerstand R_x aus dem Ausschlag des Meßinstruments bestimmt werden. Da er dem Strom proportional ist, gilt

$$I = \frac{U_Q}{R_x + R_v + R_{im}}. \tag{11}$$

Soll bei $R_x = 0$ der *Maximalausschlag*, also I_m erreicht werden, so muß gelten

$$R_v + R_{im} = \frac{U_Q}{I_m}. \tag{12}$$

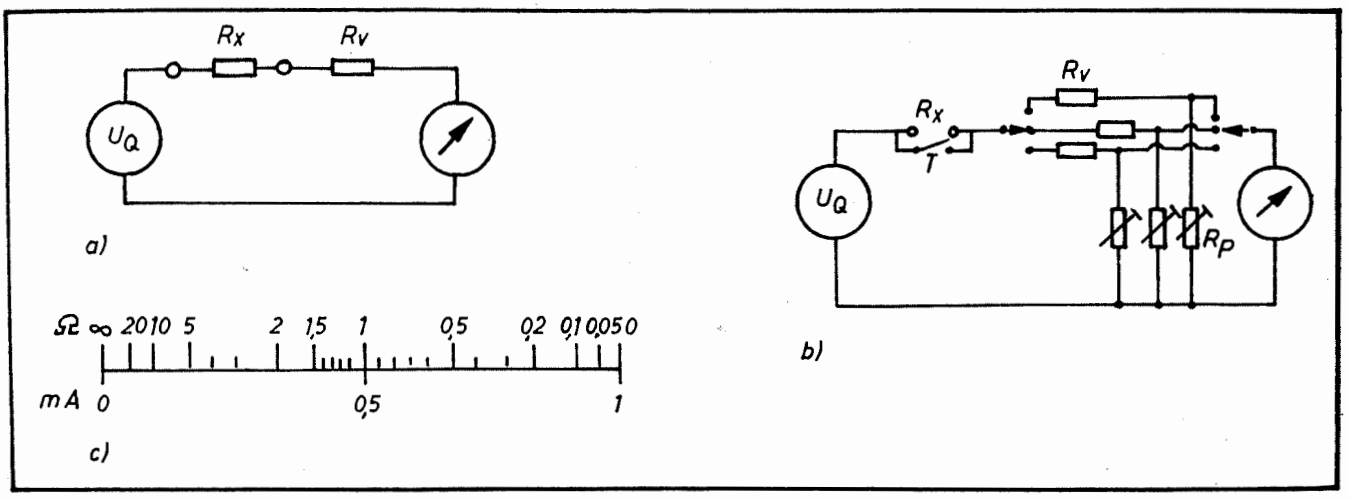


Bild 7.7.7. Zur Messung von Widerständen

a) Prinzipschaltung
b) Übliche Schaltung mit mehreren Meßbereichen
c) Umgeeichte Skale

Wird der Anzeigestrom I auf den *Maximalstrom* I_m bezogen, so folgt

$$I/I_m = \frac{1}{1 + \frac{I_m}{U_Q} R_x}. \tag{13}$$

Hieraus folgt die *nichtlineare Skale* für R_x, wie sie Bild 7.7.7 c) zeigt. Die größte Genauigkeit wird für

$$R_x \approx R_v + R_{im} = U_Q/I_m \tag{14}$$

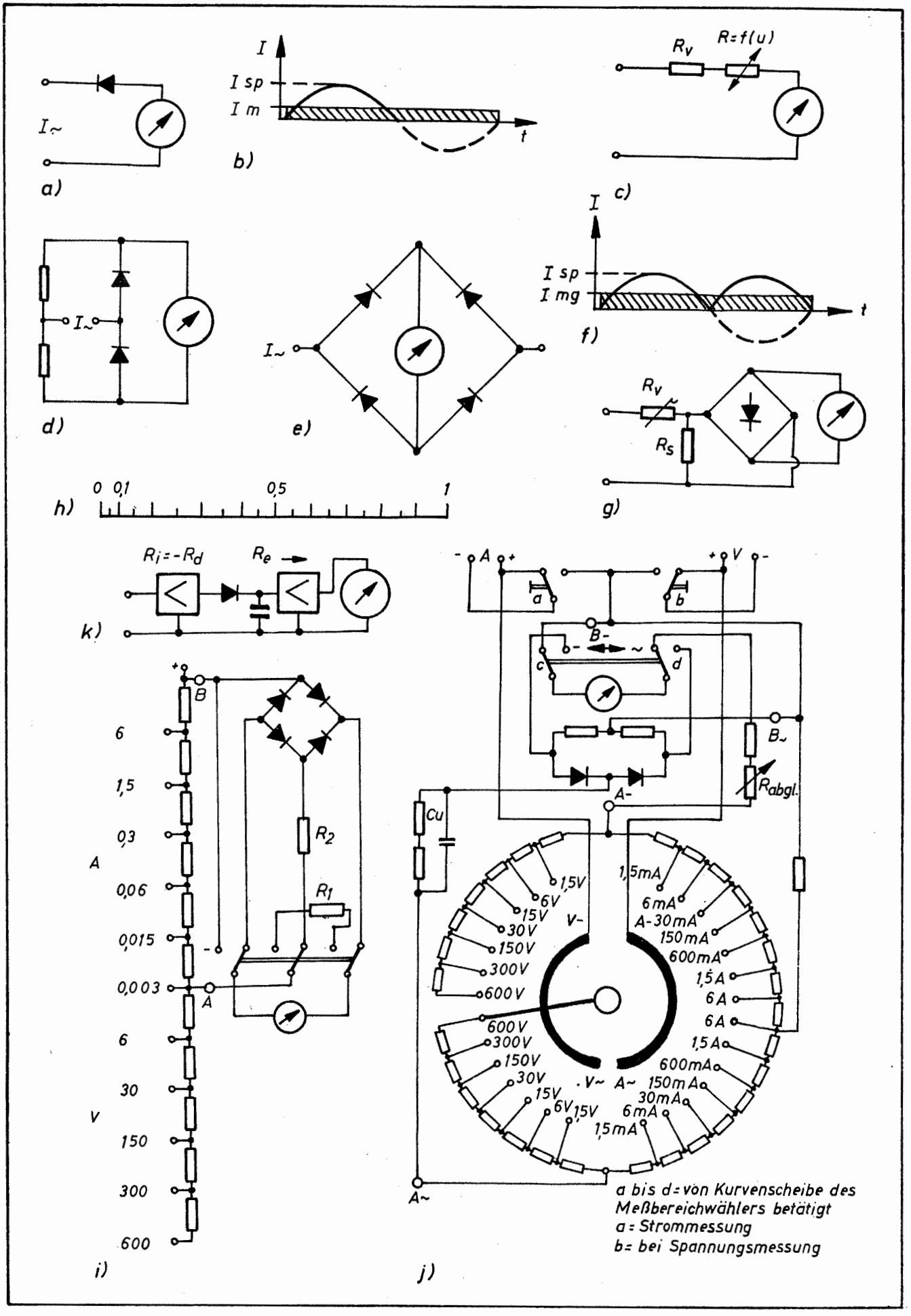
erreicht. An den Skalenenden liegen die Widerstandswerte 0 und ∞.

Bei den üblichen Instrumenten wird U_Q durch eine Batterie erzeugt. Sie ändert im Laufe der Zeit ihre Spannung, und dies erfordert einen Nachgleich. Wird er nun durch R_v vorgenommen, so stimmen nur die Werte 0 und ∞, aber die Mitte ist ungenau. Vorteilhaft ist es daher, R_v groß gegen R_{im} zu wählen, was meist relativ leicht zu realisieren ist. Dann kann der Abgleich durch einen Widerstand R_p parallel zum Instrument erfolgen. Nach diesem Prinzip läßt sich auch leicht gemäß Bild 7.7.7 b) eine Meßbereichsschaltung realisieren.

Schließlich sei noch darauf hingewiesen, daß es mit dem Operationsverstärker gemäß Tab. 7.5.7, Zeile 6, auch möglich ist, eine lineare Widerstandsskale zu erreichen.

7.7.2.4. Wechselstrommessung und Vielfachinstrumente

Mittels *Gleichrichter* kann mit einem Drehspulinstrument auch *Wechselstrom* gemessen werden. So entsteht die einfachste Schaltung von Bild 7.7.8a). Bei sinusförmigem


a)
I~
b)
I
I sp
I m
t
c)
Rv
R=f(u)
d)
I~
e)
I~
f)
I
I sp
I mg
t
g)
Rv
Rs
h)
0 0,1 0,5 1
k)
Ri=-Rd
Re
i)
B
6
1,5
0,3
A
0,06
0,015
0,003
A
6
30
V
150
300
600
R2
R1
j)
- A +
+ V -
a
b
B-
c
d
B~
Rabgl.
A-
Cu
1,5V
6V
15V
30V
150V
300V
600V
V-
1,5mA
6mA
A-30mA
150mA
600mA
1,5A
6A
6A
1,5A
600mA
150mA
30mA
6mA
1,5mA
15V
6V
1,5V
30V
150V
300V
600V
V~ A~
A~
a bis d= von Kurvenscheibe des Meßbereichwählers betätigt
a = Strommessung
b= bei Spannungsmessung

Wechselstrom wird die Fläche einer Halbwelle auf das gestrichelte Rechteck in Bild 7.7.8b) verteilt. Gemäß Abschnitt 1.4.9. beträgt der mittlere Gleichstrom dann

$$I_{sp}/\pi \approx 0{,}317 I_{sp} \quad \text{bzw.} \quad I_{eff}\sqrt{2}/\pi \approx 0{,}450 I_{eff}\,.$$

Leider tritt dieser ideale Fall nur auf, wenn der *Durchlaßwiderstand klein gegen* R_{im} und der *Sperrwiderstand sehr groß gegen* R_{im} ist. Dies ist nur bedingt der Fall. Für kleine Meßspannungen ist dies praktisch nie erfüllt. Deshalb besitzen Instrumente mit einem Gleichrichter keine Bedeutung. Dies wird noch deutlicher, wenn über einen Vorwiderstand gemäß Teilbild c) der Spannungsbereich umgeschaltet werden soll. Sehr viel vorteilhafter sind dagegen Meßinstrumente mit *Brückengleichrichter* nach Bild d) und e). Sie besitzen außerdem sofort den Vorteil, daß beide Halbwellen gemäß Teilbild f) genutzt werden können. Dadurch hat die schraffierte Fläche die doppelte Höhe im Vergleich zu Bild b), also $I_{eff}\sqrt{8}/\pi \approx 0{,}9003 I_{eff}$. Aber auch hier bleibt das Problem der *nichtlinearen Skale für kleine Spannungen* bestehen (vgl. Bild 5.4.4b)). Es läßt sich weitgehend mit einer Schaltung gemäß Teilbild g) reduzieren. Der *Shunt-Widerstand* R_s muß dann so gewählt werden, daß durch ihn bei Vollausschlag etwa der 5- bis 10fache Strom, verglichen mit dem Instrument, fließt. Dann ergibt sich eine Skale, wie sie Teilbild h) zeigt, und sie ist nahezu gleich, unabhängig davon, ob das Instrument mit verschiedenen Vorwiderständen oder Parallelwiderständen für unterschiedliche Strom- und Spannungsbereiche verwendet werden soll.

Damit dürften auch die Schaltungen von *zwei typischen Vielfachinstrumenten* gemäß Bild 7.7.8i) und j) leicht zu erklären sein. Zunächst sei in der Teilschaltung i) die Instrumentenanschaltung zwischen den Punkten A und B sowie die gezeichnete Schaltung für Wechselstrom betrachtet. Dann entsprechen die sechs Stromparallelwiderstände (Shunts) R_s in Bild g), und R_2 ist ein interner Vorwiderstand zum Gleichrichter, der dessen Kennlinie gemäß Teilbild c) etwas linearisiert. An + und den Spannungsklemmen können mit den fünf Vorwiderständen die Spannungsbereiche 6 bis 600 V erreicht werden. An den Stromwiderständen und + können die Strombereiche von 3 mA bis 6 A realisiert werden. Wird der dreifache Schalter auf Gleichstrom umgeschaltet, so entfällt der Gleichrichter, und R_1 bewirkt, daß der Endausschlag bei den Gleichstrombereichen gleich dem der Wechselstrombereiche ist.

Die Schaltung des Instrumentes von Bild 7.7.8j) unterscheidet sich von der zuvor behandelten in mehreren Punkten. Zunächst verwendet sie eine Gleichrichtung gemäß Bild 7.7.8d). Ferner ist die Schaltung so ausgelegt, daß bei Gleichwerten die volle Empfindlichkeit des Meßwerkes ausgenutzt wird. Dadurch entsteht bei Gleichstrom-

Bild 7.7.8. Wechselstromschaltungen und Vielfachinstrument

a) Triviale Wechselstromschaltung
b) Die positive Halbwelle wird in das flächengleiche Rechteck überführt
c) Zur Erklärung der unterschiedlichen Skalen-Nichtlinearität bei unterschiedlichem ohmschen Vorwiderstand R_V. Die Nichtlinearität der Diode gemäß $R = f(U)$ wird um so weniger wirksam, je größer R_V im Vergleich dazu ist
d) Brückengleichrichter mit zwei Widerständen, die zugleich die Linearität verbessern — im Vergleich zur
e) Brückenschaltung für Meßgeräte
f) Die beiden Halbwellen werden in das flächengleiche gestrichelte Rechteck umgewandelt. Es ist doppelt so hoch als bei Einweggleichrichtern gemäß b)
g) Vor- und Serienwiderstand zur Realisierung einer gleichen Skale für mehrere Meßbereiche
h) Typische Wechselstromskale von Gleichrichterinstrumenten
i) Schaltung eines der ersten Vielfachinstrumente vom Typ Multavi II
j) Schaltung eines modernen Vielfachinstrumentes, bei dem jeweils in den Gleich- und Wechselstrombereichen eine bestmögliche Empfindlichkeit erreicht wird und bei dem die Strom- und Spannungsbereiche voll entkoppelt sind
k) Prinzipschaltung für ein Instrument mit echter Spitzenwertanzeige

und Gleichspannungsmessungen ein erheblich geringerer Rückwirkungsfehler. Andererseits ist aber der Aufwand an Shunts und Vorwiderständen auch doppelt so groß. Der Abgleich aller Gleichstrombereiche erfolgt dabei mit dem einmalig vom Werk eingestellten Widerstand $R_{\text{abgl.}}$. Bei Wechselstrom dient die Kombination C_u sowohl zum Abgleich als auch zur teilweisen Korrektur des Frequenzganges der Gleichrichteranordnung.

Gemäß den Betrachtungen im Zusammenhang mit Bild 7.5.13 ist es möglich, auch *echte Effektivwerte* zu messen, indem eine entsprechende *quadratische Kennlinie* geschaffen wird.

Die Messung von *Spitzenwerten* ist jedoch wegen des stets endlichen *Stromflußwinkels* Θ (Bild 5.4.4) sehr schwierig. Sie läßt sich mit einer Schaltung gemäß Bild 7.7.8k) erreichen, indem der erste Verstärker einen negativen Widerstand besitzt, der den Durchlaßwiderstand der Diode aufhebt, und der folgende Verstärker einen sehr hohen Eingangswiderstand hat.

Schließlich sei noch darauf hingewiesen, wie vorteilhaft bezüglich einer exakt linearen Skale die *Gleichrichterschaltungen mit Operationsverstärkern* gemäß Bild 7.5.16 sind.

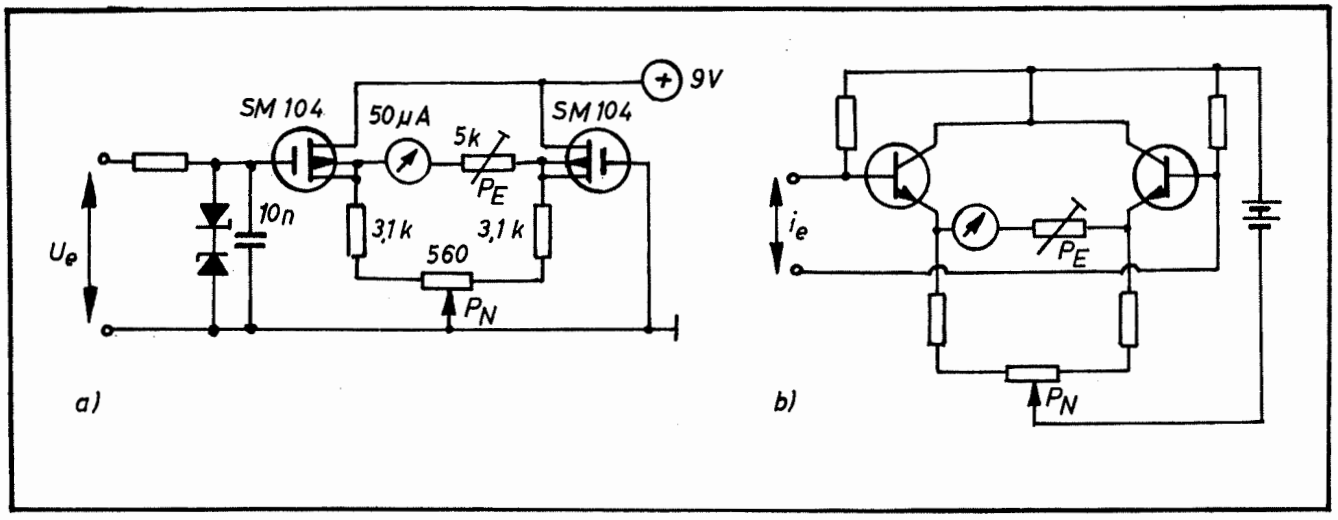


Bild 7.7.9. Prinzipieller Aufbau von zwei typischen einfachen Gleichwertmeßverstärkern, die oft unmittelbar in Instrumente eingebaut werden, a) mit MOS-FET, b) mit bipolaren Transistoren. Mittels des Potentiometers P_N wird der Nullpunkt und mittels P_E die Empfindlichkeit abgeglichen

7.7.2.5. Meßverstärker

Infolge der Kleinheit und des geringen Stromverbrauchs der Transistoren ist es jetzt üblich, in viele Meßinstrumente, ja zuweilen sogar in Vielfachmesser, Verstärker zur Verbesserung der Eigenschaften oder zur Erweiterung der Einsatzmöglichkeiten einzubauen. Hierbei sind Gleichstrom- bzw. Wechselstromverstärker und Verstärker im Sinne von Meßschaltungen zu unterscheiden.

In Bild 7.7.9 sind zwei häufig benutzte Schaltungen für Gleichspannungs- bzw. -stromwerte gezeigt. Beide Schaltungen erreichen als vereinfachte Differenzverstärker eine hohe Stabilität gegenüber Temperatur- und Betriebsspannungseinflüssen. Die Schaltung a) zeichnet sich noch durch eine zusätzliche Gegenkopplung aus. Sie ist wegen ihres sehr hohen Eingangswiderstandes vor allem für Spannungsmessungen ideal. Die beiden in Reihe geschalteten Z-Dioden bewirken einen Schutz des Gate für zu hohe Spannungen. In der Schaltung b) wird die Stromverstärkung der Transistoren ausgenutzt. Sie führt also zu einem guten Strommeßgerät.

Tabelle 7.7.1. Zusammenstellung wichtiger Analyseverfahren für Substanzen

Methode	Signalquellen	Veränderliche Größe	Meßsignal	Empfänger
Hochfrequenzspektroskopie a) Kernresonanz (-spin) b) Elektronenresonanz (-spin)	Magnetfeld 0 ... 5 T Hochfrequenz a) 10^5 ... 10^9 Hz b) 10^8 ... 10^{12} Hz	meist Magnetfeld zuweilen Frequenz	HF-Amplitude	a) Spule b) Mikrowellen-resonator
Optische Spektroskopie • Absorption • RAMANN-Streustrahlung • Emission anderer Wellenlänge • Fluoreszenz	Licht verschiedener Wellenlänge 10^{-4} ... 10^{-7} m • Ultraviolett • sichtbar • Infrarot	meist Frequenz zuweilen Ort (z. B. hinter Prisma)	Strahlungsintensität	SEV Fotohalbleiter Bolometer Fotoplatte
Röntgenspektroskopie a) R.-Fluoreszenz b) R.-Primäranalyse c) R.-Kristallstruktur d) Mikroanalyse (-sonde)	a), c) Röntgenstrahlung ca. 10^{-10} m 10^4 ... 10^5 eV b), d) Elektronenstrahl a), b), d) Emission charakteristischer Strahlung c) Beugung der Strahlung	a), b) Wellenlänge c) Winkel d) Ort	Röntgenquantenzahl	Szintillationszähler Zählrohr Halbleiterdetektor
Elektronenspektroskopie (-strahl) a) LEED-Spektroskopie b) AUGER-Spektroskopie c) ESCA-Spektroskopie d) Elektronenmikroskop	a), b) Elektronenstrahl 10^2 ... 10^3 eV d) Elektronenstrahl 10^3 ... 10^6 eV c) Elektromagnet. Welle 10^{-6} ... 10^{-10} m	a), b), c) Elektronenenergie d) Ort (als Raster oder parallel)	Elektronenstrahl	SEV Bildschirm Fotoplatte
Chromatographie in Gas oder Flüssigkeit (auf Papier)	Substanzimpuls (chemische Substanz)	Laufzeit (Ort)	Konzentration (Färbung)	Wärmeleitzelle Ionisationsdetektor
Elektrochemische Verfahren a) pH-Wert b) Polarographie c) Coulometrie	a) entsteht selbst b), c) elektr. Feld 0 ... 2 V	a), c) Zeit, Ort b) Potential	a) Potential b) Strom (elektr.) c) Ladung	a) Meßelektrode b), c) Meßzelle
Massenspektroskopie	Elektronen zur Ionisierung ca. 100 eV	Magnetfeld, HF-Feld Laufzeit	Ionenstrom	SEV
MÖSSBAUER-Spektroskopie (Absorption charakteristischer Strahlung)	γ-Quelle 10^5 ... 10^6 eV mechanische Geschwindigkeit 10^{-4} ... 10^{-1} m/s	mechanische Geschwindigkeit	Röntgenquantenzahl	SEV

7.7.3. Ausgewählte komplexe Meßschaltungen

7.7.3.1. Allgemeines Meßprinzip

Einer Vielzahl von Meßmethoden liegt das Grundprinzip von Bild 7.7.10 zugrunde. Einem zu messenden Objekt, z.B. Verstärker, Schaltung, Gerät, physikalische oder chemische Substanz oder biologisches Objekt, wird ein Signal aufgeprägt. Die elektrische (oder andere) Größe der Signalquelle kann u. a. eine Sinusschwingung, Impulsfolge, ein Einzelimpuls oder Rauschen sein. In vielen Fällen wird dieses Eingangssignal noch in definierter Weise zeitlich bzw. örtlich verändert. In einigen Fällen wird das Meßobjekt auch mit mehreren Eingangssignalen gleichzeitig beaufschlagt, z.B. bei der Hochfrequenzspektroskopie mit einem Hochfrequenzfeld und einem Gleichmagnetfeld. Unter dem Signaleinfluß können am Meßobjekt bestimmte Ausgangssignale gewonnen werden. Sie sind meist von Störungen vielfältiger Art überlagert. Die folgende *Meßschaltung* hat dann vor allem *zwei Aufgaben*:

- das Signal in eine so auswertbare Form zu bringen, daß möglichst unmittelbar die interessierenden Eigenschaften des Meßobjektes zu erkennen sind;
- das Signal möglichst weitgehend von den Störungen zu befreien.

Es kann hier keine vollständige Liste aller nach diesem Prinzip realisierten oder realisierbaren Methoden aufgezeigt werden. In den Tab. 7.7.1 und 7.7.2 seien jedoch zwei ausgewählte Gebiete betrachtet. Es sei darauf hingewiesen, daß in gewissem Sinne hier ganz ähnliche Verhältnisse wie bei der Modulation gemäß Bild 7.2.1 vorliegen.

Tabelle 7.7.2. Überblick zu wichtigen Meßprinzipien der Elektronik

Darstellungsart	Veränderung	Beispiele	Abschnitt
Zeitabläufe einer Schwingung	z. B. Spannung	Registrierung Oszillograph Meßwertspeicher	7.7.4., 6.2.3.
Einschwingvorgang	z. B. Spannung	Impulstechnik Übertragung von Rechteckschwingungen	1.5.2., 2.2.3.
Frequenzgang Amplitude und/oder Phase	$f(\omega)$	Wobbeltechnik Sweepfrequenzsignal	—
Spektrum	Frequenz	Suchtonverfahren	5.4.9.
Kennlinien z. B. von Röhren, Transistoren Hysteresiskurven	$y = f(x)$	Oszillograph Plotter xy-Schreiber	7.7.4., 6.2.3. 6.3.3.6.
Laufzeit	Zeit	Radar	7.7.3.8.
Auto- bzw. Kreuzkorrelation	Laufzeit	Korrelatoren	7.7.3.6.
Wahrscheinlichkeitsdichte	Amplitudenschwelle	Statistische Untersuchungen	—

Für die experimentellen Wissenschaften hat die Untersuchung von Stoffen, z. B. des Anteils von bestimmten Substanzen oder ihrer Struktur, große Bedeutung. Deshalb hat sich eine Vielzahl derartiger Methoden gebildet und zu speziellen, oft unikalen Geräten geführt. Einen Überblick zu den wohl wichtigsten Methoden gibt Tab. 7.7.1. Sie zeigt, ohne vollständig zu sein, schon eine beträchtliche Vielfalt.

In der elektrischen und elektronischen Meßtechnik besitzen u. a. die in Tab. 7.7.2 aufgezeigten Verfahren besondere Bedeutung. In gewissen Fällen werden diese Verfahren auch als Teilverfahren z. B. in der Meßschaltung von Bild 7.7.10 eingesetzt. Einige von ihnen sind bereits besprochen, andere werden noch behandelt.

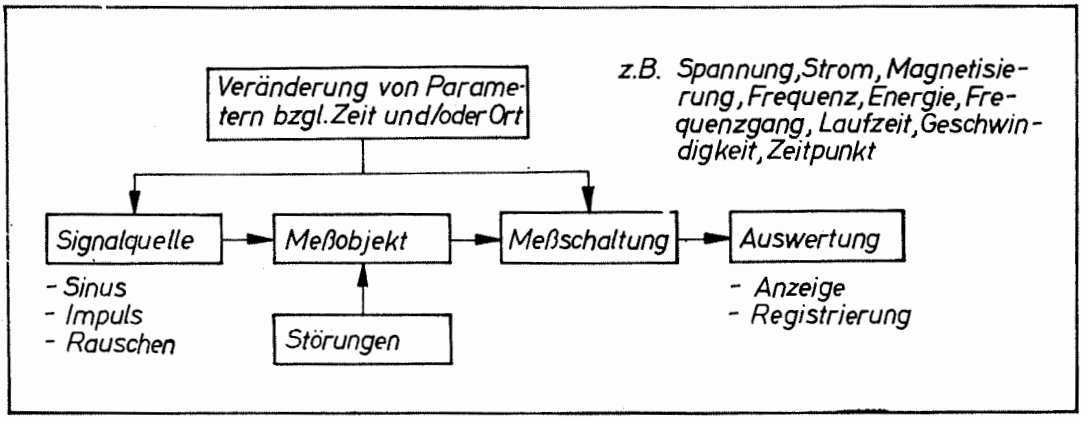


Bild 7.7.10. Grundprinzip vieler Meßmethoden zur Untersuchung von Objekten

Weiter sei darauf hingewiesen, daß diese Betrachtungsweise auch durchaus für biologische, medizinische, physiologische, psychologische u. a. Untersuchungsmethoden geeignet ist. Als Eingangssignale dienen dabei u. a. körperliche Belastung oder Leistung, Licht- oder Schallsignale, mit semantischem Gehalt belegte Bilder oder Werte beziehungsweise zu lösende Aufgaben. Als Ausgangssignale werden z. B. Hautwiderstand, Blutdruck, Herzfrequenz gemessen bzw. EKG, EEG, Myogramm oder Pupillenreflex abgeleitet. Als Beispiel sei nur die objektive Audiometrie beschrieben. Dem Ohr wird dabei ein Schallsignal bestimmter Frequenz mit definiert zunehmender Amplitude zugeführt. Von der Kopfhaut wird in der Nähe der primären Hörrinde des Gehirns das EEG abgeleitet. Sobald die Hörschwelle überschritten wird, entsteht ein akustisch evoziertes Potential. Es ist allerdings sehr stark vom Störpegel verdeckt und muß durch komplizierte Maßnahmen, auf die noch einzugehen ist, herausgefiltert werden.

Die *Filterung eines Nutzsignals* aus dem Störpegel ist heute für eine sehr große Anzahl von Meßprinzipien von entscheidender Bedeutung. Daher sind auch eine Vielzahl von Methoden entwickelt worden. Sie können auf die unterschiedlichste Art eingeteilt werden. Hier können nur einige besonders wichtige Methoden der Störbefreiung behandelt werden. Ihre Einteilung soll nach Signaleigenschaften erfolgen, die bevorzugt das Ausgangs- und teilweise auch das Eingangssignal betreffen.

7.7.3.2. Frequenzbereichfilterung

Es liege ein sich kontinuierlich veränderndes Ausgangssignal vor. Der Sonderfall, daß Störung und Nutzsignal weitgehend unterschiedliche Frequenzbereiche einnehmen, ist trivial. Denn dann sind sie mit einem relativ einfachen Filter zu trennen. Nehmen beide, was meist vorliegt, nahezu den gleichen Frequenzbereich ein, so zeigt Bild 7.7.11 die Zusammenhänge. Die Filtereigenschaften sind dann optimal, wenn das Filter je Frequenzbereich so sperrt, daß das Verhältnis von Störung zu Nutzsignal besonders günstig ist. So entsteht eine Filterfunktion, die auch Teile des Nutzsignals sperrt und selbst zu Meßfehlern führt. Die *Wahl eines* derartigen *Filters* kann also nur nach dem Kriterium der Meßfehler erfolgen, und zwar der Summe von

— Fehlern durch Störungen,
— Fehlern durch Frequenzbandbeschneidung.

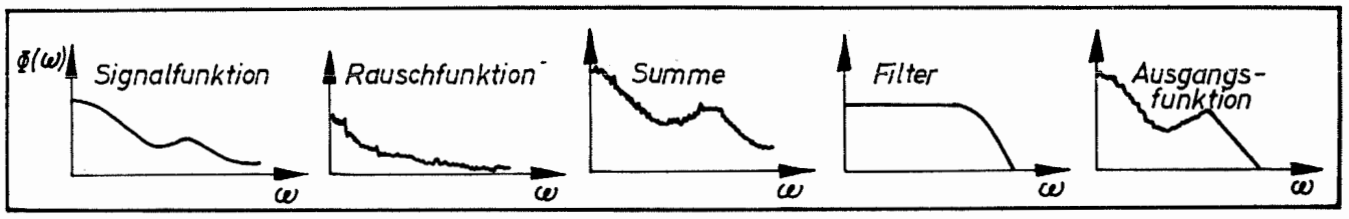


Bild 7.7.11. Zur Wirkung einer Frequenzfilterung auf das Spektrum der Signalfunktion

7.7.3.3. Beispiele aus der Audiotechnik

Bisher wurde davon gesprochen, daß eine Störbefreiung am Ausgangssignal erfolgt. Aber bereits einleitend wurde in diesem Abschnitt auf den Zusammenhang mit der Modulation hingewiesen. Dies bedeutet, daß im Prinzip auch das Eingangssignal günstig gewählt werden kann. In der Meßtechnik sind dem natürlich Grenzen durch Eigenschaften des Meßobjektes bzw. seine zulässige Belastbarkeit gegeben. Sie liegen gewiß anders als bei der Modulation, dennoch lassen sich hier zuweilen beachtliche Verbesserungen erreichen. Dies sei an Beispielen aus der Audiotechnik gezeigt. Für Sprache, noch mehr aber für Musik sind die hohen Frequenzen sehr wichtig. Zugleich besitzen sie eine relativ geringe Intensität. Dies wird sowohl bei der UKW- als auch bei der Magnetband- und Schallplattentechnik dazu benutzt, ihre Intensität vor der Übertragung bzw. bei der Aufzeichnung in optimaler, d. h. für alle drei Fälle in unterschiedlicher, aber genormter Weise anzuheben. Bei der Wiedergabe kann dann eine entsprechende Absenkung mit Störbefreiung erfolgen. Doch damit haben sich die Techniker nicht begnügt. Sie haben die Technik auch noch dynamisch gestaltet. Da die hohen Signale bzw. der Pegel generell sehr stark schwankt, sind Verfahren entwickelt worden, die auch bezüglich dieser Änderungen ständig optimale Verhältnisse schaffen. Das Signal selbst steuert dann also zusätzlich die Aufzeichnungsverhältnisse bei der Aufzeichnung und Wiedergabe. Hier können nur noch einige Verfahren, wie Dolby, DNL und High-Com genannt werden.

7.7.3.4. Trägerfrequenzverfahren und phasenempfindlicher Gleichrichter

In erster Näherung ist das Rauschen der Störquelle proportional der Bandbreite des Meßkanals (s. 5.7.):

$$U_{\mathrm{R}} \sim B\,. \tag{15}$$

Die Informationsmenge bei einer Messung ergibt sich andererseits proportional zur Meßzeit und ebenfalls zur Bandbreite:

$$I_{\mathrm{M}} \sim TB\,. \tag{16}$$

Diese Aussage entspricht auch der KÜPFMÜLLER-Beziehung Gl. (1.4.15). Aus den beiden letzten Gleichungen läßt sich folgern, daß durch eine verlängerte Meßzeit auf Kosten der Bandbreite bessere Meßwerte erhalten werden. Hierzu muß das Eingangssignal aber entsprechend langsamer verändert werden. Damit gelangt der Meßvorgang in den Bereich der tiefen Frequenzen, wo nicht mehr Gl. (15) gilt, sondern ein Anstieg der Störungen wirksam wird. Dies zeigt z.B. Bild 5.7.6. Die Verhältnisse werden sofort günstiger, wenn die Signale in einen Trägerfrequenzbereich verschoben werden. Hierzu dienen Anordnungen, wie sie bereits im Abschnitt 5.4.9. behandelt wurden. Mit derartigen Anordnungen ist auch relativ leicht das Verfahren des phasenempfindlichen Gleichrichters (Lock-in-Schaltung) zu koppeln (vgl. Bild 5.4.14). Dann sind praktisch exakt die oben genannten Voraussetzungen erfüllt, und außerdem sind auch sehr leicht extrem schmalbandige Filter zu realisieren. Über die Verlängerung der Meßzeit ist so also im Prinzip eine fast ideale Störbefreiung zu erreichen. Natürlich bestehen für die Länge der Meßzeit praktische Grenzen.

7.7.3.5. Periodische Eingangssignale und Akkumulation

Bereits bei den Trägerfrequenzverfahren liegt im gewissen Sinne ein periodisches Signal vor, nämlich die Trägerfrequenz. Das eigentliche Signal verändert sich hier jedoch noch meist einfach zeitlich linear. Es gibt jedoch auch Verfahren, bei denen die Signalanregung periodisch oder zu definierten Zeitpunkten immer wieder wiederholt wird. Dann entsteht die Möglichkeit zu einem neuen Prinzip, das mit mehreren praktisch gleichwertigen Namen belegt ist: Signal-Averaging, Signalmittelungstechnik, Akkumulation, Zeitbereichsfilterung und Abtastverfahren. Ihr Prinzip beruht darauf, daß die sich wiederholenden Ausgangssignale selbst einen definierten Zeitablauf zeigen, während die überlagerten Störungen statistisch sind. Sie nutzen also die Gesetze der Fehlerfortpflanzung bzw. die Leistungsaddition bei nichtkohärentem Rauschen aus (vgl. 5.7.). Deshalb vergrößern sich bei einer n-fachen Überlagerung (Addition) der Signale das Nutzsignal mit dem Faktor n, die Störung dagegen nur

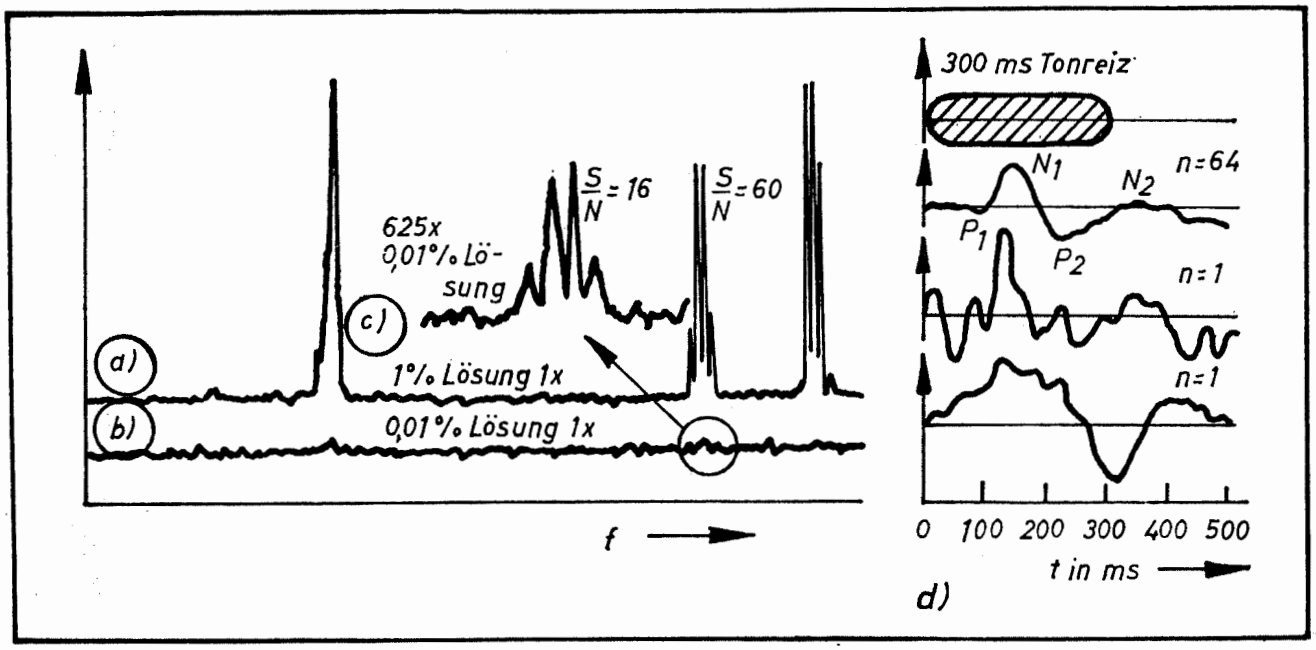


Bild 7.7.12. Beispiele zur Akkumulation von stark gestörten Signalen

a) bis c) Kernresonanzspektrum von Äthylbenzol
a) 1%-Lösung
b) 0,01%-Lösung ohne Akkumulation
c) 0,01%-Lösung als Ausschnitt mit 625facher Akkumulation
d) Akkumulation bei der objektiven Audiometrie; von oben nach unten: Hüllkurve des Tonreizes, typisches gemitteltes evoziertes Potential (AEP) und zwei Einzelpotentialverläufe des EEG nach einmaliger Reizung

mit dem Faktor $\sqrt{n}$. Der Störabstand wird also bei n-facher Überlagerung um den Faktor $\sqrt{n}$ verbessert. Um dieses Verfahren ausführen zu können, sind Verzögerungsglieder entsprechend der Periodenlänge oder Speicher für die einzelnen Signalverläufe notwendig. Es sind deshalb auch digitale Prinzipien mit Speichern der Datenverarbeitung und analoge Verfahren mit speziellen Laufzeitketten oder Magnetbandspeichern bekannt. Auch dieses Verfahren gewinnt den Störabstand sichtlich durch die Meßzeit, allerdings mit dem Faktor $\sqrt{T}$. Die Beispiele nach diesem Prinzip (Bild 7.7.12) stammen aus der Kernresonanzspektroskopie und aus der objektiven Audiometrie entsprechend der am Ende von Abschnitt 7.7.3.3. erwähnten Technik.

7.7.3.6. Auto- und Kreuzkorrelation

Ganz kurz wurde auf die Autokorrelationsfunktion (AKF) bereits im Zusammenhang mit Bild 1.4.8 eingegangen. Für die Störbefreiung besitzt sie den Vorteil, auch eine

Mittelwertsbildung zu realisieren. Sie erfolgt jetzt aber über die gesamte Meßzeit und wird durch Laufzeitsteuerung τ in diesem Prozeß verändert. Dabei ist bedeutsam, daß unterschiedliche Signale auch meist verschiedene Autokorrelationsfunktionen aufweisen. Insbesondere das inkohärente Rauschen weist nur in der Nähe von $\tau = 0$ größere Komponenten auf. Deshalb werden alle länger korrelierten Signale insbesondere aber periodische Signale gut von dem Rauschen getrennt. Zu diesem Effekt kommt noch der mögliche Gewinn durch die Meßzeit. Den prinzipiellen Aufbau eines Korrelators zeigt Bild 7.7.13. Hierbei ist er zugleich mit dem Meßobjekt und -prozeß

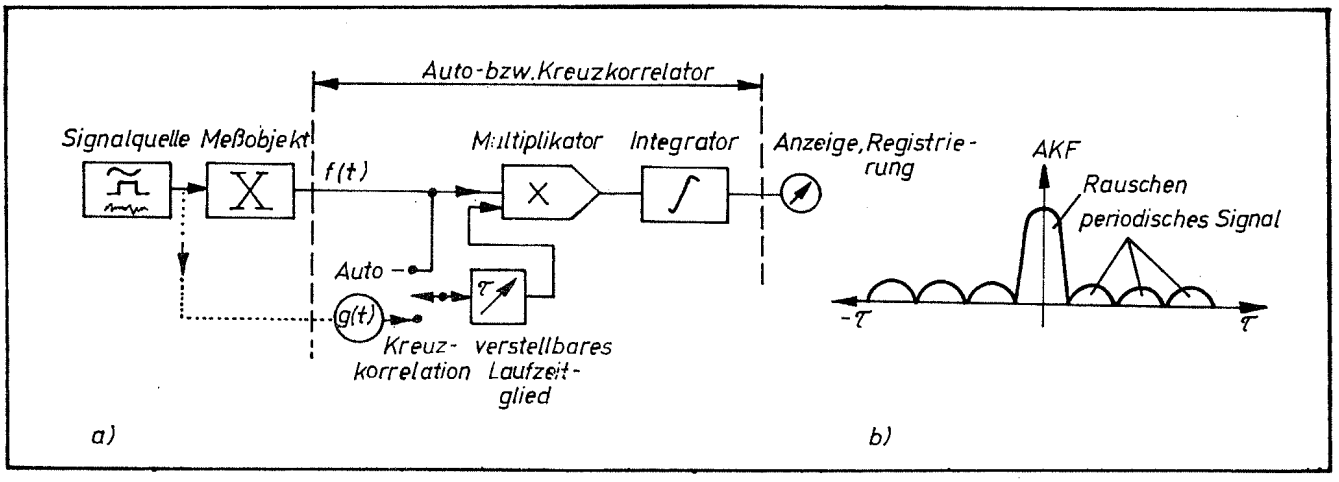


Bild 7.7.13. a) Zum Prinzip der Auto- und Kreuzkorrelation in Verbindung mit der Analysentechnik, b) Verlauf einer speziellen Autokorrelationsfunktion (AKF)

Tabelle 7.7.3. Gegenüberstellung von phasenempfindlichem Gleichrichter und Autokorrelator

Größe, Struktur	Phasenempfindlicher Gleichrichter	Autokorrelator
Eingangssignal	streng periodisch	beliebig häufig breitbandig auch Rauschen
Hilfszweig	einstellbare Phase	zeitlich veränderliche Laufzeit
Mischglied bildet	Frequenzdifferenz	Multiplikation
Siebglied	Tiefpaß	Integrator
Unabhängig veränderliche Größe	Eingangssignal	Laufzeit

gekoppelt gezeigt. Insbesondere mittels der punktiert eingezeichneten Linie läßt sich ein Vergleich zum phasenempfindlichen Gleichrichter gemäß Bild 5.4.14 anstellen. Die sich dabei ergebenden wesentlichen Unterschiede weist Tab. 7.7.3 aus. Ansonsten sind beide Strukturen fast gleich. Neben der Autokorrelationsfunktion

$$\mathrm{AKF} = \frac{1}{2T} \int_{-T}^{+T} f(t) \cdot f(t + \tau)\, \mathrm{d}t \tag{17}$$

hat zuweilen auch die Kreuzkorrelation

$$\mathrm{KKF} = \frac{1}{2T} \int_{-T}^{+T} f(t) \cdot g(t + \tau)\, \mathrm{d}t \tag{18}$$

Bedeutung. Sie besitzt etwa zwei Vorteile, nämlich einmal ist bei passender Wahl von $g(t)$ die Laufzeitverstellung leichter zu realisieren, und zum anderen kann $g(t)$ störfrei gehalten werden. Entsprechend der punktierten Linie kann als $g(t)$ auch das Eingangssignal verwendet werden.

Ein Hauptproblem der Korrelationselektronik ist das verstellbare Laufzeitglied. Bis vor einiger Zeit war ein spezielles Zweispurmagnetbandgerät noch oft die beste Lösung. In der letzten Zeit haben sich aber Halbleiterschaltungen entwickelt, die meist weitaus bessere Lösungen zulassen. Auf sie wird noch eingegangen werden. Andere effektive Möglichkeiten bietet natürlich die Rechentechnik.

7.7.3.7. Phasenregelkreis (PLL)

Eine besondere Schaltung, die im gewissen Umfang zur Störbefreiung genutzt wird, aber auch andere vielfältige Vorteile hat, ist der Phasenregelkreis, engl. phase-locked loop = PLL (in der Phase eingerastete Schleife). Obwohl das Prinzip ziemlich alt ist (erste Veröffentlichung 1922, erste realisierte Schaltung 1932), hat es erst in den letzten Jahren mit der Entwicklung der integrierten Schaltkreise größere Bedeutung erlangt. Den Aufbau zeigt Bild 7.7.14a). Einer phasenempfindlichen Schaltung, z.B. einem Ringmodulator gemäß Abschnitt 5.4.9 wird ein Signal ω zugeführt. Ferner erhält er

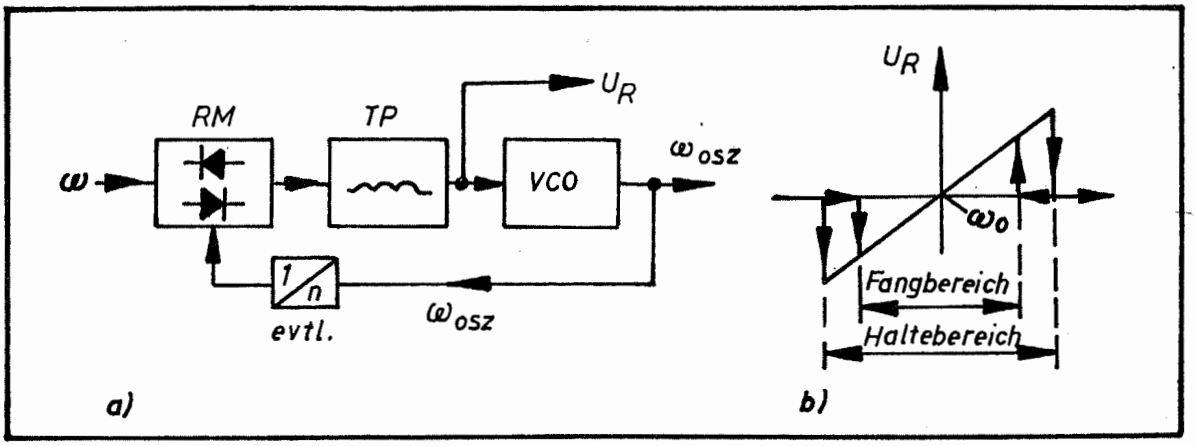


Bild 7.7.14. Prinzip des Phasenregelkreises (PLL)

a) Strukturbild; es bedeuten RM Ringmodulator (Phasendiskriminator), TP Tiefpaß, VCO voltage-controlled oszillator = spannungsgesteuerter Oszillator variabler Frequenz

b) Verhalten der Regelspannung U_R in Abhängigkeit von der Eingangsfrequenz

Auf der schrägen Geraden ist der Oszillator eingerastet. Bei Erreichen des Fangbereiches erfolgt die Einrastung, während bei Überschreiten des Fangbereiches im eingerasteten Zustand der eingerastete Zustand im Haltebereich stabil ist

seine zweite Spannung von einem in der Frequenz steuerbaren Oszillator (voltage-controlled oscillator = VCO), der z.B. gemäß Bild 7.3.23 aufgebaut sein könnte und ohne Eingangsspannung, also freilaufend ω_0 erzeugt. Der Tiefpaß läßt nur Differenzfrequenzen $|\omega - \omega_0|$ bis zu einer bestimmten Größe passieren. Hierbei tritt, wenn ω_0 und ω hinreichend nahe sind, eine ständige Frequenzänderung des VCO auf. Sobald einmal diese Frequenz ω_{osz} gleich ω ist, rastet der Oszillator (meist mit Einschwingvorgängen) auf ω ein. Die Regelspannung ist dann ein Maß für die Abweichung des Oszillators von der Leerlauffrequenz ω_0 und auch für die Phasendifferenz zwischen ω_{osz} und ω. Im Betrieb liegt dann eine Kurve vor, wie sie Bild 7.7.14b) zeigt. Es sind der Fang- und Haltebereich zu unterscheiden. Die Eigenschaften der Schaltung hängen u. a. von der Kreisverstärkung und der Grenzfrequenz des Tiefpasses ab. Je nach dem Anwendungszweck wird entweder U_R oder ω_{osz} weiterverwendet. In einigen Fällen wird gemäß Bild 7.7.14 noch ein Frequenzteiler $1:n$ eingeschaltet. Der Ringmodulator kann auch eine Multiplikationsschaltung sein.

7.7.3.8. Impulsverdichtung

Bei der Korrelationstechnik sind beliebige Signale zugelassen. Damit ist der Übergang zu Impuls- und Breitbandsignalen erfolgt. Hier entwickelte sich in dem letzten Jahrzehnt eine bedeutsame neue Meßtechnik. Hierbei wird das Meßobjekt nicht mit einer Frequenz ständig belastet, sondern einem Spektrum ausgesetzt und/oder nur impulsbelastet. Dadurch ergeben sich vielfältige Möglichkeiten und Vorteile, z.B. paralleler Empfang mehrerer (n) Ausgangssignale in verschiedenen Frequenzbereichen, wodurch sich eine erhebliche Verkürzung der Meßzeit auf $1/n$ verwirklichen läßt. Hier soll jedoch nur ein Beispiel der Impulstechnik angedeutet werden. Bei Impulsanregung liegt der wesentliche Teil der Nutzenergie in den ersten Abschnitten der Impulsantwort. Ein bekanntes Verfahren ist daher, die Dauer der Impulsantwort mit einer e-Funktion zu bewerten. Hierdurch kann eine erhebliche Störbefreiung mit geringer Signalverfälschung erreicht werden. Eine noch effektivere Methode besteht aber darin, dem Meßobjekt einen relativ langen Impuls zuzuführen und ihn später zeitlich zu verdichten. Ein besonders anschauliches Beispiel bietet sich aus der Frequenzmodulation. Das Eingangssignal zeige einen Verlauf wie in Bild 7.7.15a). Der Impuls-

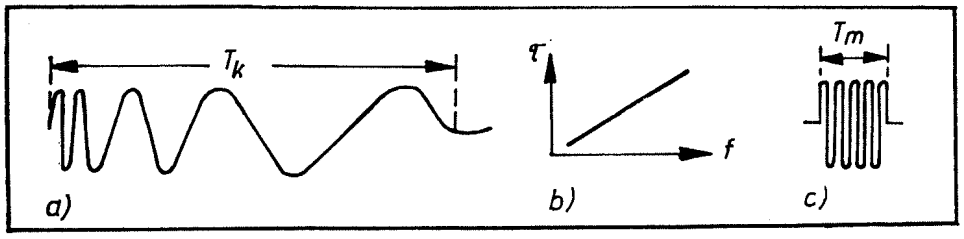


Bild 7.7.15. Anschauliches Prinzip zur Impulsverdichtung. Das Eingangssignal ist eine speziell frequenzmodulierte Welle (a). Sie wird an einem Filter, dessen Laufzeit τ mit der Frequenz zunimmt (b), zur Schwingung (c) verdichtet

verdichter muß dann eine Abhängigkeit der Laufzeit von der Frequenz wie in Teilbild b) besitzen. Dann benötigen die anfangs gesendeten hohen Frequenzen mehr Zeit als die später gesendeten Tiefen. So entsteht das Ausgangssignal von Bild c). Der Störabstand verbessert sich etwa um

$$\eta = T_k/T_m \,. \tag{19}$$

Weitaus wirksamere Verfahren dieser Art lassen sich natürlich in der Digitaltechnik realisieren.

Aus den vorangegangenen Fakten wird deutlich, daß auch der Gestalt des Eingangssignals beachtliche Bedeutung zukommt. Wird nur ein gegebenes Eingangssignal verwendet und soll das Optimum mit einem nur nachgeschalteten Filter realisiert werden, so liegt die *klassische Filtertheorie* vor. Wird dagegen auch das optimale Eingangssignal mitbestimmt, so liegt das Gebiet der Signalsynthese vor. Es kann hier nicht weiter behandelt werden.

7.7.4. Oszillographentechnik

Die Oszillographenröhre ist in den Abschnitten 6.2.9., 6.2.10. und 6.2.17. behandelt. Während ein Meßinstrument auf einer Skale einen analogen Wert anzeigt, kann bei ihr der Leuchtpunkt in zwei Richtungen — x und y — abgelenkt werden. Dies ermöglicht weitaus komplexere Darstellungen, die insbesondere für Kurvendarstellungen herangezogen werden. Da darüber hinaus mit der Oszillographenröhre nur Genauigkeiten

von ca. 2 bis 20% erreichbar sind, gilt im Vergleich zum Meßinstrument die folgende Aussage:

Ein Oszillograph ist besonders für die Darstellung von ganzheitlichen, flächigen Kurven geeignet, welche Einblick in den zeitlichen oder funktionellen Zusammenhang von Größen ermöglichen, während ein Meßinstrument daraufhin ausgelegt ist, möglichst exakte Zahlenwerte von Messungen anzuzeigen.

Beim Oszillographen sind zwei recht unterschiedliche Betriebsweisen zu unterscheiden. Werden den beiden Ablenkplattenpaaren zwei Spannungsverläufe zugeführt, zwischen denen der Zusammenhang interessiert, so liegt eine echte *zweifache Ablenkung* vor (7.7.4.1. und 7.7.4.2.).

Im zweiten Fall wird dem X-Plattenpaar eine *zeitlineare Sägezahnspannung* zugeführt. Hierdurch kann der *zeitliche Verlauf* der an die *Y-Platten* angelegten Meßspannung dargestellt werden (7.7.4.3. bus 7.7.4.5.). Dies ist die häufigste Anwendung des Oszillographen.

Einige Spezialfälle sind noch in Abschnitt 7.7.4.6. zu behandeln.

7.7.4.1. Ellipsenauswertung

Gemäß Bild 7.7.16a) werde ein linearer Vierpol mit dem Oszillographen zusammengeschaltet. Am Eingang des Vierpols und an den Y-Platten liege die Spannung

$$u_{\sim} = \hat{U}_{01} \cos \omega t \,. \tag{20}$$

Am Ausgang des Vierpols und damit an den X-Platten liegt eine Spannung, die um den Phasenwinkel φ verschoben und auf die Amplitude $\hat{U}_{02}$ verstärkt bzw. abgeschwächt ist:

$$u_x = \hat{U}_{02} \cos (\omega t + \psi) \,. \tag{21}$$

Als Folge erscheint auf dem Oszillographenschirm eine Ellipse, wie sie z.B. Bild 7.7.16b) zeigt. Aus dieser Ellipse sind relativ leicht das *Verhältnis der Amplituden*, also die Verstärkung bzw. Dämpfung, und der *Phasenwinkel* zu bestimmen. Besonders günstig dafür ist die *Fotografie*. Hierbei werden zunächst die Achsen ge-

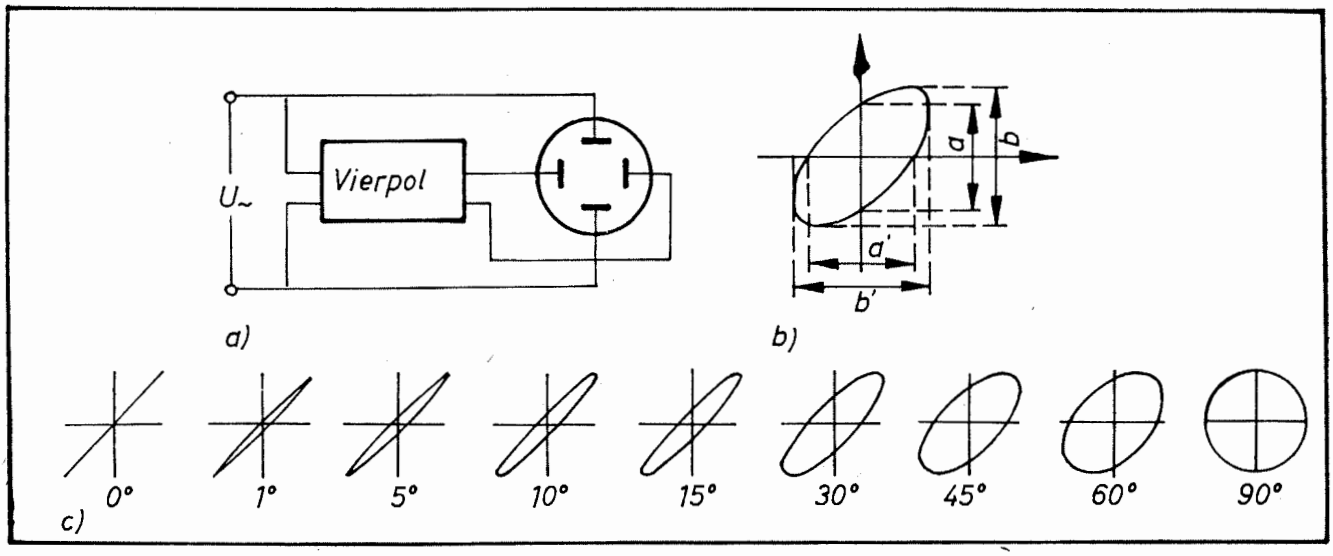


Bild 7.7.16. a) Schaltung und b) Ellipsenbeispiel zur Bestimmung von Phasenverschiebung und Übertragungsmaß eines Vierpols
c) Ellipsenbeispiele zur groben Bestimmung des Phasenwinkels. Dabei ist vorausgesetzt, daß die X- und die Y-Ablenkung genau gleich groß sind

schrieben, indem jeweils eine der beiden Spannungen zum Oszillographen hin unterbrochen wird. Die Achsen nehmen dabei genau die Längen b bzw. b' an. Werden beide Spannungen zur dritten Belichtung eingeschaltet, so entsteht die Ellipse. Ihre Schnittpunkte mit den Achsen führen zu den Werten a und a'. Für den Phasenwinkel gilt

$$\sin \varphi = \frac{a}{b} = \frac{a'}{b'}. \tag{22}$$

Damit wird der Winkel auf einen Bereich von 0° bis 90° eingegrenzt. Größere Werte müssen durch zusätzliche Methoden, z.B. Umlaufrichtung der Ellipse, bestimmt werden. Für den Fall, daß beide Ablenkungen gleich groß gewählt werden (z.B. mittels eines Verstärkers), gibt Bild 7.7.16c) einige Ellipsenbeispiele mit den zugehörigen Winkeln wieder. Die Genauigkeit ist also um 0° besonders groß und in der Nähe von 90° besonders gering.

Für die Bestimmung des Amplitudenverhältnisses ist noch zusätzlich das Verhältnis beider Ablenkempfindlichkeiten zu berücksichtigen. Es sei gleich A, dann gilt

$$\frac{\hat{U}_{01}}{\hat{U}_{02}} = A \frac{b}{b'} = A \frac{a}{a'}. \tag{23}$$

Die Ellipsenmessung ist nicht unbedingt auf die Vierpolmessung beschränkt. Mit ihr können zwei beliebige Spannungen gleicher Frequenz bezüglich ihrer Phase und ihres Amplitudenverhältnisses bestimmt werden.

7.7.4.2. Untersuchung von Nichtlinearitäten

Es sei jetzt für Bild 7.7.16a) ein linearer Vierpol ohne Phasenfehler angenommen, dann entsteht auf dem Leuchtschirm eine geneigte Gerade. Besitzt der Vierpol Nichtlinearitäten, so entsteht eine entsprechend gekrümmte Linie. Aus der Abweichung von der Geraden kann auf die Verzerrungen geschlossen werden. Wird anstelle des Vierpols ein Transistor eingefügt, so erscheint auf dem Schirm direkt dessen Kennlinie. Durch schnelles Umschalten von Parametern kann sogar eine vollständige *Kennlinienschar* dargestellt werden. Die entsprechende Prinzipschaltung für einen FET zeigt Bild 7.7.17a). Bei kleinem R_a werden die *statischen Kennlinien* z.B. gemäß Bild 5.5.1 geschrieben. Bei großem R_a können sofort die Arbeitskennlinien erhalten werden (Bild 5.6.3).

Ein anderes Beispiel ist die Darstellung der *Hysteresekurve* mit der Schaltung von Bild 7.7.17b). Mit Hilfe der Wechselspannung $U_\sim$ und der Spule S_1 wird die Probe wechselmagnetisiert. Am Widerstand R_I fällt eine Spannung ab, die dem Strom und damit der Feldstärke proportional ist. Die Spule S_2 besitzt sehr kleine Abmessungen

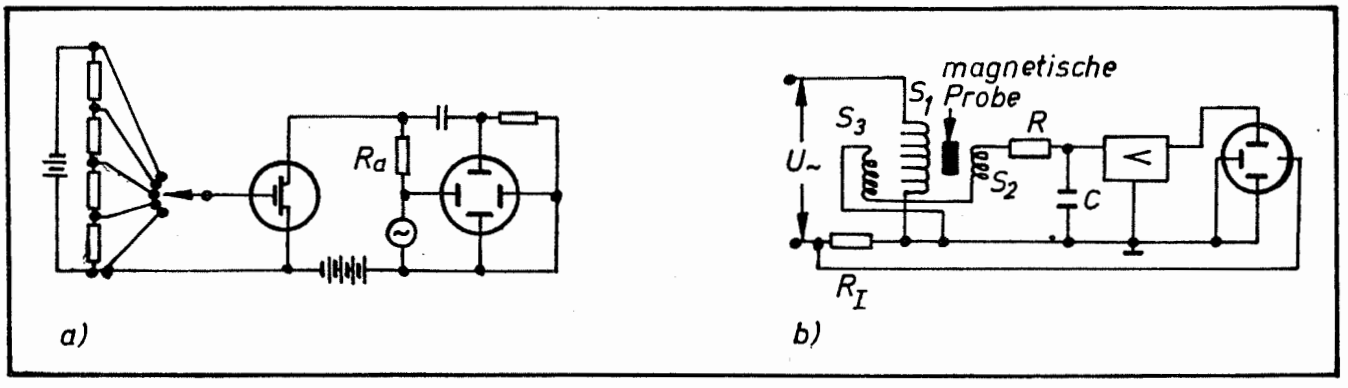


Bild 7.7.17. Prinzipschaltung zur Aufnahme a) der Ausgangskennlinienschar eines FET und b) der Hysteresekurven von magnetischen Materialien

und liegt eng um die Probe. Sie enthält meist viele Windungen. Eine zweite gleichartige Spule S_3 befindet sich ebenfalls im Feld von S_1, jedoch ohne Probe. Die Spulen S_2 und S_3 sind so geschaltet und abgeglichen, daß ohne Probe an ihrem gemeinsamen Ende keine Spannung entsteht. Mit Probe entsteht dann die Spannung

$$\boldsymbol{U}_{\mathrm{ind}} = K \frac{\mathrm{d}(\boldsymbol{B} - \boldsymbol{H})}{\mathrm{d}t}. \tag{24}$$

Die Größe K ist eine Konstante, die vom Aufbau, R, C und den Windungszahlen der Spulen abhängt. Mit der Kombination aus R und C wird die Spannung von Gl. (24) wieder integriert und anschließend über den Verstärker verstärkt. An den Y-Platten tritt daher eine Spannung auf, die der Größe $\boldsymbol{B} - \boldsymbol{H} = \boldsymbol{M}$ proportional ist. Da an den X-Platten eine $\boldsymbol{H}$-proportionale Größe liegt, entsteht auf dem Schirm eine geschlossene Kurve, die genau der Hysteresekurve der Probe entspricht. Ein Beispiel zeigt u. a. Bild 1.2.6. Mit stufenweiser Änderung der Amplitude von $U_{\sim}$ oder zusätzlicher Gleichmagnetisierung der Probe können auch Kurven gemäß Bild 1.2.7 erhalten werden.

Im Prinzip können die bisher genannten Beispiele bereits als LISSAJOUS-*Figuren* betrachtet werden. Im engeren Sinne deutet dieser Begriff jedoch auf geschlossene Figuren hin, die entstehen, wenn unmittelbar oder über dazwischenliegende Schaltungen an die Ablenkplatten Spannungen verschiedener Frequenz angelegt werden. In der Regel werden die LISSAJOUS-Figuren dann zur Bestimmung des Frequenzverhältnisses benutzt. Dazu gibt es vielfältige Möglichkeiten, die hier nicht weiter behandelt werden können.

7.7.4.3. Prinzipaufbau von Oszillographen

Der prinzipielle Aufbau eines Oszillographen geht aus Bild 7.7.18a) hervor.

Der *Y-Weg* ist sehr einfach. Mittels eines stabilisierten, breitbandigen Verstärkers (meist mehrere MHz) und geeichter Spannungsteiler, die reproduzierbar und geeicht grob und fein verstellt werden können, gelangt das Eingangssignal zu den Y-Platten. Am Ende des Verstärkers besteht noch die Möglichkeit, den Ruhepunkt des Y-Wertes über den Schirm nach oben und unten verschieben zu können. Dadurch können Details des Schwingungsverlaufs beobachtet werden.

Der *X-Weg* ist meist vielfältig gestaltet. Am Ende befindet sich wieder ein stabiler und breitbandiger Verstärker. In vielen Fällen ist er identisch mit den Eigenschaften des Y-Verstärkers. Dies ist für viele Anwendungen der doppelten Ablenkung gemäß den vorigen Abschnitten wichtig. Auch ihm folgt meist eine Punktverschiebung in x-Richtung über den Schirm. Der Eingang des *X-Verstärkers* ist umschaltbar, in der Regel auf drei Varianten:

- direkte X-Ablenkung über geeichte, einstellbare Spannungsteiler,
- periodische, zeitlineare X-Ablenkung,
- einmalig ausgelöste X-Ablenkung.

Die *Ablenkung* erfolgt durch ein spezielles *Kippgerät*. Seine Funktion geht auf das u. a. bereits im Zusammenhang mit Bild 7.3.24c) erklärte Prinzip zurück und ist hier in Bild 7.7.18b) gezeigt: Eine kontinuierlich einstellbare Stromquelle lädt einen (umschaltbaren) Kondensator C auf. Der Spannungsanstieg ist dabei linear. Sobald am Kondensator eine bestimmte Spannung erreicht wird, schließt sich der Schalter, und die Spannung wird sprungartig Null. Nach dem Öffnen des Schalters kann der Ladevorgang neu beginnen.

Bei *periodischer Ablenkung* öffnet sich der Schalter automatisch, wenn der Spannungswert Null erreicht ist. So entsteht die Ablenkungsspannung, wie sie Bild 7.7.18e)

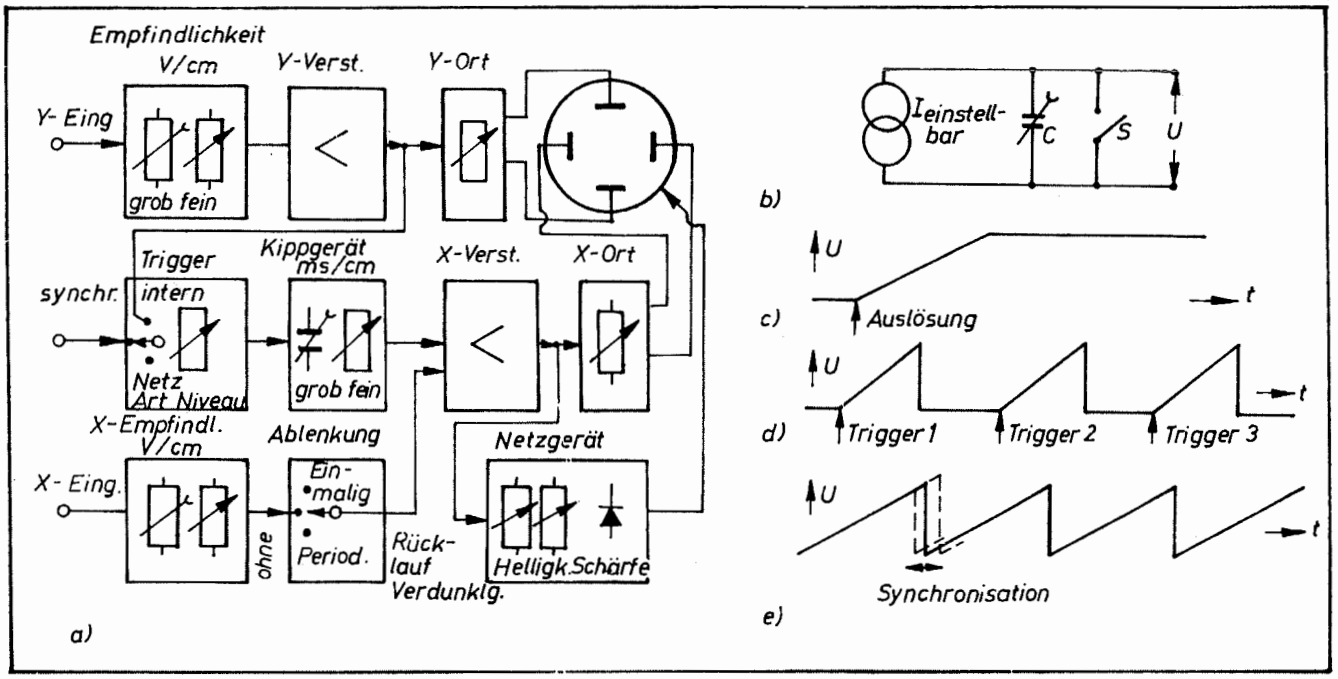


Bild 7.7.18. Aufbau und Funktion eines Oszillographen

a) Prinzipschaltung
b) Grundprinzip des triggerbaren Sägezahngenerators
c) Einmalige Ablenkung durch einen Triggerimpuls
d) Periodische Ablenkung durch Triggerimpulse
e) Periodische Ablenkung mit Synchronisation

zeigt. Die Synchronisation (Triggerung) mit dem Eingangssignal kann dadurch erzwungen werden, daß das Schließen des Schalters auf einen größeren oder kleineren Spannungswert verschoben wird. Der Einflußgrad kann dabei über das Triggerniveau eingestellt werden. Die Herkunft der Synchronisation ist meist auf drei Quellen einstellbar:

- intern, d. h. Ableitung aus dem Ausgang des Y-Verstärkers,
- extern, über ein frei wählbares Eingangssignal,
- Netz (aus dem Wechselstromnetz).

Bei der *einmaligen Ablenkung* wird entsprechend den soeben beschriebenen drei Wegen aus einem Signal ein Triggerimpuls abgeleitet und dadurch der Schalter geöffnet. Die so entstandene Ablenkung zeigt Bild 7.7.18c).

Zwischen der einmaligen und der periodischen Ablenkung liegt eine *periodische Triggerung*. Ihr Prinzip zeigt Bild 7.7.18d). Aus dem Signal am Triggereingang werden periodische (oder einzelne) Triggerimpulse abgeleitet. Sie öffnen den Schalter und bewirken so den zeitlinearen Spannungsanstieg. Hierbei sind der Auslösezeitpunkt und die Anstiegsgeschwindigkeit unabhängig voneinander wählbar. So lassen sich bestimmte Abschnitte aus dem Eingangssignal hervorheben.

In Bild 7.7.18 ist noch das *Netzgerät* angedeutet. Es bewirkt die Stromversorgung aller Baugruppen und der Oszillographenröhre. Hierbei kann die *Helligkeit* und *Schärfe* des Strahls eingestellt werden. Weiter greift hier die Verdunkelung des Strahlrücklaufs beim Schließen des Schalters ein.

7.7.4.4. Beispiel: gedämpfte Schwingung

Für den Oszillografen existieren so viele Anwendungen, daß selbst Spezialwerke wie z.B. [C 6, M 15] nur einige Beispiele anführen können. Hier können noch weniger

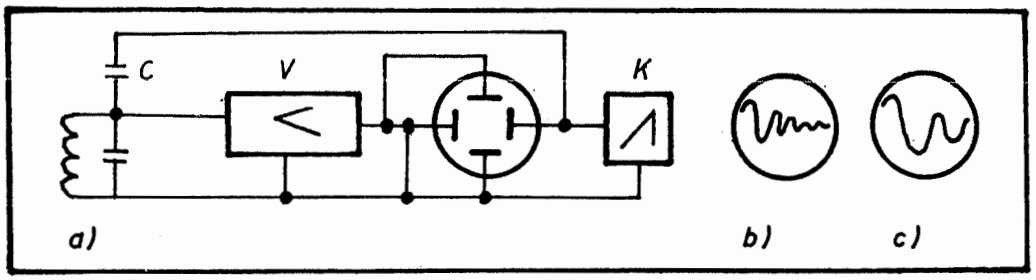


Bild 7.7.19. a) Schaltung zur Erzeugung einer gedämpften Schwingung, b) und c) Beispiele ihrer Darstellung auf dem Leuchtschirm

genannt werden. Ein recht aufschlußreiches Beispiel ist die Darstellung der *gedämpften Schwingung* gemäß Bild 7.7.19. Vom Kippgenerator K wird mittels des Kondensators C eine Differentiation des Rücklaufimpulses vorgenommen und mit dem so entstehenden Impuls der Schwingkreis angestoßen. Während der Zeitablenkung führt der Schwingkreis dann seine gedämpfte Schwingung aus, die über den Verstärker den Y-Platten zugeführt wird. Je nach der Steilheit der Zeitachse kann die gesamte Schwingung (b) oder ein Ausschnitt (c) sichtbar gemacht werden. Aus dem Verlauf der Kurve kann auf die Dämpfung des Kreises gemäß Abschnitt 2.2. geschlossen werden. Mittels Helltastung mit einer Bezugsfrequenz kann so auch leicht die Resonanzfrequenz des Schwingkreises bestimmt werden.

7.7.4.5. Dreidimensionale Oszillographie

Im Prinzip ist mit dem Oszillographen auch ein dreidimensionaler Zusammenhang darstellbar. Dabei werden die X- und Y-Größen direkt verstärkt den X- und Y-Platten zugeführt. Die dritte, die Z-Größe, wird additiv in entsprechenden Anteilen den X- und Y-Signalen aufgeprägt und steuert zugleich den WEHNELT-Zylinder der Katodenstrahlröhre. Dadurch erscheinen die weiter hinten (in z-Richtung) liegenden Meßwerte dunkler und heben so die Räumlichkeit des Bildes hervor. Dieses Prinzip, auf die Darstellung von Kurvenscharen eines FET angewendet, zeigt Bild 7.7.20. Infolge der Einfachheit liegt die z-Achse (u_G) hier in einem Winkel um 45°. Für viele Anwendungen sind weitaus kompliziertere Schaltungen entwickelt worden. Sie erlauben mit speziellen Cosinuspotentiometern eine beliebige Raumdrehung aller Koordinaten.

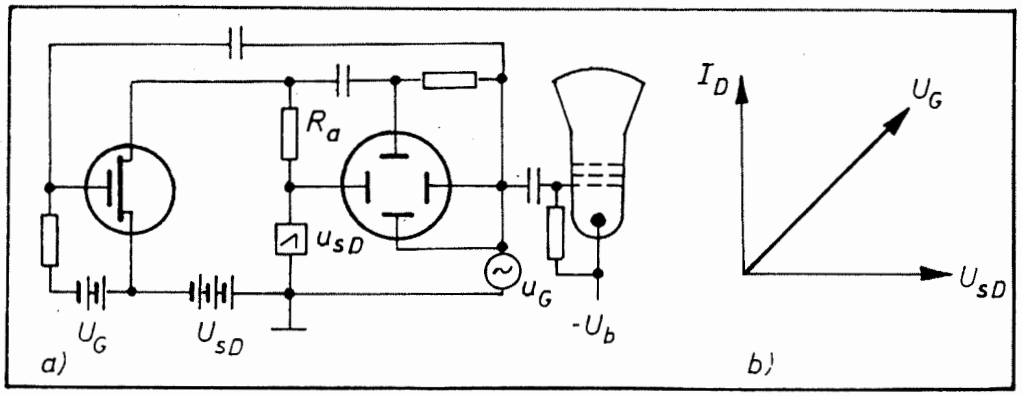


Bild 7.7.20. Zur dreidimensionalen Darstellung von Oszillogrammen bei zwei Abhängigen. Mittels des Kippgenerators, der Gleichstrom durchlassen soll, wird die Kennlinie $I_D = f(U_{SD})$ geschrieben. Sie ist eine Funktion der Gatespannung. Sie wird mittels der Quelle u_G sehr langsam gegenüber der Kippfrequenz geändert. Sie wird dabei auch beiden Plattenpaaren zugeführt und beschreibt so eine Verlagerung des Ursprungs in Richtung des Pfeiles u_G im Teilbild b). Alle Kennlinien werden entsprechend verlagert. Da die Spannung u_G außerdem dem WEHNELT-Zylinder zugeführt wird, entsteht in Richtung u_G in Bild b) eine Verdunklung. Sie verstärkt die dreidimensionale Wirkung des Bildes auf dem Schirm

7.7.4.6. Mehrstrahl-, Sampling- und Projektionsoszillograph

Für viele Anwendungen ist es notwendig, Signale an verschiedenen Stellen einer Schaltung gleichzeitig und zeitsynchron abzunehmen und darzustellen. Hierzu wurden *Mehrstrahlröhren* entwickelt. Sie enthalten zwei oder mehr komplette Strahlerzeugungs- und Ablenksysteme in einem Kolben hinter einem Leuchtschirm. Sie werden meist aus einem Kippgerät gespeist, enthalten aber je einen Verstärker für die einzelnen Systeme zur Y-Ablenkung. Eine indirekte Ausführung der Mehrfachoszillographie verwendet *elektronische Umschalter*. Sie schalten nacheinander mit einer Frequenz f_n die einzelnen Signale dem Y-Verstärker zu (Bild 7.7.21a)). Auf dem Schirm entsteht dann z. B. Bild 7.7.21b). Im allgemeinen wird nun aber die Umschalt-

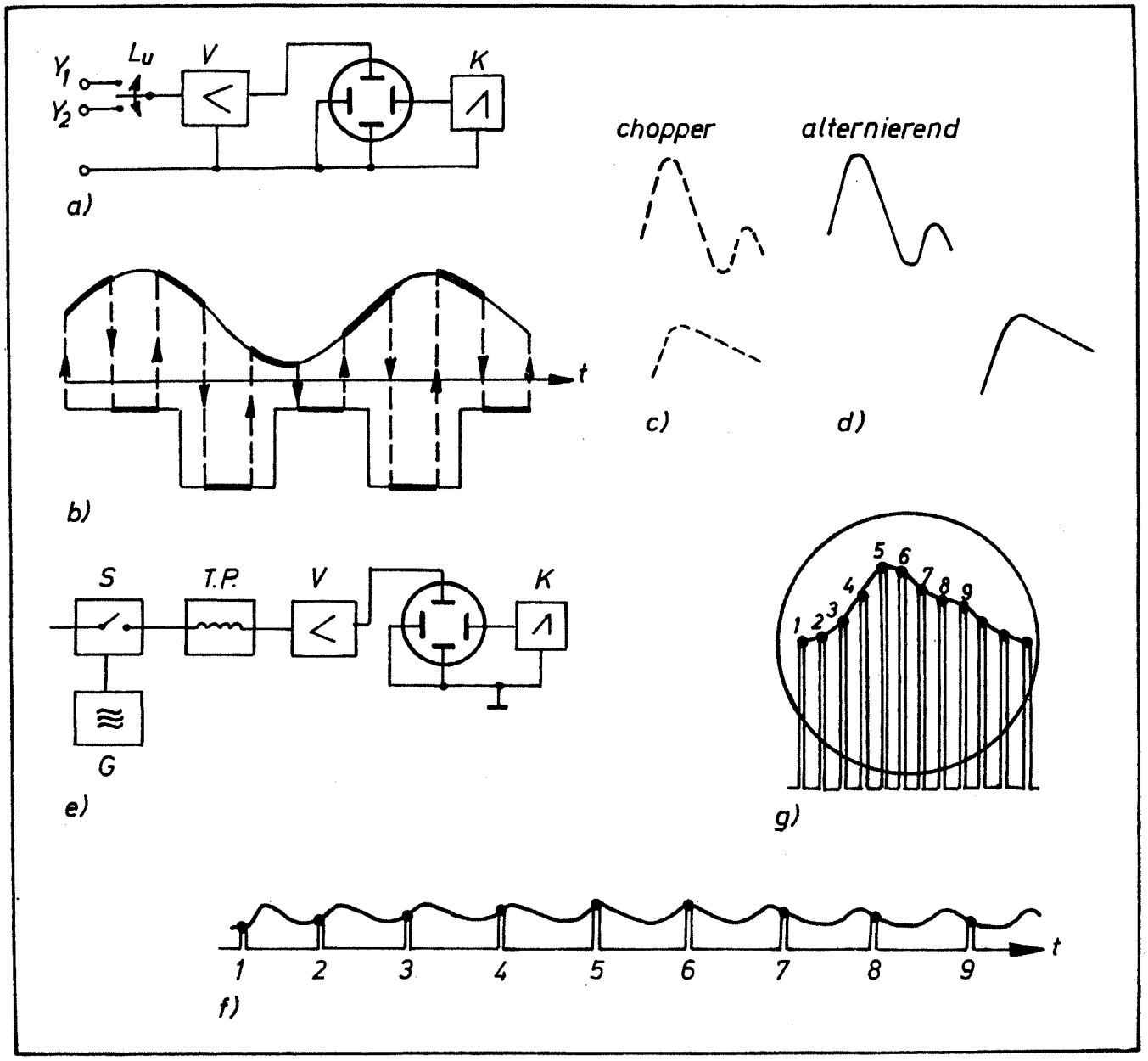


Bild 7.7.21. Zum Mehrstrahl- und Samplingoszillographen

a) Prinzip der periodischen Umschaltung zwischen zwei Signalquellen
b) Einmaliger Ausschnitt aus einem Schirmbild
c) Chopper-Betrieb
d) Alternierender Betrieb
e) Prinzip des Samplingoszillographen
f) Erzeugung der Abtastwerte aus einer sehr hochfrequenten Schwingung
g) Zusammensetzung der Abtastwerte zum Bild

frequenz in keinem rationalen Verhältnis zur Sägezahnspannung stehen. Dann befinden sich die Umschaltstellen bei jedem einzelnen Ablenkvorgang an einer anderen Stelle. Dadurch erscheinen beide Kurven für das Auge nicht mehr kommutiert, sondern zusammenhängend.

Gemäß Bild 7.7.21c) und d) können durch die Wahl des Frequenzverhältnisses von Zeitablenkung und Umschaltung die beiden Betriebsweisen

chopper und alternierend

unterschieden werden. Beim alternierenden Betrieb geht aber die zeitliche Zuordnung zwischen den beiden Signalen verloren.

Für sehr hohe Frequenzen bereitet sowohl die Verstärkung der Signale als auch die Strahlablenkung und damit die hohe Schreibgeschwindigkeit Schwierigkeiten. Hier bietet das *Samplingprinzip* einen effektiven Ausweg. Gemäß Bild 7.7.21g) werden aus der hochfrequenten periodischen Schwingung einzelne Spannungswerte ausgetastet, die etwas weiter als eine Periode entfernt sind. (Teilbild f)). Zu diesem Zweck wird der elektronische Schalter S von einem Generator G gesteuert, dessen Frequenz ein wenig höher als die des Signals eingestellt wird. Durch den Tiefpaß T. P. werden die Impulse (1, 2, 3, ...) zu einem neuen, erheblich tieferfrequenten Signal vereinigt. Bei günstiger Wahl der Tastpunkte gibt es genau den zeitlichen Verlauf des ursprünglichen Signals wieder (Teilbild g)). Dieses Signal kann folglich leicht verstärkt und dargestellt werden.

In Schulen und Universitäten müssen Oszillogramme einem großen Kreis sichtbar gemacht werden. Früher dienten hierzu *Projektionsoszillographen* mit großer Helligkeit und speziellen lichtstarken Optiken. Heute wird hierzu hauptsächlich das *industrielle Fernsehen* eingesetzt.

7.7.4.7. Digital- und Speicheroszillograph

Bei vielen Anwendungen ist es notwendig, einen einmaligen Vorgang längere Zeit auf dem Oszillographenschirm sichtbar zu halten. Hierfür existieren schon lange die *Speicheroszillographen*. Sie verwenden spezielle speichernde Röhren, die im Abschnitt 6.2.3.3. beschrieben sind.

Eine wesentlich bessere Methode entwickelte sich mit der *Digitaltechnik*. Durch sie ist es möglich, die analogen Eingangssignale mittels AD-Wandler (9.2.) in digitale Signale umzuwandeln. Dann sind sie relativ leicht zu speichern, z.B. in Halbleiterspeichern, auf Floppy disc usw. Sie können von dort jederzeit wieder abgerufen werden, mittels DA-Wandler in analoge Werte zurückgewandelt und sichtbar gemacht werden. Der *Vorteil* dieser digitalen Speicheroszillographen geht jedoch viel weiter: Der *Zeitmaßstab* kann auch hinterher beliebig geändert werden, *mehrere Signal* können nicht nur addiert, sondern beliebig mit den digitalen Methoden *verarbeitee* werden; einzelne Punkte können genau nach Zeit und Amplitude als digitale Werte bestimmt werden. Dies sind nur einige Vorteile, die mit dem sonst recht hohen Aufwand unmittelbar zur Verfügung stehen.

Zu Beginn der Entwicklung kamen die digitalen Zusätze zu analogen Oszillographen als *Transientenrecorder* bzw. *Eventoskope* auf den Markt. Heute werden analoge Oszillographen, AD- und DA-Wandler, Speicher, Digitalvoltmeter, digitale Informationsverarbeitung usw. als Einheit produziert: der *digitale Speicheroszillograph*.

7.7.5. Meßwert- und Datenspeicherung

Bereits in mehreren Beispielen wurde die Bedeutung der Speicherung von Meßwerten erwähnt. Die Besonderheit ihres Prinzips wurde bereits im Abschnitt 6.3.3.6. heraus-

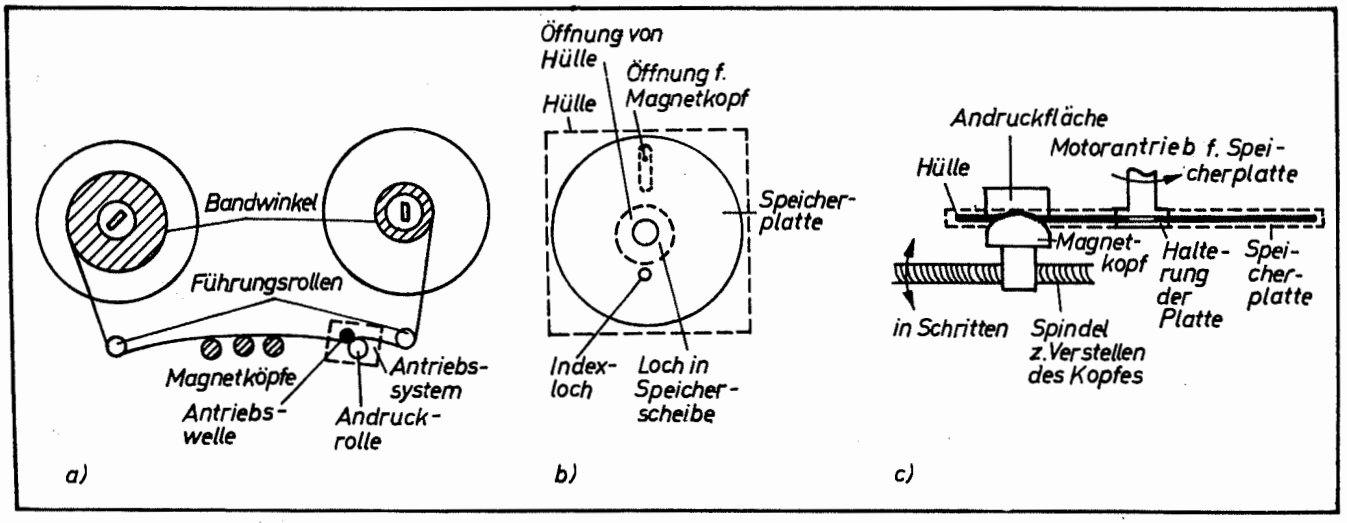


Bild 7.7.22 a) Prinzipaufbau eines Matnetbandspeichers
b) Flexible Folie (Floppy disc) in ihrer Hülle (gestrichelt)
c) Eingespannte Folie mit Antrieb sowie Magnetkopf mit Gegenandruckfläche. Kopf und Andruckfläche wirken auf die ständig rotierende Folie nur bei Aufzeichnung und Wiedergabe ein. Die Positionierung des Magnetkopfes erfolgt über die Spindel mit Schrittmotoren auf die genau definierten einzelnen Spuren

gearbeitet. Weiter sei nur auf ihre Notwendigkeit bei den Korrelationsverfahren (7.7.3.6.) und bei den Digitaloszillographen (7.7.4.7.) verwiesen. Wegen dieser und vieler anderer Anwendungen sowie wegen ihrer generellen Bedeutung in der Digitaltechnik seien hier kurz zwei wichtige Prinzipien behandelt, die *nicht auf Halbleiterspeichern* beruhen:

- die Magnetbandspeicherung und
- die Speicherung auf einer flexiblen Folie, einem Floppy disc.

Beide zeigt in ihrem prinzipiellen Aufbau Bild 7.7.22.

7.7.5.1. Magnetbandspeicherung

Die Magnetbandspeicherung kann analog und digital mit und ohne Modulation erfolgen. Hier wird nur der einfache analoge Fall beschrieben. Von einer Vorratsrolle mit Magnetband wird das Magnetband über eine Umlenkrolle an drei Magnetköpfen vorbeigeführt. Ein spezieller Antrieb sorgt für seine möglichst konstante Geschwindigkeit. Heute werden relative Geschwindigkeitsabweichungen bis unter 10^{-3} erreicht. Der Antrieb gestattet meist auch, die Bandgeschwindigkeiten in den Verhältnissen von 1:2:4:8 ... bis zu etwa 1:128 in Stufen umzuschalten. Auf einer zweiten Spule wird über eine zweite Umlenkrolle das Band wieder aufgewickelt. Der Löschkopf dient dazu, die alte Aufzeichnung auf dem Magnetband zu beseitigen. Mit dem Aufzeichnungskopf wird das zu registrierende Signal dem Magnetband magnetisch eingeprägt. Der Wiedergabekopf erzeugt aus dem magnetisch gespeicherten Signal wieder ein elektrisches. Lösch- und Aufzeichnungskopf sind elektrisch bei der Wiedergabe abzuschalten. Die *Magnetbandspeicherung* besitzt für die Meßtechnik etwa die *folgenden Vorteile*:

— Aufzeichnung und Wiedergabe kann unabhängig voneinander zu den jeweils günstigsten Zeiten und an den jeweils günstigsten Orten erfolgen.
— Aufzeichnung und Wiedergabe kann mit unterschiedlicher Geschwindigkeit erfolgen. Dadurch kann das Meßsignal den speziellen Analysiergeräten durch multiplikative Zeit- und damit Frequenztransponierung angepaßt werden.

- Der Wiedergabevorgang kann auch prinzipiell sofort nach, ja noch während der Aufzeichnung erfolgen, da keine Entwicklung wie bei Fotofilm oder ähnlichem notwendig ist.
- Der Wiedergabevorgang und damit indirekt der Meßvorgang ist beliebig oft wiederholbar und kann daher nach sehr verschiedenen Gesichtspunkten in Ruhe optimal ausgewertet werden. Auch im Meßvorgang unerwartete Ereignisse sind so meist ohne Komplikation gut auszuwerten.
- Spezielle, besonders interessante Ausschnitte aus dem Meßvorgang können als Bandabschnitt herausgeschnitten und zu einer endlosen Schleife zusammengeklebt und dann für die Wiedergabe ständig periodisch wiederholt und damit besonders genau analysiert werden.
- Magnetbandgeräte gestatten, einen sehr großen Frequenzbereich bei hohem Störabstand zu speichern. Durch Wahl der Transportgeschwindigkeit und zusätzlicher Modulations- und Demodulationseinrichtungen sind beide Größen weitgehend allen Forderungen anzupassen.
- Magnetspeichergeräte sind relativ robust zu gestalten. Sie ermöglichen häufig, mehrere Signale gleichzeitig zu registrieren. Sie sind leicht transportabel und gestatten, eine sehr große Informationsmenge zu speichern.
- Magnetbänder sind gut archivierbar, und die Aufzeichnungen erleiden selbst über Jahrzehnte kaum irgendwelche Änderungen. Durch die Löschbarkeit des Magnetbandes ist eine häufige Wiederverwendung und damit ein relativ ökonomischer Betrieb möglich.

Neben dieser verdichteten Aufzählung sei hier das mittlerweile sehr umfangreiche Spezialgebiet der magnetischen Signalspeicherung nicht weiter behandelt. In der deutschsprachigen Literatur wird es meist mit Meßwertspeicherung bzw. *Analogspeichertechnik* bezeichnet. Im gewissen Umfange zählen neuerdings aber auch die Digitalbandspeicher dazu. Im Angelsächsischen bilden derartige Geräte einen Hauptteil der sogenannten „Instrumentation". Daher stammt auch ihre Bezeichnung *instrumentation-recorder*.

7.7.5.2. Folienspeicher, Floppy disc

Der Folienspeicher verwendet eine flexible Folie von ca. 75 µm Dicke, die beidseitig mit einem magnetisierbaren Material wie ein Magnetband beschichtet ist. Genormt sind heute vor allem Foliendurchmesser von 20 cm (Standard), 13 cm (Mini) und 9 cm (Mikro). Die Folie befindet sich in einer etwa quadratischen Schutzhülle und bleibt auch in dieser bei Betrieb. Sie besitzt ein zentrales Loch, mit dem sie von einem Transportwerk aufgenommen und in Rotation (mit meist 360 Umdrehungen je Minute) versetzt werden kann. Auf der Folie können eine Vielzahl (oft 77) *kreisförmige Spuren* aufgezeichnet und wiedergegeben werden. Es liegen also im Gegensatz zur Schallplatte nicht eine spiralig angeordnete Spur, sondern einzelne, endlose Spuren an genau definierten Radien vor. Der Zugriff zu den einzelnen Spuren erfolgt durch genaue *Positionierung des Magnetkopfes* auf diesen Radien mittels Schrittmotoren und er dauert 3 bis 10 ms. Damit kann eine bestimmte Informationsmenge an bestimmten (meist 32) *Sektoren* dieser Spur abgelegt und wiedergefunden werden. Es kann also im Gegensatz zum Magnetband keine kontinuierliche Speicherung und Wiedergabe erfolgen. Nachdem eine Spur aufgezeichnet bzw. gelesen wurde, muß erst der Magnetkopf auf die nächste Spur positioniert werden. Dieses Verfahren ist für die bisher behandelte analoge Meßtechnik wenig vorteilhaft. Für die digitale Meßtechnik hat es jedoch den Vorteil, daß gewünschte Daten schnell nach den Koordinaten Spur-Nr. und Sektor-Nr. aufgezeichnet und wiedergefunden werden können. Dieses Problem wird beim Mikrorechner noch deutlich zutage treten (8.8.).

7.8. Ladung verschiebende Schaltungen

Um das Jahr 1970 entstanden etwa gleichzeitig und unabhängig voneinander zwei neue Prinzipien für integrierte Schaltungen. Infolge gewisser Ähnlichkeiten werden beide zusammengefaßt

CTD = Charge Transfer Devices genannt. Entsprechende deutsche Bezeichnungen wären Ladung transportierende oder verschiebende Bauelemente bzw. Schaltungen.

Das eine Prinzip BBD = Bucket Brigade Device geht im wesentlichen auf SANGSTER aus dem Jahre 1969 zurück. Es wird im Deutschen meist mit Eimerkettenspeicher übersetzt. Dies rührt von der Vorstellung her, daß die Ladung hierbei wie Wasser bei einer Feuerlöschkette in Eimern weitergeschickt wird.

Das zweite Prinzip wurde 1971 von BOYLE und SMITH vorgestellt und fand erstaunlicherweise nicht sofort Beachtung. Es wird CCD = Charge Coupled Device genannt und ist wohl nie übersetzt worden, da sich international sofort der Begriff CCD durchgesetzt hat (vgl. 3.3.8. und 6.1.5.).

Inzwischen gibt es zu beiden Prinzipien, insbesondere aber zu den CCD, eine Vielzahl von Varianten und Weiterentwicklungen. Da die Entwicklung hier noch stark im Fluß ist, können auch vielfältige neue Lösungen auftauchen. Dennoch wurde in Bild 7.8.1 ein Versuch zur Übersicht gemacht. Im gewissen Sinne steht zwischen beiden Techniken der SCT = Surface Charge Transistor. Im wesentlichen stellt er eine abgewandelte *CCD-Struktur* dar. Sobald aber beide Techniken nach aktivem, also teil-

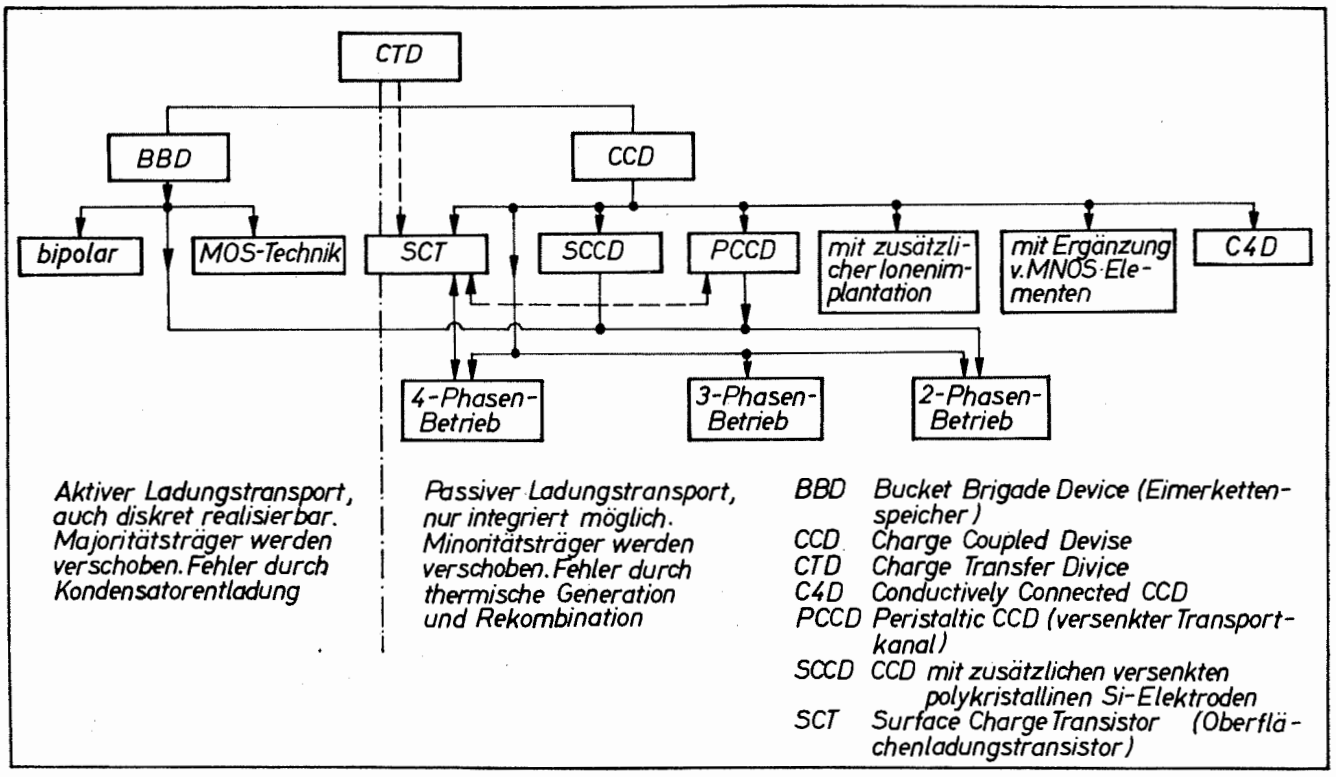


Bild 7.8.1. Versuch einer vorläufigen Einteilung von Ladung verschiebenden Schaltungen

weise verstärkendem oder passivem Ladungstransport eingeteilt werden, gehört er mehr zu den BBD. Andererseits ist er in der Struktur praktisch identisch mit einer CCD-Zweiphasenstruktur, bei der versenkte polykristalline Si-Elektroden zusätzlich verwendet werden. Diesem Fakt wurde im Bild versucht Rechnung zu tragen. Insgesamt wird sich das Bild aber in der nächsten Zeit sehr wahrscheinlich noch erheblich verändern.

Die Anwendungen beider Prinzipien sind vielfältig. Auf dem noch im Kapitel zu behandelnden digitalen Gebiet stehen die Schieberegister und Speicher im Vordergrund. Bei den analogen Anwendungen herrscht eine weitaus größere Vielfalt. Heute zeichnen sich etwa folgende Anwendungsmöglichkeiten ab:

- Analoge Schieberegister zur
  - — Frequenztransponierung,
  - — Zeitkompression oder -dehnung,

— Zeitsynchronisation (Zeitfehlerkorrektur),
— Vermeidung elektroakustischer Echos,

- Multiplexer, d. h. Übergang zwischen seriellen und parallelen Signalen und umgekehrt,
- neue Möglichkeiten der Filtertechnik (7.9.),
- neue Bildaufnahmetechnik (6.1.5.).

Insgesamt liegen damit neue Möglichkeiten zum Verzögern, Verändern, Schalten, Sortieren und Verarbeiten von analogen Informationen vor.

7.8.1. Eimerkettenspeicher

Das eigentliche Prinzip der Eimerkettenschaltung stammt aus dem Anfang der fünfziger Jahre. Es entspricht einer Prinzipanordnung von Bild 7.8.2. Es arbeitet nach dem Zweiphasenprinzip. Zunächst tragen die ungeradzahligen Kondensatoren die Information, und die geradzahligen seien leer. Im ersten Takt wird dann durch die aktiven Schalter S_1, S_3 usw. die Ladung als Träger der Information an die geradzahligen Kondensatoren weitergeleitet, wobei die ungeradzahligen leer werden sollen. Im zweiten Takt wird dem ersten Kondensator neue Information zugeführt und gleichzeitig mittels der geradzahligen Schalter die Ladung von den geradzahligen zu den ungeradzahligen Kondensatoren weitergeleitet. Dieses Prinzip forderte so aufwendige Schalter, daß selbst eine integrierte Realisierung für kurze Kettenlängen nicht möglich war. Es fand daher auch kaum praktische Realisierung.

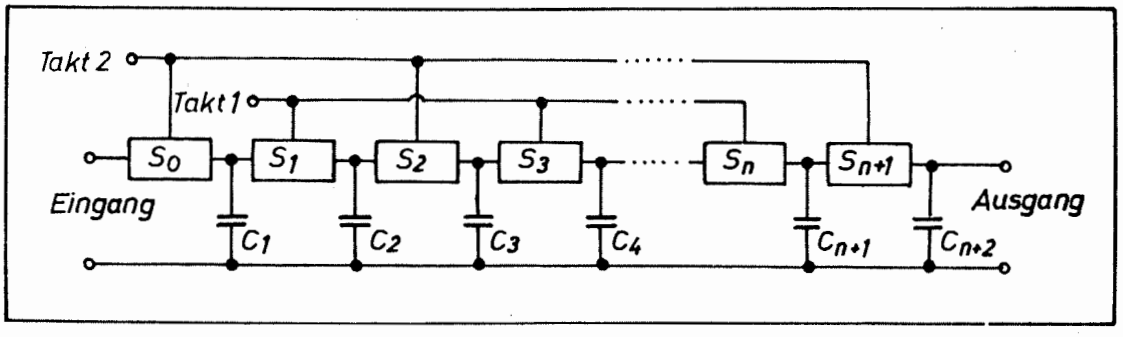


Bild 7.8.2. Prinzipanordnung des ursprünglichen Eimerkettenspeichers

7.8.2. Ladung rückwärts, Information vorwärts

Die entscheidende Idee von SANGSTER war es, Ladung und Information entgegengesetzt zu transportieren. Dies kann am besten mit einer Eimerkette gemäß Bild 7.8.3 veranschaulicht werden. Entsprechend der Reihe a) ist zunächst jeder geradzahlige Eimer voll gefüllt und trägt also keine Information. Die Information bei den ungeradzahligen steckt in der Menge fehlenden Wassers, also dem Volumen V. Im ersten Takt der Reihe b) wird nun soviel Wasser von den gefüllten geradzahligen Eimern nach links in die informationstragenden ungeradzahligen geschüttet, bis sie voll sind. Nach diesem Takt, Reihe c), ist die Fehlmenge Wasser (Information) genau und fehlerlos in die geradzahligen Eimer, also einen Schritt nach rechts gewandert. Im zweiten Takt wird nun das Wasser aus den gefüllten, jetzt ungeradzahligen ebenfalls nach links in die geradzahligen gefüllt. Die Information wandert damit wieder um einen Eimer nach rechts. Gleichzeitig wird im zweiten Takt aus dem ersten Eimer

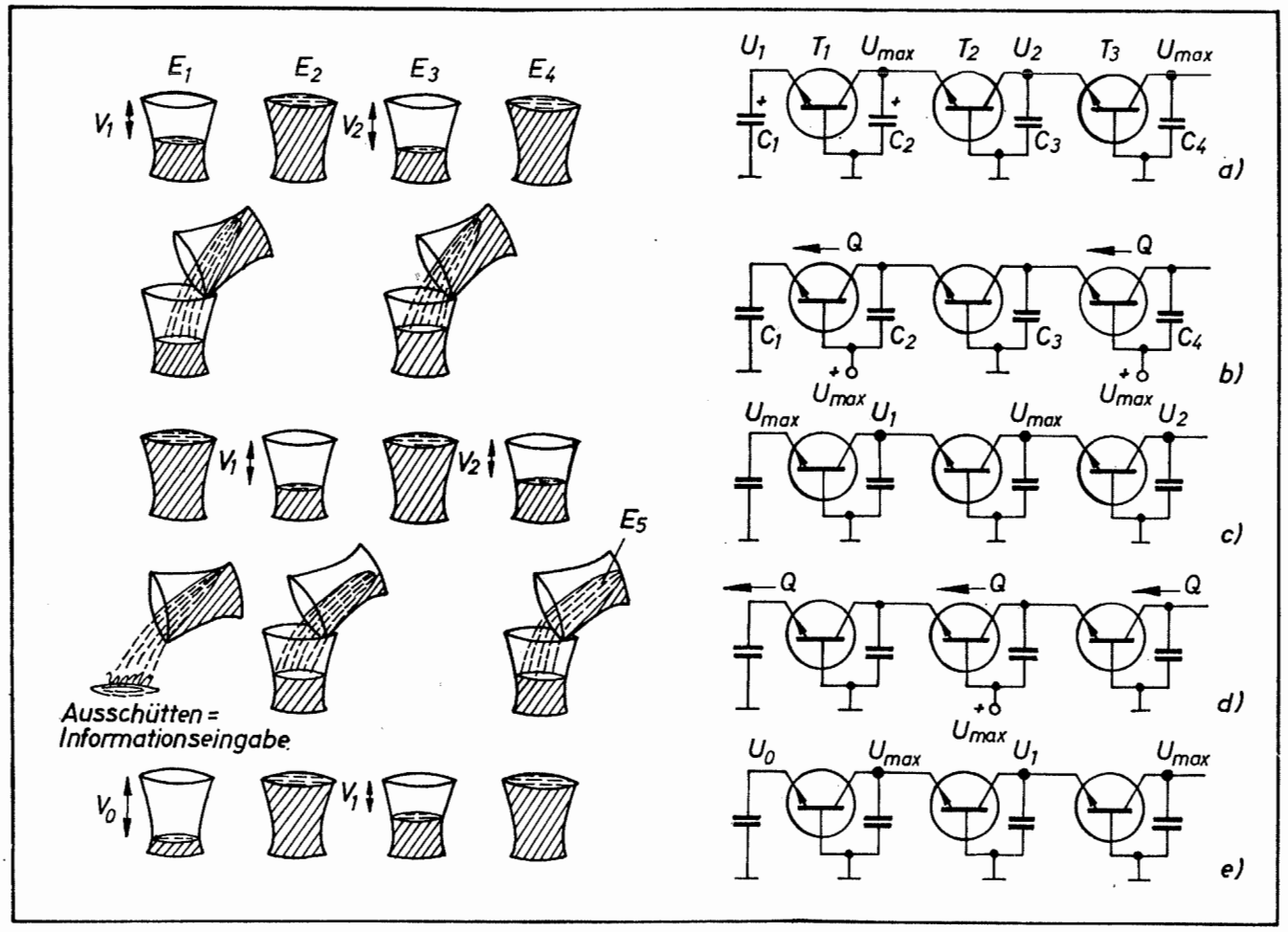


Bild 7.8.3. Schematisierung der Vorgänge bei Eimerkettenspeichern; links: Die Information ist die bis voll fehlende Wassermenge; rechts: Schema in bipolare Transistoren (richtig zugeorgnet) umgesetzt

soviel Wasser ausgeschüttet, wie der neuen Information entspricht. Es entsteht das Verlust-Volumen V_0.

Am Ende der Kette wird die Information — dies ist nicht mehr dargestellt — dadurch erhalten, daß der letzte Eimer gefüllt und die dabei notwendige Wassermenge bestimmt wird. Sie entspricht genau der Wassermenge, die vor der entsprechenden Anzahl von Takten aus dem Eimer E_1 ausgeschüttet wurde, also der Eingangsinformation.

Dieses soeben demonstrierte Prinzip läßt sich mit relativ einfachen Mitteln ins Elektronische übertragen. Dies sei zunächst an einer diskreten Schaltung erklärt. Sie ist rechts in Bild 7.8.3 dargestellt. Dabei ist zu beachten, daß die Spannungen U_1, U_2 usw. stets kleiner als $U_{\max}$ sind. In beiden Anordnungen entsprechen sich:

Eimer	Kondensatoren C
Wassermengen	Ladungen Q
Umkippen von Wasser	Anheben des Basispotentials und des Kondensators auf den definierten Betrag $U_{\max}$

Damit brauchten bei der Transistorschaltung nur die beiden Zustände Basispotential auf Null oder auf $U_{\max}$ betrachtet zu werden. Liegt das Basispotential auf Null, so ist der Transistor nach beiden Seiten gesperrt. Die Ladungen auf den Kondensatoren behalten, wenn von Verlusten abgesehen wird, ihren Wert bei. Wird der Basisanschluß und der Kondensator auf $U_{\max}$ gehoben, so wird die Emitterbasisstrecke leitend, und das obere Potential des Kollektorkondensators nimmt gegenüber Masse den Wert $U_{\max}$

an. Es sei zur Vereinfachung eine unendlich hohe Stromverstärkung des Transistors angesetzt, d. h., es fließt praktisch kein Basisstrom. Dann fließt vom Kollektorkondensator Ladung zum Emitterkondensator über. Dies hält solange an, wenn auch von der Schwellenspannung abgesehen wird, bis der Emitter das Potential U_{max} erreicht. Dann sperrt nämlich der Transistor wieder. Er regelt also automatisch, daß der „Eimer" C_1 nicht überläuft. Gleichzeitig fehlt nach diesem Vorgang C_2 die Ladung, die vorher in C_1 fehlte.

7.8.3. Störeinflüsse

Die Schwellenspannung ist bei allen Transistoren nicht gleich groß. Dies bedeutet also eine überall gleiche Ladungsfehlmenge in den Kondensatoren (Eimern). Deshalb bleibt dieser Effekt ohne Einfluß. Ebenso läßt sich zeigen, daß auch die Kondensatoren nicht genau gleich zu sein brauchen. Es wird in jedem Fall die Ladungsmenge ja nur am Anfang und am Ende der Kette auf eine Spannung bezogen. Transportiert wird aber, sofern nicht Grenzwerte überschritten werden, genau die Eingangsladungsmenge, und zwar unabhängig von der Kondensatorgröße (Eimergröße). Verluste bringt dagegen der nicht vernachlässigbare Basisstrom. Er könnte beim Umgießen als wegspritzendes Wasser interpretiert werden. Er ruft einmal eine systematische Signaländerung hervor, die einer Signaldämpfung entspricht. Da dieser Strom aber auch von der Umladungsmenge selbst abhängt, bewirkt er nichtlineare Verzerrungen. Ist dies schon schwierig im Wassereimermodell zu erklären, so gibt es für die Emitter-Kollektor-Kapazität kaum noch eine Erklärung. Sie wirkt sich wie ein Nebensprechen der einzelnen Signalwerte aus. Weitere Störungen rufen der endliche Widerstand des Transistors in Basisschaltung und die nur exponentiell erfolgende Ladungsübertragung hervor. Es fließt anfangs ein recht großer Strom, der aber theoretisch erst nach unendlich langer Zeit gegen Null geht. Andererseits muß aber die Taktung in endlichen Zeiten erfolgen. Aus diesem Grunde wird die Zeit für die Phasen a), c), e) usw. praktisch zu Null, und die eigentlichen Takte b) und d) müssen in einem gewissen Bereich liegen. Die obere Taktfrequenz folgt vor allem aus den Verzerrungen durch die nicht vollständige Umladung und durch das Nebensprechen infolge der Emitter-Kollektor-Kapazität. Die untere Grenze wird durch Restströme der Transistoren und die Kapazitätsgröße festgelegt. Frequenzbereiche um 1 kHz bis 30 MHz sind durchaus möglich.

7.8.4. FET-Eimerkette

Ähnlich wie mit bipolaren Transistoren läßt sich die Eimerkettenschaltung auch mit MOSFET realisieren. Die integrierten Strukturen sowie die Schaltungen und Taktspannungen zeigt Bild 7.8.4.

Dabei ist auffallend, daß die Kapazitäten bereits durch geringfügige Strukturänderungen erhalten werden. Bei bipolaren Transistoren ist lediglich die Übergangsfläche Kollektor-Basis hinreichend zu vergrößern. Bei der MOS-Schaltung wird das Gate hinreichend weit über den p-Bereich gezogen, der zugleich das Drain des vorangegangenen und die Source des folgenden FET darstellt. Hierdurch ergeben sich beim MOSFET sehr einfache Halbleiterstrukturen. Beim FET ist außerdem von Vorteil, daß die Gates keinen Strom aufnehmen und damit dieser Verlustanteil entfällt. Nachteilig ist jedoch die höhere Schwellenspannung und der größere Übergangswiderstand

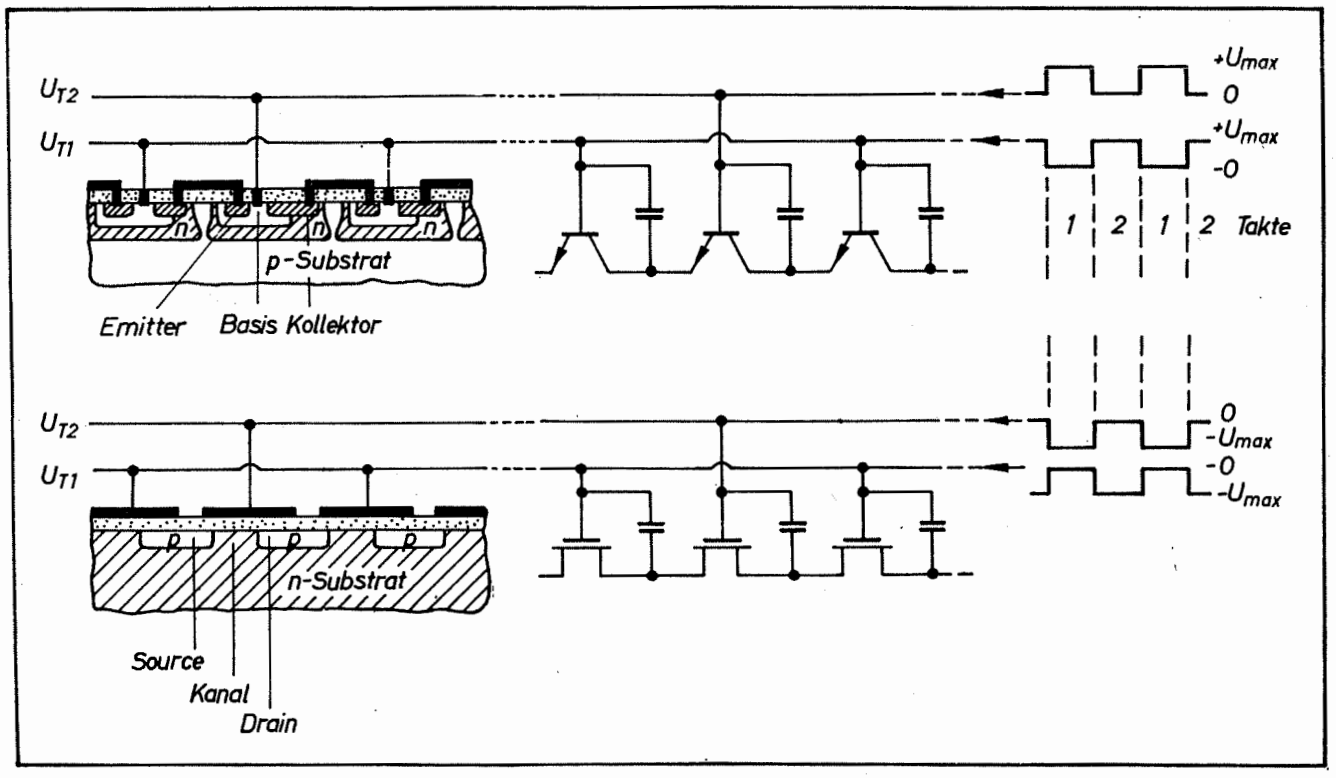


Bild 7.8.4. Struktur (links), Schaltungsprinzip (Mitte) und Taktung (rechts) von Eimerkettenspeichern; oben bipolar und unten in MOS-Technik (heute herrscht die n-Kanal-Technik mit p-Substrat vor)

sowie die prinzipiell geringere obere Grenzfrequenz. Es sind Taktfrequenzen bis zu etwa einigen MHz möglich. Die tiefsten Frequenzen liegen bei einigen 10 bis 100 Hz. Es werden hier Eimerkettenspeicher bis zu etwa 500 Stufen angeboten.

7.9. Nichtklassische Filter

7.9.1. Abgrenzung und Überblick

Hier geht es vor allem um jene Flter, die insbesondere in den letzten zehn bis zwanzig Jahren an Bedeutung gewonnen haben und noch keine Digitalfilter im engeren Sinne sind. Die *Digitalfilter* werden im Abschnitt 9.3. behandelt. Sie nutzen die vielfältigen Möglichkeiten der Rechentechnik, verwenden aber dabei Strukturen, die schon in diesem Abschnitt eingeführt werden.

Nicht zu behandeln sind hier, da bereits behandelt, die *RC- und LC-Filter* (2.3. bzw. 2.5.) und die *Aktivfilter* (7.5.8.). Nicht behandelt wird hier das *Optimalfilter*, auch angepaßtes Filter oder (engl.) matched Filter genannt, einschließlich des Spezialfalls des KALMAN-BUCY-Filters. Der Grund liegt darin, daß seine Aufgabenstellung von jener der üblichen Filter abweicht. Es sind nicht primär bestimmte Frequenzbereiche zu sperren und andere durchzulassen. Es kommt bei ihm vielmehr darauf an, aus dem Signal-Störgemisch das Signal optimal erkennbar herauszufiltern. In weitgefaßtem Sinne wurden solche Aufgaben bereits in den Abschnitten 5.4.9. (Phasenempfindlicher Gleichrichter) und 7.7.3. (Frequenzbereichfilterung, Dynamikregelungen, Akkumulation, Korrelationsverfahren, PLL, Impulsverdichtung usw.) behandelt. Die vier hier wichtigen Grundprinzipien bzw. -strukturen zeigt Bild 7.9.1.

Das *Mehrwegfilter* beruht darauf, daß hinter den einzelnen Teilfiltern unterschiedliche Phasen und Amplituden für eine Eingangsfrequenz auftreten. Durch die Summation mit unterschiedlicher Bewertung der Kanäle kann ihre Ausgangsamplitude dann bis auf Null abgesenkt, nur geschwächt werden oder aber mit voller Amplitude passieren. Als Teilfilter werden relativ oft *RC*-Glieder verwendet.

Beim *Transversalfilter* durchläuft das Signal mehrere einfache Teilfilter ($H_n(s)$). Nach jedem Teilfilter wird ein mittels des Koeffizienten c_n festgelegter Signalanteil aufaddiert. Dieses Prinzip ist heute in sehr vielfältiger Weise realisiert. In kontinuierlicher Betriebsweise kommt es bei den *Oberflächenwellen* (6.3.5., insbesondere Bild 6.3.19) gleich zweimal nacheinander zur Anwendung. Es kann aber auch mittels bestimmter Abtastwerte realisiert werden, dann entstehen die *Abtastfilter*. In diesem Fall sind die Teilfilter einfache Verzögerungsglieder für einen Zeittakt und enthalten im wesentlichen Speicherkondensatoren. Daher rührt auch der Name *Speicherfilter*.

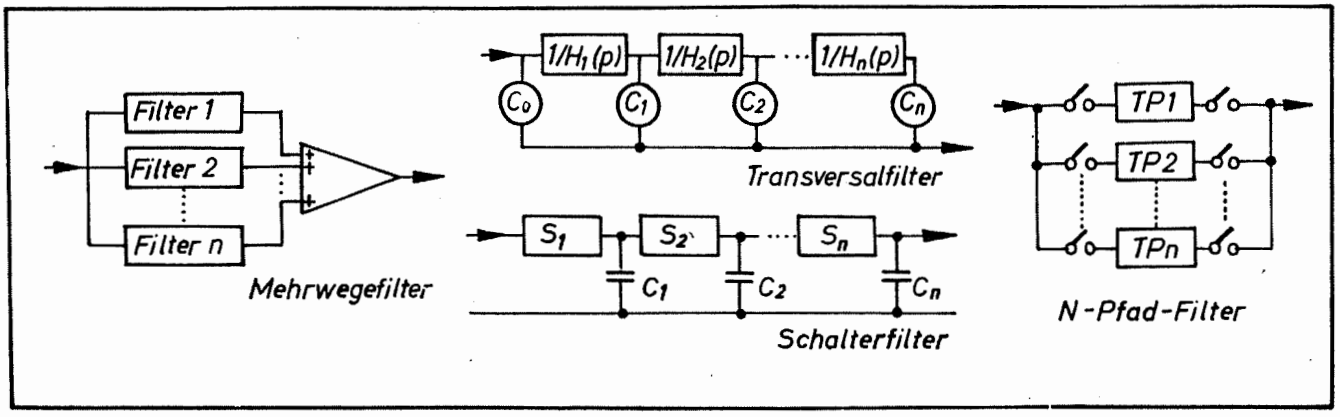


Bild 7.9.1. Prinzipielle Strukturen wichtiger Filtertypen

Die Realisierung erfolgt meist auf der Basis von CCD- und BBD-Schaltungen. Schließlich gehen auf dieses Prinzip auch die Digitalfilter zurück. Dies ist der Grund, daß dieses Prinzip in den nachfolgenden Abschnitten etwas ausführlicher behandelt wird.

Das *Schalterfilter*, auch *SC-Filter* (engl. switched capacitor) genannt, hat formal eine ähnliche Struktur wie die Transversalfilter, seine Wirkungsweise ist jedoch völlig verschieden. An den Stellen S_i befinden sich Schalter, welche getaktet und z. T. phasenverschoben die Kapazitäten umladen. Dadurch fließen bestimmte Pulsströme, welche wie z. B. Widerstände und Induktivitäten wirksam werden. Die Theorie dieser Filter ist nicht einfach, ihre praktische Bedeutung wächst z. Z. jedoch beachtlich, da dieses Prinzip gut integrierbar ist. Es steht damit in ernstzunehmender Konkurrenz zu den Transversalfiltern auf Basis der CCD-Technik.

Die *N-Pfad-Filter* verwenden ebenfalls Schalter, die zwischen verschiedenen Tiefpaßfiltern — meist *RC-Filtern* — periodisch umschalten. Hierdurch lassen sich schmalbandige Bandpässe und Bandsperren realisieren. Dieser Filtertyp wurde vor 1970 intensiv bearbeitet, seitdem ist seine Bedeutung jedoch erheblich zurückgegangen.

Schließlich seien hier noch die *ETC-Filter* (engl. electrothermal circuits) genannt. Sie sind an keine der bisher gegebenen Strukturen gebunden. Ihre Funktion beruht vielmehr auf der Wechselwirkung zwischen thermischen und elektrischen Eigenschaften des Siliciums. Dabei lassen sich Filter im Bereich von ca. 1 bis 1000 Hz realisieren.

7.9.2. Grundlagen der Verzweigungsnetzwerke

Der Begriff des Verzweigungsnetzwerks wurde von SCHÜSSLER (siehe auch Frequenz 22 (1968) H. 8) als Verallgemeinerung für viele andere Strukturen eingeführt. Er berücksichtigt u. a. Analogrechenschaltungen, Echoentzerrer, LEE-WIENER-Filter, Transversalfilter, Oberflächenwellenfilter (OFW) und Digitalfilter. Er führte dabei *vier* verschiedene *kanonische Formen* ein, von denen die ersten beiden hier in Bild 7.9.2 gezeigt werden. Alle diese kanonischen Formen sind, ähnlich wie im Abschnitt 2.4., bei *LC*-Filtern, äquivalent, bedienen sich nur unterschiedlicher Gleichungen und Strukturen.

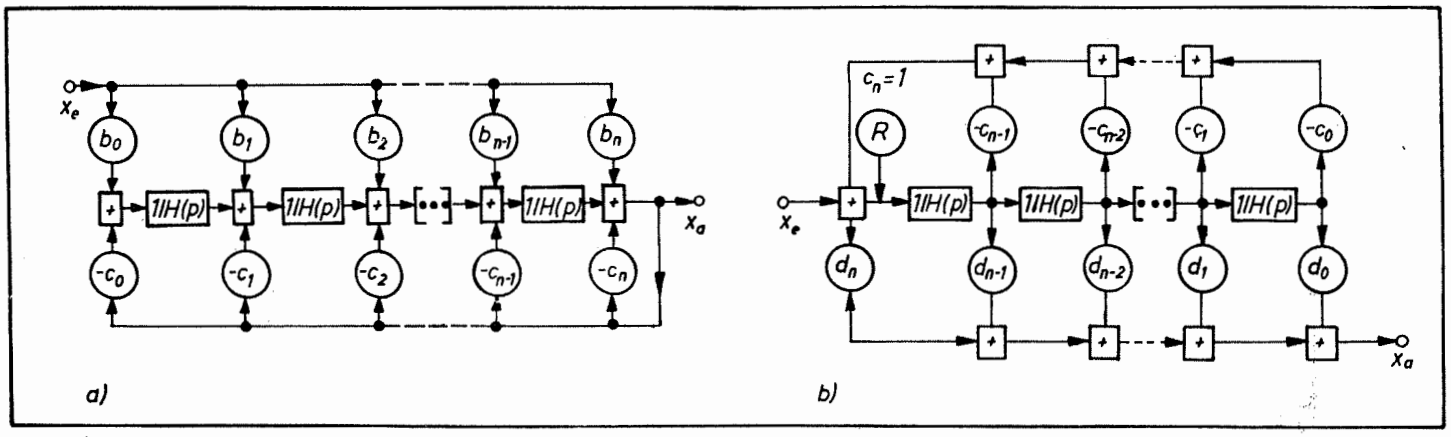


Bild 7.9.2. Verzweigungsnetzwerk a) erster und b) zweiter Art

Die Verallgemeinerung der zweiten kanonischen Form (Bild 7.9.2b)) gegenüber dem einfachen Transversalfilter von Bild 7.9.1 wird ersichtlich, wenn die Koeffizienten c_i Null gesetzt werden. Für ein *Verzweigungsnetzwerk* werden *folgende Bauelemente* verwendet:

- *Addierglieder*, welche Signale rückwirkungsfrei zusammenfügen;
- Koeffizientenglieder, welche die Signale um einen bestimmten, festen bzw. einstellbaren Faktor abschwächen;
- *Teilfilter* $1/H(p)$, die meist sehr einfach aufgebaut und untereinander gleich sind.

Die Übertragungsfunktion der Teilfilter bestimmt neben der Struktur wesentlich die Eigenschaft des Gesamtfilters. *Zwei Funktionen* sind von *besonderer Bedeutung*:

- *Verzögerungsglied* mit der Zeit τ, z. B. bei den OFW bzw. bei getakteten Schaltungen die Speicherzeit zwischen zwei Takten;
- *Integratorglied* mit $H(p) = p$, vor allem bei rein analogen Schaltungen.

Je nachdem, ob alle Koeffizienten c_i in Bild 7.9.2 Null sind oder nicht, unterscheidet man

nichtrekursive und *rekursive Filter*.

Die nichtrekursiven Filter sind die Transversalfilter. Die rekursiven Filter sind allgemeiner. Dies zeigt allein schon eine Betrachtung der Stabilität. Dabei sei vorausgesetzt, daß die Teilfilter selbst stabil sind. Bei den nichtrekursiven Filtern kann das Ausgangssignal nach Abklingen des Eingangssignals grob maximal n-mal die Ausschwingzeit des einzelnen Teilfilters dauern. *Beim rekursiven Filter dagegen besteht die Möglichkeit*, daß Teile des *Ausgangssignals wieder zum Eingang* gelangen. Dies kann dazu führen, daß nach Abklingen des Eingangssignals ein *periodisches Ausgangssignal* entsteht.

Allgemein läßt sich zeigen, daß ein rekursives Filter durch einen nichtrekursiven Teil und einen streng (ausschließlich) rekursiven Teil ersetzt werden kann.

7.9.3. Grundgleichungen

Unmittelbar aus Bild 7.9.2a) folgt

$$x_a = \sum_{\nu=0}^{n} b_\nu \big(1/H(p)\big)^{n-\nu} x_e - \sum_{\nu=0}^{n} c_\nu \big(1/H(p)\big)^{n-\nu} x_a . \tag{1}$$

Wird $c_n = 1$ gesetzt und mit $H^n(p)$ multipliziert, so folgt mittels Trennung der Ein- und Ausgangsgrößen

$$H^+(p) = \frac{x_a}{x_e} = \frac{\sum_{\nu=0}^{n} b_\nu H^\nu(p)}{\sum_{\nu=0}^{n} c_\nu H^\nu(p)} . \tag{2}$$

Dabei entspricht $H^+(p)$ der Gesamtübertragungsfunktion, und sofern nur Integrationsglieder für $H(p)$ verwendet werden, ist sofort die Gleichheit mit Gl. (2.4.5) zu erkennen. Von den Indizes abgesehen, entsprechen sich also die Koeffizienten b_i direkt und die c_i den a_{n-i}. Für andere Teilfilterfunktionen muß noch erst die Auflösung nach den Potenzen von p in Gl. (2) vollzogen werden. Aber als wesentlich bleibt festzustellen, daß die Verzweigungsnetzwerke gemäß Bild 7.9.2a) den allgemeinen Systemen von *RCL*-Gliedern entsprechen und damit diese auch in ihrem Verhalten äquivalent sind.

Für die zweite kanonische Form in Bild 7.9.2b) gelten an den beiden Summiergliedern und bezüglich des Signals x am Eingang des ersten Teilfilters die beiden Gleichungen

$$x_a = x \Big[\sum_{\nu=0}^{n} d_\nu \big(1/H(p)\big)^{n-\nu}\Big] , \tag{3}$$

$$x = x_e - x \Big[\sum_{\nu=0}^{n-1} c_\nu \big(1/H(p)\big)^{n-\nu}\Big] . \tag{4}$$

Werden $c_n = 1$ gesetzt und über x die beiden Gleichungen ineinander eingesetzt, so folgt nach Umnumerierung der Indizes

$$H^+(p) = \frac{x_a}{x_e} = \frac{\sum_{\nu=0}^{n} d_{n-\nu} \big(1/H(p)\big)^\nu}{\sum_{\nu=0}^{n} c_{n-\nu} \big(1/H(p)\big)^\nu} , \tag{5}$$

also wieder ein Ausdruck, welcher der ersten kanonischen Form entspricht.

7.9.4. Prinzip der Abtastfilter

Mit den vorangegangenen Betrachtungen wurde gezeigt, daß die Verzweigungsstrukturen mit analogen Teilfiltern zu Übertragungsfunktionen führen, die identisch mit den aus der *RCL*-Technik bekannten sind. Eine neue Qualität tritt auf, wenn mit *Abtastwerten zu bestimmten Zeittakten* und mit Verzögerungsgliedern an der Stelle von $1/H(p)$ gearbeitet wird. Dies sind die Abtastfilter.

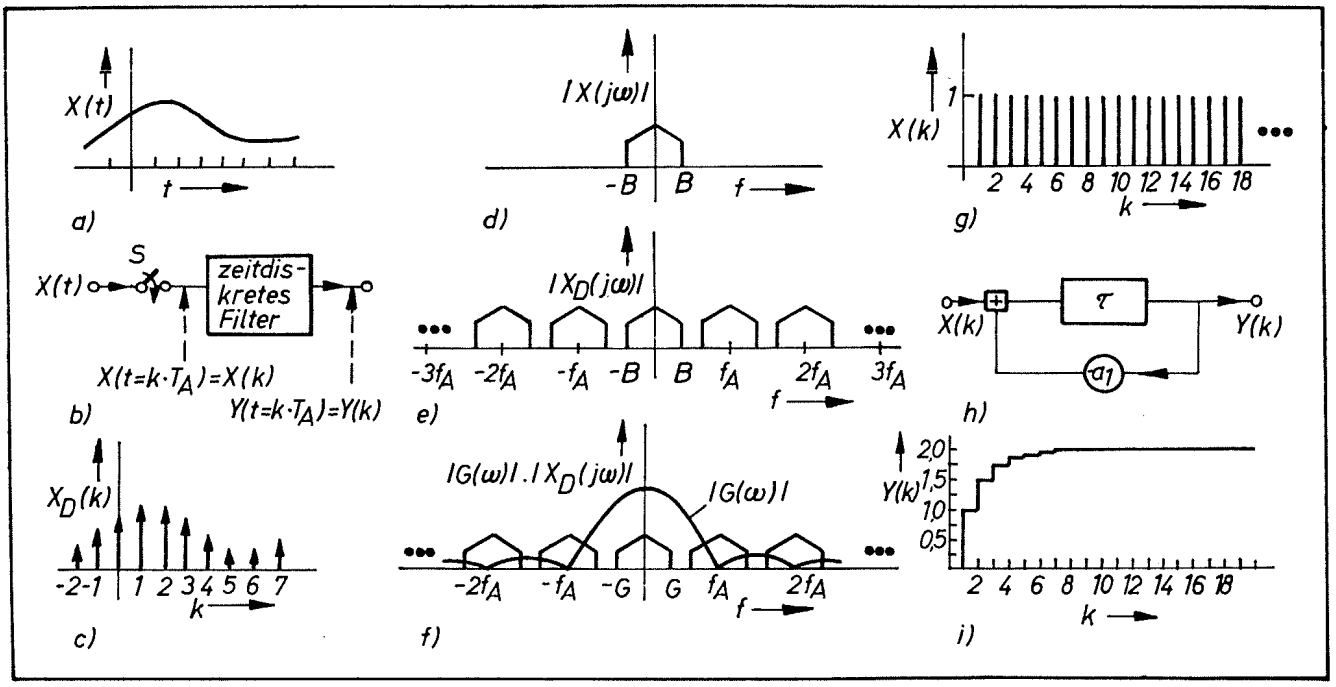


Bild 7.9.3. Zur Wirkungsweise von Abtastsystemen

a) Angenommener Verlauf eines Eingangssignals
b) Wirkung des Schalters, der die Abtastungen am Eingang vornimmt. Es entstehen Folgen von Signalwerten, die über die Abtastperiode T_A nur durch die natürliche Zahl k bestimmt sind
c) Aus dem Eingangssignal abgeleitete Signalfolge $X_D(k)$, wobei D andeuten soll, daß es sich um Delta-Impulse handelt
d) Spektrum des kontinuierlichen Eingangssignals a)
e) Spektrum der diskreten Eingangsfolge c)
f) Multiplikative Veränderung des diskreten Spektrums mit der Gewichtsfunktion $G(\omega)$, wenn die Abtastimpulse eine endliche Länge besitzen
g) Eingangsfolge konstanter Amplitude für $k \geqq 1$
h) Rekursives Abtastsystem 1. Ordnung mit der Verzögerung τ, also um einen Takt
i) Einschwingverhalten des Systems h) mit $a_1 = -0{,}5$ bei der Eingangsfolge g)

Bild 7.9.3b) zeigt das Grundprinzip eines Abtastfilters. Das kontinuierliche Eingangssignal $X(t)$ gemäß Teilbild a) wird mittels des Schalters S zu genau definierten Taktzeiten kT_A (mit k ganzzahlig) weitergeleitet. So entsteht die Folge von Abtastwerten, wie sie Teilbild c) zeigt. Mittels der FOURIER-Analyse läßt sich zeigen, daß *durch eine solche Abtastung das Spektrum* von $X(t)$, also $X(j\omega)$, in Teilbild d) *sich periodisch wiederholt.* Dies demonstriert der Übergang zum Teilbild e). Dabei ist allerdings vorausgesetzt, daß die abgetasteten Impulse DIRAC-Impulse sind, also ihre Zeitdauer unter Beibehaltung der Impulsfläche gegen Null konvergiert. Besitzen die Impulse dagegen eine endliche Länge T_B und Höhe A, so wird dem periodischen Spektrum die *Gewichtsfunktion*

$$G(\omega) = A \frac{\sin 2\pi T_B/T_A}{T_B/T_A} \tag{6}$$

eines solchen Impulses gemäß Teilbild f) multiplikativ überlagert, und die entfernteren periodischen Spektralkomponenten erhalten dadurch eine geringere Intensität.

Um aus den Abtastwerten wieder das Eingangssignal zu erhalten, müssen *alle „störenden" Seitenspektren* am Ende eines Abtastfilters *durch einen Tiefpaß abgetrennt* werden. Bei der Signalverarbeitung innerhalb des Abtastfilters sind sie jedoch notwendig.

Zur weiteren *Veranschaulichung* des Geschehens bei Abtastfiltern werde eine besonders einfache Struktur eines rekursiven Systems (1. Ordnung) gemäß Bild 7.9.3h)

angenommen. Es besitzt nur einen Koeffizienten a_1. Im folgenden gelte $-1 < a_1 < 0$. Der Zusammenhang zwischen Ein- und Ausgangssignalen kann unmittelbar über die Taktung mit der Verzögerung um einen Takt gemäß dem Bild geschrieben werden:

$$Y(k+1) = X(k) - a_1 Y(k) . \tag{7}$$

Wird nun ein konstantes Eingangssignal, also die Folge

$$X(k) = 1 \quad \text{für} \quad k = 1, 2, 3, \ldots , \tag{8}$$

vorausgesetzt, so entsteht für das Ausgangssignal eine geometrische Reihe, weil immer erneut ein Teil des zuletzt vorhandenen Ausgangssignals zum Eingangswert addiert wird. Sie kann leicht aus der Auflösung der rekursiven Gl. (6) abgeleitet werden,

$$Y(k+1) = \sum_{\nu=0}^{k} (-a_1)^\nu \tag{9}$$

bzw.

$$Y(k) = \frac{1 - a_1^k}{1 - a_1} , \tag{10}$$

und besitzt den Grenzwert

$$Y(k)_{k\to\infty} = a_\infty = \frac{1}{1 - a_1} . \tag{11}$$

Folglich gilt

$$Y(k) = a_\infty (1 - \mathrm{e}^{k \ln a_1}) . \tag{12}$$

Dies ist, von diskreten k abgesehen, der Einschwingvorgang eines RC-Tiefpasses gemäß Gl. (1.5.14), wodurch mit der Zeitkonstanten $T = RC$ gilt

$$k \ln a = -t/T . \tag{13}$$

Sie wird also durch die Taktzeit und den Koeffizienten a_1 festgelegt.

Den diskreten Verlauf des Ausgangssignals für ein $a_1 = -0{,}5$ zeigt Bild 7.9.3i).

7.9.5. Beschreibung mittels z-Transformation

Bei den Abtastsystemen sollen nach den vorangegangenen Betrachtungen die Teilfilter in den Verzweigungsnetzwerken (Bild 7.9.3) *reine Verzögerungsglieder* sein. Sie müßten dementsprechend eine *Übertragungsfunktion*

$$H(p) = \mathrm{e}^{-pT} \tag{14}$$

besitzen. Sie ist mit diskreten Bauelementen nur näherungsweise in einem begrenzten Frequenzbereich zu realisieren. Für Abtastsysteme genügt jedoch ein *Kondensator, der einen Funktionswert* während dieser Zeit, d. h. bis zum nächsten Takt, *speichert.*

Mit dem Exponentialfaktor von Gl. (14) geht die FOURIER-Transformation in die LAPLACE-Transformation über. Eine weitere formale Vereinfachung tritt ein, wenn ein z gemäß

$$z = \mathrm{e}^{pT} \tag{15}$$

eingeführt wird. Zunächst ist es dann sinnvoll, auch die Übertragungsfunktion mit dem Argument z zu schreiben:

$$H(p) \Rightarrow F(1/z) . \tag{16}$$

Damit kann die aus der *diskreten* LAPLACE-*Transformation* hervorgehende z-Transformation wie folgt geschrieben werden:

$$F(1/z) = \sum_{n=0}^{\infty} Y_n z^{-n} \tag{17}$$

bzw. als Umkehrung

$$Y_n = \frac{1}{n!} \frac{\mathrm{d}^n\big(F(1/z)\big)}{\mathrm{d}z^n}\bigg|_{z=0}. \tag{18}$$

Damit lassen sich nun alle mittels der Verzweigungsstrukturen realisierbaren *Übertragungsfunktionen* als Funktion von z schreiben:

$$F(z) = \frac{\sum_{\nu=0}^{m} A_\nu z^\nu}{\sum_{\nu=0}^{n} B_\nu z^\nu}. \tag{19}$$

Formal stimmt sie mit Gl. (2) bzw. (5) für die Übertragungsfunktionen bei kontinuierlichen Systemen überein. Gl. (19) ist aber infolge von Gl. (15) *eine transzendente Funktion.* Der Übergang zwischen beiden ist anschaulich als Transformation in der GAUSSschen Ebene interpretierbar (Bild 7.9.4). Sie ist durch *folgende Eigenschaften* gekennzeichnet:

- Das Innere des Einheitskreises der z-Ebene wird (unendlich oft) auf die linke p-Halbebene abgebildet.
- Im Sinne der mehrblättrigen RIEMANNschen Fläche entsprechen einem Punkt in der z-Ebene unendlich viele Punkte der p-Ebene. Die Punkte besitzen die gleiche σ-Koordinate und wiederholen sich periodisch im Abstand $2\pi/T$ (vgl. auch Bild 7.9.3e)).
- Der Ursprung der p-Ebene entspricht dem Punkt $z = 1$.

Mittels dieser Transformation lassen sich auch die Betrachtungen der Pole und Nullstellen in die z-Ebene übertragen. Dies wird am Beispiel verschiedener Systeme eben-

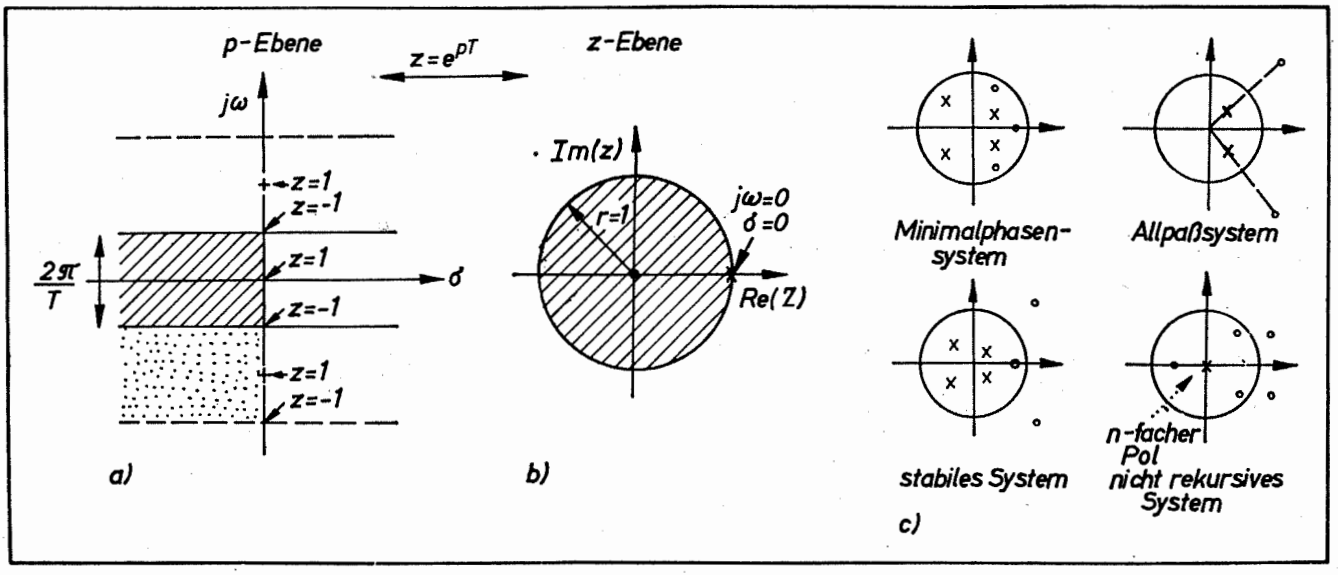


Bild 7.9.4. Zur z-Transformation und zu speziellen Abtastsystemen

a) Der Einheitskreis aus der z-Ebene wiederholt sich periodisch in der linken Hälfte der p-Ebene. Der Grundbereich ist schraffiert, ein weiterer Bereich punktiert

b) Der Einheitskreis in der z-Ebene geht aus einem $2\pi/T$ breiten Streifen der linken Hälfte der p-Ebene hervor. Der Ursprung der p-Ebene bildet sich dabei im Punkt $z = 1$ ab

c) Beispiele von Polen und Nullstellen für spezielle Abtastsysteme

falls in Bild 7.9.4 gezeigt. Daraus wird auch deutlich, daß den nichtrekursiven Systemmen (Transversalsystemen) nur ganz spezielle, unübliche klassische Systeme entsprechen können.

Weitere Aussagen zur Theorie enthalten u. a. [D 5, F 8, F 24, K 26, L 1, S 30, S 66, U 2, U 3].

7.9.6. Beispiel: CCD-Filter

Im Prinzip ist die folgende Beschreibung ohne größere Probleme auf Oberflächenwellenfilter gemäß Abschnitt 6.3.5. und auf BBD-Strukturen gemäß Abschnitt 7.8. übertragbar. Das Prinzip CCD ist also nur ein konkretes Beispiel.

Das vollständige Prinzipschaltbild eines CCD-Filters zeigt Bild 7.9.5. Da CCD-Strukturen einen optimalen Pegelbereich besitzen, wird dieser mit der Eingangsstufe über eine Referenzspannung erzeugt. Gleichzeitig erfolgt in der Eingangsstufe die Abtastung des Eingangssignals. Sie wird durch einen (eventuell extern steuerbaren) Taktgenerator bewirkt. Das Vorfilter blendet die störenden Oberwellenanteile aus (vgl. Bild 7.9.3f)).

Der CCD-Kanal mit den positiven und negativen Gewichten muß die elektronische Nachbildung des Ausdrucks von Gl. (19) bewirken. Dies erfolgt zunächst gemäß Bild 7.9.5b). Zwischen den Taktelektroden φ_1 und φ_2 befindet sich jeweils eine gespaltene Meßelektrode (Splitelektrode). Ihre beiden Teillängen L_+ und L_- geben je einen Strom ab, welcher der Ladung des jeweiligen Abtastwertes und den Teillängen proportional ist. Diese beiden Ströme werden von den einzelnen *Splitelektroden* zusammen dem invertierenden und nichtinvertierenden Eingang eines *Differenzverstärkers* zugeführt. Das Gewicht jeder Splitelektrode ist folglich L_+/L_-. Ist L_+ länger, entsteht ein positives Gewicht, bei längerem L_- ein negatives; sind beide gleich lang, ist das Gewicht Null.

Dem Differenzverstärker folgt eine *Abtast- und Halteschaltung*. Sie bewirkt, daß die einzelnen Impulse bis zum Eintreffen des nächsten bestehen bleiben. So entstand auch die Kurve in Bild 7.9.3i). Eventuell werden anschließend die steilen Übergänge noch mit einem *Tiefpaß* abgerundet.

Das Schema der Splitelektroden gemäß Bild 7.9.5b) hat den Nachteil, daß es bezüglich kleiner Koeffizienten sehr empfindlich ist. Deshalb existieren auch andere Prinzipien, z.B. mit zwei Spalten, wobei der mittlere Teil auf Referenzpotential gelegt wird. Schließlich gibt es auch Schaltungen mit äußerer Einstellung der Empfindlichkeit der einzelnen Elektroden, z.B. über steuerbare Verstärker.

7.9.7. SC-Filter

Schalter-Kondensator-Filter (SC-Filter) sind seit etwa 1952 bekannt. Ihre Theorie wurde 1936 von FETTWEISS in seiner Dissertation entwickelt.

Der *Grundgedanke* besteht darin, daß *beim Abtasten mittels Kondensatoren Ströme auftreten*, deren Hüllkurven als Impedanzen interpretiert werden. So lassen sich z.B. die Grundbauelemente gemäß Bild 7.9.6a) nachbilden. Insbesondere für den Widerstand gibt es noch viele andere Nachbildungen. In der Regel erfolgt bei SC-Filtern nun aber die Realisierung nicht mit den dargestellten Nachbildungen. Es werden viel häufiger *zwei* andere *spezifische Bauelemente*, der *SC-Integrator* und der *SC-Differenzintegrator*, verwendet. Mit ihnen kann nämlich in relativ einfacher Weise eine Kettenleiterschaltung in eine SC-Schaltung umgesetzt werden. Dies demonstriert

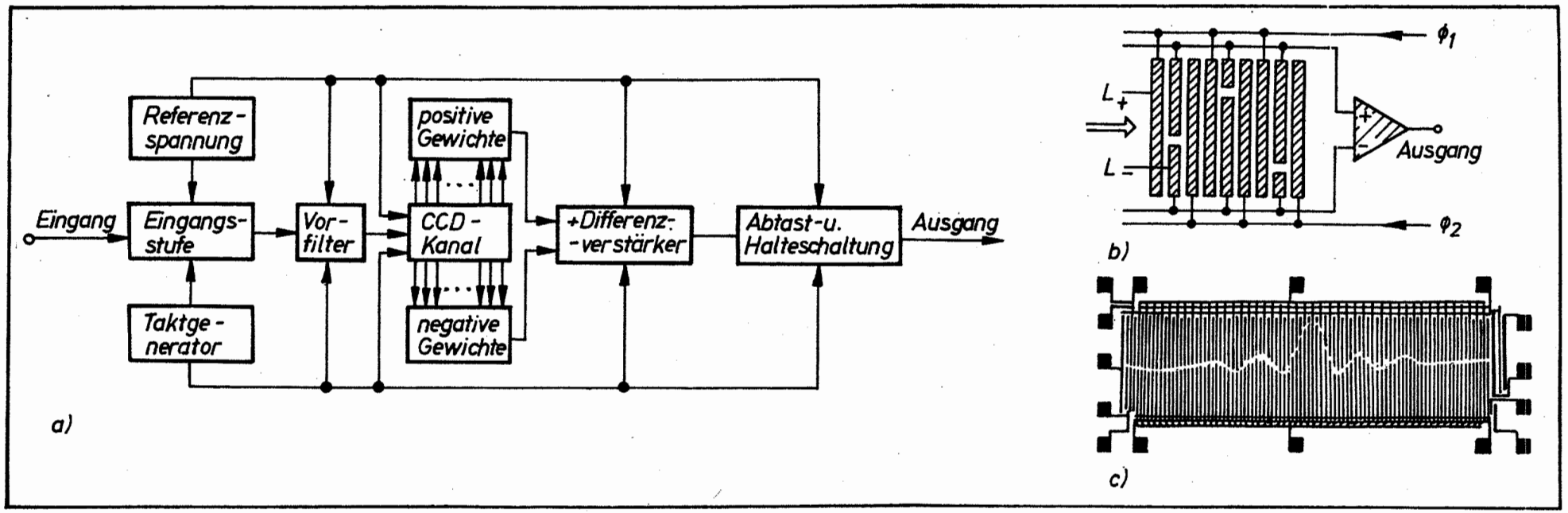


Bild 7.9.5. Beispiel eines CCD-Abtastfilters

a) Prinzipschaltung
b) Prinzip der Splitelektroden mit Zuführung zu einem Differenzverstärker
c) Beispiel einer CCD-Struktur

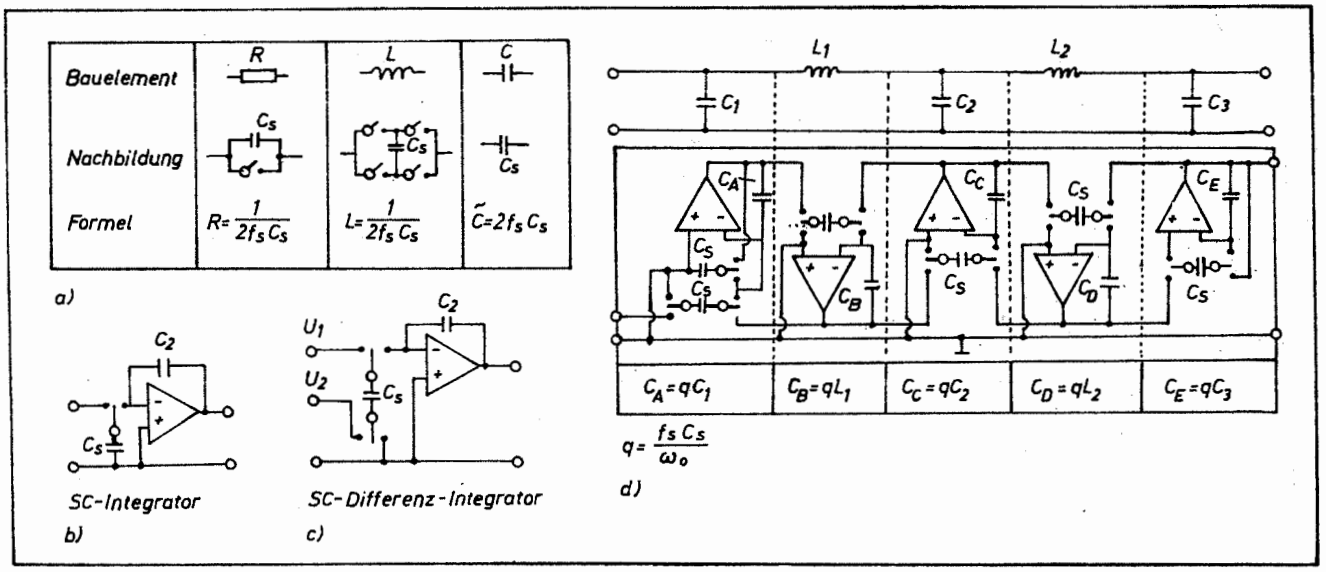


Bild 7.9.6. Zur Gestaltung von Schalter-Kondensator-Filtern (SC-Filter). Es bedeuten C_S = geschalteter Bezugskondensator; f_S = Schaltfrequenz; $q = f_S C_S/\omega_0$ = Bezugswert

a) Erzeugung der drei Grundbauelemente
b) SC-Integrator
c) SC-Integrator mit Differenzeingang
d) Ersatz eines Tiefpasses 5. Ordnung durch einen SC-Filter mit SC-Differenzintegratoren

Bild 7.9.6d) und zeigt, wie relativ einfach solche Strukturen in der MOS-Technik zu realisieren sind. Die Schalter sind dann durch getaktete MOSFET zu ersetzen.

7.9.8. Grober Vergleich der Filtertypen

Ein Vergleich der verschiedenen Filtertypen ist nur relativ komplex durchzuführen, da es unterschiedliche *Bewertungsmöglichkeiten* gibt. Hierzu gehören u. a. ohne Wichtung der Reihenfolge

- Frequenzbereich
- Flankensteilheiten ($f_0/\Delta f$)
- Verluste
- erreichbare Dämpfungen
- Integrierbarkeit
- Temperaturkoeffizient und Langzeitkonstanz
- Volumen und Masse
- Parameterempfindlichkeit.

Dementsprechend fallen auch die unterschiedlichen Vergleiche in der Literatur aus. Hier sei nur beispielhaft der Frequenzbereich herangezogen.

Im tiefen Frequenzbereich bis ca. 1000 Hz bewähren sich immer mehr die *Digitalfilter*. Größere Bedeutung könnten noch die im Abschnitt 7.9.1. genannten *ETC-Filter* erlangen, wenn kontinuierliche Funktionen erwünscht sind. Die *aktiven Filter* gelangen bis zu Frequenzen um 100 kHz. Dies ist auch die Grenze der *mechanischen Filter*, die sich aber vor allem durch hohe Güte und steile Flanken auszeichnen. Die *Abtast-, Schalter- und Speicherfilter* eignen sich besonders im Bereich von ca. 1 kHz bis 1 MHz. *Spulenfilter* überstreichen den Frequenzbereich von ca. 100 kHz bis zu mehreren MHz, realisieren aber nur Flankensteilheiten kleiner etwa 10^3. *Quarzfilter* betreffen etwa den gleichen Frequenzbereich, jedoch wesentlich größere Flankensteilheiten. *Oberflächenwellenfilter* sind besonders im Frequenzbereich von 10^7 bis 10^9 Hz vorteilhaft.

8. Digitale Schaltungen

8.1. Kodes und Zahlendarstellungen

8.1.1. Zum Begriffspaar analog — digital

Jede natürliche Sprache lebt, und daher verändert sich ständig die Bedeutung von Begriffen. Um den Inhalt der Begriffe besser verstehen zu können, ist es oft günstiger, auf die Entwicklung ihres Inhalts zurückzugreifen. Die beiden Begriffe analog und digital haben heute einen im wesentlichen noch undefinierten Inhalt. Jeder Techniker verwendet sie aber meist intuitiv richtig. Sie stehen in einem gewissen Zusammenhang mit den Begriffen binär, diskret, dual, kontinuierlich, quantisiert, stetig. Im folgenden sei versucht, den Inhalt dieser Begriffe etwas vom Ursprung her, aber vor allem im Zusammenhang zu verstehen.

8.1.1.1. Analogie, analog

Der Wortstamm für analog ist das griechische logos, etwa Vernunft. Hierzu gehören auch das griechische Wort logikos, etwa vernunftsmäßig, aus dem unser heutiger Begriff Logik stammt. Da die Logik aber meist binäre, d. h. zweiwertige Logik mit Ja-Nein-Entscheidungen darstellt, weist dieser Stamm auch bereits in das Gegenteil des heutigen Begriffs analog. Heute wird analog, dem griechischen gemäß, meist mit übereinstimmend, angeglichen, angepaßt, vergleichbar interpretiert. Die lateinische Vorsilbe ana- bedeutet auf-, wieder-, aufwärts, nach oben, zurück. So entsteht das lateinische Wort analogia, was soviel wie nach der Vernunft, übereinstimmend, aber auch Gleichmäßigkeit (insbesondere im Zusammenhang mit dem Sprachgebrauch) bedeutet. Aus diesem Stamm drang analog um 1800 in die deutsche Sprache ein.

Das Substantiv von analog ist die Analogie. Sie wird heute außer in der Umgangssprache für Entsprechung, Ähnlichkeit, Gleichwertigkeit und Übereinstimmung in wenigstens drei Fachgebieten genutzt. In der Technik und besonders in der Kybernetik wird dann von Analogie gesprochen, wenn zwei Systeme trotz unterschiedlichem Substrat oder auch unterschiedlicher Struktur funktionell gleichartiges Verhalten zeigen. Ein Musterbeispiel hierfür ist der Analogrechner. Aber auch die elektromechanischen Analogien gehören in dieses Gebiet. Mit ihnen wird z. B. die Ausbreitung von Schallwellen über elektrische Modelle aus Widerständen, Kondensatoren und Induktivitäten besser verständlich gemacht.

Hier wurde im Abschnitt 5.7.6. von der Analogie zwischen Wärmeleitung und stationärem Strom gemäß Tab. 5.7.9 Gebrauch gemacht. In etwa entsprechen auch Ersatzschaltungen und Modelle sowie mathematische Beschreibungen Analogien.

Ähnlich verwendet die Biologie den Begriff Analogie im Sinne von analogen Organen: z. B. Augen der Wirbeltiere, der Tintenfische, der Gliederfüßler (z. B. Insekten) mit stark unterschiedlicher Entwicklung (und Struktur). Auch in der Rechtsprechung

gibt es den Begriff Analogie. Er wird z.B. dann angewendet, wenn ein rechtsmäßig nichtgeregelter Tatbestand auf einen etwa wesensgleichen bezogen wird.

Allgemein kann daraus gefolgert werden, daß analog immer dann richtig angewendet wird, wenn ein Signal, ein Gerät usw. Eigenschaften hat, die in etwa funktionell oder strukturell mit einem zweiten, zu vergleichenden übereinstimmen. Die übliche (Analog-) Uhr entspricht mit ihren sich bewegenden Zeigern der Bewegung des Schattens der Sonnenuhr oder, noch weiter herausgesetzt, dem Planetenmodell. In diesem Sinne ist die Bezeichnung Analoguhr voll berechtigt, ja sogar dann, wenn sich die Zeiger nicht gleichmäßig bewegen, sondern im Abstand von 1/2 oder 1 min weiterspringen. In diesem Sinne sind auch die Mikrofonspannungen dem Schallsignal, das sie hervorruft, analog. Jedoch kann die daraus abgeleitete frequenzmodulierte Schwingung eines UKW-Senders nur noch bedingt als analog bezeichnet werden. In gewissem Sinne gilt das zwar für die jeweilige Momentanfrequenz, aber nicht für das Spektrum. In jedem Fall enthält das Sendesignal aber eine analoge Größe des ursprünglichen Schalls.

8.1.1.2. Kontinuierlich, stetig

Das Wort kontinuierlich ist lateinischen Ursprungs. Continens bzw. continuus bedeuten zusammenhängend, angrenzend an, unmittelbar folgend, ununterbrochen, jemandem zunächst stehend. Continuo: aneinanderfügen, verbinden, fortsetzen, verlängern, gleich darauf, ohne weiteres. Contingo: berühren, kosten, streuen, jemandem nahe sein, beeinflussen.

In der Mathematik wird vom Kontinuum der reellen Zahlen gesprochen. In der Physik ist das Kontinuum besser durch lückenlos zusammenhängend, ohne Berücksichtigung der Mikrostruktur zu erklären (Kontinuumsmechanik). Vielfach wird kontinuierlich auch mit dem Begriff stetig gleichgesetzt. Er stammt von stehen ab und befindet sich in Nachbarschaft von stet, stets und unstet. Umgangssprachlich wird er am besten mit beharrlich, ununterbrochen beschrieben. In der Mathematik wird stetig (in einem Punkt) im Zusammenhang mit dem Funktionsbegriff spezieller benutzt. Es muß der Grenzwert $\lim_{x \to x_0} f(x)$ existieren und gleich $f(x_0)$ sein.

In der Elektronik wären also kontinuierliche Signale dadurch zu kennzeichnen, daß beliebige Zwischenwerte angenommen werden können. Der Begriff stetig ist in diesem Zusammenhang kaum gebräuchlich. Es sei noch betont, daß Signale in zwei Eigenschaftswerten kontinuierlich sein können: bezüglich der Zeit und bezüglich der Amplitude. Das wird dann besonders deutlich, wenn wie im folgenden das Gegenteil ausgesagt wird.

8.1.1.3. Diskret

Das Wort diskret ist lateinischen Ursprungs: discretion = Unterscheidungsvermögen, Urteil und Entscheid; discretus = abgesondert, getrennt; discerno = scheiden, trennen, unterscheiden, beurteilen, entscheiden.

Diskret drang im 16. Jahrhundert in die deutsche Sprache ein. Allgemein bedeutet es heute taktvoll, rücksichtsvoll, zurückhaltend, unauffällig, unaufdringlich, vertraulich, geheim, verschwiegen. In der Physik weist es auf eine Größe hin, die nur in endlichen Schritten veränderlich ist. Es wird dann auch von unstetig gesprochen. In der Mathematik bedeutet diskret in einzelne Punkte oder Elemente zerfallend. Ein diskretes Signal ist damit nur endlich vieler, meist in ihrer Größe genau definierter Werte fähig.

Damit sind vier Arten von Signalen zu definieren. Sie sind in den Werten der Amplitude oder der Zeit jeweils kontinuierlich oder diskret. Diese Möglichkeiten faßt

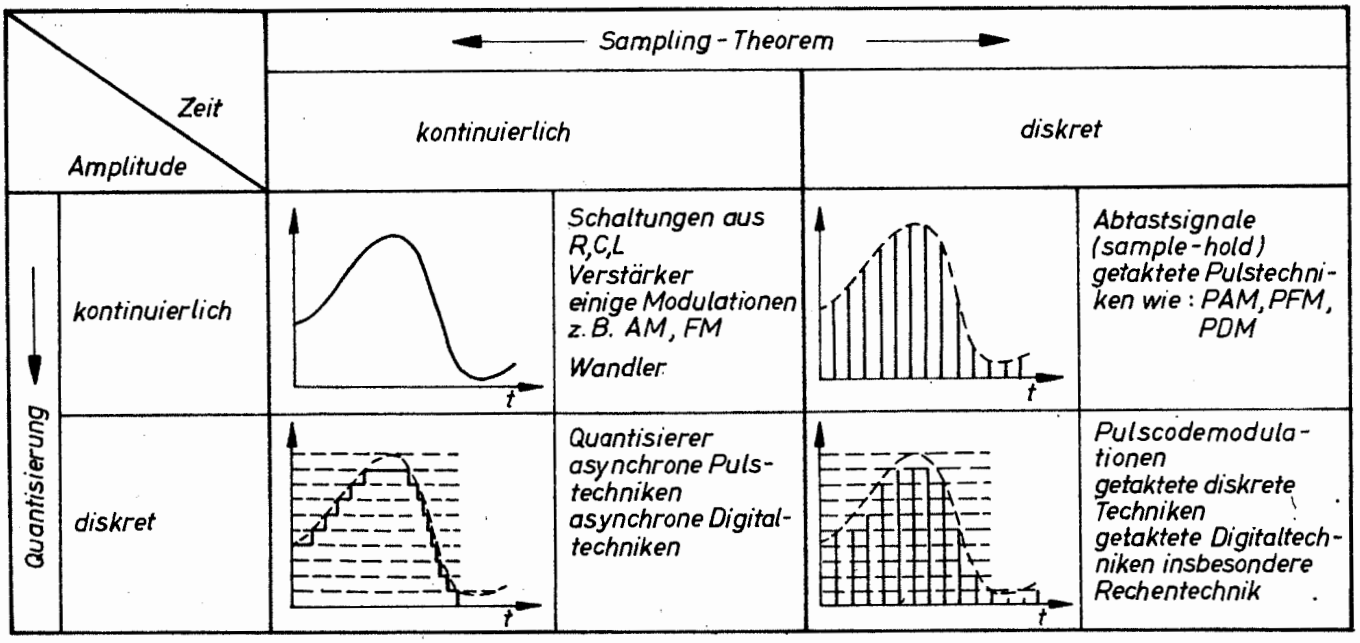


Bild 8.1.1. Einteilung von Signalen bezüglich der beiden Parameter Zeit und Amplitude in die Zustände kontinuierlich und diskret. Bedeutsam ist, daß bezüglich der Zeit gemäß dem Samplingtheorem und entsprechender technischer Realisierungen bandbreitenbegrenzte Signale exakt zwischen diskreten und kontinuierlichen Zeitwerten in beide Richtungen gewandelt werden können. Bezüglich der Amplitudenwerte ist nur der nichtumkehrbare Übergang von kontinuierlich nach diskret möglich

anschaulich Bild 8.1.1 mit Hinweisen auf entsprechende Abschnitte in diesem Buch zusammen.

8.1.1.4. Quantisiert

Auch quantisiert ist lateinischen Ursprungs; es kam im 16. Jahrhundert in die deutsche Sprache: quantitas = Größe, Anzahl; quantum = wieviel, so viel wie, inwieweit, irgendwie.

In der Physik erlangte das Wort Quant durch die Untersuchungen von PLANCK im Sinn von Lichtquant, Energiequant besondere Bedeutung, und es wurde vielfach angewendet, z.B.: Quantenphysik, Quantenmechanik, Quantenchemie, Quantenbiologie, aber auch Quantentheorie, Quantenzahl. Ferner ist auch das dialektische Paar Quantität — Qualität für den Begriffsinhalt von quantisiert zu beachten.

Zwischen quantisiert und diskret besteht also vom Inhalt her kein wesentlicher Unterschied. In der Nachrichtentechnik bezeichnet diskret jedoch mehr den Zustand des abgestuften Signals, während quantisiert mehr den Vorgang vom kontinuierlichen zum diskreten Signal, vor allem in den Amplituden, betrifft. (Vgl. 7.2.3.) Das liegt allein darin begründet, daß die entsprechenden Anordnungen Quantisierer heißen. An ihrem Eingang steht also ein kontinuierliches Signal, am Ausgang ein diskretes Signal. In der Zeitachse heißen die entsprechenden Anordnungen dagegen Abtastschaltungen (sample and hold, vgl. 7.3.6.). Sie erzeugen zeitdiskrete, aber nicht zeitgequantelte Signale. Dem Quantitativen haftet zusätzlich noch die Meßbarkeit an. Quantisieren heißt daher etwa auf einen (von endlich vielen) genau definierten Meßwert bringen. Diskret enthält dagegen nur das Gegenteil von stetig, kontinuierlich.

8.1.1.5. Digital, dual, binär

Es sei betont, daß diese drei Worte im Deutschen umgangssprachlich nicht gebräuchlich sind. Sie sind also Fachtermini. Digital stammt vom lateinischen digitus = Finger.

Digitus bedeutet aber auch Fingerbreite als Maß mit 18,5 mm. Es hängt also mit zählen, ziffernmäßig, in Zahleneinheiten zusammen. Auch das Bit ist gemäß dem englischen binary digit = Binärzahl als hieraus abgeleitet zu verstehen (digit (engl.) = Finger, Fingerbreite, Ziffer). Dual und binär stammen von Zahlwörtern. Im Prinzip gibt es davon fünf Gruppen:

- Grundzahlen (Kardinal-), z.B. eins, zwei, drei, . . .
- Ordinalzahlen, z.B. erstens, zweitens, drittens, . . .
- Verteilungszahlen (Distributiv-) z.B. je ein, je zwei, je drei, . . .
- Vervielfältigungszahlen, z.B. einfach, zweifach, dreifach, . . .
- Zahladverbien, z.B. einmal, zweimal, dreimal, . . .

Im Lateinischen bedeutet duo die Kardinal- und bini die Distributivzahl für zwei. Bi- ist außerdem die Vorsilbe für zwei, ähnlich dem Deutschen zwie-, und binär bedeutet aus zwei Teilen bestehend, zweigliedrig. Die Ableitung von duo kommt auch in Duett, Duell und Dualismus (z.B. von Materie und Welle) vor. Aus diesem Grund ist der überwiegende Gebrauch der Begriffe binär und dual sinnfällig. Bei binären Systemen wird sowohl für die verwendeten Bauelemente, Signale usw. als auch für das Zahlensystem die Basis 2 verwendet. Bei dualen Systemen sind dagegen nur die Bauelemente, Signale usw. zweiwertig. Die Basis des Zahlensystems kann dagegen beliebig sein.

8.1.1.6. Struktur der Begriffe

Aus dem Gesagten ergibt sich der Zusammenhang der Begriffe von Bild 8.1.2. Analog bezieht sich also auf das abgebildete Original, und das Gegenteil ist daher nichtanalog.

Unabhängig davon kann eine Einteilung nach kontinuierlich und diskret erfolgen. Sie betrifft die Signalamplituden und/oder die Zeit gemäß Bild 8.1.1. Aus den beiden unabhängigen Begriffspaaren analog — nichtanalog und kontinuierlich — diskret läßt sich eine Matrix mit vier Feldern erzeugen, wie sie Tab. 8.1.1 zeigt. Sie weist deutlich aus, daß es zwischen den analogen Techniken im engen Sinne und den digitalen (d. h. auf Zahlen bezogenen) noch eine Vielzahl anderer Techniken gibt, die mit dem üb-

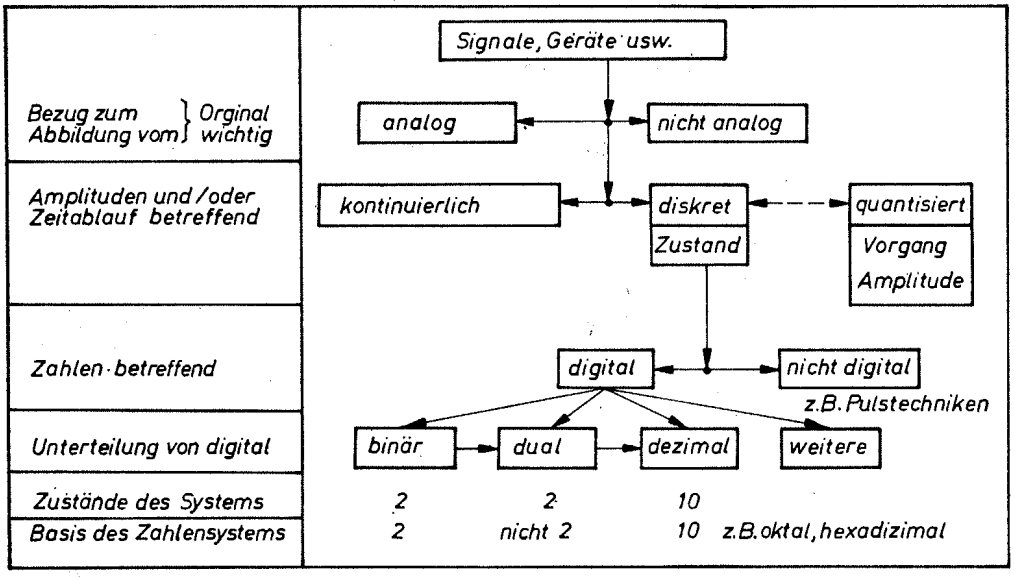


Bild 8.1.2. Strukturzusammenhang von analogen und digitalen Techniken im engen Sinne. Damit existieren im strengen Sinne Techniken, die weder analog noch digital sind. Ein Teil davon sind diskrete Techniken, die nicht als Zahlenkodierungen interpretierbar sind. Der andere Teil betrifft kontinuierliche Signale, die nicht als analoge Abbildung aufzufassen sind

Tabelle 8.1.1. Zum Zusammenhang von analog und kontinuierlich. Die Einteilung erfolgt nach dem Abbildungsverhältnis und nach den Energiezuständen der Signale. Eingetragen sind Beispiele.

Abbildung	Energiezustände	
	kontinuierlich	diskret
analog	„analoge" Technik im engen Sinne, z. B. Audiotechnik, Oszillograph, mechanische Uhr	quantisierte Signale, Fernsehen infolge der Zeilenstruktur, Rasterdruck bei Bildern, Bahnhofsuhr mit Minutensprung
nichtanalog	nur wenige Beispiele vorhanden, z. B. visible speach, Frequenzmodulation	alle codierten Darstellungen, also auch mit Zahlen und Buchstaben, z. B. viele Pulsmodulationen, km-Zähler, Digitalvoltmeter, aber auch Schriftsprache und im gewissen Umfang infolge der Phoneme Sprache überhaupt

lichen Begriffspaar analog — digital nicht eindeutig klassifiziert werden. Deshalb hat sich z. T. auch schon der Begriff diskrete Signale, diskrete Techniken (nicht: diskrete Bauelemente, die das Gegenteil der integrierten sind) durchgesetzt. Beispiele hierzu waren u. a. in den Abschnitten 6.1.5., 6.1.6., 6.1.7., 6.2.3., 6.2.4., 6.4., 6.5.4., 7.2.2., 7.2.3., 7.3.6., 7.3.7.2., 7.3.8., 7.4.6.2., 7.5.7.4., 7.5.7.6., 7.5.7.8., 7.6.6., 7.8.

Das Diskrete (und nur dies) kann dann in digital und nichtdigital unterteilt werden, je nachdem, ob auf ein Zahlensystem übergegangen wird oder nicht. Schließlich läßt sich (je nach der Basis des Zahlensystems und der Zustandszahlen) eine weitere Einteilung von digital vornehmen. Wichtige Größen sind hier insbesondere 3 (ternär), 8 (oktal) und 16 (hexadezimal, auch sedezimal oder kurz hex-Zahl).

8.1.1.7. Vorteile analoger und digitaler Technik

In der letzten Zeit hat die digitale Technik rasant an Anwendungen zugenommen. Ein Grund hierfür ist, daß es jetzt mittels der Mikroelektronik sehr viel einfacher ist, komplizierte Schaltungen ökonomisch aufzubauen. Dennoch wird es *nicht möglich sein*, die *Elektronik vollständig zu digitalisieren*. Hierfür gibt es viele Gründe, z. B. ist der Mensch in erster Linie gewohnt, seine *Umwelt analog wahrzunehmen*. Die digitale Uhr mit analoger Anzeige ist dafür nur ein einfacher Hinweis. Die meisten physikalischen Größen besitzen kontinuierliche Eigenschaften, denen das analoge Messen besser als das digitale angepaßt ist. Zumindest wird also der primäre Wandler analog, bestenfalls diskret sein. Darüber hinaus gilt das seit 1932 bekannte GOEDEL-*Theorem*, wonach es in einem hinreichend großen System stets Aussagen gibt, die weder wahr noch falsch sind. In solchen Systemen versagen also Möglichkeiten der Berechenbarkeit. Sie sind nicht mit *Ja-Nein-Entscheidungen*, d. h. über Zahlen (digital), ausreichend zu erfassen. Solche allgemein gültigen Aussagen machen es immer wichtiger, die Vor- und Nachteile der beiden Techniken möglichst deutlich zu zeigen. Wenn dies gelingt, existieren automatisch auch Aussagen für die Grenzgebiete des Diskreten und nichtanalogen Kontinuierlichen. Tab. 8.1.2 gibt einen groben Überblick zu einigen wichtigen Eigenschaften. Etwas ausgeführt sei noch die mögliche *Meßgenauigkeit*. Analoganzeigen werden in der Regel über Längen realisiert. Bei Zeigerinstrumenten gleitet dabei der Zeiger über eine Skale. Ihre Länge L wird meist zwischen 2 und 20 cm liegen (Bild 8.1.3a)). Der kleinste feststellbare Ausschlag A beträgt dagegen etwa 0,05 bis 1 mm. Der kleinere Wert erfordert bereits Lupen-

Tabelle 8.1.2. Eigenschaften von digitaler und analoger Technik

	Analog	Digital
Physikalischer Bezug	kontinuierlich veränderliche Größe	definierte diskrete Zustände
Anzeige	bevorzugt durch Länge oder Weg	durch Ziffern, evtl. Buchstaben oder Symbole (z. B. +, −)
Anzeigequalität	anschaulich z. B.: zu schnell, zu kalt, ziemlich voll	abstrakt
Erkennen von Änderungen, Tendenzen usw.	gut	schlecht
Kurvendarstellung	unmittelbar durch Schreiber	erst nach Umsetzung
Grenzwertschaltung	leicht realisierbar	schwieriger
Genauigkeit	% bis ‰	weit steigerbar
Bevorzugt entstehende Ablesefehler	Interpolationsfehler	grobe Fehler
Aufwand	kleiner	größer
Messen und Stellen	primär analog	meist indirekt
Geschwindigkeit	größer	kleiner
Fehlerwirkung	ungenauer vor Ausfall	meist sofort Totalausfall

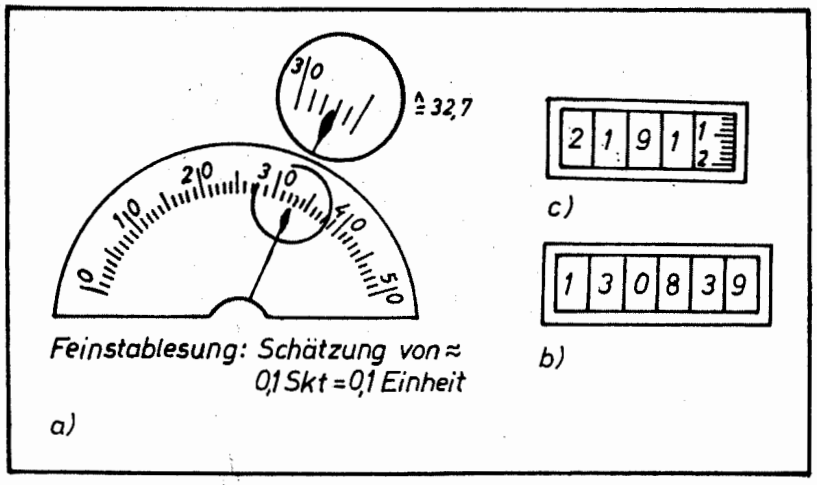


Bild 8.1.3. Zum Vergleich von a) analoger und b) digitaler Anzeige, c) eine Kombination beider

ablesung. Die maximale Ablesegenauigkeit berechnet sich (bei sonst richtiger Anzeige) zu

$$G = A/L\,. \tag{1}$$

Sie liegt mit den obigen Bestwerten bei max. 0,25‰. Allein hieraus ist ersichtlich, daß es immer aufwendig ist, Analoganzeigen mit Genauigkeiten von besser als 1% zu realisieren. Ganz anders sind die Verhältnisse bei digitaler Anzeige (Bild 8.1.3b)). Hier ist es prinzipiell immer möglich, weitere Stellen hinzuzufügen. In einigen Fällen hält sich als Kompromiß noch die kombinierte digitale und analoge Anzeige gemäß Bild 8.1.3c). Beispiele sind Gas- und Stromzähler sowie das Tachometer.

8.1.2. Zahlensysteme

Wir sind es gewohnt, dezimal zu zählen. Eine Zahl Z wird dabei mit den *Ziffern* $C = 0, 1, 2, \ldots, 9$ geschrieben. Darüber hinaus existiert ein *Stellenwert*. Eine Zahl kann

dementsprechend wie folgt geschrieben werden:

$$3\,5\,8\,2 = 3 \cdot 10^3 + 5 \cdot 10^2 + 8 \cdot 10^1 + 2 \cdot 10^0 . \tag{2}$$

Rein theoretisch, und mit der Digitaltechnik auch praktisch, kann dieses Prinzip statt der obigen Basis 10 auf *andere Basiswerte B*, z.B. 2, 8, 16, übertragen werden:

$$Z = C_n B^n + C_{n-1} B^{n-1} + \ldots + C_1 B^1 + C_0 . \tag{3}$$

Bei der sich daraus ergebenden *Schreibweise* können sich Probleme ergeben, wenn die Basis in der Zifferndarstellung nicht bekannt ist. Deshalb hat es sich eingebürgert, sie als speziellen Buchstaben anzufügen. Einen Überblick hierzu gibt Tab. 8.1.3. Darin fällt weiter auf, daß das Zahlzeichen Null, um es mit Sicherheit gegen den Buchstaben O (wie Otto) abzugrenzen, schräg durchgestrichen wird. Aus demselben Grunde wird als Kennzeichen für eine Oktalzahl zuweilen auch Q statt 0 verwendet. Die Bezeichnung 0 und L für Binärziffern ist inzwischen weitgehend verschwunden. Bei den Hexadezimalzahlen existieren keine Zahlzeichen für 10 bis 15. Hier werden jetzt allgemein die Buchstaben A bis F verwendet. Wenn dann aber eine Hex-Zahl mit einem Buchstaben wie im Beispiel von Tab. 8.1.3 beginnt, ist es wichtig, eine Ø vorzuschreiben, damit erkannt wird, daß es eine Zahl und kein Text ist.

Im Sinne des Beispiels gilt jetzt auch

$$1101\ 1111\ 1110\,\mathrm{B} = 3582\,\mathrm{D} = \text{ØDFE}\,\mathrm{H} . \tag{4}$$

Tabelle 8.1.3. Kennzeichen von Ziffern und Zahlen

Basis	Name	Kennzeichen	Ziffern	Zahlenbeispiel
2	Binär	B	Ø1	1101 1111 1110 B
3	Ternär	T	Ø12	1122 0200 T
8	Oktal	O oder Q	Ø1234567	6776 O
10	Dezimal	D	Ø123456789	3582 D
16	Hexadezimal (Sedezimal)	H	Ø123456789ABCDEF	ØDFE H

8.1.3. Bezeichnungsaufwand

Aus Tab. 8.1.3 geht deutlich hervor, daß die gleiche Zahl mit einem *unterschiedlichen Ziffernvorrat* (entsprechend der gewählten Basis) und einer dem angepaßten *Stellenzahl n* geschrieben werden kann. Das Produkt aus beiden wird *Bezeichnungsaufwand* der Zahlendarstellung genannt:

$$y = n \cdot B . \tag{5}$$

Er beträgt bei der Zahl 3582 D folglich:

$$12 \times 2 = 24(\mathrm{B}); \qquad 8 \times 3 = 24(\mathrm{T}); \qquad 4 \times 8 = 32(\mathrm{O});$$

$$4 \times 10 = 40(\mathrm{D}); \qquad 3 \times 16 = 54(\mathrm{H}).$$

Für andere konkrete Zahlen ergeben sich andere Werte, z. T. auch mit Verschiebung der Rangfolge. Für die Praxis ist es wichtig zu wissen, wo, allgemein betrachtet, das Minimum des Bezeichnungsaufwandes liegt. Hierzu ist von der Anzahl der Zahlen N auszugehen, die mit 1 bis n Stellen zur Basis B darstellbar sind. Die Abzählung dieser Reihe ergibt

$$N = B^n \tag{6a}$$

bzw., nach n aufgelöst,

$$n = \frac{\log N}{\log B}, \tag{6b}$$

und, in Gl. (5) eingesetzt, folgt

$$y = \frac{\log N}{\log B} \cdot B . \tag{7}$$

Dieser Wert besitzt sein Minimum, wie man durch Differenzierung nach B und Nullsetzen leicht zeigen kann, bei

$$B = \mathrm{e} \approx 2{,}71828 . \tag{8}$$

Damit ist es naheliegend, die Basis 2 oder 3 als nächstliegende ganze Werte zu verwenden. Die Entscheidung ist auf 2 gefallen (allerdings bevor diese Betrachtung gemacht wurde), weil technische Gebilde besonders gut 2 und weniger gut 3 Zustände stabil annehmen können. Darüber hinaus dürfte auch unser logisches Entscheidungsdenken mit ja — nein, falsch — richtig beteiligt gewesen sein.

8.1.4. Kodierungen

Die Ziffern einer Zahl müssen für die Elektronik elektrischen Signalen und/oder Zuständen zugeordnet werden. Der Einfachheit halber seien hierzu Schalter verwendet. In der Praxis werden jedoch nicht Schalter, sondern elektronische Schaltungen verwendet, die Strom bzw. nicht Strom führen. Hierauf wird im Abschnitt 8.2.3. eingegangen. Für die weiteren Betrachtungen beschreiben wir die „Schalter" mit Bit (bit). Ist ein Bit eingeschaltet, so möge es eine 1, ausgeschaltet eine Ø darstellen. Werden jetzt mehrere Bits in einer Reihe angeordnet, so kann jedem Bit ein Stellwert, z. B. gemäß einer Binärzahl, zugeordnet werden, also z. B.

Schalter:	S_4	S_3	S_2	S_1	S_0
Stellwert:	2^4	2^3	2^2	2^1	2^0
Zahlenwert:	16D	8D	4D	2D	1D

Damit bilden diese Bits genau die Binärzahldarstellung ab. Deswegen wird auch von der Binär- (früher Dual-) Kodierung der Zahlen nach diesem Prinzip gesprochen. Es werden, um die Seite zu kennzeichnen, unterschieden

MSB = most significant bit (höherwertiges Bit)
LSB = least significant bit (niederwertiges Bit).

Für andere Zahlensysteme ist eine so einfache Darstellung wie oben in der Regel nicht möglich, da es keine Bauelemente mit entsprechend vielen Zuständen gibt. So müssen diese Zahlendarstellungen in irgendeiner optimalen Zuordnung auf Bits abgebildet werden. Dieses Verfahren heißt Kodierung und die Abbildungsvorschrift ist der Kode.

8.1.4.1. BCD-Zahlen

Wenn dezimale Zahlen mittels binärer Bauelemente dargestellt werden sollen, ist es vom Menschen aus gesehen wünschenswert, der Dezimaldarstellung möglichst nahe zu bleiben. Dies führte zum BCD (= Binärer Kode für Dezimalziffern). In Anlehnung

an Gl. (2) gilt dann

$$3582\mathrm{D} = \underbrace{\text{ØØ11B}}_{3\mathrm{D}} \times 10^3 + \underbrace{\text{Ø1Ø1B}}_{5\mathrm{D}} \times 10^2 + \underbrace{\text{1ØØØB}}_{8\mathrm{D}} \times 10 + \underbrace{\text{ØØ1ØB}}_{2\mathrm{D}}. \tag{9}$$

Es werden also nur die ersten zehn kombinatorischen Bit-Zustände von 4 *bit* (einer *Schaltertetrade*) verwendet. Die restlichen sechs Zustände werden als Pseudotetraeden bezeichnet. Sie dürfen nicht auftauchen. Die Möglichkeiten der vier Bits sind folglich nicht ganz ausgenutzt. Der Kode BCD ist redundant. Die Redundanz (Weitschweifigkeit) gibt an, wieviel Zustände relativ unbenutzt sind. Allgemein gilt für die Redundanz:

$$R = 1 - \frac{\log C}{\log Z}. \tag{10}$$

Darin bedeuten Z die vorhandenen und C die kodierten, d. h. verwendeten Zustände. Für BCD gilt $R \approx 0{,}17$.

8.1.4.2. Tetradenkodes

Beim BCD sind die zehn Dezimalziffern relativ einfach den 16 Zuständen der vier Bits zugeordnet. Im Prinzip bestehen für die Zuordnung aber viele Möglichkeiten. Sie lassen sich wie folgt berechnen. Es seien die 16 Zustände entsprechend der Reihenfolge der 16 Binärzahlen geordnet. Dann können ihnen zunächst ohne Lücke von vorn die zehn Dezimalziffern in $10! = 3628800$ verschiedenen Möglichkeiten zugeordnet werden. Nun lassen sich die Pseudotetraden in beliebiger Kombinatorik einfügen. Dies multipliziert die vorhandenen Fälle mit $\binom{16}{6} = 8008$, so daß insgesamt 29059430400 verschiedene Tetradenkodes existieren. Diese Anzahl ist so groß, daß sie nicht einmal alle durchprobiert werden können. In der Praxis haben sich dennoch eine Vielzahl von nützlichen Kodes entwickelt. Sechs wichtige sind links unten in Bild 8.1.4 sowie in Zuordnung zu den geordneten Bitzuständen (= Kodewert) in

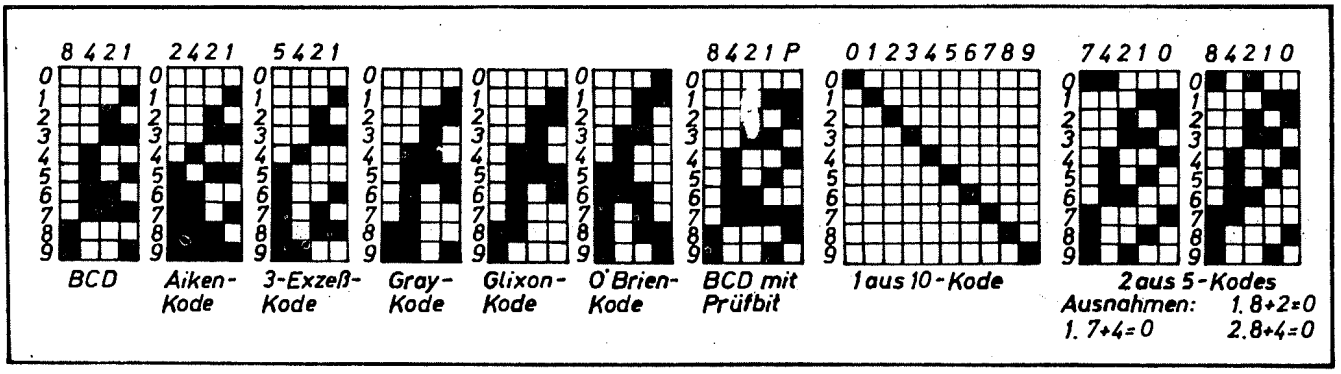


Bild 8.1.4. Häufig verwendete dezimal-binäre Kodes

Tab. 8.1.4 enthalten. Weitere Kodes enthält u. a. [B 24]. Die unterschiedliche Kodierung hat natürlich praktische Gründe, die durch spezielle *Eigenschaften des Kodes* gegeben sind.

8.1.4.3. Kodeeigenschaften

Die folgenden Kodeeigenschaften sind meist genereller Natur, gelten also nicht nur für Tetradenkodes. Sie sind am besten aus Bild 8.1.4 bzw. Tab. 8.1.4 erkennbar.

Tabelle 8.1.4. Vergleich wichtiger Dezimal-Binärkode mit der Stellenzahl vier

Kodewort	BCD	Dreiexzeß Stibitz	Aiken	Gray	Glixan	O'Brien
0000 B	0	—	0	0	0	—
0001 B	1	—	1	1	1	0
0010 B	2	—	2	3	3	2
0011 B	3	0	3	2	2	1
0100 B	4	1	4	7	7	4
0101 B	5	2	—	6	6	—
0110 B	6	3	—	4	4	3
0111 B	7	4	—	5	5 Symmetrie-linie	—
1000 B	8	5	—	—	9	—
1001 B	9	6	—	—	—	9
1010 B	—	7	—	—	—	7
1011 B	—	8	5	—	—	8
1100 B	—	9	6	8	8	5
1101 B	—	—	7	9	—	—
1110 B	—	—	8	—	—	6
1111 B	—	—	9	—	—	—

Ein Kode heißt *wägbar*, gleichwichtig oder Stellenwertkode, wenn jedem bit ein bestimmter Zahlenwert zugeordnet wird und auf diese Weise der durch den Kode dargestellte Zahlenwert so unmittelbar bestimmt ist. Dies ist u. a. für den BCD-, Aiken- und Dreiexzeßkode gegeben. Nicht wägbare Kodes heißen *Anordnungskodes*.

Symmetrisch heißt ein Kode dann, wenn die Lage der Pseudotetraden symmetrisch zur Symmetrielinie ist. Dies ist für den Aiken- und Dreiexzeßkode erfüllt und hat den Vorteil der schnellen Erkennbarkeit von Pseudotetraden.

Ein Kode ist *komplementierbar*, wenn beim Übergang von 0B ↔ 1B (Komplementbildung) eine neue Zahl kodiert entsteht, welche die alte zu 9 ergänzt. Diese Eigenschaft ist bei der Subtraktion wichtig, die dann mit einer einfachen Komplementbildung durch Addition vollzogen werden kann. Der Aiken- und Dreiexzeßkode sind komplementierbar.

Für die *Arithmetik* sind weiter *wichtig*:

- leicht erkennbar, ob die Zahl gerade oder ungerade ist; z.B. am letzten bit (erfüllen BCD-, Dreiexzeß- und Aiken-Kode);
- leicht *abrundbar*: ein bit zur Entscheidung, ob Zahl $\geqq 5$ (Aiken-, Dreiexzeßkode);
- *übertragsgerecht*: Wenn bei Addition Zahlen größer 10 entstehen, muß relativ leicht Übertrag und die verbleibende Zahl aus dem Kode zu gewinnen sein. Beim Dreiexzeßkode entsteht z.B. eine Pseudotetrade, zu der nur 0011 zu addieren ist, um wieder die gültige Ziffer zu erhalten.

Bei der *Übertragung von Kodes* ist die Taktung wichtig. Wären in den zugelassenen Bitkombinationen solche vorhanden, die nur aus Nullen bzw. Einsen bestehen, so wäre das Kodewort, d. h. zum Beispiel die vier bit einer Tetrade, nicht als Takt erkennbar. Kodes, die so etwas vermeiden, heißen *selbsttaktend* oder *maskierbar*.

Für die direkte Anzeige oder Abtastung bei Kodescheiben ist es wichtig, daß mit steigender oder fallender Zahl sich immer genau ein Bit ändert (vgl. 9.2.). Solche Kodes heißen *einschrittig*. Zu ihnen gehören u. a. der GRAY-, GLIXAN- und O'BRIEN-Kode, wobei die beiden letzten darüber hinaus noch *zyklisch* sind. Hierbei ändert sich auch beim Übergang von Ø nach 9 nur ein Bit.

8.1.4.4. Prüfbare Kodes

Bei der Übertragung der einzelnen Bits können, wenn auch mit unterschiedlicher Fehlerrate, Fehler auftreten. Dies ist bei vielen Anwendungen unzulässig. Deshalb hat es sich als notwendig erwiesen, hierzu zusätzliche Maßnahmen zu entwickeln. Sie werden in zwei Gruppen geteilt:

- *Fehlererkennende Kodes* liegen vor, wenn aus der jeweiligen Bitkombination ein Fehler bemerkt wird. In der Regel wird dann die Zahl zurückgewiesen und muß noch einmal übertragen werden.
- *Fehlerkorrigierende Kodes* ermöglichen dagegen, den Ort des Bitfehlers über Rechenverfahren zu ermitteln und damit zu korrigieren.

Beide Methoden sind nur möglich, wenn die Redundanz entsprechend erhöht wird, wenn also für eine Dezimalzahl nicht eine Tetrade, d. h. 4 Bits, sondern mehr Bits verwendet werden. Ein Beispiel zeigt Bild 8.1.4 als BCD mit Prüfbit (engl. parity check). Das P-Bit ist dabei so bestimmt, daß die Anzahl der gesetzten Bits immer geradzahlig ist. Eine andere mögliche Vereinbarung wäre auch Ungeradzahligkeit.

Eine wichtige Methode für die Fehlererkennung sind die *n-aus-m-Kodes*. Sie sind so gestaltet, daß von m vorhandenen Bitstellen in jedem zulässigen Kodewort genau n gesetzte Bits auftreten dürfen. Die Anzahl der zulässigen Kodeworte berechnet sich gemäß

$$C = \binom{m}{n} = \frac{m(m-1)\,(m-2)\,\ldots\,(m-n+1)}{1 \cdot 2 \cdot 3 \ldots n}. \tag{11}$$

Dies führt zu zwei wichtigen Möglichkeiten für Dezimalzahlen, nämlich $m = 5$; $n = 2$ bzw. $m = 5$; $n = 3$ und $m = 10$; $n = 1$. Drei Beispiele für diese beiden Kodes zeigt Bild 8.1.4. Die Redundanzen dieser Kodes betragen gemäß Gl. (10) und wegen $Z = 2^m$ ca. 0,336 bzw. 0,668. Diese Redundanzen sind größer als notwendig, denn im Idealfall müssen nur genau so viele falsche wie richtige Kodeworte so angeordnet werden, daß ein Bitfehler zu einem falschen Kodewort führt. Die zugehörige minimale Redundanz berechnet sich aus Gl. (10) zu

$$R = \log 2/\log Z = \frac{1}{1 + \log C/\log 2}. \tag{12}$$

Für Dezimalzahlen beträgt $C = 10$ und $Z = 20$, und so folgt ein Wert von ca. 0,231.

8.1.4.5. Fehlerkorrigierende Kodes

Für dreistellige Bitwörter läßt sich Fehlererkennung und Fehlerkorrektur gemäß Bild 8.1.5 geometrisch veranschaulichen. Die einzelnen Kodeworte liegen dann auf den Ecken eines Würfels. Sie sind so angeordnet, daß benachbarte Worte sich nur in einem bit unterscheiden. Der minimale Weg auf den Kanten (Änderung von jeweils 1 bit) kann von einem Wort (z.B. 101) zu einem anderen (z.B. 011) über 001 oder 111 erfolgen. D es sind zwei Schritte. Die Anzahl der minimalen Schritte heißt HAMMING-

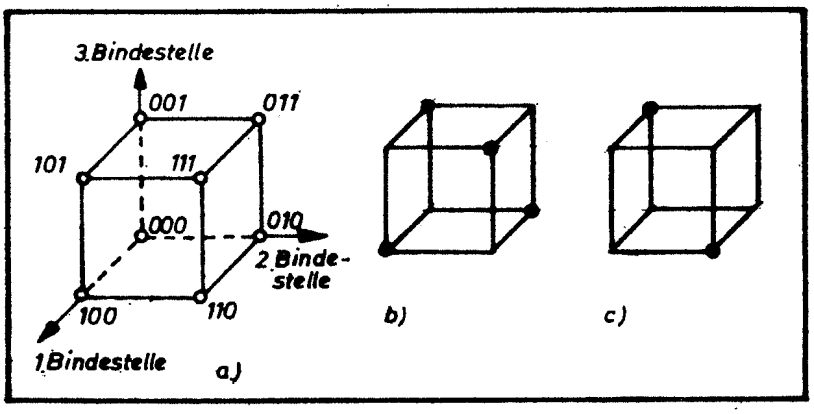


Bild 8.1.5. Zur Erklärung des HAMMING-Abstandes bei dreistelligen Bitworten
a) Belegung der Kanten eines Würfels
b) Bitworte mit dem HAMMING-Abstand 2
c) Bitworte mit dem HAMMING-Abstand 3

Abstand. Er beträgt im Beispiel 2. Für fehlererkennende Systeme muß er mindestens 2 sein. Hierfür gilt die Anordnung in Bild 8.1.5b). Soll der Fehler auch noch korrigierbar sein, so muß er 3 sein. Hierfür gilt Bild 8.1.5c). Tritt dann nämlich ein Fehler auf, so bleibt das neue Kodewort immer noch dem richtigen näher als allen anderen. Für den Aufbau solcher Kodes werden den m Informationsbits k Prüfbits hinzugefügt, so daß die Bitlänge

$$n = m + k \tag{13}$$

beträgt. Mit den k Prüfbits können 2^k Fehlerzustände unterschieden werden. Sie müssen sowohl die Störung überhaupt als auch das gestörte Bit kennzeichnen. Folglich gilt

$$2^k \geqq m + k + 1\,. \tag{14}$$

Kodes, die nach diesem Prinzip arbeiten, heißen ein-Fehler-korrigierende HAMMING-Kodes. Eine andere Art der Fehlerkorrektur beruht darauf, daß mehere Bitworte zu einem Block zusammengefaßt werden und nach der Sendung der Bitworte ein Kodewort zusätzlich gesendet wird. Hierbei kann bei den einzelnen Bitworten über Parity check festgestellt werden, ob sie falsch oder richtig sind. Aus dem Abschlußwort kann durch Aufsummierung der einzelnen Stellen dann festgestellt werden, an welcher Stelle der Fehler sich befindet und womit er korrigiert werden kann. Hieraus wird deutlich, daß alle bisher genannten Verfahren nur einen Fehler erkennen bzw. korrigieren. Die Begründung liegt darin, daß die Fehlerrate sehr klein ist. Wenn ein Fehler z.B. mit der Wahrscheinlichkeit 10^{-5} auftritt, werden zwei Fehler eben nur mit der Wahrscheinlichkeit 10^{-10} auftreten. Bei größeren Fehlerraten müssen Doppelfehler berücksichtigt werden. Dies ist auch möglich, erfordert aber erheblich höheren Aufwand. Anderseits gilt aber: Ein Verfahren für die Erkennung eines Fehlers weist die Fehler entsprechend der obigen Rate von 10^{-5} aus. Dagegen bleiben bei ihm die 10^{-10} Doppelfehler unbemerkt. Ähnliches gibt es bei der Fehlerkorrektur. Die 10^{-5} Fehler werden richtig korrigiert, die 10^{-10} Doppelfehler werden falsch korrigiert oder erweisen sich als unkorrigierbar. Diese Betrachtungen gelten für den Fall, daß Fehler nicht gehäuft auftreten können. Solche Bursts treten u. a. bei Fernübertragungen und bei der Magnetbandspeicherung durch dropout auf. Hierfür existieren spezielle Korrekturverfahren.

Bezüglich weiterer Kodierungsprobleme sei auf [H 26, S 45, S 54, Bd. 2, und V 12, Bd. 1] verwiesen.

8.1.4.6. Byte, Kbyte

Bisher wurde davon ausgegangen, daß eine Dezimalziffer dargestellt werden soll. In der Rechentechnik hat sich aus vielerlei Gründen jedoch das Byte (byte) durchgesetzt. Es besteht aus 8 bit, realisiert damit $2^8 = 256$ Zustände und kann also im obigen Sinne zwei Dezimalzahlen tragen oder eine mit Korrekturbit usw. In der Praxis liegt jedoch meist eine speziellere Kodierung (vgl. 8.1.4.9.) vor.

Eine weitere wichtige Größe in der Rechentechnik ist der Vorsatz K. Er ist deutlich von dem Dezimalvorsatz k zu unterscheiden. Er ergibt sich, wenn 10 bit zur Kodierung verwendet werden, und besitzt entsprechend den dann gegebenen $2^{10} = 1024$ Zuständen einen Wert um etwa 1000 (k). Ein Kbyte enthält folglich 1024 byte = 8192 bit, während ein Kbit = 1024 bit sind.

8.1.4.7. Zahlen mit Vorzeichen

Bei den Dezimalzahlen ist es üblich, das Vorzeichen durch + oder − mit anzugeben. Bei Binärdarstellungen muß hierfür ein bit verwendet oder eine spezielle Kodierung gewählt werden. Ursprünglich war es üblich, das MSB, also das erste bit als Vorzeichen zu verwenden. Mit einem Byte sind dann die Zahlen −127 bis +127 darstellbar. Nachteilig bei diesem Verfahren ist, daß +ØD = ØØØØ ØØØØ B und −ØD = = 1ØØØ ØØØØ B sind und daß bei Fortschreiben im Binärzahlenbereich nach +127D die Zahlen −ØD, −1D usw. folgen. Deshalb wurden im Laufe der Zeit andere Kodierungen entwickelt. Eine davon ist die offsetbinäre Darstellung, bei der additiv die Null verschoben wird. Die Zählung beginnt dann mit −128D und endet bei +127D. Eine zweite ist das Einerkomplement. Durchgesetzt hat sich jedoch das daraus entwickelte Zweierkomplement [A 28, J 11, S 67]. Hierfür gilt

ØØØØ ØØØØB = ØØH = ØD; ØØØØ ØØØ1B = Ø1H = +1D; ØØØØ ØØ1ØB = Ø2H = +2D;
Ø111 1111B = 7FH = +127D; 1ØØØ ØØØØB = 8ØH = −128D;
1ØØØ ØØØØB = 81H = −127D;
1111 111ØB = FEH = −2D; 1111 1111B = FFH = −1D.

Bei einer n-stelligen Binärzahl z gilt für das Zweierkomplement

$$Z' = 2^n - Z\,.$$

Für ein Byte wird das Zweierkomplement gemäß

$$Z' = 256 - Z$$

eine Zahl erhalten. In der Bitkombination läßt sich dies wie folgt realisieren: Von rechts beginnend, bleiben alle Nullen und die erste 1 erhalten, alle weiteren bit werden komplementiert. Hier zwei Beispiele: ←

Zahl:	Ø11Ø 1ØØØ	ØØØ1 ØØ11
Zweierkomplement:	1ØØ1 1ØØØ	111Ø 11Ø1

Der Vorteil der Zweierkomplementzahlen besteht darin, daß die Subtraktion auf die Addition zurückgeführt werden kann. Von der zu subtrahierenden Zahl wird nur zuvor das Zweierkomplement gebildet. Hierfür ist wichtig, daß auch die folgende Regel gilt: Ist eine negative Zahl bereits als Zweierkomplement gegeben und soll sie nun abgezogen werden, so führt das erneute Bilden eines Zweierkomplements wieder auf die richtige positive Zahl zurück.

8.1.4.8. Gleitkommazahlen

Gleitkommazahlen (floating point) können in zweierlei Art dargestellt werden, entweder durch Setzen eines Dezimalpunktes oder in der Schreibweise aus Mantisse und Exponent. Bisher wurden nur ganze Zahlen (integer) betrachtet. Darüber hinaus haben auch gebrochene Zahlen (real) große Bedeutung. Sie können mit einem Dezimalpunkt ebenfalls als binäre Zahlen dargestellt werden. So kommen dann die Anteile 2^{-n}, also 0,5; 0,25; 0,125; 0,0625 usw., hinzu. Ähnlich wie die Dezimaldarstellung echter Brüche, z.B. $^1/_3$, zu unendlichen Darstellungen führt, z.B. 0,333 ... , kann auch die Binärdarstellung endlicher Dezimalbrüche, z.B. 0,1, oder echter Brüche, z.B. $^1/_3$, unendlichstellige Binärzahlen erfordern. Da jeder Rechner oder jedes digitale System aber nur eine endliche Stellenzahl besitzt, können sich hierbei die dann notwendigen Abrundungen störend auswirken. So wird bei einer endlichen Binärdarstellung der Folge $-2 + n \times 0{,}1$ niemals exakt der Wert 0 erreicht. Dies ist u. a. ein Grund, weshalb in neueren Rechnern, auch Mikrorechnern, immer mehr zur Tetradendarstellung auf der Basis des Dezimalsystems übergegangen wird. Es gibt aber auch andere Methoden, dieses Problem zu umgehen. Ein Beispiel für das Format einer Gleitkommadarstellung zeigt Bild 8.1.6.

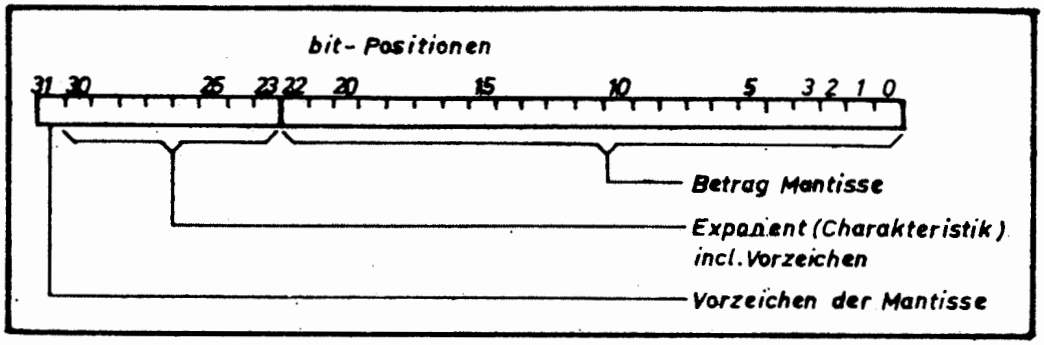


Bild 8.1.6. Beispiel für das Format einer Gleitkommazahl

8.1.4.9. Kodierung von Buchstaben, Befehlen usw.

Den Zuständen der bit können nicht nur Zahlen, sondern beliebige andere Symbole zugeordnet werden. Besondere Bedeutung hat die Kodierung von Texten. Hierbei wird den Zuständen eines Bytes mit Ausnahme jener, die MSB bewirken, das große und kleine Alphabet sowie Zahlen, Sonderzeichen und Steuerbefehle zugeordnet.

Der international genormte Kode ist also ein 7-bit-Kode. Er trägt bei geringfügigen Abweichungen u. a. die folgenden Bezeichnungen: ASCII, Internationaler CCITT Nr. 5, ISO 7-bit-Kode, SIF 1000, DIN 66003 und ist in Tab. 8.1.5 dargestellt. Eine andere Bedeutung gewinnen die gleichen Bitkombinationen, wenn sie als Befehle eines Mikrorechners auftreten. Beim U 880 bedeutet 48 H = Ø1ØØ 1ØØØ B z.B. LD C, B, also: Lade das Register C ins Register B. In anderen Zusammenhängen wird dem Wert eine wiederum andere Bedeutung gegeben werden. Wesentlich ist, daß die Binärzahlen, genauer, Zustände von bit (Schaltern), das einzige sind, womit der Rechner oder allgemeiner die digitale Technik arbeitet. Ihre richtige Interpretation als Binärzahl, als Dezimalzahl im Zweierkomplement, als Text, als Befehl, Adresse usw. wird erst durch den Kontext gegeben. Er bestimmt die richtige Kodierung und Dekodierung der Bitzustände. Vgl. Abschnitt 8.6.6.

Tabelle 8.1.5. Darstellung des internationalen 7-bit-Kodes für Texte. Die Bedeutung der Steuerzeichen NULL, SOH usw. wird hier nicht weiter erklärt [B 24]. Die Zeichen (◄) mit den Dezimal-Nr. 91 bis 94 und 123 bis 126 können international unterschiedlich belegt werden. Im Deutschen liegen hier zuweilen die Umlaute und ß.

								0	0	0	0	1	1	1	1
								0	0	1	1	0	0	1	1
								0	1	0	1	0	1	0	1
Bit	b6	b5	b4	b3	b2	b1	b0	Spalte	0	1	2	3	4	5	6	7
								HEX Zeile								
				0	0	0	0	0	NUL	(TC7) DLE	SP	0	@(§)	P	’	p
				0	0	0	1	1	(TC1) SOH	DC1	!	1	A	Q	a	q
				0	0	1	0	2	(TC2) STX	DC2	”	2	B	R	b	r
				0	0	1	1	3	(TC3) ETX	DC3	#(£)	3	C	S	c	s
				0	1	0	0	4	(TC4) EOT	DC4	$	4	D	T	d	t
				0	1	0	1	5	(TC5) ENQ	(TC8) NAK	%	5	E	U	e	u
				0	1	1	0	6	(TC6) ACK	(TC9) SIN	&	6	F	V	f	v
				0	1	1	1	7	BEL	(TC10) ETB	’	7	G	W	g	w
				1	0	0	0	8	(FE0) BS	CAN	(	8	H	X	h	x
				1	0	0	1	9	(FE1) HT	EM	)	9	I	Y	i	y
				1	0	1	0	10; A	(FE2) LF	SUB	*	:	J	Z	j	z
				1	0	1	1	11; B	(FE3) VT	ESC	+	;	K	[(Ä)	k	{(ä)
				1	1	0	0	12; C	(FE4) FF	(IS4) FS	,	<	L	\(Ö)	l	|(ö)
				1	1	0	1	13; D	(FE5) CR	(IS3) GS	–	=	M	](Ü)	m	}(ü)
				1	1	1	0	14; E	SO	(IS2) RS	.	>	N	^	n	~(ß)
				1	1	1	1	15; F	SI	(IS1) US	/	?	O	←	o	DEL

8.2. Kombinatorische Schaltungen

In der digitalen Technik ist es sinnvoll, drei grundlegende Schaltungen zu unterscheiden:

— *Kombinatorische Schaltungen* sind im Prinzip ohne Zeiteinfluß zu betrachten. An ihrem Eingang stehen gleichzeitig mehrere binäre Signale, aus deren Kombination ein oder mehrere Ausgangssignale abgeleitet werden.

— *Speicher* mit Schaltungen, die eigentlich erst durch die Mikroelektronik als besondere elektronische Schaltungen deutlich hervortreten. Mit ihnen ist es möglich, digitale Signale über bestimmte Zeiten auch dann, wenn das Eingangssignal bereits nicht mehr vorhanden ist, zu erhalten und weiterhin bereitzustellen.

— *Sequentielle Schaltungen* bestehen aus kombinatorischen Schaltungen und Speichern. Sie sind dadurch in der Lage, aus einer Folge von digitalen Signalen eine andere Folge zu erzeugen. Daher heißen sie auch Folgeschaltungen. Sie sind die Grundlage aller Automaten, insbesondere der Rechner. Je nachdem. ob die Schaltungen in bestimmter Zeitfolge von einem Takt gesteuert werden oder nicht, werden bei ihnen synchrone (getaktete) oder asynchrone Schaltungen unterschieden.

Zuweilen werden die Speicher bereits den sequentiellen Schaltungen zugerechnet. Hier werden die sequentiellen Schaltungen noch enger begrenzt, nämlich in Bezug zur Automatentheorie und zum Mikrorechner aufgefaßt. Dies führt dazu, daß den relativ einfachen sequentiellen Schaltungen, wie dem Schieberegister, Zähler, Frequenzteiler und Bussystem, ein besonderer Abschnitt gewidmet wird. Ferner sei hier bereits darauf hingewiesen, daß diese bewußt von der Struktur der Schaltun-

gen abgeleitete Einteilung vor allem historisch bestimmt ist. Die eigentlich interessierende funktionelle Seite von Schaltungen ist zum großen Teil unabhängig davon. Hierauf kann aber erst im Kapitel 9. eingegangen werden.

Weiter sei erwähnt, daß die Benennung der Schaltungen international nicht einheitlich ist. Aus diesem Grunde gibt Tab. 8.2.1 eine Zusammenstellung von im wesentlichen synonymen Begriffen für die beiden Hauptgruppen.

Tabelle 8.2.1. Synonyme von Begriffen der digitalen Technik. Zum Teil bestehen dabei geringfügige Inhaltsunterschiede

Kombinatorische Schaltungen	*Sequentielle Schaltungen*
Logische Schaltungen	Folgeschaltungen
Schaltnetze	Schaltnetzwerke
Statische Logik	Dynamische Logik
Binäre Logik	
Zuordner	
Theorie dazu	
Schaltalgebra	Automatentheorie
Boolesche Algebra	

8.2.1. Mathematische Grundlagen

8.2.1.1. Aussagenlogik

Die Logik geht auf den griechischen Philosophen Aristoteles (384—322 v. u. Z.) zurück. Er hat den Formalismus der Zweiwertigkeit einer Aussage, wahr oder falsch, eingeführt. Die entscheidenden theoretischen Entwicklungen in Analogie zur Algebra erfolgten durch A. de Morgan (1806—1871) und G. Boole (1815—1864). Andere wichtige Grundlagen, die nicht in so enger Beziehung zur Algebra standen, entwickelten G. Freege (1848—1925) und G. Peano (1858—1932).

Die Logik besteht im wesentlichen aus der Aussagen- und Prädikatenlogik. Im Zentrum des Prädikatenkalküls stehen die Prädikate. Prädikat stammt vom lateinischen praedicatum (zu praedicare = laut sagen), also Ausgesagtes. In der Sprachlehre (Satzbau) ist Prädikat die Satzaussage. Prädikate der Logik dienen zur Bezeichnung von Eigenschaften der Objekte oder Beziehungen zwischen ihnen. Der Prädikatenkalkül hat für die Schaltalgebra so gut wie keine Bedeutung.

Die Aussagenlogik geht von Aussagen wie

A: „Warschau liegt an der Weichsel“
B: „Die Weichsel fließt durch Warschau“
C: „5 ist eine Primzahl“
D: „$2 + 2 = 5$“

aus. Dabei interessiert nicht der eigentliche Inhalt der Aussage, sondern nur, ob sie „falsch“ oder „wahr“ ist.

Durch Verknüpfung verschiedener Aussagen entsteht eine neue Aussage, die wiederum „falsch“ oder „wahr“ sein kann. Solche Verknüpfungen können zunächst umgangssprachlich: nicht, und, oder, entweder — oder, sowohl — als auch, wenn — dann usw. sein. Also z.B.: „wenn 5 eine Primzahl ist, dann ist $2 + 2 = 5$“. Die Verknüpfungsoperatoren heißen auch Funktoren. Es gibt ein- und zweistellige Funktoren. Von den obengenannten ist nur „nicht“ einstellig. Bezüglich der Wahrheitswerte (falsch — wahr) der Einzelaussagen und der durch Funktoren verknüpften

Tabelle 8.2.2. Beispiele gebräuchlicher Wahrheitswerte

Beispiel	Wahrheitswert	
Aussagenlogik	wahr richtig	falsch nicht wahr
Symbolisch	1	Ø
Strom Widerstand Leitwert	groß	klein
Relais	angezogen	abgefallen
(Arbeits-) Kontakt	geschlossen	geöffnet
Potential	positiv hoch H (high)	negativ niedrig L (low)
Transistor	leitend	gesperrt
Lochkarte	Loch	Nicht-Loch

Tabelle 8.2.3. Beispiele für Funktoren

Bezeichnung	Umgangssprachliche Formulierung	Symbole	Andere Symbole	Polnische Notation
Negation (NOT)	es ist nicht wahr, daß x	$\bar{x}$	$\neg x$; x'; $\sim x$	Nx
Konjunktion (AND)	x_1 und x_2	$x_1 \wedge x_2$	$x_1 \cdot x_2$; x_1x_2; x_1 & x_2 $x_1 \cap c_2$ Min (x_1, x_2)	Kx_1x_2
Disjunktion = Alternative (OR)	x_1 oder x_2 x_1 und/oder x_2	$x_1 \vee x_2$ ($\vee$ = vel = oder)	$x_1 + x_2$; $x_1 \cup x_2$ Max (x_1, x_2)	Ax_1x_2
Implikation	wenn x_1, so (dann) x_2; x_1 ist hinreichend dafür, daß x_2	$x_1 \Rightarrow x_2$	$x_1 > x_2$; $x_1 \rightarrow x_2$	Cx_1x_2
Äquivalenz	x_1 dann und nur dann, wenn x_2; x_1 ist hinreichend und notwendig dafür, daß x_2	$x_1 \Leftrightarrow x_2$	$x_1 \equiv x_2$; $x_1 \leftrightarrow x_2$ $x_1 \rightleftarrows x_2$; $x_1 \sim x_2$	Ex_1x_2
Scheffersscher Strich (NAND) = negierte Konjunktion	nicht x_1 oder nicht x_2	$x_1 \mid x_2$		Dx_1x_2
Gleichzeitige Negation (NOR) = Peirce-Funktion = negierte Disjunktion	weder x_1 noch x_2	$x_1 \downarrow x_2$	$x_1 \downarrow x_2$	Xx_1x_2
Antivalenz = symmetrische Differenz = ausschließende Alternative Exklusiv-ODER (XOR)	entweder x_1 oder x_2	$x_1 \nsim x_2$	$x_1 \oplus x_2$; $x_1 \not\equiv x_2$ $x_1 \dot{-} x_2$	Jx_1x_2

Aussagen gibt die Aussagenlogik Rechenregeln. Da z.B. die Aussage C wahr und die Aussage D falsch ist, folgt, daß die Aussagen „A und B" falsch, aber „A oder B" wahr sind.

In der Logik wird der soeben dargestellte Inhalt formalisiert. Dabei sind zu unterscheiden:

— *Variable* (= Aussagen), meist mit A, B, aber auch mit p, q oder $x_1, x_2, \ldots, x_n$ und $y_1, \ldots, y_n$ bezeichnet.
— *Wahrheitswerte*: falsch — wahr. Hierzu gibt Tab. 8.2.2 einen Überblick.
— *Funktoren*. Beispiele enthält Tab. 8.2.3. Die polnische Notation in der letzten Spalte geht auf LUKASIEWICZ zurück. Sie hat die Eigenschaft, daß der Funktor vorangestellt ist und so selbst bei komplizierten Ausdrücken keine Klammern benötigt werden.

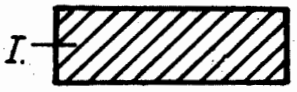
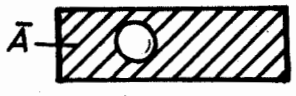
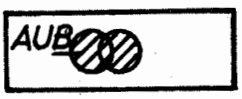
8.2.1.2. Abstrakte Algebra

Algebra stammt von dem arabischen al gabr und bedeutet soviel wie das Herüberbringen der Glieder einer Gleichung und hängt damit ursprünglich mit dem Auflösen von Gleichungen, dem Finden von Wurzeln zusammen. Im Laufe der Geschichte unterlag die Algebra vielen und grundlegenden Änderungen. Über die Theorie der Gleichungen rückten im Laufe der Zeit Begriffe wie Zahlkörper, irreduzibles Polynom, Theorie der Gruppe, Theorie der Permutationen, Isomorphismen, Operationen, Ringe usw. ins Zentrum. Die moderne Algebra wurde von VAN DER WAERDEN 1930 durch einen axiomatischen Aufbau begründet. Ein Teilgebiet der Algebra ist die BOOLEsche. Sie wird noch genauer behandelt.

8.2.1.3. Mengentheorie

Die Mengentheorie wurde von G. CANTOR (1845—1918) begründet. Eine Menge ist eine Zusammenfassung von bestimmten, wohlunterscheidbaren Elementen zu einem

Tabelle 8.2.4. Zusammenstellung wichtiger Mengenbegriffe und Beziehungen zur Logik

Begriff	Symbol	Analog zu	Veranschaulichung
Relative Allmenge Einsmenge	I	1	
Leere Menge Nullmenge	0	$\emptyset$	
Teilmenge	$A \subset I$; $B \subset I$	—	
Komplementäre Menge	$\bar{A}$	Negation	
Durchschnitt zweier Mengen	$A \cap B$	Konjunktion	
Vereinigungsmenge	$A \cup B$	Disjunktion	

Ganzen. Elemente sind Objekte unserer Anschauung oder unseres Denkens. Eine Menge läßt sich damit durch Aufzählung der Elemente oder durch ihre gemeinsamen Eigenschaften definieren. Um Widersprüche (Antinomien) zu vermeiden, müssen alle Aussagen in einer relativen Allmenge I erfolgen. In dieser Menge gibt es dann Teilmengen, Vereinigungsmengen usw. Veranschaulicht werden derartige Zusammenhänge im VENN- oder EULER-Diagramm. Einen Überblick zu den hier bedeutsamen Mengenbegriffen und zu den Beziehungen zur Logik zeigt Tab. 8.2.4.

8.2.2. BOOLEsche Algebra

Erste Versuche, auf Schalter- und Relaisstrukturen mathematische Methoden anzuwenden, erfolgten in den zwanziger Jahren. Wegen der damaligen relativ geringen Komplexität derartiger Schaltungen war ihnen kein Erfolg beschieden. Dies änderte sich in den dreißiger Jahren vor allem durch die Fernsprech- und Telegrafentechnik. Unabhängig und gleichzeitig entstanden dabei praktikable Lösungen durch NAKASHIMA und HANZAWA in Japan, durch SCHESTAKOW in der UdSSR und durch SHANNON in den USA. Die für die Zukunft fruchtbarste Lösung wurde von SHANNON gefunden.

8.2.2.1. Mögliche Funktionen

Bei reellen Zahlen kennen wir zwei Gruppen von Funktionen: Funktionen, die von *einer Variablen* abhängen,

$$f(x)\,,\ \text{z.B. } x^2;\ \sqrt{x};\ \sin x;\ \log x;\ \mathrm{e}^x;\ \int_0^x f(\xi)\,\mathrm{d}\xi;\ x!\,,$$

und Funktionen von *zwei Variablen*,

$$f(x, y)\,,\ \text{z.B. } x + y;\ x - y;\ x \cdot y;\ x/y;\ \log_y (x);\ x^y;\ \sqrt[x]{y}\,.$$

Im Prinzip lassen sich bei den reellen Zahlen unendlich viele Funktionen beider Gruppen definieren. Dies liegt u. a. wesentlich darin begründet, daß die Variable x bzw. y und auch ihr Ergebnis, die Funktion $f(x)$ bzw. $f(x, y)$, *unendlich vieler Werte* fähig sind.

Bei der BOOLEschen *Algebra* (auch Schaltalgebra genannt) liegen die Verhältnisse anders. Bei ihr bestehen für die Variablen x, y und das Ergebnis der Funktionen *nur zwei Möglichkeiten*. Sie können jeweils nur die Werte Ø und 1 annehmen (vgl. Tab. 8.2.2). Dadurch ist es möglich, alle zulässigen (möglichen) Werte tabellarisch aufzuschreiben. Für die BOOLEschen Funktionen mit einer Variablen bestehen auf diese Weise nur die vier in Tab. 8.2.5 dargestellten Möglichkeiten. Davon ist offensichtlich nur eine Funktion, nämlich die Negation (NOT), nicht trivial.

Werden *Funktionen von zwei Eingangsvariablen* x_1 und x_2 (oben wurde x und y geschrieben) verwendet, so lassen sich überhaupt nur vier mögliche Kombinationen der Eingangsgrößen unterscheiden. Sie sind links in Tab. 8.2.6 zusammengestellt und können als die vier zulässigen Eingangszustände aufgefaßt werden. Werden dabei die

Tabelle 8.2.5. Die einstelligen BOOLEschen Funktionen

x	1 Ø	Bezeichnung
y	Ø Ø	Konstante Ø
	Ø 1	Negation
	1 Ø	Regeneration
	1 1	Konstante 1

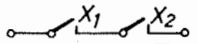
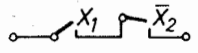
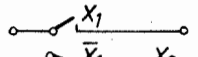
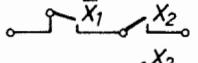
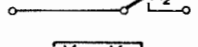
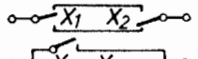
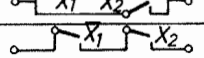
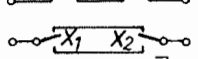
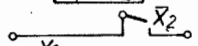
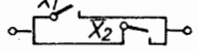
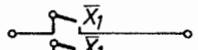
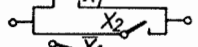
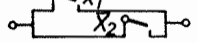
Tabelle 8.2.6. Zur Interpretation der zweistelligen BOOLEschen Funktionen

Mögliche Zustände des Eingangs			Als Dezimalzahl interpretiert		Beispiel Ausgang bei OR
x_1	x_2	⇒	x_D	⇒	$y = x_1 \vee x_2$
Ø	Ø		Ø D		Ø
Ø	1		1 D		1
1	Ø		2 D		1
1	1		3 D		1
					⇓
					Als Dezimalzahl: 7 D

Größen x_1 und x_2 gemeinsam betrachtet, so können sie auch als *zweistellige Binärzahl* gemäß Abschnitt 8.1.2. und 8.1.4. interpretiert werden. Dies ermöglicht die *Umrechnung in eine Dezimalzahl* und führt zu den Werten in der Mitte von Tab. 8.2.6. Diese Interpretation hat darüber hinaus, insbesondere wenn noch mehr als zwei Variablen verwendet werden, den Vorteil, daß alle zulässigen Zustände in einer *eindeutigen Reihenfolge* angeschrieben werden können.

Die verschiedenen zweistelligen Funktionen entstehen nun dadurch, daß jedem der vier Eingangszustände ein Ausgangszustand y aus den beiden Möglichkeiten Ø oder 1 zugeordnet wird. Für die Funktion der Disjunktion (OR) ist dies in Tab. 8.2.6 beispielhaft geschehen. Werden nun die so erhaltenen vier Ausgangswerte (entsprechend den Eingangszuständen) wiederum zusammengefaßt, so entsteht eine vierstellige binäre Zahl. Sie kann ebenfalls ins Dezimale umgerechnet werden und führt für die OR-Funktion zu 7D. Sie kann auch als *Funktions-Nr.* interpretiert werden. Da eine vierstellige binäre Zahl nur 16 verschiedene Werte annehmen kann, existieren folglich auch nur *16 verschiedene zweistellige* BOOLEsche Funktionen. Sie sind in

Tabelle 8.2.7. Zusammenstellung aller zweistelligen BOOLEschen Funktionen

Funktions Nr.		Ausgang y bei x_1: 0011, x_2: 0101	Bezeichnung	Mathem. Darstellung mittels ∧,∨,-	Kontaktnetzwerk	Symbol. Darstellung	Mengentheoret. Darstellung	Kornaugh-Diagramm
Ø		0000	Konstante 0 (niemals)	0		—		
1	Normale Formen	0001	Konjunktion (AND)	$X_1 \wedge X_2$		&		
2		0010	Inhibition II	$X_1 \wedge \bar{X}_2$		&		
3		0011	Identität (X_1)	$\bar{X}_1$		—		
4		0100	Inhibition I	$X_1 \wedge X_2$		&		
5		0101	Identität (x_2)	X_2		—		
6		0110	Antivalenz (XOR)	$(\bar{X}_1 \wedge X_2) \vee (X_1 \wedge \bar{X}_2) \equiv (\bar{X}_1 \vee \bar{X}_2) \wedge (X_1 \vee X_2)$		=1		
7		0111	Disjunktion (OR)	$X_1 \vee X_2$		1		
Symmetrieebene								
8		1000	Peirce-Fktn. (NOR)	$\bar{X}_1 \vee \bar{X}_2 = \bar{X}_1 \wedge \bar{X}_2$		1		
9		1001	Äquivalenz	$(\bar{X}_1 \wedge \bar{X}_2) \vee (X_1 \wedge X_2) \equiv (X_1 \vee \bar{X}_2) \wedge (\bar{X}_1 \vee X_2)$		=		
10		1010	Negation ($\bar{X}_2$)	$\bar{X}_2$		—		
11	Negierte Formen	1011	Implikation II	$X_1 \vee \bar{X}_2$		1		
12		1100	Negation ($\bar{X}_1$)	$\bar{X}_1$		—		
13		1101	Implikation I	$\bar{X}_1 \vee X_2$		1		
14		1110	Schefferscher Strich (NAND)	$\overline{X_1 \wedge X_2} \equiv \bar{X}_1 \vee \bar{X}_2$		&		
15		1111	Konstante 1 (immer)	1		—		

Tab. 8.2.7 zusammengestellt. Dabei werden in den verschiedenen Spalten unterschiedliche Darstellungsmethoden für die einzelnen Funktionen verwendet, auf die noch einzugehen ist. Ein Teil dieser Funktionen ist wiederum, wie bei den einstelligen Funktionen, trivial. Dies sind die Funktionen mit den Nummern 0, 3, 5, 10, 12 und 15. Sie stellen entweder Konstanten dar oder sind eigentlich einstellige Boolesche Funktionen. Besonders sei noch auf die Funktion Nr. 6, die *Antivalenz*, hingewiesen. Sie entspricht dem umgangssprachlichen „*Entweder — Oder*". Sie hebt sich damit deutlich von der Funktion 7 der *Disjunktion* ab. Diese Funktion ist umgangssprachlich schwierig zu erfassen, denn sie müßte lauten „*Und oder Oder*". Im Patentrecht wird dabei und/oder geschrieben. Damit gibt es drei Funktionen, die in der Kurzform das mögliche Oder (OR) enthalten: XOR (6) — von exklusives Oder, OR (7) und NOR (8) — verneintes Oder bzw. NOT-OR.

8.2.2.2. Vielzahl der Darstellungsformen

Aus den vorangegangenen Betrachtungen wird deutlich, daß es für den gleichen Sachverhalt einer Boolеschen Funktion mehrere Darstellungsformen gibt. Sie sind auf verschiedenen Gebieten historisch entstanden und haben für die Anwendung der Boolеschen Algebra in verschiedenen Gebieten und bei verschiedenen Aufgaben unterschiedliche Vor- und Nachteile. Deswegen sei hier versucht, die einzelnen Methoden aufzuzählen und grob zu vergleichen.

Die *verbale* Darstellungsform geht im wesentlichen auf die Aussagenlogik (8.2.1.1.) zurück und verwendet vor allem Wörter, wie: wenn — dann, oder, entweder — oder, folgt aus usw. (vgl. auch Tab. 8.2.3). Diese Formulierungen werden oft der Anfang einer zu lösenden Aufgabe sein. Dann ist in der Regel das Problem noch unscharf gefaßt und bedarf einer weiteren Präzisierung.

Die *Wertetabelle* ist eine Darstellung, die bei nicht zu großen Problemen vorteilhaft ist. Sie wird im folgenden noch mehrfach verwendet. Ihr Prinzip geht aus dem linken und dem rechten Teil von Tab. 8.2.6 hervor. Für alle einstelligen Funktionen zeigt sie Tab. 8.2.5.

Die *mathematische* oder *symbolische* Darstellung ist für mehrere Funktionen und in verschiedenen Notationen in Tab. 8.2.3 gezeigt. Sie ist in der Regel die Voraussetzung dafür, daß Boolеsche Funktionen im Rechner behandelt werden können.

Das Karnaugh-*Diagramm* (zuweilen auch Karnaugh-Veitsch-Diagramm oder Wahrheitstafel genannt) ist eigentlich nur eine spezielle Darstellungsform der Wertetabelle (vgl. letzte Spalte in Tab. 8.2.7). Es hat jedoch Vorteile, wenn eine Minimierung des Aufwandes einer aus der Boolеschen Algebra abgeleiteten Schaltung erfolgen soll. Da heute solche Probleme meist im Rechner mit der symbolischen Darstellung erfolgen, hat es an Bedeutung verloren.

Das Venn-*Diagramm* ist die mengentheoretische Betrachtung Boolescher Funktionen. Es besitzt bei einfachen Zusammenhängen den großen Vorteil einer guten Übersichtlichkeit. Sein Prinzip ist in Tab. 8.2.4 gezeigt. In Tab. 8.2.7 sind in der vorletzten Spalte alle zweistelligen Funktionen auf diese Weise dargestellt.

Der technischen Realisierung besonders nahe ist eine Veranschaulichung der Boolеschen Funktionen mittels einer *Schaltung aus Kontakten*. Wenn durch sie Strom fließt, liegt eine Eins vor, sonst eine Null. Für alle zweiwertigen Funktionen zeigt solche Schaltungen die mittlere Spalte in Tab. 8.2.7. Heute haben Kontaktdarstellungen nur noch untergeordnete Bedeutung.

Besonders wichtig ist die Darstellung mit *Schaltsymbolen*. Sie leiten nämlich unmittelbar zur technischen Realisierung über. Dies ist vor allem seit dem Bestehen der TTL-Technik wichtig, weil dabei in der Regel einem Schaltsymbol ein einfacher Bau-

stein entspricht. Die wichtigen einfachen Symbole enthält Tab. 8.2.7. Dabei sind folgende *Kennzeichen in und an den Kästchen* zu beachten:

○ Der offene Kreis am Eingang bewirkt, daß diese Eingangsgröße, bevor sie im Kästchen weiter beachtet wird, erst negiert wird.
Am Ausgang bedeutet er, daß nach der Verarbeitung der Information im Kästchen anschließend eine Negation der Ausgangsgröße noch vorgenommen wird.

Die Zeichen im Kästchen berücksichtigen die Eingangsgröße also erst nach dem Kreis. Sie bedeuten dann:

& Alle Eingangsgrößen müssen 1 sein, damit eine 1 am inneren Ausgang auftritt (AND).
1 Zumindest eine der Eingangsgrößen für eine interne 1 muß 1 sein (OR). Zuweilen wird hierfür auch $\geqq 1$ geschrieben.
= Die Eingangsgrößen müssen für eine interne 1 untereinander gleich sein (Äquivalenz).
$=1$ Die Summe aller Eingangsgrößen muß für eine interne 1 genau 1 sein (Antivalenz).

Bereits bei dieser Beschreibung ist deutlich geworden, daß die Schaltsymboldarstellung leicht auf mehr als zwei Eingangsvariablen übertragbar ist.

Bei der Darstellung ist zu beachten, daß früher u. a. Symbole mit Halbkreisen gültig waren und in den USA eine spezielle Darstellung entwickelt wurde. Eine vergleichende Zusammenstellung zeigt Tab. 8.2.8.

Tabelle 8.2.8. Zusammenstellung einiger älterer gebräuchlicher Symbole für BOOLEsche Funktionen

Negation	1		-1	NOT		
Konjunktion	&			AND		
Disjunktion	≥1			OR	+	
Antivalenz	=1			XOR		
SCHEFFER-Funktion	&			NAND		
PEIRCE-Funktion	≥1			NOR		

8.2.2.3. Mögliche Grundfunktionen

Die 16 in Tab. 8.2.7 dargestellten BOOLEschen Funktionen sind nicht unabhängig voneinander. Es gibt vielmehr einige Möglichkeiten, sie durch Zusammenschalten aus einigen wenigen Grundfunktionen zu realisieren. Größere Bedeutung haben *fünf* solcher *Sätze*:

- Die drei Funktionen $\vee$, $\wedge$, $-$ (OR, AND, NOT) sind klassische Grundfunktionen. Sie haben heute in der Praxis nur noch geringe Bedeutung als solche.
- Die SHELGAKIN-*Algebra* verwendet die Konjunktion (AND) und die Antivalenz (XOR). Sie hat hervorragende Anwendung dort gefunden, wo dynamische Vorgänge mit der BOOLEschen Algebra ausgedrückt werden sollen, z. B. in [B 33].
- Eine wenig benutzte Auswahl entsteht mittels *Inhibition* und *Implikation*.

- Größte Bedeutung als einzige Grundfunktion, aus der alle anderen aufgebaut werden, hat das NAND. Dies ist darauf zurückzuführen, daß die Grundfunktion der TTL-Schaltkreise (8.2.4.) diese Funktion ist.
- Schließlich sei als ebenfalls allein genügende Grundfunktion das NOR erwähnt. Es ist bei den ECL- und I^2L-Schaltkreisen vorherrschend (8.2.5. und 8.2.6.).

8.2.2.4. Höherstellige Funktionen

Aus den vorangegangenen Betrachtungen ist ersichtlich, daß die BOOLEschen Funktionen auch leicht für mehr als zwei Eingangsvariablen definiert werden können. Insbesondere behalten hierbei die Grundfunktionen des vorigen Abschnittes ihre volle Gültigkeit. Dies bedeutet u. a., daß alle höherstelligen Funktionen ausschließlich z.B. aus NAND-Gliedern zu realisieren sind. *Bei den höherstelligen Funktionen* geht jedoch ein wesentlicher *Vorteil der* BOOLE*schen Algebra verloren*, nämlich daß alle Funktionen leicht *per Tabelle aufgezählt* werden können. Dies wird erst deutlich, wenn die Anzahl der mehrstelligen Funktionen berechnet wird. Dazu müssen zunächst einmal die möglichen *Eingangszustände* bestimmt werden. Für n Eingänge entsteht eine n-stellige Zahl von Eingangszuständen in Analogie zu Tab. 8.2.6. Sie besitzt genau 2^n mögliche Werte. Jedem dieser Werte kann nun wieder ein Ausgangszustand zugeordnet werden. Die sich dadurch ergebenden Kombinationen bestimmen die Anzahl der *möglichen Funktionen*. Dies sind $2^{(2n)}$, Tab. 8.2.9 faßt hierzu die berechneten Zahlenwerte zusammen. Daraus geht einmal hervor, daß es spätestens ab $n = 4$ praktisch nicht mehr möglich ist, alle Funktionen z.B. als Tabelle aufzuschreiben. Zum anderen geht daraus aber auch hervor, daß mit BOOLEschen Funktionen höherer Ordnung eine *praktisch unausschöpfbare Vielzahl von Realisierungen* besteht. Die mehrstellige BOOLEsche Algebra hält alle für die meisten Anwendungen praktikablen Lösungen bereit. Dies gilt auch unter der Bedingung, daß unter den vielen Funktionen immer einige triviale enthalten sind.

Tabelle 8.2.9. Anzahl der möglichen BOOLEschen Funktionen mit n Eingangsvariablen

Variable n	Eingangszustände 2^n	Funktionen 2^{2n}
0	1	2
1	2	4
2	4	16
3	8	256
4	16	65536
5	32	$4{,}29 \cdot 10^9$
6	64	$1{,}84 \cdot 10^{19}$
7	128	$3{,}40 \cdot 10^{38}$
8	256	$1{,}16 \cdot 10^{77}$
9	512	$1{,}34 \cdot 10^{154}$
10	1024	$1{,}80 \cdot 10^{308}$

8.2.2.5. Schwellenwertlogik

Eine Zeitlang hatten Schwellenwertelemente eine große Bedeutung. Zu ihnen gehört die Schwellenwertlogik. Sie wird in der Regel so definiert, daß die Eingangs- und Ausgangsgrößen wie in der BOOLEschen Algebra nur der Werte Ø und 1 fähig sind. Intern erfolgt aber die Verarbeitung in zwei Stufen. Zunächst geschieht eine gewöhnliche Addition (und z. T. auch Subtraktion) der Eingangsgrößen. Dann erfolgt eine Entscheidung, ob ein bestimmter Schwellwert, z.B. 3, über- bzw.

unterschritten wird. Je nach dem Ergebnis dieser Entscheidung wird dann an den Ausgang eine Ø oder 1 ausgegeben. Es läßt sich zeigen, daß eine solche Schwellwertlogik stets mit BOOLEschen Funktionen zu realisieren ist. Dies ist für einige Fälle in Tab. 8.2.10 dargestellt. Durch die schnelle Entwicklung der Digitaltechnik in den letzten Jahren ist daher die Schwellwertlogik bedeutungslos geworden.

Tabelle 8.2.10. Beispiele für Schwellwertfunktionen

Schwellwertlogik	Zugehöriger BOOLEscher Ausdruck
$x_1 + x_2 \geqq 1$	$x_1 \vee x_2$
$x_1 + x_2 \geqq 2$	$x_1 \wedge x_2$
$x_1 + x_2 \leqq Ø$	$\bar{x}_1 \wedge \bar{x}_2$
$x_1 + x_2 + x_3 \geqq 1$	$x_1 \vee x_2 \vee x_3$
$x_1 + x_2 + x_3 \geqq 2$	$(x_1 \wedge x_2) \vee (x_1 \wedge x_3) \vee (x_2 \wedge x_3)$
$x_1 + x_2 + x_3 \geqq 3$	$x_1 \wedge x_2 \wedge x_3$
$2x_1 + x_2 + x_3 + x_4 \geqq 3$	$(x_1 \wedge x_2) \vee (x_1 \wedge x_3) \vee (x_1 \wedge x_4)$ $\vee (x_2 \wedge x_3 \wedge x_4)$

8.2.2.6. Rechenregeln

Für das Rechnen mit BOOLEschen Ausdrücken gibt es wichtige Gleichungen, die im folgenden zusammengestellt werden sollen. Zunächst seien Operationen mit 1 und Ø angeschrieben:

$$\bar{Ø} = 1\,, \qquad \bar{1} = Ø\,, \tag{1}$$

$$x \wedge Ø = Ø\,, \qquad x \vee Ø = x\,, \tag{2}$$

$$x \wedge 1 = x\,, \qquad x \vee 1 = 1\,. \tag{3}$$

Wie hier bereits begonnen, werden links die Ausdrücke mit Konjunktion und rechts mit Disjunktion geschrieben. Bezüglich beider besteht eine Dualität. Werden bei einer Gleichung Konjunktion und Disjunktion vertauscht, so bleibt die Gleichung richtig, wenn auch 1 und Ø vertauscht werden. Der eigentliche Inhalt liegt, wie Gl. (13) zeigen wird, noch tiefer.

Bei den BOOLEschen Ausdrücken gibt es *Tautologien* bzw. *Redundanzen*, die das Entstehen von Ausdrücken wie $2x$ oder x^2 entsprechend der Zahlenalgebra verhindern:

$$x \wedge x = x\,, \qquad x \vee x = x\,. \tag{4}$$

Ähnlich wirkt auch das *Absorptions-* oder *Verschmelzungsgesetz*

$$x_1 \wedge (x_1 \vee x_2) = x_1\,, \qquad x_1 \vee (x_1 \wedge x_2) = x_1\,. \tag{5}$$

Für die *Negation* folgt hieraus

$$x \wedge \bar{x} = Ø\,, \qquad x \vee \bar{x} = 1 \tag{6}$$

und für die *doppelte Negation*

$$\bar{\bar{x}} = x\,. \tag{7}$$

Weiter gelten die mit Gl. (5) oft als *Vereinfachungsidentitäten* oder Kürzungsregeln bezeichneten Ausdrücke:

$$x_1 \wedge (x_1 \vee x_2) = x_1 \wedge x_2\,, \qquad x_1 \vee (x_1 \wedge x_2) = x_1 \vee x_2\,, \tag{8}$$

$$(x_1 \vee x_2) \wedge (\bar{x}_1 \vee x_2) = x_2\,, \qquad (x_1 \wedge x_2) \vee (\bar{x}_1 \wedge x_2) = x_2\,. \tag{9}$$

Ähnlich wie bei Zahlen gelten zwei *kommutative (Vertauschungs-) Gesetze*

$$x_1 \wedge x_2 = x_2 \wedge x_1\,, \qquad x_1 \vee x_2 = x_2 \vee x_1\,, \tag{10}$$

zwei *assoziative (Verbindungs-) Gesetze*

$$\begin{aligned} & x_1 \wedge x_2 \wedge x_3 && x_1 \vee x_2 \vee x_3 \\ &= (x_1 \wedge x_2) \wedge x_3 && = (x_1 \vee x_2) \vee x_3 \\ &= x_1 \wedge (x_2 \wedge x_3)\,, && = x_1 \vee (x_2 \vee x_3) \end{aligned} \tag{11}$$

und im Gegensatz zur Rechnung bei Zahlen auch zwei *distributive (Verteilungs-) Gesetze*

$$\begin{aligned} & x_1 \wedge (x_2 \vee x_3) && x_1 \vee (x_2 \wedge x_3) \\ &= (x_1 \wedge x_2) \vee (x_1 \wedge x_3)\,, && = (x_1 \vee x_2) \wedge (x_1 \vee x_3)\,. \end{aligned} \tag{12}$$

Die zuvor erwähnte Symmetrie faßt das *Gesetz von* DE MORGAN zusammen:

$$\overline{x_1 \wedge x_2} = \bar{x}_1 \vee \bar{x}_2\,, \qquad \overline{x_1 \vee x_2} = \bar{x}_1 \wedge \bar{x}_2\,. \tag{13}$$

Allgemeiner besagt es, daß bei Negation eines Ausdruckes die Negation seiner Glieder vorgenommen werden kann, wenn zusätzlich Disjunktion und Konjunktion vertauscht werden. Deshalb kann dieses Gesetz auch in dem SHANNONschen *Satz* verallgemeinert werden:

$$\overline{F(x_1, x_2, \ldots, x_n, \vee, \wedge)} = F(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n, \wedge, \vee)\,. \tag{14}$$

Die kommutativen, assoziativen und distributiven Gesetze lassen sich ohne Schwierigkeiten ebenfalls auf viele Variable ausdehnen. Wegen seiner großen Bedeutung sei hier auch noch der *Entwicklungssatz von* SHANNON angeführt:

$$\begin{aligned} F(x_1, \ldots, x_i, \ldots, x_n) = {} & [x_i \wedge F(x_1, \ldots, x_{i-1}, 1, x_{i+1}, \ldots, x_n)] \\ & \vee [x_i \wedge F(x_1, \ldots, x_{i-1}, \emptyset, x_{i+1}, \ldots, x)]\,. \end{aligned} \tag{15}$$

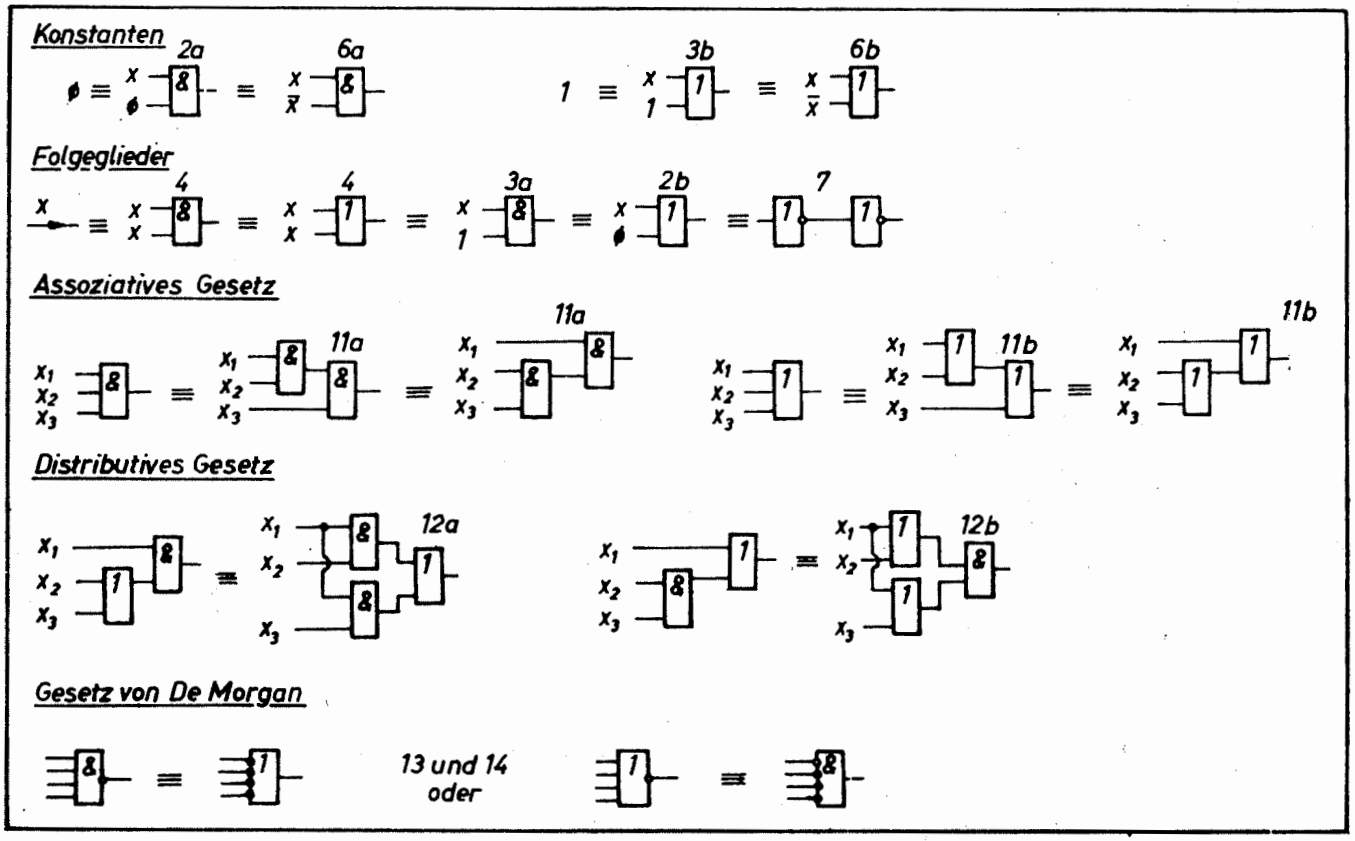


Bild 8.2.1. Beispiele für die symbolische Darstellung von Identitäten der BOOLEschen Algebra
Die Zahlen beziehen sich auf die entsprechenden Gleichungen, wobei *a* auf die linke und *b* auf die rechte Seite hinweisen

Alle genannten Gleichungen (Identitäten) lassen sich natürlich auch durch die äquivalenten Darstellungsformen: Wertetabelle, Kontaktnetzwerk, mengentheoretisch, im KARNAUGH-Diagramm oder in der symbolischen Darstellung angeben. Einige Beispiele in symbolischer Darstellung zeigt Bild 8.2.1. Die anderen sind nach diesem Prinzip leicht aufzustellen.

8.2.2.7. Normalformen

Infolge der vielfältigen möglichen, untereinander aber identischen Darstellungen einer BOOLEschen Funktion ist es notwendig, eine Bezugs-, d. h. Normalform zu besitzen. Allgemeine Bedeutung haben die disjunktive und konjunktive Normalform erlangt. Eine disjunkte Normalform besteht also aus disjunktiv verknüpften Konjunktionen. Ein Beispiel ist

$$y = (x_1 \wedge \bar{x}_3) \vee (x_4 \wedge x_3 \wedge \bar{x}_2) \vee (\bar{x}_1 \wedge \bar{x}_5) . \tag{16}$$

Bei der konjunktiven Normalform sind disjunktive Terme konjunktiv zusammengefaßt, also z.B.:

$$y = (x_1 \vee x_3) \wedge (\bar{x}_4 \vee \bar{x}_3 \vee x_2) \wedge (x_1 \vee x_5) . \tag{17}$$

Die jeweils in Klammer gesetzten Terme, also z.B. $x_1 \wedge \bar{x}_3$ bzw. $x_1 \vee x_3$, heißen konjunktive bzw. disjunktive Fundamentalterme oder auch gekürzte Terme, was später verständlich wird.

Für jede konkrete BOOLEsche Funktion gibt es viele solche Normalformen. Eine ausgezeichnete Form entsteht, wenn in allen Fundamentaltermen alle Variablen enthalten sind, also z.B.: $x_1 \vee \bar{x}_2 \vee x_3 \vee \bar{x}_4 \vee x_5$. Ein derartiger Term heißt dann *Vollterm* und die entstehende Normalform vollständig oder kanonisch. Oft wird dies vollständig oder kanonisch nicht besonders betont, sondern stillschweigend vorausgesetzt. Zuweilen wird statt Vollterm auch der Begriff Elementarterm verwendet. Er ist wegen der leichten Verwechslung mit Fundamentalterm jedoch nicht glücklich. Wie die *vollständige Normalform* aus der Wertetabelle zu gewinnen ist, zeigt Tab. 8.2.11. Entsprechend den Zeilen der Wertetabelle lassen sich Vollterme bilden. Sie enthalten alle Variablen x_1 bis x_n entweder in konjunktiver oder disjunktiver Verknüpfung. Die konjunktive Verknüpfung ist nur bei den Zeilen zulässig, wo der y-Wert (Ausgangswert) 1 ist. Aus ihnen entstehen Vollkonjunktionen, wenn überall dort, wo in der Zeile für x_n eine 1 steht, x_n gesetzt wird, und wo eine Ø steht, $\bar{x}_n$ gewählt wird. Es entstehen also genau so viele Vollkonjunktionen, wie Zeilen mit $y = 1$ existieren. Vollkonjunktionen heißen auch *Minterme*. Durch ihre disjunktive Verknüpfung entsteht schließlich die *disjunktive Normalform* (*DNF*). Sie wird zuweilen auch alternative Normalform (ANF) genannt. Soll die Vollständigkeit betont werden, so wird auch von kanonischer disjunktiver bzw. alternativer Normalform (KDF, KAF) gesprochen.

In Analogie zu dem eben geschilderten Vorgang kann entsprechend Tab. 8.2.11 über *Maxterme* (Volldisjunktionen) auch die *konjunktive Normalform* (KNF) erstellt werden. Beide Normalformen sind identisch. Dies läßt sich über den Satz von DE MORGAN beweisen. Daraus folgt, daß es meist vorteilhaft ist, jene Normalform zu entwickeln, die weniger Vollterme besitzt. Wenn die y-Spalte mehr 1 als Ø enthält, ist es die konjunktive, sonst die disjunktive Normalform. In Tab. 8.2.11 wurde bewußt ein Beispiel gewählt, in dem beide Normalformen gleich aufwendig sind. Enthielte nur die 7. Spalte bei y eine 1 und alle anderen eine Ø, so existiert nur der eine Minterm m_7, und die disjunktive Normalform würde lauten

$$y = x_1 \wedge x_2 \wedge x_3 . \tag{18}$$

Tabelle 8.2.11. Beispiel zur Erstellung der disjunktiven (alternativen) bzw. konjunktiven Normalform aus einer Wertetabelle mit zugehörigen Schaltungen

Wertetabelle

Zeile	x_1	x_2	x_3	y
Ø	Ø	Ø	Ø	1
1	Ø	Ø	1	1
2	Ø	1	Ø	Ø
3	Ø	1	1	Ø
4	1	Ø	Ø	Ø
5	1	Ø	1	Ø
6	1	1	Ø	1
7	1	1	1	1

Auswahl je nachdem, ob in Spalte y
weniger 1 oder weniger Ø

↙ ↘

Vollkonjunktion ← *Vollterm* → Volldisjunktion bzw. Vollalternative

Minterm enthält alle Eingangsvariablen *Maxterm*

Auswahl der Zeilen mit $y = 1$
$\bar{x}_n$ dann, wenn x_n in Tabelle Ø

↓

$m_0 = \bar{x}_1 \wedge \bar{x}_2 \wedge \bar{x}_3$
$m_1 = \bar{x}_1 \wedge \bar{x}_2 \wedge x_3$
$m_6 = x_1 \wedge x_2 \wedge \bar{x}_3$
$m_7 = x_1 \wedge x_2 \wedge x_3$

↓

Auswahl der Zeilen mit $y = Ø$
$\bar{x}_n$ dann, wenn x_n in Tabelle 1

↓

$M_2 = x_1 \vee \bar{x}_2 \vee x_3$
$M_3 = x_1 \vee \bar{x}_2 \vee \bar{x}_3$
$M_4 = \bar{x}_1 \vee x_2 \vee x_3$
$M_5 = \bar{x}_1 \vee x_2 \vee \bar{x}_3$

↓

vollständige oder kanonische *Normalform*

↓

vollständige oder kanonische disjunktive bzw. alternative Normalform
DNF
(DAF, KDF, KAF)

↓

$y = m_0 \vee m_1 \vee m_6 \vee m_7$

↘

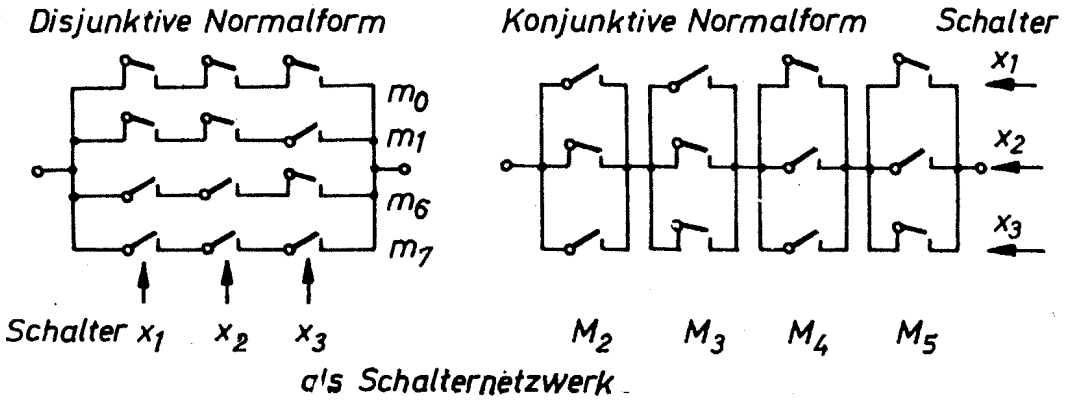
vollständige oder kanonische konjunktive Normalform
KNF
(KKF)

↓

$y = M_2 \wedge M_3 \wedge M_4 \wedge M_5$

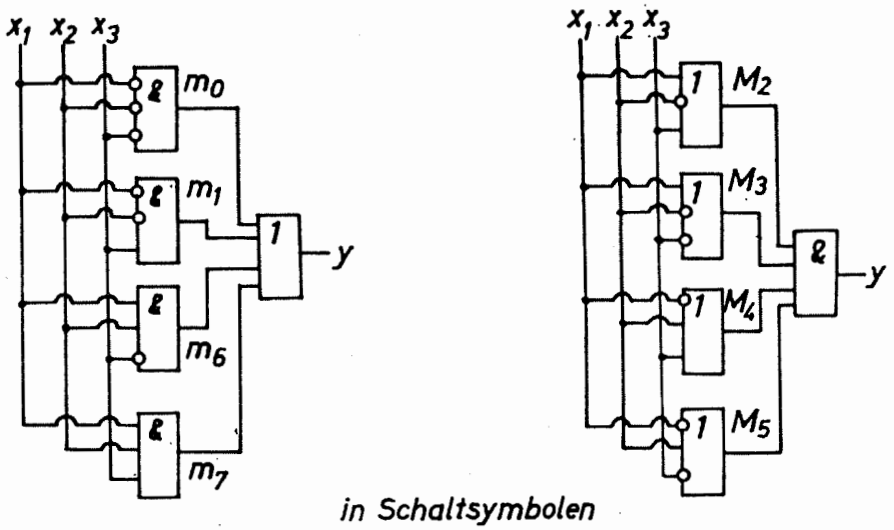
↙

beide Formen sind identisch



als Schalternetzwerk

Tabelle 8.2.11. (Fortsetzung)



Sie enthielte dann ausnahmsweise und paradoxerweise gar keine Disjunktion und wäre zugleich die minimalste Form.

Ergänzend zum geschilderten Ablauf zur Gewinnung der Normalformen sind in Tab. 8.2.11 noch die zu ihnen gehörenden Schaltnetzwerke und symbolischen Darstellungen gezeigt. Bei den Schaltnetzwerken entstehen also Parallel-Reihen-Schaltungen und bei den binären Elementen zweistufige Anordnungen aus Disjunktionen und Konjunktionen.

8.2.2.8. Minimierung

Die beiden kanonischen Normalformen sind in der Regel sehr aufwendig. Wie schon gesagt, dienen sie unter anderem zum Vergleich verschiedener Boolescher Funktionen. Für mehrere Verfahren sind sie aber auch Ausgangspunkt, um die *optimale Normalform* zu erhalten. Verfahren, die zu einer optimalen Form führen, heißen Kürzungs- oder Minimierungsmethoden. Es gibt heute eine größere Anzahl solcher Verfahren. Sie können hier nicht einmal aufgezählt werden. Am Beispiel von Tab. 8.2.11 sei jedoch ein prinzipieller Weg angedeutet. Ausgangspunkt sei die kanonische disjunktive Normalform. Sie lautet ausgeschrieben

$$y = (\bar{x}_1 \wedge \bar{x}_2 \wedge \bar{x}_3) \vee (\bar{x}_1 \wedge \bar{x}_2 \wedge x_3) \vee (x_1 \wedge x_2 \wedge \bar{x}_3) \vee (x_1 \wedge x_2 \wedge x_3)\,. \tag{19}$$

Mittels des assoziativen Gesetzes (11) lassen sich je einmal $\bar{x}_1 \wedge \bar{x}_2$ und $x_1 \wedge x_2$ ausklammern:

$$y = [(\bar{x}_1 \wedge \bar{x}_2) \wedge (\bar{x}_3 \vee x_3)] \vee [(x_1 \wedge x_2) \wedge (\bar{x}_3 \vee x_3)]\,. \tag{20}$$

Wegen (6b) bezüglich $\bar{x}_3 \vee x_3$ folgt weiter

$$y = [(\bar{x}_1 \wedge \bar{x}_2) \wedge 1] \vee [(x_1 \wedge x_2) \wedge 1]\,. \tag{21}$$

Mit (3a) gilt so schließlich

$$y = (\bar{x}_1 \wedge \bar{x}_2) \vee (x_1 \wedge x_2)\,, \tag{22}$$

die Beziehung für die Äquivalenz (Spalte 9 in Tab. 8.2.7). Es ist also x_3 ersichtlich redundant.

Nach der eben gegebenen Methode lassen sich prinzipiell immer einfachste Formen gewinnen, doch ist der Weg stark von Erfahrung und Intuition abhängig. Er kann also so nicht dem Rechner übertragen werden. Deshalb sind systematische Methoden

Tabelle 8.2.12. Zur Erklärung des Kürzungsverfahrens von QUINE und MCCLUSKEY

Minterme ohne Zeichen	1. Kürzungen							Einmal gekürzte Minterme	2. Kürzungen		Zweimal gekürzte Minterme
	1	2	3	4	5	6	7		1	2	
$\bar{x}_1\ \bar{x}_2\ x_3\ \bar{x}_4$	×							$\bar{x}_1\ x_3\ \bar{x}_4$			$\bar{x}_1\ x_2$ ←
$\bar{x}_1\ x_2\ \bar{x}_3\ \bar{x}_4$		×		×				$\bar{x}_1\ x_2\ \bar{x}_3$	×		$\bar{x}_1\ x_2$ ←
$\bar{x}_1\ x_2\ \bar{x}_3\ x_4$		×		×	×			$\bar{x}_1\ x_2\ \bar{x}_4$		×	
$\bar{x}_1\ x_2\ \bar{x}_3\ \bar{x}_4$	×		×			×					
$\bar{x}_1\ x_2\ x_3\ x_4$				×		×		$x_2\ \bar{x}_3\ x_4$			
$x_1\ \bar{x}_2\ \bar{x}_3\ x_4$							×	$\bar{x}_1\ x_2\ x_3$	×		
$x_1\ x_2\ \bar{x}_3\ x_4$					×		×	$x_1\ \bar{x}_3\ x_4$			

Ergebnis: $y = (\bar{x}_1 \wedge x_3 \wedge \bar{x}_4) \vee (x_2 \wedge \bar{x}_3 \wedge x_4) \vee (x_1 \wedge \bar{x}_3 \wedge x_4) \vee (\bar{x}_1 \wedge x_2)$

entwickelt worden. Eine ist die Methode von QUINE und MCCLUSKEY. Damit sie durchsichtig wird, ist zumindest von vier Variablen auszugehen. Das Schema demonstriert dann Tab. 8.2.12. In der ersten Spalte werden alle Minterme zur Vereinfachung ohne $\wedge$-Zeichen eingetragen. Dann wird nach Mintermen gesucht, die bis auf eine Variable übereinstimmen. Diese beiden werden abgehakt (z.B. bei 1 $\bar{x}_2$ und x_2), und der um diese Variable gekürzte Ausdruck wird als neuer Minterm eingetragen. Im Beispiel sind so sieben Kürzungen möglich, die zu sieben gekürzten, d. h. dreistelligen Mintermen führen. Hier wiederholt sich dieser Vorgang. Er ist zweimal durchführbar. Am Ende dieses Vorganges werden die nicht gekürzten (abgehakten) Minterme mit Pfeilen gekennzeichnet, und aus ihnen folgt eine optimale Normalform. Für die Beispielgleichung (19) entsteht so das ganz einfache Bild der Tab. 8.2.13.

Tabelle 8.2.13. Verfahren nach QUINE und MCCLUSKEY für Gl. (19)

Minterme	Kürzungen	Neue Minterme
$\bar{x}_1\ \bar{x}_2\ \bar{x}_3$	×	$\bar{x}_1\ \bar{x}_2$ ←
$\bar{x}_1\ \bar{x}_2\ x_3$	×	$x_1\ x_2$ ←
$x_1\ x_2\ \bar{x}_3$	×	
$x_1\ x_2\ x_3$	×	

Ergebnis: $y = (\bar{x}_1 \wedge \bar{x}_2) \vee (x_1 \wedge x_2)$

8.2.2.9. Schaltungen mit mehreren Ausgängen

Alle bisherigen Betrachtungen waren darauf ausgerichtet, daß von mehreren Eingängen ein einzelnes Signal zu einem Ausgang gebildet wird. Dies ist für die Praxis relativ selten erfüllt. Meist werden von mehreren Eingangsgrößen $x_1, \ldots, x_n$ mehrere unterschiedliche Ausgangssignale $y_1, \ldots, y_m$ erzeugt. Es entsteht dabei eine allgemeine Schaltung, wie sie Bild 8.2.2a) zeigt. Im Prinzip kann eine solche Schaltung aus m Schaltungen mit je einem Ausgang zerlegt und entsprechend berechnet werden. Hierdurch entsteht jedoch mit großer Wahrscheinlichkeit ein zu hoher Aufwand. Im Innern der einzelnen Schaltungen sind nämlich verschiedene Fundamentalterme zu bilden, die durchaus bei mehreren Schaltungen auftreten können. Sie brauchten in der verbundenen Schaltung aber nur einmal realisiert zu werden. Deshalb ist es meist

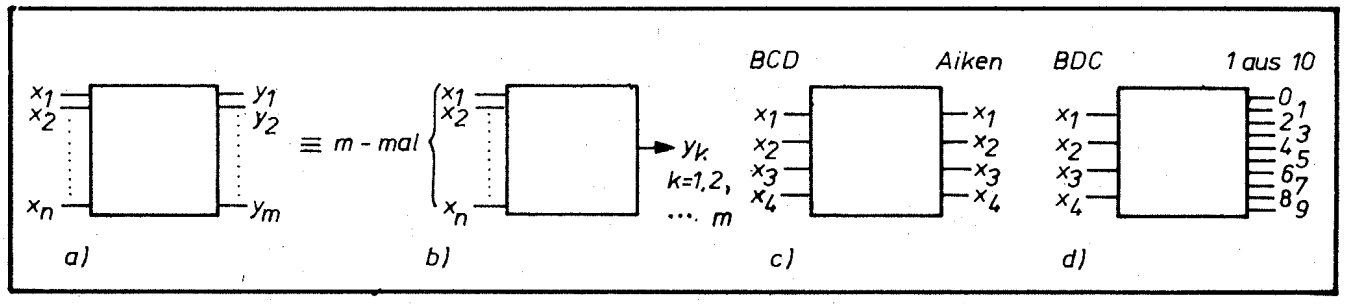


Bild 8.2.2. a) Kombinatorische Schaltung mit mehreren Ein- und Ausgängen und b) Ersatz durch kombinatorische Schaltung mit nur einem Ausgang. Beispiele für Kodeumwandler c) von BCD nach AIKEN, d) von BCD nach 1 aus 10

sinnvoller, die komplexe Schaltung mit allen Ausgängen als Einheit zu realisieren. Bild 8.2.2 zeigt auch zwei einfache Beispiele für solche kombinatorischen Schaltungen.

8.2.3. Prinzipien technischer Realisierungen

Ursprünglich bestanden digitale Schaltungen als mechanische Gebilde. Noch K. ZUSE baute seine Rechenmaschine Z 1 1942 mit z. T. *mechanischen, bistabilen* Elementen.

Mit der Fernsprech- und Telegrafentechnik gewann jedoch schon um die Jahrhundertwende das *Relais* einen entscheidenden Einfluß. Noch heute werden die meisten technischen Automatisierungsanlagen, zumindest aber in der Kleinautomatisierung und viele Fernsprechvermittlungen, mit Relais aufgebaut. Das liegt an einigen Vorteilen: Durch die Kontakte kann der Strom ungehindert in beiden Richtungen fließen. Von einer Spule können gleichzeitig viele Kontakte (Arbeits- und Ruhekontakte) betätigt werden, und schließlich liegen hier umfangreiche Einsatz- und Entwurfserfahrungen vor. Nachteile des Relais sind jedoch seine begrenzte Schaltgeschwindigkeit und Zuverlässigkeit sowie der große Leistungsverbrauch.

Mit der Entwicklung der *Halbleitertechnik* und insbesondere der integrierten Schaltkreise ist hier eine neue Situation entstanden. Es wurde eine Vielzahl von Schaltkreisfamilien entwickelt, welche direkt BOOLEsche Funktionen realisieren. Es wird eingeschätzt, daß ca. 50 solche Familien vorhanden sind bzw. z. T. schon wieder der Vergangenheit angehören. Die allgemein besonders wichtigen Familien TTL, ECL, I^2L und die verschiedenen MOS werden im folgenden getrennt behandelt. Danach folgt ein genereller Vergleich. In diesem Abschnitt seien zuvor einige Grundprinzipien auf der Basis der nur noch bedingt vorhandenen Dioden- bzw. RTL-Logik eingeführt.

8.2.3.1. Diodenschaltungen

Bereits im Abschnitt 5.4.3. wurde gezeigt, daß einer Diode im wesentlichen zwei Zustände zugeordnet werden können: leitend und gesperrt. In Idealisierung zeigt dies jetzt Bild 8.2.3a). Hierauf bauen die binären Schaltungen mit Dioden auf. Sie werden auch verkürzt, aber sprachlich unrichtig mit Diodenlogik bezeichnet. Die entsprechenden Grundschaltungen zeigt Bild 8.2.3b) und c). Zur einfacheren Betrachtung werden hier die möglichen Eingangskodierungen als zeitliche Folge angenommen. Dann entsteht am Ausgang ebenfalls eine zeitliche Folge.

Die Wirkung der Disjunktionsschaltung b) ist unmittelbar einsichtig. Sobald nur einer der beiden Eingänge positiv wird, ist die entsprechende Diode leitend und

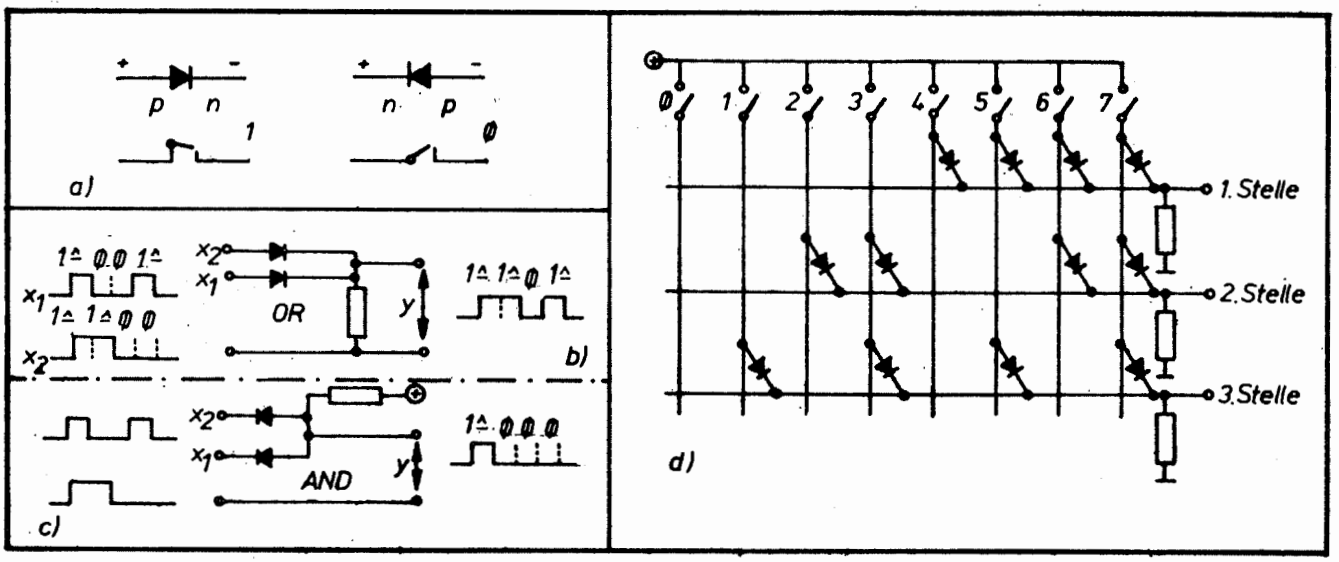


Bild 8.2.3. Zur Diodenlogik

a) Eine Diode wirkt je nach Polung als Schalter und realisiert die logischen Werte Ø und 1
b) OR-Schaltung aus zwei Dioden und Wirkung einer Signalfolge
c) Dito für AND-Schaltung
d) Aufbau einer Diodenmatrix, mit der die dezimalen Werte Ø bis 7 in die entsprechenden dreistelligen binären Zahlen gewandelt werden

gibt das Signal zum Ausgang weiter. Die andere Diode wird dabei (noch stärker) gesperrt. Dies ist wichtig, damit die Eingänge entkoppelt bleiben. Sind beide Eingänge negativ, so sind beide Dioden gesperrt. Sind beide Eingänge positiv, so sind beide Dioden leitend. In diesem Fall kann die gegenseitige Beeinflussung keine Störungen hervorrufen.

Bei der Konjunktionsschaltung c) sind die Dioden andersherum gepolt, und außerdem ist der Ausgang über den Widerstand positiv vorgespannt. Diese Spannung wird am Ausgang jedoch nur wirksam, wenn beide Dioden gesperrt sind, also an beiden Eingängen eine mindestens gleich große positive Spannung liegt. In jedem anderen Fall ist zumindest eine der Dioden leitend. Dann wird das negativere Eingangspotential zum Ausgang durchgeschaltet.

Nach dem Vorangegangenen ist mit Dioden relativ leicht die Disjunktion und Konjunktion zu realisieren. Für ein vollständiges Logiksystem fehlt also noch die Negation. Sie ist nicht auf Diodenbasis realisierbar. Deshalb sind reine Diodenschaltungen nur für begrenzte Anwendungsfälle brauchbar. Beispiele hierfür sind u. a. Diodenmatrizen zur Kodierung. Eine entsprechende Schaltung zeigt Bild 8.2.3d). Schließlich sei erwähnt, daß komplexe Diodenschaltungen aus Disjunktion und Konjunktion kaum mehr praktische Bedeutung besitzen, da jede Diodenschaltung eine gewisse Dämpfung aufweist.

8.2.3.2. Transistor als Negator

Bereits im Abschnitt 5.7.5. wurde der Transistor als Schalter behandelt. Im wesentlichen kommt es darauf an, zwei relativ genau bestimmte Zustände zu definieren: leitend und gesperrt. Im einen Fall fließt durch das Bauelement ein relativ starker Strom bei möglichst kleinem Spannungsabfall. Im anderen Fall ist der Stromfluß nahezu Null, dafür liegt an ihm aber nahezu die gesamte Betriebsspannung. Dies ist im Bild 5.7.12 und 13 auf der Kennlinie durch entsprechend gekennzeichnete Punkte hervorgehoben. Wichtig für den Betrieb von Transistoren als binäre Bauelemente ist, daß bereits die Eingangsgrößen (Steuergrößen) so gewählt werden, daß nur diese

beiden Gebiete im Betrieb angenommen werden können. Bei echt binären Signalen ist nur der Pegel passend zu wählen. In anderen Fällen sind vorher Schaltungen zur Pegelanpassung anzuordnen. Besondere Bedeutung besitzen hierbei die verschiedenen Impulsschaltungen gemäß Abschnitt 7.3.8. Unter diesen Voraussetzungen ergibt sich die Grundschaltung des Transistors als Negator gemäß Bild 8.2.4. Neben Negator sind hierfür auch die Begriffe Inverter oder Buffer gebräuchlich.

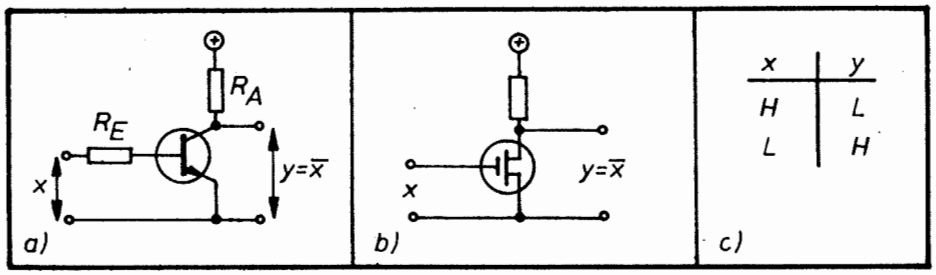


Bild 8.2.4. Prinzip des Negators
a) mit bipolarem Transistor,
b) mit FET
c) Wertetabelle

Nach den Betrachtungen der Funktionsweise des Transistors ist die Wirkungsweise dieser Schaltung unmittelbar einsichtig: Hohes positives Potential am Eingang macht den Transistor leitend, und die Betriebsspannung fällt am Widerstand R_A ab. Am Ausgang liegt also ein relativ niedriges Potential, nämlich die Restspannung (Kniespannung) des Transistors. Der Widerstand R_E dient zur Begrenzung des Eingangsstromes. Beim FET entfällt er. Niedriges Potential am Eingang läßt den Basisstrom so klein werden, daß der Transistor praktisch gesperrt ist. Folglich fällt an R_A nur eine ganz geringe Spannung ab, und am Ausgang liegt fast die gesamte Betriebsspannung.

8.2.3.3. Positive und negative Logik

Im Abschnitt 8.1.4., insbesondere in 8.1.4.9., wurde deutlich, daß im Prinzip eine relativ willkürliche Zuordnung zwischen den binären Zahlen und deren Bedeutung möglich ist. So konnten dieselben Zahlen u. a. dezimale Zahlen, Adressen, Befehle oder Buchstaben bedeuten. In genau demselben Sinn tritt dies nun noch einmal bezüglich der Potentialzustände einer Schaltung auf. Es existieren dabei zunächst generell zwei Pegel:

H High; Hoch (näher $+\infty$)
L Low; Tief (näher $-\infty$).

In diesem Sinne ist bereits die Pegeltabelle in Bild 8.2.4 belegt. Diese Bezeichnung ist deshalb so gewählt, weil sie ohne Einschränkung für alle Schaltungen gültig ist. Dies zeigen die folgenden Beispiele:

−2 V	+3 V	+5 V	H
−10 V	−2 V	+1 V	L

Damit besteht nun die Möglichkeit der Interpretation dieser Pegel, und es sind dabei zwei Varianten zulässig:

positive Logik: H → 1 L → Ø: H aktiv
negative Logik: H → Ø L → 1: L aktiv.

Sie werden also mit positiver bzw. negativer Logik oder, was äquivalent ist, mit H bzw. L aktiv bezeichnet. Im allgemeinen wird — insbesondere wenn nichts anderes betont wird — die positive Logik verwendet. Die negative Logik hat jedoch u. a. große praktische Bedeutung beim IEC-Bus (8.7.). Mit dem *Übergang von positiver zu negativer Logik* wird der Satz von DE MORGAN wirksam. Dies bedeutet, daß sich u. a. folgende BOOLEsche *Funktionen vertauschen*:

AND ↔ OR
NOT ↔ NOT
NAND ↔ NOR.

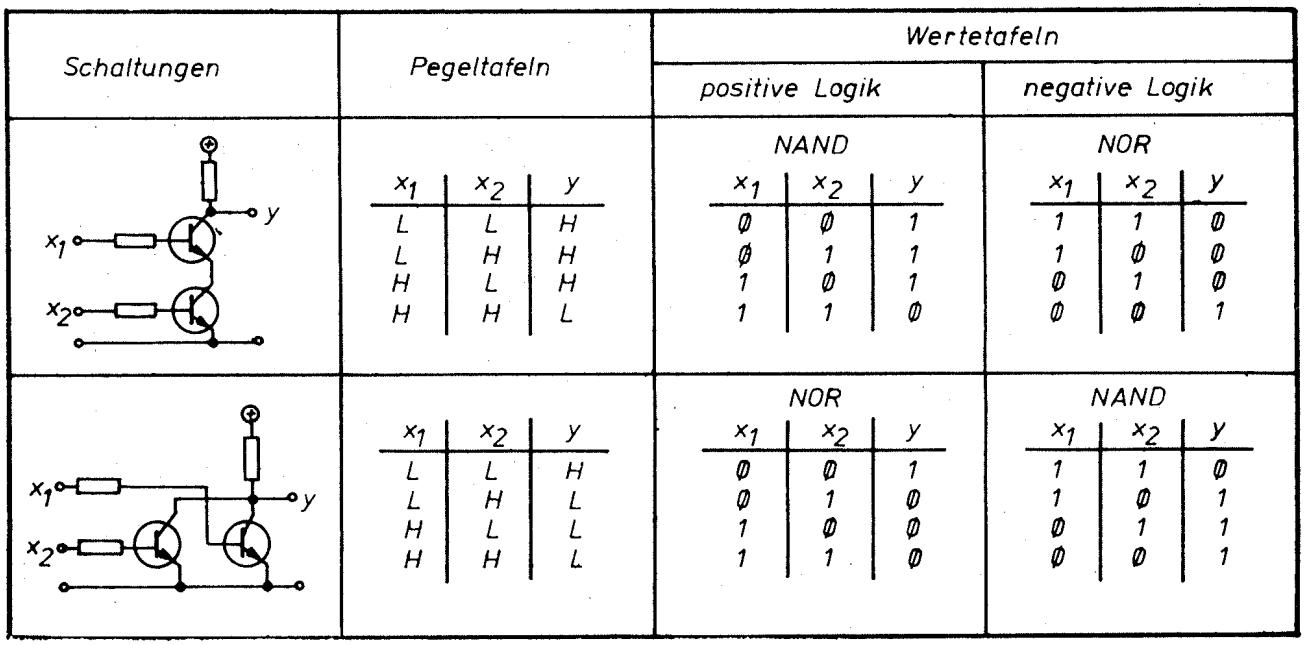


Bild 8.2.5. Zur Erklärung von positiver und negativer Logik am Beispiel von zwei binären Schaltungen

Am Beispiel für zwei Schaltungen ist dieser Zusammenhang in Bild 8.2.5 gezeigt. Dies bedeutet, daß die Funktion einer Schaltung erst dann vollständig bestimmt ist, wenn auch die bei ihr genutzte Logik (positiv oder negativ) mit angegeben ist. In diesem Sinne gelten die beiden Angaben bei den Diodenschaltungen in Bild 8.2.3 auch nur für positive Logik. Würde dagegen der Pegel Ø V = L als 1 und +5 V = H als Ø interpretiert, so käme dort die negative Logik zur Anwendung, und die Bezeichnungen OR und AND müßten vertauscht werden. Es sei weiter darauf hingewiesen, daß Pegeltafeln nur die Werte L und H verwenden und daher von der Logik (positiv bzw. negativ) unabhängig sind. Die *Wertetafeln*, auch Wahrheitstafeln genannt, verwenden dagegen die binären Werte Ø und 1, die auch als Wahrheitswerte interpretiert werden können. Sie haben dann schon die Auswahl der Logik berücksichtigt. Sie müßte bei ihnen im strengen Sinne angegeben werden. Wenn jedoch nichts erwähnt wird, gilt stillschweigend die positive Logik.

8.2.3.4. Pegelbereiche und Fan-out

Bisher wurde von den Pegeln L und H gesprochen. In Wirklichkeit handelt es sich aber hierbei um Pegelbereiche. Denn es dürfen sich nicht Änderungen der Betriebsspannung, der Temperatur und der Bauelemente darauf auswirken, ob eine logische Ø oder 1 vorliegt. Dies bedeutet weiter, daß zwischen den beiden Pegelbereichen ein *verbotener Bereich* liegen muß, den die Ausgangs- und Eingangspegel statisch nicht

annehmen dürfen, sondern nur beim Umschalten schnell durchlaufen müssen. Doch damit sind die Pegelbereiche immer noch nicht ausreichend bestimmt. In der Anwendung werden nämlich von dem Ausgang einer Schaltung unterschiedlich viele andere Eingänge anzusteuern sein. Der Ausgang einer Schaltung kann also unterschiedlich belastet sein. Auch die so verschiedenen Belastungen müssen noch garantieren, daß die entsprechenden *Toleranzfelder*, in denen eindeutig H und L definiert ist, nicht verlassen werden. Faßt man alle diese Einflüsse zusammen, so ergibt sich ein schematisches Pegelbild, wie es Bild 8.2.6 zeigt. Es existieren also unterschiedliche

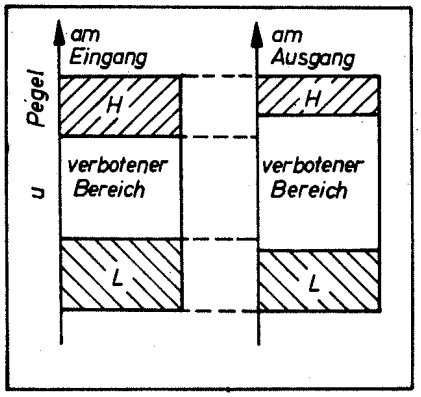


Bild 8.2.6. Prinzipielle Definition und Zuordnung der Pegelbereiche für H und L am Eingang und Ausgang einer Schaltung

Pegelbereiche am Eingang und Ausgang der Schaltung, wobei der Ausgang immer enger toleriert ist. Die zugehörigen Grenzen können wie folgt interpretiert werden: Solange der Pegel am Eingang einer Schaltung — ein einzelner Eingang heißt auch Gatter — im zulässigen H- bzw. L-Bereich bleibt, wird intern die Schaltung voll durchgesteuert. Die Schaltung arbeitet also intern exakt fehlerfrei. Am Ausgang entstehen dann Pegel, die in den Ausgangstoleranzfeldern für H und L liegen. Dies gilt unabhängig von zulässigen Änderungen der Betriebsspannung und Temperatur, von Alterung und Belastungen. Da dieser Pegelbereich enger als jener der Eingangswerte ist, wird bei einer Zusammenschaltung selbst unter Einschluß eventueller zusätzlicher Verluste die Funktion der hiermit angesteuerten weiteren Schaltung (Gatter) garantiert. Die *Anzahl von Schaltungen* (Gatter), die maximal an den Ausgang einer Schaltung angeschaltet werden dürfen, *heißt Fan-out*, was soviel wie Auffächerung bedeutet.

8.2.3.5. Innere und äußere Logik

Die vorangegangenen Pegelbetrachtungen machten deutlich, daß es notwendig ist, in einer logischen Schaltung zwei Teile zu unterscheiden: jenen Teil, der die logische Verkopplung der *Eingangssignale* realisiert, und die Endstufe, welche das dabei entstehende Eingangssignal hinreichend verstärkt und zugleich einen so kleinen Innenwiderstand besitzt, daß genügend Gatter weiterer Schaltungen ohne Verlassen der Pegelbereiche angeschaltet werden können. Die *Endstufe* verbraucht damit in der Regel eine beachtliche Energie. Anderseits sind die Pegelbereiche, wie sie in Bild 8.2.6 gezeigt sind, mit der relativ großen verbotenen Zone aber auch nötig, um von außen einwirkende Störungen nicht wirksam werden zu lassen. Alle diese Betrachtungen gelten folglich genau dann, wenn *einzelne logische Schaltungen* existieren, die *über Leitungen miteinander* zu einer komplexen Schaltung *verbunden* werden sollen. So etwas wird als *äußere Logik* bezeichnet.

Die Verhältnisse werden z. T. wesentlich anders, wenn mehrere *logische Teilschaltungen auf einem Chip*, als großintegrierter Schaltkreis verbunden, zusammengefaßt werden. Dann sind die Leitungen extrem kurz, die Kapazitäten der Leitungen folglich sehr klein. Weiter können auf diese Leitungen kaum Störungen von außen dringen. Dadurch kann der verbotene Bereich verkleinert werden. Weiter kann die Belastung beim Entwurf der Schaltung genau berücksichtigt werden, denn sie ist nicht mehr von außen beeinflußt. Diese Fakten bewirken zusammen, daß für die logischen Verschaltungen auf einem Chip in der Regel einfache Forderungen gelten. Deshalb wird in solchen Fällen von *innerer Logik* gesprochen. Sie ist in der Regel *wesentlich einfacher*. Allerdings müssen bei solcher Logik an den Anschlüssen, die nach außen gehen (wo wieder die äußere Logik wirksam wird), Abschlußstufen höherer Leistung vorgesehen werden, die Pegel und Störabstand (verbotene Zone) auf die Standardwerte anpassen. In der Regel ist das heute der TTL-Pegel. Alle folgenden Schaltfamilien mit Ausnahme der I^2L gehören im wesentlichen zur äußeren Logik.

8.2.4. TTL-Technik

TTL rührt von der Abkürzung *Transistor-Transistor-Logik* her und weist damit aus, daß ihre logische Funktion durch die Wechselwirkung von Transistoren erreicht wird.

Die TTL-Technik entstand als erste leistungsfähige integrierte Technik 1962. Inzwischen sind daraus mindestens fünf verschiedene Typenreihen hervorgegangen, wobei bedeutsam ist, daß wichtige Parameter die einzelnen Reihen zueinander kompatibel und z. T. sogar austauschbar machen. Die Fortschritte, die bei der TTL-Technik in den vergangenen zwanzig Jahren erreicht wurden, sind beachtlich. Auch auf anderen Gebieten wurden große Fortschritte gemacht. Dennoch hat sich das Grundkonzept der TTL-Technik hervorragend für viele Anwendungen bewährt, und es hat den Anschein, daß dies auch weiterhin der Fall ist. Viele Parameter aller integrierten Schaltungen werden heute auf TTL-Pegel und TTL-Lastfaktoren bezogen. Es ist also nicht nur die am breitesten angewendete Logikfamilie, sondern auch viele ihrer Daten sind Bezugspunkte für die meisten Betrachtungen. In diesem Sinne sind zu betrachten:

- die mit TTL spezifisch realisierte Logikfunktion, einschließlich geringfügiger Abwandlungen durch die Weiterentwicklung;
- die verschiedenen Varianten von Endstufen;
- die entscheidenden Daten, einschließlich jener der verschiedenen Typenreihen.

8.2.4.1. Die typische Eingangsstufe

Eine typische Eingangsstufe der TTL-Technik zeigt Bild 8.2.7. Zur Vereinfachung sei zunächst von der Diode D und dem Widerstand R_1' abgesehen. Dann existiert ein Multiemittertransistor T_1. In fast allen Fällen sind es 4 Emitter, die um eine Basis B' herum gelagert sind. Ob davon 2, 3 oder 4 verwendet werden, hängt von dem jeweiligen TTL-Gatter ab. Es gibt noch ein zweites Lay-out mit acht Eingängen. Wichtig für alle Anwendungen ist, daß die einzelnen Emitter gleich groß und gleich wirksam sind.

Es können generell *zwei Betriebszustände* unterschieden werden:

- ein oder mehrere Eingänge liegen auf L, das ist etwa 0 V
- alle Eingänge liegen auf H, das ist etwa $U_B \approx +5$ V.

Sofern auch nur ein Eingang auf L liegt, wird die Basis-Emitter-Strecke von T_1 leitend, und damit liegt auch der Punkt K auf tiefem Potential, d. h. T_2 ist gesperrt,

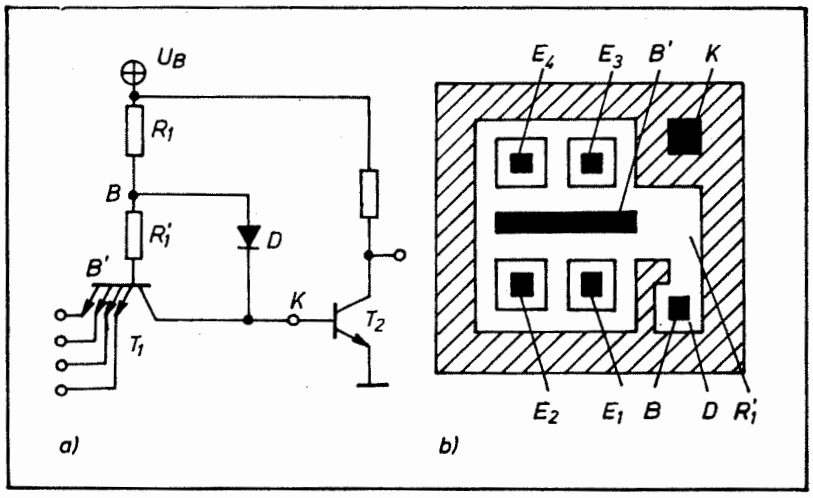


Bild 8.2.7. a) Schaltbild und b) Struktur der Eingangsstufe bei Standard-TTL-Schaltungen

und der Ausgang liegt auf H. Für diesen Zustand ist es offensichtlich gleichgültig, ob 1, 2, 3 oder alle Emitter auf L liegen; mehr als leitend kann T_1 nicht werden. Dies wird oft so ausgedrückt:

L am Eingang greift durch und bewirkt am Ausgang H.

Wenn *alle* Eingänge an H liegen (gleichberechtigt damit ist: überhaupt nicht angeschlossen sind), ist T_1 gesperrt. Für die Basis-Emitter-Strecke von T_2 besteht damit die Möglichkeit, die Kollektor-Basis-Strecke in den leitenden Zustand zu versetzen. Dies bedeutet, daß T_1 invers betrieben wird: Kollektorseite leitend, Basisseite gesperrt. Dieser Strom bringt T_2 in den leitenden Zustand, und der Ausgang wird L:

Nur wenn alle Eingänge H besitzen (bzw. nicht angeschlossen sind), entsteht am Ausgang L.

Faßt man diese beiden Aussagen zusammen, so liegt mit dem Grundgatter der TTL-Schaltung bei positiver Logik also ein

NAND

vor.

Die Diode D und die Aufteilung von R_1 in Bild 8.2.7a) garantieren, daß beim inversen Betrieb von T_1 zwischen Kollektor und Basis maximal die etwa 0,7 V ihrer Durchlaßspannung abfallen können. Dadurch wird vom Punkt K der Stromverstärkungsfaktor auf 5 bis 10 begrenzt. Der Widerstand R_1' ist klein gegen R_1. In allen üblichen Schaltbildern wird zur Vereinfachung D und R_1' weggelassen.

8.2.4.2. Ausgangsstufe der Standard-TTL

Der Eingangsstufe, die für die logische Verknüpfung der Eingangsgrößen sorgt, ist in jedem heutigen TTL-Schaltkreis eine Ausgangsstufe nachgeschaltet. Sie hat erstens dafür zu sorgen, daß möglichst viele Gatter nachgeschaltet werden können. Wie im Abschnitt 8.2.3.4. gezeigt wurde, besteht hierfür oft die Notwendigkeit. Es wird in diesem Zusammenhang von Auffächerung, oder wenn es auf den Schaltkreis bezogen wird, von Ausgangslastfaktor oder *Fan-out* gesprochen. Dieser Wert gibt an, wieviel Eingänge an einen Ausgang geschaltet werden dürfen. Für die üblichen Ausgangsstufen beträgt dieser Wert zehn.

Zum zweiten muß die Ausgangsstufe einen *geringen dynamischen Innenwiderstand* besitzen. Er ist notwendig, damit kapazitive Lasten nicht die Verzögerungszeit des Schaltkreises unnötig vergrößern.

Zum dritten *erhöht* die Endstufe den *statischen Störabstand* beträchtlich. Die Eingangsschaltschwelle des Transistors T_1 liegt nämlich bei etwa 0,7 V. Hier geht er in den inversen Betrieb über. Andererseits liegt die Restspannung des Transistors bei etwa 0,2 V. So ergäbe sich ein Sicherheitsspielraum von nur 0,5 V.

Die meist gebrauchte Standardausgangsschaltung zeigt Bild 8.2.8, und zwar gleich in den beiden typischen Zuständen mit angeschriebenen Strömen und Potentialen sowie bei Belastung durch mehrere nachfolgende Schaltkreise. Der Transistor T_2 erzeugt über die Widerstände R_2 und R_3 zwei gegenphasige Spannungen. Mit ihnen werden die beiden Endtransistoren T_3 und T_4 so angesteuert, daß zu jedem Eingangszustand jeweils nur einer leitend ist.

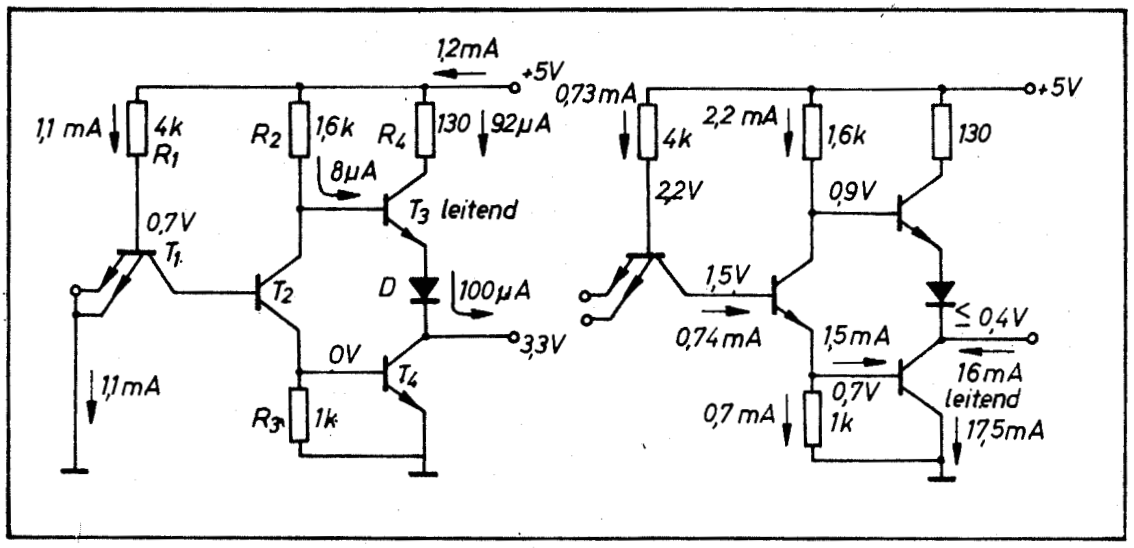


Bild 8.2.8. Die beiden Zustände der Standard-TTL-Schaltung bei Belastung durch weitere TTL-Stufen (alle Transistoren sind vom Typ *npn*)

a) Eingang L, Ausgang H b) Eingang H, Ausgang L

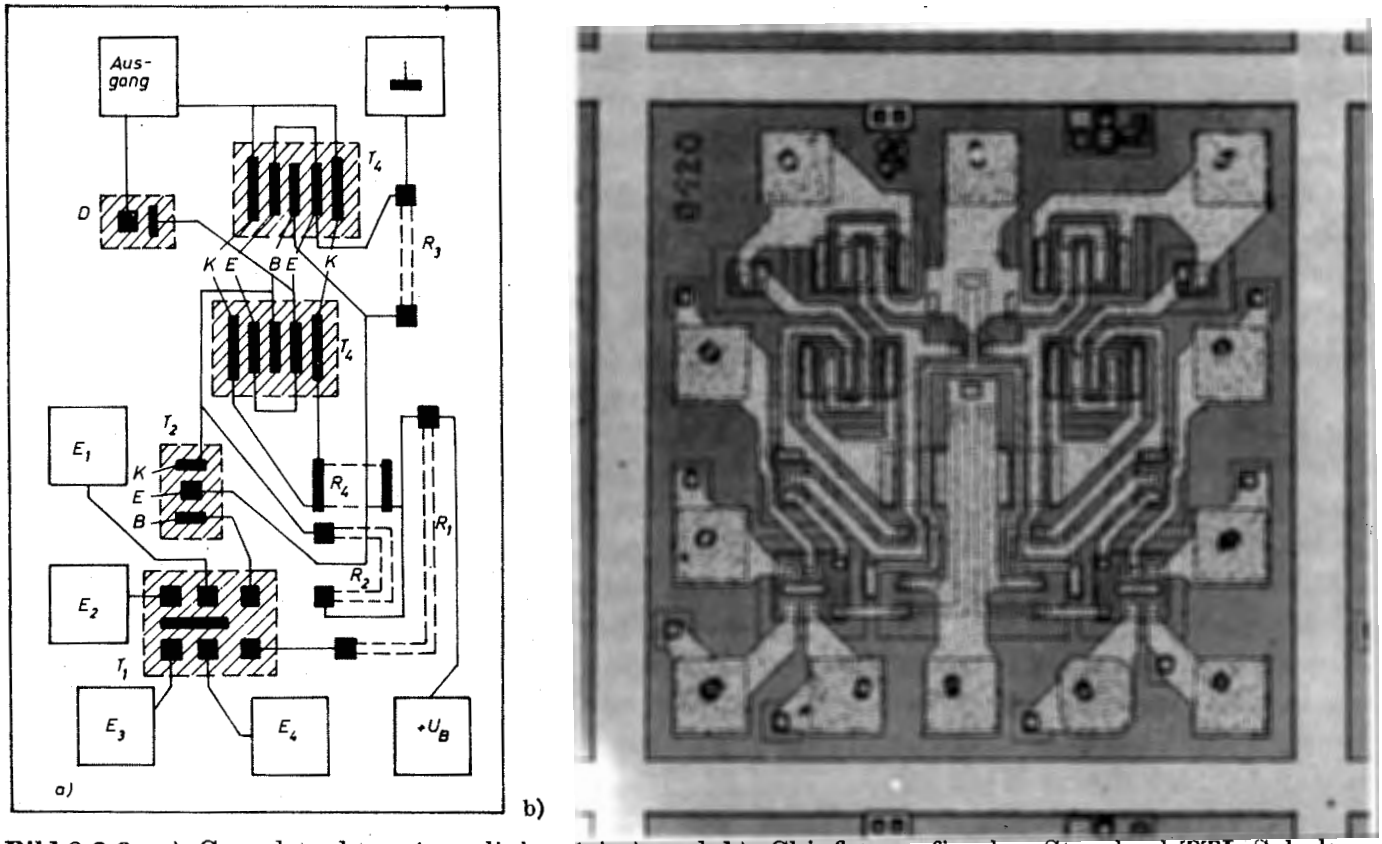


Bild 8.2.9. a) Grundstruktur (nur linke Seite) und b) Chipfotografie des Standard-TTL-Schaltkreises D 120

Liegt am Eingang das Potential L (8.2.8a)), so ist T_2 gesperrt. Dadurch wird T_3 in den leitenden und T_4 in den gesperrten Zustand gebracht. Am Ausgang erscheint das Potential H.

Liegen dagegen alle Eingänge an H, so ist T_2 leitend (Teilbild b)). Daher sinkt das Basispotential von T_3. Er wird gesperrt. Andererseits steigt aber das Basispotential von T_4, und er wird leitend. Am Ausgang entsteht L.

Eine Chipfotografie (b) und das entsprechend zur besseren Übersicht umgezeichnete Lay-out (a) der linken Hälfte eines Schaltkreises D 120 zeigt Bild 8.2.9. Es ist daraus sowohl die Grundstruktur des Multiemittertransistors T_1 von Bild 8.2.7 als auch die Anordnung und Größe der anderen Bauelemente zu erkennen.

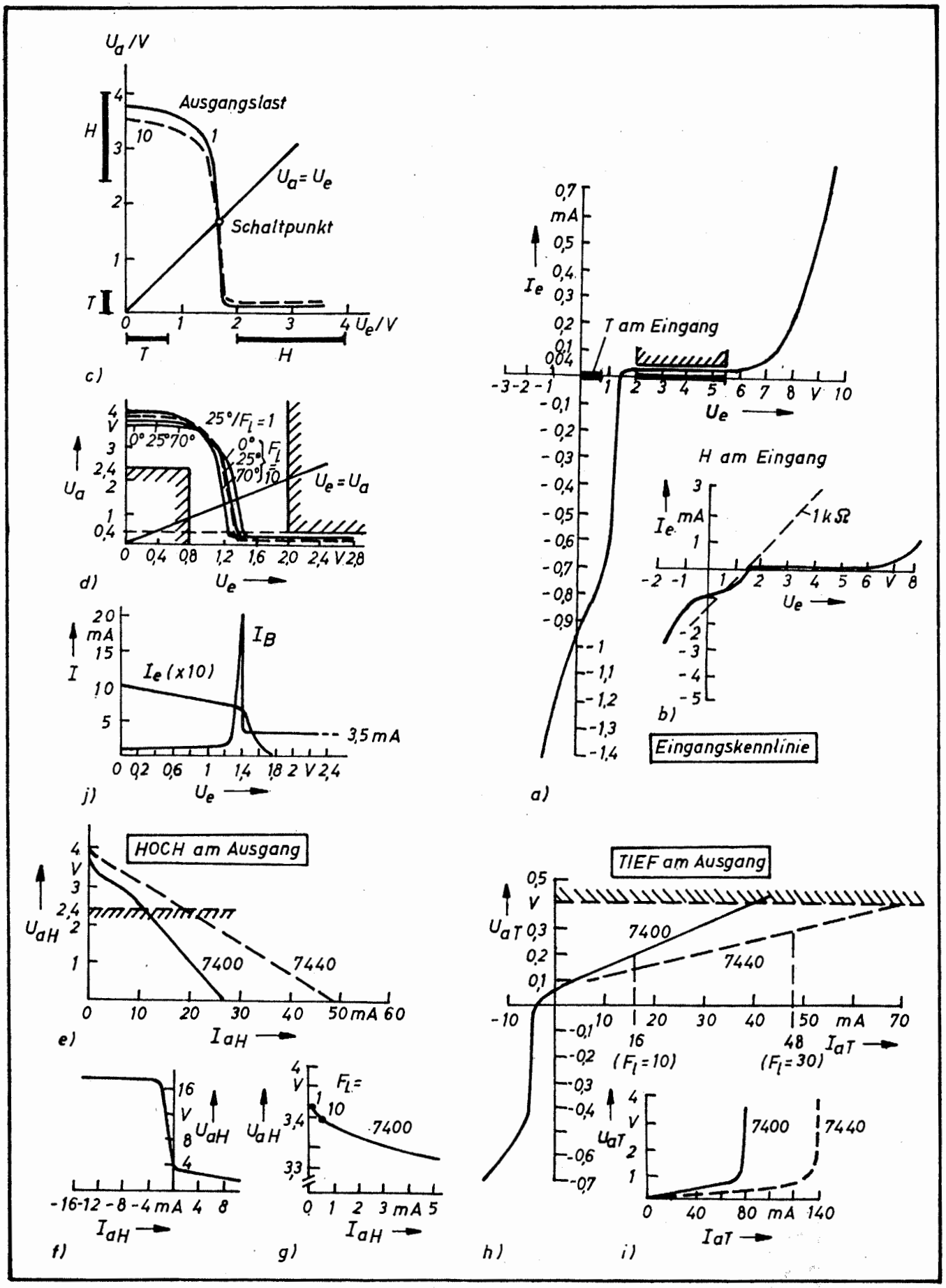
8.2.4.3. Kennwerte der Standard-TTL

Tab. 8.2.14 faßt die wichtigsten Daten und Bild 8.2.10 wichtige Kennlinien zusammen.

Die *Eingangskennlinie* von TTL-Schaltkreisen zeigt Bild 8.2.10a). Solange die Emitterspannung größer als etwa 1,5 V (Potential H) ist, beträgt der Eingangsstrom etwa 20 µA. Garantiert werden im Bereich von 2 bis 5,5 V höchstens 4 µA. Das bedeutet gleichstrommäßig etwa 100 kΩ. Oberhalb von etwa 7 V erfolgt der Z-Durchbruch der Emitterdiode, und es entsteht ein starker Stromanstieg. Die genaue Durchbruchspannung hängt stark von Serie und vom Hersteller ab. Deshalb sind Eingangsspannungen größer +5,5 V und größere Spannungen zwischen den Emittern verboten.

Tabelle 8.2.14. Zusammenstellung der Kennwerte von Standard-TTL-Schaltkreisen

	Grenzwerte für am Ausgang		typische Werte
	L	H	
Eingangsspannung	$\geqq$ 2 V	$\leqq$ 0,8 V	—
Ausgangsspannung	$\leqq$ 0,4 V	$\geqq$ 2,4 V	0,07 ... 0,2 bzw. 3,2 V
Stromaufnahme	$\leqq$ 5,5 mA	$\leqq$ 2 mA	3 bzw. 1 mA
Eingangsstrom	$\leqq$ 40 µA	$\leqq$ −1,6 mA	2 ... 25 µA bzw. 1 mA
Ausgangslastfaktor	—	—	10
Umschaltpunkt	—	—	1,37 ... 1,43 V
Durchbruch Eingangsemitter	—	—	8,3 ... 8,7 V
Durchbruch Ausgangskollektor	—	—	10 ... 18 V
Speisespannung	−0,5 ... +7 V		5 V ± 0,25 V
Verlustleistung je Baustein	400 ··· 450 mW		—
Eingangsgleichspannung	−1,5 ... +5,5 V		—
Ausgangsgleichspannung	−0,8 ... +5,5 V		—
Gleichspannung zwischen zwei Gattern	<5,5 V		—


Ua/V
Ausgangslast
10
1
Ua = Ue
Schaltpunkt
H
T
Ue/V
c)
25°/Fl = 1
0° 25° 70°
0°
25°
70°
Fl
10
Ue = Ua
Ua
d)
Ue
20
mA
I
IB
Ie (x10)
3,5 mA
j)
Ue
HOCH am Ausgang
UaH
7400
7440
e)
IaH
16
V
UaH
-16 -12 -8 -4 mA 4 8
f)
IaH
Fl =
1
10
7400
UaH
0 1 2 3 mA 5
g)
IaH
Ie
mA
T am Eingang
-3 -2 -1 1 2 3 4 5 6 7 8 V 10
Ue
H am Eingang
1 kΩ
-2 -1 1 2 3 4 5 6 V 8
Ue
b)
Eingangskennlinie
a)
TIEF am Ausgang
UaT
7400
7440
-10 10 20 30 40 50 mA 70
16
(Fl = 10)
48
(Fl = 30)
IaT
UaT
7400
7440
0 40 80 mA 140
IaT
h)
i)

Beim Potential Tief (L) fließt etwa ein Strom von 1 mA aus dem Schaltkreis heraus. Im Bereich von 0 bis 0,8 V werden Werte von kleiner als 1,6 mA garantiert. Auch hier erfolgt ein steiler Anstieg des Stromes, jedoch bei Werten kleiner als −1,5 V. Hierbei wird der *pn*-Übergang des Mehremittertransistors zum Substrat leitend.

Für Wechselspannungen zum Schalten beträgt der Eingangswiderstand der TTL-Schaltung, wie Bild 8.2.10b) zeigt, etwa 1 kΩ. Die Wechselstromwiderstände bei H betragen etwa 2 MΩ und beim Eingangspotential L etwa 5 kΩ.

Die *Übertragungskennlinie* kennzeichnet das Schaltverhalten einer Schaltung. Für binäre Schaltungen sollte sie im Idealfall rechteckig verlaufen. Die reale Kennlinie zeigt Bild 8.2.10c). Mittels der Geraden $U_e = U_a$ wird der Schaltpunkt bestimmt. Er liegt bei etwa 1,5 V. Die Zustände H und L sind ebenfalls eingezeichnet. Sie berücksichtigen bereits verschiedene Last- und Temperatureinflüsse. Dabei zeigt sich, daß die Eingangsspannung jeweils um 0,4 V Sicherheit gegenüber den entsprechenden Ausgangswerten hat. Diese Störsicherheit ist statischer Natur und heißt auch Worst-case-Störabstand. So ergeben sich die verbotenen Bereiche im Kennlinienfeld, wie sie Bild 8.2.10d) zeigt. Daraus ist auch der Temperatur- und Lasteinfluß zu erkennen. F_L bedeutet Lastfaktor, also die Anzahl der nachschaltbaren Eingänge.

In Bild 8.2.10e) bis i) sind die Verhältnisse am Ausgang des TTL-Schaltkreises dargestellt, und zwar getrennt voneinander für die beiden *Ausgangszustände*. Teilbild e) zeigt, wie der Ausgangspegel bei H am Ausgang vom entnommenen Laststrom abhängt. Das Gatter ist in diesem Zustand kurzschlußfest. Der Strom wird dabei durch den Widerstand R_4 in Bild 8.2.8 begrenzt. Die Neigungen der Kennlinien entsprechen diesen Werten mit ca. 130 und 75 Ω. Bedeutsam ist weiter, daß von außen auch eine größere Spannung zugeführt werden kann. Wie das Bild 8.2.10f) zeigt, beginnt der Durchbruch des Transistors dann erst bei etwa 16 V. Diese Spannung unterliegt jedoch stark Fabrikationsstreuungen. Zum besseren Erkennen des Lasteinflusses ist der Anfangsteil der Kennlinie vergrößert in Teilbild g) herausgezeichnet.

Die *Ausgangskennlinie* für Tief am Ausgang zeigt Teilbild h). Hier nimmt die Ausgangsspannung mit steigender Belastung zu. Die Steigung der Kennlinie beträgt etwa 10 bzw. 5 Ω. Bei Kurzschluß zur Masse muß in die Schaltung ein Strom von etwa 3 mA fließen. In diesem Zustand ist allerdings ein Kurzschluß zur Betriebsspannung gefährlich. Gemäß Teilbild i) fließen bei Spannungen über etwa 0,5 V relativ große, wenn auch begrenzte Ströme.

Der *Stromverbrauch* eines Gatters ist vom Zustand H oder L abhängig. Dies zeigt Bild 8.2.10j). Im Ausgangszustand H fließen etwa 1 mA, im Ausgangszustand L dagegen 3,5 mA. Weitaus kritischer ist jedoch die Stromspitze in der Umgebung des Schaltpunktes von ca. 20 mA. Sie kann statisch in beiden Richtungen sowohl von H nach L als auch von L nach H gemessen werden. Da dieser starke Strom das Gatter relativ hoch belastet, muß der Schaltpunkt sehr schnell durchlaufen werden. Für die

◀

Bild 8.2.10. Wichtige Kennlinien der Standard-TTL-Technik

a) Eingangskennlinie
b) Eingangskennlinie mit Hervorhebung des effektiven Wechselstromwiderstandes
c) Übergangskennlinien für verschiedene Ausgangslasten
d) Einfluß der Temperatur auf die Übertragungskennlinie
e) Ausgangskennlinien bei H am Ausgang für normale Endstufe (ausgezogen) und Leistungsgatter (gestrichelt)
f) Kennlinie bezüglich zusätzlicher Spannungen am Ausgang im H-Zustand
g) Ausgangsspannung bei H am Ausgang für geringe Belastungen
h) Ausgangskennlinie bei L am Ausgang (bei normalem Gatter ausgezogen, beim Leistungsgatter gestrichelt)
i) Ausgangskennlinien für sehr hohe Belastungen
j) Verlauf der Stromaufnahme eines Gatters und des steuernden Eingangsstromes

Standard-TTL-Gatter wird eine Übergangszeit zwischen beiden Zuständen von höchstens 10 µs zugelassen. Andernfalls treten thermische Instabilitäten auf.

Bei *dynamischem Schalten* wird die *Stromspitze* noch durch die Umladung von Kapazitäten erhöht. Dies bedeutet, daß der Stromverbrauch von TTL-Schaltungen bei Frequenzen über ca. 10 MHz zunimmt.

Infolge des unterschiedlichen Stromverbrauchs und der Stromspitze beim Schalten werden an die *Versorgungsspannung* von TTL-Schaltungen *hohe Anforderungen* gestellt. In der Regel muß jeder einzelne Baustein, um die Ausbreitung von Störimpulsen zu vermeiden, mit einigen nF kurzgeschlossen werden. Ferner muß die Betriebsspannung ziemlich niederohmig sein.

Nicht nur auf den Versorgungsleitungen treten Stromänderungen auf. Auch die Verbindungsleitungen zwischen den Gattern führen Impulse. Deshalb ist in Bild 8.2.10j) der Verlauf des Eingangsstromes eingezeichnet. Insgesamt können beim Betrieb von TTL-Schaltungen auf den Leitungen der Stromversorgung und der logischen Zusammenschaltung *Störspitzen* aufgetreten, die ein einwandfreies Schalten verhindern. Hiergegen sind Sondermaßnahmen zu ergreifen, die u. a. durch Kapazitäten nach Masse, Abschluß mit Wellenwiderständen und Verdrillen von Leitungen realisiert werden [K 30].

8.2.4.4. Weitere Ausgangsstufen

Bei der Standard-TTL-Technik werden normale Ausgangsstufen und *Leistungsendstufen* unterschieden. Beide besitzen vor allem unterschiedliche Stromaufnahme, Stromabgabe und einen anderen Leistungsverbrauch. Sie ermöglichen dadurch aber auch ein unterschiedliches Fan-out. Diese beiden Fälle sind bereits in den Ausgangskennlinien der Bilder 8.2.10e), h) und i) enthalten. Darüber hinaus gibt es unterschiedliche Gatter mit z. T. nur wenig veränderten Eigenschaften. Sie sind in Bild 8.2.11 zusammengestellt.

Die *Diode D* im Bild 8.2.8 sorgt dafür, daß bei niedrigem Ausgangspotential L (Teilbild b)) der Transistor T_3 mit Sicherheit gesperrt bleibt. Sie hebt sein Emitterpotential um etwa 0,7 V an. Hierzu gibt es auch andere Schaltungsvarianten, die vor allem bei sogenannten Leistungsausgangsstufen angewendet werden. Sie zeigt Bild 8.2.11b). Es ist der zusätzliche Transistor T_5 eingefügt. Seine Schaltwelle liegt um die Basis-Emitter-Spannung (etwa 0,7 V) höher. Außerdem hat die DARLINGTON-Schaltung aus T_5 und T_3 eine höhere Verstärku g. Ob R_5 oder R_6 in dieser Schaltung verwendet wird, ist nicht so entscheidend. Es wird aber immer nur einer verwendet.

Für Sonderanwendungen werden die Anschlüsse zur Endstufe E als gemäß Bild 8.2.11a) herausgeführt. Dann können dort spezielle Eingangsstufen aus T_1 und T_2, die nur R_1 enthalten, parallelgeschaltet werden. Derartige Schaltungen heißen Erweiterungsschaltungen.

In Bild 8.2.11c) ist die Endstufenvariante mit ausschließlich dem Transistor T_4 gezeigt. Sie dient zur Realisierung eines *verdrahteten ODER* über einen außen angelegten gemeinsamen Widerstand R_C (wired OR).

Eine einfache Parallelschaltung der Gatter mit Endstufe ist nämlich nicht möglich. Wenn ein Gatter H und ein anderes L eingenommen hätte, würde über den leitenden oberen Transistor des einen und den leitenden unteren Transistor des anderen ein so starker Strom fließen, daß keine Betriebssicherheit mehr bestände. Mit der vereinfachten Schaltung ist dies möglich. Der Widerstand R_C muß nach der Anzahl der Gatter gewählt werden. Sein Wert bestimmt den gemeinsamen Innenwiderstand aller Gatter bei hohem Ausgangspotential. Er ist meist erheblich höher als bei einer üblichen Ausgangsstufe. Bei tiefem Ausgangspotential ist dagegen mindestens ein T_4 mit kleinem Innenwiderstand leitend.

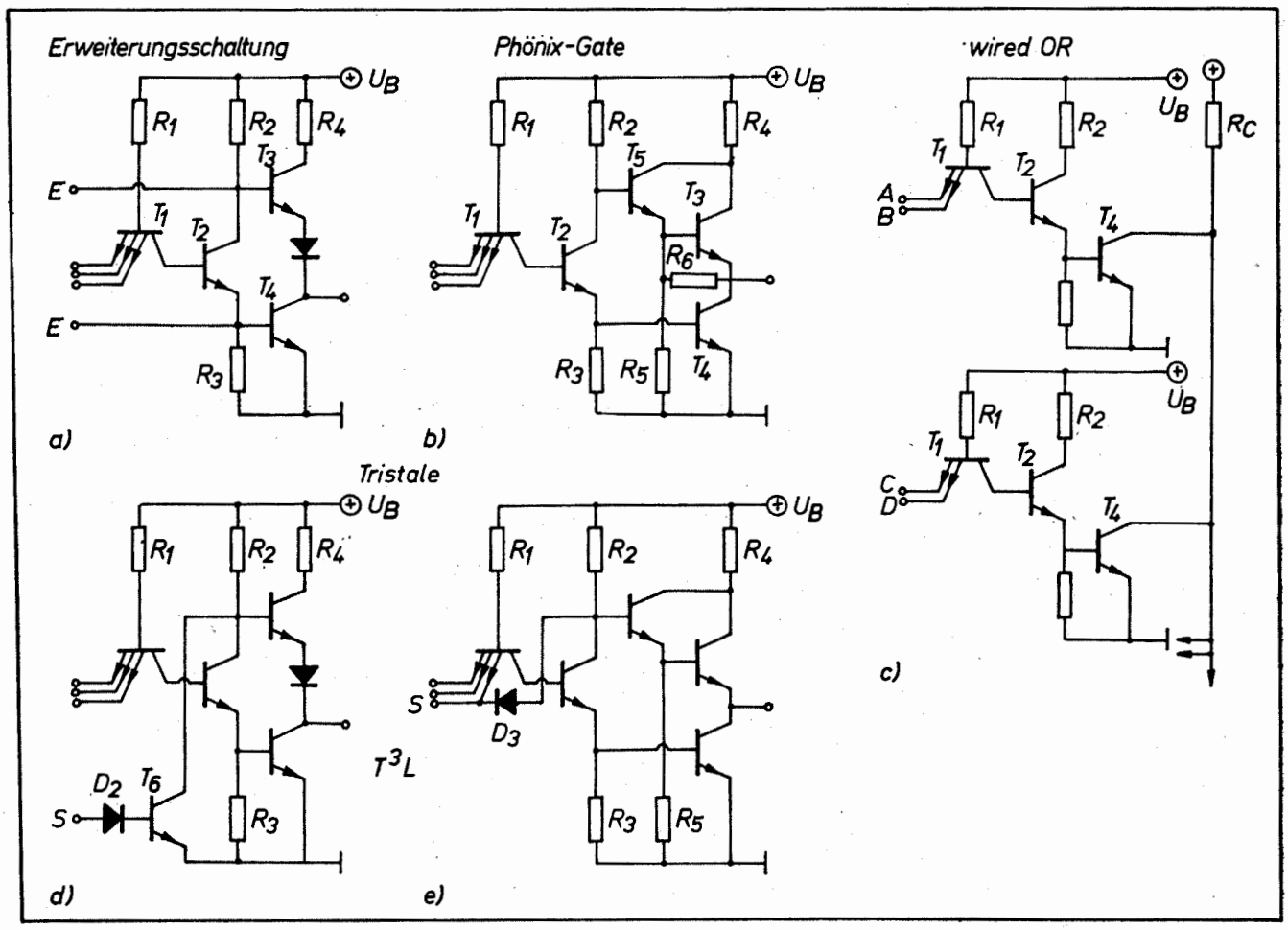


Bild 8.2.11. Unterschiedliche Ausgangsstufen bei TTL-Schaltungen

a) Erweiterungsschaltung mit Herausführung der Anschlüsse für die Ausgangsstufe

b) Sogenanntes Phönix-Gatter (Leistungsgatter), weil es erstmalig 1962 in der Phönix-Rakete verwendet wurde. Es existiert nur entweder R_5 oder R_6. R_5 ist üblicher

c) Gatter mit der Eigenschaft, daß die Ausgänge parallel geschaltet werden können

d) und e) Gatter mit drei Ausgangszuständen (T³L)

Die Parallelschaltung von TTL-Ausgängen kann auch realisiert werden, wenn es durch eine zusätzliche Steuerung gelingt, beide Ausgangstransistoren T_3 und T_4 gleichzeitig in den gesperrten Zustand zu bringen. Dann entstehen Schaltungen, wie z. B. Bild 8.2.11d) und e) zeigen.

Wird bei der Schaltung nach Bild 8.2.11d) dem Eingang S niedriges Potential L zugeführt, so ist der Transistor T_6 gesperrt, und die Schaltung arbeitet wie üblich. Erhält aber T_6 hohes Potential H, so wird er leitend. Dadurch werden beide Ausgangstransistoren in den gesperrten Zustand versetzt. Die Diode D_2 dient lediglich dazu, den Pegel für S auf den üblichen TTL-Pegel zu bringen.

Bei einem Gatter gemäß 8.2.11b) reicht sogar nur die Diode D_3 (8.2.11e)) aus, um die beiden Ausgangstransistoren in den gesperrten Zustand zu bringen. In diesem Fall muß dazu der Eingang S aber tiefes Potential besitzen.

Schaltungen gemäß Bild 8.2.11d) und e) besitzen also *drei Zustände am Ausgang*:

- Sie sind auf H geschaltet: dann leitet der obere Endstufentransistor.
- Sie sind auf L geschaltet: dann leitet der untere Endstufentransistor.
- Sie sind inaktiv: beide Endstufentransistoren sind hochohmig.

Im dritten, inaktiven Zustand können sie folglich ohne wesentliche Stromaufnahme beliebig auf L oder H von dem Ausgang anderer Gatter geschaltet werden. Infolge dieses zusätzlichen Zustandes heißen entsprechende Schaltungen Tristate-TTL oder T³L.

8.2.4.5. Teilschaltungen zur Verbesserung

Drei *Veränderungen* in der Schaltungstechnik haben die TTL-Technik *wesentlich vorangebracht*:

- Eingangsdioden zur Begrenzung negativer Impulse
- der SCHOTTKY-Transistor
- der aktive Basisableitwiderstand.

Diese Prinzipien faßt Bild 8.2.12 zusammen. Im Teilbild a) ist die Anschaltung der *Dioden* zu sehen (vgl. Bild 8.2.8). Sie bewirken, daß keine negativen Eingangsspannungen an dem Multiemitter auftreten können, die größer als etwa −0,7 V sind. Dies ist die Schwellenspannung der Dioden. Solche Dioden sind vor allem bei hohen Frequenzen wichtig, wo die Impulse einmal Resonanzen anstoßen können oder wo auf den Leitungen Reflexionen ähnliches bewirken können. Sie verkürzen solche Vorgänge und tragen damit zu kurzen Schaltzeiten bei.

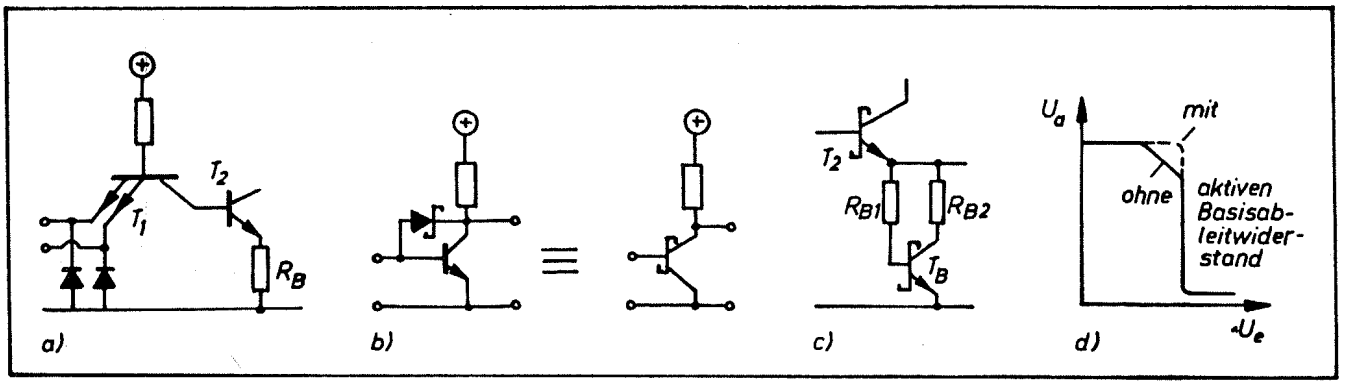


Bild 8.2.12. Verbesserungen an TTL-Schaltungen

a) Eingangsdioden zur Unterdrückung der negativen Impulsspitzen

b) Schaltbild und Schaltzeichen des SCHOTTKY-Transistors. Sein Lay-out enthält Bild 5.6.7

c) Ersatz des ohmschen Basisableitwiderstandes R_B in Teilbild a) durch einen dynamischen aus T_B, R_{B1} und R_{B2}

d) Wirkung des dynamischen Basisableitwiderstandes auf die Übertragungskennlinie (vgl. Bild 8.2.10c) und d))

Die Schaltzeit von Transistoren hängt u. a. davon ab, wieviel Träger erzeugt bzw. ausgeräumt werden müssen. Die entsprechenden Zeiten sind dann besonders lang, wenn der Transistor voll in die Sättigung gesteuert wird. Eine günstige Alternative hierzu ist, diesen Effekt bereits früh zu begrenzen. Dies kann dadurch erfolgen, daß eine SCHOTTKY-*Diode* zwischen Basis und Emitter geschaltet wird (Teilbild b)). Die technisch einfache Realisierung (Schalter) eines solchen Transistors mit Diode — eben eines SCHOTTKY-*Transistors* — zeigte bereits Bild 5.6.7. Vereinfacht wird ein SCHOTTKY-Transistor im rechten Teil von Bild 8.2.12b) dargestellt. Der Widerstand R_B in Bild 8.2.12a) ruft bei der Übertragungskennlinie einen unangenehmen Knick hervor (Bild 8.2.12d) sowie 8.2.10c) und d)). Er läßt sich völlig vermeiden, wenn er durch einen *dynamischen Wechselstromwiderstand* ersetzt wird. Dies zeigt Bild 8.2.12c) als Teilschaltung mit dem Transistor T_B und den beiden Widerständen R_{B1} und R_{B2}. Die Wirkung geht aus Bild 8.2.12d) deutlich hervor.

8.2.4.6. Schaltzeit-Leistungsprodukt; TTL-Familien

Theoretisch sind die kombinatorischen Schaltungen unabhängig von der Zeit. Sobald die Eingangssignale anliegen, sollte am Ausgang das richtige Signal bereitstehen. Dies ist schon rein physikalisch nicht möglich, da Kapazitäten umzuladen sind. In der analogen Technik (vgl. u. a. 5.5.4. und 7.5.3.7.) erfolgt so eine Begrenzung gemäß der

Güte und/oder *Slew-Rate*. In der Digitaltechnik hat eine ähnlich universelle Bedeutung das Produkt aus der Schaltzeit und dem Leistungsaufwand in pJ. Es heißt u. a. Geschwindigkeit-Leistungsprodukt oder Schaltzeit-Leistungsprodukt. Für die bisher behandelten Standard-TTL-Schaltungen beträgt die Laufzeit (Eingang — Ausgang) ca. 10 ns (Lastkapazität 15 pF), und die Leistungsaufnahme eines Gatters liegt bei 10 mW. Folglich beträgt das Geschwindigkeit-Leistungsprodukt 100 pJ. Die *Weiterentwicklung* der TTL-Schaltungen hatte nun *drei Tendenzen*:

- Verkürzung der Schaltzeit für schnelle Vorgänge (eventuell bei Erhöhung der Leistung),
- Minderung der Leistung für stromarme Schaltungen (eventuell bei verlängerten Schaltzeiten),
- Senkung des Geschwindigkeit-Leistungsproduktes für generell bessere Lösungen.

Fordern die ersten beiden Punkte spezielle Auslegungen, vor allem der Endstufen, so sind bei der dritten Forderung neue Technologien notwendig. Aus diesen Forderungen entstanden einmal spezielle TTL-Varianten gemäß Punkt 1 und 2, zum anderen in zeitlich großem Abstand unterschiedliche *Generationen* von TTL-Schaltungen. Ganz wesentlich ist, daß TTL-Schaltungen kombiniert mit allen diesen Varianten aufgebaut werden können. Ja, darüber hinaus konnte z. T. sogar erreicht werden, daß sie einfach direkt gegeneinander ausgetauscht werden können. Die einzelnen Auslegungen bzw. Generationen von TTL-Schaltungen haben also deutlich exakt aufeinander abgestimmte Daten. Sie sind z. T. sogar bezüglich der Anschlußbelegungen austauschbar (*pin-kompatibel*). Diese Entwicklung trug zusätzlich zur großen Verbreitung der TTL-Technik bei. Insgesamt existieren heute drei Generationen von TTL-Schaltungen und mindestens acht TTL-Familien. Einige Grundschaltungen für TTL-Gatter faßt Bild 8.2.13 zusammen. Es enthält auch die entsprechenden Punkte

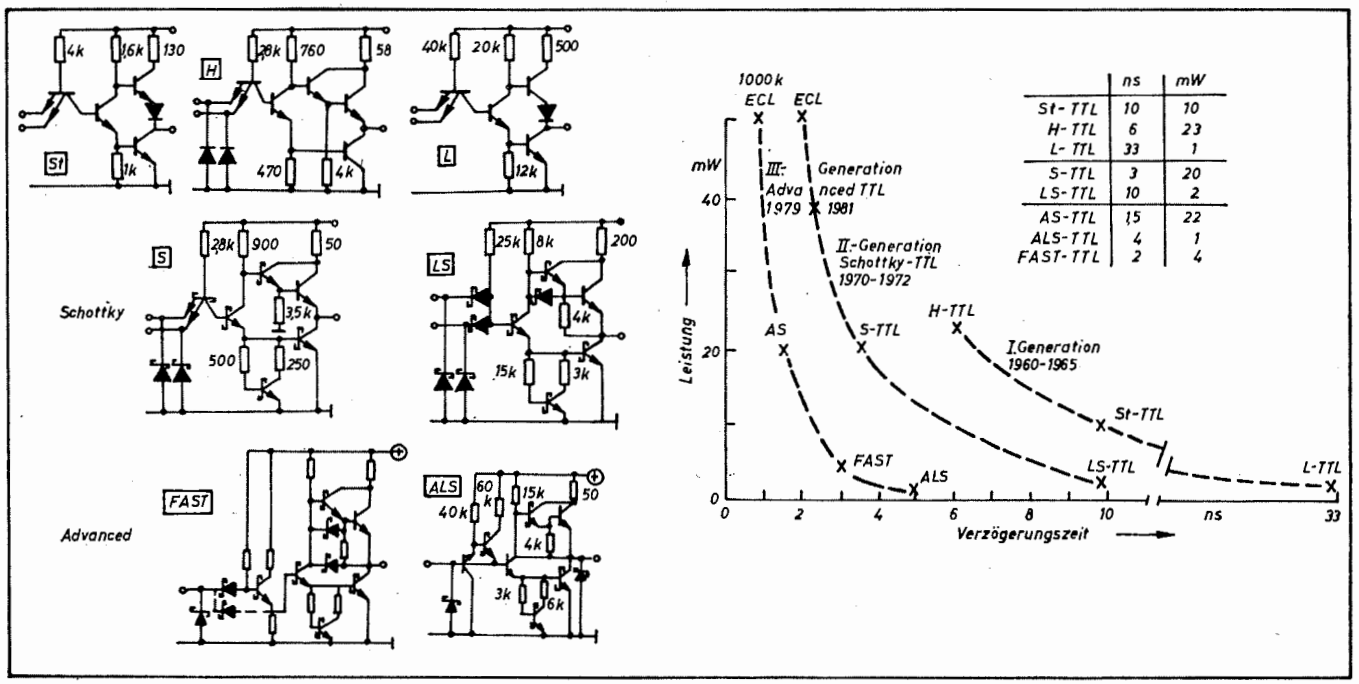


Bild 8.2.13. Zu den TTL-Familien: Links sind die Grundschaltungen wichtiger Familien und rechts ihre Einordnung in das Geschwindigkeit-Leistungsdiagramm dargestellt. Hierbei heben sich deutlich die drei Generationen heraus.

I: St = Standard; H = High-speed; L = Low-power
II: SCHOTTKY-TTL: S = SCHOTTKY; LS = Low-power-SCHOTTKY
III: Advanced: FAST = FAIRCHILD-Advanced-SCHOTTKY-TTL; AS = Advanced SCHOTTKY; ALS = Advanced Low-Power-SCHOTTKY
(Die beiden letzten von Texas Instruments)

im Leistungs-Schaltzeit-Diagramm. Aus dieser Darstellung wird deutlich der Fortschritt erkennbar, der mit der jeweils neuen Generation erreicht werden konnte. Hierbei wirkten sich neben den schaltungstechnischen Verbesserungen vor allem die erheblichen Fortschritte in der Technologie der Mikroelektronik aus. Hierzu zählen u. a. Verkleinerung der Abmessungen, Implantation statt Dotierung und bessere Isolationstechniken.

Die Anzahl der TTL-Schaltkreise für verschiedene Funktionen ist so groß, daß nur noch mit speziellen Tabellenbüchern ein leidlicher Überblick zu halten ist, z. B. [H 28, K 30]. In der Standardausführung existieren weit über 200 Typen. Die nächsthöhere Typenvielfalt liegt bei den Low-power-SCHOTTKY mit ca. 150. Dabei wurden z. T. Typen entwickelt, die nicht in der ersten Generation existieren. Die LS-TTL werden immer mehr zur Standardfamilie. Bei der dritten Generation wurden zunächst die wichtigsten Grundtypen wiederholt. Es liegen bereits (1983) gegen 100 Typen vor.

8.2.5. ECL-Technik

Etwa 1962 entstand die *Emitter coupled logic*, ECL. Sie wird zuweilen auch CML = *current mode logic* genannt. Ihr Prinzip beruht auf einem abgewandelten Differenzverstärker (Bild 8.2.14a)), der jedoch nicht analog, sondern nur in den beiden Grenzen stromlos — stromführend betrieben wird. Da hierbei infolge der *Stromquelle* keiner der Transistoren in die Sättigung gerät, wird zuweilen auch von *ungesättigter Stromschaltung* gesprochen. Dies hat den Vorteil, daß extrem kurze Schaltzeiten bis unter 1 ns zu realisieren sind. Deshalb hat die ECL-Technik bei schnellen Rechnern, in der PCM-Technik und verwandten Gebieten große Bedeutung. 1980 nahm die ECL-Technik 3% als Marktanteil der integrierten Schaltkreise ein.

Der eine Eingang des Differenzverstärkers liegt stets an einer Referenzspannungsquelle, so daß nur der zweite Eingang des Differenzverstärkers für ein logisches Signal zur Verfügung steht. Ihm können aber gemäß Bild 8.2.14b) weitere Transistoren parallelgeschaltet werden, die dann eine OR-Verknüpfung bewirken. Da am Ausgang des Differenzverstärkers stets das normale und das negierte Signal auftreten, realisiert eine ECL-Schaltung (bei positiver Logik) immer zugleich NOR und OR. Unter anderem durch diese Dopplung benötigt eine ECL-Schaltung relativ viel Leistung.

In Bild 8.2.14b) sind weiter die Endstufen als Emitterfolger und die Erzeugung der Referenzspannung gezeigt. Auffällig gegenüber den meisten logischen Schaltungen ist u. a., daß die Betriebsspannung $-5{,}2$ V, also negativ, ist. Aus der Kennlinie gemäß Bild 8.2.14c) gehen folgende typischen Pegel hervor:

$$-0{,}85\ \text{V} \mathrel{\hat{=}} \text{H} \quad \text{und} \quad -1{,}70 \mathrel{\hat{=}} \text{L}\,.$$

Der Hub beträgt also nur 0,85 V. Da der Umschaltbereich der Kennlinie nur ca. 0,1 V beträgt, ist ein ausreichender Störabstand gewährleistet.

Bei der ECL-Technik lassen sich komplizierte *logische Verknüpfungen auf vielfältige Weise* erreichen. Eine Möglichkeit besteht darin, Ausgänge vor verschiedenen Gattern einfach entsprechend wired-OR zu verbinden (vgl. Bild 8.2.11 bez. TTL). Eine andere besteht in *Mehrebenen-Raumstrukturen* nach dem Prinzip von Bild 8.2.14d). Das konkret ausgeführte Beispiel für ein XOR (Exklusiv-OR: vgl. Tab. 8.2.3 und 8.2.7) zeigt Bild 8.2.14e). Hier wird deutlich, wie z. B. mittels Transistor T_A der Pegel auf das Niveau der unteren Ebene angepaßt wird. Ebenso sind mehrere Referenzspannungen notwendig.

Genauso wie die TTL-Technik wurden auch für die ECL-Technik verschiedene Varianten und Weiterentwicklungen möglich.

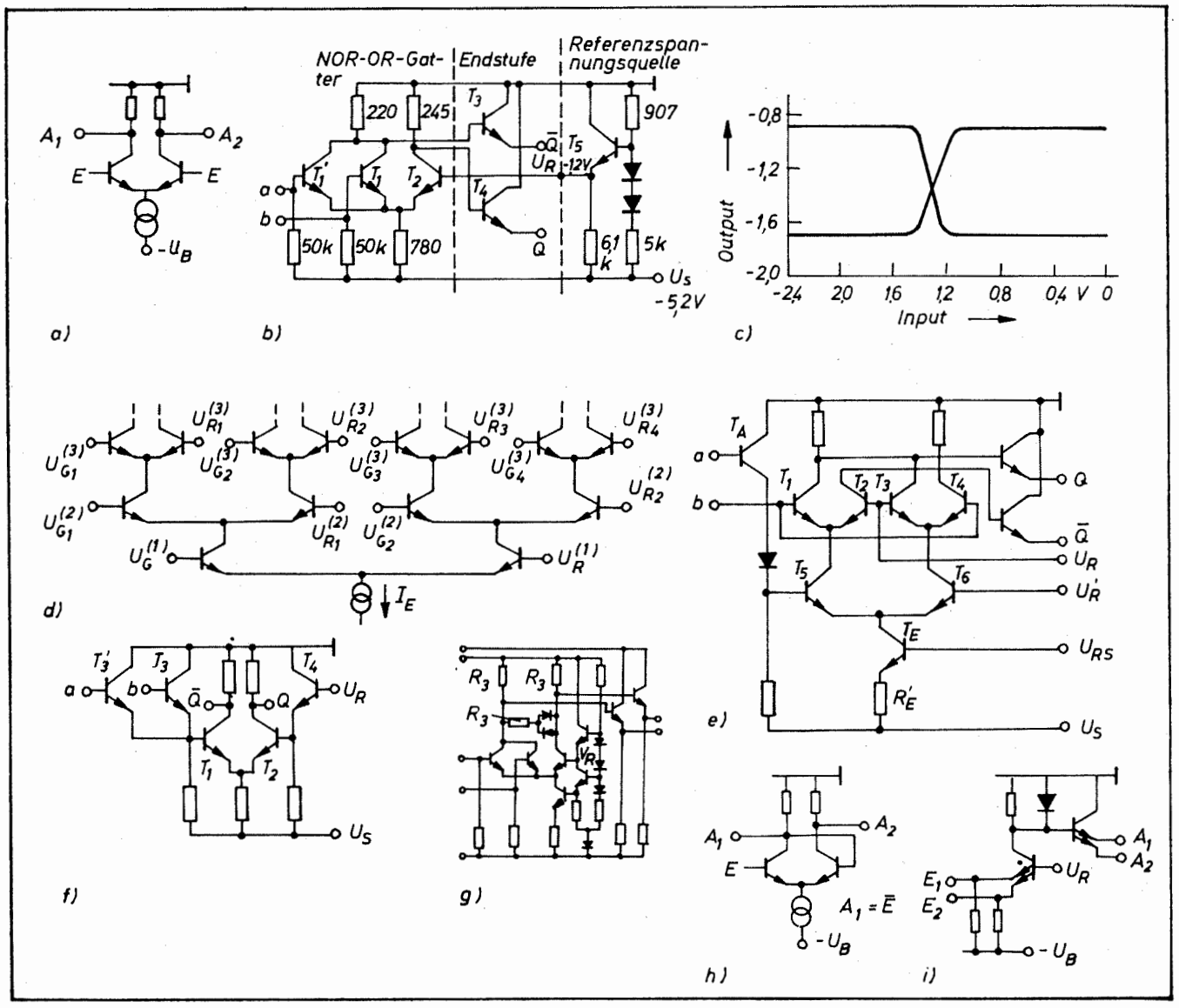


Bild 8.2.14. Zur ECL-Technik

a) Grundprinzip aus Stromquelle und Differenzverstärker
b) Beispiel einer ECL-OR-NOR-Schaltung mit Referenzquelle und Endstufen. Die Werte entsprechen dem konkreten Typ MECL 10000 (Motorola)
c) Typische Übertragungskennlinie eines ECL-Gatters
d) Prinzip der Mehrebenen-Baumstruktur
e) Beispiel einer ECL-Zweiebenen-Schaltung für XOR und $\overline{\text{XOR}}$
f) Prinzipschaltung des E²CL-OR-NOR-Gatters
g) Beispiel einer temperaturkompensierten ECL-Schaltung (FAIRCHILD)
h) Prinzip der rückgekoppelten ECL-Technik
i) Prinzip der EFL-Technik

Etwa parallel zur ECL-Technik besteht die E²CL (emittergekoppelte ECL). Hierbei ist der Differenzverstärker selbst die Endstufe, und die Emitterfolger (sonst Endstufen) sind vorgeschaltet. So ergibt sich das Bild 8.2.14f). Obwohl ECL und E²CL in der Leistungsfähigkeit etwa gleichwertig sind, hat sich international die ECL-Technik durchgesetzt.

Im allgemeinen sind die Kennlinien der ECL-Schaltungen relativ stark durch die *Temperatur* beeinflußt. So entstand eine voll temperaturkompensierte ECL, deren Prinzipschaltung Bild 8.2.14g) zeigt. Der Aufwand ist aber erheblich höher.

Eine andere Variante ist die *rückgekoppelte* ECL (RECL). Hierbei wird der zweite Ausgang des Differenzverstärkers nicht mehr an eine Referenzspannung gelegt, sondern an den Kollektor des anderen Transistors (Bild 8.2.14h)).

Im Vergleich einiger Daten von 1962 und 1969 sei der schnelle Fortschritt der ECL-Technik gezeigt. Die Verzögerungszeit verkürzte sich von ca. 8 auf 0,9 ns, die Leistung fiel von 37 auf 13 mW, die Fläche sank von 50×80 auf $15 \times 28\,\mu m^2$ je Emitter. Heute werden im Labor Zeiten bis etwa 0,3 ns bei 2,2 mW/Gate erreicht.

Als innere Logik, also logische Verarbeitung auf dem Chip, ist aus der ECL-Technik die EFL-Technik (*Emitter-Funktions-Logik*) entstanden. Ihr Grundprinzip zeigt Bild 8.2.14i). Das Prinzip der Stromverteilung mit dem Differenzverstärker wird erst dann sichtbar, wenn der Eingang eines Gatters mit dem Ausgang eines anderen verbunden wird.

8.2.6. I^2L-Technik

Die I^2L-Technik *(integrierte Injektions-Logik)* ist relativ spät entstanden. Die entscheidende Idee geht unabhängig voneinander auf HART und SLOB, Philips (Niederlande), sowie auf T. WIEDEMANN und H. BERGER (IBM-Boeblingen — BRD) zurück.
Die erste technologische Reife wurde bei Philips (Niederlande) von N. C. DE TROY 1971/72 erreicht. Die I^2L-Technik hat als interne Logik heute einen hohen Stand und breite Anwendung erreicht. Die I^2L-Technik wird zuweilen auch mit MTL (merged transistor logic) bezeichnet.

8.2.6.1. Das Wirkprinzip

Bei der I^2L-Technik werden im Gegensatz zu den anderen Schaltungen im Halbleitersubstrat vorhandene bzw. bewußt erzeugte (injizierte) Ladungsträger verwendet. Deshalb sind *I^2L-Schaltungen nicht in diskreter Technik realisierbar.* Zur Veranschaulichung sei aber von einer diskreten Struktur gemäß Bild 8.2.15a) ausgegangen. Die Stromquelle I_0 umschreibt den später zu behandelnden Injektionsstrom. Ihr Strom kann entweder in die Basis von T_2 oder in den Kollektor von T_1 fließen. Zwischen den beiden extremen Fällen ist natürlich eine Stromverteilung möglich. Sie hat aber für die weiteren Betrachtungen nur untergeordnete Bedeutung.

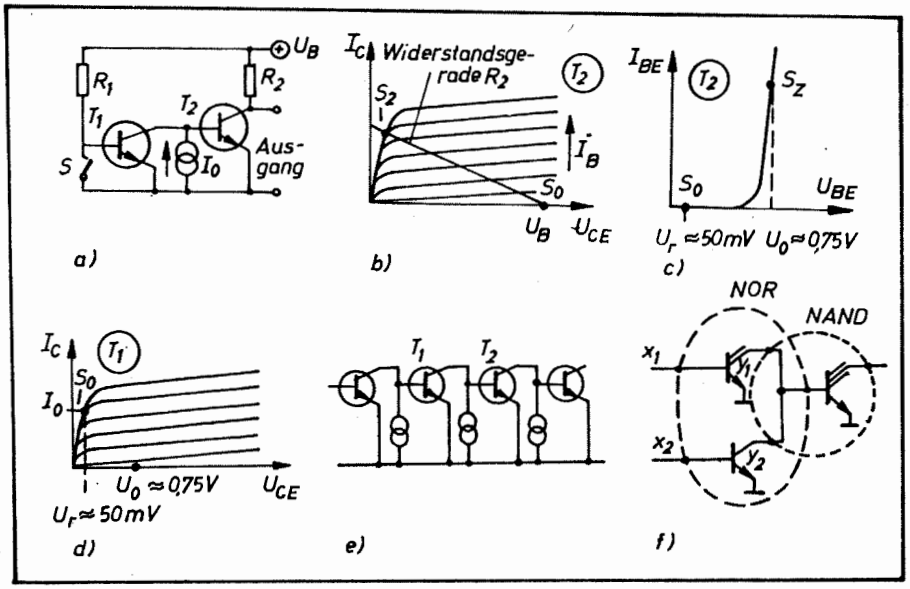


Bild 8.2.15. Zur I^2L-Technik

a) Diskrete Struktur zur Veranschaulichung
b) Ausgangskennlinie von T_2
c) Eingangskennlinie von T_2
d) Ausgangskennlinie von T_1
e) Reihenschaltung von I^2L-Invertern
f) Logische Funktionen bei der I^2L-Technik

Zunächst werde der Schalter S geschlossen angenommen (Zustand S_z, von zu), dann ist der Transistor T_1 gesperrt, d. h., es kann kein Kollektorstrom fließen. Deshalb fließt der gesamte Strom der Stromquelle I_0 in die Basis von T_2 und macht dessen Kollektor leitend. Am Ausgang entsteht also ein niedriges Potential. Dies zeigt im Ausgangskennlinienfeld Bild 8.2.15b). Für das Eingangskennlinienfeld von T_2 gilt Teilbild c). Es stellt sich also entsprechend der Kniespannung der Basis-Emitter-Diode eine Spannung von etwa 0,76 V ein.

Nun werde der Schalter geöffnet (Zustand S_0, von offen). Dadurch fließt in die Basis von T_1 ein Strom über R_1. Er macht die Kollektorstrecke von T_1 niederohmig, und zwar niederohmiger als die Basis-Emitter-Strecke von T_2. Das bedeutet, daß praktisch der gesamte Strom von I_0 in den Kollektor von T_1 fließt. Dabei entsteht ein Arbeitspunkt auf der Grenzkennlinie von T_2 (Teilbild d)). Der Spannungsabfall erreicht eine Restspannung von etwa 50 mV. Bei so niedrigen Spannungen ist die Basis-Emitter-Diode von T_2 (Teilbild c)) bereits sehr hochohmig. Damit wird T_2 gesperrt, und die Ausgangsspannung der Schaltung nimmt entsprechend Teilbild b) einen hohen Pegel an.

Aus den bisherigen Betrachtungen ist offensichtlich, daß mittels einer Stromquelle zwischen zwei Transistoren logische Signale ohne Spannungsquelle weitergeleitet und negiert werden können. An sich ist dies fast selbstverständlich, denn eine Stromquelle kann durch eine hinreichend große Spannungsquelle mit einem entsprechend großen Innenwiderstand ersetzt werden. Dies führt hier aber gedanklich nicht weiter, denn im Halbleiterkristall entsprechen der Stromquelle ja injizierte Ladungsträger. Damit können die beiden Widerstände R_1 und R_2 mit der Spannungsquelle durch zwei Stromquellen entsprechend Teilbild e) ersetzt werden. An der Funktion der Schaltung ändert sich dadurch nichts. Es kann sogar eine unendliche Kettenschaltung angenommen werden. Für jeden Transistor gelten dann die Kennlinien von Teilbild c) und d) für den Ein- bzw. Ausgang mit den beiden möglichen Zuständen, die ganz zu Beginn der Schaltung eingeschaltet werden. Aus den Betrachtungen folgt weiter (vgl. z.B. Teilbild b) und d)), daß die Zustände S_z und S_0 sich von Transistor zu Transistor vertauschen. Es werden also nicht nur Zustände weitergeleitet, sondern auch negiert. Ein Transistor mit einer Stromquelle stellt also in der Kette genau einen Negator dar.

8.2.6.2. Logische Verknüpfungen

Logische Verknüpfungen sind bei der I^2L-Technik relativ einfach möglich, indem Kollektoren einfach verbunden werden. Gemäß Bild 8.2.15f) ergeben sich daraus die Möglichkeiten des NOR und NAND. Wenn zwei Kollektoren auf eine gemeinsame Basis des Folgetransistors einwirken, so genügt es, einen Kollektor leitend zu machen, damit der Folgetransistor gesperrt ist. Bei positiver Logik bedeutet dies für die Größen x_1 und x_2, daß nur eine der beiden oder beide den Pegel H zu besitzen brauchen. Dies ist also das NOR in Bild 8.2.15f). Wird dagegen der Eingang des Folgetransistors betrachtet, so müssen beide vorangehenden Kollektoren für y_1 und y_2 die Werte H besitzen, damit der Ausgang L ist. Dies ist ein NAND. Die I^2L-Technik benötigt folglich nur Vielkollektortransistoren, da Kollektoren und Basisanschlüsse in definierter Weise für die zu realisierende Funktion zu verbinden sind. Damit ergibt sich insgesamt eine sehr einfache Struktur.

8.2.6.3. Injektionsquellen

Die bisherigen Betrachtungen ließen sich, wenn von den Mehrfachkollektortransistoren abgesehen wird, noch diskret verstehen und realisieren. Anders wird es, wenn die

Stromquelle als Injektionsquelle für Ladungsträger gestaltet werden soll. Dies kann einmal durch Licht, Trägerpaarbildung mittels Quanten und zum anderen mittels eines in Durchlaßrichtung betriebenen *pn*-Übergangs erfolgen. Die Ladungsträgererzeugung durch Licht könnte vielleicht einmal im Weltraum Bedeutung besitzen. Vielleicht war dies sogar geschichtlich der Ausgangspunkt aller Betrachtungen. Zumindest lassen die ersten Publikationen darauf schließen.

Es sei von einer Struktur ausgegangen, wie sie Bild 8.2.16a) zeigt. Hierbei wird bereits wie generell bei I^2L jeder Transistor im Gegensatz zu seiner sonstigen Anordnung invers betrieben. Es handelt sich also eigentlich um Multiemittertransistoren mit ihren schon von TTL-Strukturen bekannten guten Realisierungsmöglichkeiten.

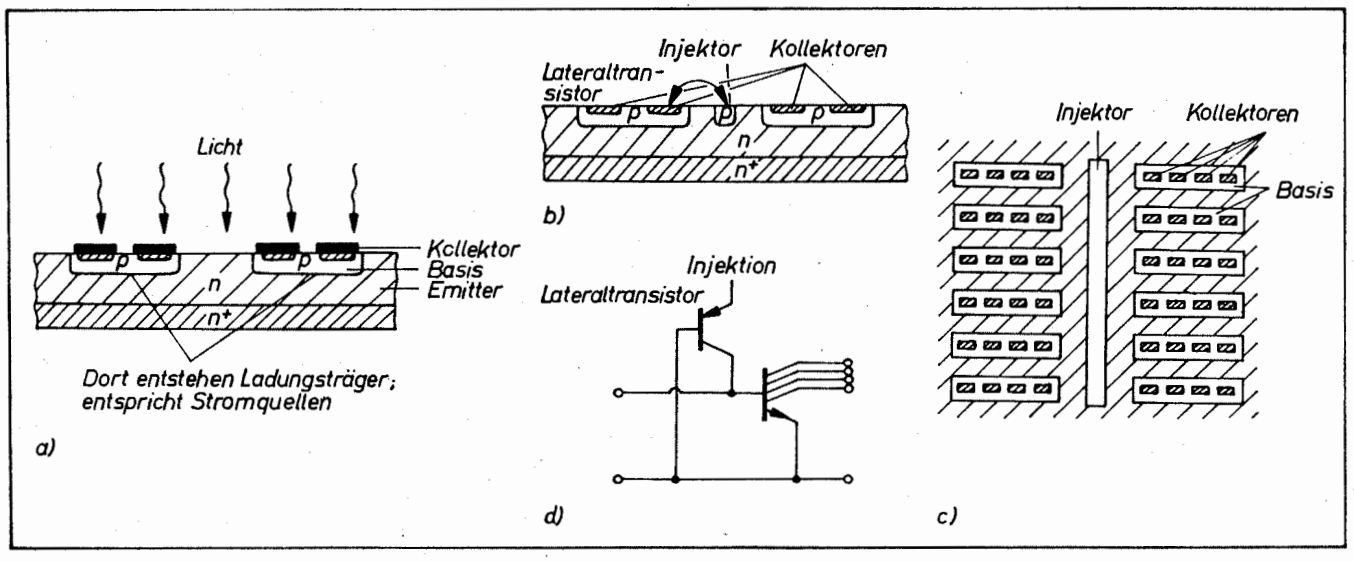


Bild 8.2.16. Zur Strukturierung der I^2L-Technik

a) Multikollektortransistor und Licht als Injektor

b) Einschaltung eines schmalen *p*-Gebietes, das als in Durchlaßrichtung betriebener *pn*-Übergang Ladungsträger erzeugt, und der durch den Injektor entstandene Lateraltransistor

c) Anordnung vieler Multikollektortransistoren rechts und links eines Injektorstreifens

d) Übliche Ersatzschaltung für die I^2L-Technik

Fällt Licht auf eine solche Struktur, so können sich Ladungsträger an allen beleuchteten Stellen, vor allem im *pn*-Übergang bilden. Infolge der undurchsichtigen Kontakte über den Kollektoren kommt nur die Grenze Basis—Emitter in Betracht. Dies ist aber nach Bild 8.2.15 genau erwünscht. Im Vergleich dazu ist auch sofort die Struktur einer solchen Schaltung klar. Für jeden Transistor ist eine *p*-Struktur zu erzeugen und in ihr die entsprechende Anzahl von Kollektoren. Die Emitterschicht ist allen Transistoren gemeinsam und braucht dabei nur als Masse aufgefaßt zu werden. Eine solche Einfachheit dürfte kaum mehr zu überbieten sein. Ihr Nachteil ist lediglich, daß sie Licht benötigt.

Für den Betrieb unter irdischen Bedingungen müssen die Ladungsträger anders erzeugt werden. Hierzu dient ein *p*-Injektor. Ihm wird in Durchlaßrichtung Strom zugeführt. Der Injektor bildet zusätzlich mit den benachbareten *p*-Gebieten (Basis der Mehrfachkollektortransistoren) einen Lateraltransistor (Teilbild b)). Deshalb wandern die Ladungsträger in das Gebiet, wo vorher durch Licht Ladungsträger erzeugt wurden, und schon ist das alte Prinzip von Teilbild a) wieder realisiert. In Nachbarschaft eines Injektors sind dann nur noch für eine effektive Flächenausnutzung möglichst viele Transistoren anzuordnen. So entsteht eine Struktur, wie sie Teilbild c) zeigt. Rechts und links von einem schmalen Injektorstreifen werden mehrere längliche Basisgebiete, in denen jeweils mehrere Kollektoren eingelassen sind, angeordnet. Die Verdrahtung mittels der Leitmaske erzeugt dann die logische Verknüpfung.

Nach den vorangegangenen Betrachtungen wird die meist gezeigte Grundschaltung der I²L-Technik entsprechend Teilbild d) verständlich. Weiter wird ersichtlich, welche hohe Packungsdichte und große Einfachheit der I²L-Schaltung innewohnt.

8.2.6.4. Schaltzeiten und Injektorstromstärke

Ein spezieller Vorteil der I²L-Technik besteht darin, daß mittels der Injektorstromstärke die Leistung und die Schaltzeit beeinflußt werden können. Es sei zunächst von so großen Strömen abgesehen, daß die Speicherheit der Ladungsträger von Einfluß ist. Dann werden nur Kapazitäten umgeladen, und die Umladezeit beträgt

$$T = C \cdot U/I\,, \tag{23}$$

wobei U der Spannungshub, also etwa 0,7 V ist. Die von einem Gatter aufzubringende Leistung beträgt

$$P = U \cdot I\,. \tag{24}$$

Daraus folgt

$$T \cdot P = C \cdot U^2 = \text{Konstante}\,. \tag{25}$$

Proportional mit dem Injektorstrom wird also sowohl die Leistung als auch die Schaltzeit beeinflußt, wobei das Produkt beider eine von der Schaltung und der Technologie abhängige Konstante ist. Dies zeigt beispielhaft Bild 8.2.17. Dabei sind die Werte von 1974 und die theoretischen Grenzwerte ausgewählt. Bei hohen Strömen wirkt sich die Speicherzeit der Ladungsträger nachteilig aus. In einer komplexen

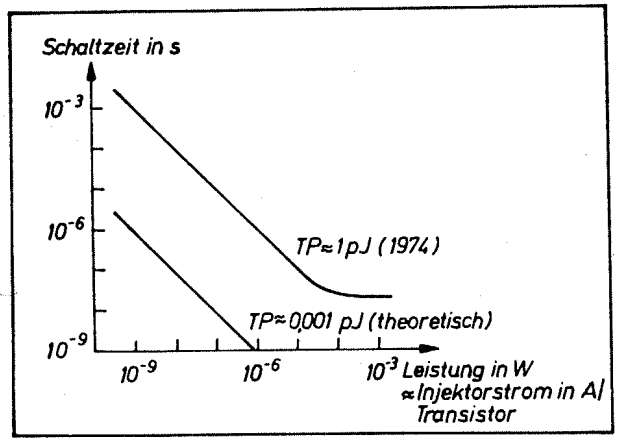


Bild 8.2.17. Prinzipieller Verlauf der Schaltzeit als Funktion der Injektionsstromstärke je Transistor, einschließlich der theoretisch erreichbaren Grenze

Schaltung ist es schließlich möglich, die Injektorstromstärke je nach der geforderten Geschwindigkeit der Teilschaltung unterschiedlich zu wählen. Die *Vorteile* der I²L-Technik sind beträchtlich. Hier seien nur einige Fakten genannt:

- hohe Packungsdichte; nur ein bis zwei Transistoren je Gatter; keine Widerstände;
- geringe Betriebsspannung: $\geqq 0{,}8$ V;
- hohes Fan-out: bis zu zehn Kollektoren zusammenschaltbar, sehr viele Baisanschlüsse zusammenfaßbar;
- wählbare bis einstellbare Geschwindigkeit:
  übliche Werte 0,02 bis 2 pJ und 5 ... 20 ns;
- geringer Prozeßaufwand: zwei Diffusionen, vier Masken.

8.2.6.5. Weiterentwicklungen

Die I^2L-Technik hat trotz ihres prinzipiell sehr guten Standes eine Vielzahl von Änderungen und Verbesserungen erfahren. Sie betreffen einmal die Technologie. Statt der Isolation über sperrend wirkende n-Streifen wurde auch die mit SiO_2 arbeitende Isoplanartechnologie angewendet. So entstand dann die Bezeichnung I^3L. In den Burried layer (n^+) wurde eine kleine p^+-Strecke eingelassen und als Injektor verwendet. Da hier der Injektorstrom vertikal wirkt, entstand die Bezeichnung VIL (vertcal injection logic). Statt des Multikollektors wurden dem Kollektor SCHOTTKY-Dioden angefügt. So entstand die SCHOTTKY-I^2L-Technik. Werden die Dioden direkt für die logische Verknüpfung ausgenutzt, spricht man von STL (SCHOTTKY-Transistor-Logik). Weiter konnte mit SCHOTTKY-Dioden der Schalthub unter 0,3 V abgesenkt werden. Diese und viele weitere Varianten trugen dazu bei, daß die I^2L-Technik mit ihren Varianten zu einer wichtigen Technik für die auf dem Chip zu realisierende Logik ist. Dabei besitzt sie noch den besonderen Vorteil, daß relativ leicht zu TTL-Pegeln überzugehen ist. Es genügen am Eingang und am Ausgang der Schaltung jeweils ein normaler npn-Transistor.

8.2.7. MOS-Techniken

Es gibt eine Vielzahl von MOS-Techniken. Sie sind nur z. T. durch die technologische Entwicklung bedingt. Prinzipiell bestehen die Möglichkeiten der p-MOS-, n-MOS- und CMOS-Technik. Für sie gibt es jeweils getrennte technologisch bedingte Varianten (vgl. z.B. Bild 5.5.10). Daneben haben noch die dynamischen MOS-Techniken Bedeutung. Sie werden bei den Schieberegistern im Abschnitt 8.5.5. behandelt. Schließlich ist noch eine spezielle MOS-Technik auf der technologischen Basis von SOS (silicon-on-sapphire) zu behandeln.

8.2.7.1. p- und n-MOS-Technik

In Bild 3.3.1 sind die verschiedenen Varianten der p- und n-MOS-Technik zusammengestellt. Bei den Anreicherungstypen besitzt die Gatespannung dasselbe Vorzeichen wie die Betriebsspannung. Dadurch werden diese Typen vorteilhaft für digitale Anwendungen. Deshalb hat auch der Sperrschicht-FET, der dem Verarmungstyp entspricht, in der Digitaltechnik keine Bedeutung erlangt.

Die älteste MOS-Technik ist die p-MOS-Technik mit metallischem Gate, meist Aluminium. Sie wurde 1966/67 produktionsreif. Da ihre Schwellspannungen bei -3 V bis -5 V liegen, werden Betriebsspannungen von -25 V bis -30 V benötigt. Deshalb wird hier von Hochvolttechnik gesprochen. Mittels zusätzlicher technologischer Schritte, u. a. Ionenimplantation, gelang es, die Schwellspannungen auf $-1{,}5$ V bis $-2{,}5$ V herabzusetzen. Dadurch genügen Betriebsspannungen von -8 V bis -13 V für diese Niedervolttechnik. Um 1970 konnte das polykristalline Si-Gate realisiert werden. Dadurch gelingt eine gezielte Einstellung der Schwellspannungen, und mit Betriebsspannungen von $+5$ V und -9 V oder -12 V gelang es, Kompatibilität zur TTL-Technik herzustellen.

Mit der Silicium-Gate-Technik gelang es schließlich, bei der n-MOS-Technik volle TTL-Kompatibilität auch bei 5 V Betriebsspannung herzustellen, und es konnte darüber hinaus infolge der 2,5fachen Beweglichkeit der Elektronen gegenüber den Löchern eine höhere Geschwindigkeit realisiert werden. Die Technologien der HMOS (high performance MOS) sowie VMOS bzw. DMOS sind technologische Weiterent-

wicklungen (vgl. Bild 5.5.10). VMOS und DMOS sind n-Kanal-Techniken, erlauben hohe Packungsdichten und Geschwindigkeiten bis zu 20 bzw. 5 ns.

8.2.7.2. Prinzipieller Aufbau der Gatter und Logik

Grundbauelement aller logischen Schaltungen ist der Inverter. Er wird in den MOS-Techniken genau wie in den bipolaren aus einem Transistor mit Arbeitswiderstand aufgebaut. Da bei MOSFET jedoch zu große Widerstände notwendig wären, wird hier — wie bereits im Zusammenhang mit Tab. 5.1.2 gezeigt — ebenfalls ein FET verwendet. Das Grundprinzip eines Inverters werde an der Schaltung mit zwei Anreicherungs-FET vom n-Kanal-Typ gemäß Bild 8.2.18a) und der Kennlinie von Bild 8.2.18b) betrachtet.

Bei kleiner Spannung am Eingang ist der untere Transistor gesperrt. Am oberen muß aber soviel Spannung abfallen, daß er leitend ist. Dies ist zumindest die Schwellenspannung U_0. So ergibt sich die obere Gerade der Kennlinie im Bild. Wird nun die Eingangsspannung positive, so erreicht sie ebenfalls den Wert U_0, und durch T_1 beginnt der Strom zu fließen. Der Spannungsabfall erfolgt zunächst linear, da T_1 im Einschnürbereich arbeitet (vgl. Bild 3.3.3a)). Sobald aber die Spannung an T_1 für diesen Betrieb nicht mehr ausreicht, geht der Verlauf parabelförmig weiter, und zum Schluß verbleibt an T_1 nur noch eine gewisse Restspannung U_r. Sie kann durch die topologische Auslegung des Verhältnisses der Kanallängen und -breiten beider Tran-

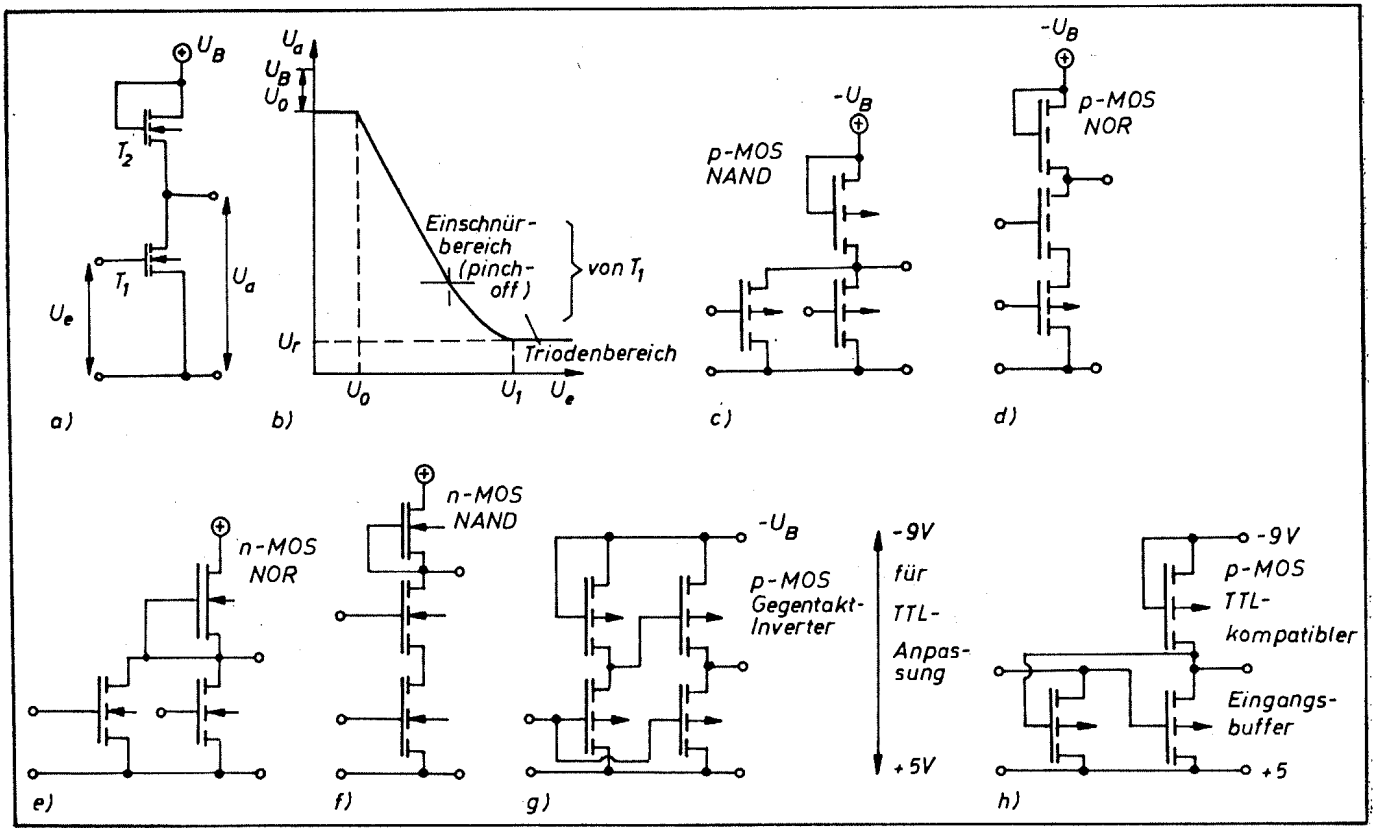


Bild 8.2.18. Prinzipieller Aufbau von Invertern in p- bzw. n-MOS-Technik

a) Prinzip eines n-MOS-Inverters
b) Kennlinie zu a)
c) NAND-Schaltung (positive Logik) bei p-MOS-Technik
d) Dito für NOR-Schaltung
e) NOR-Schaltung (positive Logik) bei n-MOS-Technik
f) Dito für NAND-Schaltung
g) p-MOS-Gegentaktinverter sowie Betriebsspannungsangaben, wenn Anpassung an TTL-Pegel erfolgen soll
h) TTL-kompatibler Eingang für p-MOS-Schaltungen

sistoren beeinflußt werden. Die Restspannung muß so klein gewählt werden, daß durch sie nachgeschaltete Stufen mit Sicherheit gesperrrt werden, also $U_r < U_0$. Es läßt sich zeigen, daß dazu T_2 hochohmiger sein muß. Sein Kanal muß also länger und schmaler als der von T_1 sein. Damit auch mit Sicherheit die Restspannung U_r erreicht wird, muß für die größtmögliche Ausgangsspannung gelten: $U_{a,max} > U_1$, wobei U_1 der Einmündungspunkt der Parabel ist. Dies bedeutet, daß eine hinreichend große Ausgangsspannung zu verwenden ist. Ferner folgt aus beiden Relationen, daß kleine Betriebsspannungen nur dann möglich sind, wenn eine kleine Schwellenspannung und große Steilheit bei den FET erreicht werden.

Die beiden prinzipiellen logischen Grundschaltungen zeigt für p-MOS Bild 8.2.18c) und d) und für n-MOS e) und f). Da p-MOS-Schaltungen eine negative Betriebsspannung benötigen, vertauschen sich bei positiver Logik gegenüber n-MOS und TTL die Anordnungen (Reihen- und Parallelschaltung) bezüglich NAND und NOR (8.2.3.3.) Dies hat zur Folge, daß bei der p-MOS-Technik recht oft mit der negativen Logik gearbeitet wird. Im Bild sind nur die logischen Funktionen für positive Logik angeschrieben. Bei den p-MOS-Schaltungen werden in der Regel auch für den MOS-Widerstand Anreicherungs-FET verwendet. Bei den n-MOS-Schaltungen wird dagegen meist ein Verarmungstyp verwendet. Dies ist möglich, da hier ohnehin eine gezielte Beeinflussung der Schwellspannung infolge des Polysiliciumgate möglich ist und so Vorteile in der Technologie und im Betrieb des Gatters entstehen. Typische Werte für die n-MOS-Technik sind heute 50 ns und 0,2 mW je Gatter.

Auf dem Chip selbst genügt in der Regel das einfache Gatter. Soll aber zu äußeren Anschlüssen übergegangen werden, so wird meist mehr Leistung benötigt (z. B. eine TTL-Last). Hierfür ist ein Gegentaktinverter gemäß Bild 8.2.18g) vorteilhaft, da bei ihm immer einer der beiden Ausgangstransistoren leitend und der andere gesperrt ist. Mit diesem Gatter kann weiter bei p-MOS-Technik ein Übergang zu den TTL-Pegeln erreicht werden. Es sind dann allerdings die beiden Betriebsspannungen von -9 V und $+5$ V erforderlich. Für den Eingang von p-MOS-Schaltungen ist dies ebenfalls mit einer speziellen Schaltung gemäß Bild 8.2.18h) möglich. Da die MOS-Gates sehr empfindlich gegenüber Spannungsspitzen größer als ca. 50 V sind, werden an den nach außen führenden Eingängen in der Regel zusätzlich Schutzdioden angeschaltet. Die MOS-Gatter benötigen statisch gesehen keine Steuerleistung. Dies ermöglicht theoretisch ein beliebig großes Fan-out. Praktisch liegen jedoch an den Ein- und Ausgängen verschiedene Kapazitäten vor, die auf- bzw. umgeladen werden. Der dafür notwendige Strom wird um so größer, je höher die Frequenzen sind. Dies begrenzt das praktisch nutzbare Fan-out bzw. die maximale Frequenz.

Neben den bisher behandelten MOS-Techniken gibt es auch noch eine dynamische MOS-Technik. Hierbei werden die Signalpegel vorübergehend in den Kapazitäten gespeichert und dann getaktet weiter verarbeitet. Auf eine Besprechung wird hier verzichtet, sie entspricht in etwa den Prinzipien der dynamischen Speicher (8.3.5.4.) und der dynamischen Register (8.5.5.7.) [M 19].

8.2.7.3. CMOS-Schaltungen

Durch die Kombination von p-MOS und n-MOS läßt sich ein CMOS-Inverter aufbauen, wie ihn Bild 8.2.19a) zeigt. Dieser Inverter hat den Vorteil, daß er statisch praktisch keinen Strom verbraucht, da immer genau einer der beiden Transistoren leitend und der andere gesperrt ist. Es fließt also im statischen Betrieb nur ein Reststrom von einigen nA. Dies hat weiter zur Folge, daß der Pegel am Ausgang sehr gut durch OV bzw. die Betriebsspannung beschrieben werden kann. In der Praxis kann so ein sehr hoher Störabstand (verbotene Zone) erreicht werden. Weiter kann bei kleinen Schwellspannungen im Bereich von 1,5 V (oder kleiner) ein sehr großer Be-

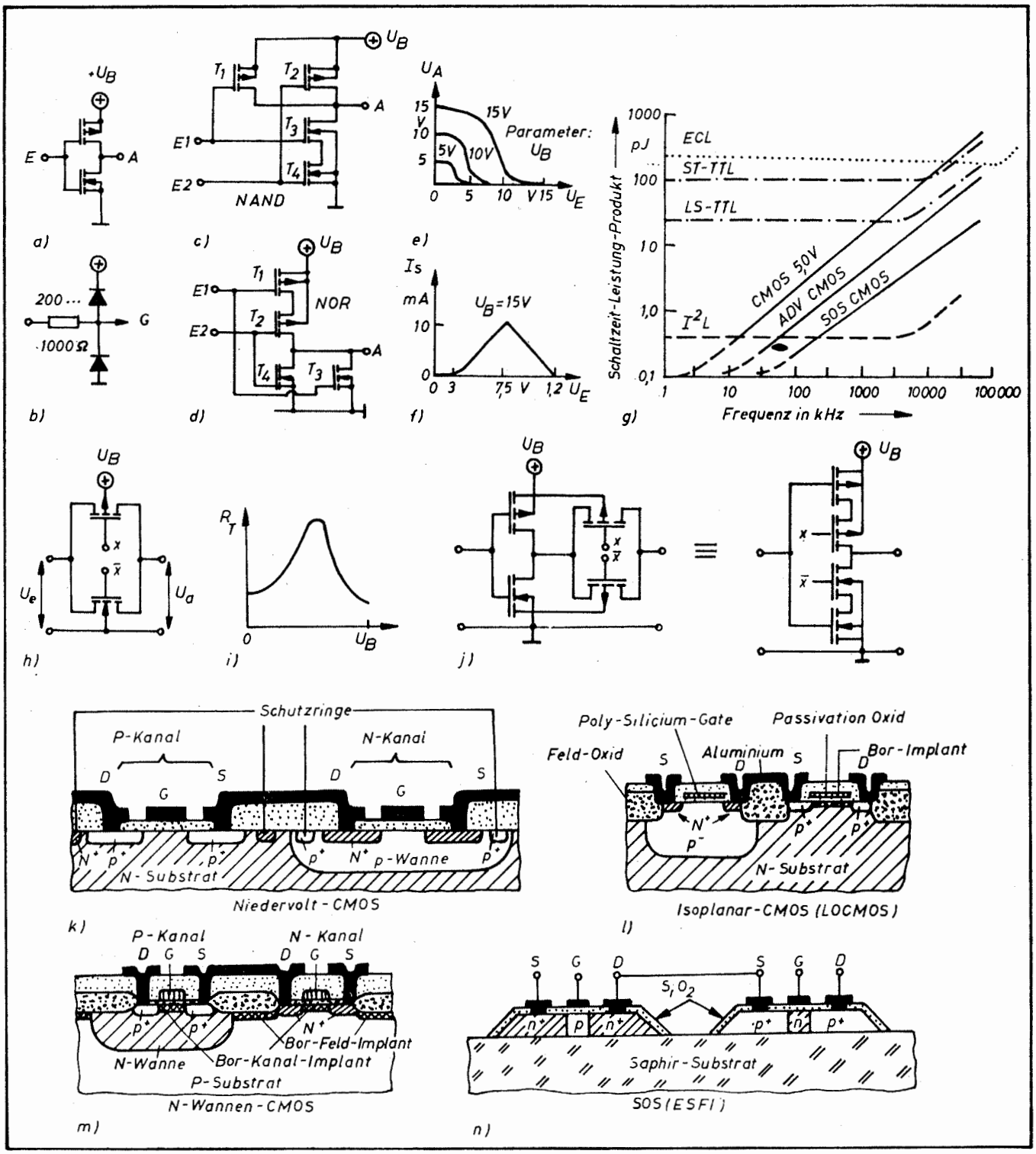


Bild 8.2.19. CMOS-Technik

a) Grundprinzip eines Inverters
b) Prinzip der Eingangsschutzschaltung
c) NAND-CMOS-Gate (positive Logik)
d) NOR-CMOS-Gate (positive Logik)
e) Übertragungskennlinien für verschiedene Betriebsspannungen eines Gate
f) Stromaufnahme eines Gate bei der Betriebsspannung 15 V
g) Schaltzeit-Leistungsprodukt für verschiedene Schaltkreisfamilien (ADV = advanced)
h) Transmissionsgate
i) Widerstandsverlauf beim Transmissionsgate
j) Äquivalente Schaltungsvarianten für einen Inverter
k) bis n) Verschiedene Strukturen bei unterschiedlichen Technologien

triebsspannungsbereich U_B von ca. 3 V (1,5 V) bis 15 V zugelassen werden. Die Pegelbereiche werden dann meist zu folgenden Werten festgelegt:

$$0 \leqq L \leqq 0{,}3 \cdot U_B \quad \text{und} \quad 0{,}7 \cdot U_B \leqq H \leqq U_B \,. \tag{26}$$

Die Übertragungskennlinie ist in der Regel gut symmetrisch (Bild 8.2.19e)).

Ein Stromverbrauch tritt bei CMOS im dynamischen Betrieb auf. Es müssen ja die Leitungskapazitäten und die Gatekapazitäten der folgenden Stufen umgeladen werden. Dies äußert sich einmal in der Stromspitze gemäß Teilbild f). Es hat aber auch zur Folge, daß der Leistungsverbrauch einer CMOS-Schaltung mit der Betriebsfrequenz zunimmt. Es gilt etwa

$$P_v = C_L U_B^2 f \,. \tag{27}$$

So entsteht ein Verlauf des Schaltzeit-Leistungs-Produkts, wie ihn Teilbild g) zeigt. Er steht zumindest bei tiefen bis mittleren Frequenzen in Widerspruch zu anderen Schaltungen. Es wird deutlich, daß CMOS in jedem Fall bei tiefen Frequenzen im Vorteil ist. Ja, in Ruhe genügt zur Aufrechterhaltung ein so kleiner Strom, daß sich ein Abschalten der Betriebsspannung oft nicht lohnt.

Die Ausführung je eines NOR- und NAND-Gatters zeigen die Teilbilder d) und c). Es gehören also immer eine Reihenschaltung in einem Zweig und eine Parallelschaltung im anderen Zweig zusammen.

Eine Besonderheit der CMOS-Technik ist die Möglichkeit von *Transmissionsgates* (Teilbild h)). Sie können einen Strom in beiden Richtungen passieren lassen oder sperren. Dadurch ist es möglich, CMOS-Schaltungen *ähnlich den Tristate-TTL* zu realisieren. Die Abhängigkeit des Ein-Widerstandes R_T von der anliegenden Spannung ist zwar vorhanden, stört aber meist nicht.

Die Zusammenschaltung eines Inverters mit einem Transmissionsgate zeigt Teilbild j). Im zweiten Teil des Bildes ist eine andere Struktur mit äquivalenten Eigenschaften gezeigt. Damit soll zugleich demonstriert sein, daß in der CMOS-Technik für komplexere Schaltungen mehrere Varianten als Lösung in Betracht zu ziehen sind. Dabei sind oft nicht jene Varianten mit der geringsten Anzahl von Transistoren am besten, denn die benötigte Fläche auf dem Siliciumsubstrat hängt auch von anderen Fakten ab. Zuvor seien jedoch noch die typischen Daten der CMOS-Schaltungen zusammengestellt. Sie wurden im wesentlichen *1976 als JEDEC-B-Standard* festgelegt. Für eine Standardschaltung gilt etwa:

Betriebsspannungsbereich	3 ... 15 V
Ruhestromverbrauch	0,01 ... 1 µA
Eingangswiderstand	10^9 ... 10^{12} Ω
Eingangskapazität	5 pF
Eingangsstrom	10 pA
Ausgangswiderstand	< 500 Ω
Restspannung	0,05 V
Statisches Fan-out	≈ 1000
Schaltzeit (ganz grob)	100 ns

Gegenüber TTL sind noch folgende *Vorteile* zu erwähnen:

- keine stabilisierte Betriebsspannung notwendig,
- großer Temperaturbereich, ca. −40 ... +85 °C.

Gegenüber den anderen MOS-Schaltkreisen bestehen neben Vorteilen auch *Nachteile*:

- relativ große benötigte Chipfläche (3—4mal),
- komplizierte Technologie.

8.2.7.4. CMOS-Technologien

In der Technologie der CMOS-Schaltkreise hat es ebenfalls eine große Entwicklung gegeben. Die Standardtechnologie seit etwa 1975 zeigt Bild 8.2.19k). Der *n*-Kanal-MOSFET befindet sich in einer *p*-Wanne. Um die Kanalausbildung in unerwüschte Bereiche zu vermeiden, sind weiter beide FET mit einem Schutzring (Kanalstopper; guard board) umgeben. Hierdurch wird eine relativ große Fläche benötigt. Mit der Isoplanartechnologie (Teilbild 1)) konnte dies durch tief eingesenkte SiO_2-Flächen (das Feldoxid) überflüssig gemacht und so viel Fläche eingespart werden. Eine Firma spricht in diesem Zusammenhang auch von LOCMOS (lokale Oxydations-CMOS).

Um 1979 entstand die zweite Generation der CMOS-Technologie. Sie verwendet nicht *p*-, sondern *n*-Wannen und nutzt stark die Möglichkeiten der Ionenimplantation aus (Teilbild m)). Sie hat neben verschiedenen weiteren Vorteilen auch den, daß mit dieser Technologie gleichzeitig sehr gute Schaltungen in *n*-MOS hergestellt werden können.

Ein Nachteil aller CMOS-Schaltungen (und nicht nur dieser) besteht in den parasitären Bauelementen, z. B. den großen Kapazitäten zwischen den einzelnen Isolationsinseln. Bei der CMOS-Technik existiert als Folge dieser Gegebenheiten z. B. das Latch-up über parasitäre Transistoren. Es entsteht folgendermaßen (Teilbild k) und l)): Vom *p*-Gebiet der Drain über das *n*-Substrat zur *p*-Wanne und zum *p*-Gebiet der Source besteht die Folge *pnpn*. Diese entspricht einem Vierschichtbauelement (vgl. 6.4.2.), und damit entsteht die Möglichkeit, daß beim Überschreiten einer bestimmten Betriebsspannung der CMOS-Inverter durchbricht, einen sehr hohen Strom zieht und so zerstört wird. Dieser und alle anderen Effekte lassen sich vermeiden, wenn von der klassischen Siliciumchiptechnologie abgegangen wird.

8.2.7.5. SOS-Technik

Die SOS-Technologie geht von einem Isolator aus, auf dem epitaktisch eine dünne Si-Schicht aufwächst (meist unter 1 μm). Diese Si-Schicht kann dann in Inseln mühelos mittels normaler fotolithografischer Prozesse zerlegt (lokales Ätzen) und in bekannter Weise strukturiert werden (Bild 8.2.19n)). Hierdurch entstehen viele Vorteile. Die meistgebräuchliche Technologie geht von α-Al_2O_3 (Korund) aus, dessen blaugefärbte Modifikation das Saphir ist. Daher der Name Silicon-on-Sapphire (SOS). Eine Firma verwendet Spinell ($MgAl_2O_4$) und den Namen Epitaxial Silicon Films on Isolator (ESFI).

Das Hauptproblem bei dieser Technologie ist das teure Substrat, das im Czochralski-Verfahren hergestellt und dann in Wafer zerteilt und poliert werden muß, bevor die epitaktische Si-Schicht aufwachsen kann. Dadurch ist die SOS-Technologie sehr teuer. In den letzten Jahren sind hier jedoch große Fortschritte erreicht worden, insbesondere um 1979 durch das EFG (= edge defined film-fed growth). Hierbei werden direkt verwendbare Bänder gezogen. Es ist zu erwarten, daß die SOS-Technologie generell, vor allem aber bei CMOS, große Bedeutung erlangen könnte.

8.2.8. Ausgewählte Beispiele

8.2.8.1. Koder — Dekoder

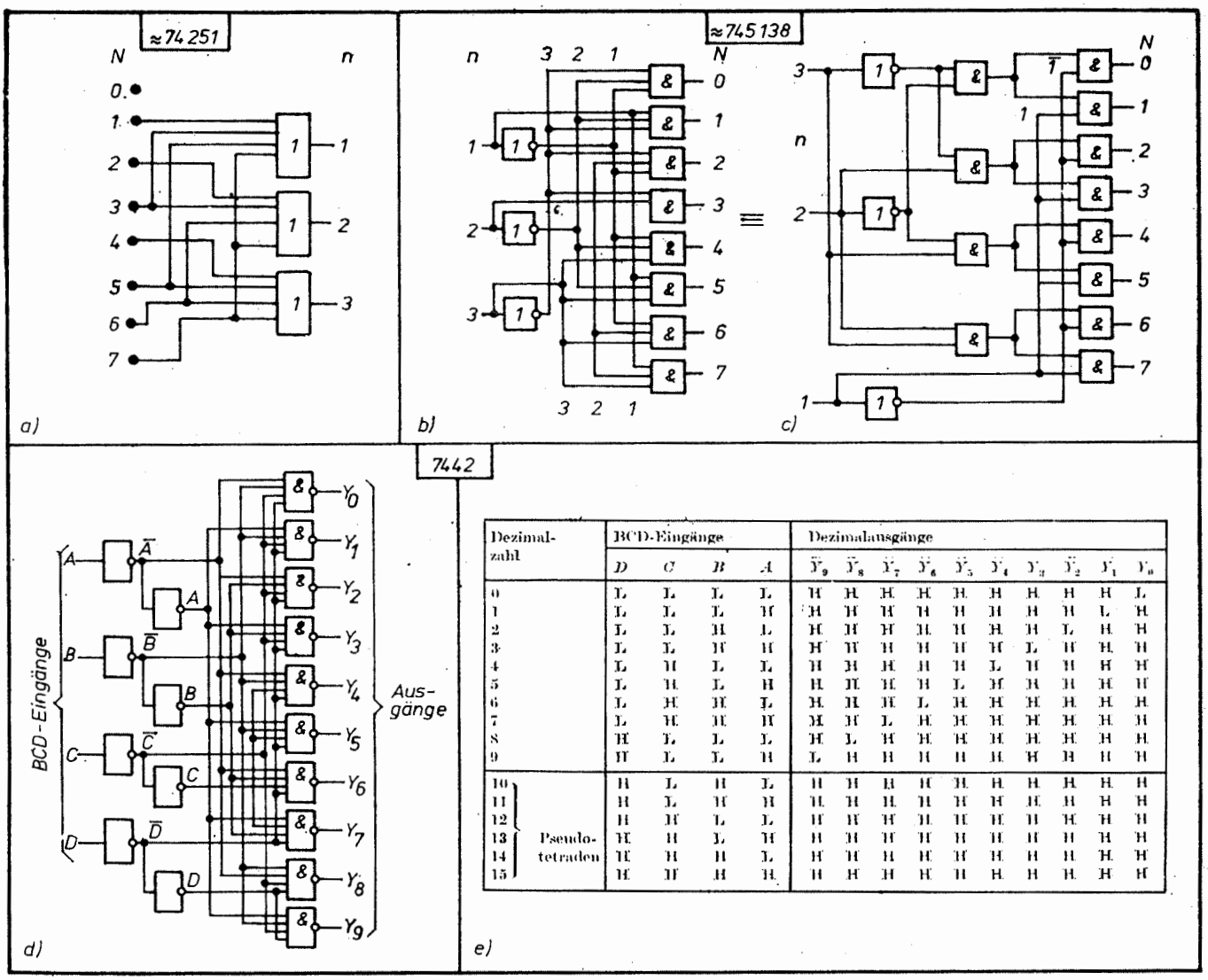
Es mögen auf der einen Seite N Zustände existieren, von denen genau nur jeweils einer möglich ist, und auf der anderen Seite ein Binärwort der Länge n, wobei an den n Stellen in jeder beliebigen Kombination die logischen Werte Ø und 1 auftreten

können. Zwischen diesen beiden Gegebenheiten soll vermittelt werden. Dies führt theoretisch auf die Problematik des Abschnitts 8.1.4. zurück. Die technische Lösung kann mit logischen Schaltungen erfolgen, die dann Kodierer bzw. Koder oder Dekodierer bzw. Dekoder heißen. Zwischen den beiden Zahlen N und n besteht die Relation

$$N \leqq 2^n \quad \text{bzw.} \quad n \geqq \operatorname{ld} N\,, \tag{28}$$

wobei das $<$ bzw. $>$ dann auftritt, wenn N eine ganze Zahl ist, die nicht als 2^n mit einer ganzen Zahl n darstellbar ist. Zum Beispiel wäre für $N = 13$ $n = \operatorname{ld} 13 \approx 3{,}7$, also muß $n = 4$ gewählt werden, womit dann im Prinzip aber $2^4 = 16$ Zustände kodiert werden könnten. Dieses Problem wurde bereits bei den Pseudotetraden im Abschnitt 8.1.4.2. für den Spezialfall $N = 10$ behandelt.

Um die Struktur derartiger Koder und Dekoder darzustellen, wurde Bild 8.2.20 gezeichnet. In Teilbild a) wird der Fall dargestellt, daß $N = 8$ Zustände existieren, die in ein 3-bit-Wort zu kodieren sind. In diesem Fall wurde eine einstufige Kodierung gewählt. Sie erfordert bei N Zuständen immer n Gatter mit je $N/2$ Eingängen. Sie



Dezimal-zahl	BCD-Eingänge				Dezimalausgänge									
	D	C	B	A	$\overline{Y}_9$	$\overline{Y}_8$	$\overline{Y}_7$	$\overline{Y}_6$	$\overline{Y}_5$	$\overline{Y}_4$	$\overline{Y}_3$	$\overline{Y}_2$	$\overline{Y}_1$	$\overline{Y}_0$
0	L	L	L	L	H	H	H	H	H	H	H	H	H	L
1	L	L	L	H	H	H	H	H	H	H	H	H	L	H
2	L	L	H	L	H	H	H	H	H	H	H	L	H	H
3	L	L	H	H	H	H	H	H	H	H	L	H	H	H
4	L	H	L	L	H	H	H	H	H	L	H	H	H	H
5	L	H	L	H	H	H	H	H	L	H	H	H	H	H
6	L	H	H	L	H	H	H	L	H	H	H	H	H	H
7	L	H	H	H	H	H	L	H	H	H	H	H	H	H
8	H	L	L	L	H	L	H	H	H	H	H	H	H	H
9	H	L	L	H	L	H	H	H	H	H	H	H	H	H
10	H	L	H	L	H	H	H	H	H	H	H	H	H	H
11	H	L	H	H	H	H	H	H	H	H	H	H	H	H
12	H	H	L	L	H	H	H	H	H	H	H	H	H	H
13	H	H	L	H	H	H	H	H	H	H	H	H	H	H
14	H	H	H	L	H	H	H	H	H	H	H	H	H	H
15	H	H	H	H	H	H	H	H	H	H	H	H	H	H

(Pseudotetraden)

Bild 8.2.20. Beispiel für Koder und Dekoder
(Die Teilbilder a) bis c) sind nur formal richtig, da sie eine relativ hohe Eingangslast besitzen)

a) Kodierung der 8 Zustände (0 bis 7) in ein 3-bit-BCD-Format
b) Dekodierung von 1 bis 8 aus 3-bit-BCD in Parallelbetrieb
c) in serieller Form
d) Logische Innenschaltung des BCD-zu-Dezimal-Dekoders 7442
e) Pegeltabelle zu d)

ist also nur sinnvoll bis $N = 16$ (8 Eingänge) durchführbar. Darüber hinaus muß die Kodierung mehrstufig vorgenommen werden. Dieses Prinzip wird bei den folgenden Dekodern erklärt. Koder werden beispielhaft bei den AD-Wandlern im Abschnitt 9.2. behandelt.

Zwei mögliche Dekoder von $n = 3$ nach $N = 8$ zeigt Teilbild b) bzw. c). Teilbild b) ist, von den Negatoren abgesehen, ein *einstufiger Dekoder*, Teilbild c) dagegen ein *mehrstufiger*, auch sequentieller, der jeweils nur Gatter mit zwei Eingängen verwendet. Bedeutsam ist es, den Aufwand dieser beiden Extremfälle zu vergleichen. Wird die *Summe aller Eingänge* einschließlich der Negatoren als Gatteranzahl A gezählt, so gilt für den allgemeinen Paralleldekoder (b)

$$A_p = n(2^n + 1) \tag{29}$$

und für den seriellen Dekoder

$$A_s = n + \sum_{k=3}^{n+1} 2^k . \tag{30}$$

Für einige n ergeben sich so die folgenden Zahlenwerte:

n	2	3	4	5	6	8	10	16
A_p	10	27	68	165	390	2056	10250	1048 592
A_s	10	27	60	125	254	1024	4098	262 152

Aus dieser Sicht ist also für $n > 3$ deutlich *der serielle Wandler überlegen. Er hat jedoch dafür zwei Nachteile*:

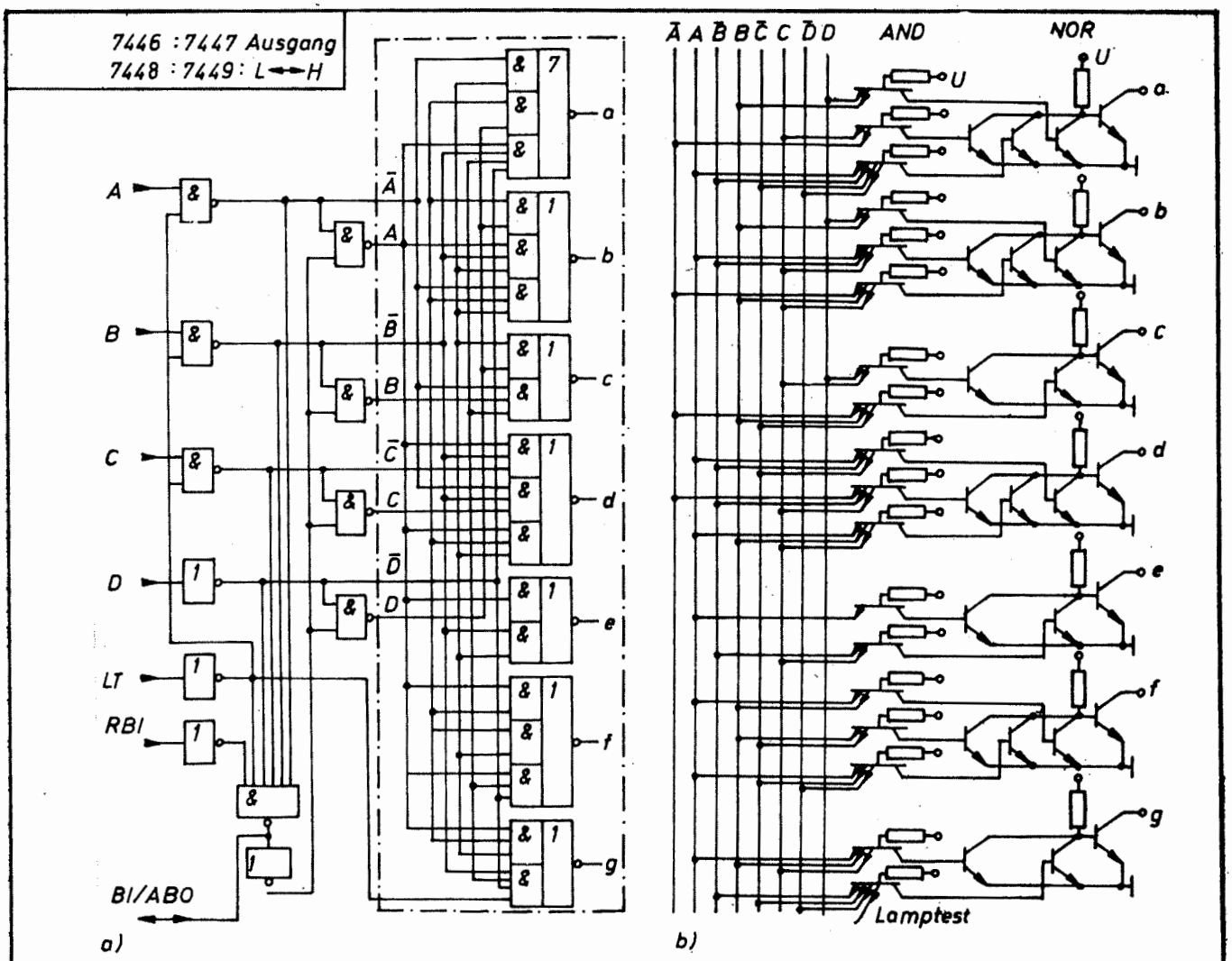
- Seine Geschwindigkeit ist durch die Mehrstufigkeit geringer.
- Es können Hazards (vgl. 8.2.8.8.) auftreten.

Wie im Abschnitt 8.1.3. bez. des minimalen Bezeichnungsaufwands gezeigt wurde, wären 3 Eingänge je Gatter (näher an e $\approx$ 2,71) eventuell noch vorteilhafter. Doch dies führt nur in Ausnahmen zu guten Lösungen (3^x meist ungleich 2^n). Vgl. auch z.B. [K 30].

Ein weiteres Beispiel eines Dekoders zeigt Bild 8.2.20d) als Schaltung und e) als Pegeltabelle. Hiermit wird eine BCD-Zahl in die entsprechende Dezimalzahl so umgewandelt, daß alle Pseudotetraden die Dezimale Null (negative Logik, L entspricht der Zahl am Ausgang) ergeben. An diesem konkreten Beispiel wird weiter deutlich, daß in der Regel die Eingänge nur eine TTL-Last aufweisen sollen. Deshalb beginnt die Schaltung mit vier Negatoren für $\overline{A}$ bis $\overline{D}$, und aus diesen werden mit jeweils vier weiteren Negatoren die Pegel für A bis D endgültig bereitgestellt. Das nötige Fan-out für die folgenden 10 NAND wird also intern erzeugt. Dies wurde bei den Bildern 8.2.20a) bis c) bewußt ausgelassen.

8.2.8.2. BCD-7-Segment-Dekoder

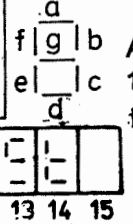
Für die Anzeige ist der BCD-Kode mittels einer 7-Segment-Anzeige darzustellen. Hierbei erweist sich eine zweimalige Umsetzung als notwendig: einmal eine Dekodierung des BCD in 1 aus 10 und dann eine Kodierung auf die 7 Segmente a bis g einer Anzeige (vgl. Bild 6.1.7 oben). Diese Zusammenhänge können besonders übersichtlich mittels der Tabelle in Bild 8.2.21 dargestellt werden. Das BCD-Signal ist mittels der Eingänge A, B, C, D (international genormt) gegeben. Die zugehörigen Zahlen Ø bis 9 sowie die Pseudotetraden 10 bis 15 bzw. A bis F enthält die erste Spalte. Die Segmentbezeichnung a bis g enthält das Bild unten bzw. die Ausgänge der Tabelle. L bedeutet

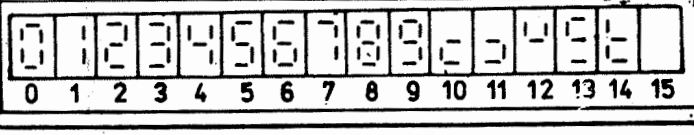


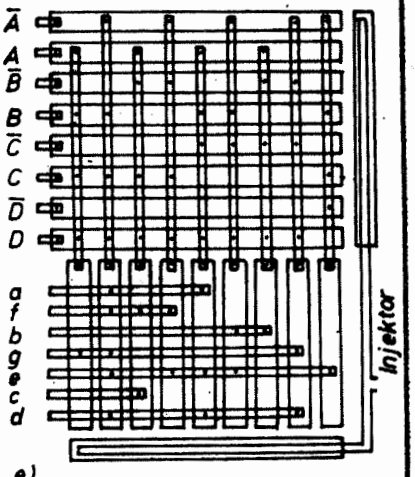
Dezimalzahlen oder Funktion	Eingänge						Ausgänge							
	LT	RBI	D	C	B	A	BI/RBO	a	b	c	d	e	f	g
0	H	H	L	L	L	L	H	L	L	L	L	L	L	H
1	H	X	L	L	L	H	H	H	L	L	H	H	H	H
2	H	X	L	L	H	L	H	L	L	H	L	L	H	L
3	H	X	L	L	H	H	H	L	L	L	L	H	H	L
4	H	X	L	H	L	L	H	H	L	L	H	H	L	L
5	H	X	L	H	L	H	H	L	H	L	L	H	L	L
6	H	X	L	H	H	L	H	L	H	L	L	L	L	L
7	H	X	L	H	H	H	H	L	L	L	H	H	H	H
8	H	X	H	L	L	L	H	L	L	L	L	L	L	L
9	H	X	H	L	L	H	H	L	L	L	L	H	L	L
10 (A)	H	X	H	L	H	L	H	H	H	H	L	L	H	L
11 (B)	H	X	H	L	H	H	H	H	H	L	L	H	H	L
12 (C)	H	X	H	H	L	L	H	H	L	H	H	H	L	L
13 (D)	H	X	H	H	L	H	H	L	H	H	L	H	L	L
14 (E)	H	X	H	H	H	L	H	H	H	H	L	L	L	L
15 (F)	H	X	H	H	H	H	H	H	H	H	H	H	H	H
BI	X	X	X	X	X	X	L	H	H	H	H	H	H	H
RBI	H	L	L	L	L	L	L	H	H	H	H	H	H	H
LT	L	X	X	X	X	X	H	L	L	L	L	L	L	L

BI Eingang Hell/Dunkel-Tastung (Intensitätssteuerung)
RBI Eingang Dunkelsteuerung der Ziffer 0
LT Eingang Lampentest
RBO Ausgang Dunkelsteuerung für Ziffer 0
X Schaltzustand beliebig (L bzw. H), don't care

c)







hier Low, kann aber auch ausnahmsweise als leuchtendes Segment verstanden werden. Mit den negierten Werten $\overline{A}$ bis $\overline{D}$ gilt der umrandete Teil in Teilbild a). In diesem Teilbild sind darüber hinaus die Sonderfunktionen der Standardschaltkreise 7446/7447 bzw. D146/D147 enthalten:

BI	Hell-Dunkel-Tastung
RBI	Dunkelsteuerung der Null (Nullunterdrückung)
RBO	Ausgang der Dunkelsteuerung für Null
LT	Lampentest (Hellsteuerung aller Segmente).

Diese Schaltung enthält bereits 88 Gatter (Eingänge aller Teilschaltungen). Die gesamte Logik läßt auch typische Anzeigen bei den Pseudotetraden, wie sie Teilbild d) zeigt, zu. Die Wunschanzeige für Hexzahlen gemäß Teilbild f) ist nur mit Sonderschaltkreisen, z. B. in CMOS, realisiert.

Den umrandeten Teil der Schaltung in Teilbild a) zeigt Teilbild b) in TTL-Technik. Es folgen nacheinander AND- und OR-Glieder. Für die doppelte Umrechnung BCD nach 7-Segment gelten so die Gleichungen

$$\overline{a} = (\overline{A} \wedge C) \vee (B \wedge D) \vee (A \wedge \overline{B} \wedge \overline{C} \wedge \overline{D}),$$

$$\overline{b} = (B \wedge D) \vee (A \wedge \overline{B} \wedge C) \vee (\overline{A} \wedge B \wedge C),$$

$$\overline{c} = (\overline{C} \wedge D) \vee (\overline{A} \wedge B \wedge \overline{C}),$$

$$\overline{d} = (A \wedge \overline{B} \wedge \overline{C}) \vee (\overline{A} \wedge \overline{B} \wedge C) \vee (A \wedge B \wedge C),$$

$$\overline{e} = A \vee (\overline{B} \wedge C),$$

$$\overline{f} = (A \wedge B) \vee (B \wedge \overline{C}) \vee (A \wedge \overline{C} \wedge \overline{D}),$$

$$\overline{g} = (A \wedge B \wedge C) \vee (\overline{B} \wedge \overline{C} \wedge \overline{D}).$$

Die entsprechenden Terme sind unmittelbar aus dem Teilbild b) zu erkennen.

Eine Lösung in I^2L-Technik für den umrandeten Teil in Teilbild a) zeigt Teilbild e). Hier wird eine andere Auflösung der logischen Gleichungen realisiert. In den waagerechten Kollektoranschlüssen wird von links nach rechts ein Ausgangssignal erzeugt, wenn die Ziffern

Ø 1 2 3 4 5 6 7 9

in BCD vorliegen (8 fehlt!). In dem unteren Teil des Bildes werden dann über die Kollektoren jene Segmente unterdrückt, die nicht leuchten sollen. Bei 8 und den Pseudotetraden wird kein Element dunkel getastet. Der Aufwand in I^2L ist also wesentlich geringer. Statt der 43 Transistoren mit 71 Emittern bei TTL sind in I^2L nur 17 Transistoren mit 54 genutzten Kollektoren erforderlich.

◀

Bild 8.2.21. Zum BCD → 7-Segment-Dekoder

a) Schaltbild mit Zusatzfunktionen, die unter der Tabelle erklärt sind
b) TTL-Ausführung des in Bild a) umrandeten logischen Teils
c) Pegeltabelle zu a)
d) Anzeige der 15 möglichen Kombinationen der Eingänge A bis D
e) Prinzip eines Lay-out für I^2L für den in a) umrandeten Teil
f) Gewünschte (nicht realisierte) Anzeige der Eingangskombinationen 10 bis 15 für einen Hex-Kode

8.2.8.3. Multiplexer — Demultiplexer

Multiplexer (auch Datenselektor, Datenwähler oder MUX) und Demultiplexer (auch DEMUX) sind zueinander invers wie Koder und Dekoder. Sie werden für *dreierlei Aufgaben* eingesetzt:

- als Umschalter bzw. Schrittschalter,
- zur Serien-Parallel- oder Parallel-Serien-Wandlung von Daten,
- zur Realisierung spezieller logischer Funktionen.

Am unmittelbarsten ist ihre Funktion an den ersten Gebieten zu ersehen. Schematisch geht dies aus Bild 8.2.22a) hervor. Der Multiplexer schaltet Daten von mehreren Eingangsleitungen wahlweise auf eine Ausgangsleitung; der Demultiplexer schaltet Daten von einer Eingangsleitung wahlweise auf unterschiedliche Ausgangsleitungen. Werden beide Schaltungen gemäß Bild 8.2.22a) in Reihe geschaltet und die beiden Schalter synchron betätigt, so werden vom Multiplexer die primär parallelen Daten sequentiell (der Reihe nach, hintereinander) auf die eine Ausgangsleitung geschaltet und können so über die eine Leitung (statt mehrerer) übertragen werden. Der Demultiplexer schaltet die Daten am Ende einer Übertragungskette wieder so auf die einzelnen Leitungen, daß sie wieder parallel genutzt werden können.

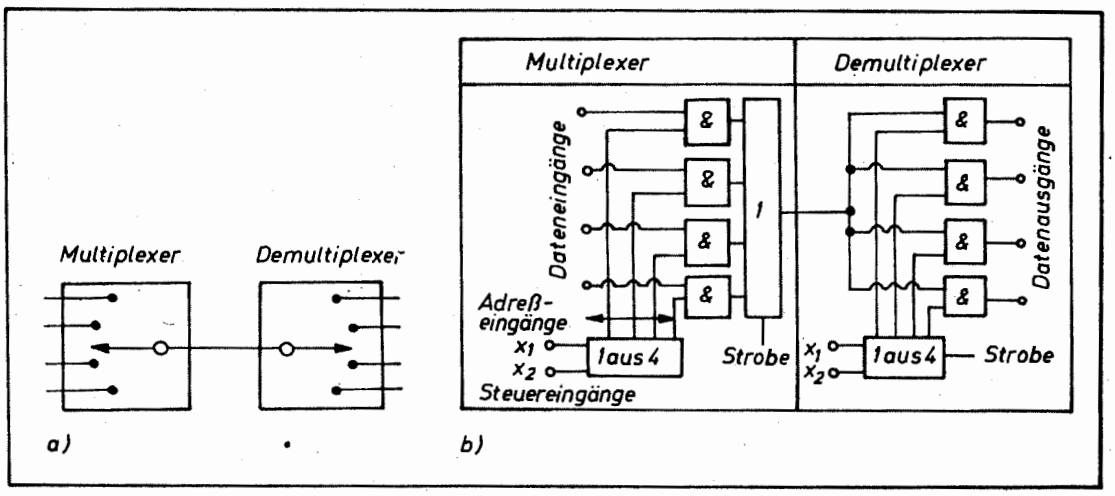


Bild 8.2.22. Multiplexer und Demultiplexer
a) in formaler Darstellung,
b) in Gatterdarstellung

Die Realisierung dieses Prinzips mit logischen Funktionen (kombinatorische Schaltungen) zeigt Bild 8.2.22b). Es werden *unterschieden*

- Datenein- bzw. -ausgänge,
- Adreßeingänge, welche die Daten z.B. beim Pegel H durchschalten,
- Steuereingänge, welche kodierte (z.B. BCD) Adreßeingänge sind, und daher weniger Leitungen benötigen (x_1, x_2 im Bild).

Schließlich existiert bei verschiedenen Multiplexern und Demultiplexern noch ein

- Strobeeingang.

Er wird zuweilen auch Steuer- bzw. Sperreingang genannt und dient dazu, generell den Durchgang von Daten zu sperren. Er ist besonders wichtig, wenn mehrere Multiplexer bzw. Demultiplexer zur Erhöhung der Anzahl von Datenein- bzw. -ausgängen zusammengeschaltet werden sollen. Weiter gibt es noch einige Multiplexer bzw. Demultiplexer mit einem

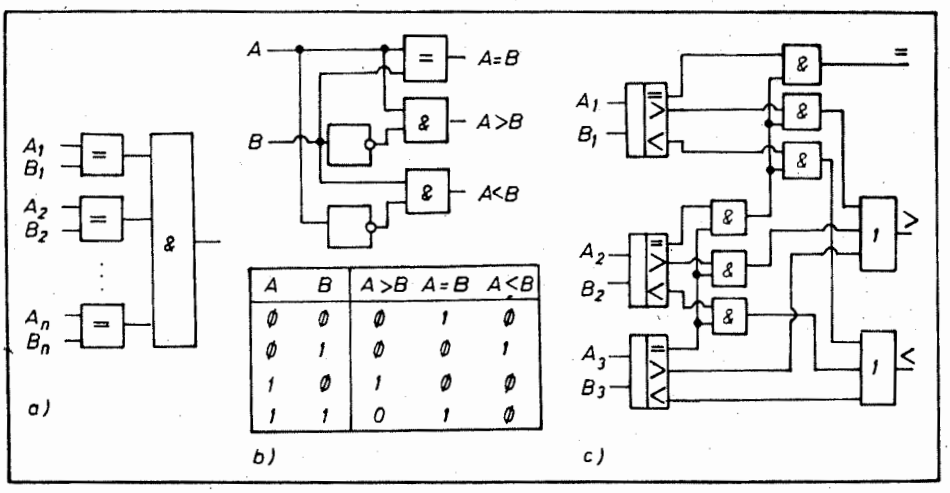
- Selecteingang.

Er ist notwendig, wenn ein Multiplexer mehrere Ausgänge bzw. ein Demultiplexer mehrere Eingänge hat. Mit dem Selecteingang wird dann zwischen diesen umgeschaltet.

Die Realisierung spezieller logischer Funktionen mit einem Multiplexer wird dann deutlich, wenn man alle Eingänge zusammen als ein Kodewort auffaßt. In diesem Sinne ist z.B. ein Demultiplexer ein Dekoder mit zusätzlichem Eingang (Dateneingang). Auf weitere Zusammenhänge hiermit wird im Abschnitt 8.2.8.7. eingegangen.

8.2.8.4. Vergleicher

Neben logischen Schaltungen werden in Rechnern auch arithmetische Schaltungen benötigt, die in der Regel allerdings aus logischen aufgebaut werden. Eine besonders einfache arithmetische Funktion ist der Vergleich von Zahlen. Dieser Vergleich erfolgt bezüglich einstelliger binärer Zahlen durch die Äquivalenz bzw. Antivalenz (vgl. Tab. 8.2.7). Soll er für mehrstellige Binärzahlen erfolgen, entsteht eine logische Schaltung, wie sie Bild 8.2.23a) zeigt. Oft werden auch die drei Relationen $=$, $>$ und $<$ benötigt. Ihre logischen Relationen zeigt Bild 8.2.23b) als Wertetabelle und logische



A	B	A>B	A=B	A<B
0	0	0	1	0
0	1	0	0	1
1	0	1	0	0
1	1	0	1	0

Bild 8.2.23. Vergleicher

a) Mehrstufig nur bezüglich Gleichheit

b) Einstufig mit den Relationen $=$, $>$ und $<$ sowie zugehöriger Wertetabelle

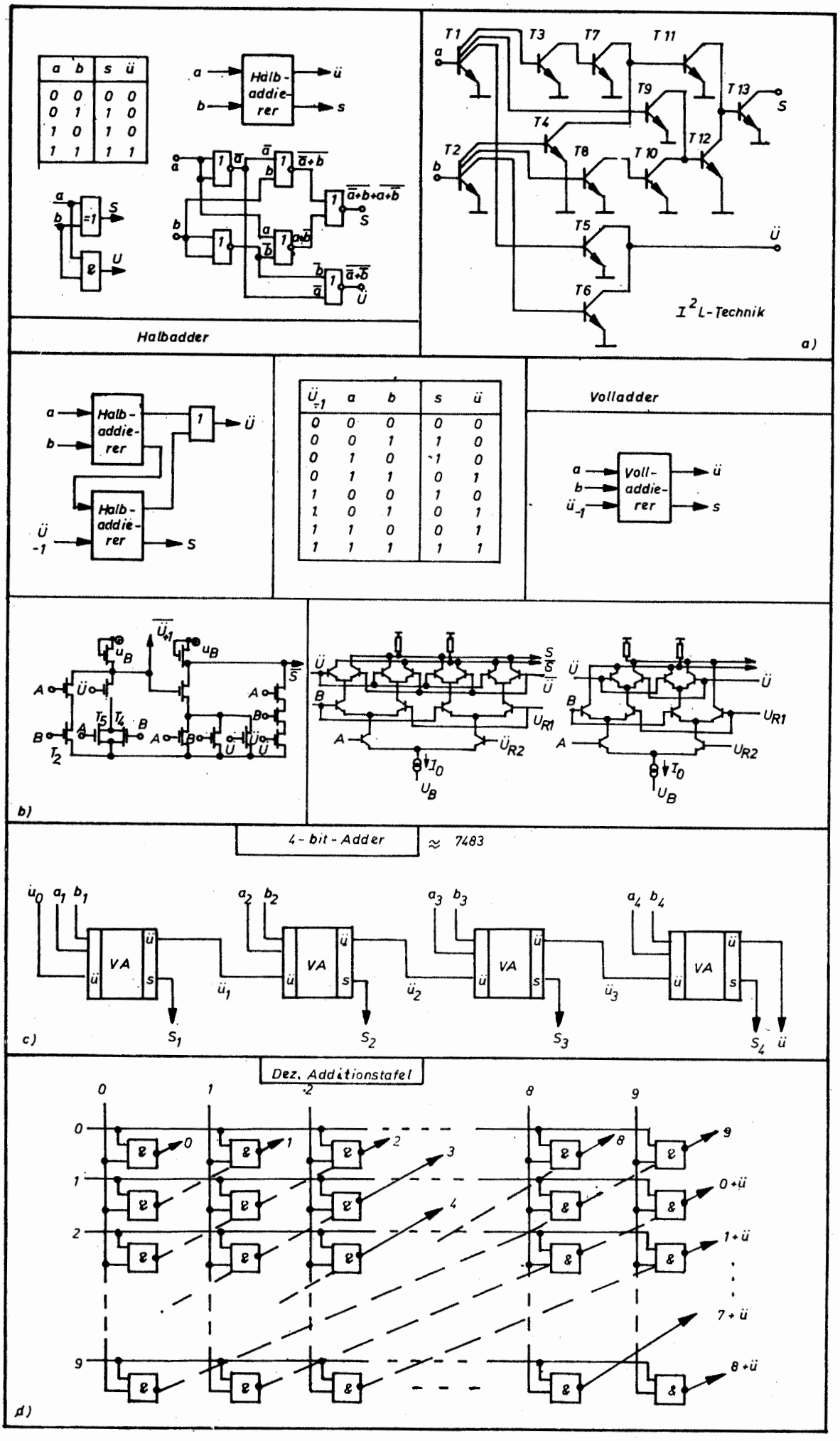
c) Mehrstufig infolge von Kaskadierung aus b) und Zusatzbauelementen für $=$, $>$ und $<$

Schaltung für einstellige Binärzahlen. Für mehrstellige Binärzahlen ist eine komplexe Kaskadierung notwendig, wie sie aus Teilbild c) für eine dreistellige Binärzahl dargestellt ist. Zunächst erfolgt der Vergleich mit einer Schaltung gemäß b) bei der höchstwertigen Binärziffer (MSB = most significant bit) A_3, B_3. Wenn hier $>$ oder $<$ auftritt, ist bereits die Entscheidung gefallen. Bei Gleichheit muß das nächstniedrigere bit herangezogen werden (A_2, B_2). Liegt nun hier $>$ oder $<$ vor, so ist die Entscheidung gefallen. Bei Gleichheit muß die nächstniedrigere Binärziffer zum Vergleich herangezogen werden usw.

Vergleicher werden oft auch Komparatoren genannt. Dies kann zu Verwechslungen mit den analogen Komparatoren führen (7.5.7.6.). Es sollte daher nur der Begriff digitaler bzw. binärer Komparator verwendet werden. Eventuell sind diese Zusätze auch beim Vergleicher sinnvoll.

8.2.8.5. Adder

Die Grundfunktion der arithmetischen Funktionen ist die Addition. Aus ihr lassen sich nämlich alle anderen Funktionen wie Subtraktion, Multiplikation usw. herleiten.


Halbadder
Halb-addierer
=1
Volladder
Voll-addierer
I²L-Technik
a)
b)
4-bit-Adder
≈ 7483
VA
c)
Dez. Additionstafel
d)

Das fundamentale Bauelement für die binäre Addition ist der Halbadder (Bild 8.2.24a)). An seinem Eingang sind zwei binäre einstellige Zahlen. Aus ihnen entsteht maximal eine zweistellige binäre Zahl. Das LSB ist die Summe s und das MSB ist der Übertrag $\ddot{u}$. So ergibt sich die Wertetabelle in Bild 8.2.24 oben. Beide Anteile können durch die logischen Funktionen

$$s = (\bar{a} \wedge b) \vee (a \wedge \bar{b}) \quad \text{— Antivalenz,}$$

$$\ddot{u} = x_1 \wedge x_2 \quad \text{— AND}$$

gebildet werden. Zuweilen ist es günstig, die Antivalenz unter Ausnutzung des Übertrags zu bilden. Dann gilt

$$s = (x_1 \vee x_2) \wedge \overline{\ddot{u}} .$$

Eine Zusammensetzung logischer Funktionen für beide Größen, die besonders für die I²L-Technik vorteilhaft ist, einschließlich der I²L-Schaltung, zeigt ebenfalls das Bild 8.2.24, wobei zur Vereinfachung die Injektionsquellen fortgelassen wurden.

Im allgemeinen Fall liegen bei einer mehrstelligen binären Zahl je bit folglich drei Größen vor: die beiden Eingangsbit und ein Übertrag von der niederwertigeren Bitstelle. Um diese Addition zu leisten, ist ein Volladder (Mitte in Bild 8.2.24) notwendig. Er läßt sich ersichtlich aus zwei Halbaddern und einem OR-Glied realisieren. Neben der Wertetabelle zeigt das Bild je eine Realisierung in MOS- und ECL-Technik.

Die Zusammenschaltung von vier Volladdern zu einem 4-bit-Adder zeigt Bild 8.2.24c). Sie hat einen prinzipiellen Nachteil: die einzelnen bit werden parallel und die Überträge sequentiell verarbeitet. Dies wird deutlich, wenn man beachtet, daß ein Übertrag aus der $a_1 b_1$-Stelle sich eventuell bis zur vierten Stelle fortpflanzen kann, und dies benötigt infolge der Verzögerungen in jedem Volladder auch die vierfache Zeit gegenüber der Bildung der einzelnen Teilsummen. Um das zu vermeiden, wurden verschiedene spezielle Additionstechniken entwickelt. Ein oft benutztes Prinzip ist die Look-ahead-carry-Methode. Bei ihr werden in einer einheitlichen, zusätzlichen Kombinationschaltung alle Summen und Überträge der einzelnen Volladder gleichzeitig und parallel zum Ergebnis aufbereitet. Der Aufwand ist natürlich höher als bei den zuvor geschilderten Prinzipien. Für spezielles Interesse sei u. a. auf [K 30, S 9, S 67, T 4] verwiesen.

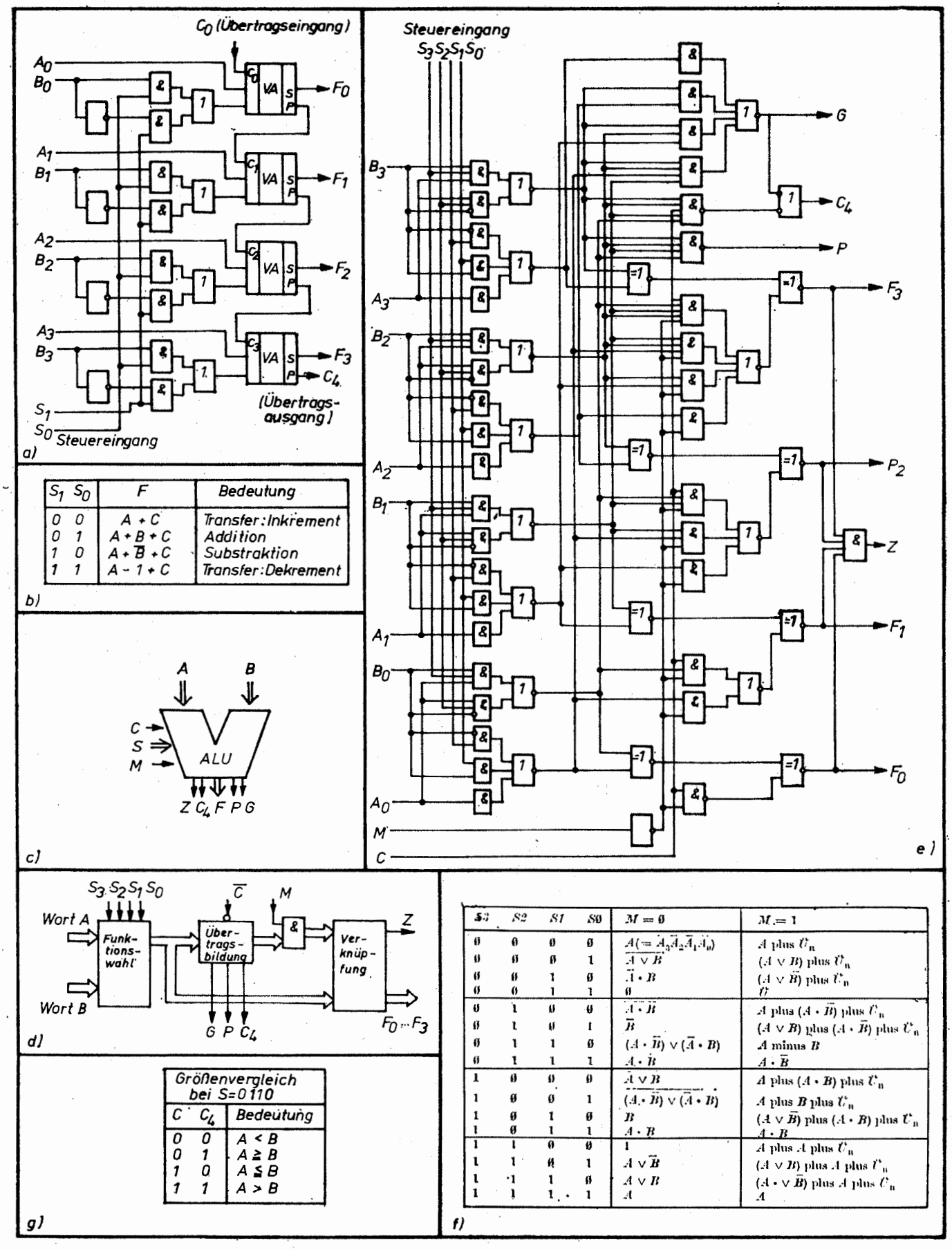
Einen völlig abweichenden Weg zeigt Bild 8.2.24d) für das Beispiel dezimaler Zahlen. In ein Bild übertragen, wird eine Matrix aufgestellt, deren Spalten bzw. Zeilen die Zahlen Ø bis 9 tragen. An jedem Kreuzungspunkt befindet sich ein AND-Glied und sein Ausgang zu mehreren OR-Schaltungen (diagonal gestrichelte Linien), wo die richtigen Ausgangswerte in Ausgängen Ø bis 9 und den Übertrag zusammengefaßt werden. Solche logisch gestalteten Additionstafeln sind extrem schnell.

Erwähnt sei noch, daß der Übertrag im Englischen Carry genannt und dann oft mit C bezeichnet wird.

◀

Bild 8.2.24. Adder

a) Halbadder: Wertetafel; Zusammensetzung aus Äquivalenz und AND; Umsetzung in I²L-Technik und Realisierung
b) Volladder: Aufbau aus zwei Halbaddern, Wertetafel, Symbol und Realisierungen in MOS- und ECL-Technik
c) 4-bit-Adder aus 4 Volladdern (VA)
d) Additionstafel für dezimale Zahlen



S_1	S_0	F	Bedeutung
0	0	$A + C$	Transfer: Inkrement
0	1	$A + B + C$	Addition
1	0	$A + \overline{B} + C$	Substraktion
1	1	$A - 1 + C$	Transfer: Dekrement

Größenvergleich bei S=0110		
C	C_4	Bedeutung
0	0	$A < B$
0	1	$A \geq B$
1	0	$A \leq B$
1	1	$A > B$

$S3$	$S2$	$S1$	$S0$	$M = 0$	$M = 1$
0	0	0	0	$A (= A_3\overline{A}_2\overline{A}_1A_0)$	A plus C_n
0	0	0	1	$\overline{A \vee B}$	$(A \vee B)$ plus C_n
0	0	1	0	$\overline{A} \cdot B$	$(A \vee \overline{B})$ plus C_n
0	0	1	1	0	C_n
0	1	0	0	$A \cdot \overline{B}$	A plus $(A \cdot \overline{B})$ plus C_n
0	1	0	1	$\overline{B}$	$(A \vee B)$ plus $(A \cdot \overline{B})$ plus C_n
0	1	1	0	$(A \cdot \overline{B}) \vee (\overline{A} \cdot B)$	A minus B
0	1	1	1	$A \cdot \overline{B}$	$A \cdot \overline{B}$
1	0	0	0	$\overline{A} \vee B$	A plus $(A \cdot B)$ plus C_n
1	0	0	1	$\overline{(A \cdot \overline{B}) \vee (\overline{A} \cdot B)}$	A plus B plus C_n
1	0	1	0	B	$(A \vee \overline{B})$ plus $(A \cdot B)$ plus C_n
1	0	1	1	$A \cdot B$	$A \cdot B$
1	1	0	0	1	A plus A plus C_n
1	1	0	1	$A \vee \overline{B}$	$(A \vee B)$ plus A plus C_n
1	1	1	0	$A \vee B$	$(A \vee \overline{B})$ plus A plus C_n
1	1	1	1	A	A

Bild 8.2.25. Arithmetisch-logische Einheit ALU

a) Vereinfachte ALU, welche 4 Funktionen (Inkrement, Addition, Substraktion und Dekrement) einschließlich des Übertrags C erfüllt

b) Steuerwort S_1S_0 für a) und dessen Bedeutung

c) Prinzip einer realen ALU 74181; К 155 ИП4

d) Blockschaltbild

e) Gatterbild

f) Funktionstabelle

g) Zusätzliche Größenvergleiche

8.2.8.6. Arithmetisch-logische Einheiten: ALU

Arithmetisch-logische Einheiten (arithmetic logic unit = ALU) sind kombinatorische Schaltungen, welche zwei Eingänge für binäre Datenworte und einen Steuereingang für Steuerworte besitzen. Das Ausgangswort ist je nach dem Steuerwort eine arithmetische oder logische Verknüpfung der beiden Eingangsworte. ALU sind also logische Schaltungen, deren logische Funktion durch ein Steuerwort aus einem bestimmten vorgegebenen Vorrat ausgewählt werden kann. Eine stark vereinfachte ALU zeigt Bild 8.2.25a). Sie besitzt die vier Grundfunktionen der Tabelle gemäß Bild 8.2.25b). Die Subtraktion erfolgt durch Komplementbildung. Von Inkrement spricht man, wenn eine binäre Größe um 1 ($C = 1$) erhöht wird; von Dekrement bei Erniedrigung um 1 ($C = Ø$). Im Prinzip kann der Übertrag C (Carry) vom Eingang her auch als zusätzliches Steuerbit betrachtet werden. Dann realisiert die einfache ALU von Bild 8.2.25a) sogar acht logische Funktionen.

Das Prinzip einer realen ALU 74181 (К 155 ИП 4) zeigt schematisch Bild 8.2.25c). Die Eingänge bedeuten:

A, B	4-bit-Eingangsworte
F	4-bit-Ausgangswort
S	4-bit-Steuerwort
C	Eingangsbit, z. B. Carry = Übertrag
G, P	bit für die Zusammenschaltung mehrerer ALU mit Look-ahead-carry (generate varry und propagate carry)
M	Steuerwort: $M = 1$ — arithmetische Funktion, $M = Ø$ — logische Funktion
Z	Anzeige Gleichheit, für $Z = 1$ gilt $A = B$
C_4	Ausgang: Übertrag: Carry

Das Blockschaltbild dieser ALU zeigt Bild 8.2.25d), das Gatterbild Teilbild e), während die möglichen Funktionen als Tabelle in Bild 8.2.25f) zusammengestellt sind. Die ALU ist in TTL-Technik ausgeführt, hat eine typische Übertragungsverzögerungszeit von 13 ns und eine Rechenzeit von 22 ns. Durch Kombination von Eingangsbit C und Ausgangsbit C_4 lassen sich bei $S = Ø11Ø$ (Subtraktion) auch die Relationen gemäß Bild 8.2.25g) unmittelbar feststellen.

8.2.8.7. Programmierbare Felder: PLA

Die programmierbaren Felder sind eine spezielle Art von Logikschaltungen. Sie sind im Prinzip universell bezüglich der logischen Funktionen. Im Gegensatz zur ALU wird die entsprechende Funktion aber nicht durch Steuersignale festgelegt, sondern durch eine strukturelle Veränderung bzw. Festlegung innerhalb der allgemein gültigen Möglichkeiten.

Die programmierbaren Felder sind seit 1973 in der Literatur gut bekannt, wurden jedoch bisher relativ wenig benutzt. Generell wird jetzt allgemein (nach Abklärung von Grenzen und Möglichkeiten der Mikroprozessoren) eine verstärkte Anwendung erwartet. Für sie existieren eine Vielzahl unterschiedlicher Abkürzungen. Hier seien nur einige Beispiele genannt:

PLA	— programmable logic array
PAL	— programmable array logic
FPLA	— field programmable logic array
FPL	— field programmable logic
IFL	— integrated fuse logic
FPLS	— field programmable logic sequencer
FPRP	— field programmable ROM patch.

Das Programmierbare soll hier zunächst formal betrachtet werden. Es entspricht den Prinzipien der programmierbaren Speicher (8.3.5.). Dementsprechend werden auch hier die beiden Hauptfälle unterschieden:

- Programmierung vom *Hersteller*, z.B. durch spezielle Masken (ROM),
- Programmierung vom *Anwender*, z.B. elektrisch (PROM).

Die Programmierung vom Anwender wird hier mit dem englischen Begriff field bezeichnet, während array im Deutschen durch Feld (besser wäre Matrix) umschrieben wird.

Theoretisch kann eine PLA immer gemäß Tab. 8.2.11 über die Minterme bzw. Maxterme, d. h. über die Normalformen DNF bzw. KNF, beschrieben werden. In der Praxis wird dies so realisiert, daß die Spalten und Zeilen der Normalform durch je eine Matrix gegeben sind, in der die Knotenpunkte je nach der zu realisierenden logischen Funktion gesetzt oder unbesetzt sind. Zwei Prinzipbeispiele solcher PLA-Strukturen zeigt Bild 8.2.26: einmal mittels Dioden und einmal mehr formal. Von

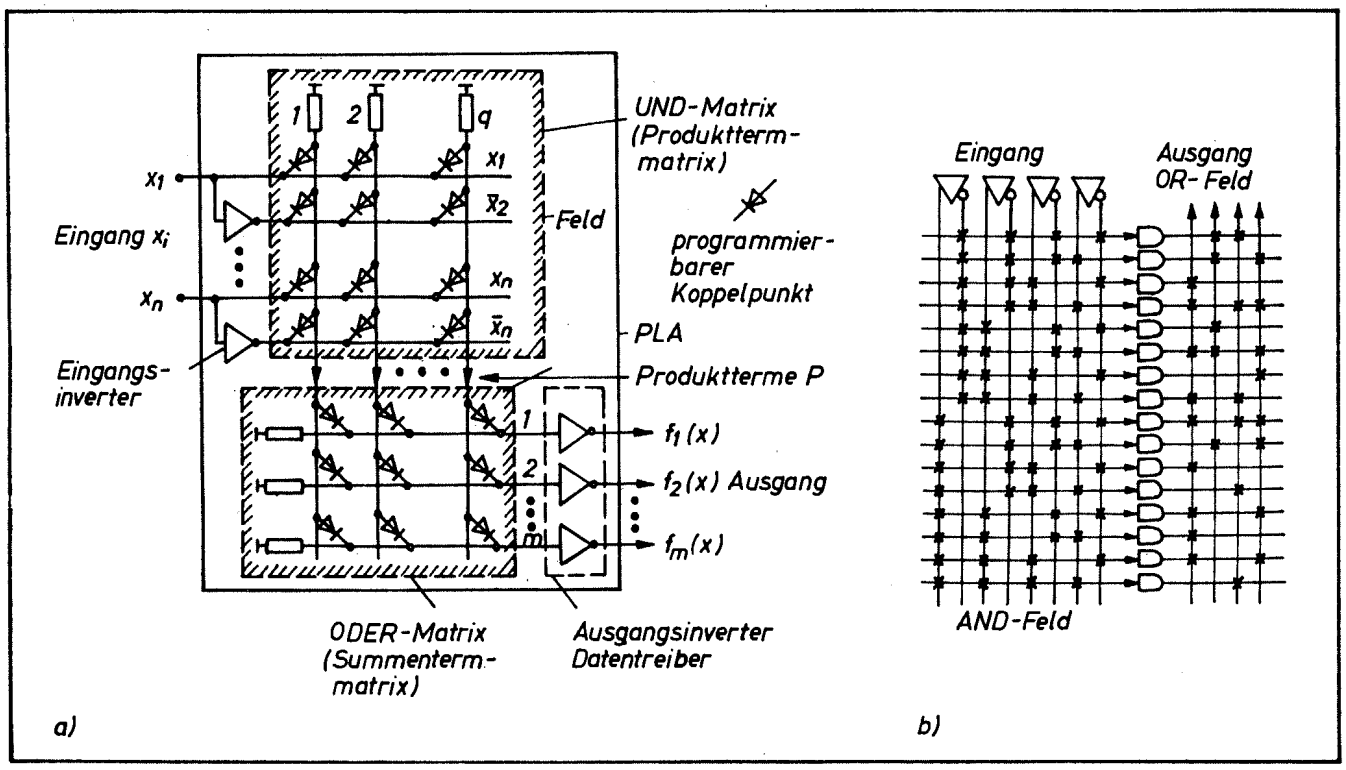


Bild 8.2.26. Zur PLA

a) Prinzip auf der Basis von Dioden, die vorhanden oder nicht vorhanden sind

b) Formales Bild der beiden Matrizen, das deutlicher die Art der zu wählenden programmierenden Punkte berücksichtigt

den n Eingangsgrößen werden zunächst die negierten zusätzlich bereitgestellt. Die Größen werden in der UND-Matrix über die setzbaren Knotenpunkte im Sinne der AND-Relationen auf q Leitungen verknüpft. So entstehen die Minterme (Produktterme). Sie können wiederum im OR-Feld über zu setzende Knotenpunkte auf m Ausgänge im Sinne von je einer Oder-Funktion verbunden werden. Damit stellt eine PLA mit n Eingängen und m Ausgängen m logische Schaltungen mit je n Eingängen dar (vgl. Bild 8.2.2b)). Die mögliche Anzahl p der Minterme legt eine gewisse Begrenzung der m Schaltungen fest. Sie ist aber in der Regel so groß, daß sie keine wesentlichen Einschränkungen bewirkt.

Neben der beschriebenen AND-OR-Form gibt es auch den Matrizenaufbau als OR-AND-Form, oder was technologisch besonders vorteilhaft ist, als NOR-NOR-Form.

Eine praxisnahe (wenn auch nicht als PLA) Verwendung der Normalform geht bereits aus dem Bild 8.2.21 hervor. Hier wird in Teilbild b) die TTL-Technik als AND-NOR entsprechend angewendet und in Teilbild c) die I^2L-Technik als NOR-NOR-Feld.

8.2.9. Ergänzungen und Vergleiche

8.2.9.1. Kontaktprellen

Störungen bei logischen Schaltungen treten in der verschiedensten Weise auf. Auf Probleme der Leitungsreflexionen, der Stromaufnahmespitzen beim Schalten von TTL, Impulskopplungen zwischen den Leitungen usw. wurde bereits im Abschnitt 8.2.2.3. hingewiesen. Hier seien noch zwei allgemeine Probleme behandelt: das Entprellen von Kontakten und das Auftreten von Hazards.

Zum Steuern von Schaltungen werden in der Regel Kontakte von Schaltern, Tasten usw. benutzt. Sie gehen aber, von Ausnahmen abgesehen, nie schlagartig in den anderen, z.B. offenen oder geschlossenen Zustand über. Im Bereich von ms findet vielmehr ein Prellen statt. In quasiperiodischer Folge öffnet bzw. schließt sich der Kontakt mehrfach. Eine vom Kontakt gesteuerte Schaltung soll dagegen aber nur einmal von dem einen in den anderen Zustand übergehen. Dies bedeutet, daß das Kontaktprellen unterdrückt werden muß, und zwar mittels spezieller Schaltung. Dies

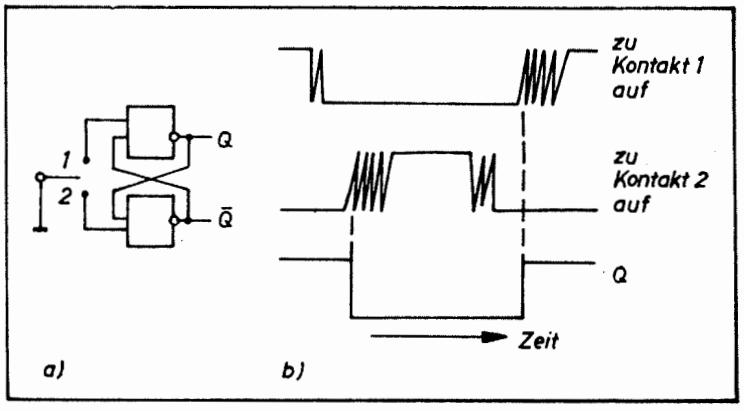


Bild 8.2.27. Zum Prinzip des Entprellens von Kontakten
a) Eine oft verwendete Schaltung bei Umschaltkontakten
b) Prinzip des Kontaktprellens und Wirkung der Schaltung a)

sei am Beispiel von Bild 8.2.27 gezeigt. Der jeweils zeitlich auftretende erste Kontaktschluß bewirkt, daß das Flipflop (vgl. 8.5.2.) in die entsprechende Stellung geschaltet wird. Die folgenden Kontaktöffnungen und -gaben desselben Kontaktes haben dann keinen Einfluß mehr. Erst wenn der zweite Kontakt das erste Mal schließt, schaltet das Flipflop zurück, und weiteres Öffnen und Schließen auch dieses Kontaktes bleibt ohne Wirkung. Sobald wieder der andere Kontakt das erste Mal schließt, schaltet das Flipflop um. Das geschilderte Prinzip gilt nur für Wechselkontakte.

Für einfache Kontakte ist eine Entprellung erheblich schwieriger. Hier werden z.B. Monoflops (7.3.8.2.) verwendet. Sie wirken über die Zeitdauer ihrer Verzögerung. Bei Rechnern erfolgt in der Regel eine zweimalige Abtastung der Kontakte in einem Zeitabstand von einigen ms. Nur wenn zu beiden Zeiten der Kontakt geschlossen war, wird er als geschlossen weiter interpretiert.

8.2.9.2. Hazards

Der Begriff Hazard ist englisch und bedeutet soviel wie Wagnis, Zufall, Gefahr. Bei den kombinatorischen Schaltungen sind Hazards Fehler, d. h. L- oder H-Signale, die bei einwandfreiem Funktionieren der Schaltung nicht auftreten dürften. Sie werden durch die Laufzeiten bzw. durch Laufzeitdifferenzen von Signalen durch mehrere Gatter hindurch hervorgerufen, insbesondere dann, wenn Signale auf unterschiedlichen Wegen über unterschiedlich viele Gatter wieder zusammentreffen.

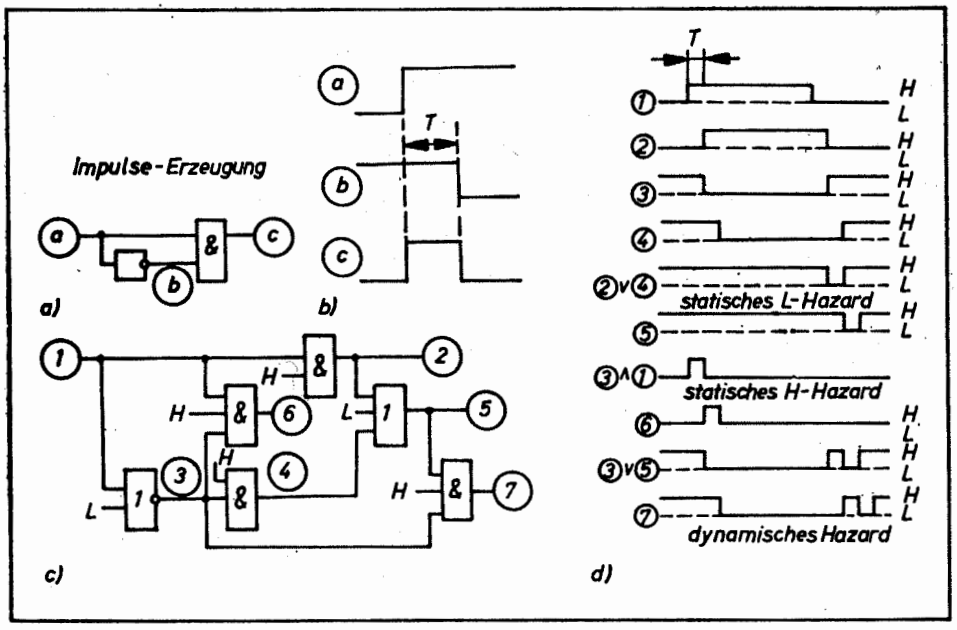


Bild 8.2.28. Zur Entstehung von Hazards

a) Schaltbild einer einfachen Impulserzeugung aus einer LH-Flanke
b) Signalbilder zu a)
c) Schaltbild zur Entstehung der verschiedenen Formen von Hazards
d) Impulsbilder zu c)

Eine übersichtliche Erklärung eines speziellen H-Hazards geht aus Bild 8.2.28a) und b) hervor. Da hier am AND-Glied immer zugleich das normale und das negierte Signal anliegt, sollte gemäß den Regeln der Schaltalgebra am Ausgang stets ein L vorhanden sein (positive Logik). Nun verzögert aber das negierende Glied um die Zeit T, und um soviel verspätet tritt folglich an seinem Ausgang auch das negierte Signal b auf. Es existiert also genau während dieser Zeit an beiden Eingängen des AND-Gliedes ein H-Signal. So entsteht das unzulässige, kurzzeitige H-Signal am Ausgang c. Es sei hier nur nebenbei erwähnt, daß dieses spezielle, sonst immer störende Signal bei einigen Flipflops (8.5.2.) gezielt genutzt wird.

Bild 8.2.28c) zeigt eine Schaltung und Teilbild d) ihre Signalverläufe, in denen alle *drei möglichen Arten* von Hazards auftreten:

— statisches L-Hazard. Das Ausgangssignal von zwei Eingangssignalen müßte eigentlich ständig H sein. Dadurch, daß eins der beiden Signale mehr verzögert ist (4 gegenüber 2), entsteht ein L-Hazard. L-Hazards entstehen an OR- bzw. NOR-Schaltungen, nicht an AND- und NAND-Schaltungen.

— statisches H-Hazard. Bei statischen H-Hazard sollte das Ausgangssignal eigentlich immer L sein. Jedoch durch unterschiedliche Laufzeiten (3 und 1) über zwei und mehr Wege entsteht an einer AND- bzw. NAND-Schaltung ein H-Hazard. H-Hazards entstehen nicht an OR- bzw. NOR-Schaltungen.

— dynamisches Hazard. Dynamische Hazards entstehen, wenn eigentlich ein Übergang von H nach L bzw. von L nach H erfolgen sollte. Auch hier sind verschieden lange Wege Voraussetzung. Aber zusätzlich muß die Schaltung statische H- und L-Hazards besitzen. (Im Bild 8.2.28 L-Hazard an 5 und L-Hazard von 3 nach 7 mit 3, 4, 5 nach 7). Eine Schaltung, die nur H- oder L-Hazards enthält, zeigt keine dynamischen Hazards.

Bisher wurden nur Hazards erwähnt, die in rein kombinatorischen Schaltungen entstehen können. Sie sind durch einen entsprechenden Entwurf prinzipiell vermeidbar. Es gibt aber auch Hazards, die durch Rückführungen hervorgerufen werden. Sie sind prinzipiell nicht vermeidbar (vgl. z. B. [G 6]).

8.2.9.3. Entwicklungsgeschichte und Vergleich

Bezüglich der im folgenden zu behandelnden Daten schwanken die Angaben in der Literatur beträchtlich. In der ersten Gruppe ist der geschichtliche Verlauf der Techniken behandelt. In diese Zahlen gehen Probleme der vollständigen Erfassung der Produktion bzw. des Verbrauchs, des Unterschieds zwischen beiden, des Geldwertes usw. ein. Aus mehreren solchen Angaben wurde in Bild 8.2.29a) versucht zu zeigen, welchen Anteil die einzelnen Schaltkreisfamilien im Laufe der Zeit hatten und wie sie sich voraussichtlich künftig weiter entwickeln. Daraus ist zu erkennen, wie z. B. DTL als erste Schaltkreisfamilie bereits um 1977 ausstarb. Für die klassischen TTL (d. h. für die erste Generation) wird dies um 1984 erwartet. Anderseits ist das Entstehen neuer Familien zu erkennen: p-MOS ca. 1970, CMOS und n-MOS ca. 1973, SCHOTTKY-TTL um 1975, HMOS als neue n-MOS-Technik um 1978.

Im allgemeinen wird die Komplexität von integrierten Schaltkreisen nach der Anzahl der Transistoren bzw. Gatter einer Schaltung eingeteilt. In etwa gelten dabei die folgenden Größenordnungen:

SSI — small scale integration:	$<$ 10 Gates,
	$<$ 100 Bauelemente
MSI — medium scale integration:	10 ... 100 gates,
	10^2 ... 10^3 Bauelemente
LSI — large scale integration:	100 ... 1000 gates,
	10^3 ... 10^4 Bauelemente
VLSI — very large scale integration:	$>$ 1000 gates,
	$>10^4$ Bauelemente

Die entsprechenden Anteile dieser Schaltungen am Gesamtwert enthält, getrennt nach bipolaren und MOS-Schaltungen, Bild 8.2.29b) bzw. c).

Die Packungsdichte von Schaltungen kann nach verschiedenen Gesichtspunkten beschrieben werden: einmal als Gatter je mm², zum anderen als Transistoren je Chip. Den Zusammenhang mit der Verzögerungszeit zeigt Bild 8.2.29d) bzw. e).

In Bild 8.2.29f) ist schließlich der Zusammenhang von Verzögerungszeit und Verlustleistung dargestellt. Hier wurden auch zwei künftige Technologien, nämlich GaAs- und Tieftemperaturbauelemente (JOSEPHSON) einbezogen. Detaillierte Diagramme hierzu enthalten die Bilder 8.2.13 bez. TTL; 8.2.17 bez. I²L und 8.2.19 bezüglich der Frequenzabhängigkeit des Schaltzeit-Leistungs-Produktes. Tab. 8.2.15 ergänzt diese Daten durch mehrere Parameter.

8.3. Speicherschaltungen

8.3.1. Allgemeine Grundlagen

Das Speichern existiert in vielen Formen, z. B. Speichern von Stoff, Energie oder Information. Das Speichern von Information ist dabei von besonders fundamentaler Bedeutung, und es können genetische, neurologische, kollektive und technische Speicher unterschieden werden [V 12]. Selbst die technische Speicherung von Information existiert in sehr vielen Formen, es seien hier nur

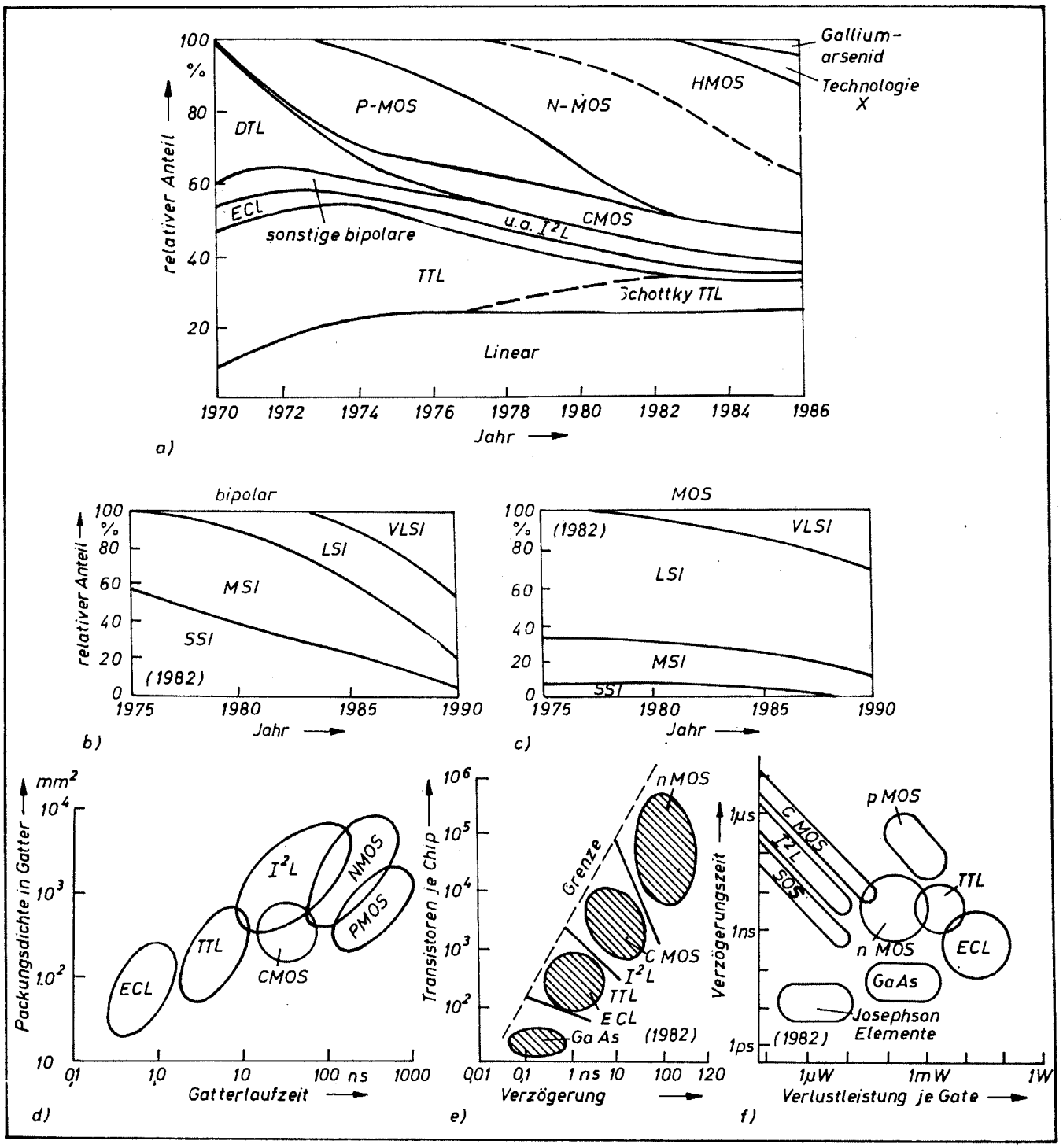


Bild 8.2.29. Entwicklung und Vergleich wichtiger logischer Familien

a) Relativer Anteil am Weltumsatz
b) Einteilung nach Integrationsgrad bezüglich bipolarer Schaltkreise,
c) für MOS-Schaltkreise
d) Zusammenhang von Packungsdichte und Gatterlaufzeit
e) Transistoren je Chip und Verzögerungszeit
f) Verzögerungszeit und Verlustleistung

beispielhaft Schrift, Buchdruck, Film, Schallplatte, Magnetband und elektronische Speicher, u. a. auf der Basis von Halbleitern, erwähnt. Diese Fülle und schließlich auch die Vielfalt der elektronischen Speicher, die hier noch genauer zu behandeln ist, machen eine Systematik der Prinzipien der Speicherung sehr schwer. Und obwohl allgemein akzeptiert wird, daß die Speicher, z. B. in der Rechentechnik, die Leistungsfähigkeit wesentlich sowohl technisch als auch ökonomisch begrenzen, gibt es bis heute bestenfalls Ansätze zu einer Theorie der Speicherung. Deshalb können hier nur gewisse Grundlagen dargestellt werden.

Tabelle 8.2.15. Überblick über die Eigenschaften wichtiger Integrationstechniken

Integrationstechnik	*p*-MOS	*n*-MOS			CMOS		TTL LS	ECL			I^2L		
		1976	1979	1984	Silicium	ESFI/SOS		1975	1978	1982	1975	1978	1982
Packungsdichte in Gatter pro mm²	40	40	150	400	40	80	30	40	100	200	100	150	400
Gatterlaufzeit in ns	80	15	5	0,5	20	4	4	2	0,5	0,1	20	5	2
Schaltzeitleistungs-Produkt in pWs	200	6	0,2	0,005	3	0,4	4	10	1	0,1	0,2	0,025	0,002
Statische Verlustleistung in mW	2	0,2			10 nW	100 nW	1	5 ... 30			0,2		
Betriebsspannung in V	10 ... 30	5 ... 15			3 ... 15	1 ... 15	5	5,2			1 ... 5		
Fan-out	20	25			50 ... 100		10 ... 75	25			1		
Anzahl der Maskenschritte	5	6			7		7	8 ... 9			5 ... 7		
Anzahl der Dotierungsschritte	2	3			4		4	4 ... 5			3 ... 4		
Bauelemente für 2 Gatter	3	3			4		10 ...15	10 ...12			3 ... 4		
Relativer Aufwand	1	1,5			5	7	2	2			1,5		
Kompatibilität der Peripherieschaltungen	schlecht	mäßig			mäßig		sehr gut	sehr gut			gut		

8.3.1.1. Speicherfunktionen

Die Grundfunktionen der Speicherung gehen aus Bild 8.3.1 a) hervor. Die aufzeichnenden Signale $f_A(t)$ sind zeitabhängig und damit in der Zeit vergänglich. Für viele Anwendungen werden sie jedoch zu späteren, frei wählbaren Zeitpunkten benötigt. Als Beispiel diene die Rede einer Persönlichkeit. Damit sie aufbewahrt werden kann, muß sie an eine zeitunabhängige Struktur gebunden werden. Dies ist allgemein eine Funktion des Ortes $f(z)$. Sie sei als Weg, z. B. eines Magnetbandes, gegeben. Dann ist mit einem Startsignal der Aufzeichnungsvorgang auszulösen, und die Schallschwingungen werden in Magnetisierungsstrukturen auf dem Magnetband gewandelt. Dies ist der Speicherzustand $f(z)$. Soll die Rede nun zu irgendeinem gewünschten Zeitpunkt wiedergegeben werden, so ist das Magnetband erneut an der richtigen Stelle zu starten und der Wiedergabevorgang auszulösen. Dabei entsteht e n Signal $f_W(t)$, welches dem Aufzeichnungssignal $f_A(t)$ möglichst ähnlich sein soll.

Die Verhältnisse sind völlig anders, wenn eine Schallplatte mit der Rede vorliegt. Dann existiert nur der Speicherzustand $f(z)$ und seine Wiedergabe einschließlich des Starts. Es interessiert dagegen

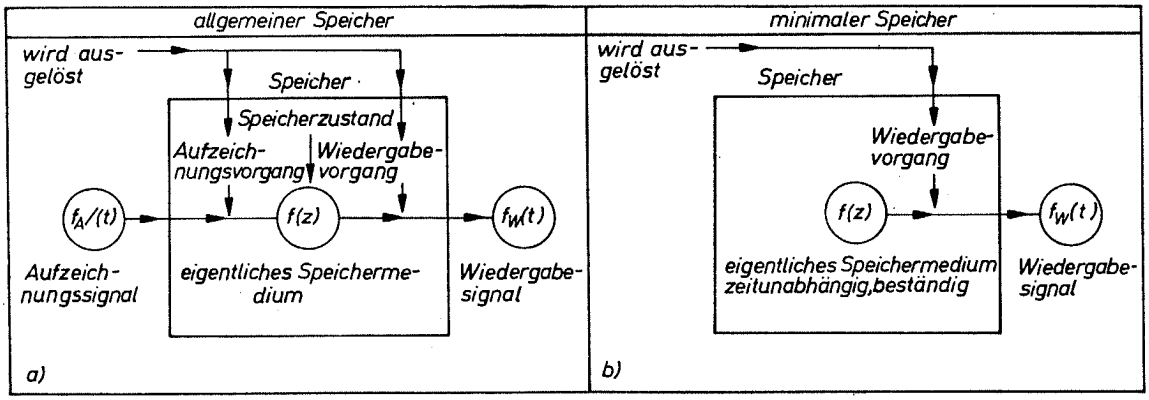


Bild 8.3.1. Speicherprinzipien in schematischer Darstellung

a) Allgemeiner Speicher

b) Minimaler Speicher

in der Regel nicht der Aufzeichnungsvorgang. Er ist ein Produktionsprozeß, der für den Anwender der Schallplatte im Prinzip belanglos ist. Solche stark vereinfachten Speicher sind Minimalspeicher (Bild 8.3.1 b)). Zwischen ihnen und den allgemeinen Speichern in Teilbild a) gibt es vielfältige Übergänge.

Bei digitalen Signalen muß der Speicherzustand einmal die einzelnen bit als Ø bzw. 1 oder als L bzw. H enthalten. Dies kann z. B. auf einen Stromfluß, auf die Ladung eines Kondensators, auf die Polung eines Magneten oder auf den Zustand eines Schalters bezogen werden. Jedes Element, das zwei unterschiedliche Zustände besitzen kann und langfristig behält, ist als Speicherzelle geeignet. Mit einer Speicherzelle ist aber auch nur ein Signalzustand festzuhalten. In der Regel werden aber viele binäre Signale notwendig, so daß auch viele Speicherzellen erforderlich sind. Doch ihr Vorhandensein allein ist nicht ausreichend. So wie auf dem Magnetband oder in der Schallplatte die Kenntnis des Ortes für eine bestimmte Textstelle notwendig ist, so müssen die einzelnen Signale genau definiert bestimmten Speicherzellen zugeordnet werden. Zu diesem Zweck werden die Speicherzellen selbst mit bestimmten Zahlen, den Adressen, gekennzeichnet. Das Signal X liegt dann in der Speicherzelle Y und kann hier zu gegebener Zeit wiedergefunden werden.

Eine Speicherzelle ist durch folgende Eigenschaften gekennzeichnet:

— Sie muß zwei Zustände langfristig annehmen können, die mit L und H gekennzeichnet seien (Speicherzustände).
— Sie muß mittels einer Adresse (binäres Signal) aus einer Fülle anderer Speicherzellen auswählbar sein (Ort z).
— Sie muß nach der Auswahl definiert in den Zustand L oder H setzbar sein (Aufzeichnungsvorgang).
— Sie muß nach der Auswahl imstande sein, ihren Zustand nach außen als Signal abzugeben (Wiedergabevorgang).

Im Falle des Minimalspeichers gemäß Bild 8.3.1 b) ist der Aufzeichnungsprozeß nicht notwendig. Im digitalen Bereich heißen derartig vereinfachte Speicher ROM (read only memory). Bei ihnen ist einer der beiden Zustände L oder H ein für alle Mal festgelegt, und mit der Auswahl der Speicherzelle erfolgt sofort die Wiedergabe des Signals.

8.3.1.2. Eigenschaften von Speichern

Die *Hauptkenndaten* eines Speichers sind

- *Kapazität*: die Anzahl der Speicherzellen = den maximal zu speichernden bit. Hierbei ist zu beachten, daß in Kbit bzw. Kbyte das K $2^{10} = 1024$ bedeutet und groß geschrieben wird. Bei Mbit bzw. Mbyte ist dies nicht einheitlich geregelt: Es kann das M $2^{20} = 1048576$ oder 10^6 bedeuten.
- *Zugriffszeit*: Sie ist ein Mittelwert, welcher die Zeit angibt, die vergeht von dem Moment an, wo die Adresse einer Speicherzelle bekannt ist, bis zu jenem, wo der Zustand der Speicherzelle als Signal vorliegt.
- *Volumen*: Es wird mit m^3 angegeben und enthält in der Regel nicht nur die Speicherzellen, sondern auch alle technischen Einrichtungen, die zum Betrieb des gesamten Speichers notwendig sind, also z. B. Stromversorgungen, Aufzeichnungs-, Wiedergabe- und Adressierungseinrichtungen.
- *Preis* des Speichers: z. T. als Anschaffungspreis, z. T. unter Einbeziehung der Betriebskosten, z. T. aber auch nur auf die Speicherzellen, also auf das Speichermedium, bezogen.

Ein *idealer Speicher* sollte folgende Grenzwerte anstreben:

Kapazität	$\to \infty$
Zugriffszeit	$\to 0$
Volumen	$\to 0$
Preis	$\to 0$.

Dies ist jedoch weder praktisch möglich noch sinnvoll. Je größer die Kapazität z.B. ist, um so umfangreicher wird die Adressenrechnung, um die gewünschte Speicherzelle zu finden. Weiter ist mit allen Signalen bereits rein physikalisch eine bestimmte Energie verbunden. Sie muß als Wärme abgeführt werden. Hierzu ist eine bestimmte Oberfläche des Speichers notwendig. Aus diesen und vielen weiteren Gründen haben sich eine Vielzahl von Speichern mit unterschiedlicher Kapazität und Zugriffszeit entwickelt, die in dem Rechner als *Speicherhierarchie* arbeiten: Der größte Speicher ist am langsamsten. Er besitzt als Archivspeicher alle benötigten Informationen. Ein schnellerer Zwischenspeicher ist kleiner und übernimmt aus dem großen nur jene Informationen, die in absehbarer Zeit benötigt werden. Schließlich existiert ein noch kleinerer, aber sehr schneller Arbeitsspeicher, der die gerade für den Rechner aktuellen Informationen zu enthalten hat.

Ein Strukturgebilde, das ein bit Information trägt, heißt Speicherzelle. Es gibt eine Vielzahl von *Speicherzellen*, die auf unterschiedlichen Technologien und Organisationsprinzipien beruhen. Sie können *eingeteilt werden* in

bipolare Speicherzellen (8.3.3.)
statische MOS-Speicherzellen (8.3.4.1.)
dynamische MOS-Speicherzellen (8.3.4.5.)
PROM-Speicherzellen (8.3.5.3.)
ROM-Speicherzellen (8.3.5.2.).

Sie werden in den genannten Abschnitten genauer erklärt und behandelt. Darüber hinaus gibt es noch *spezielle Speichertypen* mit anderem Verhalten. Dies sind einmal die

Umlaufspeicher, insbesondere CCD und Bubbles (8.3.6.)

sowie die

Assoziativspeicher (8.3.7.).

Tabelle 8.3.1. Ungefähre Hauptdaten wichtiger Datenspeicher

Typ	Halbleiter-speicher	Floppy disc	Kassetten-speicher	Platten-speicher	Band-speicher	Bandbiblio-theken
Kapazität/bit	ca. 1 Mbit	mehrere 100 Kbit	mehrere Mbit	um Gbit	je Spule einige 10 Mbit	Terabit
Zugriffszeit	10 ns bis 1 µs	um 100 ms	Minuten	10 bis 100 ms	mehrere Sekunden	Sekunden

Damit sind also vor allem Kapazität und Zugriffszeit die wichtigsten Parameter. Einen Überblick hierzu gibt Tab. 8.3.1 und Bild 8.3.2., wobei die Halbleiterspeicher noch global zusammengefaßt werden und im Teilbild a) auch die genetischen, neuronalen und kollektiven Speicher berücksichtigt sind, während Teilbild b) eine Unterteilung der Halbleiter versucht. Sie ist immer etwas unsicher, da hier z. Z. eine extrem schnelle Entwicklung erfolgt. Unter anderem deshalb wurde die Jahreszahl bei den Bildern als Richtwert für die ungefähr richtige Einschätzung hinzugefügt. In den Bildern 8.3.2c) ist noch der Zusammenhang von Bitzellengröße und Zugriffszeit für verschiedene Technologien und in d) der Zusammenhang von Zugriffszeit und Kapazität dargestellt.

Um den zeitlichen Einfluß in Vergangenheit und Zukunft bezüglich verschiedener Speicherdaten deutlich zu machen, wurde Bild 8.3.3 aus einer Vielzahl von verschiedenen Daten zusammengestellt und ausgewählt. Teilbild a) zeigt, wie im Laufe der Zeit die immer größer werdende Kapazität je Speicherchip, also auf ein bit bezogen, die billigste ist. Daraus ist auch die allmähliche Ablösung des Kernspeichers im Arbeitsspeicher von Rechnern zu erklären. Sie wurde in Teilbild i)

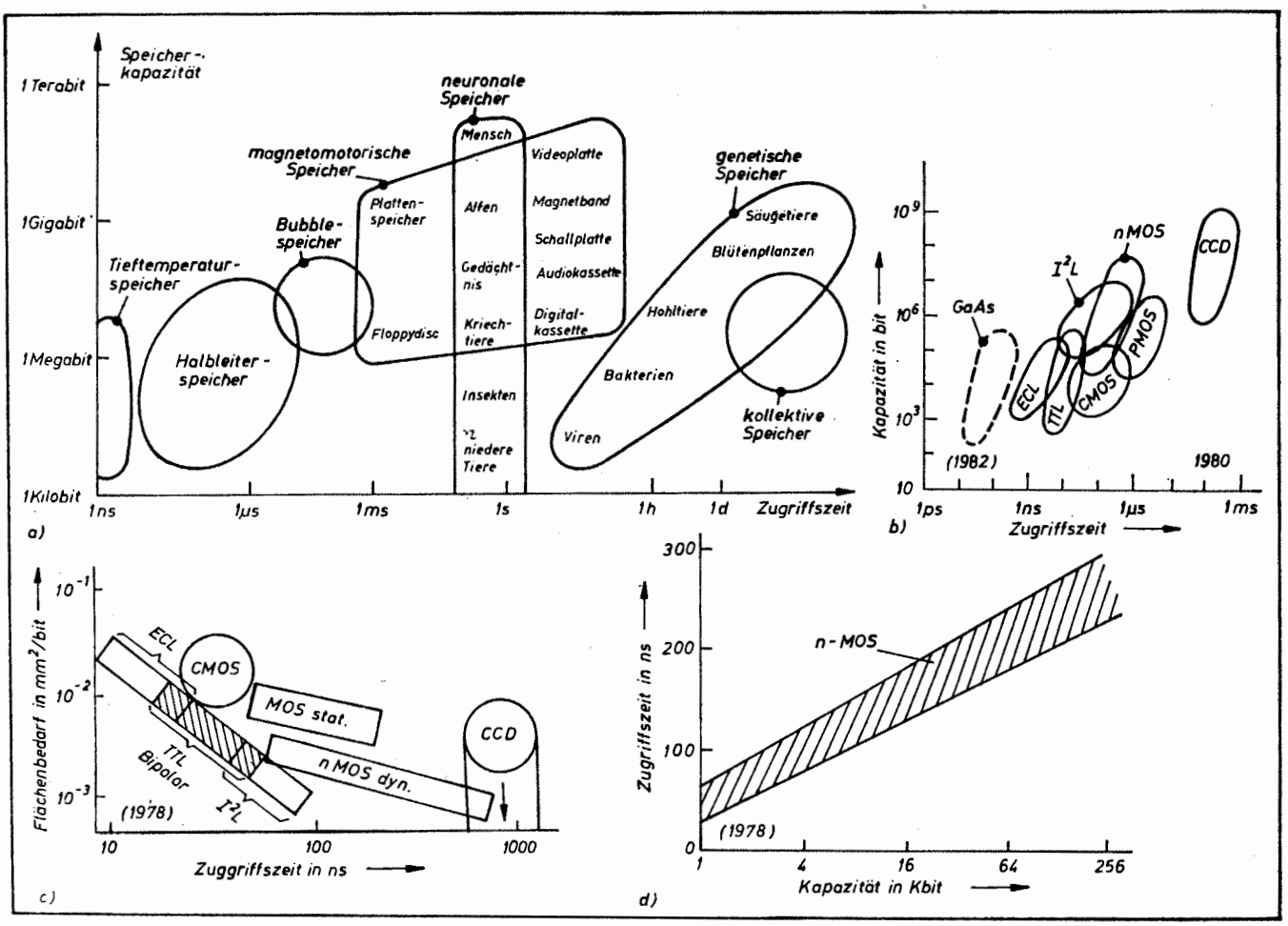


Bild 8.3.2. Speicher bezüglich Zugriffszeit und Kapazität

a) Unter Einbeziehung genetischer, neuronaler und kollektiver Speicher
b) Versuch einer Trennung der verschiedenen Halbleiterspeicher
c) Zusammenhang von Speicherzellengröße und Zugriffszeit
d) Zugriffszeit und Kapazität bei *n*-MOS-Speichern

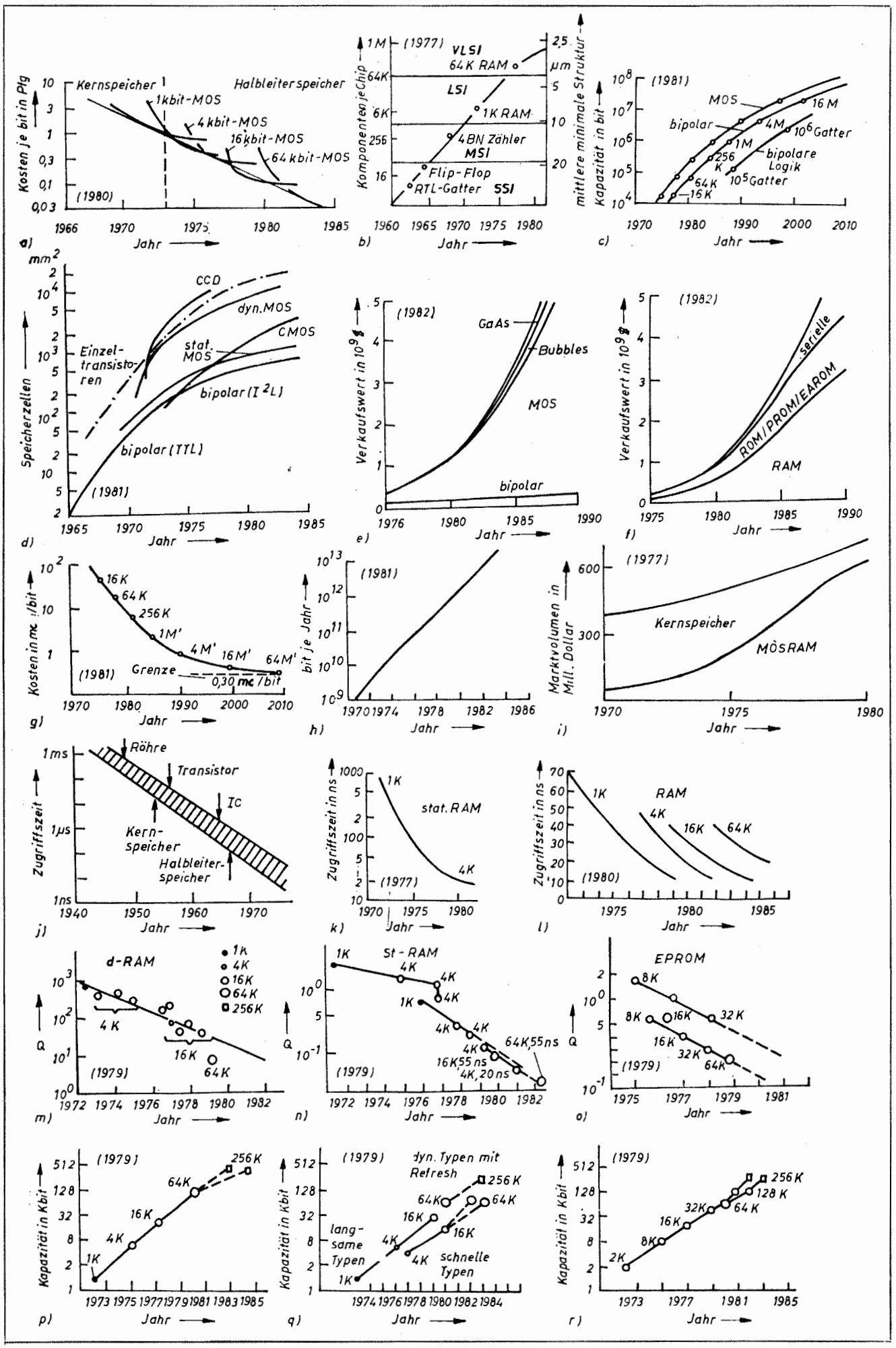
durch die weltweiten Umsatzziffern für die Jahre 1970 bis 1980 deutlich dargestellt, während Teilbild g) den voraussichtlichen Trend bis zum Jahre 2010 prognostiziert, wobei ein asymptotischer Grenzwert von ca. 0,3 mc/bit erreicht wird.

Teilbild b) zeigt auf, wie generell die Anzahl der Komponenten in Schaltkreisen zunimmt und die minimalen Linienbreiten notwendig abnehmen. Es liefert also den Bezug zu generellen Größen der Halbleitertechnologie. Teilbild c) gibt eine Prognose wieder, wann langfristig welche Speicher-

▶

Bild 8.3.3. Speicherdatenentwicklung und -prognosen

a) Kosten als Funktion der Kapazität und der Zeit
b) Komponenten je Chip und mittlere minimale Struktur
c) Voraussichtlich zu erwartende Einführung verschiedener Kapazitäten
d) Größe der Bitzelle bei verschiedenen Speichertypen
e) Weltweiter Speicherumsatz nach Technologien
f) Weltweiter Speicherumsatz nach Speicherarten
g) Voraussichtliche Entwicklung der Bitkosten (vgl. a))
h) Benötigte totale Kapazitäten bei Halbleiterspeichern
i) Ablösung der Kernspeicher durch Halbleiterspeicher bei den Arbeitsspeichern
j) Reduzierung der Zugriffszeiten über einen langen Zeitraum
k) Abnahme der Zugriffszeit bei mittleren Halbleiterspeichern
l) Voraussichtliche Entwicklung der Zugriffszeiten
m), n), o) Entwicklung der Güte Q (siehe Text) bei dynamischen, statischen RAM und EPROM,
p), q), r) bezüglich der Kapazität (vgl. c))


a)
Kosten je bit in Pfg
Kernspeicher
Halbleiterspeicher
1kbit-MOS
4kbit-MOS
16kbit-MOS
64kbit-MOS
(1980)
Jahr
b)
Komponenten je Chip
(1977)
VLSI
64K RAM
LSI
1K RAM
4BN Zähler
MSI
Flip-Flop
RTL-Gatter
SSI
mittlere minimale Strukturbreite in µm
Jahr
c)
Kapazität in bit
(1981)
MOS
bipolar
16M
4M
1M
256K
64K
16K
10⁶ Gatter
10⁵ Gatter
bipolare Logik
Jahr
d)
Speicherzellen in mm²
CCD
dyn. MOS
CMOS
stat. MOS
Einzeltransistoren
bipolar (I²L)
bipolar (TTL)
(1981)
Jahr
e)
Verkaufswert in 10⁹ $
(1982)
GaAs
Bubbles
MOS
bipolar
Jahr
f)
Verkaufswert in 10⁹ $
(1982)
serielle
ROM/PROM/EAROM
RAM
Jahr
g)
Kosten in mc /bit
16K
64K
256K
1M'
4M'
16M'
64M'
(1981)
Grenze
0,30 mc /bit
Jahr
h)
bit je Jahr
(1981)
Jahr
i)
Marktvolumen in Mill. Dollar
(1977)
Kernspeicher
MOSRAM
Jahr
j)
Zugriffszeit
1ms
1µs
1ns
Röhre
Transistor
IC
Kernspeicher
Halbleiterspeicher
Jahr
k)
Zugriffszeit in ns
1K
stat. RAM
4K
(1977)
Jahr
l)
Zugriffszeit in ns
1K
RAM
4K
16K
64K
(1980)
Jahr
m)
Q
d-RAM
1K
4K
16K
64K
256K
(1979)
Jahr
n)
Q
1K
St-RAM
4K
64K,55ns
16K,55ns
4K,20ns
(1979)
Jahr
o)
Q
EPROM
8K
16K
32K
64K
(1979)
Jahr
p)
Kapazität in Kbit
(1979)
1K
4K
16K
64K
256K
Jahr
q)
Kapazität in Kbit
(1979)
dyn. Typen mit Refresh
langsame Typen
schnelle Typen
Jahr
r)
Kapazität in Kbit
(1979)
2K
8K
16K
32K
64K
128K
256K
Jahr

kapazitäten bei MOS- bzw. bipolaren Speichern erwartet werden. Vergleichsweise ist die zu den jeweiligen Zeiten erreichbare Gatterzahl je Chip angegeben.

Teilbild d) zeigt, wieviel Speicherzellen (Kapazität) je mm^2 Chipfläche für verschiedene Speichertechnologien vorhanden waren bzw. sind. Als Vergleichsbezug ist die Größe eines Transistors aufgetragen. Lediglich CCD unterschreitet diese Grenze (wobei allerdings neuere Daten fehlten).

In Teilbild e) ist der weltweite Umsatz von Speichern einschließlich einer Prognose bis 1990 aufgezeigt. Auffallend ist der relativ kleine Anteil von bipolaren Speichern sowie künftig von Bubblespeichern und GaAs. Eine andere Einteilung unter denselben Voraussetzungen zeigt eine Aufteilung nach RAM, ROM/PROM und seriellen Speichern.

Wie groß die Gesamtkapazität an gebrauchten bzw. benötigten Halbleiterspeichern ist, weist Teilbild h) aus.

Teilbild j) zeigt, daß über einen Zeitraum von ca. 40 Jahren die Zugriffszeit von Speichern um rund 5 Zehnerpotenzen mit ständig verbesserten Technologien fiel. Für einen kleineren Zeitraum und bei mittleren Speichern untersetzt dies Teilbild k), während l) eine Prognose darstellt. Sie weist aus, daß langfristig generell etwa 20 bis 40 ns erreicht werden. Ein Wert, der z. Z. (1983) von ECL-Speichern eingehalten und sehr bald für n-MOS erwartet wird.

In den Teilbildern m) bis o) wird für die drei Speichertypen dynamische und statische RAM sowie EPROM ein allgemeiner Qualitätsparameter

$$Q = P \cdot A \cdot T/K$$

betrachtet. Darin bedeuten

- P die Leistung in mW
- A die Chipfläche in 10^3 mil^2 ($\approx 25{,}4 \times 25{,}4\ \mu m^2 \approx 650\ \mu m^2$)
- T die Zugriffszeit in ns
- K die Kapazität in bit.

Die Teilbilder p) bis r) zeigen untersetzt zu den Teilbildern b), c) und g) die Entwicklung für die drei Typen dynamische und statische RAM sowie EPROM.

8.3.2. Organisationsformen

Ein Speicher besteht aus einer Vielzahl von Speicherzellen. Sie müssen technisch so realisiert sein, daß sie in definierter Weise mit Information versehen (Aufzeichnen) bzw. die Informationen gelesen (Wiedergeben) werden können. Dies muß mit einer begrenzten Anzahl von Leitungen, entsprechend der pin-Zahl der Gehäuse, erfolgen können. Durch diese begrenzte pin-Zahl (ca. 16 bis 40) ist es notwendig, den Zugriff zu den einzelnen Speicherzellen in unterschiedlicher Weise zu organisieren. Zur Erklärung dieser Problematik ist zu beachten, daß ein Speicherchip *Zuleitungen* u. a. für folgende Anschlüsse *benötigt*:

- Betriebsspannung (2 bis 4),
- Adressierung der Speicherzellen (wird im folgenden behandelt),
- Leitungen für die aufzuzeichnenden bzw. wiederzugebenden Signale,
- Eingang zum Schreiben (Aufzeichnen) bzw. Lesen (Wiedergeben), meist als W bzw. $\overline{W}$ bezeichnet,
- Chip Select ($\overline{CS}$ bzw. CS) für die Aktivierung des ausgewählten Bausteines,
- evtl. weitere Anschlüsse wie CAS, RAS (vgl. Abschnitt 8.3.4.6.).

Bei den folgenden Betrachtungen werden nur die Anschlüsse für die Adressierung betrachtet, die anderen sind für einen realen Speicher zu ergänzen. Dazu gibt Bild 8.3.4 einen Überblick. Gemäß ① kann jede Speicherzelle direkt aufgerufen werden. Dies bedeutet, daß für m bit Kapazität auch n Adreßleitungen notwendig sind. Eine solche Anordnung ist aber nur bei ca. 4 bis 16 bit Kapazität effektiv; sie hat den Vorteil, sehr schnell zu sein. In der Regel führt diese Organisation aber unmittelbar zu den

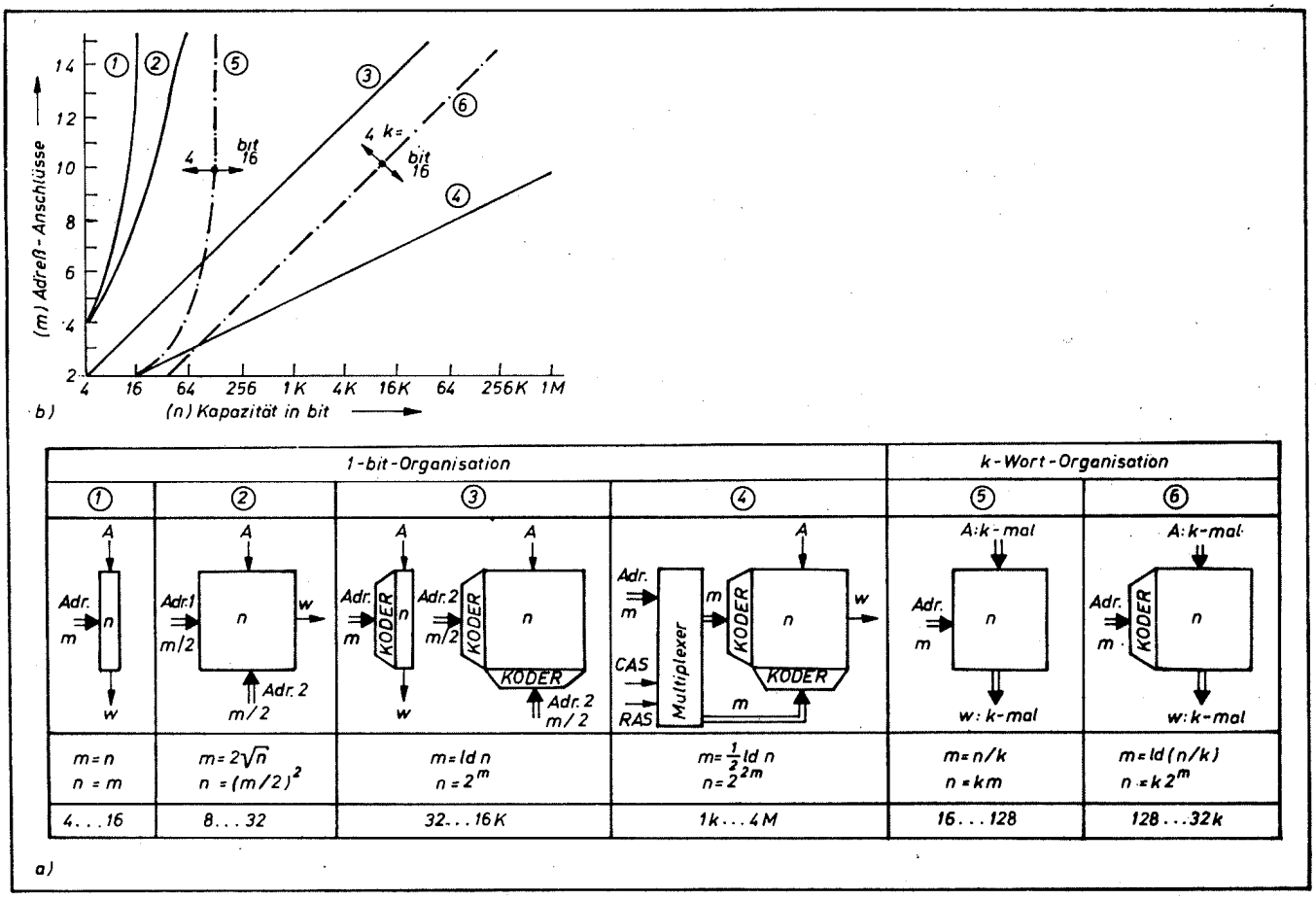


Bild 8.3.4. Zur Organisation von Speichern. Es bedeuten n Bitkapazität, m Adreßleitungen, k Wortbreite

a) Strukturelle Prinzipien im Vergleich
b) Zusammenhänge von Kapazität und Adreßleitungen

Schieberegistern, die zusätzliche Funktionen besitzen und im Abschnitt 8.5.5. noch genauer behandelt werden.

Eine gewisse Kapazitätserhöhung ergibt sich, wenn jede Speicherzelle über ein AND aus zwei Leitungen ausgewählt wird. Dies entspricht der Matrixorganisation von ②, und für n bit sind nur noch $2\sqrt{n}$ Adreßleitungen notwendig. Der günstige Kapazitätsbereich liegt hier bei ca. 8 bis 32 bit.

Ein wesentlicher Kapazitätsgewinn wird erreicht, wenn gemäß ③ die Adressen der Speicherzellen über Dekoder ausgewählt werden (vgl. 8.2.8.1.). Dann sind nämlich nur ld n Adreßleitungen notwendig. Bedeutsam ist, daß dieser Wert unabhängig davon gilt, ob Speicherzellen dann direkt gemäß ① oder im Matrixprinzip gemäß ② über die Kodierung ausgewählt werden. Dabei existiert nur ein Unterschied in der Kodierungstiefe und -breite und damit in gewissem Umfang auch bezüglich der Verzögerungszeit und des Kodierungsaufwandes. Mit dieser Methode kann optimal ein Speicherbereich von etwa 32 bit bis 16 Kbit realisiert werden.

Für noch größere Speicher werden die Adressen zusätzlich über einen Demultiplexer gemäß ④ in nieder- und höherwertigen Anteilen nacheinander (sequentiell) ausgewählt. Dadurch sind nur $\frac{1}{2}$ ld n Adreßleitungen am Chip notwendig. Dieses Prinzip ist für Speicherkapazitäten von etwa 1 Kbit bis 4 Mbit effektiv. Es reicht also für alle Entwicklungen bis etwa zum Jahr 2000 aus.

Alle bisherigen Organisationsformen waren darauf orientiert, jeweils genau eine Speicherzelle bei der Aufzeichnung bzw. Wiedergabe auszuwählen. Dieses Prinzip heißt Bitorganisation. Daneben gibt es eine k-Wort-Organisation, wobei k bevorzugt 4 oder 8 ist. In diesem Fall werden also gleichzeitig 4 bzw. 8 Speicherzellen mit einer

Adresse ausgewählt. Dieses Prinzip hat eventuell Vorteile, wenn ein entsprechender Bus (vgl. 8.7.) benutzt wird. Es existieren dabei die zu ① und ③ gehörenden Organisationen gemäß ⑤ und ⑥. Sie sind damit ohne weiteres verständlich, benötigen aber nicht eine Datenleitung, sondern k, d. h. 4 bzw. 8. Sie sind deshalb für 16 bis 128 bit bzw. für 128 bit bis 32 Kbit geeignet.

In Bild 8.3.4b) ist graphisch der Zusammenhang zwischen Kapazität von Speichern und den für die einzelnen Organisationsprinzipien notwendigen Adreßleitungen dargestellt.

Eine speziell in den letzten Jahren auftauchende Speicherform ist das Nibble-Prinzip. Ein Nibble ist meist ein Halbbyte, also 4 bit. In diesem Fall wird damit aber eine bestimmte interne Speicherorganisation gemeint, die in extrem kurzen zeitlichen Abständen, meist ca. 20 bis 50 ns, 4 bit nacheinander ausgibt. Dies ist vor allem bei Bildschirmbetrieb, z.B. für hochauflösende Grafik, nützlich.

Zuweilen wird die Frage gestellt, warum man nicht zu höherdimensionalen Matrixprinzipien übergeht. In der Anordnung ③ wird dadurch kein Vorteil sichtbar. Bei r-Dimensionalität gilt in der Form ② zunächst je Dimension $\sqrt[r]{n}$. Wird dieser Wert kodiert, folgt $(\mathrm{ld}\, n)/r)$, und da dies für jede Dimension erfolgen muß, ergibt sich wieder ld n für alle Adreßleitungen. Es würden also nur Vorteile in der Kodiertiefe und -verzögerung sowie ihrem Aufwand entstehen. Anderseits entstehen aber Nachteile im Lay-out, denn der Chip ist flächig, also zweidimensional. Deshalb wird fast immer die rechte Variante in ③ angewendet. Konsequenzen hieraus für den wortorganisierten Speicher werden noch im Abschnitt 8.3.4.3. behandelt.

8.3.3. Bipolare Speicher

Es gibt eine Vielzahl von bipolaren Speicherzellen. Im folgenden werden ein TTL-Speicher und das I^2-L-Prinzip genauer erklärt. Anschließend werden weitere Prinzipien kurz ergänzt.

8.3.3.1. Latch, Speicherzelle

Das Grundprinzip der einfachsten Speicherzelle, eines Latch (engl.; Klinke, also einklinken, einrasten), zeigt Bild 8.3.5b). Es besteht aus zwei stark gleichstrommäßig mitgekoppelten Transistoren T_1 und T_2. Diese Schaltung besitzt zwei stabile Zustände:

T_1 leitend; T_2 gesperrt; $Q = \mathrm{L}$; $\overline{Q} = \mathrm{H}$

T_1 gesperrt; T_2 leitend; $Q = \mathrm{H}$; $\overline{Q} = \mathrm{L}$.

Sie können den logischen Werten Ø und 1 zugeordnet werden. Als Symbol einer Speicherzelle kann das eines RS-Flipflop gemäß dem oberen Teil in Bild 8.3.5c) genutzt werden. Den Aufbau eines Latch aus TTL (oder anderen logischen Gattern) zeigt Bild 8.3.5c) bis e) in den drei möglichen Varianten

c) aus zwei NOR
d) aus zwei NAND
e) aus je einem AND und OR.

Die Eingänge R und S dienen zum Rücksetzen (reset) und Setzen (set) des Latch. Setzen bedeutet immer $Q \rightarrow \mathrm{H}$. Weitere damit zusammenhängende Probleme werden bei den Flipflops (8.5.2.) behandelt.

Die Kennlinienkonstruktion für das Latch aus zwei NAND-Gliedern (c) zeigt Teilbild a). Dabei ist einmal die typische TTL-Übertragungskennlinie gemäß Bild 8.2.10c) bzw. d) vorausgesetzt. Entsprechend der Zusammenschaltung der

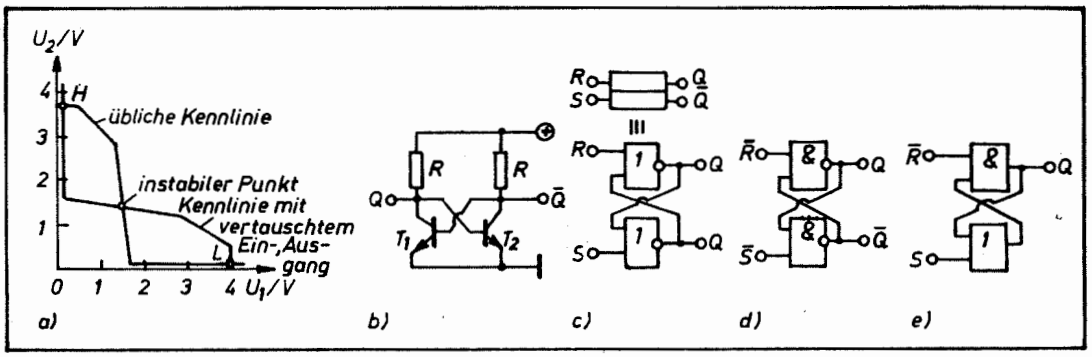


Bild 8.3.5. Prinzip einer Speicherzelle (Latch)

a) Zwei stabile Punkte L und H sowie ein labiler Punkt entstehen bei Rückkopplung von zwei NAND-TTL-Kennlinien (vgl. d))

b) Aufbau eines Latch aus zwei rückgekoppelten Transistoren, von denen jeweils einer leitend ist

c) Zusammenschaltung von zwei NOR zu einer Speicherzelle, darüber Symbol eines äquivalenten Flipflop gemäß Abschnitt 8.5.

d) Dito aus zwei NAND

e) Dito aus einem AND und einem OR

beiden Gatter in Bild 8.3.5d) muß sie ein zweites Mal für das zweite Gatter mit vertauschten Ein- und Ausgangspegeln ergänzt werden. Dann entstehen drei Schnittpunkte,

bei H, bei L und ein instabiler.

Beim instabilen Punkt genügt eine ganz geringe Abweichung, und es wird einer der beiden stabilen Zustände L bzw. H angenommen. Am R- bzw. S-Eingang ist also nur eine Spannung zuzuführen, damit der instabile Punkt überschritten und der gewünschte stabile Zustand eingenommen wird. Dieser Fakt ist für alle Speicherzellen und -anwendungen wichtig.

8.3.3.2. Statisches RAM, 16 bit, Typ D 181

Den Aufbau und Teilschaltungen eines statischen 16-bit-RAM auf TTL-Basis zeigt Bild 8.3.6. Die Speicherzelle enthält Teilbild a). Sie besitzt zwei Dreiemittertransistoren, wobei je ein Emitterpaar zum Leseverstärker, zur x- und y-Auswahl führt. Es sind also nicht die Q- und $\overline{Q}$-Anschlüsse herausgeführt, sondern statt dessen der jeweils dritte Emitter. Dies bedingt u. a. Vorteile für die Geschwindigkeit der Speicherzelle. Ein der Speicherzelle sehr ähnliches Lay-out (allerdings nur zwei Emitter) zeigt Teilbild b). Die matrixförmige Speicherzellenanodenung gemäß Bild 8.3.4a), Nr. 2 zeigt Bild 8.3.6c).

Befindet sich der Speicher im Ruhezustand, so liegen alle x- und y-Leitungen auf niedrigem Potential. Die Ströme des jeweils geschalteten Transistors in einer Zelle fließen über die beiden entsprechenden Emitter. Wird ein Emitter z.B. von der entsprechenden x-Leitung auf hohes Potential gehoben, so übernimmt der andere Emitter den gesamten Strom, und an seiner Größe ändert sich nichts. Erst wenn auch die zugehörige y-Leitung auf hohes Potential gebracht wird, muß der dritte Emitter den Strom übernehmen. Er fließt dann über den Widerstand R_L bzw. R_0 des zugehörigen Schreibleseverstärkers. Bezüglich dieser Widerstände bestehen Probleme in der Stabilität des Flipflop. Hierzu zeigt Bild 8.3.6e) die Spannungsverhältnisse in dem vereinfachten Flipflop, wenn die Emitter auf Masse liegen. Der rechte Transistor T ist leitend. Werden jetzt Widerstände eingeschaltet, so fließt in beiden Transistoren Strom, und der Zustand des Flipflop ist nicht mehr eindeutig. Mit Hilfe eines über zwei antiparallele Dioden gegengekoppelten Verstärkers (Teilbild f)) kann jedoch der Span-

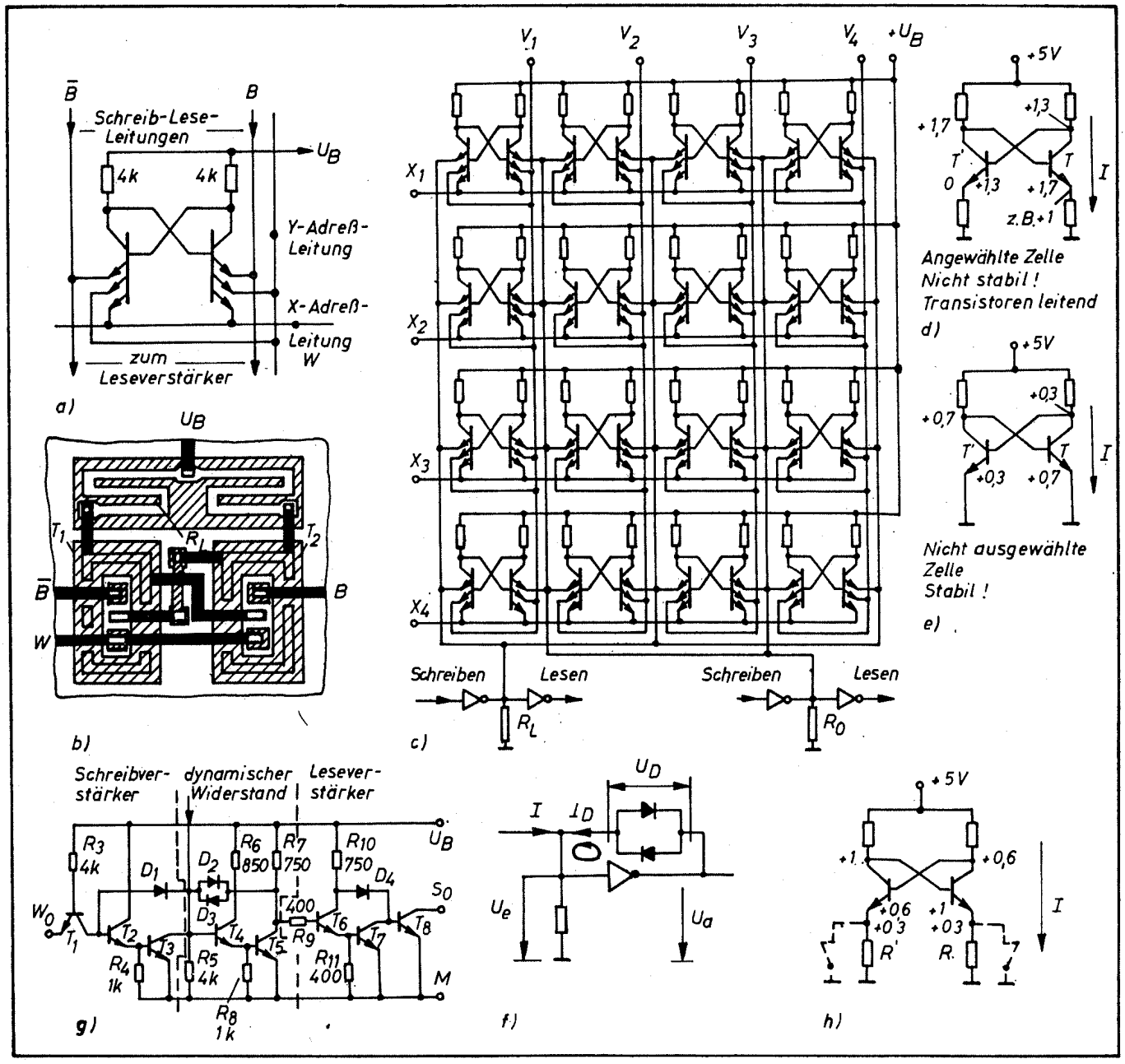


Bild 8.3.6. Der TTL-Speicher D 181 (7481)

a) Grundaufbau einer Speicherzelle
b) Vereinfachtes Prinzip-Lay-out für eine TTL-Speicherzelle, jedoch nur mit zwei Emittern
c) Schaltung der Matrix und Anordnung der beiden Schreib-Lese-Verstärker
d) Ausgewählte Speicherzelle, die durch den Spannungsabfall an den Widerständen instabil werden könnte, wenn nicht der dynamische Widerstand gemäß f) verwendet würde
e) Speicherzelle nicht ausgewählt: Ruhestand: beide Emitter auf Masse
f) Prinzip des dynamischen Widerstandes für die Lese-Schreib-Emitter
g) Schaltung der Schreib-Lese-Verstärker
h) Reale Spannungsverhältnisse an der Speicherzelle mit dynamischem Emitterwiderstand. Das Schreiben (Aufzeichnen) erfolgt durch Betätigen der Schalter, die im linken Teil des Schreib-Lese-Verstärkers nach g) elektronisch realisiert sind

nungsabfall an den Widerständen unabhängig vom Strom auf etwa 0,3 V konstant gehalten werden. Dies ergibt sich wie folgt: Ist v die Verstärkung, so gilt

$$u_a = U_B - v u_e \,. \tag{1}$$

Wegen des KIRCHHOFFschen Gesetzes gilt weiter in der eingezeichneten Schleife

$$u_a - u_e = u_D = \text{konstant} \,. \tag{2}$$

Wird u_a in Gl. (1) ersetzt, so folgt

$$U_B - vu_e - u_e = u_D \tag{3}$$

oder aufgelöst

$$u_e = \frac{U_B - u_D}{v + 1}. \tag{4}$$

Da alle rechts stehenden Größen konstant sind, ist es auch u_e. Der Strom für den entstehenden konstanten Spannungsabfall kann aus der Speichermatrix oder aus den Dioden durch R fließen. Dementsprechend ändert sich die Spannung u_a. Sie zeigt also an, ob die entsprechende Seite des Flipflop leitend oder gesperrt ist. Damit ist der Leseverstärker in Bild 8.3.6g) verständlich.

Die Spannung an R ist immer etwa 0,3 V positiv. Es kann von der Matrix her also nur ein Strom hineinfließen, wenn eine Speicherzelle angewählt wird. Es liegen sonst ein oder beide Emitter auf dem tieferen Potential L und übernehmen den Strom des leitenden Transistors.

Beim Schreiben wird am Eingang des Schreibverstärkers ein H-Signal angelegt. Dadurch werden die Transistoren T_2 und T_3 leitend. Dies bedeutet einen Kurzschluß des Widerstandes R_5. Die Wirkung für die Speicherzelle ist im Bild 8.3.6h) gestrichelt eingezeichnet. Das Flipflop ist dann auf der zugehörigen Seite leitend. Infolge des kurzgeschlossenen Widerstandes kann während der Schreibzeit nicht gelesen werden.

8.3.3.3. I^2L-Speicher

Die I^2L-Technik wurde bereits im Abschnitt 8.2.6. behandelt. Die dort genannten Vorteile wiederholen sich bei den Speichern. Wieder werden nicht einzelne Transistoren integriert, sondern die gesamte Funktion. Ein Transistorenersatzschaltbild gibt daher nur teilweise richtige Aussagen.

Einen Ausschnitt aus einem I^2L-Speicher zeigt Bild 8.3.7 von der Oberfläche (a) und im Schnitt (b). Eine Speicherzelle ist getrennt im Teilbild c) herausgezeichnet. Dabei ist zu beachten, daß die p-Gebiete I_1 (Injektor) und E_L bzw. E_H (Emitterzustand L bzw. H) auch noch zu der jeweils rechten und linken Speicherzelle gehören. Die oberen und unteren p^+-Streifen isolieren die Zelle gegenüber darüber- und darunterliegenden. Das eigentliche Latch bilden die beiden mittleren Strukturen, die als T_1 und T_2 bezeichnet sind und im Vergleich zu üblichen Transistoren invers betrieben werden. Die anderen drei Strukturen I_1, E_L und E_H können zunächst zur Vereinfachung als Injektoren aufgefaßt werden. Sie stellen zwei Zweifachemitter-Lateral-Transistoren, jeweils zu T_1 und T_2, dar. So entsteht dann die Ersatzschaltung des Teilbildes d). Die vereinfachten Verhältnisse bei der Wiedergabe zeigt Teilbild e), wobei T_2 als leitend und T_1 als gesperrt angenommen sind. Ferner ist die eigentliche Injektion I_1 als Stromquelle gezeichnet. Das $I_1/4$ ergibt sich durch die Aufteilung auf die vier benachbarten Transistoren. Infolge des hohen Pegels an der Basis von T_2 ist auch die Emitter-Basis-Strecke von T_L leitend, und über E_L (Schreib-Lese-Leitung für Zustand L) kann ein Strom von außen fließen (T_L arbeitet invers). Beim Schreibvorgang könnte im Prinzip bei E_L ein Strom eingeführt werden (Teilbild f)). Dann bricht die Spannung an der Basis von T_2 über den Kollektor von T_L zusammen, und die Speicherzelle schaltet um. T_1 wird leitend. Durch die ineinandergeschachtelten Speicherzellen sind die Verhältnisse noch komplizierter als eben dargestellt. In Wartestellung liegen die x-Leitungen auf 0 V und als I fließt allen Zellen etwa je 1 µA zu. Außerdem fließt aus jeder Zelle über E_H oder E_L je nach Zustand der Zelle ebenfalls ein Strom von ca. 1 µA. Beim Aufruf für Lesen und Schreiben wird das x-Gebiet der entsprechenden Zeile auf −0,3 V abgesenkt. Dadurch werden die anderen x-Zeilen entkoppelt. Beim

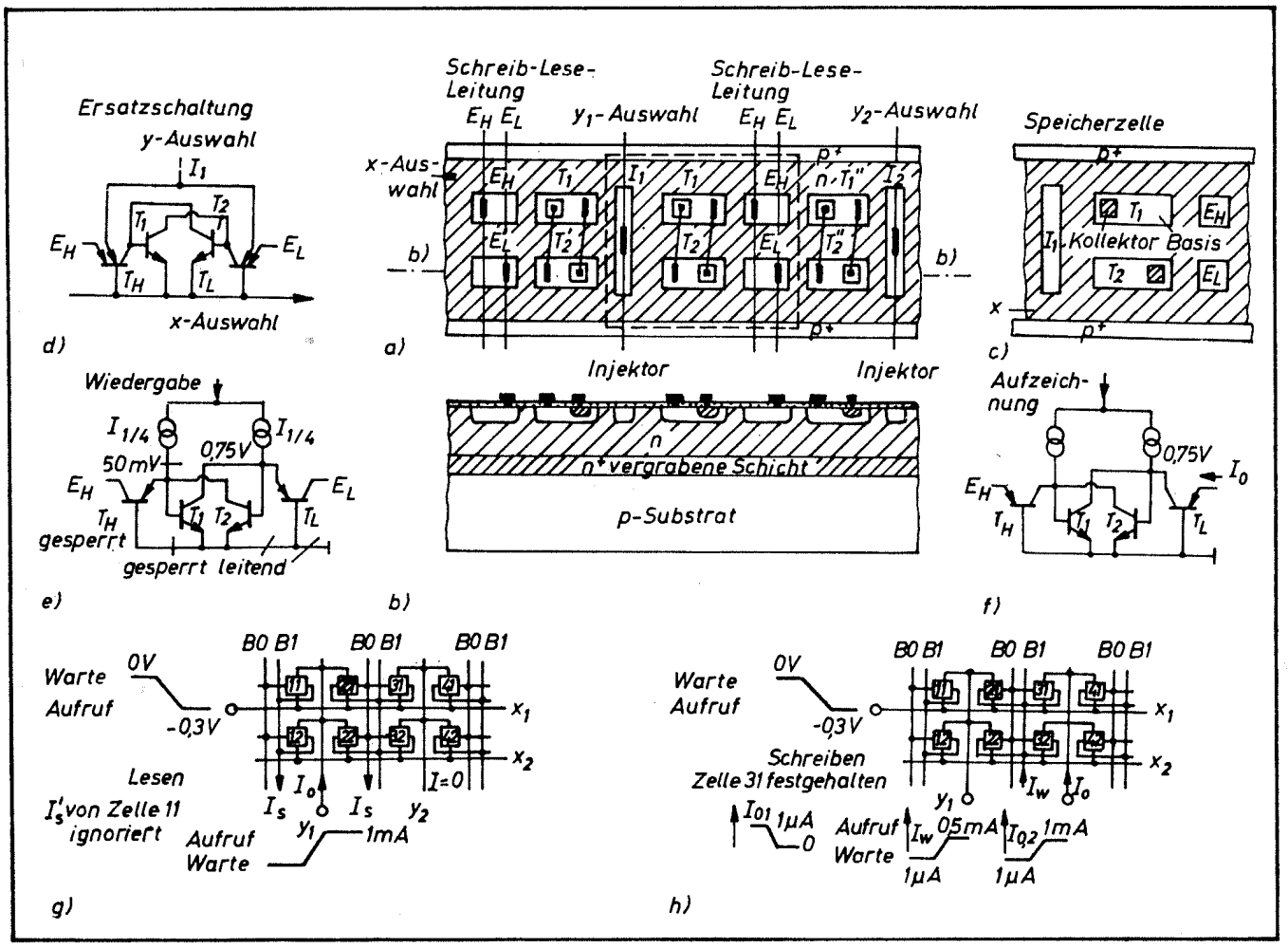


Bild 8.3.7. Zur I²L-Speicherzelle

a) Oberflächenstruktur eines I²L-Speichers
b) Schnitt zur Verdeutlichung der Tiefenstruktur
c) Herausgezogene Speicherzelle, wobei die p-Gebiete I_1, E_L und E_H auch für die Nachbarzellen wirksam sind
d) Diskretisierte Ersatzschaltung der Struktur
e) Vereinfachter Betrieb beim Lesen
f) Vereinfachter Betrieb beim Schreiben
g) Gesamtzusammenhänge beim Lesen
h) Gesamtzusammenhänge beim Schreiben

Lesen wird gleichzeitig der Strom durch den Injektor von 1 µA auf 1 mA erhöht. Dadurch erhalten die entkoppelten Zeilen etwa den gleichen Strom wie zuvor, aber durch die Zellen 11 und 12 im Bild 8.3.7g) fließt ein erhöhter Strom, der an E_H oder E_L bzw. E'_H oder E'_L entnommen werden kann. Durch diese Auswahl ist die jeweilige Zelle exakt bestimmt.

Beim Schreiben wird der Strom durch y_1 auf Null gebracht. Dies darf nur so kurzzeitig erfolgen, daß die Zellen über ihre Kapazitäten den Zustand beibehalten. Der Nachbarstrom y_2 wird dagegen auf 1 mA erhöht. Ebenso wird der Strom in E_L bzw. E_H je nachdem, welcher Wert eingeschrieben werden soll, erhöht, jedoch nur auf etwa 0,5 mA. Dadurch bleibt die Stromrichtung in 31 und deren Zustand erhalten; anders jedoch in der Zelle 21.

8.3.3.4. Weitere Speicherzellen, minimaler Strom

Das Grundprinzip aller bipolaren Speicherzellen leitet sich aus der Schaltung von Bild 8.3.5b) ab. *Unterschiede von Speicherzellen* bestehen vor allem in *folgenden Punkten*:

- Wahl der beiden rückgekoppelten Inverter,
- Wahl der Lastwiderstände R,
- Ankopplung der Auswahlschaltung für die einzelnen Speicherzellen,
- Ein- bzw. Auskopplung der Schreib- bzw. Lesesignale.

Bei der TTL-Schaltung von Abschnitt 8.3.3.2. wurden die TTL-Prinzipien mit Multiemittertransistoren voll genutzt. Die Lastwiderstände waren rein ohmsch, und die Auswahl für x und y erfolgte genauso wie die der Schreib-Lese-Technik über je einen Emitter. In der Regel werden aber, wie bereits bei der I^2L-Zelle behandelt, die x-Anwahl und die Schreib-Lese-Technik zusammengefaßt. So entsteht die Schreib-Lese-Leitung, auch Datenleitung oder Bitleitung genannt. Sie wird jetzt generell mit B und $\overline{B}$, getrennt für die beiden Zustände, genannt. Die y-Leitung heißt dann Wortleitung W. So entsteht die Standard-TTL-Zelle für größere Kapazitäten gemäß Bild 8.3.8a).

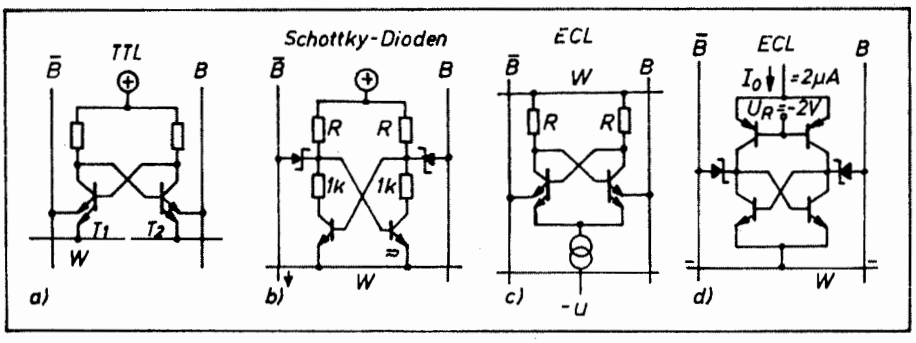


Bild 8.3.8. Weitere Speicherzellen

a) Vereinfachte TTL-Zelle
b) Speicherzelle mit SCHOTTKY-Dioden
c) ECL-Speicherzelle mit Stromquelle
d) ECL-Zelle mit SCHOTTKY-Dioden

Für schnelle Speicherzellen wurde sie durch die SCHOTTKY-*Dioden-Technik* gemäß Bild 8.3.8b) ersetzt. An der SCHOTTKY-Diode entsteht ein kleinerer Spannungsabfall als an einer Basis-Emitter-Strecke. Außerdem wird der Platz für zwei Emitter eingespart, da der Anschluß für B bzw. $\overline{B}$ zugleich die Chipfläche der SCHOTTKY-Diode ist. Schließlich wird Geschwindigkeit gewonnen, da die Transistoren wegen der SCHOTTKY-Dioden nicht in die Sättigung gelangen. Dies führt wieder zur noch schnelleren ECL-Technik gemäß Bild 8.3.8c), bei der Wortleitung W und Spannungsquelle als Anschlüsse vertauscht sind und zusätzlich statt der Spannungsquelle eine Stromquelle zur Anwendung gelangt. Eine weiter verbesserte ECL-Technik zeigt Bild 8.3.8d).

Von großer Bedeutung — ganz besonders für große Kapazitäten — ist es, *wieviel Strom eine Speicherzelle verbraucht.* Hierfür lassen sich einfache Abschätzungen angeben. Der Strom I_0 muß mindestens so groß sein, daß gemäß Teilbild 8.3.5a) die drei Schnittpunkte der Kennlinien mit hinreichendem Störabstand zu unterscheiden sind. Hierzu muß der Spannungsabfall $I_0 R$ am Arbeitswiderstand R größer als die doppelte Temperaturspannung U_T sein. Für Zimmertemperatur gilt daher die zugeschnittene Größengleichung

$$I_0 > \frac{52\,\mu\mathrm{A}}{R/\mathrm{k}\Omega}. \tag{5}$$

Es sind also möglichst große Widerstände erforderlich. Sie können dabei auch als dynamische Widerstände, wie z.B. in Bild 8.3.8d), realisiert werden.

8.3.4. MOS-Speicher

8.3.4.1. Statische Speicherzellen

Eine Zusammenstellung einiger typischer statischer Speicherzellen und zugehöriger Lay-outs zeigt Bild 8.3.9. Teilbild b) zeigt die typische 6-Transistorzelle. Die Transistoren T_1 und T_2 bilden mit den als Lastwiderstände wirkenden Transistoren T_3 und T_4 die beiden rückgekoppelten Inverter in Analogie zu Bild 8.3.5b). Die Transistoren T_5 und T_6 koppeln diese Zelle mit den beiden Bitleitungen B und $\overline{B}$. Die Steuerung erfolgt über die Wortleitung W. Das Lay-out einer älteren Struktur zeigt Bild 8.3.9a), während ein modernes Lay-out Teilbild c) zeigt. Die Auslegung eines Speichers nach dieser Schaltung enthält Abschnitt 8.3.4.2.

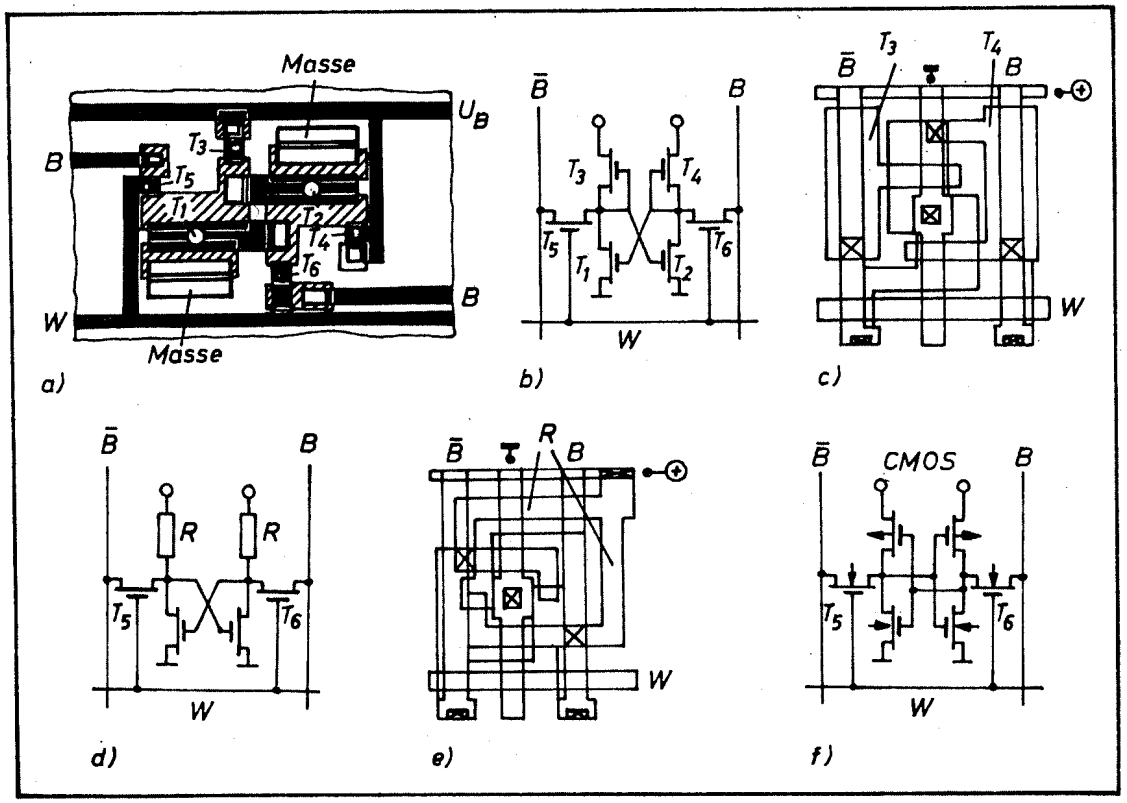


Bild 8.3.9. Statische MOS-Speicherzellen

a) Veraltetes Lay-out einer 6-Transistor-Zelle
b) Schaltbild der 6-Transistor-Zelle
c) Modernes Lay-out
d) Schaltbild mit ohmschen Widerständen
e) Lay-out für d) in Analogie zu c)
f) Beispiel einer statischen CMOS-Speicherzelle

Mit der Entwicklung der Polysiliciumtechnik wurde es möglich, große Widerstände im MΩ-Bereich zu realisieren. So entstand die Transistorzelle von Teilbild d) mit dem zu c) analogen Lay-out von Bild 8.3.9e). (Ein Lay-out mit VMOS zeigt das spätere Bild 8.3.11.) Durch Polysiliciumwiderstände anstelle der MOS-Widerstände kann Chipfläche gewonnen werden (vgl. Teilbild c)).

Im Bild 8.3.9f) ist schließlich noch eine statische Speicherzelle auf der Basis von CMOS dargestellt, wie sie z. B. im INTEL 5101 verwendet wird. Die Speicherzelle aus den Transistoren T_1 bis T_4 verbraucht hier nur einen Ruhestrom von ca. 1 nA. Er ist so klein, daß ein solcher Speicher selbst bei größerer Kapazität ständig von einer Batterie gepuffert werden kann und dabei in derselben Größenordnung wie der Leckstrom der Batterie liegt.

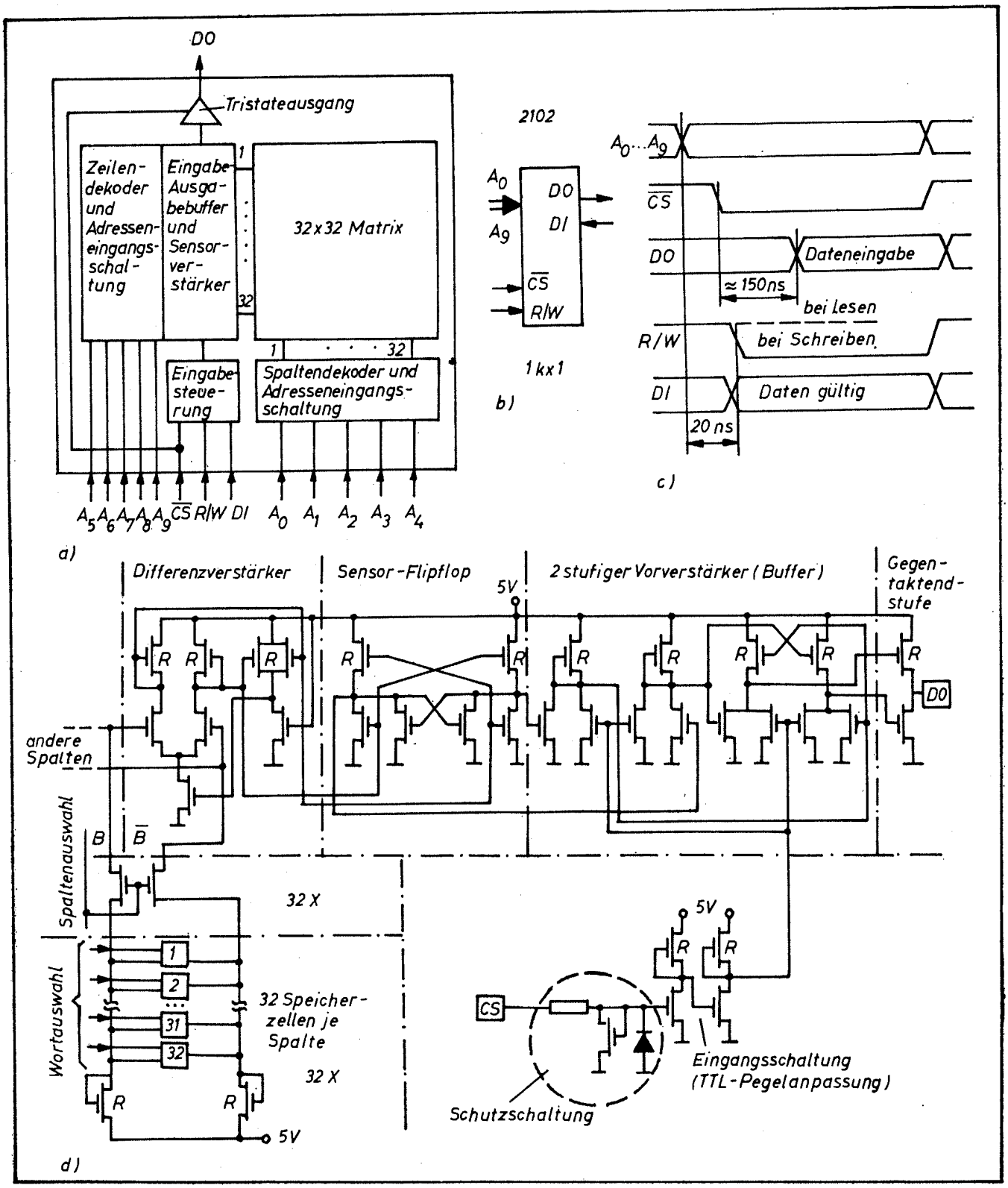


Bild 8.3.10. Zum (1 K × 1 bit)-statischen MOS-Speicher

a) Prinzipschaltbild
b) Schaltsymbol
c) Zeitbedingungen der Signale
d) Schaltbild für Zeilen-, Spaltenauswahl und Wiedergabeverstärker, einschließlich Chip-Select

8.3.4.2. Statischer (1 K × 1 bit)-MOS-Speicher 2102

Den Prinzipaufbau eines Speichers 2102 zeigt Bild 8.3.10a). Im Zentrum liegt die Speichermatrix, die 32 × 32 organisiert ist und über zwei Dekoder von den 10 Adreßleitungen A_0 bis A_9 angesteuert wird (vgl. Bild 8.3.4, Nr. 3, rechts). In diesem Sinne

stellt Teilbild b) das übersichtliche Kurzschaltbild des Speichers dar. In Teilbild d) sind weitere Schaltungseinzelheiten zu sehen. Vor allem fehlen die Dekoder, welche die Verbindung von den Adreßleitungen A_0 bis A_9 zur Spalten- und Wort- (Zeilen-) Auswahl realisieren, sowie der Aufzeichnungsverstärker. Die Speicherzellen einer Spalte sind durch die Kästchen 1; 2 und 31; 32 engedeutet. Sie enthalten genau die 6 Transistoren gemäß Bild 8.3.9b). Mittels der Spaltenauswahl wird dann das Ausgangssignal oder das Schreibsignal zu oder von den Bitleitungen B und $\overline{B}$ über je einen FET weiter durchgeschaltet. Der gegengekoppelte Differenzverstärker verstärkt das Wiedergabe- (Lese-) Signal und führt es dann dem Sensor-Flipflop zu, der eindeutige Pegel für L oder H aus dem Signal entsprechend den MOS-Pegeln der ausgewählten Zelle realisiert. Schließlich folgt ein zweistufiger Vorverstärker, der das Signal soweit vergrößert, daß eine Gegentaktendstufe angesteuert werden kann, die das Ausgangssignal DO (data out) im TTL-Pegel bereitstellt. Diese Gegentaktendstufe kann mit dem Chip-Select-Signal $\overline{CS}$ im Tristate-Fall auf einen sehr hochohmigen Wert gebracht werden. Dann ist der Baustein vom DO-Ausgang praktisch abgeschaltet. Wie aus Teilbild a) zu ersehen ist, geht die Chip-Select-Steuerung auch an die Eingabe (den Schreibverstärker).

Im Teilbild c) ist noch der notwendige zeitliche Signalablauf für den Betrieb des Speichers aufgezeigt. Nachteilig bei diesem Speicherbaustein ist zuweilen, daß der Read-write- (Schreib-Lese-) Eingang R/W beim Schreiben nicht die Ausgangssignale abschaltet.

8.3.4.3. Ein VMOS-(2 K × 8 bit)-Speicher

Einen wortorganisierten (2 K × 8 bit)-Speicher von INMOS aus dem Jahr 1980 zeigt Bild 8.3.11. Seine Struktur entspricht in etwa Bild 8.3.4, Nr. 5. Da die Matrix für 16 K jedoch 128 × 128 organisiert ist, werden die Zeilen direkt kodiert durch die Adressen A_4 bis A_{10}, während die Spalten aufgeteilt sind, und zwar in 16 Worte zu 8 bit. Hier werden die acht Ausgangspuffer bei WE (write enable = Schreiben möglich) in den hochohmigen Zustand (Tristate) gebracht. Außerdem gibt es hierfür einen zusätzlichen Anschluß $\overline{OE}$ (output enable). Abgesehen davon, daß VMOS-Transistoren und Polysilicium-Arbeitswiderstände verwendet werden (Bild 8.3.9d)), weicht die Schaltung in Bild 8.3.11a) prinzipiell kaum von der vorigen in Bild 8.3.10 ab. Im Wiedergabeverstärker fehlt das Sensor-Flipflop, er ist dafür dreistufig. Entsprechend dem fortgeschrittenen Technologiestand ist die Speicherzelle (Teilbild b) und c)) mit $22 \times 38 \approx 850\ \mu m^2$ sehr klein. Die Zugriffszeit des Speichers liegt bei rund 150 ns, der Leistungsverbrauch bei 370 mW. Die Größe des Chips liegt bei $5{,}9 \times 3{,}4\ mm^2$. Auf dieser Fläche sind insgesamt mehr als 100000 Transistoren untergebracht.

8.3.4.4. Absenkung der Verlustleistung bei statischen RAM

Bereits bei der CMOS-Speicherzelle gemäß Bild 8.3.9f) wurde darauf hingewiesen, daß — wie überhaupt bei der CMOS-Technik — im Ruhefall, also ohne Schreiben und Lesen, eine Speicherzelle extrem wenig Energie verbraucht. Dies gilt nur z. T. für den gesamten Speicher, da bei ihm die Zusatzschaltungen auch erhebliche Energie verbrauchen können. So wird bei CMOS-Speichern mit mehr als 4 Kbit der Leistungsverbrauch im Betrieb (operating power) und in Ruhe (standby power) i. a. mit folgenden Beispielwerten angegeben: 200 mW/10 μW; 75 mW/530 μW oder 15 mW/50 nW.

Bei den statischen MOS-Speichern besteht eine Möglichkeit, die Leistung wesentlich zu reduzieren, darin, daß die Betriebsspannung an den Speicherzellen in einem Taktverhältnis von 1:3 bis 1:1000 gepulst wird. Dann geschieht gemäß Bild 8.3.12 folgen-

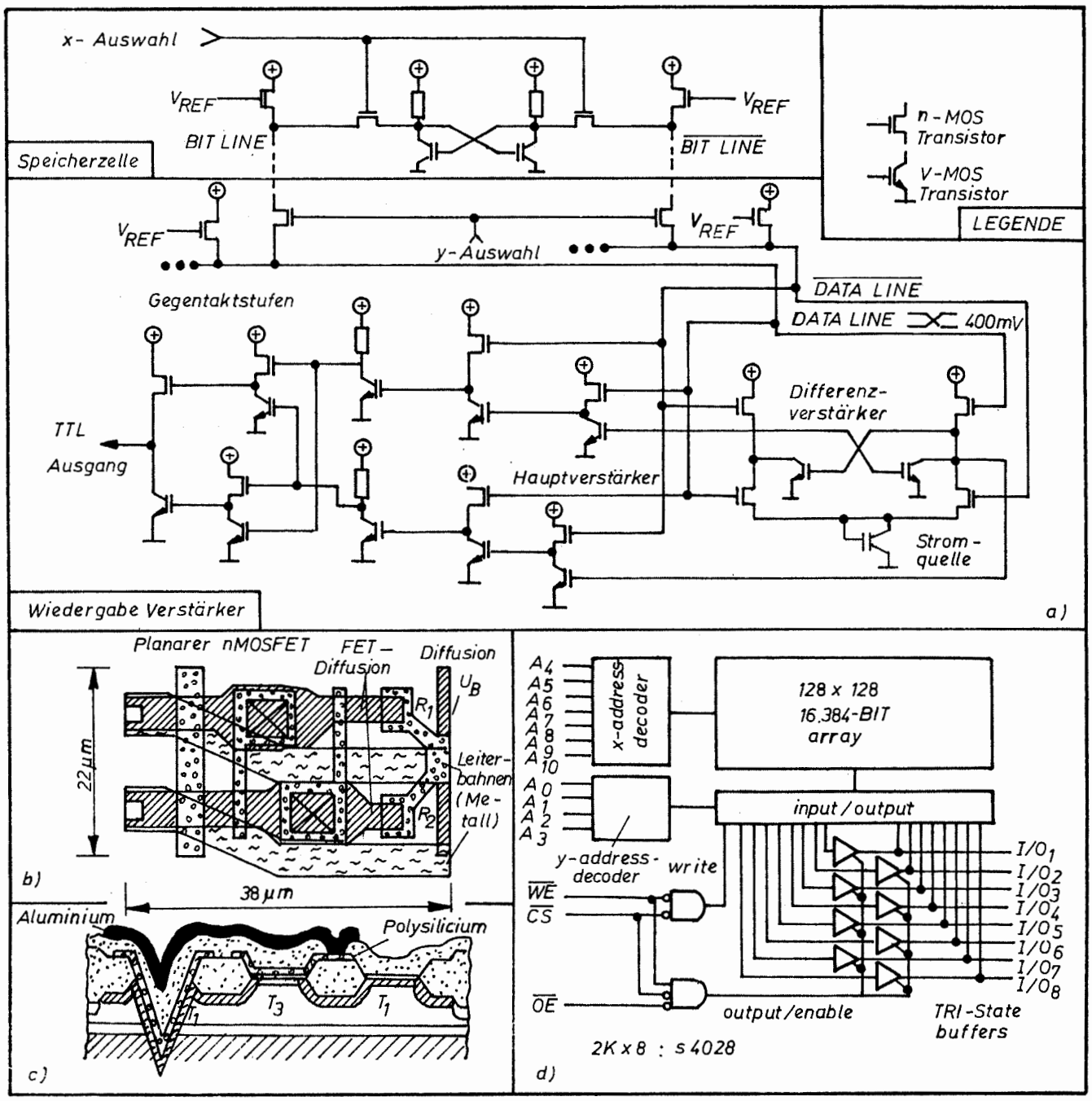


Bild 8.3.11. Beispiel eines wortorganisierten (2 K × 8 bit)-Speichers in VMOS-Technik

a) Speicherzelle und Wiedergabeverstärker (beachte die unterschiedliche Darstellung der VMOS- und der normalen Transistoren)

b) Lay-out einer Speicherzelle von oben,

c) im Schnitt

d) Prinzipschaltung des Speicherchips

des: Mit der Betriebsspannung hat die Speicherzelle aus T_1 und T_2 einen bestimmten Zustand, z.B. T_1 leitend, T_2 gesperrt. Wird jetzt die Betriebsspannung abgeschaltet, so behalten noch die Gatekapazitäten von T_1 und T_2 ihre Ladung eine Weile bei. Lediglich die Leckströme vermindern sie langsam. Wenn sie nach einer hinreichend kurzen Zeit (im ms-Bereich) wiederkehrt, sind die beiden Gateladungen noch so unsymmetrisch, daß sich sofort der richtige Zustand voll einstellt. Dabei werden die Ladungen auf den Gates voll generiert. Hierfür genügt sogar eine relativ kurze Zeit (µs-Bereich). Danach kann die Betriebsspannung wieder für längere Zeit (ms) abgeschaltet werden usw. Bei einem so gestalteten Betrieb wird viel Leistung eingespart. Der Nachteil ist aber, daß hierbei kein Schreiben und Lesen erfolgen kann. Dazu muß die Betriebs-

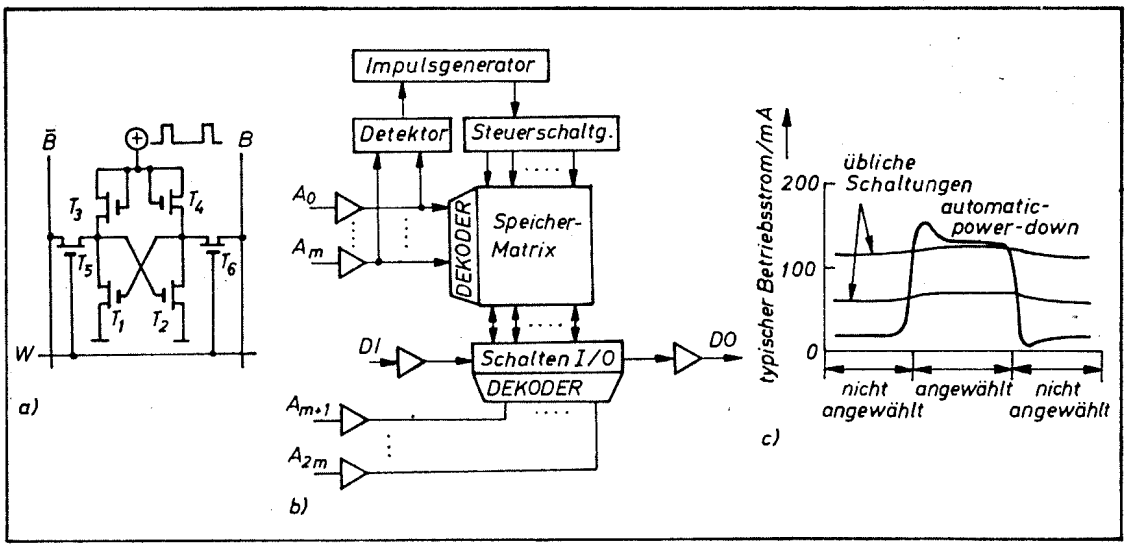


Bild 8.3.12. Energiereduzierung bei statischen MOS-Speichern

a) Durch Pulsbetrieb der Betriebsspannung

b) Prinzip des Automatic-power-down. Es werden mittels der Detektorschaltung die Adressen beobachtet. Sobald die Speichermatrix angewählt wird, wird der Impulsgenerator umgeschaltet und die Speicherzellen der Matrix erhalten soviel Energie, wie zum Schreiben bzw. Lesen notwendig ist. Dies realisiert die Steuerschaltung

c) Stromaufnahme von üblichen statischen Speichern und Speichern mit Automatic-power-down

spannung voll anliegen. So werden die beiden Zustände Betrieb und Wartestellung notwendig. Es sind auf diese Weise Leistungsverhältnisse von z.B. 500/20; 900/100 (mW) oder ähnlich zu erreichen.

Eine andere Variante des gepulsten Betriebes besteht darin, die beiden Transistoren T_3 und T_4 oder die entsprechenden Widerstände je Speicherzelle wegzulassen. Dann kann das Pulsen zum Auffrischen der Gatekapazitäten dadurch erfolgen, daß gleichzeitig beide Bitleitungen B und $\overline{B}$ auf das entsprechende Potential gehoben werden und über die Wortleitung das Aufladen realisiert wird.

Der Aufwand für die Umschaltung zwischen den Zuständen läßt sich reduzieren, wenn die Adressen beobachtet werden. Sobald eine bestimmte Gruppe von Speicherzellen über die Adresse angesprochen wird, erfolgt das Umschalten für eine bestimmte Zeit auf Betriebsmodus. Eine hierfür übliche Schaltung (automatic-power-down oder adress-activation) zeigt Bild 8.3.12b). Mit dieser Schaltung tritt zwar eine gewisse Verlängerung der Zugriffszeit ein, sie kann aber bei modernen Schaltungen im Bereich von 10 bis 100 ns gehalten werden. Ein weiterer Nachteil entsteht dadurch, daß der Betriebsstrom starken Veränderungen unterworfen ist (Bild 8.3.12c)). Insgesamt bleiben aber die Vorteile solcher Schaltungen beachtlich. Dies weisen u. a. die folgenden Leistungswerte für Betrieb/Ruhe bei relativ großen Kapazitätswerten ($>$ 8 Kbit) aus: 800/100 (mW) oder 400/40 (mW).

8.3.4.5. Dynamische Speicherzellen

Im vorigen Abschnitt wurde der Vorteil des gepulsten Betriebes gezeigt. Hierbei ist der Zustand L oder H nicht mehr durch leitende bzw. gesperrte Transistoren realisiert, sondern durch Ladungen von Kapazitäten, die allerdings infolge von Verlustströmen periodisch aufgefrischt werden müssen. Nach diesem Prinzip ist eine beträchtliche Vereinfachung der Speicherzelle möglich. Es genügt nämlich eine Gatekapazität C zur Speicherung. Durch Fortlassen von T_1, T_3 und T_4 entsteht so aus Bild 8.3.12a) die *dynamische 3-Transistor-Speicherzelle* von Bild 8.3.13a).

Mit einem Wortimpuls wird der Transistor T_5 geöffnet, und je nach dem Pegel auf der Schreibleitung B gelangt ein H oder L auf die Gatekapazität und wird dort ge-

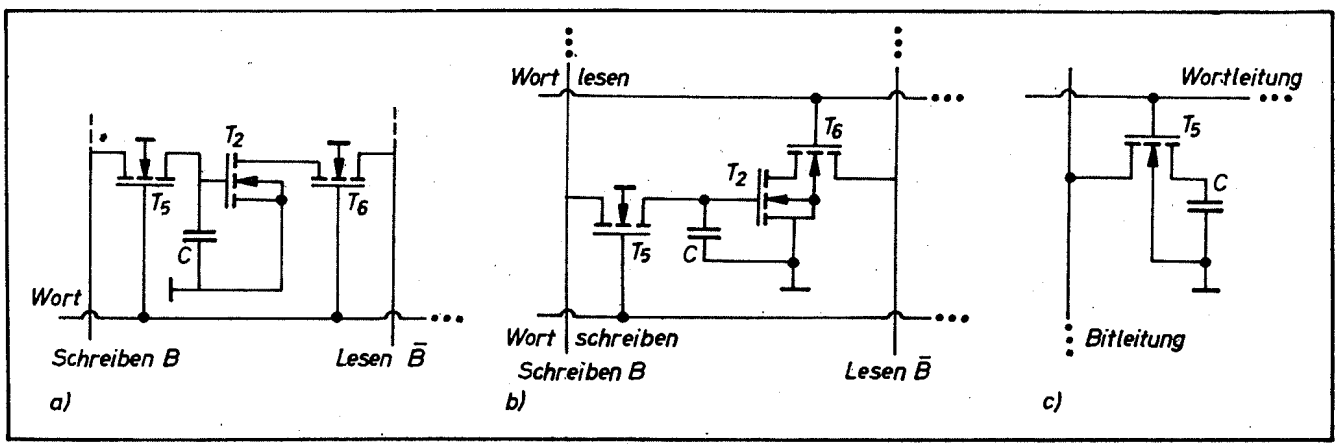


Bild 8.3.13. Dynamische Speicherzellen

a) Aus der 6-Transistor-Zelle von Bild 8.3.12a) abgeleitete 3-Transistor-Zelle

b) Trennung von Schreib- und Leseleitungen

c) Prinzip der 1-Transistor-Zelle

speichert. Es macht den Transistor T_2 gesperrt oder leitend. Dies wird jedoch erst auf der Leseleitung $\overline{B}$ wirksam, wenn vor der Wortleitung T_6 geöffnet wird. Bei dieser 3-Transistoren-Zelle ist es daher sinnvoll, die Wortleitung in zwei Teile — eine für das Schreiben und eine für das Lesen — zu zerlegen, so wie es Bild 8.3.13b) zeigt. Der Nachteil dieser Zelle ist, daß die Regeneration des Speicherzustandes nicht unmittelbar bei Wiedergabe erfolgt. Es muß hierzu vielmehr ein spezieller Refresh- bzw. Restore-Zyklus (Auffrischung bzw. Rückspeichern) vorgesehen werden, der den Kondensator neu auf die volle oder leere Ladung bringt. Hierzu sind spezielle Auffrischungsverstärker notwendig.

Die 3-Transistoren-Zelle läßt sich weiter vereinfachen, und es entsteht *eine 1-Transistor-Zelle*, wie sie Bild 8.3.13c) zeigt. Über Taktung der Wortleitung und je nach Betriebsart der Bitleitung wird entweder der Kapazität C ein H oder L zugeführt (Schreiben) oder der Zustand zurückgewonnen, gelesen. Hierbei treten Probleme für das Wiedergabesignal auf, da sich die Ladungsmenge von C bei der Wiedergabe auf die Kapazität der Bitleitung überträgt und so der Spannungspegel relativ gering werden kann. Der wesentliche Nachteil dieser Speicherzelle ist jedoch, daß ihr gespeicherter Inhalt beim Lesen zerstört wird. Es liegt also ähnlich wie beim Ferritkernspeicher DRO-Betrieb vor.

DRO stammt von destructive read out (zerstörendes Auslesen) und steht im Gegensatz zum bisher behandelten NDRO (ND = nondestruktiv, also zerstörungsfreie Wiedergabe).

Infolge des zerstörenden Lesens muß bei der 1-Transistor-Zelle anschließend in einem darauffolgenden Takt das gespeicherte Signal verstärkt wieder eingeschrieben werden. Hierzu müssen bei derartigen Speicherzellen *Auffrischverstärker mit Speicherregister* vorgesehen werden. Die Einfachheit der Speicherzelle wird also durch einen hohen Aufwand an zusätzlicher Elektronik für die Auffrischtechnik und für das kleine Lesesignal erkauft.

Die dynamische 3-Transistor-Zelle hatte bei den Halbleiterspeichern nur vorübergehend Bedeutung, z.B. im Baustein 1103 für 1 K $\times$ 1 bit. Alle heutigen dynamischen Speicher werden als 1-Transistor-Zellen ausgeführt. Im wesentlichen existieren dabei die drei Varianten, wie sie Bild 8.3.14 zeigt.

In der Einfach-Polysiliciumtechnik wird der Kondensator durch einen Polysiliciumbelag realisiert. Da hierbei ein Abstand zum FET bestehen bleibt, muß die Verbindung durch die n^+-Zone der Source S realisiert werden. Bei der Zweiebenen- oder Doppel-Polysiliciumtechnik können Gate und Kapazität teilweise überlappend gestaltet

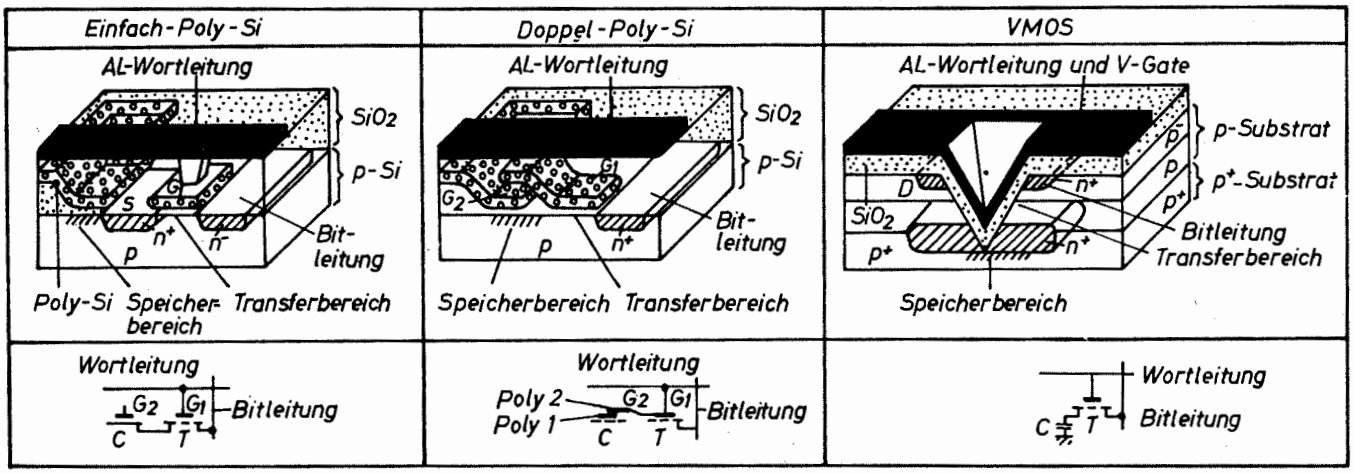


Bild 8.3.14. Drei Typen von 1-Transistor-Zellen

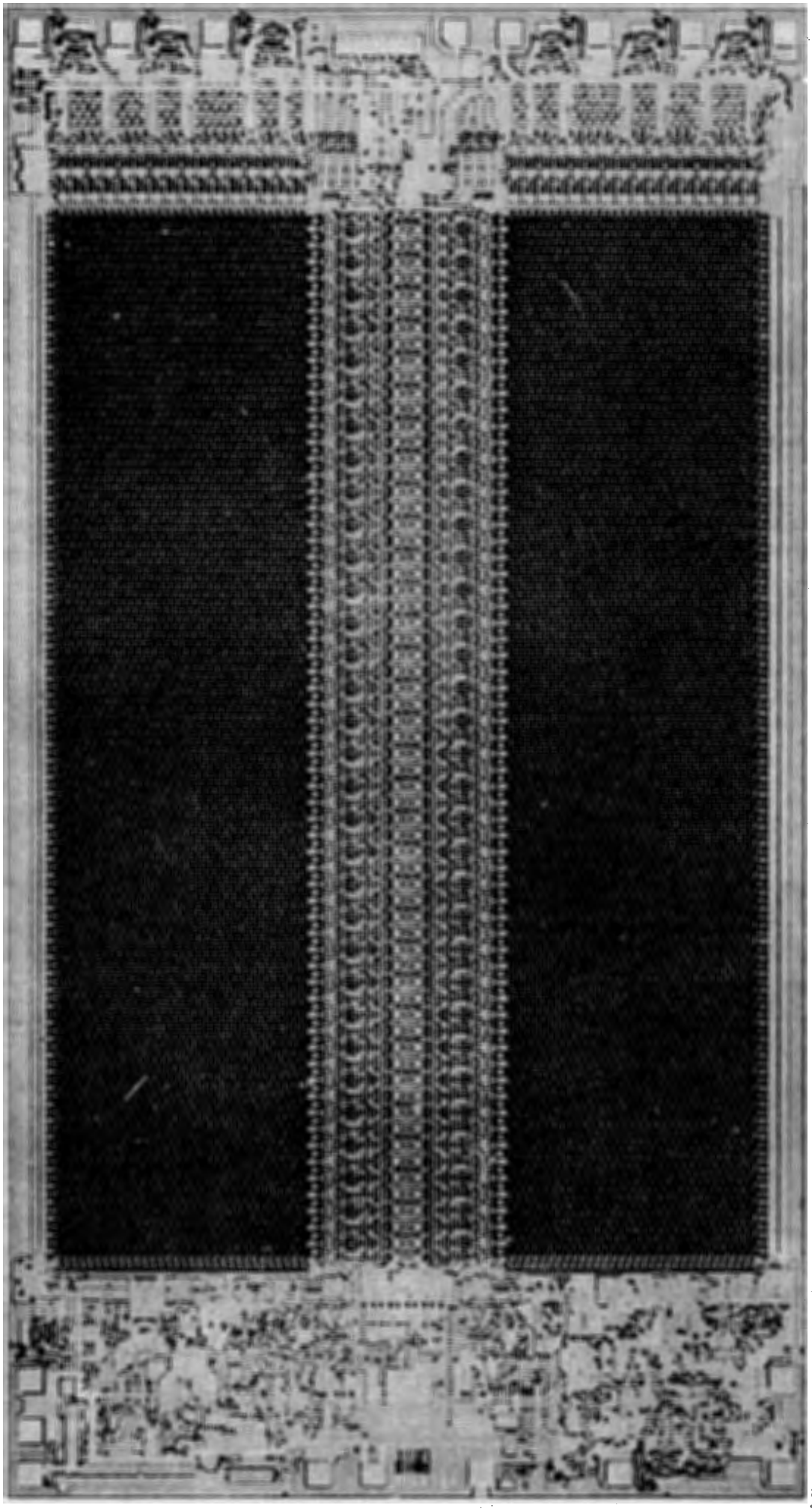
werden. Dadurch wird die Struktur kleiner und die eine n^+-Dotierung kann entfallen. Eine weitere Verringerung der Abmessungen ist bei der VMOS-Technik möglich. Hier ist einmal der FET bereits in die Tiefe verlegt, jedoch noch darunter befindet sich eine vergrabene n^+-Schicht, die zugleich Source und Kapazitätsbelag darstellt.

8.3.4.6. Dynamisches 16-K-RAM U256D bzw. 4116

Das dynamische (16 K × 1 bit)-RAM ist zum Standardspeicher vieler Mikrorechnersysteme geworden. Es enthält rund 38000 Transistoren und existiert in vielen, von Firma zu Firma etwas abweichenden Ausführungen, z. B. mit mehreren oder nur einer 5-V-Betriebsspannung und unterschiedlichen Zugriffszeiten im Bereich von ca. 100 bis 300 ns.

Eine Chipfotografie des 4116 von MOSTEK zeigt Bild 8.3.15a), während b) schematisch die interne Organisation und c) die doppelte Standardspeicherzelle (n-MOS mit zwei Polysiliciumebenen gemäß der Mitte in Bild 8.3.14) zeigt. Es wird hier also die Adreßorganisation gemäß Bild 8.3.4, Nr. 4, verwendet. Die Speichermatrix ist in zwei Teilmatrizen von je 64 × 128 bit zerlegt. Dies ist notwendig, weil in der Mitte dazwischen die Sensor- und Refreshverstärker angeordnet sind. Hierdurch werden sowohl relativ kurze Leitungen als auch viele Vorteile bezüglich der kleinen Signalwiedergabespannungen erreicht. Am äußeren Rand der Speicherzellenmatrizen liegt je Spalte eine Dummy-Speicherzelle (Blindzelle). Sie dient der Symmetrie beim Lesen. Wenn die obere Matrix gelesen wird, ist zusätzlich die Dummy-Zelle der unteren Matrix aktiviert und umgekehrt. Dies sei im folgenden beispielhaft mittels der Teilschaltung von Bild 8.3.15d) und des Zeitdiagramms von Teilbild e) beschrieben.

Gemäß Bild 8.3.15b) werden die 14 Adressen A_0 bis A_{13} in zweimal nacheinander A_0 bis A_6 bzw. A_0' bis A_6' zerlegt und über den Adressenmultiplexer mit den Befehlen CAS (column-adress-select = Spaltenauswahl) und RAS (row-adress-select = Zeilenauswahl) den Spalten- bzw. Zeilendekodern zugeführt. Beim Schreiben wird das Bit etwa gleichzeitig mit dem CAS-Befehl in die Matrix übertragen. Beim Lesen ist das dynamische Flipflop aus den Transistoren T_1 bis T_3 und die Taktsteuerung mit Φ_1 bis Φ_4 wichtig. Zu Beginn liegt das $\overline{\text{RAS}}$-Signal auf H, und im Takt Φ_1 werden die Kapazitäten der Bitleitungen C_Z sowie die der Dummy-Zelle mit $\frac{1}{2} C_S$ auf U_{DD} aufgeladen. Wenn das $\overline{\text{RAS}}$ auf L geht, erfolgt dies auch für Φ_1. In der Folge wird die Zeilenadresse übernommen, dekodiert und mittels W eine der 128 Zeilen (Wortadresse) ausgewählt. Damit erfolgt der Ladungsabgleich zwischen Speicher- und Bitleitungskapazität auf der linken Seite, also zwischen C_S und C_B am Punkt A, und rechts zwischen Dummy-



a)

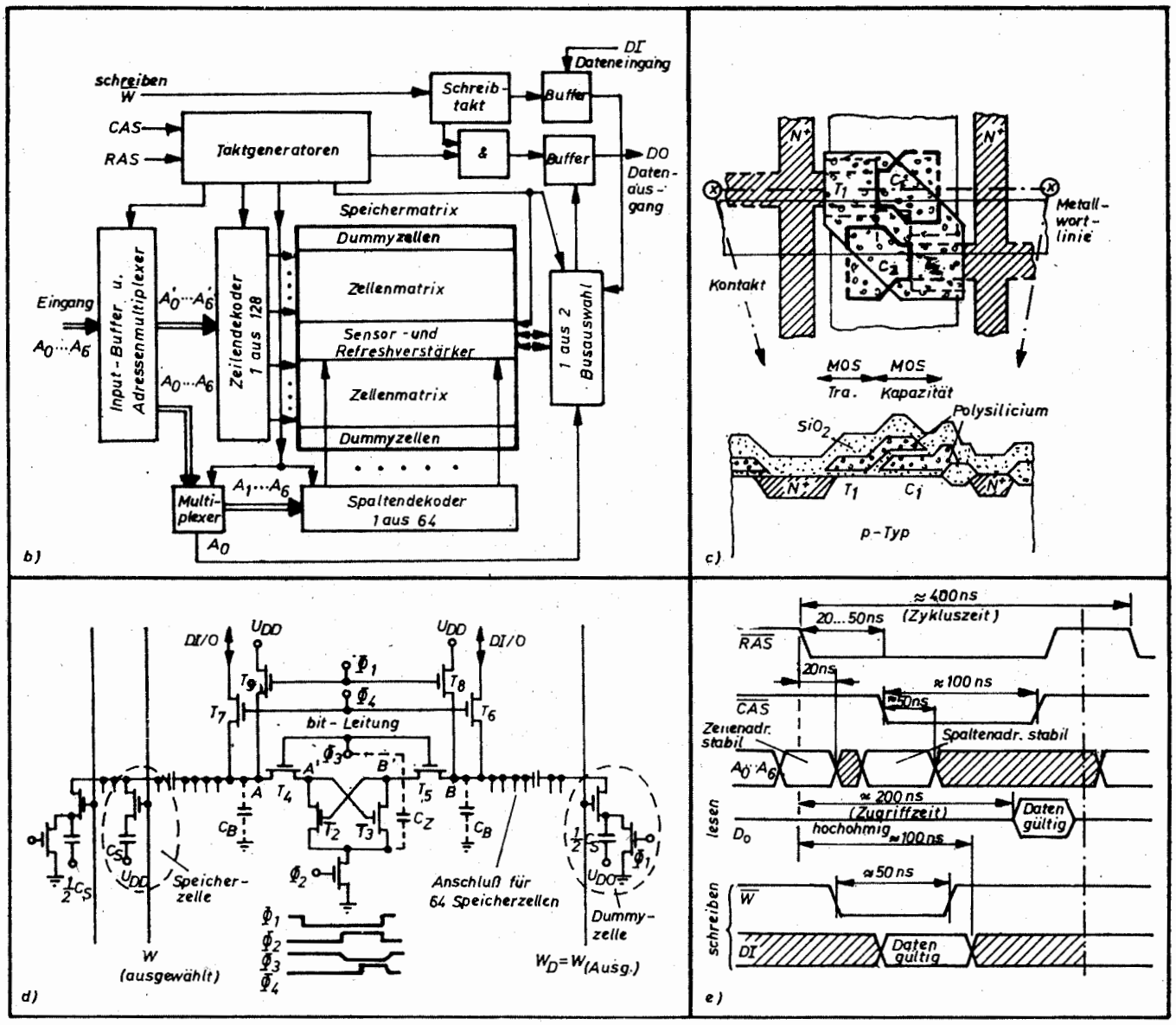


Bild 8.3.15. Dynamisches (16 K × 1)-RAM U 256 bzw. 4116

a) Chipfotografie (Seite 745)
b) Blockschaltbild
c) Standardzelle von ca. 400 μm^2
d) Prinzip des Sensorverstärkers und der Dummy-Zellen
e) Zeitbedingungen

Zellen- und Bitleitungskapazität $\frac{1}{2} C_S$ und C_B am Punkt B. Dabei entsteht ein Potentialunterschied zwischen A und B, der von der Ladung, also dem Zustand von C_S, abhängt. Dieser wird über die Transistoren T_4 und T_5 zum Flipflop an die Punkte A' und B' übertragen. Mit dem Takt Φ_2 wird der Flipflop dann in die entsprechende Lage gezwungen. Gleichzeitig wird mit C_Z etwas entladen, und die Transistoren T_4 und T_5 werden über Φ_3 hochohmiger. Hierdurch kippt das Flipflop schneller, als es gemäß den Bitleitungskapazitäten möglich wäre. Nach dem Kippen des Flipflop und dem Ladungsausgleich auf den Bitleitungen erfolgt die Auskopplung des detektierten Zustandes mittels Φ_4 über die Transistoren T_7 und T_8 und gleichzeitig das Wiederauffrischen des Speicherinhalts von C_S.

Das *Refresh* der Speicherzellen muß infolge von Leckströmen *mindestens alle* 2 ms durch einen internen Lesezyklus erfolgen. Dazu werden nacheinander alle Zeilen (alle Spalten) gleichzeitig intern über das Sensor-Flipflop ausgelesen und wieder neu eingeschrieben. Dies erfolgt allein über den $\overline{\text{RAS}}$-Eingang. Mit ihm werden nacheinander die acht niederwertigen Adressenbits angesteuert.

Für die Leistungsbilanz des RAM gilt folgende Aufstellung:

Taktgeneratoren	330 mW	60,5%
Leseverstärker	139 mW	25,5%
Ausgangstreiber	37 mW	7,0%
Übrige Schaltungen	37 mW	7,0%

Für die Flächeneinteilung gilt dagegen etwa:

Speichermatrix	45%
Leseverstärker	7%
Dekoder	15%
Taktgeneratoren	10%
Restliche Schaltungen	10%
Übrige Fläche	13%

8.3.4.7. Entwicklung zu hohen Kapazitäten, Redundanz

Gemäß Bild 8.3.3c) geht die Speicherentwicklung zu immer höheren Kapazitäten. Bereits ab 1978 entstanden die ersten dynamischen 64-K-RAM und ab etwa 1983 die ersten 256-K-RAM. Hierbei wurden und werden die Abmessungen der Speicherzellen immer kleiner und die Chipabmessungen nur wenig größer. 1982 liegen für einen 64-K-RAM die Zellen bei 150 bis 240 μm^2 und die Chipgrößen bei 25 bis 30 mm^2, während für den 256-K-RAM angestrebte Abmessungen bei 40 bis 100 μm^2 für die Zelle und 35 bis 50 mm^2 für den Chip liegen. Der 64-K-RAM konnte im wesentlichen durch Dimensionsverkleinerung (scale down) realisiert werden, während beim 256-K-RAM einige *neue Wege* beschritten werden mußten, die z. T. auch schon beim 64-K-RAM angewendet wurden. *Hierzu zählen* u. a.

- mehrfache (> 2) Aufteilung der Matrizen,
- zusätzliche Eigenschaften, wie z. B. internes automatisches Refresh (das allerdings z. T. auch abgeschaltet werden kann),
- komplizierte oder andersartige Gestaltung der Gate- und Kapazitätsstrukturen, z. B. mittels Dreifach-Polysiliciumschichten, Einsatz von Molybdän oder $MoSi_2$ bzw. Erwägen des Dielektrikums Ta_2O_5, und schließlich
- Einsatz von Redundanz, um z. B. durch Programmierung in der Wafertestphase fehlerhafte Teilstrukturen durch die redundanten zu ersetzen (Reparatur vor dem Einbau!).

Die *Idee der Reparatur* wurde bereits 1964 von IBM beschrieben und erstmals 1980 für 64-K-RAM und dann für 256-K-RAM eingesetzt. Es wurde herausgearbeitet, daß eine Reparatur nur sinnvoll ist, wenn ohnehin die Ausbeute hoch ist, und daß sie optimal durch Reservezeilen und/oder -spalten der Matrix erfolgt, da diese zuerst von Fehlern betroffen werden. Auf dem Chip werden etwa 2 bis 4 Zeilen und Spalten sowie eine zusätzliche Programmierung mit einer relativen Chipfläche von 5 bis 10% vorgesehen. Die zusätzliche Leistungsaufnahme beträgt ca. 3%, während die Zugriffszeit vernachlässigbar verlängert wird. Beim Wafertest wird untersucht, ob alle Zeilen und Spalten in Ordnung sind. Eine gestörte Zeile/Spalte wird in Adressierung durch die zu programmierende Inhibitfunktion ausgeschaltet, und die Redundanzzeile bzw. -spalte erhält durch Programmieren diese Adresse. Dies erfolgt wie bei den PROM im Abschnitt 8.3.5., jedoch während des Wafertests an einem Anschluß, der später nicht herausgeführt wird. Die Anzahl der programmierbaren Elemente ist hierbei mit ca. 20 bis 100 sehr klein.

8.3.4.8. Soft-Errors

Während um 1978 die Ursache der sogenannten Soft-Errors noch recht unsicher war, waren sie 1979 bereits weitgehend erkannt. Sie gehen auf die Bildung von α-Teilchen der Uran- oder Thorium-Zerfallsreihe zurück, die u. a. als Verunreinigungen im Keramikgehäuse, im Lötzinn oder in dem Silicium selbst enthalten sind. Sie besitzen eine typische kinetische Energie von 5,5 MeV. Insbesondere beim schrägen Einfall (Bild 8.3.16a)) erzeugen sie bei einer Eindringtiefe von ca. 25 µm rund 1,5 Millionen Ladungspaare, die sich im Feld der Speicherkapazität oder unter Bitleitungen in verschiedene Richtungen bewegen, also trennen. Den zeitlichen Stromverlauf hierzu zeigt Bild 8.3.16b). So können die gespeicherten Ladungen soweit verfälscht werden,

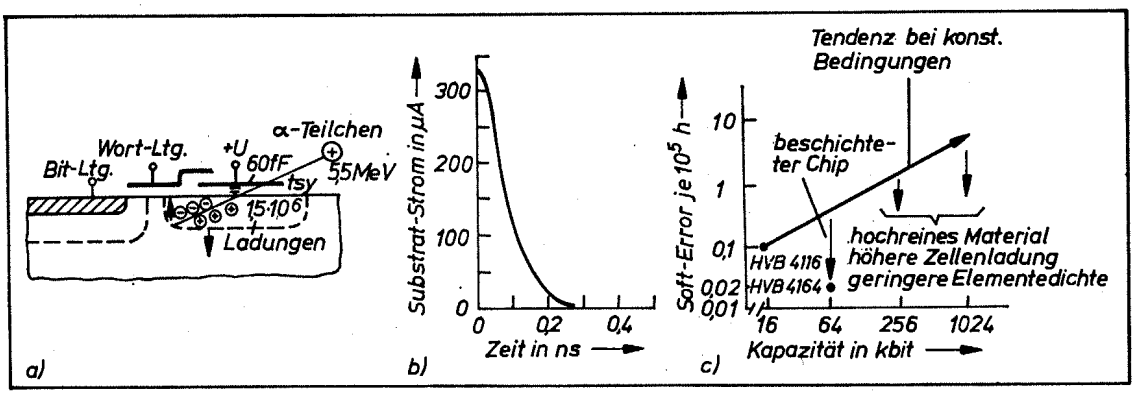


Bild 8.3.16. Soft-Errors

a) Ein α-Teilchen von ca. 5,5 MeV dringt schräg in den Bereich der Speicherkapazität ein und erzeugt so rund $1{,}5 \cdot 10^6$ Ladungspaare

b) Dadurch auftretender Substratstrom

c) Größenordnung der Soft-Error-Rate sowie Einflußmöglichkeiten auf ihre Senkung

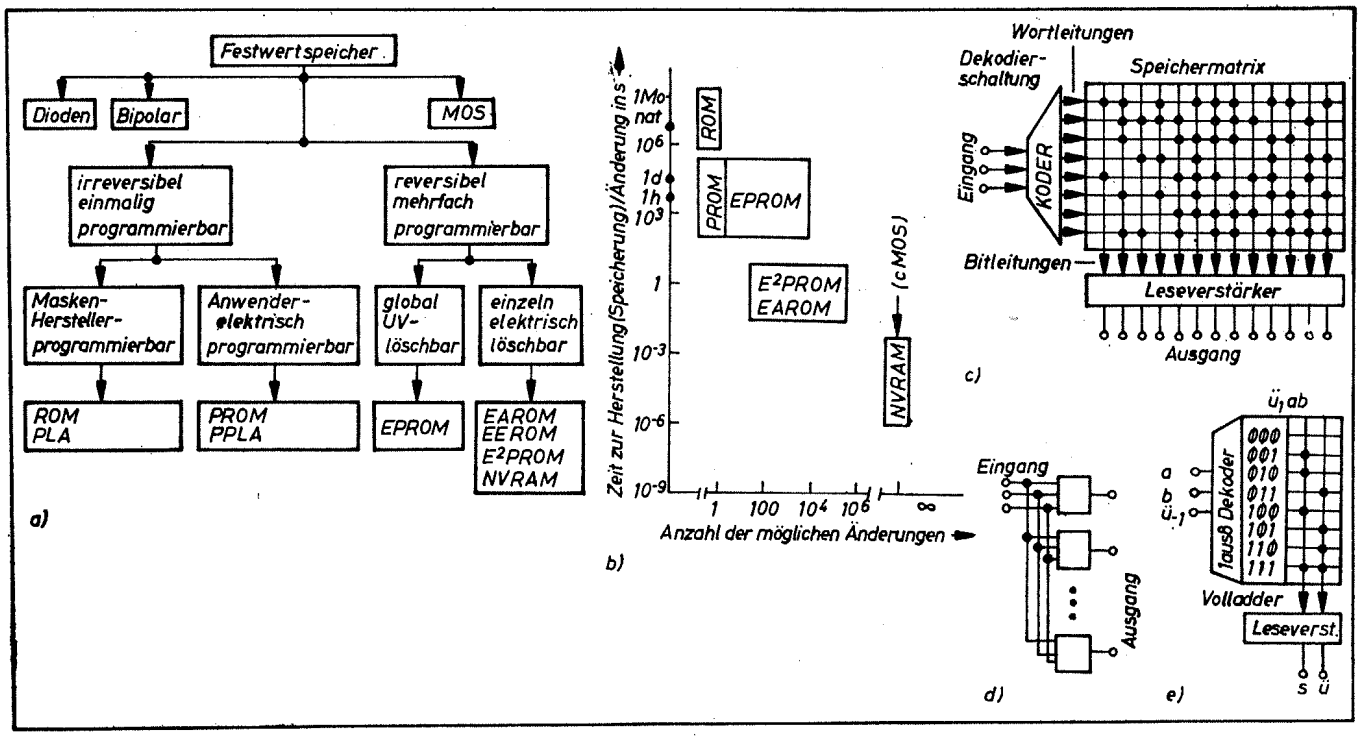


Bild 8.3.17. ROM/PROM

a) Systematische Einteilung

b) Einteilung nach zwei Eigenschaften

c) Typische Struktur

d) Äquivalente Schaltung mit logischen Bauelementen

e) Volladder als ROM

daß ein einmaliger Bitfehler auftritt. Die Häufigkeit solcher Fehler hängt von mehreren Fakten ab. Ein wichtiger ist die in der Kapazität gespeicherte Ladung. Sie hängt von der Spannung und der Kapazität ab. Mit der Strukturverkleinerung sank der Wert z.B. vom 16-K-RAM mit ca. 250 fF auf ca. 40 fF beim 256-K-RAM. Beim zu erwartenden 1 Mbit werden dann nur noch ca. 200000 Elektronen als Ladung gespeichert.

Einige typische Fehlerraten dieses Effektes zeigt Bild 8.3.17c). Daraus ist ersichtlich, daß er erst ab 16 K bzw. 64 K deutlich merkbar wurde. Nachdem er erkannt war, wurden auch verschiedene Methoden zur Reduzierung des Fehlers angewendet. Hierzu gehören u. a. Änderung des Lay-outs (z.B. gefaltete Metall-Bitleitungen), Überzug des Chips mit Materialien, die α-Teilchen absorbieren, z.B. Polyamid oder Silikon, sowie hochreine Gehäusematerialien. Dies ist auch in Bild 8.3.16c) gezeigt.

8.3.5. Programmierbare Speicher — Festwertspeicher

8.3.5.1. Überblick, Definition

Die bisher behandelten Speicher waren vom allgemeinen Typ gemäß Bild 8.3.1a). Die folgenden gehören mehr dem Prinzip des minimalen Speichers an. Sie sind also vor allem Speicher, die nur wiedergegeben werden. Sie tragen dementsprechend auch den Namen ROM (read only memory). Weiter sollen hier aber auch Speicher behandelt werden, bei denen eine einmalige oder mehrmalige Aufzeichnung bei häufiger Wiedergabe erfolgt. Hierfür wurde im Laufe der Zeit ein vielfältiges Spektrum von Arten und Typen entwickelt, deren aus typischen Eigenschaften abgeleitete Systematik Bild 8.3.17a) zeigt, während eine Zusammenstellung der entsprechenden Begriffe und ihrer Bedeutung wie folgt aussieht:

RAM (random access memory)	Speicher mit wahlfreiem Zugriff
RWM (read write memory)	Schreib-Lese-Speicher
ROM (read only memory)	Nur-Lese-Speicher
PROM (programmable ROM)	programmierbarer ROM
EPROM (erasable PROM)	löschbarer PROM
EAROM (electrically alterable ROM)	elektrisch veränderbarer ROM
EEROM (electrically erasable ROM)	elektrisch löschbarer ROM
E^2PROM (electrically erasable PROM)	elektrisch löschbarer PROM
NVRAM (non volatile RAM)	beständiger RAM
RMM (read mostly memory)	vorrangiger Lesespeicher

Zwei *wichtige Kennzeichen* sind

- die zum Programmieren notwendige Zeit,
- die mögliche Anzahl von Löschungen und neuen Programmierungen.

Das hierzu zusammengestellte Diagramm zeigt Bild 8.3.17b).

In Analogie zu Bild 8.3.4 kann ein ROM bzw. PROM im Prinzip in allen sechs dort gezeigten Möglichkeiten existieren. Meistens existiert er jedoch nur in der Form von Nr. 6. So entsteht das typische Bild eines ROM/PROM, wie es Bild 8.3.17c) zeigt. Zuweilen kann auch der Leseverstärker entfallen, wobei die Knoten durch spezielle Bauelemente zu belegen sind. Sie werden in Analogie zu den Speicherzellen der RAM in den folgenden Abschnitten in Anlehnung an Bild 8.3.17a) behandelt.

Ein in den Eigenschaften dem ROM/PROM äquivalentes Bauelement läßt sich auch durch logische Schaltungen realisieren. Dies geht aus Bild 8.3.17d) hervor. ROM/PROM sind also gegenüber logischen Schaltungen mit mehreren Ausgängen vor allem durch ihren strukturellen Aufbau gemäß Bild 8.3.17c) gekennzeichnet. Dieser wechselseitige Zusammenhang sei am Beispiel des Volladders gemäß Bild 8.2.24b)

demonstriert. Die dort gezeigte Wertetafel läßt sich direkt in einen ROM/PROM umsetzen, und es entsteht Bild 8.3.17e). Der 1-aus-8-Dekoder kann z.B. gemäß Bild 8.2.20b) oder c) realisiert sein. Aus diesem Grunde kommt es auch vor, daß spezielle ROM/PROM zuweilen als Zuordner bzw. Kodeumsetzer bezeichnet werden.

Während ein RAM beliebige, vorher nicht bekannte Daten usw. für eine bestimmte Zeit zu speichern hat, liegen für einen ROM/PROM meist andere Forderungen vor. Es gibt in der Digitaltechnik bzw. in der Rechentechnik u. a.

- Konstanten, wie π, e,
- allgemeine Daten,
- feste Wertetabellen,
- feste Programmteile,
- Kodeumsetzungen,
- Darstellungen von Zeichengestalten,

die *vor der Entwicklung festliegen und immer wieder benötigt* werden. Es ist daher sinnvoll, sie in einem Speicher einmal fest abzulegen. Sie können dann weder bei Stromausfall noch durch andere Störungen verlorengehen. Sie sind außerdem leicht verfügbar, da sie in Massen billig hergestellt werden. Hier liegt die Bedeutung der ROM/PROM. Infolge der heute verfügbaren großen Kapazitäten von mehreren Kbyte können auch ganze Programmiersprachen für spezielle Rechner in ROM bereitgestellt werden.

8.3.5.2. ROM

Gemäß Bild 8.3.17a) werden ROM vom Hersteller des Bauelements programmiert. Sie besitzen also sofort beim Verkauf bzw. bei der Anwendung in einer Schaltung die spezifische Information. Hierfür ist entscheidend, daß die Grundstruktur aller ROM einer Typenreihe identisch ist und die Information durch Variation einer oder zwei Masken spezifisch eingeprägt wird. In der Regel wird also die Belegung (der Kreuzungspunkte) in Bild 8.3.17c) durch die spezielle Maske festgelegt. Wie dies im Prinzip erfolgt, zeigt Bild 8.3.18. In Teilbild a) ist eine *Diodenmatrix* gezeigt. In jedem

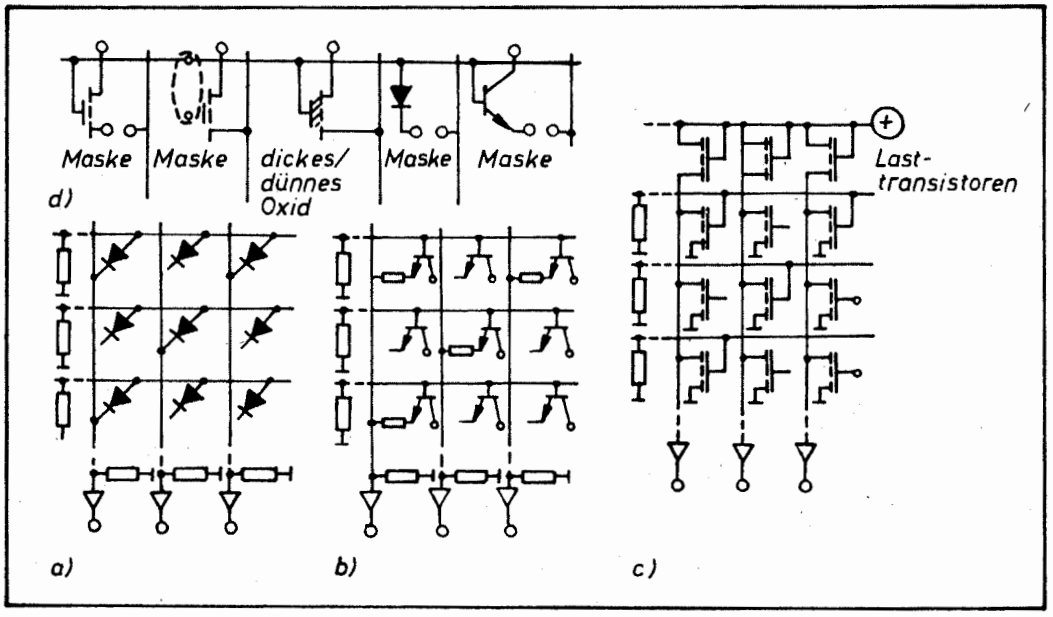


Bild 8.3.18. Aufbauprinzipien von ROM

a) als Diodenmatrix (vgl. auch Bild 8.2.3d)),
b) mit bipolaren Transistoren, c) mit MOSFET
d) Zusammenstellung der wichtigsten Maskentechniken

Knotenpunkt befindet sich eine Diode, die immer an die (waagerechten) Wortleitungen angeschlossen ist. Dies alles erfolgt durch die universellen Masken dieses ROM-Typs, die spezifische Maske, welche den Anschluß an die (senkrechten) Bitleitungen realisiert, enthält die Information dadurch, daß einige Dioden kontaktiert und andere nicht kontaktiert werden. Dies kann z.B. dadurch realisiert werden, daß Kontaktfenster erzeugt werden bzw. die Oxidisolation bestehen bleibt. Für jede spezielle Information ist also eine und nur eine zugeordnete Maske notwendig. Ihre Herstellung ist nur rentabel, wenn viele ROM gleicher Belegung benötigt werden. Die Herstellung einer solchen Maske ist teuer und dauert relativ lange. Daher der Ort der ROM in Bild 8.3.17b).

In Bild 8.3.18 sind weitere Prinzipien von ROM dargestellt. Das ROM mit bipolaren Transistoren, bei dem das *Emitterfenster* (ja oder nein) die Programmierung bewirkt (Teilbild b)); das MOS-ROM, wobei u. a. unter dem *Gate dickes Oxid* (ca. 1 µm) oder dünnes Oxid (ca. 10 nm), nicht FET:FET, die Information enthält (Teilbild c)). Andere MOS-Varianten in einer breiteren Prinzipzusammenstellung enthält Teilbild d).

Eine besondere Anwendung der ROM-Technik liegt bei den bereits im Abschnitt 8.2.8.7. besprochenen PLA vor. Hier werden zwei ROM-Matrizen — z.B. eine AND- und eine OR-Matrix — in Reihe geschaltet angewendet (Bild 8.2.26).

8.3.5.3. PROM

Gemäß der Begriffszusammenstellung im Abschnitt 8.3.5.1. kann PROM unter *zwei Gesichtspunkten* verstanden werden:

- einmal als Überbegriff für alle irgendwie programmierten ROM,
- dann speziell als einmalig, irreversibel vom Anwender (Kunden) programmierbarer Speicher.

In diesem Abschnitt wird der zweite Inhalt behandelt, auf andere PROM wird in späteren Abschnitten eingegangen.

Für die *einmalige Programmierung* gibt es zwei Varianten:

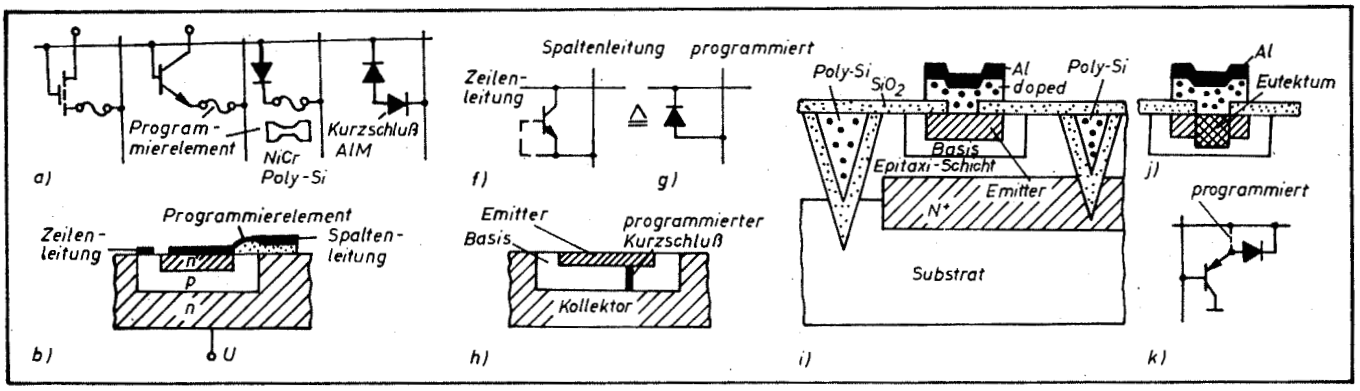
- fusable link (FL) $\triangleq$ lösbare Verbindung durch Durchbrennen, analog einer Schmelzsicherung,
- avalanche induced migration (AIM) $\triangleq$ Materialtransport (migration) durch Avalanche-Effekte.

Beim FL existiert ein sehr dünner (z.B. 2,5 µm) Streifen aus NiCr bzw. Polysilicium mit einem Widerstand von ca. 100 Ω. Mit einem kurzen Stromstoß von ca. 20 mA und wenigen µs verdampft das NiCr bzw. oxydiert Si zu SiO, SiO_2. Dadurch wird diese Leitung unterbrochen. So entstehen die drei Prinzipien von Bild 8.3.19a) links, während die Teilbilder b) die Grundstruktur, c) eine durchgebrannte NiCr-Verbindung, d) eine noch vorhandene Polysiliciumschicht und e) eine oxydierte Poly-Si-Schicht zeigen.

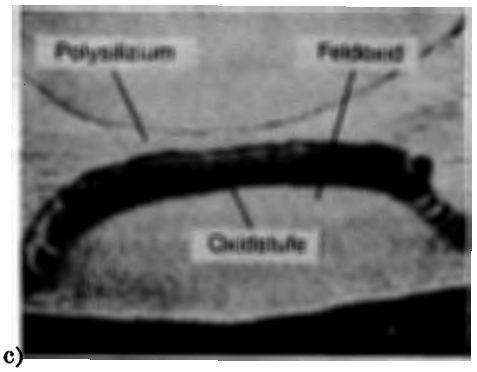
Bei diesen ROM ist zu beachten, daß der Betriebsstrom ca. 2 mA nicht übersteigt, da sonst die „Sicherung" ungewollt im Normalbetrieb durchbrennt.

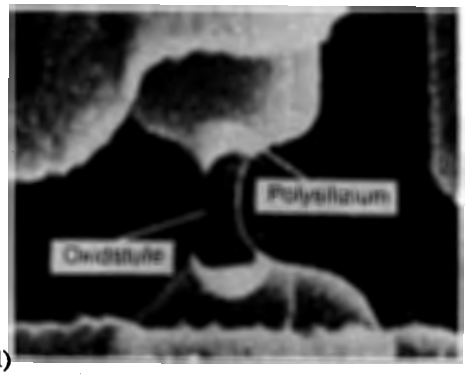
Die NiCr-Variante hat heute stark an Bedeutung verloren, da das verdampfende Metall anderweitig Störungen auf dem Chip hervorrufen kann, die bei Polysilicium ausgeschlossen sind, da es oxydiert wird.

Beim AIM-PROM liegt gemäß Bild 8.3.18f) ein Transistor zwischen Zeilen- und Spaltenleitung, wobei die Basis nicht angeschlossen ist. Der Transistor ist dadurch gesperrt ($\geqq$ 100 kΩ). Zur Programmierung wird die Kollektor-Basis-Diode in Fluß-



c)





d) e)

Bild 8.3.19. Die wichtigsten PROM-Prinzipien

a) Die Speicherzellen
b) Schnittbild eines bipolaren Transistors mit Fusable link
c) Durchgebranntes NiCr-Element
d) Poly-Si-fusable link im leitenden Zustand (Leitungsweg von rechts nach links)
e) Dito, jedoch programmiert als Si-Oxid (Leitungsweg von oben nach unten)
f) Prinzip des AIM. Gepunktete Linie wird durch Programmierung als Kurzschluß realisiert
g) Ersatzdiode nach Programmierung
h) Schema für den Kurzschluß durch Migration
i) DEAP-Element von FUJITSU (unprogrammiert)
j) Entscheidender Ausschnitt mit nach Programmierung gebildetem Eutektikum
k) Schema des DEAP-Elements

richtung und die Emitter-Basis-Diode in Sperrichtung betrieben, und zwar mit einer Spannung von mehr als 10 V, so daß ein Avalanche-Strom entsteht, der auf ca. 200 mA und einige µs begrenzt werden muß. Dieser Strom transportiert Al in die Emitter-Basis-Diode und schließt sie so kurz; es verbleibt nur die Basis-Emitter-Diode (Teilbild g) bzw. h)). Hierbei sind allerdings strenge Forderungen zu erfüllen. Sowohl zu viel wie auch zu wenig Strom mal Zeit sind störend. Deshalb muß die Programmierung intensiv mit kurzen Impulsen erfolgen, und nach jedem Impuls ist zu messen, ob der Sollwert von ca. 200 Ω erreicht ist.

Im allgemeinen wird die Meinung vertreten, daß die AIM-Speicher gegenüber FL-Speichern an Bedeutung verlieren. Doch um 1982 brachte die Firma Fujitsu eine neue Variante auf den Markt, die in Bild 8.3.19i) bis k) gezeigt ist. Oberhalb des Emitters befindet sich eine Diode Al-Poly-Si-*n*Si. Sie stellt mit dem Teiltransistor (Basis-Emitter-Strecke) einen gesperrten Übergang dar. Durch Stromfluß wird die Temperatur dieser Diode so erhöht, daß ein Eutektikum entsteht, das die Diode überbrückt. Dieses Verfahren nennt die Firma DEAP.

Während 1973 gerade ca. 1-Kbit-PROM-Speicher existierten, sind um 1982 16-K-PROM (2 K × 8 bit) bereits eingeführt. Daraus können schon beträchtliche Programmteile (Monitor usw.) realisiert werden. Die Zugriffszeiten konnten z. T. auf etwa 25 ns gesenkt werden. Gemäß Abschnitt 8.2.8.7. existieren zweistufige PROM als FPLA.

8.3.5.4. Floating gate = FG

Viele reversibel, d. h. mehrfach programmierbare ROM verwenden als Grundstruktur ein Floating gate (FG). Es befindet sich wenige nm über dem Kanal eines *n*- oder *p*-MOSFET, besteht aus Polysilicium und ist vollständig von SiO_2 isolierend umgeben, d. h., es ist zu keiner Struktur hin elektrisch verbunden. Hierdurch können dem FG auf unterschiedliche Weise Ladungen, d. h. Löcher oder Elektronen, aufgeprägt werden. *Diese Ladungen können infolge der sehr guten Isolation nicht entweichen*, bleiben also über Jahre erhalten und bestimmen so die Leitfähigkeit des darunterliegenden Kanals. Für digitale Speicher interessieren nur die beiden Zustände leitend und nicht leitend. Entscheidend für die Funktion des FG ist nun, daß es *vier Effekte* gibt, mittels derer der *Ladungszustand* des FG *verändert* werden kann:

- AI = *avalanche injection* = Lawineninjektion. Der *pn*-Übergang des FET wird im Durchbruchbereich betrieben. Dadurch entstehen viele heiße (schnelle, energiereiche) Elektronen und Löcher. Insbesondere können so die Elektronen die Barriere Si-SiO_2 mit etwa 3,25 eV überwinden und auf das Gate gelangen. Das Gate wird also mittels AI negativer geladen.
- KI = *Kanalinjektion*. Bei Kurzkanal-FET ($L < 5\,\mu m$) entstehen in einigen Betriebsfällen heiße Elektronen. Sie können so auf das Gate gelangen, das mittels KI also negativer wird.
- Ex = *Extraktion*. Dieser Effekt erfordert ein zusätzliches Gate in Nachbarschaft zum FG. Wenn dieses zweite Gate bei positiver Spannung eine hinreichend hohe Feldstärke (einige MV/cm) gegenüber dem FG erzeugt, fließen Elektronen vom FG zum Zusatzgate. Das FG wird durch Ex also positiver geladen.
- UV = *UV-Strahlung*. Mittels Fotoeffekten können überzählige Elektronen vom FG abfließen. Hierdurch ist es möglich, das FG in einen ungeladenen Zustand zu überführen.

Diese Effekte bzw. einige von ihnen können für unterschiedlich reversible PROM genutzt werden. Hierauf wird in den folgenden Abschnitten eingegangen.

Auf der Basis des FG existieren einige, meist relativ ähnliche MOS-Transistoren, die mit bestimmten Namen belegt sind, z. B.:

FAMOS = Floating-gate-avalanche-injection-MOS
SAMOS = Stacked-gate-avalanche-injection-MOS
SIMOS = Stacked-gate-injection-MOS
ATMOS = Adjustable-threshold-MOS
DIFMOS = Dual-injector-floating-gate-MOS.

8.3.5.5. EPROM

Ein EPROM ist gemäß Bild 8.3.17a) dadurch gekennzeichnet, daß es global für alle bit *mit UV-Licht* gelöscht wird. Die ersten EPROM (um 1972) waren *p*-Kanal-Typen (Bild 8.3.20a) und b)). Heute werden fast ausschließlich *n*-Kanal-Typen (Bild 8.3.20c)) verwendet. Das Gate wird mittels Avalanche geladen. Dabei wird bei den *p*-Typen der Kanal leitend und bei den *n*-Typen gesperrt. Dies bedeutet, daß die *p*-Typen im gelöschten Zustand bei positiver Logik eine Ø und die *n*-Typen eine 1 repräsentieren. Das Programmieren erzeugt also 1 bzw. Ø. Die ersten EPROM benötigten drei Betriebsspannungen, z. B. +5 V; −5 V; +12 V bzw. +12 V; −12 V; −5 V, und eine zusätzliche Spannung beim Programmieren. Die jetzigen EPROM kommen mit einer Betriebsspannung von 5 V aus. Im praktischen Aufbau befindet sich der komplette Chip unter einem UV-durchlässigen Quarzfenster, was zusätzlichen Aufwand und

damit höhere Preise bedingt. Gelöscht wird ein EPROM meist mit einer Wellenlänge von ca. 250 nm (2500 Å), einer Energiedichte um 10 W/cm² und in einer Zeit von ca. 30 min (Bild 8.3.20d)). Beim Löschen entsteht dadurch soviel Wärme, daß empfohlen wird, vor dem Programmieren das EPROM erst abkühlen zu lassen. Da beim Programmieren mit ca. 25 V und 2 mA ebenfalls Wärme entsteht, muß jede Bitzelle oft mehrmals nacheinander mit kurzen Impulsen von ca. 50 ms programmiert werden. Das Programmieren eines 16-K-EPROM dauert ca. 2 Minuten, beim 128-K ca. 30 Minuten.

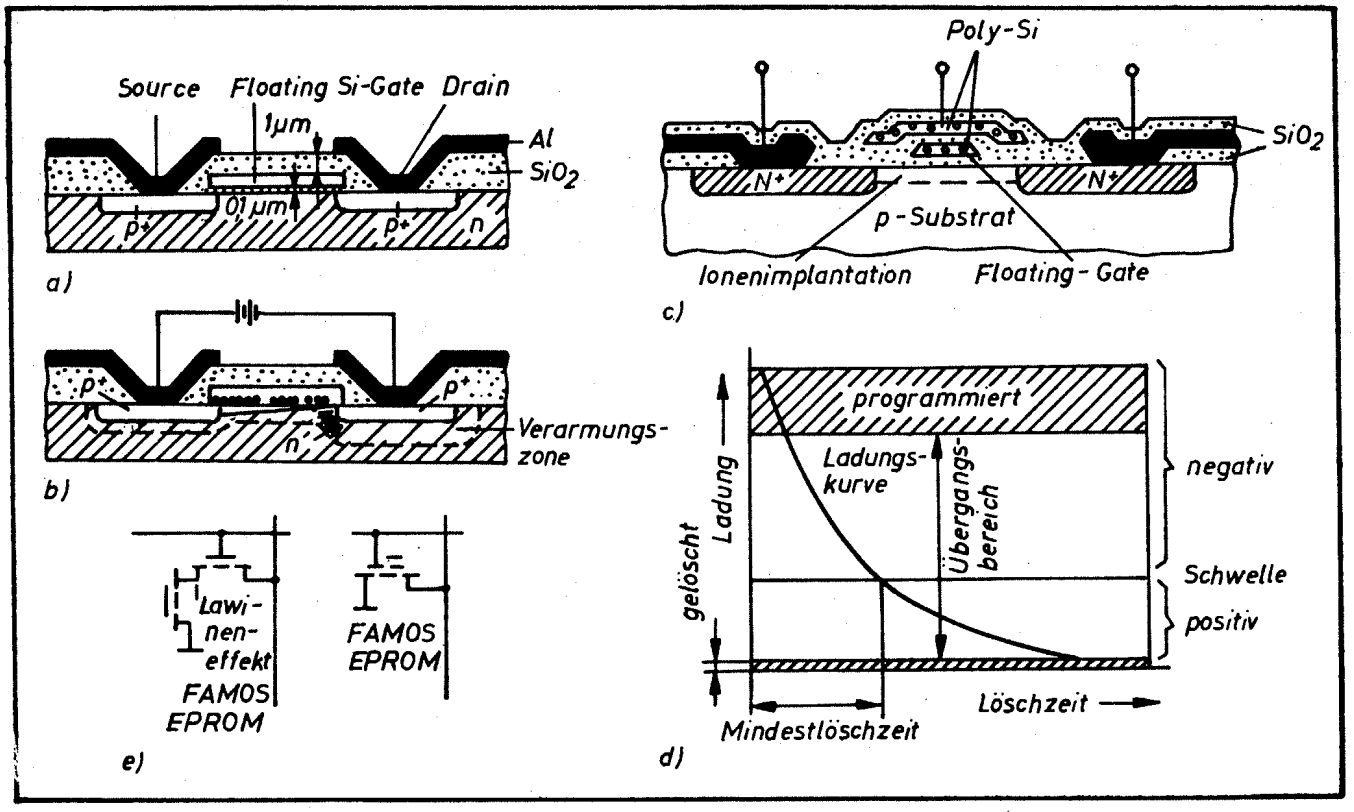


Bild 8.3.20. EPROM-Prinzipien

a) Aufbau von *p*-Kanal-EPROM
b) Programmiervorgang beim *p*-Kanal-EPROM
c) Aufbau des *n*-Kanal-EPROM. Das obere zweite Gate führt zur Auswahlleitung
d) Schematische Darstellung des Löschvorgangs mit Ladungsgrenzen für das Floating gate
e) Zwei Möglichkeiten der Speicherzellen von EPROM

Ohne Lichteinfluß und Röntgenstrahlung kann eine programmierte Ladung zwischen 10 (Garantiewert) und 1000 Jahren (unterschiedliche Herstellerangaben) beständig sein. Der übliche Leckverlust soll unter 1 Elektron je Jahr liegen. Der Leckstrom wird auch zu 10^{-36} A/m² angegeben. Löschung erfolgt bei einer Wellenlänge von kleiner als 400 nm. Direkte Sonneneinstrahlung kann ein EPROM innerhalb von ca. drei Tagen teilweise löschen, eine Leuchtstoffröhre kann dies in etwa drei Wochen bewirken. Deshalb wird empfohlen, das Quarzfenster nach der Programmierung mit schwarzem Klebeband zuzukleben. Röntgenstrahlung bewirkt bei einer Dosis von ca. 10 rad erste Bitausfälle.

Den Aufbau von Speicherzellen zeigt Bild 8.3.20e). Es sind zwei Varianten zu unterscheiden: Entweder wird das Gate über dem FG verwendet oder ein zusätzlicher FET für die Anschlüsse der Auswahlleitungen.

8.3.5.6. E²PROM

Hier werden *elektrisch programmierbare Speicherzellen* mit Floating gate (FG) gemäß Bild 8.3.21 behandelt; jenen mit Haftstellen (EAROM) ist der nächste Abschnitt gewidmet.

Die SAMOS-Zelle gemäß Bild 8.3.21a) verwendet je einen p-Kanal für die Speicherzelle und die Auswahlleitung. Bei einer Gate-1-Spannung von ca. +20 ... +30 V und normalem Source-Drain-Betrieb wird das FG mittels AI negativ geladen und so die Zelle programmiert. Mit ca. +70 V und ohne Source-Drain-Spannung kann dagegen das FG mittels Ex (vgl. 8.3.5.4.) negativer gemacht werden, indem Elektronen zum Gate 1 übergehen.

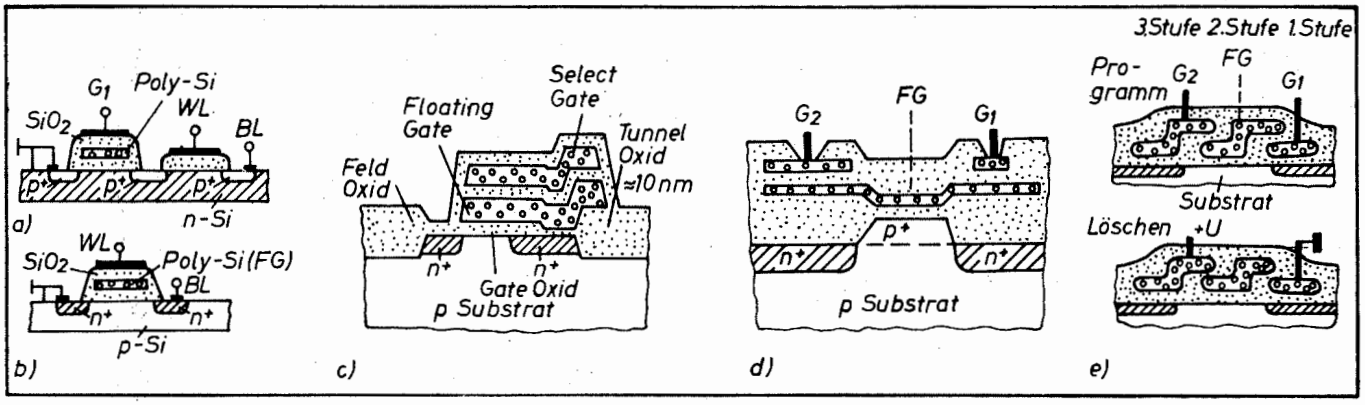


Bild 8.3.21. E²PROM-Prinzipien mit Floating gate

a) SAMOS-Speicherzelle (p-Kanal) mit zusätzlichem FET
b) SIMOS-Speicherzelle (n-Kanal)
c) Zelle mit Tunnelzone, sonst ähnlich b)
d) Doppelgatezelle (alte Prinzipausführung)
e) Neuer Typ mit drei Poly-Si-Stufen und Programmierungs- sowie Löschprinzip

Die SIMOS-Zelle (Teilbild b)) kann sowohl elektrisch als auch mittels UV-Licht gelöscht werden. Zur elektrischen Löschung werden hier die Elektronen des FG zum Source-Bereich abgeführt (WL = OV). Dieser Effekt wird verbessert durch ein extrem dünnes Tunneloxid gemäß Teilbild c). Diese Struktur ist u. a. typisch für das 4-Kbyte-E²PROM 2816 von Intel. Die Betriebsspannung dieses und ähnlicher Typen beträgt 5 V, zum Programmieren und Löschen werden 21 V bei max. 15 mA für einige ms und 8 bit = 1 byte benötigt. Im aktiven Betrieb werden 500 mW, im Standby-Modus nur 165 mW Leistung benötigt. Ein Byte kann nur maximal ca. 10000mal umprogrammiert werden. Dies dürfte für die meisten Anwendungen jedoch genügen.

Bild 8.3.21 zeigt eine Variante mit zwei zusätzlichen Steuergates. Die Aufladung des FG erfolgt, wenn beide Gates (1 und 2) an eine positive Spannung gelegt werden. Beim Löschen wird G_2 an Masse gelegt. Durch diese unterschiedliche Beschaltung der beiden Gates ist also ein Umprogrammieren von FG möglich. Das gleiche Prinzip mit zwei Steuergates und dem FG existiert auch in anderen Ausführungen, z.B. in einem Dreiebenen-Poly-Si-Prozeß (Bild 8.3.21e)).

Der Hauptvorteil der E²PROM liegt im Entfallen des Quarzfensters und darin, daß sie weder zum Löschen noch zum Umprogrammieren aus dem Sockel herausgenommen werden müssen. Sie sind jedoch sehr teuer und nur für kleine Kapazitäten verfügbar.

8.3.5.7. EAROM bzw. MNOS

Bei den bisher behandelten E²PROM war das Floating gate der Ladungsträger, mit dessen Hilfe ein MOSFET beständig in die beiden Zustände leitend — nichtleitend überführt werden kann. Es besteht aber auch die Möglichkeit, hierfür Haftstellen an Grenzstellen zu verwenden. So entstand 1976 der erste 256-bit-MNOS-EAROM (metal nitrid oxid semiconductor). Dieses Prinzip ist strukturell in Bild 8.3.22a) erklärt. Entscheidend ist bei ihm der Schichtaufbau der Gateisolation. Sie besteht aus einer sehr dünnen (1 ... 2 nm) SiO_2-Schicht und einer darüberliegenden normalen, d. h. dickeren (50 ... 70 nm) Si_3N_4-Schicht. Dadurch, daß beide Gitter nicht genau aufeinander passen, entstehen an der Grenzschicht Haftstellen mit Donatorencharakter (Teilbild b)). Die hier eingefangenen positiven Ladungsträger wirken mit der Spannung am Gate gemeinsam auf das Kanalgebiet ein. Je nachdem, ob die Haftstellen neutral oder besetzt sind, entsteht also eine andere Schwellspannung für den MNOS-FET (Teilbild c)). Durch eine hohe Gate-Substrat-Spannung von ca. 30 bis 40 Volt kann die Ladung der Haftstellen über den Tunneleffekt verändert werden (Teilbild d)). Sie muß ca. 50 µs angelegt werden. Damit kann der MOSFET in zwei Zustände gebracht werden, die eine relativ hohe Stabilität von ca. 10^4 Stunden aufweisen. Um sie deutlich zu unterscheiden, liegt die Gateelektrode bei der Wiedergabe auf etwa 6 bis 8 V. Es liegt ein gewisser Einfluß der Leseoperationen vor. Nach ca. 10^{11} Lesezyklen kann die Ladung der Haftstellen soweit verändert sein, daß nicht mehr eindeutig Ø und 1 zu unterscheiden ist. Ähnlich besteht eine begrenzte Anzahl von maximal ca. 10^4 bis 10^6 Schreiboperationen.

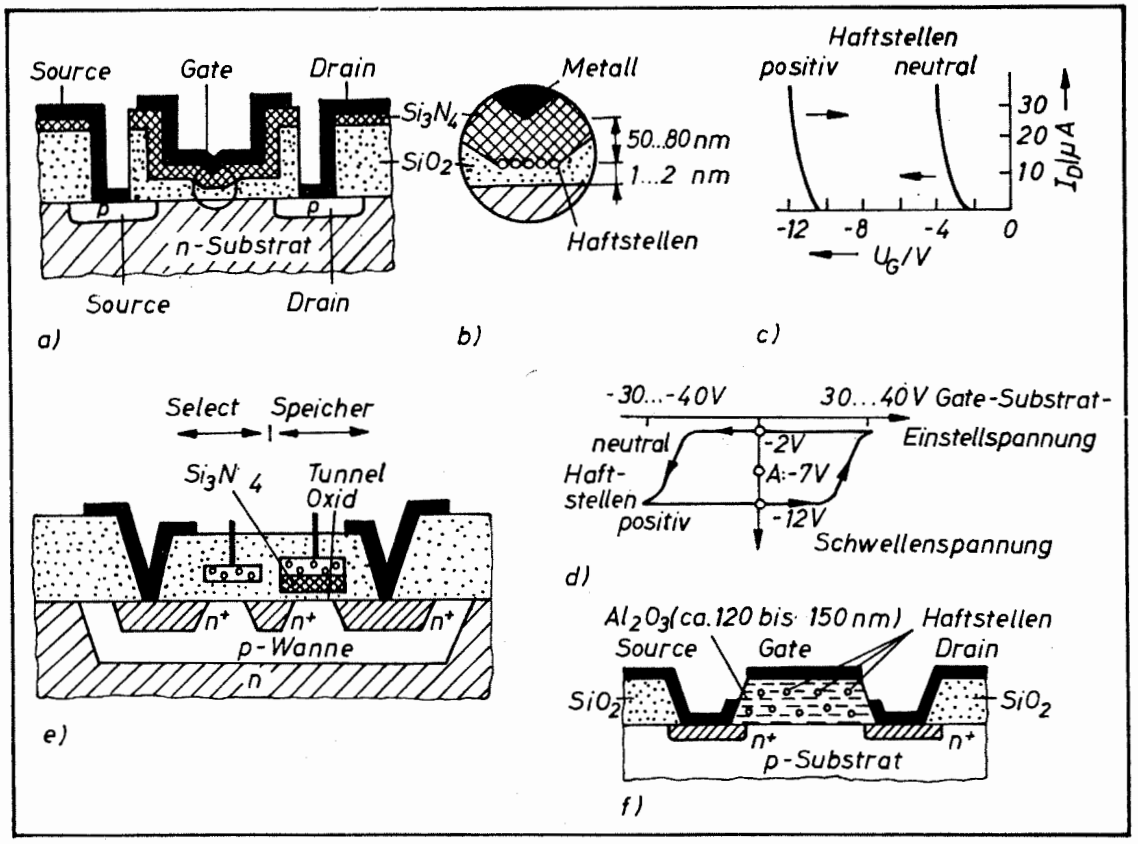


Bild 8.3.22. Beispiele von EAROM

a) *p*-Kanal-EAROM
b) Vergrößerter Ausschnitt aus der Umgebung des Gate
c) Kennlinien je nach Besetzung der Haftstellen
d) Schreiben und Löschen der Haftstelle als Hysteresekurve
e) Moderner MNOS-EAROM mit *n*-Kanal und Poly-Si-Gate und Auswahl- sowie Speicher-FET
f) MASFET für EAROM

Eine Verbesserung konnte auch hier durch die n-Kanal-Technik gemäß Bild 8.3.22e) erreicht werden. Der Arbeitspunkt A ist von -7 V auf ziemlich genau ± 0 V festlegbar. Dadurch entfällt die Begrenzung der Lesezyklen. Weiter ist es bequem möglich, den Zusatztransistor für die Zellenanwahl in die gleiche p-Wanne zu legen. Dadurch wird die Zelle sehr klein, so daß Kapazitäten um 16 Kbit erreicht werden. Die Schreiboperation für ein Byte dauert etwa 10 ms, während die Zugriffszeit beim Lesen nur 300 ns beträgt. Die Betriebsspannung konnte auf 5 V eingestellt werden. Die Ladung an den Haftstellen besteht länger als 10 Jahre.

Eine andere Variante mit Haftstellen im gesamten Material ist der MASFET (metal aluminia semiconductor). Sein Schema zeigt Bild 8.3.22f). Er konnte bisher keine größere Bedeutung erlangen.

8.3.5.8. NVRAM

Der Begriff NVRAM wird z. T. unterschiedlich gebraucht. Zuweilen werden hierunter auch einige E²PROM, z. B. gemäß Bild 8.3.22d), gezählt. Im Prinzip könnten darunter auch übliche RAM aufgeführt werden, denen durch Unsymmetrie der beiden Latch-Transistoren indirekt ein ROM überlagert ist. Schließlich gewinnen in der letzten Zeit Bausteine ähnlich dem 2817 von INTEL einige Bedeutung. Sie enthalten ein kleines RAM und ein größeres E²PROM bzw. EAROM. Die Änderung wird zunächst schnell in das RAM eingespeichert und dann langsamer an die gewünschte Stelle des E²PROM übertragen. Solche Bausteine heißen auch richtiger RAM-E²PROM. Ein völlig anders geartetes NVRAM wurde 1982 von der Firma Xicor vorgestellt. Es enthält in jeder Speicherzelle ein statisches RAM und ein E²PROM. Im Normalbetrieb wird mit dem schnellen RAM gearbeitet (Schreiben und Lesen). Im Hintergrund, zeitlich langsamer und daher verzögert, wird in eine E²PROM-Zelle gemäß Bild 8.3.21e) die gerade im RAM existierende Information übernommen. Solche Speicherzellen lassen zum System hin nur das RAM mit seinen Vorteilen bezüglich der kleinen Zugriffszeit in Erscheinung treten. Sobald aber das System abgeschaltet wird bzw. zusammenbricht, bleibt der faktisch letzte Zustand des RAM (nämlich der von ca. 10 ... 50 ms zuvor) im E²PROM gespeichert.

8.3.6. Umlaufspeicher

Bei einem Umlaufspeicher ist die Information in ständiger Bewegung im Speichermaterial. Der Zugriff zur Information erfolgt an einem speziellen Punkt und kann daher nur zeitlich nacheinander beim Vorbeilaufen der Information, also seriell, erfolgen. Deshalb heißen solche Speicher auch

SAM = serial access memory = Serielle Speicher .

Durch den Umlauf wird die Information natürlich gedämpft und muß nach einer gewissen Zeit regeneriert werden. Im Grunde wird bei diesem Speicherprinzip die Laufzeitverzögerung durch ständige periodische Wiederholung ausgenutzt. *Als Laufzeitspeicher können u. a. angewendet werden*

- akustische Wellen in Materialien
- CCD-Strukturen
- magnetische Bubbles.

Akustische Wellen hatten in den Anfängen der Speichertechnik eine gewisse Bedeutung. Die OFW gemäß Abschnitt 6.3.5. haben für die Speicherung keine Bedeutung

erlangt. Deshalb werden in den beiden folgenden Abschnitten nur die CCD-Speicher und Bubbles behandelt.

8.3.6.1. CCD-Speicher

Das Grundprinzip der CCD ist im Zusammenhang mit Bild 3.3.10 behandelt, während Abschnitt 6.1.5. einige Grundprinzipien und vor allem die CCD-Lichtempfänger-Matrizen behandelte. Einen allgemeineren Überblick der Ladungstransfer-Bauelemente (CTD), einschließlich der verwandten BBC, enthielt weiter Abschnitt 7.8. Für die Speicherung sind jedoch nur die CCD wichtig, denn sie besitzen für Frequenzen größer als 100 kHz den geringsten Leitungsbedarf. Darüber hinaus zeichnen sie sich durch größte technologische Einfachheit (keine *pn*-Übergänge) und höchste Packungsdichte aus. Ihr Nachteil ist aber, daß von Stufe zu Stufe der Signalpegel abnimmt. Deshalb ist nach einer gewissen Stufenzahl eine Regeneration durch Verstärker (refresh-amplifier) notwendig. Die Stufenzahl zwischen zwei Regenerationsverstärkern N_R hängt vor allem von der technologischen Perfektion und der Taktfrequenz ab. Wurden anfänglich nur 5 bis 30 Bitlängen erreicht, so werden heute bis zu 500 angewendet.

Um ein Optimum an Speicherkapazität und Zugriffszeit zu erhalten, sind vier Organisationsformen für CCD-Speicher entstanden. Ihr Prinzip zeigt Bild 8.3.23.

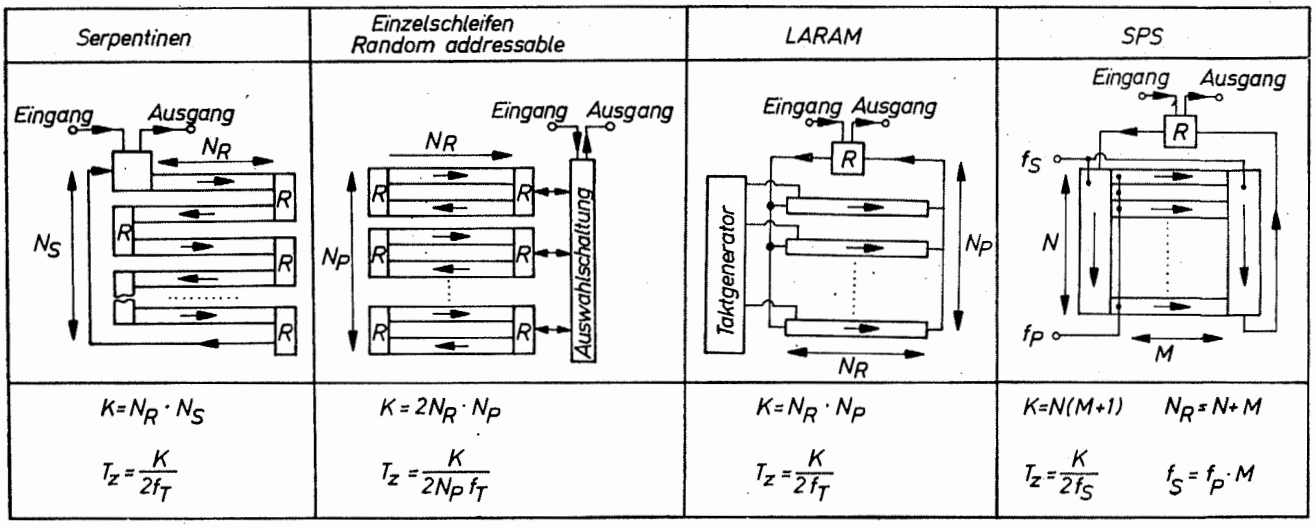


Bild 8.3.23. Organisationsprinzipien von seriellen Spe. hern

Es bedeuten R Regenerationsverstärker, N_R Stufenzahl für Regeneration, N_P Anzahl paralleler Schleifen, K Kapazität, T_Z Zugriffszeit, f_T Taktfrequenz. Die Steuerung der Auswahl der bit wurde zur Vereinfachung weggelassen. Die Anordnung mit Einzelschleifen wird zuweilen auch BORAM (= Block-orientierter RAM) genannt.

Die Grundform der CCD-Speicherschaltungen ist die Anordnung in Serpentinen. Sie läßt die CCD-Ketten relativ dicht anordnen und fügt im notwendigen Abstand die Regenerationsverstärker ein. So können relativ lange Register realisiert werden. Sie werden von einem Takt gesteuert. Würde ein Speicher nur aus solch einer Anordnung aufgebaut, so hätte er stets eine sehr lange Zugriffszeit. Deshalb erfolgt meist zusätzlich eine Parallelorganisation. Bei zu Schleifen (engl. loop) verkürzten Serpentinen zeigt dies Teilbild b). Ein solcher Speicher heißt synchronously clocked random addressable loop memory. Mit der Auswahlschaltung wird entschieden, welche Schleife für das Schreiben oder Lesen ausgewählt wird. Gegenüber der Serpentinenschaltung wird bei gleicher Kapazität die Zugriffszeit um den Faktor der parallel betriebenen Schleifen N_P verringert.

Bei der Schaltung 8.3.23 b) ist es möglich, mit zwei Takten zu arbeiten. Einen Takt T_R für die Schleifen, die nur regeneriert zu werden brauchen. Er wird möglichst

niedrig gewählt, damit die Leistungsaufnahme gering bleibt. Der zweite Takt T_A wird möglichst hochfrequent gewählt und dem ausgewählten Register zugeführt, um eine möglichst geringe Zugriffszeit zu erreichen. Der so erhaltene Vorteil wird durch erhöhten Aufwand erkauft.

Eine Weiterentwicklung dieses Prinzips führt zum LARAM (line addressable random access memory). Hierbei wird immer nur ein Register mit dem hochfrequenten Takt versorgt. Die anderen Register ruhen während dieser Zeit vollständig. Es muß also in sehr schneller Folge die Information jedes Registers genau einmal zur Regeneration über den Regenerationsverstärker bewegt werden. Die Folge muß so dicht sein, daß in den einzelnen Registern während der Ruhepause noch kein Verlust eintritt. Bei sehr geringer Leistung wird so nur ein Regenerationsverstärker benötigt. Dafür ist aber der Taktgenerator komplizierter.

Ein anderes Prinzip ist der SPS-Speicher (serial-parallel-series memory). Hier werden N Schieberegister parallel mit einem Paralleltakt der Frequenz f_P betrieben. Sie werden verkoppelt über zwei Serienregister mit der Taktfrequenz

$$f_S = f_P M \tag{6}$$

und N Speicherplätzen. Jede Information durchläuft die M Speicherplätze der Parallelregister und, aufgeteilt auf die beiden Serienregister, noch einmal N Plätze. Für die Regenerationslänge gilt daher

$$N_R = M + N\,. \tag{7}$$

In dem Serienregister werden hier also die Informationen der Parallelregister ineinander verschachtelt. In der überwiegenden Fläche des Speichers, nämlich in den Parallelregistern, wird die Information langsam bewegt. Dadurch ergibt sich ein guter Kompromiß bei dieser Schaltung. Sie ähnelt in der Struktur einer Schaltung der CCD für die Bildwandler (vgl. Bild 6.1.33c)). In vielen Speichern wird dieses Grundprinzip mit einem zusätzlichen Parallelbetrieb mehrerer solcher Anordnungen verwendet. Derartige Speicher heißen dann SPS random addressable memory (SPS-RAM).

Ende der 70er Jahre wurde den CCD-Speichern eine große Anwendungsbreite vorausgesagt. Damals lagen Speichertypen von 16 und 64 Kbit vor. Eine wesentliche Weiterentwicklung ist seitdem wohl nicht erfolgt. Es sei deshalb kurz der Intel-Speicher 2416 beschrieben. Er verwendet das Prinzip der Einzelschleifen von Bild 8.3.23b). Es werden 64 Schleifen mit je 256 bit benutzt. Der Umlauf einer Schleife wird mit einem 750-ns-Takt gesteuert, so daß der Zugriff zu einem bit innerhalb einer Schleife max. 200 µs dauert. Der Zugriff zu 64 Schleifen erfolgt mit einem 6-bit-Wort. Im Ruhezustand können die Schleifen langsamer getaktet werden (etwa 9-µs-Takt). Dadurch wird weniger Energie verbraucht.

8.3.6.2. Bubbles

Bubbles sind spezielle magnetische Strukturen, die in einigen dünnen Einkristallschichten, z.B. den Orthoferriten ($YFeO_3$, $Sm_{0,55}Tb_{0,45}FeO_3$), hexagonalen Ferriten, synthetischen Garnetkristallen und einigen amorphen Metallen auftreten. Solche Schichten werden epitaktisch, z.B. auf GGG-Substrate, abgeschieden. Spontan bilden sich in diesen Schichten magnetische Strukturen, wie es beispielhaft Bild 8.3.24a) andeutet. Wirkt auf diese Schichten senkrecht ein Magnetfeld ein, so schrumpfen in dem Beispiel die (oben positiven) magnetischen Strukturen zusammen, bis schließlich nur noch einzelne zylindrische Strukturen, die Bubbles, übrigbleiben (Teilbild b)). Bei weiterer Erhöhung des Feldes verschwinden auch sie. Um Bubbles zu erhalten, ist also ein relativ eng toleriertes magnetisches Stützfeld notwendig. Dann

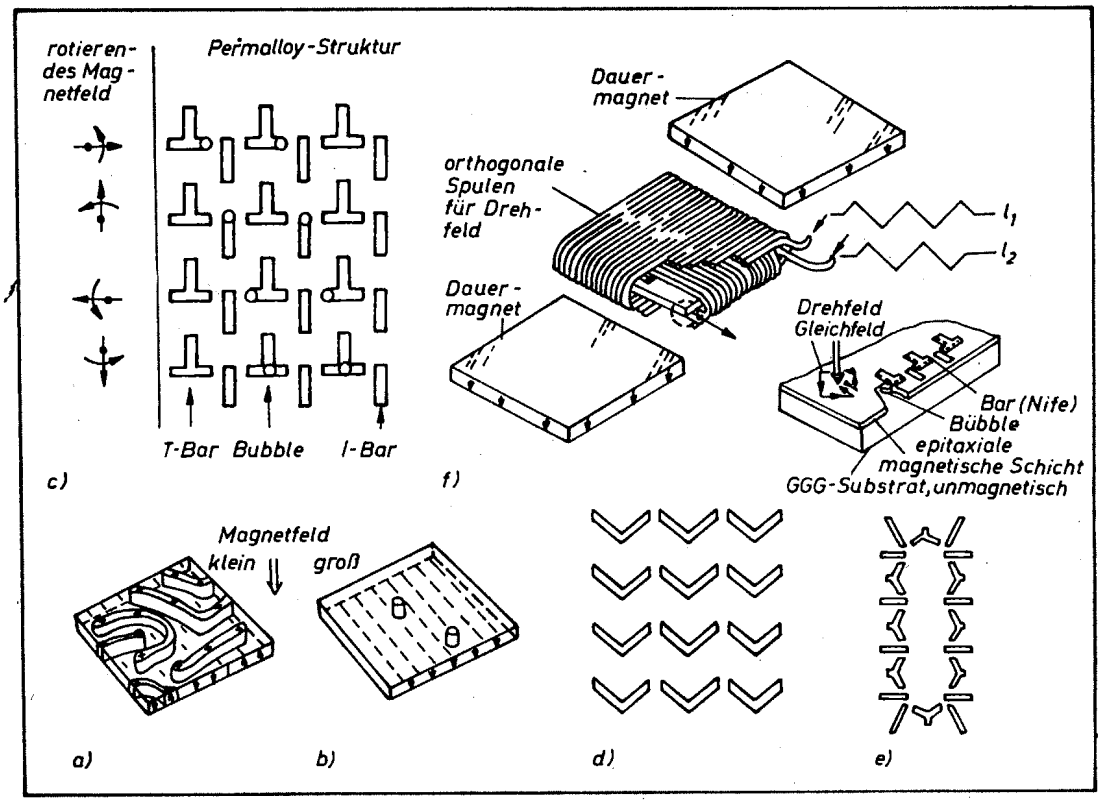


Bild 8.3.24. Bubblespeicher

a) Magnetische Schicht, z.B. aus Orthoferrit, mit spontanen magnetischen Strukturen
b) Zusammenziehen der magnetischen Strukturen zu Bubbles bei einem vergrößerten senkrechten Feld
c) I-T-Bar aus Permalloy auf der magnetischen Schicht lassen mit einem zusätzlichen Drehfeld die Bubbles definiert bewegen
d) V-Bar
e) Eine Schleife aus I-Y-Bar
f) Aufbau eines Bubblespeicher in schematischer Explosionszeichnung

sind die Bubbles in der Schicht durch kleine Feldgradienten leicht zu bewegen. Um sie definiert zu bewegen, werden auf der Bubble-Schicht Leitstrukturen (Bar) aus NiFe angebracht. Ein Beispiel hierfür sind die I-T-Bar gemäß Teilbild c). In diesem Bild wird auch gezeigt, wie sich die Bubbles durch ein kleines Drehfeld definiert fortbewegen lassen. Andere Leitstrukturen sind z.B. die V- bzw. I-Y-Bar von Teilbild d) und e). Bei Teilbild e) ist darüber hinaus eine kurze geschlossene Schleife angedeutet, in der ein Bubble mit dem Drehfeld mit 8 magnetischen Drehzyklen genau einmal umläuft. Wird Vorhandensein eines Bubble als 1 und Fehlen als Ø betrachtet, so ist sofort die Analogie zu den CCD-Speichern offensichtlich. Bubblespeicher werden auch in den Organisationsformen gemäß Bild 8.3.23 mittels der Permalloystrukturen organisiert.

Den Aufbau eines Bubblespeichers zeigt Bild 8.3.24f). Das GGG-Substrat trägt die magnetische Bubbleschicht und darauf die Barstrukturen. Dieses eigentliche Chip wird von zwei gekreuzten Spulen für das Drehfeld umgeben, und für das Stützfeld werden darüber zwei Dauermagnete angeordnet.

Ein wichtiger Vorteil der Bubblespeicher ist, daß die eingespeicherte Information auch bei Stromausfall (also Drehfeld) voll erhalten bleibt. Sie stellen also RAM mit beständiger Information (NVRAM) dar. Die Nachteile sind der relativ komplizierte Aufbau, die umfangreiche Randelektronik zur Steuerung usw. sowie die relativ niedrige Taktfrequenz von ca. 100 kHz. Anderseits sind mit Bubbles sehr große Kapazitäten von einigen 100 Kbit bis zu Mbit je Chip zu erreichen. Bisher wurden sie aber nur in wenigen Fällen angewendet.

8.3.7. Assoziativspeicher

Die bisher behandelten Speicher werden durch eine Adresse in Wort- oder Bitorganisation abgefragt. Solange Rechenanlagen für wissenschaftliche oder numerische Berechnungen verwendet werden, ist dieses Prinzip akzeptabel und effektiv. Anders werden die Verhältnisse bei Alphanumerischer Information und insbesondere dann, wenn Informationen aus einem großen Speicher gesucht, geordnet, gemischt oder verglichen werden sollen. Hierzu ein Beispiel: Auf einem Flugplatz landen und starten eine große Anzahl von Flugzeugen. Sie seien in einer Liste gemäß Tab. 8.3.2 (hier nur kleiner Auszug) geordnet. Gesucht wird ein Flug von Moskau nach Berlin. Dann besitzen alle Spalten außer Start und Ziel keine Bedeutung. Sie werden maskiert. Oft wird auch gesagt, für sie gilt don't care (nicht darum kümmern). In Start und Ziel müssen nun alle Orte mit den Suchworten verglichen werden, und wo Gleichheit

Tabelle 8.3.2. Flugplan als Beispiel für assoziative Speicherung

	lfd. Nr.	Flugnummer	Zeit	Start	Ziel	Marke
Inhalt	1	SU 111	08.10	Moskau	Berlin	*
	2	SU 122	08.45	Leipzig	Moskau	
	3	IF 300	08.50	Dresden	Moskau	
	4	OK 724	10.00	Prag	Berlin	
	5	LZ 111	11.10	Berlin	Stockholm	
	6	IF 302	14.50	Berlin	Leningrad	
	7	IF 605	20.35	Moskau	Berlin	*
Maske	+	+	+			
Suchwort				Moskau	Berlin	

besteht, wird in die letzte Spalte eine Marke eingefügt. Unter den zwei so gefundenen Flügen ist jetzt der unter anderen Gesichtspunkten günstigere auszuwählen. Hier handelt es sich um ein Vorrangproblem.

Eine andere Aufgabe wäre die Ordnung der Liste nach den Fluggesellschaften oder Abflughäfen. In jedem Fall muß dann erst eine Suchoperation in einer oder mehreren Spalten mit einem Vergleich vorgenommen werden, um danach eine Umordnung des Speicherinhaltes zu realisieren.

Solche Probleme sollen und können mit Assoziativspeichern gelöst werden. Der Gedanke hierzu entstand Mitte der fünfziger Jahre.

Der Assoziativspeicher ist im Laufe seiner Entwicklung mit verschiedenen Namen belegt worden, u. a.: inhaltsadressierter Speicher, content addressed memory (CAM), content adressable memory, correlation addressed memory, search, parallel-search, self sorting. Er hat immer eine gewisse Faszination dadurch besessen, da vermutet wird, daß unser Gehirn in etwa so arbeiten könnte. Er besitzt aber bis heute viele technische Probleme.

8.3.7.1. Grundprinzip

Das Grundprinzip eines Assoziativspeichers zeigt Bild 8.3.25a). Es ist auf neun Speicherzellen vereinfacht, von der nur die mittlere voll ausgeführt ist. Hier wird deutlich, daß die assoziative Speicherzelle aus vier Funktionsteilen besteht:

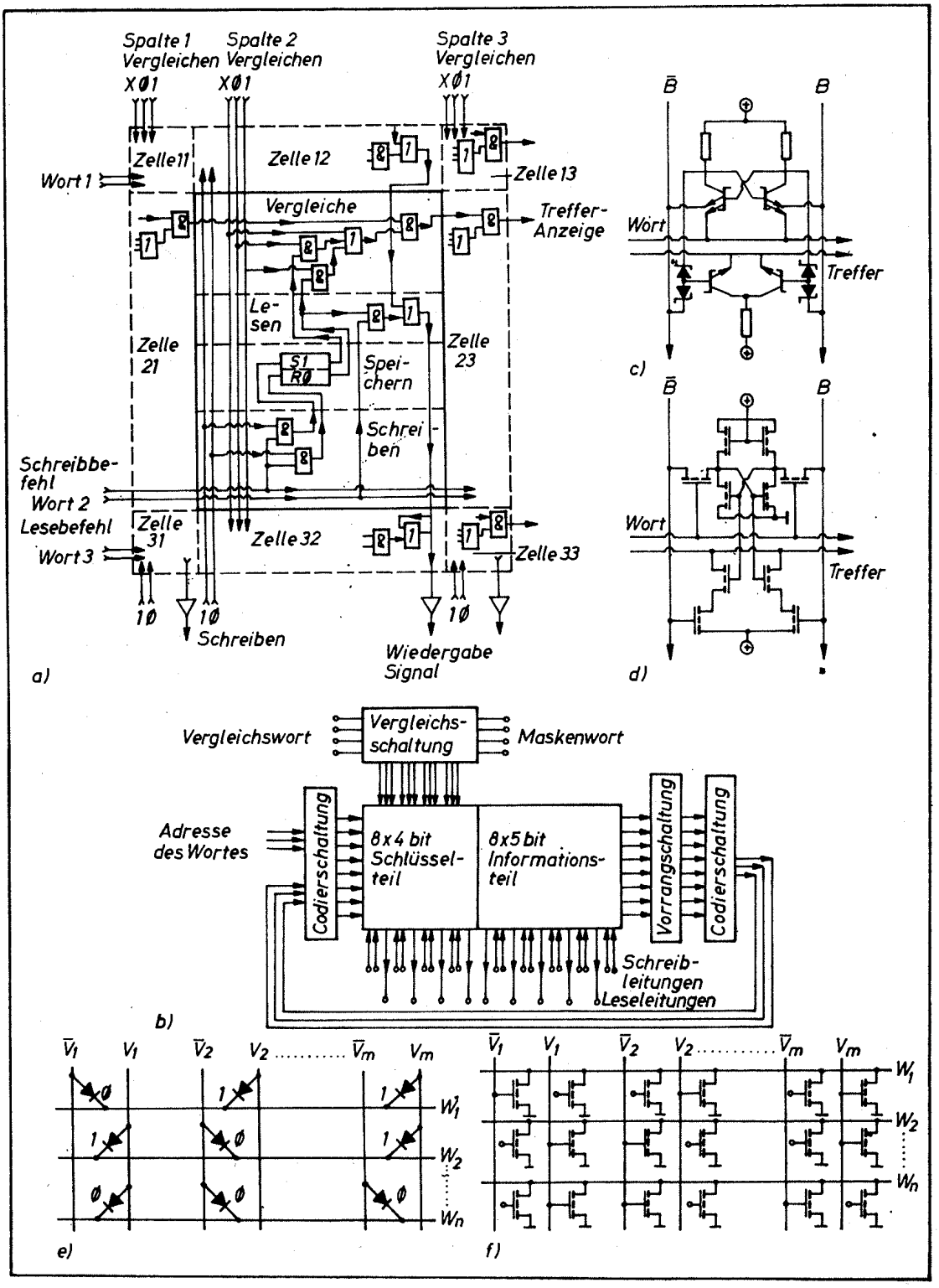


Bild 8.3.25. Assoziativspeicher

a) Prinzip am Beispiel mit drei Speicherzellen. Nur die mittlere ist voll ausgeführt. (Das rechteckige Symbol im Feld Speichern ist ein RS-Flip-Flop; vgl. 8.5.2. bzw. gemäß Bild 8.3.5c).)
b) Vollständiges Schema
c) SCHOTTKY-TTL-Assoziativspeicherzelle
d) Dito in MOS-Technik
e) Assoziativer Festwertspeicher in Diodenlogik
f) Dito in FET-Technik

— eigentlicher Speicher, Flipflop, Latch
— Funktionsteil zum Schreiben,
— Funktionsteil zum Lesen,
— Vergleichsteil mit Trefferausgang.

Die assoziative Speicherzelle ist also um eine binäre Schaltung umfangreicher als der übliche Schreib-Lese-Speicher. Sie benötigt daher in der Regel etwa die doppelte Chipfläche. Es wird aus dem Bild aber auch ein weitaus größerer Nachteil ersichtlich, nämlich die wesentlich größere Anzahl von Anschlüssen der Speicherzelle und des Speichers überhaupt.

Beim Schreiben und Lesen verläuft der Vorgang in der Speicherzelle genauso wie beim üblichen wortorganisierten Speicher. Hierfür ist also keine neue Beschreibung notwendig. Bei der Vergleichsoperation geschieht folgendes: Den einzelnen Spalten wird das Vergleichswort aus mehreren Ø oder 1 eingegeben. Spalten, die nicht verglichen werden sollen, erhalten ein Signal am X-Eingang. Diese Spalten sind dann maskiert. Ein mögliches Vergleichswort könnte also z. B. 1X1 oder Ø11 oder XX1 sein. Liegt nun bezüglich 1 oder Ø Übereinstimmung mit dem Zustand des Flipflop vor, so schalten die beiden UND-Glieder im Vergleichsteil durch. Das Dreifach-ODER-Glied reagiert unabhängig davon direkt auf die Maskierung X. Sein Ausgangssignal gelangt in jedem Fall mit dem Ausgangssignal der vorangehenden Speicherzelle des gleichen Wortes zu einem UND-Glied. Sind beide Signale 1, so wird es zur dritten Zelle des gleichen Wortes weitergegeben, wo auch die Vergleichsoperation durchgeführt wird. Am Ende eines jeden Wortes steht also genau dann ein 1-Signal zur Verfügung, wenn an allen Spalten des Wortes zwischen Vergleichssignal und Flipflop-Zustand Übereinstimmung herrscht. Damit liegt aber erst das Ergebnis vor, daß dieses Wort, oft sind es mehrere, den Vergleichskriterien genügt. Noch nicht bekannt sind damit die Inhalte, d. h. die 1Ø-Folge der Worte. Sie muß nun mit einer Leseoperation ausgegeben werden. Dafür ist aber zunächst noch die Rangfolge der Ausgabe für die erkannten Worte festzulegen. Hierzu gibt es meist nicht zum Speicher gehörende Vorrangsteuerungen. Sie arbeiten nach verschiedenen Kriterien: z. B. in der Reihenfolge der Worte oder nach einer Bedeutungsfolge aus dem Vergleichswort. Die Vorrangsteuerung löst danach wortweise den Lesebefehl des Speichers aus.

Oft ist es auch üblich, daß ein Wort eines Assoziativspeichers aus zwei Teilen besteht, dem Schlüssel- und dem Informationsanteil. Dann ist der Informationsanteil ein zusätzlich angefügter Teil mit üblichen Speicherzellen ohne Vergleichsteil. Das Schema eines unter diesen Gesichtspunkten vollständigen Speichers zeigt Bild 8.3.25 b).

8.3.7.2. Typische assoziative Speicherzellen

Es gibt heute eine größere Anzahl von assoziativen Speicherzellen. In Bild 8.3.25 c) und d) sind je eine typische aus MOS- und bipolaren Transistoren ausgewählt. Bei beiden sind die Bitleitungen $\overline{B}$ und B für Schreiben, Lesen und Vergleichen zusammengefaßt. Oben befindet sich der Speicherteil. Er ist wie üblich aufgebaut. Der unten dargestellte Vergleicherteil bei der MOS-Zelle besitzt je zwei in Reihe geschaltete MOS. Wenn das Signal an den Bitleitungen mit dem Speicherzustand im oberen Flipflop übereinstimmt, sind alle vier Transistoren leitend, und an der Trefferanzeige kann ein Signal abgenommen werden. Es genügt auch eine Übereinstimmung. Dies ist wegen der Maskierung über $\overline{B}$ und B notwendig. Ganz ähnlich wirken die beiden Dioden in der bipolaren Zelle.

8.3.7.3. Assoziative Festwertspeicher

Zunächst mutet es seltsam an, von assoziativen Festwertspeichern zu sprechen. Sie sind aber schon lange bekannt und finden auch Anwendung. Der entscheidende

Gedanke für den assoziativen Festwertspeicher liegt darin, jede Spalte doppelt auszuführen, einmal normal und einmal negiert. Der assoziative Festwertspeicher ist vom üblichen Standpunkt daher 100% redundant. Seine Funktion sei mittels Bild 8.3.25e) erklärt. Er enthält nur noch die Vergleichseingänge V und den negierten $\overline{V}$ für die Spalten und die Wortleitungen. Zur Suche nach einer 1 erhalten die linken $\overline{V}_m$ positive Impulse, für eine Ø die rechten V_m. Die Dioden sind genau entgegengesetzt angeschlossen. Dadurch gelangt auf die Wortleitung genau dann kein Impuls, wenn Trefferanzeige erfüllt ist. Maskiert wird durch keinen Impuls. Für die ebenfalls gezeigte FET-Schaltung in Teilbild f) gelten genau die gleichen Bedingungen. Es ist offensichtlich, daß so aus einem umfangreichen Speichervorrat schnell eine spezielle Information gefunden werden kann. Um das Wort dann zu erhalten, wird der Speicher von den Wortleitungen aus zurück betrieben, und das richtige Signal entsteht an den Leitungen V_m.

8.3.8. Kellerspeicher

Abgesehen von Assoziativspeichern werden Speicherzellen ihrer Adresse nach aufgerufen. Ein spezielles Prinzip dieser Art ist der Kellerspeicher. Er trägt auch die Bezeichnung Stack, Last-in-first-out, bzw. abgekürzt Lifo-Stack, push-down-store oder Stapelspeicher. Sein mechanisches Modell zeigt Bild 8.3.26a). In einen Kasten können nur von oben mit Information beschriebene Karten eingelegt (push) bzw. abgenommen (pop) werden. Die Reihenfolge der Karten auf dem Stapel bleibt dabei immer erhalten. Das, was zuletzt eingelegt wurde, muß zuerst abgenommen werden (last in first out), bevor die darunter liegenden Karten erreichbar werden. In der

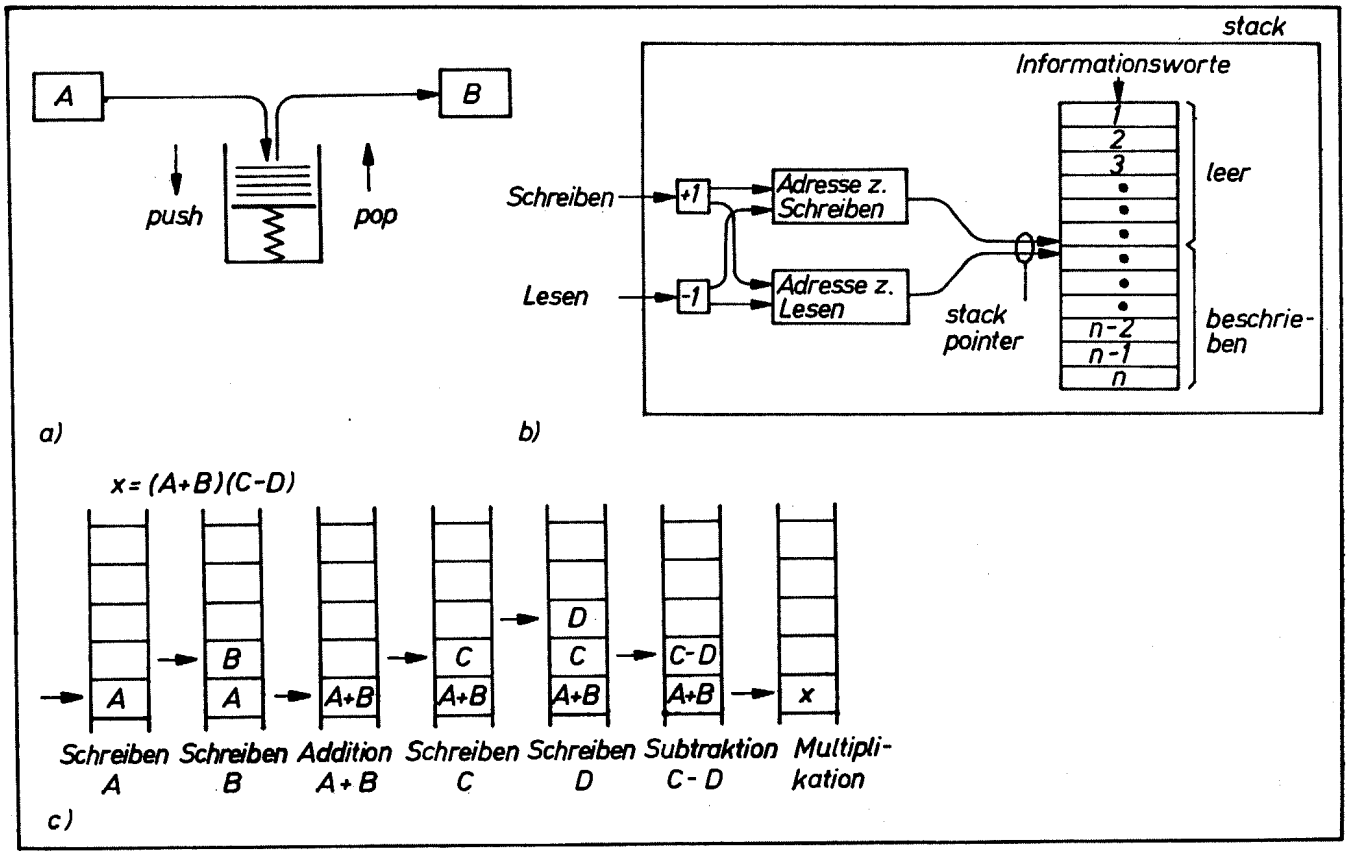


Bild 8.3.26. Prinzip des Kellerspeichers (Stack)

a) Mechanisches Analogon

b) Organisation eines Kellerspeichers mit Stackpointer

c) Beispiel einer Rechenaufgabe, bei der ein Stack vorteilhaft anzuwenden ist

Praxis wird ein Kellerspeicher aus einem wortorganisierten Speicher durch Zusätze zu der dann nach außen nicht mehr zugänglichen Adreßbildung realisiert. Dies zeigt Bild 8.3.26b). Mittels der Erhöhung oder Erniedrigung des Adressenzeigers (-Zählers, Stackpointers) ist immer nur jeweils ein Informationswort für das Lesen und ein „darüberliegendes" für das Schreiben zugänglich.

Das Prinzip der günstigen Arbeitsweise des Kellerspeichers sei am Ausdruck

$$X = (A + B) \cdot (C - D) \tag{8}$$

gemäß Bild 8.3.26c) beschrieben. Das zweite von oben gesehene Register dient hier also immer gleich als Akkumulator.

8.4. Grundlagen sequentieller Schaltungen

Sequentielle Schaltungen setzen sich aus kombinatorischen Schaltungen und Speichern zusammen. Dadurch ist ihr Ausgangssignal nicht nur von den Eingangssignalen, sondern auch von den internen Zuständen im Speicher abhängig. Man kann denselben Inhalt auch so ausdrücken: Das oder die jeweiligen Ausgangssignale sind von den jeweils vorhandenen und dazu von den vorangegangenen Eingangssignalen abhängig. Dies bedeutet, daß die sequentiellen Schaltungen im Gegensatz zu den kombinatorischen auch die Vorgeschichte berücksichtigen. Bezüglich der Zeitbedingungen des Verhaltens sind *zwei Typen von sequentiellen Schaltungen* zu unterscheiden:

- *ungetaktete* (asynchrone)
- *getaktete* (synchrone).

Bei den synchronen Schaltungen existieren periodische Zeittakte, zu denen etwas in der Schaltung geschehen kann. In den Zwischentaktzeiten sind die Schaltungen bezüglich der Eingangssignale unempfindlich. Bei den ungetakteten Schaltungen schalten immer die Flanken der Eingangssignale, sobald sie auftreten.

Eine andere Unterscheidung hängt davon ab, ob das *Ausgangssignal unmittelbar von den inneren Zuständen* abgeleitet wird oder noch kombinatorisch mit den Eingangssignalen verknüpft wird. Dementsprechend werden *unterschieden*:

MOORE-*Automat* und MEALY-*Automat*.

Im folgenden sei versucht, die Theorien und Beschreibungen der Grundlagen der sequentiellen Schaltungen in möglichst einfacher Weise darzustellen. Dabei wird von einem einfachen Spielmodell ausgegangen, das vor allem das Problem der Analyse zeigen soll. Danach wird an einem einfachen Beispiel die Vielfalt der Beschreibungsmöglichkeiten aufgezeigt. Schließlich wird auf einige Probleme der Berechenbarkeit, d. h. der Grundlagen von Algorithmen, eingegangen.

8.4.1. Zwei Modelle zur Analyse

8.4.1.1. Erstes Versuchsgerät

Es liege ein kleiner Kasten gemäß Bild 8.4.1 vor. Er enthält *drei Tasten* x_1, x_2, x_3. Sie können gedrückt werden und stellen daher die Eingangsgrößen dar. Beim Drücken einer Taste leuchtet jeweils eine der *drei Lampen* mit den farbigen Gläsern gelb, rot bzw. blau auf. Das Aufleuchten kennzeichnet die Ausgangsfunktion. Wir wollen versuchen, den Zusammenhang zwischen der gedrückten Taste und der dadurch angezeigten Farbe zu ergründen. Dazu schreiben wir beim Versuch ein Protokoll. Die obere Reihe trägt die Tastenbezeichnung, und darunter steht die dadurch ausgelöste Farbe (rot = r, gelb = g, blau = b). Beim ersten *Versuch* ergibt sich z. B. die folgende

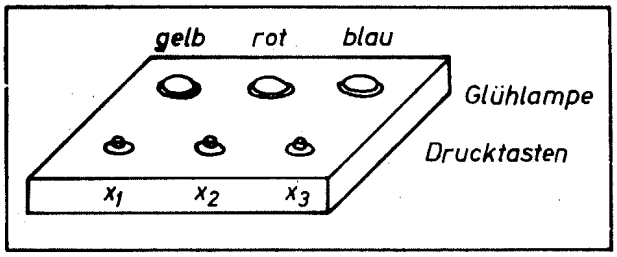


Bild 8.4.1. Einfaches Modell für einen Automaten, der durch die Tasten x_1, x_2 und x_3 betätigt werden kann und als Ausgabe die Anzeigen gelb, rot und blau besitzt

Protokollreihe:

$$x_1, x_1, \ldots, x_3, x_1, \ldots, x_2, x_2, x_2, \ldots, x_3, x_3, x_3$$
$$g, \quad g, \ldots, r, \quad g, \ldots, g, \quad r, r, \ldots, b, \quad r, r\,.$$

Auffallend ist, daß offensichtlich einer Taste nicht nur eine Farbe zugeordnet ist. Der Zusammenhang zwischen Farben und Tasten, d.h. Aus- und Eingang ist also komplizierter. Das Verhalten des Gerätes bedarf einer systematischen Analyse. Dazu sind in der Regel eine größere Anzahl von Versuchen notwendig. Hier sei nur das Ergebnis mitgeteilt:

— x_1 löst immer *gelb* aus,
— x_2 kann *rot* oder *gelb* auslösen,
— x_3 kann *rot* oder *blau* auslösen,
— *gelb und blau* treten nur einmal auf, wenn x_2 bzw. x_3 hintereinander betätigt wird,
— *blau* erscheint bei Taste x_3, wenn vorher x_2 gedrückt wurde,
— *gelb* erscheint bei Taste x_2, wenn vorher x_1 oder x_3 gedrückt wurde.

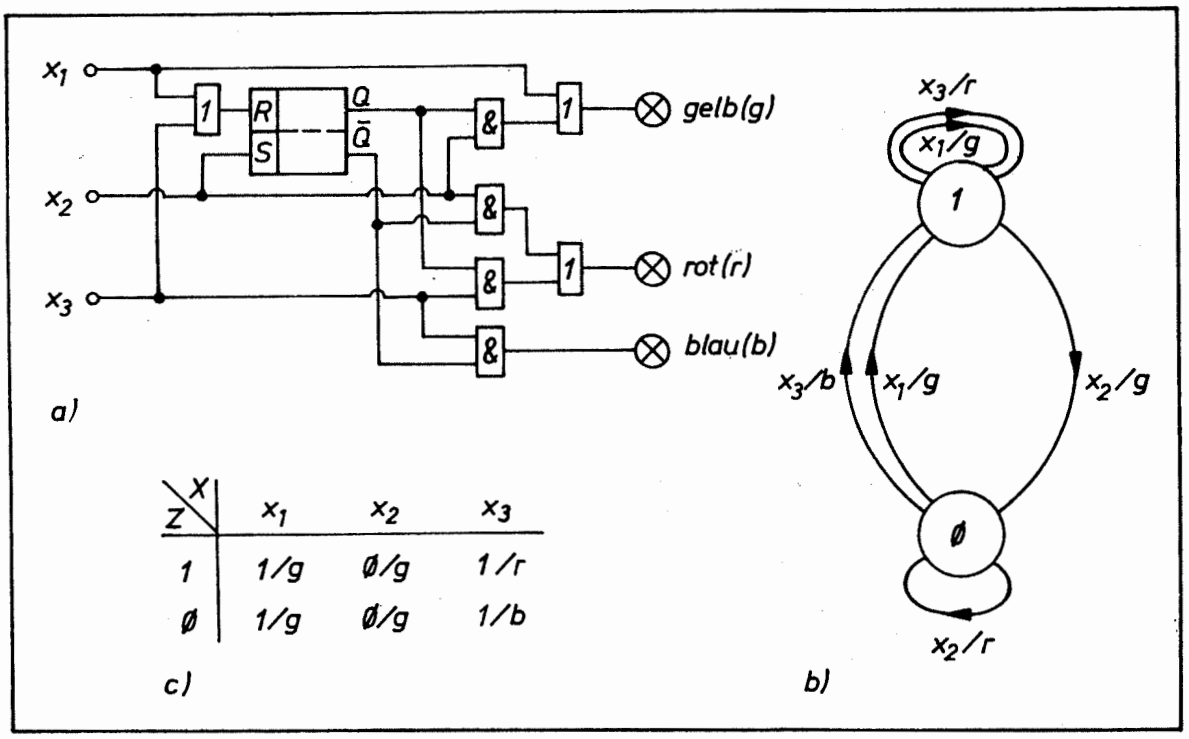
Besonders aus den beiden letzten Aussagen folgt, daß in dem Kasten ein Gedächtnis (Speicher) sein muß. Es merkt sich, was das letzte Mal vorher gedrückt wurde, und nimmt dabei einen inneren Zustand ein.

Es gibt nun zwei Wege weiterzugehen:

— Es wird ein elektronisches Modell geschaffen, das dem Kasten entspricht. Dies ist eine Synthese eines Automaten aus den gegebenen Eigenschaften und kann zu verschiedenartigen Automaten gleichen Verhaltens führen.
— Es wird der Kasten geöffnet und die Schaltung aufgenommen. Dabei handelt es sich um eine spezielle und nicht immer mögliche Variante der Analyse. Eine andere Variante der Analyse war unser Versuchsablauf.

Hier soll die fertige Schaltung gemäß Bild 8.4.2a) betrachtet werden. Sie enthält ein RS-Flipflop und etliche Gatter. Durch den Tastendruck wird der Flipflop eingestellt. Das Verhalten der Schaltung ist also nur von den beiden möglichen Zuständen des Flipflop abhängig. Es muß deshalb eine einfachere Beschreibung als die obige verbale geben. Sie setzt aber voraus, daß wir den Zustand des Flipflop kennen. Er existiert aber nur innerhalb des Gerätes und ist damit nicht direkt der Beobachtung zugänglich. Dies war schon die Folgerung, die weiter oben auf Grund der verbalen Beschreibung gezogen wurde. Gemäß der Schaltung können die beiden Zustände 1 und Ø des Flipflop existieren, und so kann die Automatentafel von Bild 8.4.2c) aufgestellt werden. Darin bedeute Z die inneren Zustände, die gerade existieren. Ist $Z = 1$ und erscheint das Signal $X = x_1$, so bleibt der Zustand $Z = 1$ erhalten, und es leuchtet die Lampe gelb. Wird bei $Z = 1$ dagegen $X = x_2$ betätigt, so leuchtet ebenfalls die Lampe g, und nach dem Betätigen entsteht der Zustand $Z = Ø$.

Eine andere Beschreibungsmethode ist das graphentheoretische Modell, wie es Bild 8.4.2b) zeigt. Es enthält die Zustände des Flipflop als entsprechende Kreise.



Z \ X	x_1	x_2	x_3
1	1/g	ø/g	1/r
ø	1/g	ø/g	1/b

c)

Bild 8.4.2. Beispiel eines einfachen MEALY-Automaten

a) Schaltbild (Der Flipflop muß ein Master-Slave-Flipflop sein (8.5.2.6.), der erst dann kippt, wenn die betätigte Taste wieder losgelassen wird.)

b) Darstellung der Funktion durch einen Graphen mit den Zuständen $Q = ø$ bzw. 1

c) Wertetabelle für den Automaten

Von jedem Zustand gehen Pfeile aus, die den Tastendrücken entsprechen. Sie sind mit dem Tastendruck und der ausgelösten Farbe gekennzeichnet und enden bei demselben oder dem (einem) anderen Zustand. Enden sie beim selben Zustand, so ist dieser stabil für diese Eingangsgröße (Tastendruck). Sie kann dauernd wiederholt werden, es leuchtet immer dieselbe Lampe, z.B. Zustand ø und Tastendruck x_2 immer rot. Der Pfeil kann auch zum anderen Zustand führen. Dann wird sich erst dort zeigen, ob dieser nun für den Tastendruck stabil ist oder nicht. In diesem einfachen Modell ist dies der Fall. Gleichzeitig zeigt sich aber dabei, wie schwer von außen die Zustände erkennbar sein können: Bei Zustand ø beginnend, werde x_1 betätigt, es erscheint gelb und wird 1 angenommen. Erneut x_1 betätigt, bleibt 1, aber auch hier wird gelb angezeigt. Wären statt der drei Ausgänge nur zwei vorhanden gewesen, indem z.B. zwei Lampen zusammengefaßt worden wären, so hätte sich die Analyse wahrscheinlich noch schwieriger gestaltet. Sie wäre auch schwieriger geworden, wenn mehrere Speicher im Automaten vorhanden gewesen wären.

8.4.1.2. MEALY- und MOORE-Automat

Die bisherige Betriebsweise des Gerätes (Automaten) von Bild 8.4.1 war, so daß Farben nur dann aufleuchten, wenn eine Taste gedrückt wird. Es ist aber auch möglich, ständig eine Taste leuchten zu lassen und dann beim Drücken einer Taste zwischen den brennenden Lampen umzuschalten. Eine entsprechende Schaltung zeigt Bild 8.4.3. Es ist auffällig, daß die Ausgangsgrößen nur noch von den inneren Zuständen abhängen. Damit ist dieser Automat im Prinzip einfacher. Im Graphen sind die Farben bei den Knoten mit eingetragen. An den Kanten stehen nur noch die Eingangsgrößen. (Der isolierte Zustand 1ø ist erst für spätere Betrachtungen vorgesehen. Er hat zunächst keine Bedeutung, da er ja nicht durch Drücken einer Taste erreichbar ist.) Die Automatentafel vereinfacht sich. In der Eingangsspalte können

Zustände und Ausgangsgrößen zusammengefaßt werden. In den Zeilen stehen nur die neuen Zustände. Die Lampen leuchten ja auch ohne x. Sie werden nur durch x umgeschaltet. Dementsprechend heißt dieser Automat zuweilen auch Zustandsautomat. Die Ausgangsgröße hängt nur von den inneren Zuständen ab. Der Automat gemäß Bild 8.4.2 heißt aus ähnlichen Gründen zuweilen auch Übergangsautomat. Nur während der Betätigung, nach deren Beendigung der Übergang in den neuen Zustand erfolgt, existiert ein Ausgangssignal. Da aber auch die Begriffe Zustand und Übergang auf die äußeren Signale bezogen werden können, lassen diese Bezeichnungen Fehlinterpretationen zu. Es ist besser, die historisch gewachsenen Begriffe MEALY-Automat für den ersten und MOORE-Automat für den zweiten Typ zu verwenden.

8.4.2. Beschreibungsmethoden

Bei den kombinatorischen Schaltungen wurden im Abschnitt 8.2.2.2. sieben *Möglichkeiten zur Darstellung* des gleichen Sachverhaltes anzugeben. Für sequentielle Schaltungen existieren die folgenden sechs, von denen schon einige verwendet wurden:

- verbale Beschreibung,
- mathematische Beschreibung durch zwei Funktionen (8.4.4.)
- Wertetabellen (Bild 8.4.2c) und 8.4.3b)),
- Schaltungen (im Sinne von Schaltbildern wie Bild 8.4.2a)
- Zustandsdiagramme bzw. -graphen (Bild 8.4.2b) und 8.4.3c))
- PETRI-Netze (8.4.3.).

Für die Wertetabellen sei noch eine allgemeine Form angegeben:

MEALY		MOORE		
	Eingänge			Eingänge
Zustände	neuer Zustand	Zustand	Ergebnis	neuer Zustand
	Ergebnis			

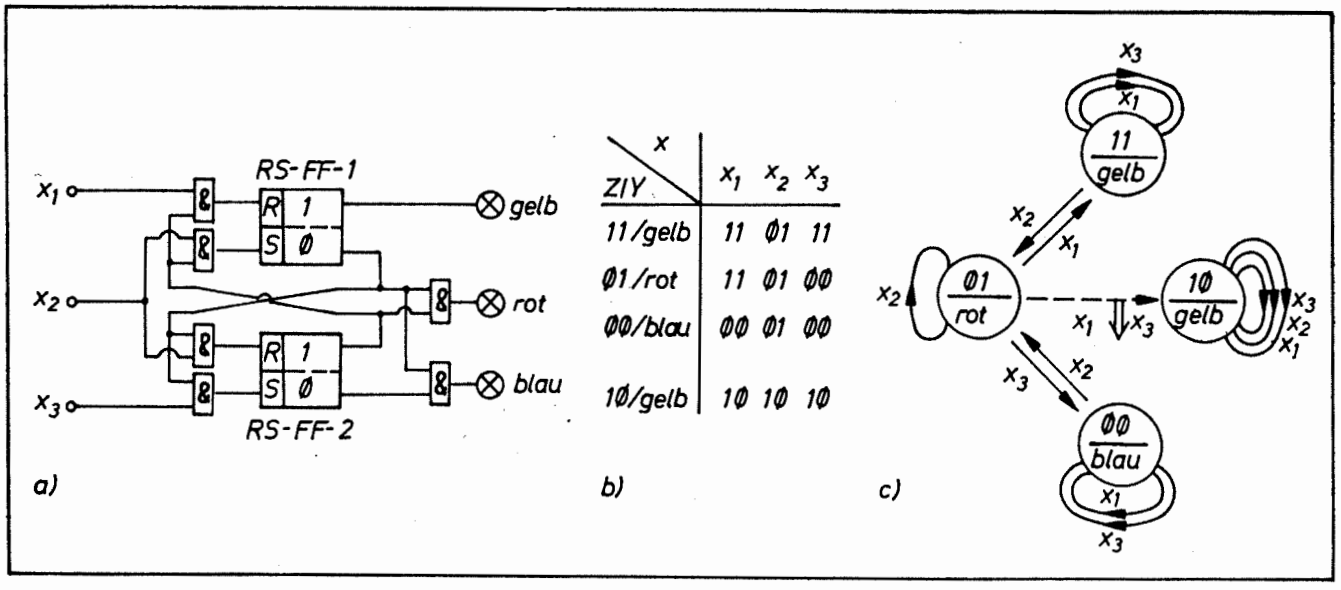


Bild 8.4.3. Beispiel eines MOORE-Automaten a) im Schaltbild, b) Wertetabelle und c) Graph. Der Graph enthält auch einen unerwünschten Zustand 10, der nur durch gleichzeitige Betätigung von x_1 und x_3 erreichbar ist (vgl. 8.4.8., Race)

Einige der Methoden seien an einem besonders einfachen Beispiel erklärt:

Verbal: Es existiert ein Fahrkartenautomat, der 10- und 20-Pfennig-Stücke zur Eingabe zuläßt. Die Fahrkarte kostet 30 Pfennige. Sie wird ausgegeben, wenn dieser Wert bei der Eingabe erreicht ist. Werden zwei 20-Pfennig-Stücke eingegeben, so wird die Karte und 10 Pfennige ausgegeben. Die Reihenfolge der Eingabe der Geldstücke darf keine Rolle spielen.

Die *Wertetabelle* dieses Automaten zeigt Bild 8.4.4a) in der Form des MEALY-Automaten, bei dem die Eingabe der Geldstücke unmittelbar die eventuelle Ausgabe bewirkt. In diese Betrachtung ist auch die Ausgabe von weder Geld noch Fahrkarte als eine Ausgabefunktion, nämlich die leere (angedeutet durch den Strich —), einzubeziehen.

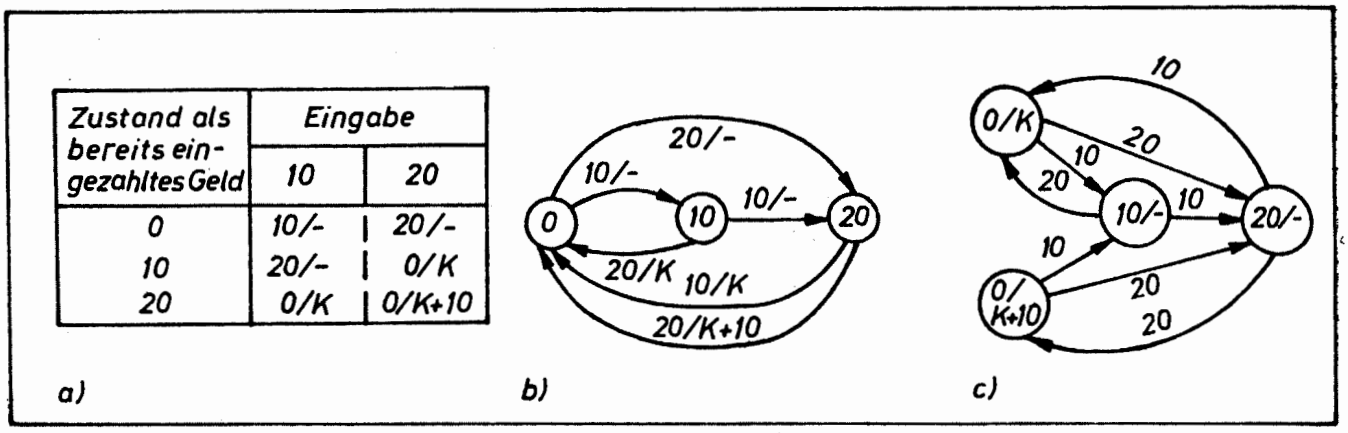


Bild 8.4.4. a) Wertetafel, b) MEALY-Graph und c) MOORE-Graph eines speziellen Fahrkartenautomaten (siehe Text)

Den *Zustandsgraphen* dieser Tabelle zeigt Bild 8.4.4b). Er entspricht der Darstellungsform von Bild 8.4.2b). Im Prinzip läßt sich aber auch in der Wertetabelle und in der Graphendarstellung dieses Problem als MOORE-Automat erfassen. Es ist so in Bild 8.4.4c) dargestellt. Als Knoten werden jetzt die Kombinationen aus im Automaten enthaltenem Geldwert (innerer Zustand) und Ausgabefunktion (Nichts, Karte, Karte +10) verwendet.

Dieses Beispiel läßt zugleich einen allgemeingültigen Zusammenhang zwischen MOORE- und MEALY-Automat beispielhaft erkennen:

> Jeder MOORE-Automat ist in einen äquivalenten MEALY-Automaten transformierbar und umgekehrt. Die Anzahl der Knoten (Zustände) ist dabei beim MOORE-Automaten stets größer gleich der beim MEALY-Automaten.

So wird verständlich, daß der MEALY-Automat unter sonst gleichen Bedingungen uns komplexer erscheint (vgl. die Beschreibungen zu Bild 8.4.2 und 8.4.3).

8.4.3. PETRI-Netze

PETRI-Netze gehen auf Arbeiten von C. A. PETRI zurück. Sie erlangten erst in den letzten Jahren Bedeutung. Ein PETRI-Netz ist ein spezieller Graph mit zwei Typen von Knoten:

- *Kreisen* = Zustandsknoten, auch Stellen oder Plätze genannt
- *Rechtecken* = Transitionsknoten, Ereignisse.

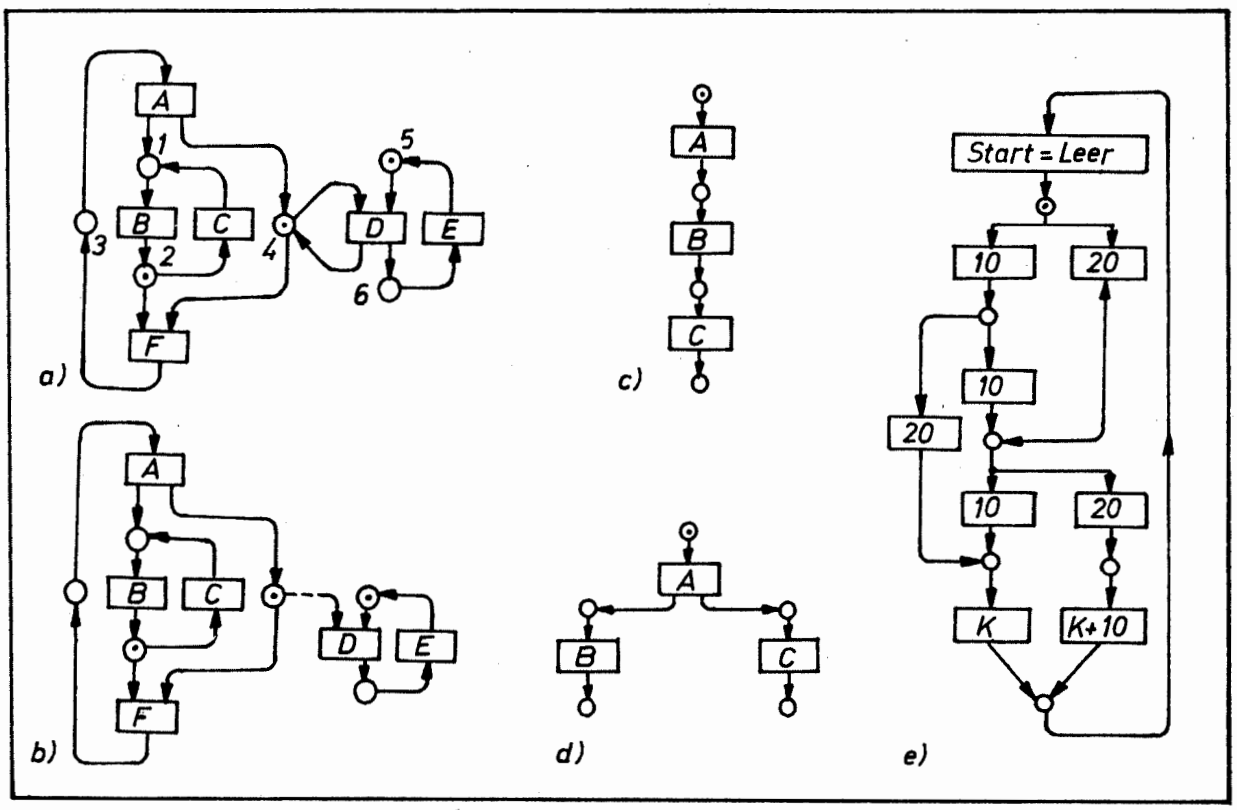


Bild 8.4.5. Zu PETRI-Netzen

a) Beispielschema
b) Dito, jedoch mit Hervorhebung der Lesestelle 4 (vgl.a))
c) Darstellung einer Kausalkette *ABC*
d) Darstellung einer Verzweigung: Nach *A* kann *B* oder *C* folgen
e) PETRI-Netz des Fahrkartenautomaten von Bild 8.4.4

Sie werden durch *gerichtete Linien* immer alternierend verbunden. Wesentlich für die PETRI-Netze ist, daß die Kreise mit einer Marke (ähnlich einem Stein bei Brettspielen) belegt werden können. Daher auch die Bezeichnung markierter Graph für ein PETRI-Netz (Bild 8.4.5).

Der Arbeitsweise des Automaten entsprechen im PETRI-Netz die *Schaltregeln* für die Transitionen:

1) Eine Transition kann genau dann schalten, wenn alle vorangehenden Kreise (Stellen), die mit Pfeilen zu ihr weisen, durch eine Marke belegt, also markiert sind. In Bild 8.4.5a) gilt dies für *D* und *F*.
2) Wenn eine Transition gemäß 1 wirklich schaltet (ausgeführt wird), dann bewirkt dies,
   - daß bei allen vorangehenden Stellen die Marken weggenommen werden, sie also leer werden,
   - daß alle der Transition folgenden (wohin ihre Ausgangspfeile zeigen) Stellen mit Marken belegt werden.

Wenn also *F* schaltet, entfallen die Marken bei 2 und 4, und 3 erhält eine Marke. Im Bild 8.4.5a) gibt es eine besondere Struktur 4 bezüglich *D*. Wenn *D* schaltet, entfallen die Marken bei 4 und 5 und neu belegt werden 4 und 6. Bei 4 geschieht also nichts, sie heißt deshalb eine *Lesestelle* bezüglich *D*. Dies läßt sich besser in Bild 8.4.5b) mit der gestrichelten Linie darstellen.

Wenn in Bild 8.4.5a) gemäß der obigen Beschreibung *F* schaltet, so ist die Stelle 4 nicht belegt, und damit ist durch das Schalten von *F* die vorher vorhandene Schalt-

bereitschaft von D beseitigt. Dies wird verbal so ausgedrückt, daß F in *Konflikt* mit D steht. Infolge der Stelle 2 steht auch F in Konflikt mit C, und in diesem Fall gilt auch die Umkehrung, daß C in Konflikt mit F steht.

PETRI-Netze können sowohl zur Darstellung von kausalen Folgen als auch sich anschließender Ereignisse verwendet werden (Bild 8.4.5c) und d)). Ein besonders einfaches PETRI-Netz zeigt Bild 8.4.5e). Es ist die Darstellung des Fahrkartenautomaten von Bild 8.4.4.

Weitere Details zu PETRI-Netzen enthält u. a. [S 54, Bd. 3].

8.4.4. Formale Beschreibung

Nach den bisherigen Betrachtungen existieren *drei Größen für einen Automaten*:

- eine endliche Zahl von *Eingangsgrößen* X: $x_1, x_2, x_3, \ldots$ Sie heißen auch das *Eingangsalphabet* (Tasten bzw. Geld).
- eine endliche Zahl von *Ausgangsgrößen* Y: $y_1, y_2, y_3, \ldots$ (das waren bisher die Farben b, r, g bzw. die Fahrkarte und das Geld). Sie heißen *Ausgangsalphabet.*
- eine endliche Anzahl von *inneren Zuständen* Z: $z_1, z_2, z_3, \ldots$ Das waren bisher die Zustände 1 und Ø der Flipflops.

Mit diesen Größen und deren Zusammenhängen ist jeder Automat vollständig zu beschreiben. Für die Zusammenhänge sind *zwei Funktionen* wichtig:

Die *Übergabefunktion*

$$Z := f(Z, X) \tag{1a}$$

gibt an, welcher neue Zustand sich durch die Eingangsgrößen, z.B. das Drücken der Taste, einstellt. Hieraus wird deutlich, daß zwei Zeiten, t_n vor dem Drücken der Taste und t_{n+1} nach dem Drücken der Taste, zu unterscheiden sind. Deshalb steht das Ergibtzeichen $:=$. Mit den Zeittakten ist die Schreibweise

$$^{n+1}Z = f(^nZ, {}^nX) \tag{1b}$$

möglich. Die Takte werden also oben nks angefügt. Die Übergabefunktion heißt auch Übergangsfunktion.

Die zweite Funktion bezieht sich auf die Ausgangsgröße. Sie heißt daher *Ausgabefunktion* oder auch Ausgangs- bzw. Ergebnisfunktion. Für sie gilt

$$Y = g(Z, X)\,. \tag{2}$$

Sie gibt an, welche Ausgangsgröße beim Zustand Z und der Empfangsgröße X auftritt. Insgesamt ist damit ein Automat durch

$$(X, Y, Z, f, g) \tag{3}$$

definiert. Zuweilen wird für f auch δ und für g dann ω geschrieben.

Eine mögliche Darstellung der Funktionen sind die Wertetabellen in den Bildern 8.4.2 bis 8.4.4.

Die Anzahl Z der *inneren Zustände* kann durch die Anzahl n der *Flipflops* ausgedrückt werden:

$$Z \leqq 2^n\,.$$

Die Belegung der Zustände mit einer Kombination der Zustände der einzelnen Flipflops kann z. T. willkürlich erfolgen und heißt *Kodierung der Zustände Z.*

8.4.5. Typen von Automaten

Automaten können einmal nach ihrem Speicherumfang und nach der Art ihres Speichers unterschieden werden. So entstehen die verschiedenen endlichen und unendlichen Automaten (z.B. Kellerautomat, beschränkter Automat, TURING-Automat [V 12, Bd. I]). Daneben existieren deterministische und stochastische Automaten.

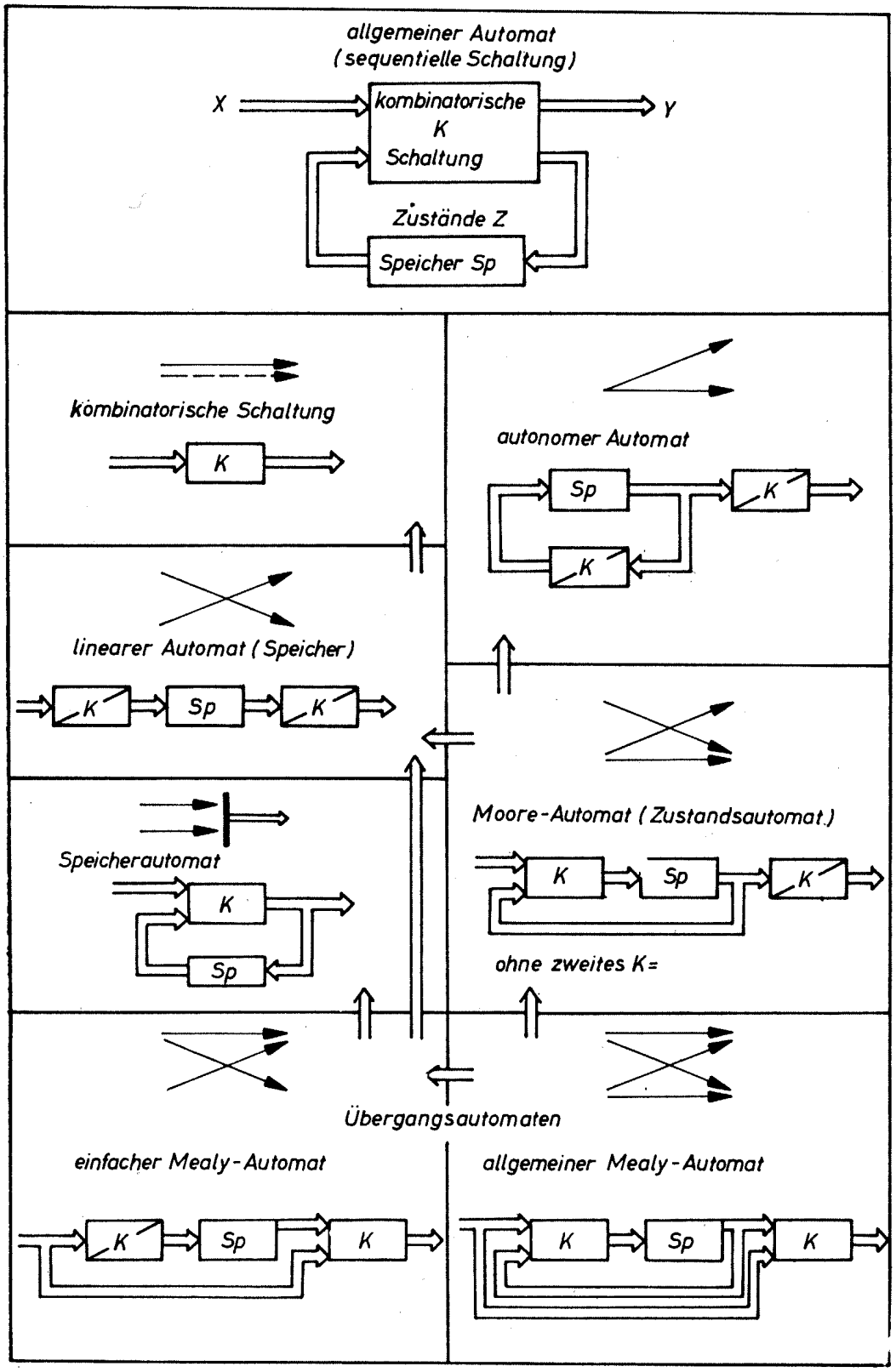


Bild 8.4.6. Ableitungen verschiedener Automatentypen aus der allgemeinen Zusammenschaltung von kombinatorischer Schaltung und Speicher. Die offenen Pfeillinien sollen andeuten, daß hier parallel mehrere Variable existieren

Hier seien nur die verschiedenen endlichen, deterministischen Automaten weiter betrachtet. Diese können grundsätzlich aus einer kombinatorischen Schaltung und endlich vielen Speicherzellen zusammengesetzt sein. Ein sehr allgemeines Schema zeigt damit Bild 8.4.6 oben. Die gedoppelten Linien mit Pfeilen sollen mehrere mögliche Größen symbolisieren. Die kombinatorische Schaltung kann nun in verschiedener Weise zwischen den zwei Eingangsvariablen (X und vom Speicher) und den zwei Ausgangsvariablen (Y und zum Speicher) vermitteln. Dadurch entstehen die sieben verschiedenen, im Bild dargestellten Spezialfälle, wobei zur weiteren Vereinfachung die diagonal durchstrichenen K auch entfallen können. Aus dem MOORE-Automaten wird dann z. B. ein MEDWEDEV-Automat. Beim Speicherautomaten sind die Ausgangssignale Y und die Eingangssignale des Speichers identisch. Bis auf die Numerierung des Taktes stellt er einen stark vereinfachten MEALY-Automaten dar. Der autonome Automat schließlich gibt ohne Eingangssignale Ausgangssignale ab. Er stellt also eine spezielle digitale Form von Generatoren dar.

Insgesamt stellt nach diesen Betrachtungen der MEALY-Automat den allgemeinen Fall dar. Aus ihm lassen sich alle anderen Fälle in Bild 8.4.6 durch Vereinfachung ableiten. Dies ist durch entsprechende Pfeile ebenfalls angedeutet. Lediglich beim Speicher geht das nicht so konsequent.

8.4.6. Beispiel: Steuerung für ein Magnetbandgerät

Für den Entwurf eines Automaten werde die Steuerung für ein Magnetbandgerät ausgewählt. Dies hat den Vorteil, daß sein Verhalten als allgemein bekannt vorausgesetzt werden kann. Bei den üblichen Geräten werden zumindest heute noch die Funktionen fast immer rein mechanisch ausgelöst. Hier sei jetzt ein Gerät gewählt, das durch elektrische Signale steuerbar ist. Die *Eingangssignale* erhalten die Großbuchstaben, also

H Halt,
V schneller Vorlauf,
R schneller Rücklauf,
W Wiedergabe,
A Taste, die vor der Wiedergabe zu betätigen ist, wenn eine Aufzeichnung zustande kommen soll.

Von diesen Eingangssignalen sind deutlich die vom Gerät jeweils ausgeführten *Funktionen* zu unterscheiden. Sie erhalten kleine Buchstaben:

h Gerät befindet sich in Haltestellung,
v Magnetband wird schnell vorwärts bewegt,
r Magnetband wird schnell rückwärts bewegt,
w Magnetband wird mit konstanter Geschwindigkeit für die Aufzeichnung und Wiedergabe transportiert,
a die Aufzeichnungs- und Löschverstärker sind eingeschaltet. Die eigentliche Aufzeichnung ist also aw.

Ferner soll *das Gerät drei Signale* als Eingangssignale für eine Steuerung *selbst erzeugen*. Sie werden ebenfalls in Großbuchstaben gekennzeichnet:

D Band bewegt sich; Wickel drehen sich,
B Band ist zu Ende oder gerissen; wird durch Lichtschranke realisiert,
S Sperre für die Aufzeichnung zum Schützen wertvoller Aufnahmen gegenüber einer Löschung.

Zwischen den eingeführten Größen bestehen nun die Relationen, die den Automaten ausmachen. Die *verbale Beschreibung* sieht etwa so aus:

H und *B* lösen immer und sofort *h* aus,

A schaltet immer die Aufzeichnungsverstärker ein. Dies wird aber sofort rückgängig gemacht, wenn eine Taste außer *W* betätigt wird oder *B* bzw. *S* erscheinen,

V, R, W führen sofort zu *v, r, w*, wenn $\overline{D}$ gilt. Wenn *D* existiert, wird zunächst *h* ausgelöst, bis $\overline{D}$ erscheint. Dann wird in die Zustände *v, r* bzw. *w* übergegangen. Dies ist also eine komplexe Verriegelung dieser Tasten zum Schutze des Bandes.

Tabelle 8.4.1. Automatentafel für eine Magnetbandsteuerung
(Besonders wichtige Übergänge sind fett gedruckt)

	H	*V*	*R*	*W*	*A*	*D*	$\overline{D}$	*B*	$\overline{B}$	*S*	$\overline{S}$
Hh	*Hh*	*Vv*	***Rr***	***Ww***	*Hha*	*Hh*	*Hh*	*Hh*	*Hh*	*Hh*	*Hh*
Vh	*Hh*	*Vh*	*Rh*	*Wh*	*Vha*	*Vh*	***Vv***	*Hh*	*Vh*	*Vh*	*Vh*
Vv	***Hh***	*Vv*	*Rh*	*Wh*	*Vva*	*Vv*	*Vv*	***Hh***	*Vv*	*Vv*	*Vv*
Rh	*Hh*	*Vh*	*Rh*	*Wh*	*Rha*	*Rh*	***Rr***	*Hh*	*Rh*	*Rh*	*Rh*
Rr	***Hh***	*Vh*	*Rr*	*Wh*	*Rra*	*Rr*	*Rr*	***Hh***	*Rr*	*Rr*	*Rr*
Wh	*Hh*	*Vh*	*Rh*	*Wh*	***AWah***	*Wh*	***Ww***	*Hh*	*Wh*	*Wh*	*Wh*
Ww	***Hh***	*Vh*	*Rh*	*Ww*	***AWaw***	*Ww*	*Ww*	***Hh***	*Ww*	*Ww*	*Ww*
Awah	*Hh*	*Vh*	*Rh*	*AWah*	*AWah*	*AWah*	***AWaw***	*Hh*	*AWah*	*AWh*	*AWah*
AWaw	***Hh***	*Vh*	*Rh*	*AWaw*	*AWaw*	*AWaw*	*AWaw*	***Hh***	*AWaw*	*AWw*	*AWaw*

Die gegebene Beschreibung verlangt einen MOORE-Automaten und führt zur *Automatentafel* von Tab. 8.4.1. Die besonders bedeutsamen Übergänge sind dort zur besseren Übersicht fett gesetzt. In der ersten Spalte für die Zustände sind die verschiedenen möglichen Kombinationen von vorher gedrückten Tasten und ablaufender Funktion des Gerätes zusammengestellt. Dies ist hier also der Einfachheit halber eine Kombination von Zuständen und Ausgangsgrößen. Es bestehen *neun* Möglichkeiten. Zu ihnen müßten eigentlich noch die ersten 6 Zustände addiert werden, die in der Spalte *A* erscheinen. Da ihnen jedoch keine spezifische Bedeutung zukommt, können sie vernachlässigt werden. Außerdem ist es für die spätere Minimierung nicht entscheidend, ob 9 oder 15 Zustände existieren. Sie erfordern in jedem Falle *mindestens 4 Flipflops*. Diese liefern $2^4 = 16$ mögliche Zustände, während 3 nur $2^3 = 8$ ergeben.

Nachdem damit nun die minimale Anzahl Flipflops feststeht, sind die 16 möglichen Zustände möglichst günstig zu kodieren, d. h. eineindeutig denen in der Automatentafel zuzuordnen. Bei *günstiger Kodierung* ist der Aufwand für die zusätzlichen kombinatorischen Schaltungen gering zu halten. Außerdem lassen sich verschiedene Störeinflüsse unterdrücken, die noch zu behandeln sind. Insgesamt bestehen hier viele Möglichkeiten, und es ist vorher schwer zu sagen, wie eine besonders günstige entsteht. Wir gehen folgenden Weg:

Aus der Wertetabelle bzw. der verbalen Beschreibung geht hervor, daß die drei Eingänge *V R W* für den Fall *D* zunächst zu speichern sind. Sie müssen also unmittelbar Flipflopzustände ansteuern. Andererseits schließen sie sich gegenseitig aus. Bei Betätigung des einen muß der andere gelöscht werden. Dies ist natürlich günstig mit den Zuständen eines Flipfloppaares zu realisieren. Dabei bleibt aber noch ein Zustand offen. Hierfür eignet sich besonders gut *Hh*, da durch ihn die anderen, *Vh, Vv, Rh, Rr, Wh* und *Ww*, gelöscht werden. Hiervon unabhängig scheint es sinnvoll, für

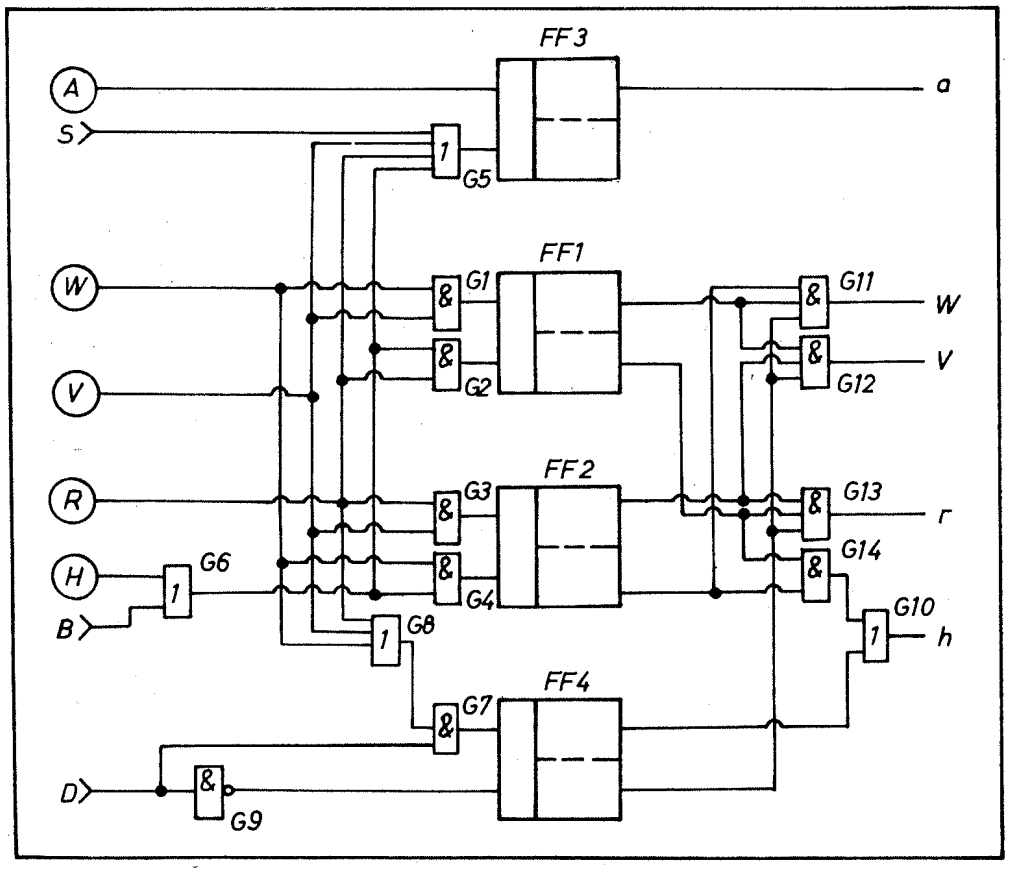


Bild 8.4.7. Beispiel einer Magnetbandsteuerung

die Taste A ein Flipflop für a und $\bar{a}$ zu wählen. Dann bleibt nur noch ein Flipflop übrig. Es hat zu realisieren, daß bei D Vh, Rh und Wh ausgelöst werden und beim Übergang zu D in Vv, Rr bzw. Ww übergehen.

Nach dieser Festlegung ist zu prüfen, ob sich eine *Struktur mit Kombinationsschaltungen realisieren* läßt. Dies müßte nicht immer der Fall sein. Das Beispiel wurde natürlich so gewählt. Die Schaltung zeigt Bild 8.4.7.

Die Gatter $G1$ bis $G4$ und $G11$ bis $G14$ kodieren die Zustände der $FF1$ und $FF2$ auf die Tasten W, V, R und H bzw. die Funktionen w, v, r und h. Der Ausgang von $FF4$ verhindert im Zustand Q über die Gatter $G11$ bis $G13$, daß w, v bzw. r wirksam werden. Gleichzeitig steuert dieser $FF4$ über $G10$ dann ein h an.

Über das Gatter $G6$ wirkt B genauso wie H. Das Gatter $G5$ realisiert die verschiedenen Rückstellungen des $FF3$ für die Aufzeichnung und zwar durch S, V, R und H.

Jetzt bleibt nur noch die Ansteuerung von $FF4$ zu beschreiben. Es sei D. Dann setzt es jede der Tasten über $G8$, $G7$ in die Stellung Q und bewirkt h über $G10$ sowie Sperrung der $G11$ bis $G13$. Sobald dadurch die Bandwickel in Ruhe kommen, erscheint $\bar{D}$ und schaltet über $G9$ den $FF4$ in $\bar{Q}$. Dadurch wird h aufgehoben und das voreingestellte Gatter von $G11$ bis $G13$ geöffnet.

8.4.7. Entwurfsprinzipien

Das vorangegangene Beispiel zeigte die wesentlichen *Stufen für den Entwurf* von Automaten auf:

— *Erfassung und Abklärung der Aufgabenstellung*

Diese Phase endet meist mit der Automatentafel. Aber auch andere Darstellungsformen wie Graph, Schaltfolgediagramm sind möglich. Die verbale Formulierung

der Aufgabe ist meist nur eine Vorstufe. In dieser ersten Stufe fällt auch die Entscheidung, ob MOORE- oder MEALY-Automat in asynchroner oder synchroner Betriebsweise.

— Der zweite Schritt ist die *Zustandsreduktion*. Sie bezieht sich immer auf die Anzahl der Flipflops. Ihr kommt aber heute durch die integrierte Technik nicht mehr so große Bedeutung wie früher zu.

— Mit der *Kodierung* erfolgt die Zuordnung der Flipflop-Zustände zu den Betriebszuständen der Schaltung. Die gewählte Kodierung beeinflußt den Aufwand der kombinatorischen Schaltung und eventuell auftretende dynamische Störungen (8.4.8.).

— Die *Festlegung der Struktur* ist die letzte Stufe des Entwurfs. Sie liefert die fertige Schaltung.

Für den Entwurf von Automaten existieren heute viele Methoden. Sie sind weitgehend rechnergestützt realisiert, verlangen jedoch an vielen Teilen Vorgaben durch den Menschen. Sie können sowohl die zu verwendenden Bauelemente als auch (der andere Extremfall) die Struktur der Schaltung betreffen. So wie bereits gezeigt wurde, daß eine kombinatorische Schaltung, z.B. der Adder, in verschiedenen Strukturen realisierbar ist, so gilt dies für Automaten in weitaus größerem Umfang. Hierauf wird u. a. noch in Abschnitt 8.5.6.1. und 8.9.3. eingegangen.

8.4.8. Fehler: Race und Rückkopplungshazards

Typische Fehler der kombinatorischen Schaltungen sind die im Abschnitt 8.2.9.2. behandelten Hazards. Für sequentielle Schaltungen treten *zwei* zusätzliche *typische Fehler* auf:

Races = Wettläufe
essentielle Hazards = Rückkopplungshazards.

Beide Fehler existieren *nur in ungetakteten Schaltungen*. Sie sind also bei den getakteten Schaltungen nicht vorhanden. *Races* können existieren, wenn bei einer sequentiellen Schaltung gleichzeitig mehrere Eingangsvariable geändert werden und diese nicht genau gleichzeitig, sondern zeitlich gegeneinander versetzt eintreffen. Dann laufen die beiden Flanken sozusagen um die Wette, wer zuerst am Eingang ist und damit zuerst

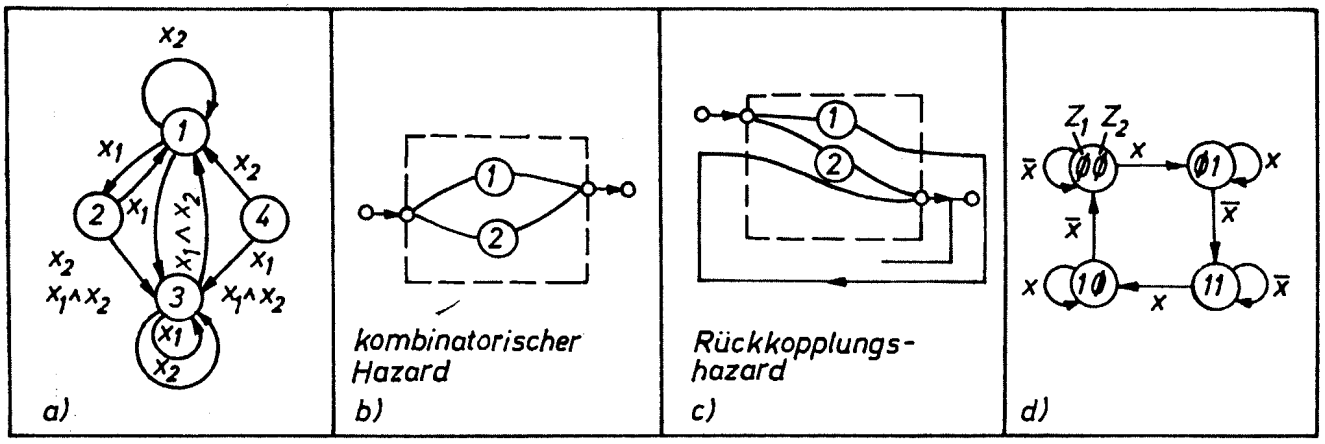


Bild 8.4.8. Zu den Fehlern in sequentiellen Schaltungen

a) Graph, bei dem nur der Zustand ① frei von kritischen Races ist
b) Kombinatorisches Hazard in schematischer Darstellung mit den verschiedenen Wegen ① und ②
c) In gleicher Darstellung wie b) Prinzip für ein Rückkopplungshazard
d) Graph mit Rückkopplungshazards

die Möglichkeit hat, die Schaltung zu beeinflussen. Dies sei am Beispiel von Bild 8.4.8a) erklärt.

Es existieren die vier Zustände 1 bis 4 und die zwei Eingangsvariablen x_1 und x_2, die einzeln oder zugleich auftreten können. Für die Fälle, wo sie zugleich auftreten können, besteht die Gefahr von Races. Die Gefahr ist in den einzelnen Ausgangszuständen unterschiedlich:

Von 1 aus gibt es keine störenden Races. Sobald x_1 und x_2 gleichzeitig vorliegen, ist der Zustand 4 erreicht und wird auch nicht verlassen, wenn ein Signal allein zeitlich nachhinkt. Der Wettlauf kann aber direkt nach 4 oder über 2 nach 4 führen. Dieser Wettlauf wird als unkritisch bezeichnet.

Für die Zustände 2, 3 und 4 existieren jeweils kritische Wettläufe. Bei 2 und 3 wird 1 oder 4, bei 4 1 oder 2 erreicht.

In diesem Sinne liegt auch bei der Schaltung gemäß Bild 8.4.3 ein kritischer Wettlauf vor. Wird im Zustand Ø1 gleichzeitig x_1 und x_2 betätigt, so wird der Zustand 1Ø angenommen, der auf keine andere Weise erreichbar und auch nicht wieder zu verlassen ist. Zuerst x_1 und dann x_2 führen nach 11, zuerst x_3, dann x_1 nach ØØ. An diesem Beispiel ist ein Grundprinzip aller Races zu erkennen. Sie treten dann ein, wenn zwei Zustände mit einem Pfeil verbunden sind, deren Zustandskodierung in mehr als einem Wert abweicht. Dies ist der Grund, weshalb die Zustandskodierung die Möglichkeit von Races beeinflußt. Races treten nicht auf, wenn nur Zustände verbunden sind, die sich in einem Zustandswert unterscheiden. Dies kann immer, eventuell unter Hinzunahme zusätzlicher Übergangszustände, erreicht werden (vgl. [S 54, Bd. 3]).

Rückkopplungshazards können immer dann auftreten, wenn eine Ausgangsgröße zugleich von der Rückkopplung und von einem Eingangssignal herrührt. Dies zeigt schematisch Bild 8.4.8c) im Vergleich zum kombinatorischen Hazard in Teilbild b). Ein Beispiel für ein solches Rückkopplungshazard zeigt Bild 8.4.8d). Es existieren vier Zustände ØØ bis 11. Sie werden von einer Eingangsvariablen x bzw. $\bar{x}$ gesteuert. Vom Zustand Ø1 gelangt man zum Zustand 11 durch $\bar{x}$. Fällt dann infolge Zeitverzögerungen kurzfristig x aus, also $\bar{x}$, so gelangt man faktisch über 1Ø nach ØØ. Rückkopplungshazards liegen im Prinzip der Schaltung begründet und können weder durch geschickte Zustandskodierung noch durch kompliziertere Schaltungsstrukturen unterbunden werden.

8.5. Einfache sequentielle Schaltungen

8.5.1. Impulsschaltungen

Die Grundlagen der verschiedenen Impulsschaltungen wurden bereits im Zusammenhang mit Bild 7.3.25 bei der analogen Schaltungstechnik behandelt. Sie können aber auch aus digitaler Sicht betrachtet werden. Dann sind zwei Grundgatter, z.B. NAND oder NOR mitgekoppelt zusammenzuschalten. Je nach der *Art der Rückkopplung*,

- galvanisch, d. h. gleichstrommäßig, oder
- über Zeitglieder, z.B. Kondensatoren,

ergeben sich *zwei Gruppen von Schaltungen*:

- Flipflop (bistabile Speicherschaltungen, Latch) bzw. SCHMITT-Trigger
- Uni- bzw. Multivibratoren.

Infolge der großen Bedeutung der Flipflops und ihrer Vielzahl an Varianten ist ihnen der Abschnitt 8.5.2. gewidmet. Für die anderen drei Bauelemente kann hier nur eine kleine Auswahl gegeben werden. Mehr Details enthalten u. a. [K 30, S 67].

Die drei hier zu behandelnden Schaltungen faßt Bild 8.5.1 zusammen. Sie können alle drei so beschrieben werden, daß sie *zwei interne Zustände* besitzen, die einzeln oder gemeinsam, *stabil* oder *quasistabil* mit Verweilzeit (dynamisch stabil) sind.

Beim SCHMITT-*Trigger* sind sie beide stabil, welcher Zustand eingenommen wird, hängt (über eine Hysterese) von der Eingangsspannung ab. Es können daher die Eingangsspannungen für H bzw. L gemäß den Formeln im Bild näherungsweise bestimmt werden.

Beim *Univibrator* ist nur ein Zustand stabil. Der andere kann durch ein Eingangssignal, z.B. einen kurzen Impuls, erzwungen werden und bleibt dann für eine Zeit T bestehen. Ein spezieller integrierter Univibrator mit guter Temperaturkonstanz ist der Schaltkreis 74121. Er besitzt drei Eingänge mit unterschiedlicher Auslösefunktion. Dies zeigt die Tabelle im Bild.

Multivibratoren können unterschiedlich aufgebaut werden. Drei Beispiele zeigt das Bild 8.5.1. Die Schaltung mit dem Steuereingang St kann mittels einer Spannung von (5 ± 0,3) V um etwa ±1% in der Frequenz verstellt werden. Die Schaltung mit dem Quarz zeichnet sich durch hohe Temperaturkonstanz aus. Sie kann mit der Kapazität um geringe Beträge gezogen werden.

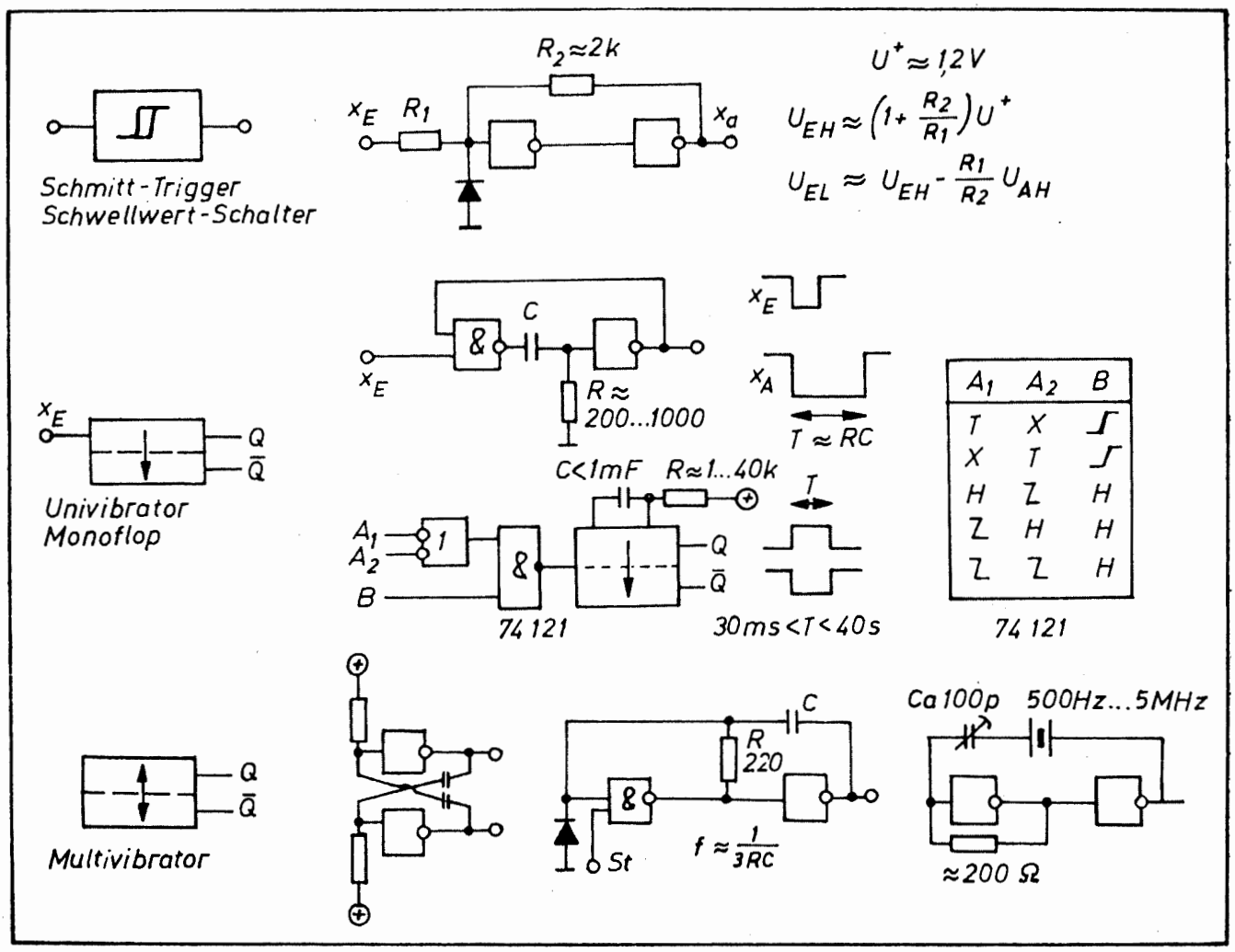


Bild 8.5.1. Zusammenstellung der wichtigsten Impulsschaltungen auf der Basis von logischen Gattern, hier insbesondere der TTL-Schaltkreise

Abschließend sei darauf hingewiesen, daß es auch universelle Schaltungen bereits integriert gibt. Ein Beispiel ist der Timer 555. Er kann z.B. als Multivibrator oder Univibrator in verschiedenen Betriebsweisen genutzt werden. Vgl. z.B. [S 67].

8.5.2. Flipflop

Es gibt eine solche Vielzahl von Flipflops (= FF), daß ihre Systematik schwierig ist. Sie kann wie folgt aufgelistet werden:

Das normale FF besitzt *zwei Ausgänge* Q und $\overline{Q}$, welche die Zustände Ø und 1 (L und H) annehmen können. Sie entsprechen den *inneren Zuständen* des FF, beim später einzuführenden Master-Slave-FF denen des Slave.

Die *Eingänge* der FF müssen in zwei Gruppen eingeteilt werden:

- *Informationseingänge.* Zu ihnen gehören die Bezeichnungen R, S, D, T, J, K, Z — welche die Eingangssignale in unterschiedlicher Weise wirken lassen (vgl. 8.5.2.1.).
- *Takt- bzw. Vorbereitungseingänge.* Sie ermöglichen, daß die Informationssignale wirksam werden können oder nicht. Hierzu gehören die Bezeichnungen C (von clock = Takt), auch T (wird vermieden wegen T als Informationseingang) und V. In diesen Begriffen ist leider nicht enthalten, wie der Takt wirkt. Dies wird im Abschnitt 8.5.2.4. behandelt.

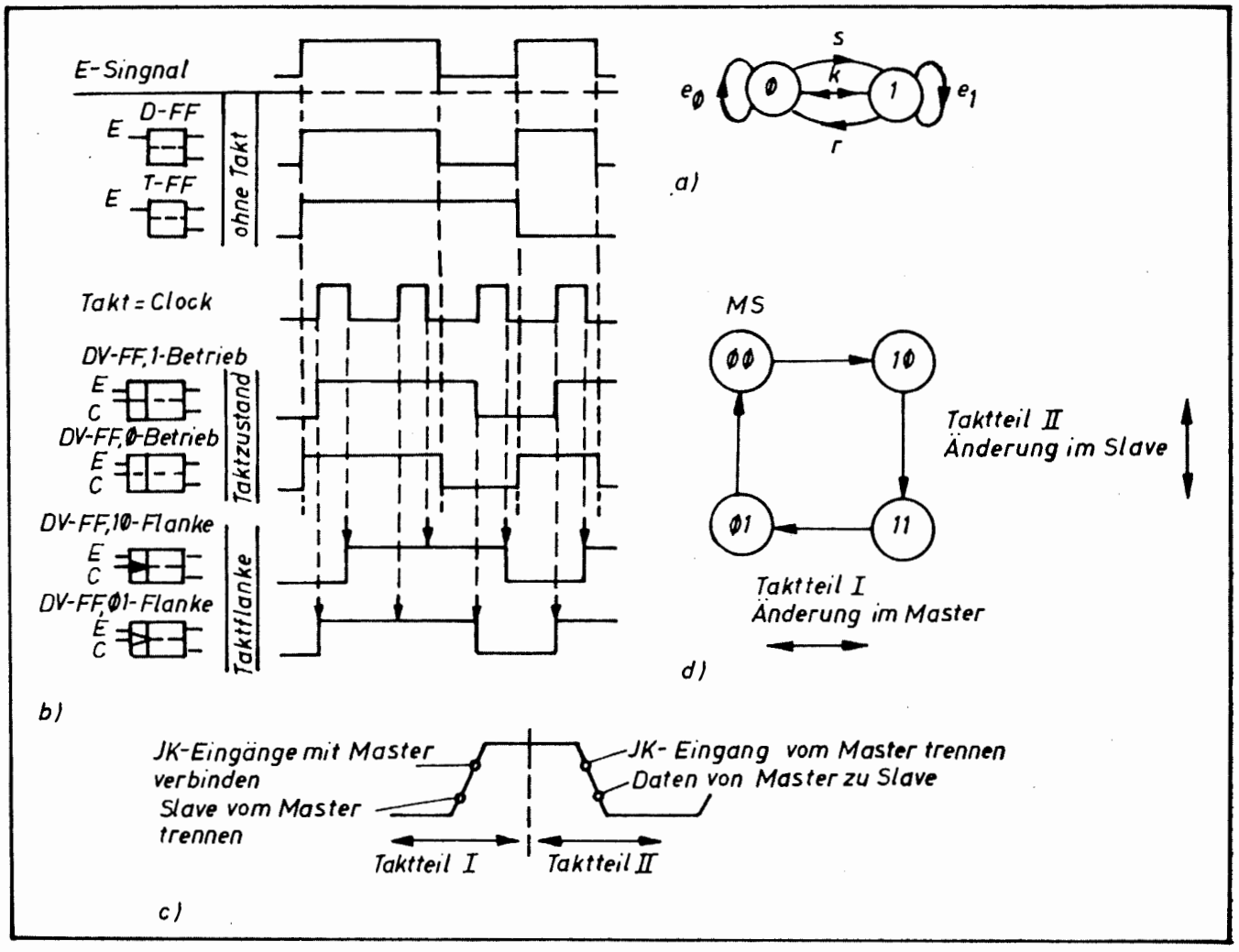


Bild 8.5.2. Zur Steuerung von Flipflops (FF)

a) Die vier Funktionen eines FF: e, k, r und s

b) Steuerung von verschiedenen FF-Typen; oben D-FF und T-FF, darunter die vier verschiedenen Taktsteuerungen

c) Wirkung der Taktsteuerung beim Master-Slave-FF (MS-FF), z.B. gemäß Bild 8.5.3j)

d) Graph für einen MS-FF

8.5.2.1. Funktionen von FF

Das einfachste FF ist das RS-FF, wie es bereits im Zusammenhang mit Bild 8.3.5 und 7.3.25 eingeführt wurde. Es trägt auch die Namen bistabile Kippstufe, Latch, bistabiler Multivibrator und Speicher. Es ist dadurch gekennzeichnet, daß es zwei stabile innere Zustände einnehmen kann. Sie seien mit Ø und 1 bezeichnet und sollen für den Ausgang Q gelten. Zwischen diesen *beiden Zuständen* gibt es ganz allgemein, also für eine beliebige Ansteuerung eines FF *vier mögliche Übergänge* bzw. *Funktionen*, wie sie Bild 8.5.2a) zeigt:

e: Das Flipflop ist im Zustand Ø oder 1, und bei einem Eingangsimpuls bleibt es auch in diesem Zustand. Dieser Fall werde mit *Erinnern von* Ø oder 1 bezeichnet.
s: Das Flipflop ist im Zustand Ø und wird durch einen Eingangsimpuls nach 1 überführt. Hier liegt *Setzen* des Flipflop vor (engl. set, zuweilen auch preset).
r: Das Flipflop ist im Zustand 1 und wird durch einen Impuls nach Ø überführt. Hier liegt *Rücksetzen* vor (engl. reset, zuweilen auch clear).
k: Das Flipflop ist in Ø oder 1 und wird durch einen Impuls in den anderen Zustand überführt. Hier liegt *Kippen* vor.

Das *RS-FF* verwendet hiervon nur die drei Funktionen e, r und s. Sie treten auf, wenn an den beiden Eingängen R, S die Werte Ø; Ø bzw. 1; Ø bzw. Ø; 1 liegen. Bei ihm ist die noch fehlende Eingangskombination 1; 1 verboten. Dabei würden nämlich Q und auch $\overline{Q}$ den Wert 1 annehmen, und das ist unzulässig. Außerdem würde das FF nach dem Aufhören dieses Signals in einen undefinierten Zustand übergehen.

8.5.2.2. Universelles FF, D-FF und T-FF

Das *universelle* und jedes beliebige FF kann gemäß Bild 8.5.3b) aus dem *RS-FF* und einer *vorgeschalteten kombinatorischen Schaltung* abgeleitet werden, wobei in die kombinatorische Schaltung zuweilen die Ausgangszustände Q und $\overline{Q}$ zurückgeführt werden. Die Anzahl n der dann außen vorhandenen Eingänge bestimmt die Vielzahl der möglichen FF.

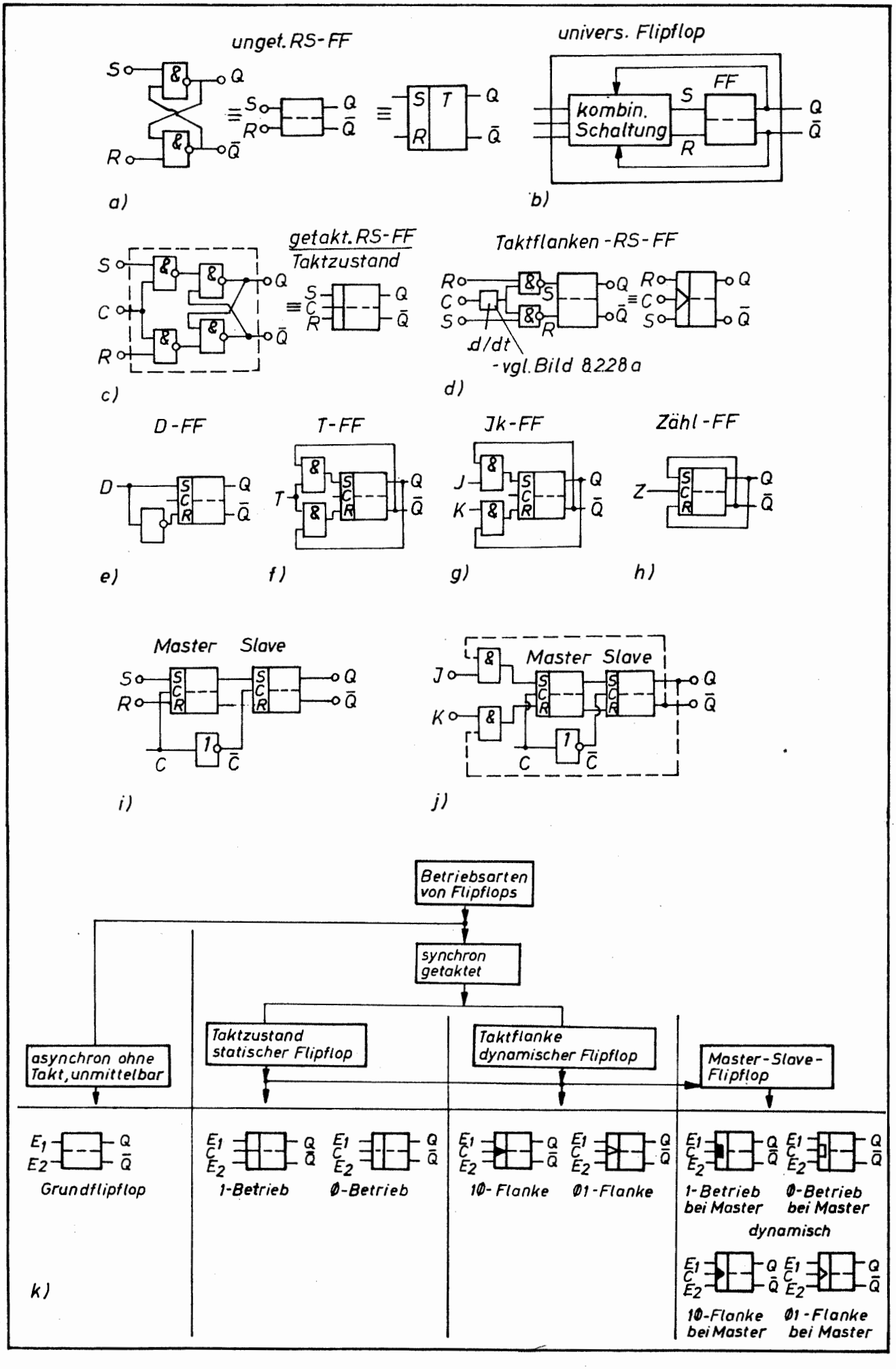
Besonders einfach ist ein FF, wenn es nur einen Eingang besitzt. Dann kann dieses Eingangssignal die beiden Zustände Ø und 1 annehmen. Ihr können in unterschiedlicher Kombination die vier Funktionen e, s, r, k zugeordnet werden. Dies bedeutet, daß $2^4 = 16$ mögliche FF existieren. Darunter sind auch solche trivialen FF wie jenes, das bei keinem Eingangssignal schaltet, bzw. jenes, das bei jedem Eingangssignal schaltet (Multivibrator). Praktische Bedeutung haben jedoch nur das D-FF und das T-FF erreicht.

Bei den *D-FF* (von delay = Verzögerung) folgt der Q-Ausgang dem Signal des *D-Eingangs*. Es heißt daher auch Auffang-FF oder Zwischenspeicher. Es entsteht gemäß Bild 8.5.3e) dadurch aus dem RS-FF, indem die beiden Eingänge über einen

▶

Bild 8.5.3. Überblick zu verschiedenen FF-Typen

a) Aufbau und zwei Symboldarstellungen für einen RS-FF. Die rechte Variante ist neuer, wird hier aber wegen einiger Probleme in der Darstellung nicht verwendet
b) Universelles FF aus RS-FF mit vorgeschalteter kombinatorischer Schaltung
c) Taktzustandsgesteuertes RS-FF
d) Taktflankengesteuertes RS-FF
e) bis h) Die Haupttypen von FF gemäß der Wirkung der Eingangsgrößen
i) und j) Zwei Beispiele von MS-FF
k) Systematische Darstellung aller Taktarten bei FF


unget. RS-FF
univers. Flipflop
kombin. Schaltung
FF
a)
b)
getakt. RS-FF
Taktzustand
Taktflanken-RS-FF
d/dt
-vgl. Bild 8.228a
c)
d)
D-FF
T-FF
Jk-FF
Zähl-FF
e)
f)
g)
h)
Master Slave
i)
j)
Betriebsarten von Flipflops
synchron getaktet
asynchron ohne Takt, unmittelbar
Taktzustand statischer Flipflop
Taktflanke dynamischer Flipflop
Master-Slave-Flipflop
Grundflipflop
1-Betrieb
Ø-Betrieb
10-Flanke
Ø1-Flanke
1-Betrieb bei Master
Ø-Betrieb bei Master
dynamisch
10-Flanke bei Master
Ø1-Flanke bei Master
k)

Inverter verkoppelt sind. Das *T-FF* (von Trigger) schaltet immer dann um, wenn der *T*-Eingang 1 ist. Dadurch entsteht, wie aus Bild 8.5.2b) (oben) zu sehen, am Ausgang die halbierte Frequenz des Eingangssignals, d. h., jeder zweite Eingangsimpuls erscheint nur an seinem Ausgang. Daher die englische Bezeichnung scale of two (= durch 2 teilen) bzw. die deutsche Frequenzteiler oder spezieller *Zähl-FF* (Z-FF), welche aber in der Regel noch einen Takteingang erhalten. Das Realisierungsschema eines T-FF aus einem RS-FF zeigt Bild 8.5.3f). Auffällig ist die oben allgemein erwähnte Rückführung vom Ausgang.

8.5.2.3. Anzahl der möglichen FF

Liegen n Eingänge vor, so existieren 2^n verschiedene Eingangszustände. Auch sie können wieder kombinatorisch den vier Funktionen zugeordnet werden. Dadurch entstehen $(2^n)^4$ mögliche FF. Dies bedeutet:

n:	1	2	3	4	5	6
FF:	16	256	4096	65536	1048576	16777216

Da es heute FF mit sogar mehr als 6 Eingängen gibt, ist die theoretische Zahl der möglichen FF extrem groß. Sie wird noch weiter vergrößert durch die verschiedenen Varianten der im nächsten Abschnitt zu behandelnden *Taktarten* und durch die zusätzliche Variante des *Master-Slave-FF*. In der Praxis wird aber die Anzahl — wie bereits beim FF mit einem Eingang gezeigt — drastisch eingeschränkt. Hierfür gelten folgende Fakten:

Weglassen trivialer FF,
Unterscheiden von Takt- und Signaleingang,
Auswählen spezieller FF, aus denen die anderen aufgebaut werden können.

8.5.2.4. Takt- und Vorbereitungseingang

Für das normale FF (ohne Takteingang) bestehen einige Probleme bezüglich der Stabilität:

Wenn das Eingangssignal, z.B. beim D-FF, kurzfristig ausfällt, so ändert sich damit auch der Zustand des FF.
Sollen mehrere Signale am Eingang eines FF verknüpft werden, so können Laufzeitdifferenzen, wie sie z.B. bei den Hazards (8.2.9.2.) besprochen wurden, ebenfalls nachteilige Folgen zeigen.
Bei der Rückkopplung über einen oder mehrere FF können sich Races bzw. Rückkopplungshazards ausbilden.

Die ersten beiden und z. T. auch der dritte Fehler lassen sich unterdrücken, wenn FF nur zu bestimmten Zeiten empfindlich gestaltet werden. Dies kann z.B. gemäß Bild 8.5.3c) bei einem RS-FF durch zwei zusätzliche NAND-Glieder erreicht werden. Der *Takteingang* wird nun unterschiedlich mit *T* (von Takt; sollte vermieden werden!) oder mit *C* (von clock) bezeichnet. Auch eine Kennzeichnung mit *V* (Vorbereitungseingang) ist möglich. Sie wird besonders dann verwendet, wenn noch ein zweiter Steuereingang vorhanden ist, der mit dem Takt durch ein logisches Und verknüpft ist. Zuweilen wird auch die Beziechnung Freigabeeingang (engl. enable) verwendet.

Nach dem Geschilderten entsteht aus einem D-FF ein DV-FF, wie es im Prinzip in Bild 8.5.3e) (oft auch nur als D-FF bezeichnet) gezeigt ist. Sein Verhalten bez. Takt und Eingangssignal zeigt die Mitte von Bild 8.5.2b). Dort sind zugleich zwei verschiedene DV-FF gezeigt: einmal wird das *E*-Signal beim 1-Signal des Taktes über-

nommen, zum anderen beim Ø-Signal. In dem gezeigten Beispiel fällt auf, daß das zweite DV-FF auch während des Taktes schaltet. Dies ist allgemein unerwünscht, da es zu Störungen führen kann. Deshalb wurde neben dem bisher behandelten taktzustandsgesteuerten FF noch das *taktflankengesteuerte* FF entwickelt. Auch dieses existiert gemäß Bild 8.5.2b) (unten) in zwei Varianten. Da es immer nur während der Flanken (mit Pfeilen gekennzeichnet) empfindlich ist, können hier die oben genannten Störungen nicht auftreten. Der technische Aufwand ist für ein flankengesteuertes FF allerdings erheblich höher als für ein taktzustandsgesteuertes. Eine Prinzipschaltung zeigt Bild 8.5.3d), wobei das Differenzierglied oft durch zwei Gatter gemäß Bild 8.2.28a) realisiert wird.

Eine andere Variante benutzt Schaltungen mit zusätzlichen FF (vgl. das spätere Bild 8.5.4). Sie werden dann auch zuweilen Vorspeicher genannt.

8.5.2.5. JK-FF

Das JK-FF ist ein recht universelles FF mit 2 Signaleingängen (Bild 8.5.3g)). Es hat die Eingänge so gestaltet, daß *alle vier Funktionen e, s, r, k* mit einer Kombination *J*; *K* gemäß ØØ; 1Ø; Ø1 und 11 realisiert werden. Es stellt also eine sinnvolle Erweiterung des RS-FF dar.

Über die Bezeichnung der Eingänge mit *J* und *K* gibt es unterschiedliche Auffassungen. Einmal wird in der Literatur der Standpunkt vertreten, daß sie PHISTER willkürlich eingeführt habe. Eine andere Meinung besagt, daß sie zu Ehren des Erfinders der integrierten Schaltung, JACK KILBY, von Texas Instruments gewählt seien.

Die hohe Variabilität des JK-FF liegt neben der Ähnlichkeit mit dem RS-FF auch darin, daß durch Zusammenschalten der beiden Eingänge der T-FF entsteht.

8.5.2.6. Master-Slave-FF

Werden zwei getaktete RS-FF in Reihe geschaltet und im Takt gegenphasig gesteuert, so entsteht das RS-MS-FF (MS von Master-Slave), wie es Bild 8.5.3i) zeigt. Üblicher ist ein daraus abgeleitetes JK-MS-FF gemäß Bild 8.5.3j). Beide besitzen gegenüber den bisher behandelten FF ein erheblich verändertes Verhalten. Dies geht aus Bild 8.5.2c) hervor. Mit dem Ansteigen der Taktflanke wird zunächst das zweite Teil-FF, der Slave (Sklave), unempfindlich gemacht, d. h., er wird vom ersten Teil-FF, dem Master (Herr, Meister), abgetrennt. Etwas später, auf der ansteigenden Flanke, wird der Master empfindlich geschaltet, d. h. mit den Eingängen, z.B. *JK*, verbunden. Er übernimmt dann die dort vorhandene Information in seinen Zustand. Mit der abfallenden Flanke wird zunächst der Master wieder von den Eingängen abgeschaltet. Danach wird der Slave empfindlich und übernimmt den Zustand des Masters als den seinen. Damit ist auch die Bezeichnung Master-Slave verständlich. Der Master wird durch die Eingangsinformation gesetzt, und der Slave kann nur tun, was der Herr befiehlt.

Ein MS-FF hat infolge der zwei Teil-FF eigentlich vier mögliche Zustände. Sie können aufgrund des Vorangegangenen aber nur in einer bestimmten Weise durchlaufen werden. Dies zeigt das Bild 8.5.2d) als Graph.

Der Vorteil der MS-FF besteht also in der weitgehenden Trennung von Eingangs- und Ausgangsvariablen. Hierdurch lassen sich verschiedene Rückkopplungshazards und ähnliche Störungen gut unterdrücken. Das im Bild 8.5.3h) gezeigte Zähl-FF ist meist aus einem JK-MS-FF abgeleitet.

Auch bei den MS-FF gibt es taktzustands- und taktflankengesteuerte Varianten. Unter dieser Berücksichtigung gibt Bild 8.5.3k) einen systematischen Überblick zu

den Steuertypen von FF, während Bild 8.5.3e) bis h) die vier Haupttypen von FF bez. des Eingangs zusammenfaßt.

8.5.2.7. Realisierte Beispiele für DV-FF und JK-MS-FF

Bild 8.5.4 zeigt verschiedene Varianten eines taktflankengesteuerten D-FF. Teilbild a) zeigt die Innenschaltung des D 174 (7474) der TTL-Serie. Dieser Schaltkreis enthält gleich zwei solcher FF. Die logische Struktur gibt Teilbild b) wieder. Es sind deutlich drei Teil-FF aus je zwei NAND-Gliedern zu erkennen. Durch diese komplexe Schaltung wird sowohl die Vorspeicherung von D als auch die Übertragung dieses Wertes bei der Taktflanke erreicht.

Bei I²L bzw. MOS werden dieselben Funktionen gemäß Teilbild c) durch NOR-Glieder realisiert. Im Gegensatz zur MOS-Schaltung (Teilbild e)) wurden bei der I²L-

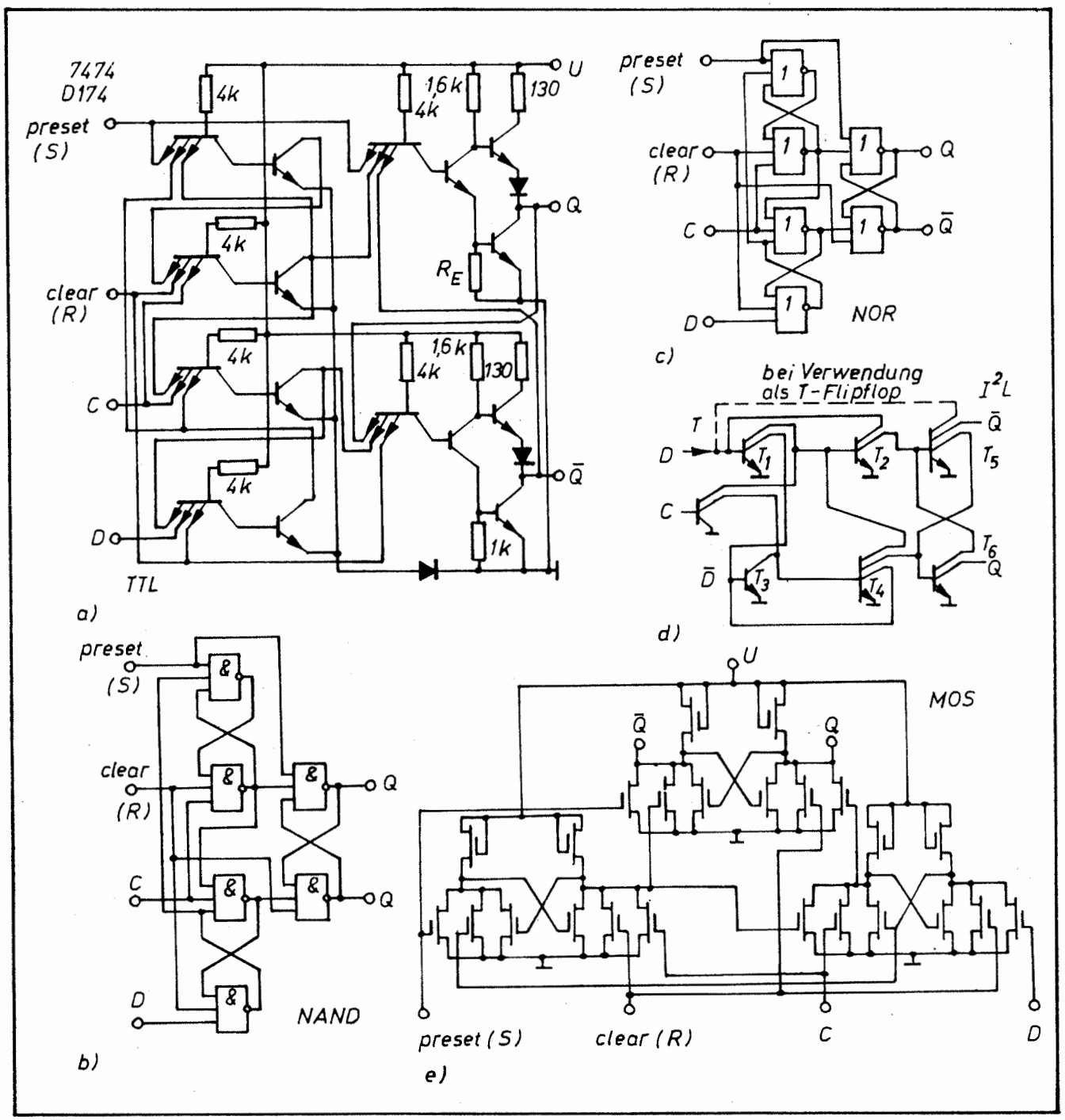


Bild 8.5.4. Varianten von taktflankengesteuerten D-FF

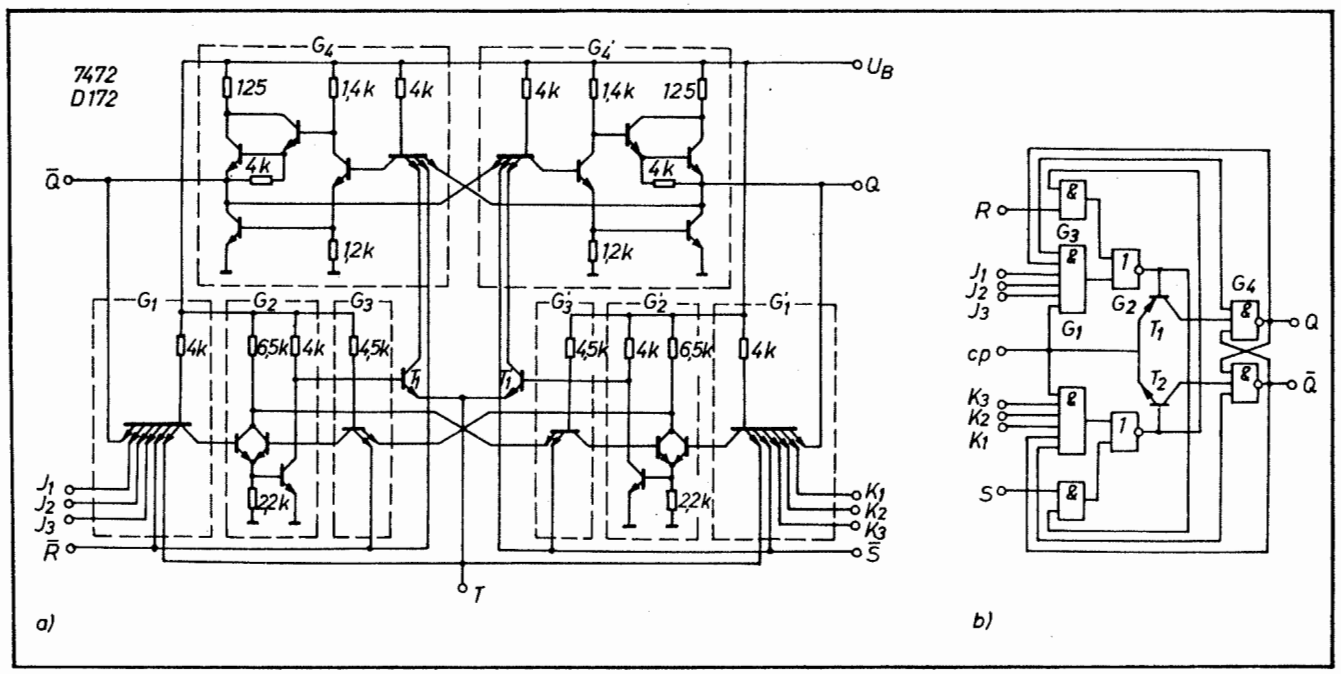


Bild 8.5.5. a) Schaltbild und b) logischer Aufbau des JK-MS-FF D 172

Schaltung die Eingänge R und S zur Vereinfachung fortgelassen. Dafür ist aber angedeutet, wie zu einem T-FF überzugehen ist.

Den Aufbau des JK-MS-FF vom TTL-Typ D 172 (7472) zeigt Bild 8.5.5a), während Teilbild b) die Logikstruktur zeigt. Dabei wird deutlich, wie mit den Transistoren T_1 und T_2 sowohl der Takt negiert als auch eine OR-Verknüpfung realisiert wird. Aus dieser Schaltung wird besonders deutlich, wie groß der schaltungstechnische Aufwand für komplexe FF ist.

8.5.3. Frequenzteiler

Frequenzteiler sind dadurch gekennzeichnet, daß sie je einen Ein- und Ausgang besitzen. Am Ausgang erscheint der Eingangstakt mit einem n-tel der Frequenz. Es werden also $n - 1$ Takte zum Ausgang hin unterdrückt. Meist wird dabei kein symmetrisches (gleiche positive und negative Intervallänge) Ausgangssignal verlangt. Wichtig ist dagegen die Unterscheidung in *asynchrone und synchrone Frequenzteiler*. Beim synchronen Frequenzteiler werden die Takteingänge aller untersetzenden FF gemeinsam aus dem Eingangssignal versorgt. Dadurch ist im Gegensatz zu den asynchronen Frequenzteilern ein erheblich höherer Schaltungsaufwand notwendig. Deshalb sei zunächst der asynchrone Frequenzteiler behandelt.

Als *Grundbauelement* aller folgenden Frequenzteiler wird das JK-MS-FF vorausgesetzt. Durch das MS-Prinzip können Stabilitätsprobleme vermieden werden. Mittels der JK-Eingänge können Taktsignale unterdrückt werden. Liegen nämlich alle JK auf H, d. h., sind sie nicht angeschlossen, funktioniert das JK-Flipflop vom Takteingang wie ein Zähl-FF.

8.5.3.1. Asynchrone Frequenzteiler

Das Grundprinzip und Grundbauelement aller Frequenzteiler ist das Zähl-FF (JK-MS-FF bzw. T-FF). Es untersetzt die Eingangsfrequenz auf die Hälfte (Bild 8.5.6a)). Eine Kettenschaltung aus ihm mit m Stufen (Bild 8.5.6b)) ermöglicht so eine maxi-

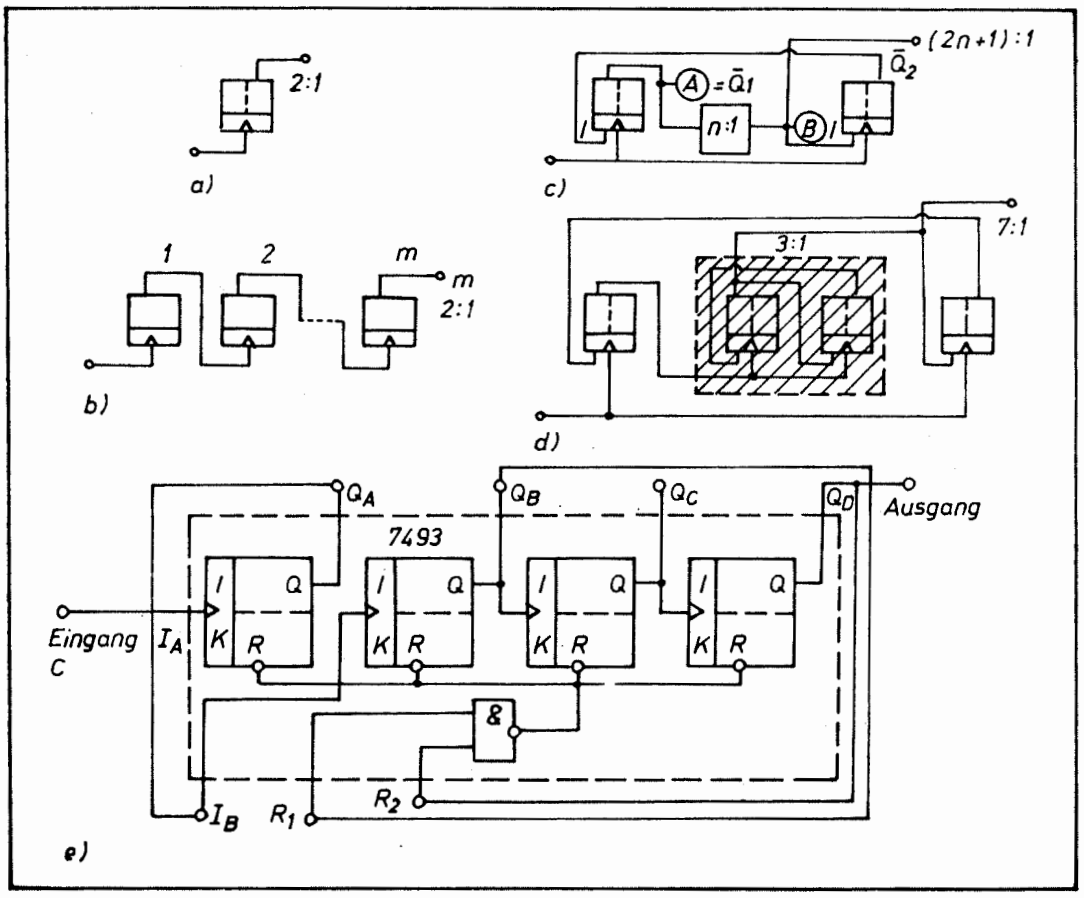


Bild 8.5.6. Grundschaltungen der asynchronen Frequenzteiler

a) Flipflop als Teiler 2:1, auch Zähl-Flipflop genannt
b) Reihenschaltung von Zähl-Flipflops mit dem Teilerverhältnis 2^m:1
c) Teiler mit dem Verhältnis $(2n + 1)$:1
d) Aus b) und c) lassen sich beliebige Teilerverhältnisse realisieren, hier zweimal c) in einandergeschaltet mit 7:1
e) 10:1-Teiler, geschaltet aus dem integrierten Schaltkreis 7493, der Verhältnisse von 1:2 bis 1:15 realisieren kann

male Frequenzteilung von 2^m:1. Durch Variation dieses Prinzips, indem bestimmte Takte unterdrückt werden, können nur kleinere Teilungsfaktoren erreicht werden. Deshalb gilt für die minimale Anzahl m von FF für einen beliebigen Teilungsfaktor n:1

$$m \geqq \mathrm{ld}\, n\,. \tag{1}$$

Eine besondere Bedeutung für beliebige Teilerverhältnisse besitzen die Primzahlen. Aus ihnen läßt sich durch Reihenschaltung (multiplikativ) jedes Teilungsverhältnis aufbauen. Hierfür ist es bedeutsam, daß mit der Schaltung von Bild 8.5.6c) eine allgemeine Schaltung für ungerade Teilerverhältnisse vorliegt. Sie besteht aus einem beliebigen inneren Teiler mit dem Verhältnis n:1 und zwei äußeren JK-FF. Wird in dieser Schaltung die Rückführung vom zweiten FF zum ersten zunächst weggelassen, so entsteht an der Stelle A eine Teilung von 2:1 und damit über den Teiler n:1 eine Teilung $2n$:1. Weiter ist daran zu erinnern, daß JK-Flipflops genau dann mit dem Takt schalten, wenn an ihrem J-Eingang ein H liegt, bei L schalten sie hingegen nicht. Da am ersten Flipflop durch ein L am J-Eingang nur Impulse unterdrückt werden können, muß die Schaltung ein Teilerverhältnis von größer als $2n$:1 besitzen.

Die Betrachtung der Schaltung werde begonnen, wo B gerade auf L liegt. In diesem Zustand arbeitet das zweite Flipflop nicht. Also wird B genau nach n Takten umschalten (Hälfte von $2n$). Dadurch wird auch das zweite Flipflop aktiv und sperrt einen Takt später das erste Flipflop. Einen weiteren Takt später wechselt wieder das zweite Flipflop sein Ausgangssignal auf H, und von nun an schalten beide Flipflops

synchron, bis nach abermals n Takten bei A wieder ein Vorzeichenwechsel auftritt. Insgesamt sind bis dahin folgende Takte vergangen:

— n Takte, wo nur das erste Flipflop schaltet,
— 1 Takt, wo das erste Flipflop gesperrt ist,
— n Takte, wo beide Flipflops synchron schalten.

Das sind $2n + 1$ Takte.

Mit der Schaltung von Bild 8.5.6c) sind durch Einfügen von b) als Frequenzteiler sofort die Verhältnisse 3:1 (ohne innere Frequenzteiler, d. h. $n = 1$); 5:1; 9:1; 17:1 usw. zu realisieren. Durch zusätzliches Vor- oder Nachschalten einer oder zweier Stufen von Schaltung b) entstehen dann 6; 10; 18; 34 bzw. 12; 20; 36; 68. Wie dieses Prinzip angewendet wird, sei noch für den Teiler 7:1 angegeben. Hierbei ist der auf Schaltung c) beruhende Teiler 3:1 noch einmal in die Schaltung d) eingefügt. Es gibt einen integrierten Schaltkreis 7493, der durch äußere Beschaltung im Bereich von 2:1 bis zu 15:1 zu programmieren ist. Sein Prinzip mit der äußeren Beschaltung für 10:1 zeigt Bild 8.5.6e).

8.5.3.2. Synchrone Frequenzteiler

Die Grundschaltung des synchronen Frequenzteilers zeigt Bild 8.5.7 oben. Durch Verschaltung zwischen den Ausgängen Q_k und $\bar{Q}_k$ mit den J- bzw. K-Eingängen ma bzw. mb ist dafür zu sorgen, daß eine bestimmte Anzahl von Takten unterdrückt wird. Im einfachsten Fall geschieht dies wie beim asynchronen Teiler des Bildes 8.5.6c). Für $n = 1$ ergibt sich bei ihm ja 3:1, und er ist dann auch ein synchroner Frequenzteiler. Die Bestimmung der Verschaltungen ist im allgemeinen nicht ganz einfach.

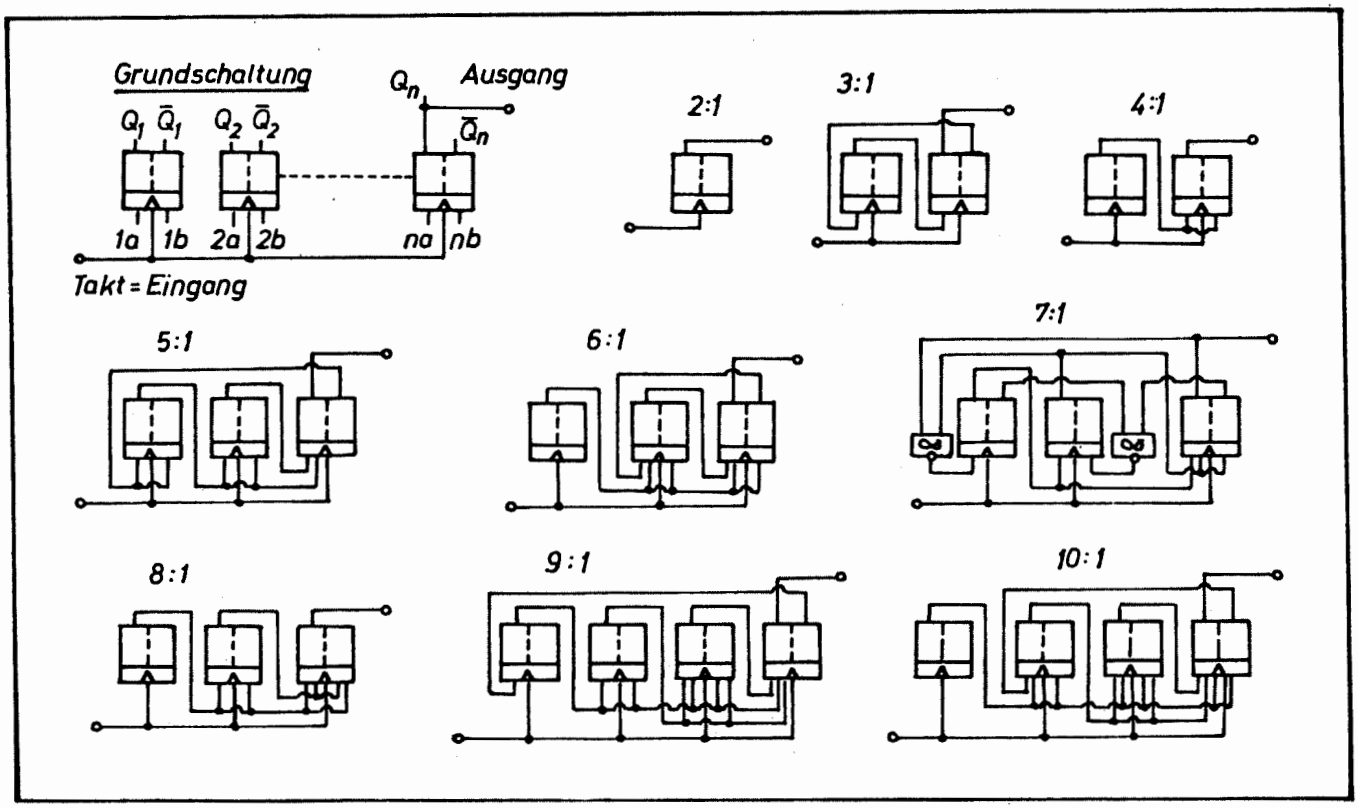


Bild 8.5.7. Prinzipschaltung und Schaltungen für die Verhältnisse 2 bis 10 zu 1 der synchronen Frequenzteiler, die auch Modulo-n-Zähler genannt werden. Durch entsprechendes, meist mehrfaches Verbinden von Ausgängen Q_k bzw. $\bar{Q}_k$ mit den J- bzw. K-Eingängen ma bzw. mb werden die Teilungsverhältnisse realisiert

Deshalb sind in Bild 8.5.7 die synchronen Frequenzteiler von 2:1 bis 10:1 angegeben. Für die synchronen Frequenzteiler gibt es leider keine Reduktionsregel wie bei den asynchronen. Sie werden um so komplizierter, je größer das Teilungsverhältnis ist.

8.5.4. Zählerschaltungen

8.5.4.1. Grundprinzip

Zählerschaltungen entsprechen weitgehend den Frequenzteilern. Sofern mit ihnen nur eine Anzahl bestimmt werden soll, genügt ein Frequenzteiler vollständig. Daher heißen die synchronen Zählschaltungen auch Modulo-*n*-Zähler. *Asynchrone Zählschaltungen sind nur bedingt geeignet.* Bei ihnen schaltet nämlich jede folgende Stufe gegenüber der vorangehenden etwas verzögert. Es besteht daher die Gefahr, daß während der mehrfachen Verzögerung die erste Stufe bereits durch inzwischen eingegangene Impulse geschaltet hat und so die Zählung falsch ist.

In der Regel haben Zähler gegenüber Frequenzteilern die Anzahl der aufgenommenen *Impulse zu jedem Zeitpunkt anzuzeigen.* Hierzu können Signale aus den Zuständen der einzelnen Flipflops parallel abgeleitet und dann nach einer *Dekodierung der Zustände* angezeigt werden. Weiter besitzt ein allgemeiner Zähler einige zusätzliche Funktionen, die in einem speziellen Zähler nicht alle vorhanden sein müssen. Das entsprechende vollständige Prinzipschaltbild eines Zählers zeigt Bild 8.5.8.

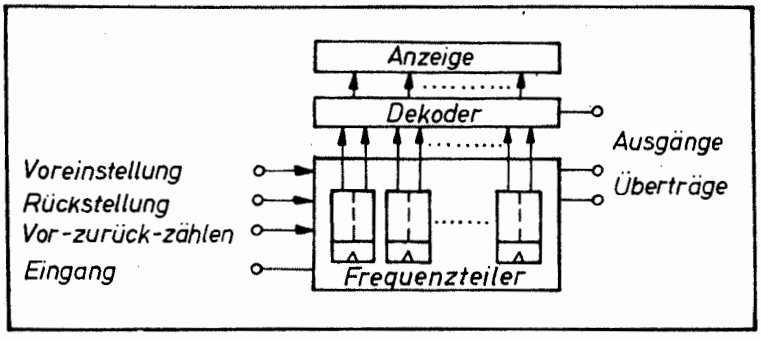


Bild 8.5.8. Grundaufbau eines allgemeinen Zählbausteins. Bei unterschiedlichen Zählbausteinen entfallen einige der hier aufgezeigten Einrichtungen

Die *Rückstellung* dient dazu, den Zähler zu Beginn einer Zählung auf Null einzustellen. Ihr Anschluß führt einfach zu den Rücksetzeingängen der einzelnen Flipflops.

Am *Ausgang* eines Frequenzteilers erscheint genau ein Signal, sobald er die seinem Untersetzungsverhältnis entsprechenden Impulse abgezählt hat. Oft wird nur ein Ausgangsimpuls erwünscht, der bei einer vorher frei einstellbaren kleineren Zahl entsteht.

Mittels der *Voreinstellung* werden hierzu die gewünschten Änderungen im Frequenzteiler oder im Zusammenhang mit dem Dekodiernetzwerk vorgenommen.

In einigen Fällen wird es gewünscht, bestimmte Impulse am Eingang zu addieren und andere zu subtrahieren. Hierzu ist es günstig, die *Zählrichtung* umschalten zu können.

In verschiedenen Fällen sollen Zähler *kaskadiert*, also für größere Zählwerte zusammengeschaltet werden. Dann ist es notwendig, Ausgangsleitungen für den Takt und den Übertrag zur nächsten Stufe zu besitzen. Bei Vor-Rückzählern sind Überträge zur höheren und zur niederen Stufe erforderlich.

Diese Zusatzeinrichtungen wurden bereits teilweise behandelt, z.B. die Anzeigen im Abschnitt 6.1., die Koder und Dekoder im Abschnitt 8.2.8.1. und die Kodearten

im Abschnitt 8.1. Daher sind hier nur spezielle Schaltungsprinzipien der Zähler berücksichtigt worden. Sie können wie folgt *eingeteilt* werden:

- bez. der Basis des verwendeten Zahlensystems, z.B. duale, dezimale oder andere Zähler,
- synchrone oder asynchrone Zähler,
- bez. der möglichen Zählrichtung, z.B. Vor-Rückwärtszähler,
- ob eine Voreinstellung möglich ist,
- bez. der Art der Ableitung aus dem Frequenzteiler gemäß Abschnitt 8.5.3. oder Ring- bzw. JOHNSON-Zähler.

8.5.4.2. Ringzähler

Ein besonders einfacher Zähler ist der Ringzähler. Sein Prinzip zeigt Bild 8.5.9a). Im Beispiel sind vier Flipflops hintereinandergeschaltet, wobei die Q- bzw. $\overline{Q}$-Ausgänge des vorangehenden mit den J- bzw. K-Eingängen des folgenden verbunden sind. Vom letzten führen die Q- bzw. $\overline{Q}$-Ausgänge zu den J- bzw. K-Eingängen des ersten Flipflop. Vor dem Zählbeginn muß der Zähler eingestellt werden. Im Beispiel wird der erste auf Ø und die anderen auf 1 gesetzt. Die Zählung erfolgt gleichzeitig und parallel auf alle Takteingänge, dabei läuft die Ø um je ein FF weiter. Wenn sie beim letzten angelangt ist, wird sie wieder zum ersten geführt. Sie läuft also im Ringzähler um, daher der Name Ringzähler. An den Ausgängen der FF können direkt die gezählten Impulse modulo n (bei n FF) abgeleitet werden. Es können also mit n FF genau bis zu $n - 1$ Impulse gezählt angezeigt werden. Infolge dieses Aufbaus hat der Ringzähler folgende *Vorteile*:

- Er arbeitet bei höchsten Taktfrequenzen, da keine logischen Verkopplungen erforderlich sind.
- Er benötigt keine Dekodierung der Zustände.

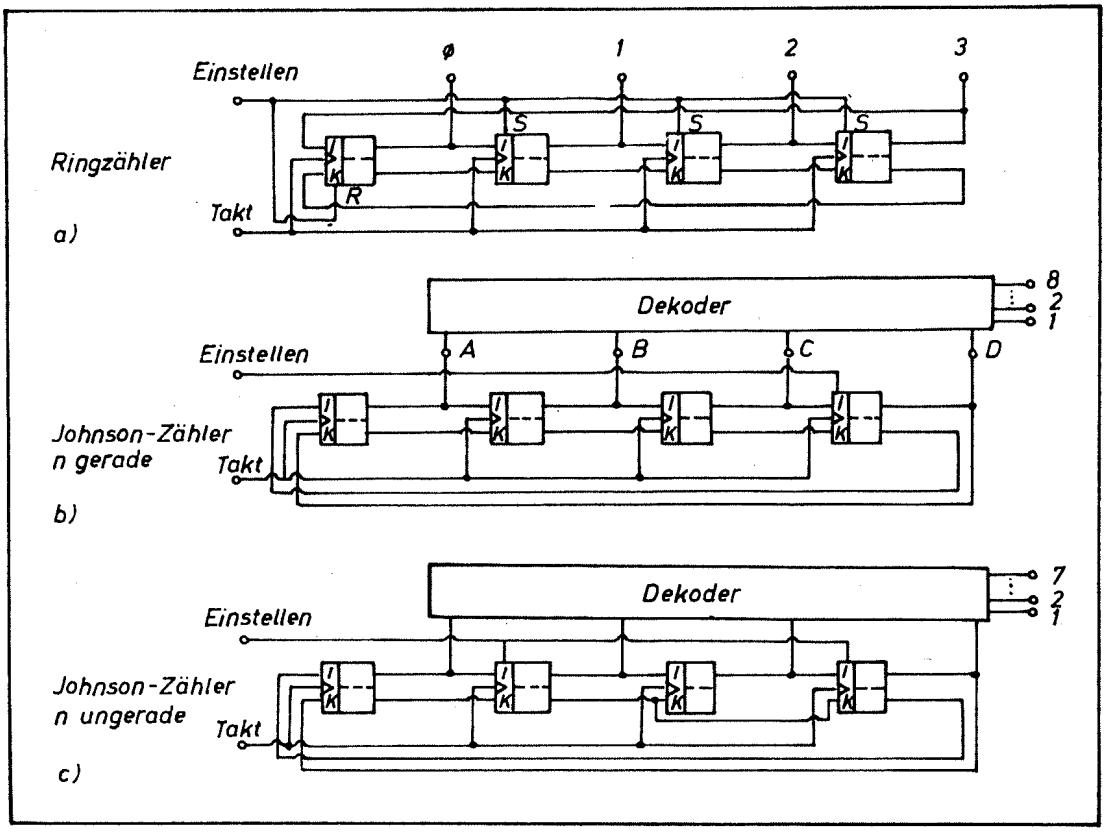


Bild 8.5.9. Prinzipien für Ring- und JOHNSON-Zähler im Beispiel mit je vier Flipflops

- Die Zählkapazität kann auf einfache Weise durch Einfügen weiterer FF vergrößert werden.
- Der Zählzustand ist leicht in negierter Form von den Ausgängen $\overline{Q}$ ableitbar.

Die *Nachteile* des Ringzählers sind:

- Es muß vor dem Start der notwendige Startzustand erzeugt werden.
- Es werden viele FF benötigt, deshalb wird der Ringzähler bis nur etwa $n = 32$ realisiert.

Änderungen des Ringzählers gegenüber dem hier dargestellten Fall gibt es in folgenden Punkten:

- Statt einer umlaufenden Ø kann bei anderem Start auch eine umlaufende 1 in Anwendung kommen.
- Es können auch D-FF angewendet werden, dann werden sie unmittelbar gekoppelt.
- Durch eine zusätzliche Logik ist es möglich, den Startzustand indirekt mittels einer bestimmten Impulszahl zu erzeugen.

8.5.4.3. JOHNSON-Zähler

Der Grundtyp des JOHNSON-Zählers (Bild 8.5.9b)) weicht *gegenüber dem Ringzähler in drei Punkten* ab:

- Von dem letzten zum ersten FF erfolgt die Rückführung gekreuzt.
- Der Startzustand wird dadurch erreicht, daß alle FF gesetzt (im anderen Fall rückgesetzt) werden.
- Er benötigt eine Dekodierschaltung.

Dadurch erfolgt die Zählung wie folgt: Beim Start liegen alle Ausgänge — im Bild A bis D — auf 1. Beim ersten Takt wird das erste FF durch Übertrag vom letzten auf Ø gesetzt. Bei jedem weiteren Takt werden dann die folgenden FF ebenfalls auf Ø gesetzt. Dies geht so lange, bis alle n (bei n Stufen) auf Ø liegen. Danach wird vom Ausgang her das erste FF auf 1 gesetzt, und nach $2n$ sind alle auf 1.

Der JOHNSON-Zähler zählt also $2n$ Impulse, d. h. doppelt so viele Impulse wie der Ringzähler. Dieser Vorteil wird dadurch erkauft, daß er einen zusätzlichen Dekoder benötigt.

Durch den Faktor 2 können zunächst nur eine gerade Anzahl von Impulsen gezählt werden. Sollen es ungeradzahlig viele sein, so ist eine kleine Änderung notwendig, die in Bild 8.5.9c) gezeigt ist. Der K-Eingang des letzten Flipflop wird nicht mit $\overline{Q}$ des vorangegangenen verbunden, sondern mit $\overline{Q}$ des zwei Stufen zuvor liegenden.

Der JOHNSON-Zähler wird zuweilen wegen seiner Analogie zum MÖBIUS-Band auch MÖBIUS-Zähler genannt.

Die Dekodierung der Zustände des JOHNSON-Zählers ist recht übersichtlich, weil der zugehörige Kode einschrittig ist.

Der JOHNSON-Zähler ist besonders effektiv als Dezimalzähler. Mit 5 FF und einem relativ einfachen Dekoder ist dann direkt eine Anzeige gemäß 1 aus 10 möglich. Ein integriertes Beispiel dieses Typs ist der CMOS-Schaltkreis 4017.

8.5.4.4. Dualzähler

Gemäß den Aussagen im Abschnitt 8.5.4.1. lassen sich aus allen Frequenzteilern Zähler aufbauen, indem die Zustände der einzelnen FF herausgeführt und dekodiert werden. So entstehen einmal die asynchronen Zähler aus den Schaltungen von

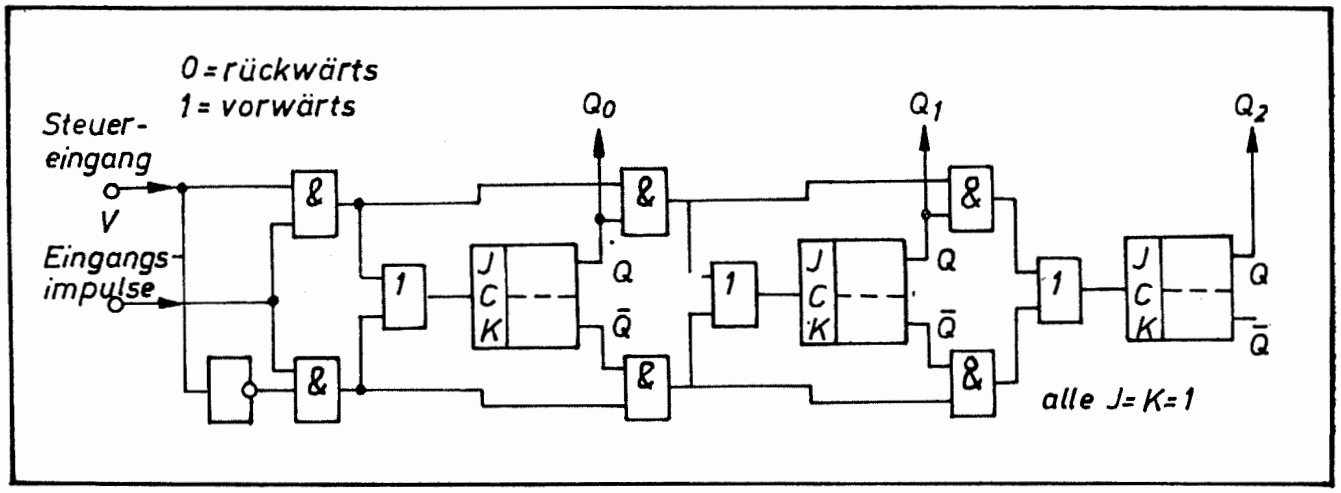


Bild 8.5.10. Prinzip des asynchronen Vor-Rückwärtszählers

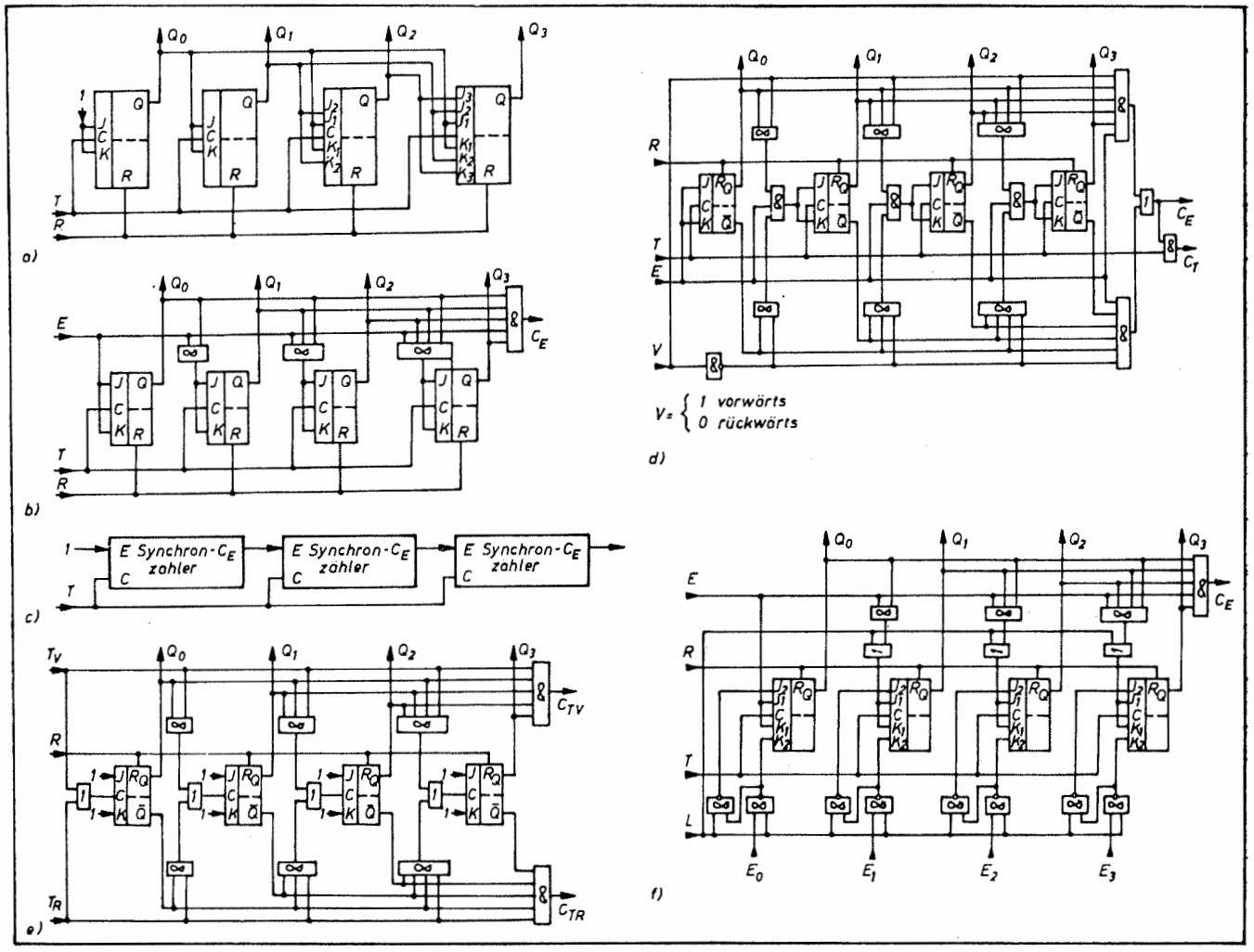


Bild 8.5.11. Synchrone Dualzähler

a) Grundprinzip mit 4 FF für 16 Zählungen. Von den Ausgängen Q_0 bis Q_3 wird die Dekodierschaltung für die Anzeige gesteuert. Mit R wird der Zähler auf Null zurückgesetzt. E ist der Zähleingang.

b) Grundprinzip wie in a), jedoch mit Verkopplung durch kombinatorische Schaltungen und Bildung eines Übertrags C_E zur Kaskadierung mehrerer solcher Zähler

c) Prinzip der Kaskadierung von Zählern gemäß b)

d) Vor- und Rückzähler, abgeleitet aus b) (zur Funktion vgl. Bild 8.5.10). Bei der Kaskadierung muß hier sowohl der Taktausgang C_T als auch der Übertragsausgang C_E verwendet werden

e) Vor- und Rückzähler mit getrenntem Takteingang für beide Zählrichtungen sowie den zugehörigen Ausgängen zur Kaskadierung

f) Prinzip eines voreinstellbaren Zählers

Bild 8.5.6 und die synchronen aus den-Schaltungen von Bild 8.5.7. Besondere Bedeutung haben hierbei die Dualzähler, welche sich insbesondere aus der Schaltung von Bild 8.5.6b) ableiten. Sie heißen dann auch Ripple-counter, da sich in ihnen der Eingangsimpuls wie eine Welle (ripple) fortpflanzt. Diese Zähler lassen sich relativ leicht in der Zählrichtung umschalten.Hierzu genügt es, das nächste FF vom Ausgang Q bzw. $\overline{Q}$ anzusteuern. Eine entsprechende Prinzipschaltung zeigt Bild 8.5.10. Die Umschaltung erfolgt durch den Pegel Ø bzw. 1 am Eingang V. Der synchrone Dualzähler wird vor allem für 4 FF, also 16 Zählungen gemäß Bild 8.5.11a), angewendet. Er stellt die Erweiterung von 2 FF für vier und von 3 FF für acht Zählungen gemäß Bild 8.5.7 dar. Infolge der vielen Verknüpfungen an den Eingängen der weiter zum Ende hin liegenden FF ist es kaum möglich, dieses Prinzip über die 4 FF hinaus zu erweitern. Hierzu ist es sinnvoll, eine Übertragungslogik gemäß Bild 8.5.11b) zu verwenden, die zusätzliche Gatter benutzt und zugleich ein Ausgangssignal für die Kaskadierung mehrerer solcher Stufen gemäß Bild 8.5.11c) benutzt. Durch Erweiterung dieser Zusatzschaltungen besteht auch die Möglichkeit, den Zähler vorwärts und rückwärts zählen zu lassen. Die entsprechende Schaltung zeigt Bild 8.5.11d). Die Umschaltung der Zählrichtung erfolgt am Eingang V. Das Prinzip ist aus dem Zusammenhang mit dem asynchronen Vor-Rückwärtszähler relativ leicht zu verstehen. Durch eine geringe Änderung ist es auch möglich, hieraus gemäß Bild 8.5.11e) einen Zähler abzuleiten, der getrennte Zähleingänge für das Vorwärts- und Rückwärtszählen besitzt. Bei den beiden Varianten der Teilbilder d) und e) ist dann auch die Kaskadierung unterschiedlich zu gestalten.

Bei dem Vor- und Rückwärtszähler gemäß Bild 8.5.11e) kann es zu Koinzidenzen bezüglich der beiden Zähleingänge kommen. Sie treten dann auf, wenn an den Eingängen Impulse in so geringem Abstand auftreten können, daß der Zähler ihnen nicht einzeln folgen kann. Für diesen Fall kann ein solcher Fehler mit zusätzlichen Antikoinzidenz-Schaltungen vermieden werden.

Eine Erweiterung des Dualzählers mit Eingängen E_0 bis E_3 zur Voreinstellung auf einen bestimmten Wert zeigt Bild 8.5.11f). In diesem Fall kann dem Übertragsausgang C_E — eventuell auch nach der Kaskadierung — noch eine andere Bedeutung zugeordnet werden. Er ändert genau dann sein Potential, wenn die Differenz Null zwischen voreingestellter Zahl und maximalem Zählwert durch die Zählimpulse erreicht wird.

Es besteht also die Möglichkeit, voreingestellte (voreinstellbare) Impulszahlen zu zählen und bei ihrem Erreichen einen Impuls abzugeben. Diese sind für verschiedene Automatisierungsvorgänge von großer Bedeutung. Es liegt damit also ein programmierbarer Zähler vor.

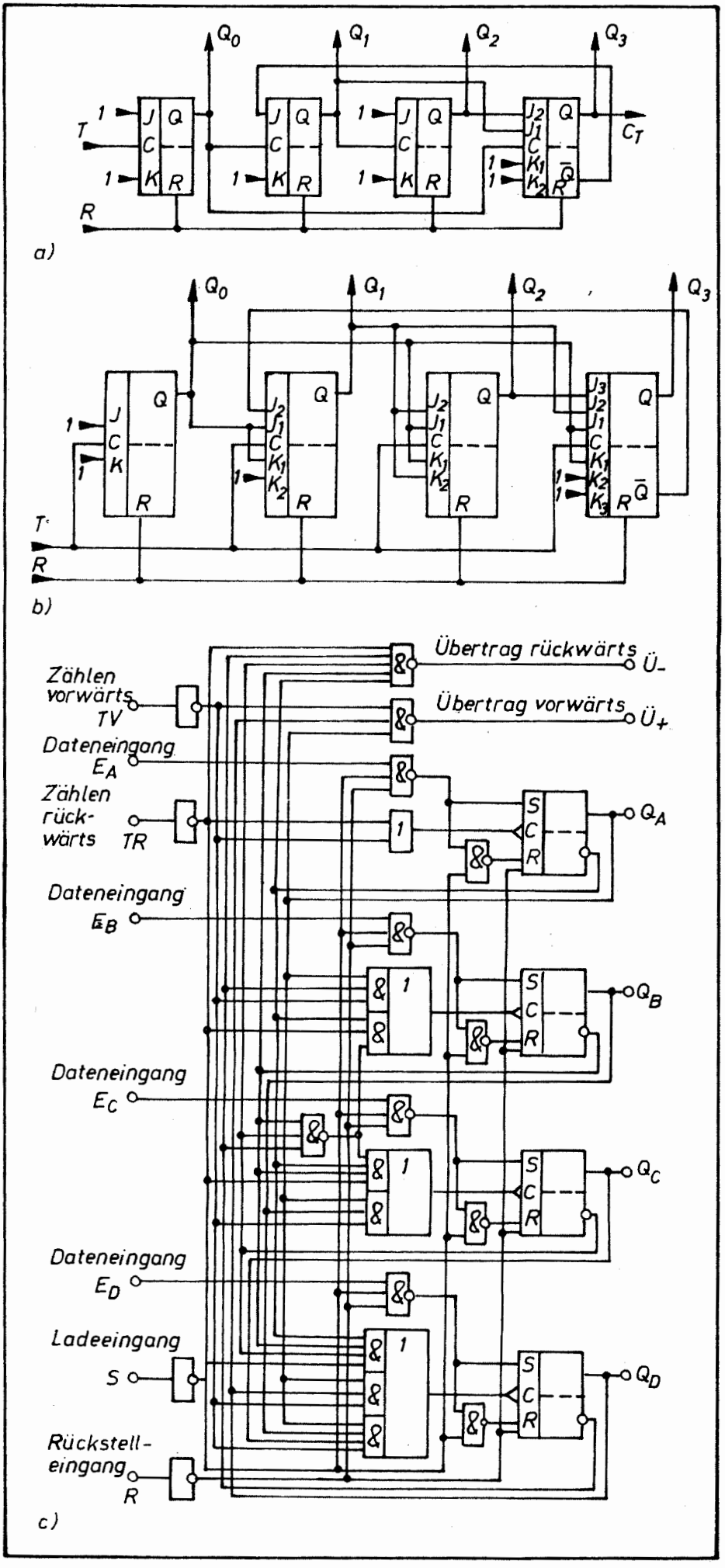
Ein anderes Prinzip der voreinstellbaren Zähler besteht darin, daß die Voreinstellung kodiert direkt mit dem Ausgang Q_0 bis Q_n auf Gleichheit verglichen wird und bei Gleichheit ein Impuls ausgelöst wird (vgl. 8.2.8.4.).

8.5.4.5. BCD-Zähler

Beim Dezimalzähler müssen Impulse so unterdrückt werden, daß er bereits nach 10 statt nach 16 Impulsen den Anfangszustand einnimmt. Das Grundprinzip hierzu ist

Bild 8.5.12. Prinzipien von BCD-Zählern

a) Asynchroner Typ
b) Synchroner Typ
c) Innenschaltung des TTL-Zählers D 192 (7492)


a)
b)
c)
Übertrag rückwärts
Übertrag vorwärts
Zählen vorwärts
TV
Dateneingang
Zählen rückwärts
TR
Ladeeingang
S
Rückstell-eingang
R

bereits bei den 10:1-Frequenzteilern für asynchronen Betrieb im Zusammenhang mit Bild 8.5.6 und für den synchronen Betrieb mit Bild 8.5.7 behandelt. Daraus lassen sich nach dem vorangegangenen Abschnitt die entsprechenden Schaltungen mittels zusätzlicher kombinatorischer Schaltungen ableiten (Bild 8.5.12a) und b)).

Um den erheblichen Aufwand eines solchen Zählers aufzuzeigen, sei die Innenschaltung des BCD-Zählers der TTL-Reihe vom Typ D 192 (7492) in Bild 8.5.12c) vorgestellt. Er enthält die beiden Zähleingänge T_V und T_R für Vor- und Rückwärtszählung, den Rückstelleingang und Dateneingänge E_A bis E_D zur Voreinstellung.

8.5.5. Register

8.5.5.1. Überblick

Der *Begriff Register* ist heute vielfältig im Gebrauch. Allgemein kann *darunter verstanden* werden:

- Es liegt eine relativ kleine Anzahl von Speichern vor.
- Sie sind in definierter Weise über kombinatorische Schaltungen verbunden.
- Sie sind in der Regel sehr schnell steuerbar; kurze Zugriffszeiten.

Es können in diesem Sinne *drei Gruppen von Registern* nach ihrem Anwendungszweck unterschieden werden (vgl. Bild 8.5.13):

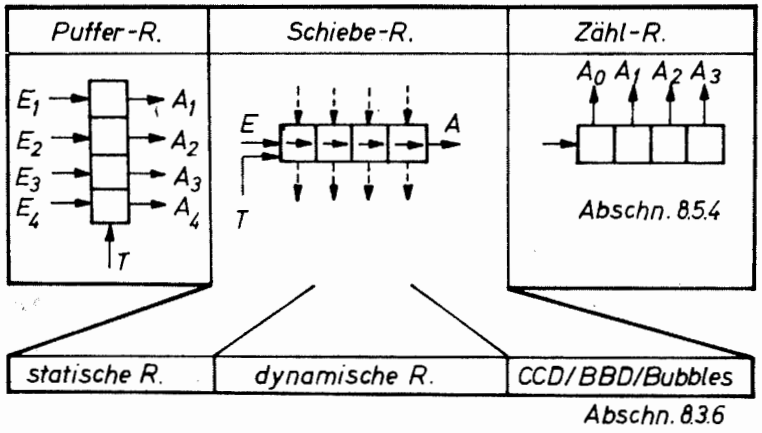


Bild 8.5.13. Einteilung der Register mit Hinweis auf schon zuvor behandelte Arten

Ein *Pufferregister* liegt dann vor, wenn n Speicherzellen — in der Regel D-FF — parallel organisiert sind und mit einem gemeinsamen Takteingang ausgestattet sind. Sie übernehmen dann parallel zum Zeitpunkt der Taktwirksamkeit die Information auf den Eingangsleitungen E_1 bis E_n. Die Information steht danach — eventuell unter Zwischenschaltung weiterer kombinatorischer Schaltungen — an den Ausgängen A_1 bis A_n zur Verfügung. Weitere Namen für solche Schaltungen sind Buffer, Latch und Auffangregister. Ein spezieller Baustein wird beispielhaft im Abschnitt 8.5.5.2. behandelt.

Schieberegister sind nicht parallel, sondern seriell organisiert. Sie übernehmen zeitlich nacheinander zu den Taktzeiten Informationen des Eingangs E und leiten sie dabei schrittweise zu den nachgeschalteten FF weiter. Am Ausgang A erscheint dann dieselbe Information um n Taktzeiten verzögert, ebenfalls sequentiell. Schieberegister existieren in vielfältigen Varianten. Zum Teil ermöglichen sie zusätzliche parallele Ein- und/oder Ausgabe (gestrichelt in Bild 8.5.12). Sie existieren weiter als statische und dynamische Register sowie in den schon im Abschnitt 8.3.6. behandelten Formen

der CCD-, BBD- und Bubblespeicher. Sie werden deshalb im folgenden nicht mehr behandelt.

Schließlich existieren noch die *Zählregister*. Dies ist nur eine andere Bezeichnung der bereits im Abschnitt 8.5.4. behandelten Zählschaltungen.

8.5.5.2. Pufferregister

Als Beispiel für ein Pufferregister (auch Parallelregister genannt) wird der LS-TTL-Baustein 8212 bzw. MH 3212 betrachtet (Bild 8.5.14). Er enthält 8 D-FF, die von den

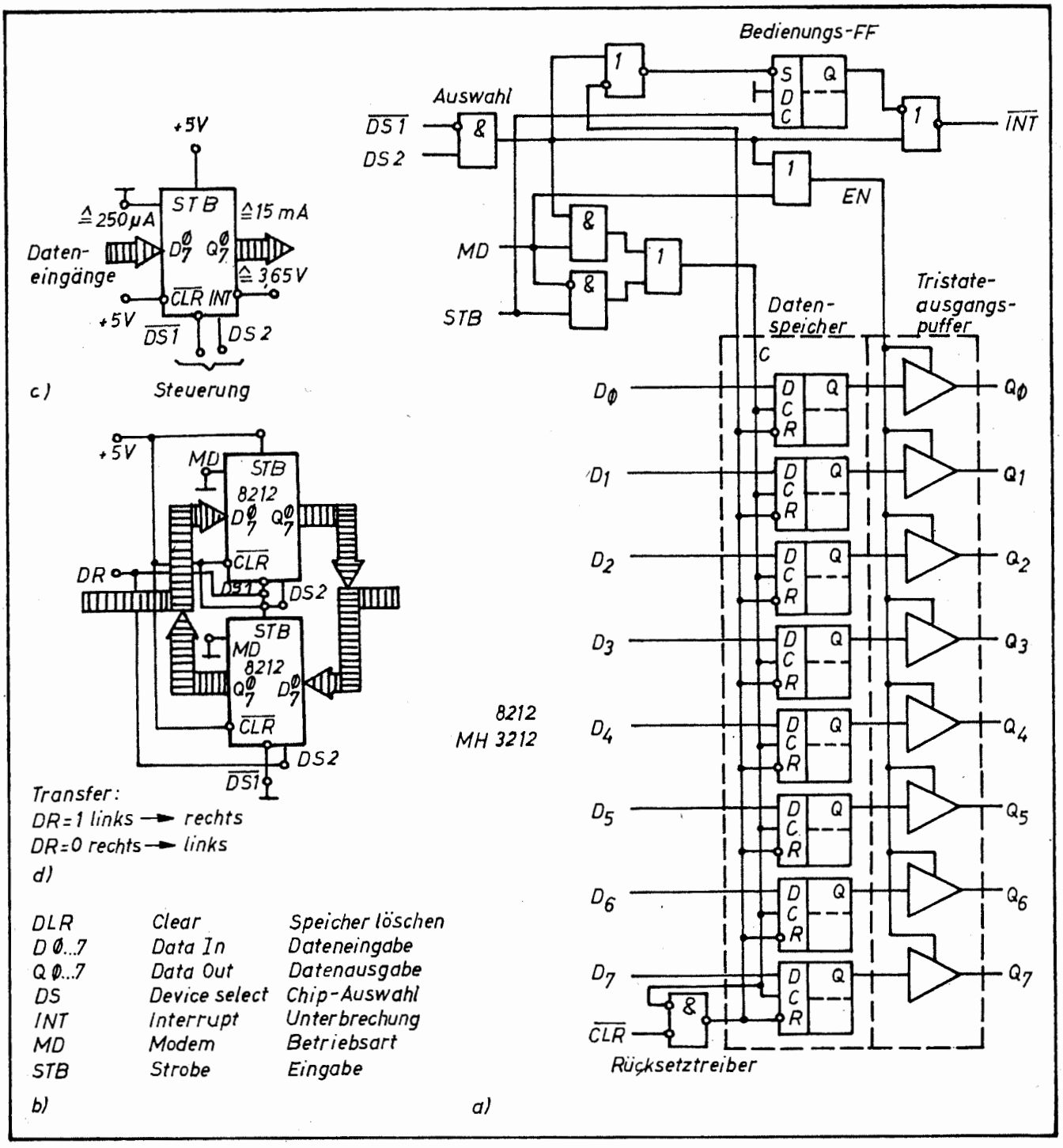


Bild 8.5.14. Unidirektionales Pufferregister 8212; MH 3212

a) Innenschaltung
b) Bedeutung der Kurzzeichen
c) Schaltung für Daten-Ein-Ausgabe (Durchschaltung)
d) Prinzip für bidirektionalen Betrieb

Dateneingängen D_0 bis D_7 gesteuert werden und dann, wenn das gemeinsame Taktsignal C auf H ist, den Eingängen folgen. Beim Übergang des Taktsignals auf L werden die entsprechenden Werte in den FF gespeichert. Am gemeinsamen Reseteingang R können alle FF zurückgesetzt werden. Die Q-Ausgänge der FF steuern je eine nichtinvertierende Tristate-Ausgangsstufe, welche maximal mit 15 mA belastet werden kann und welche mit dem gemeinsamen Signal EN auf L in den hochohmigen Tristate-Zustand versetzt werden können.

Zur Steuerung der Schaltung sind neben dem eigentlichen Pufferregister und der Tristate-Stufe noch einige kombinatorische Schaltungen sowie ein Bedienungs-Anforderungs-FF enthalten. Dadurch besitzt der Schaltkreis etliche zusätzliche Ein- und Ausgänge, die seiner universellen Anwendbarkeit in Mikrorechnerschaltungen dienen. Hier genügt im wesentlichen die Kurzerklärung im Bild.

Die *Ausgabe* des Speicherinhalts kann wahlweise über $\overline{DS1} \wedge DS2$ oder über MD gleich H realisiert werden.

Die *Eingabe* kann wahlweise über $STB \wedge \overline{MD}$ oder über $MD \wedge \overline{DS1} \wedge DS2$ erfolgen. Das Bedienungs-Anforderungs-Flipflop wird von CLR, $\overline{DS1}$, $DS2$ und STB beeinflußt, und bei $STR \wedge \overline{DS1} \wedge DS2$ auf L ist der Ausgang $\overline{INT}$ auf H.

8.5.5.3. Betriebsarten statischer Schieberegister

Bei den Schieberegistern steht die serielle Ein- und Ausgabe von Informationen vom Namen her im Vordergrund. Bei den statischen Schieberegistern, die mit FF aufgebaut sind, bestehen aber weitere Möglichkeiten der Informationsein- und -ausgabe.

Die wichtigsten Möglichkeiten sind in Bild 8.5.15 zusammengestellt. Die obere Reihe (a) bis d)) betrachtet den Informationsfluß bez. der seriellen bzw. parallelen Ein- bzw. Ausgabe. Die Variante SISO gemäß Teilbild a) zeigt die typische Eigenschaft der Schieberegister, von der auch ihr Name herrührt. Die beiden Varianten b) und c) sind Wandler für Informationsströme von seriell nach parallel bzw. umgekehrt. Die vierte Variante entspricht dem Pufferregister vom vorigen Abschnitt. Viele Schieberegister ermöglichen mehrere, einige sogar alle Betriebsarten. Daraus ergibt sich die Vielfalt ihrer Einsatzmöglichkeiten.

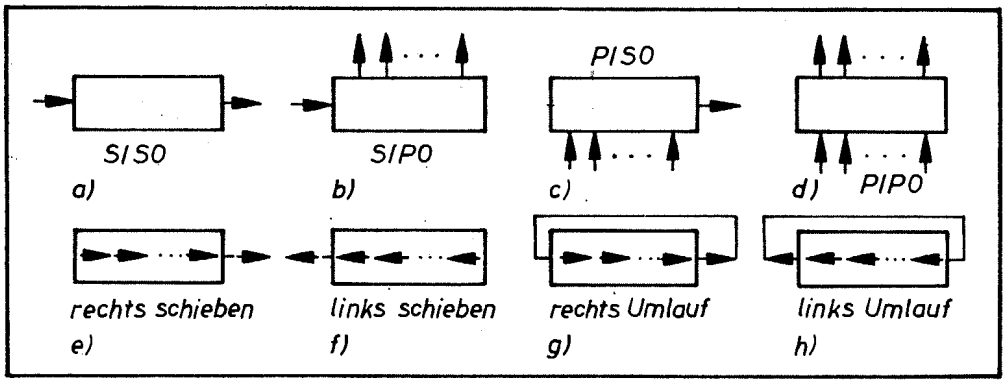


Bild 8.5.15. Betriebsarten von statischen Schieberegistern

a) SISO = serial in, serial out
b) SIPO = serial in, parallel out
c) PISO = parallel in, serial out
d) Wie Pufferregister gemäß Abschnitt 8.5.5.2.
e) Rechts schieben: links neues Bit, rechts geht ein Bit verloren
f) Links schieben: rechts neues Bit, links geht ein Bit verloren
g) Umlaufregister rechts herum,
h) links herum

Die zweite Zeile von Bild 8.5.15 (e) bis h)) bezieht sich auf die Konsequenzen und Richtungen der Schiebefunktionen. Bezüglich der Schieberichtung können rechts = vorwärts (e) und g)) und links = rückwärts (f) und h)) unterschieden werden. Diese Funktionen sind deutlich anders als die im Abschnitt 8.5.4.5. behandelte Vorwärts- bzw. Rückwärtszählung. Dort interessierte der kodierte Zählwert. Hier die Anordnung der Bits in ihrer Abfolge.

Bezüglich der Schiebefunktion ist die reine Schiebefunktion (e) und f)) und der Umlauf (g) und h)) zu unterscheiden. Bei der reinen Schiebefunktion gelangt ein neues Bit an einem Ende in das Register und das am entgegengesetzten Ende stehende geht verloren. Beim Umlauf werden alle Bits um ein FF nach rechts bzw. links verschoben, und das ganz rechts (links) stehende Bit gelangt nach ganz links (rechts). Es kommt also keine zusätzliche Information hinzu, sondern die im Register vorhandene rotiert im Kreis, ähnlich wie dies bereits bei den Ringzählern für ein einzelnes Bit beschrieben wurde.

8.5.5.4. Beispiele statischer Schieberegister

Ein *typisches Schieberegister* ist der TTL-Schaltkreis D 195. Seine Innenschaltung zeigt Bild 8.5.16a). Es *enthält*

- vier getaktete MS-RS-FF
- vier parallele Eingänge A bis D
- vier parallele Ausgänge Q_A bis Q_D
- einen seriellen Eingang S_E
- je einen Takteingang für parallele Eingabe T_2 und serielle Eingabe T_1
- einen Eingang MC (mode control) für die Betriebsart (MC = H entspricht parallel MC = L entspricht seriell).

Die MS-RS-FF übernehmen bei positivem Taktteil (T_1 oder T_2) die Information in den Master und geben bei negativem Taktteil zum Slave. Von da ab stehen sie bis zum Beginn des nächsten negativen Taktteils an den Ausgängen zur Verfügung. In diesem Sinne kann am Ausgang Q_D die Information auch seriell abgenommen werden. Damit erlaubt dieses Schieberegister die Betriebsarten a) bis e) des Bildes 8.5.15 in der Variante rechts schieben (e). Wird Q_D mit S_E verbunden, ist auch die Variante Umlauf rechts (g) möglich. Eine Umschaltmöglichkeit zwischen Schieben und Umlauf zeigt als Teilschaltung, die zusätzlich notwendig ist, Bild 8.5.16b). Die Umschaltung erfolgt am Steuereingang U. Je nach seinem Zustand erfolgt Dateneingabe (seriell) oder Rückführung (Umlauf). Nicht möglich sind beim D 175 dagegen die Varianten links schieben und Umlauf. Um dies zu erreichen, sind prinzipielle Änderungen bezüglich der kombinatorischen Schaltungen erforderlich. Dies zeigt als Teilschaltung Bild 8.5.16c) für drei MS-D-FF. Je nach der Einschaltung von Li:H bzw. Re:H erfolgt die Links- bzw. Rechtsverschiebung. Entsprechend sind auch ein linker und ein rechter Eingang, E_{Li} und E_{Re}, vorhanden.

8.5.5.5. Realisierungstechniken

Bisher wurden nur TTL-Register berücksichtigt. Register existieren aber in allen Schaltkreisfamilien, z.B. sehr schnelle Schaltungen in ECL-Technik. Wie bei den Speichern läßt sich bei den MOS-Registern wiederum eine erhebliche Verringerung des Aufwandes dadurch erreichen, daß die Information nicht als Zustand einer Schaltung, sondern in Kapazitäten gespeichert wird. So entstehen die quasistatischen und dynamischen Register mit ihren Vor- und Nachteilen, wobei ein Nachteil in der ständigen

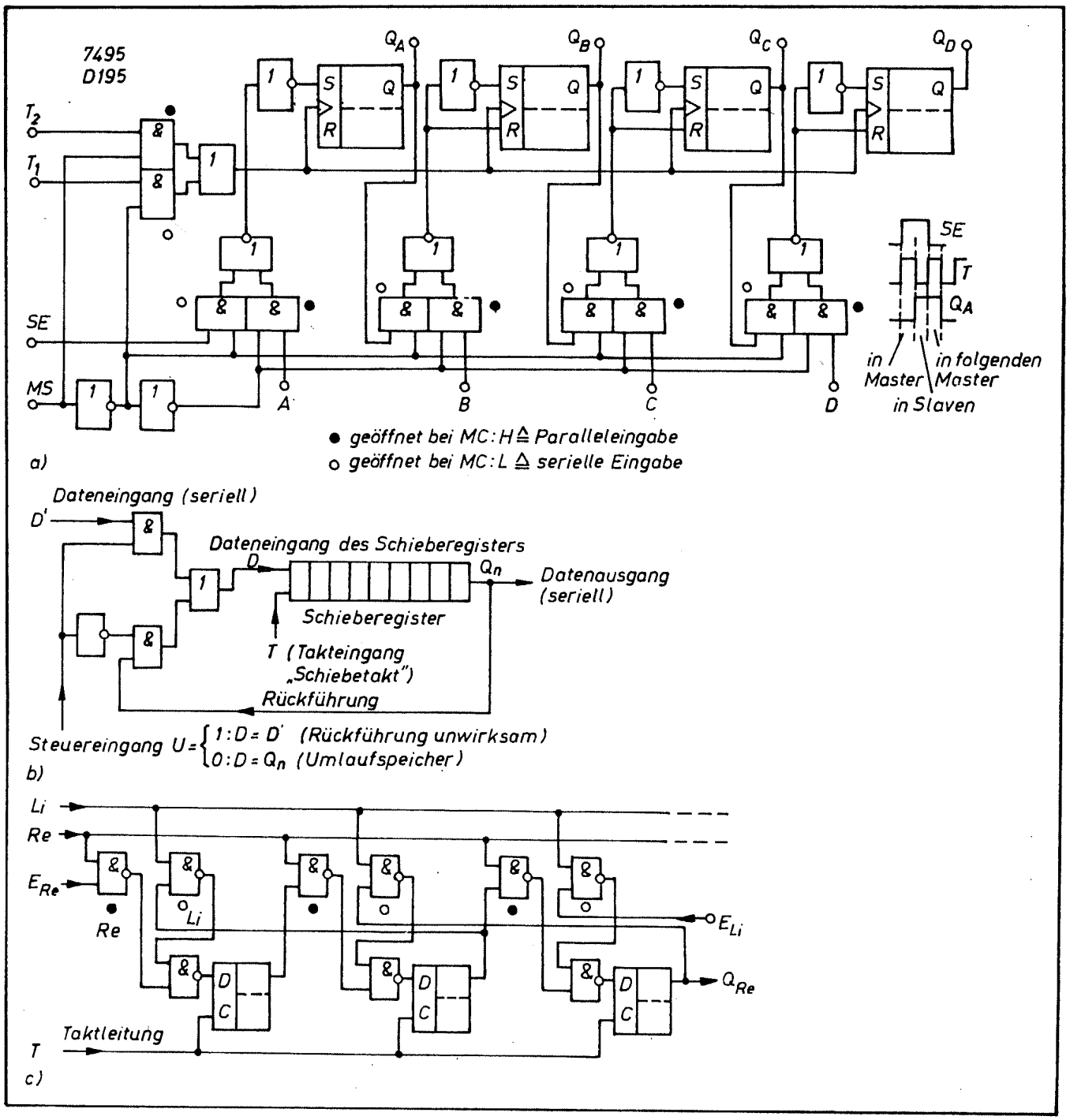


Bild 8.5.16. Beispiele statischer Schieberegister

a) Innenschaltung des D 195, 7495
b) Zusatzschaltung zur Umschaltung zwischen Schiebe- und Umlaufregister über den Steuereingang *U*
c) Teilschaltung für die Möglichkeit der Rechts- und Linksschiebung

Taktung (Refresh) mit Frequenzen im Bereich von etwa 1 kHz bis 10 MHz besteht. Aus den dynamischen Registern ist dann aber auch leicht der Übergang zu den CCD-Registern verständlich. Neben dem Vorteil der wenigen Bauelemente bei den dynamischen Registern sei noch der Vorteil der geringen Leistung von ca. 5 bis 20 μW/bit genannt.

8.5.5.6. Quasistatische Register

Für ein Master-Slave-Flipflop sind viele Bauelemente notwendig. Es benötigt also relativ viel Chipfläche. Für den Aufbau umfangreicher Schieberegister ist es also wenig

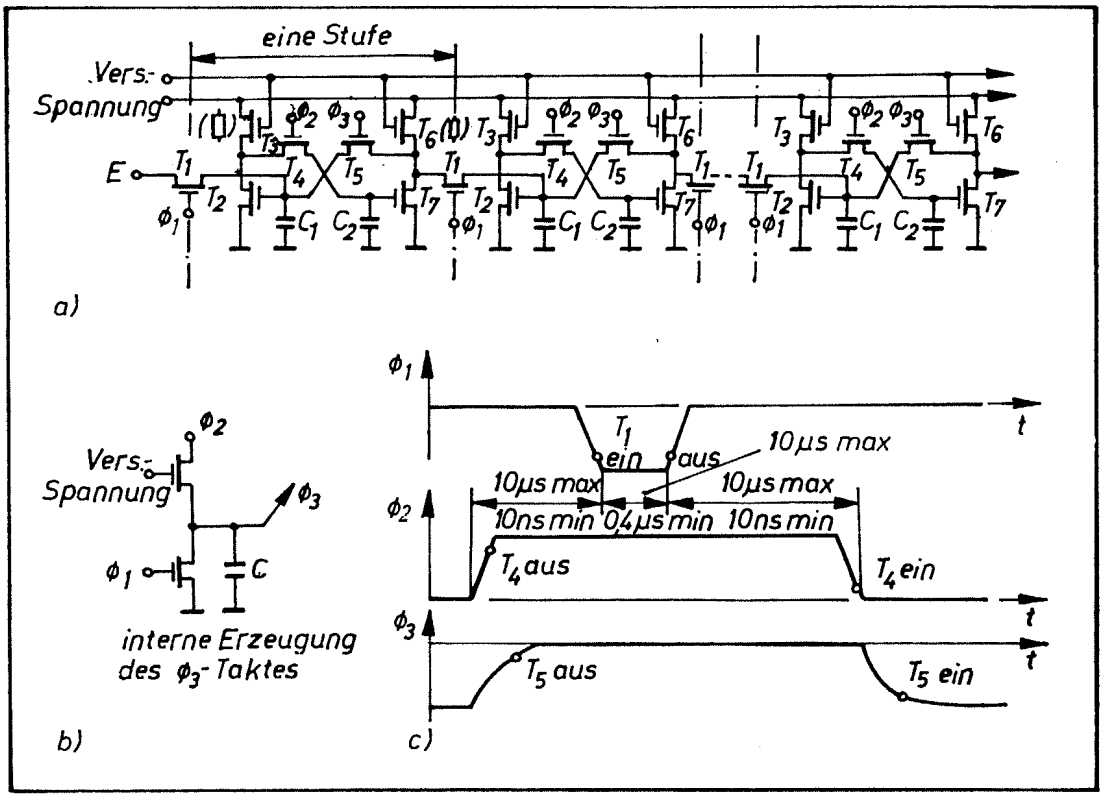


Bild 8.5.17. Vereinfachtes quasistatisches Schieberegister

a) Schaltung

b) Interne Erzeugung des Φ_3-Taktes

c) Zeitlicher Verlauf der Taktsignale

geeignet. Es sind dann Vereinfachungen notwendig. Ein Beispiel zeigt Bild 8.5.17. Es enthält nur sieben Transistoren und zwei Gatekapazitäten je Zelle. Im Vergleich zum JK-Master-Slave-Flipflop in TTL-Technik ist also eine drastische Reduzierung erfolgt. Es benötigt 40 Transistorfunktionen und 20 Widerstände. Infolge dieser Reduzierung sind drei Taktsignale notwendig, wobei der dritte Takt aus den anderen beiden intern mittels Schaltung b) erzeugt wird.

Die beiden Transistoren T_3 und T_6 sind nur als Widerstände wirksam. Die Transistoren T_4 und T_5 sind Schalter, die im Ruhebetrieb geschlossen sind und das Flipflop aus T_2 und T_7 rückkoppeln. T_1 ist schließlich der Übergabeschalter zur folgenden Stufe (Transmissionsgate).

Die Funktion der Schaltung wird am besten mittels der Taktverläufe aus Teilbild c) verständlich. Zunächst wird T_4 und damit die Rückkopplung der Flipflops geöffnet. Infolge der Ladung auf den Gatekapazitäten behalten sie ihren Zustand bei. Hieran ändert sich auch nichts, wenn etwas später T_5 öffnet. Dann schließt T_1, und es kann über ihn der an T_7 vorhandene Zustand zur nächsten Stufe nach T_2 weitergeleitet werden. Ebenso wird vom Eingang der neue Zustand auf C_1 geleitet. Jetzt muß zum Umladen von C_2, die den neu eingestellten Zustand im rückgekoppelten Flipflop festhält, zunächst T_4 schließen und dann T_5.

8.5.5.7. Dynamische Register

Das Schieberegister von Bild 8.5.17 funktioniert bei periodischem Takt auch noch, wenn T_5 und damit auch Φ_3 fehlt. Durch ihn wird ja nur der eingespeicherte Zustand stabilisiert. Ohne ihn liegt also ein dynamisches Zweiphasenschieberegister in Ratiotechnik vor. Seine Schaltung zeigt Bild 8.5.18a). Von Ratiotechnik wird immer dann

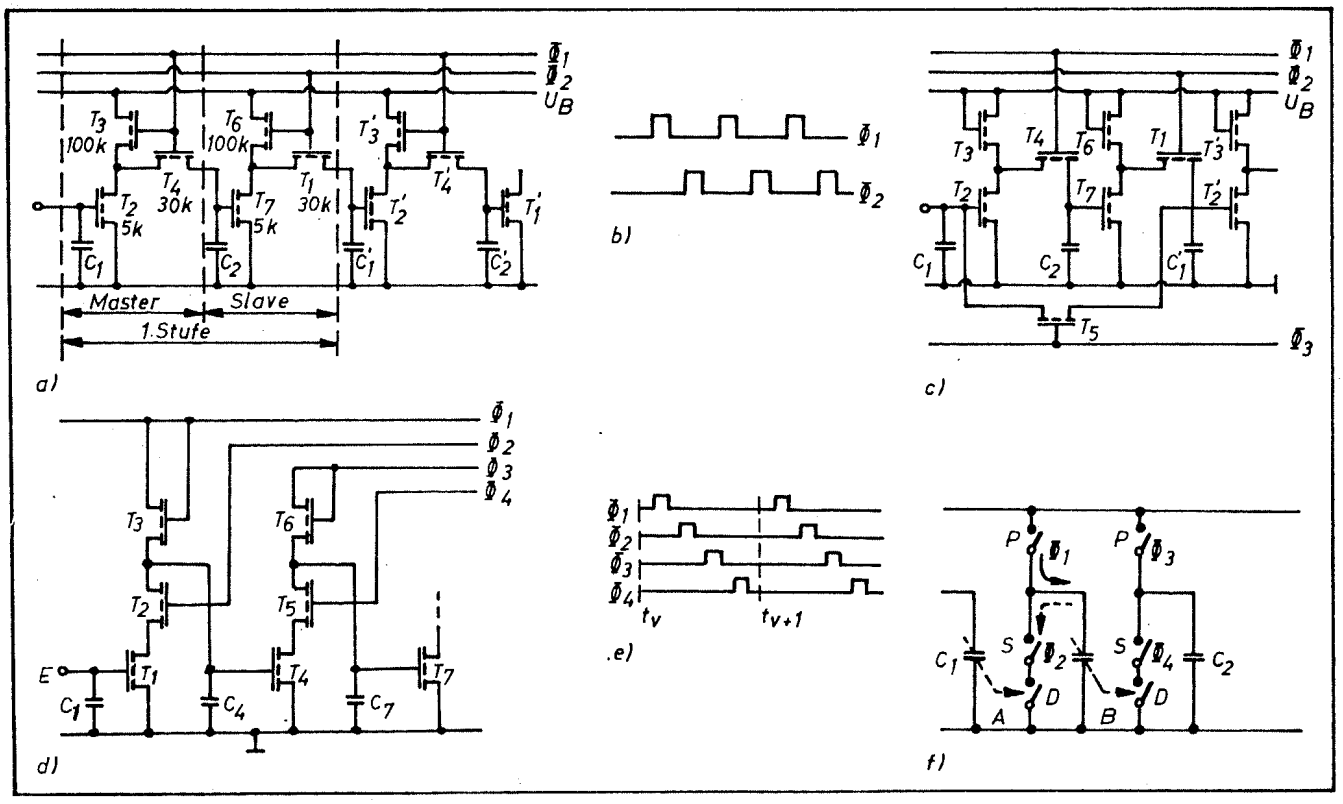


Bild 8.5.18. Beispiele für dynamische Schieberegister

a) Schaltung für Zweiphasenbetrieb
b) Nichtüberlappende Takte für die Schaltung a)
c) Mit Rückkopplungstransistor T_5 entsteht statisches Schieberegister. Dazu wird Takt Φ_2 abgeschaltet, und Φ_3 erhält ständig Spannung
d) Schaltung eines Vierphasenregisters
e) Taktverläufe für d)
f) Funktionsprinzip von d)

gesprochen, wenn ein Inverter, wie z.B. T_2, T_3 oder T_7, T_6 in ihrem Widerstandsverhältnis gesteuert wird.

Zur Funktion der Schaltung dürften sich die beiden Takte nicht überlappen (siehe Teilbild b)). Während der Phase Φ_1 sind die beiden Transistoren T_3 und T_4 leitend. T_2 und T_3 wirken als Inverter und der Kondensator C_2 wird auf das entgegengesetzte Potential von C_1 geladen. Danach werden T_3 und T_4 wieder gesperrt. Etwas später folgt der Takt Φ_2 und macht T_6 und T_1 leitend, dadurch wird C_1' so, wie C_1 ursprünglich war, aufgeladen. Das Eingangssignal ist also um eine Stufe weitergeschoben. Wegen des Weiterschiebens in zwei Etappen werden die beiden Inverter mit dem folgenden Schalttransistor zuweilen auch als Master und Slave bezeichnet. Entsprechend der Gatekapazität C von etwa 0,1 pF und dem Sperrstrom des Schalttransistors von etwa 1 nA beträgt die tiefste Taktfrequenz etwa 1 kHz.

Das dynamische Schieberegister läßt sich wieder durch Einfügen von T_5 in ein zeitweilig statisches verwandeln. So entsteht die Schaltung von Bild 8.5.18c). Die Gates von T_3 und T_6 müssen dazu allerdings an die Betriebsspannung gelegt werden. Dies bedeutet höhere Verluste, da jeder Inverter, der leitend ist, pausenlos Strom verbraucht. Für den Haltebetrieb wird Φ_2 ab- und Φ_3 dauernd eingeschaltet. Während der Impulse von Φ_1 wird das Flipflop T_2, T_7 aufgefrischt.

Ein Nachteil vom Zweiphasenschieberegister besteht darin, daß ständig während der Taktzeiten Φ_1 bzw. Φ_2 durch die Inverter T_2, T_3 bzw. T_7, T_6 Strom fließt, obwohl eigentlich nur die jeweils folgende Kapazität umzuladen ist. Dies ist der Nachteil aller Ratioschaltungen. Er kann durch einen vollständigen Schalterbetrieb, also ratiofreie

Schaltungen vermieden werden. Dazu ist allerdings in der Regel ein Vierphasenbetrieb notwendig. Eine entsprechende Schaltung zeigt Bild 8.5.18d). Ihre Funktion wird unmittelbar aus dem Taktverlauf e) und dem Schalterersatzbild f) verständlich. Die Umladeenergie beträgt je Kondensator dabei nur etwa 50 pJ. Der Leistungsverbrauch einer solchen Schaltung nimmt also proportional mit der Frequenz zu und erreicht bei 100 kHz ca. 5 μW.

Im Prinzip gibt es noch eine beträchtliche Zahl von Varianten der dynamischen Flipflops. Die beiden hier beschriebenen sind jedoch die gebräuchlichsten.

Ergänzend sei erwähnt, daß mittels dynamischer Register auch eine dynamische Logik zu realisieren ist. Auf ihre Beschreibung sei hier verzichtet.

8.5.5.8. Erzeugung pseudozufälliger Folgen

Für verschiedene Meßtechniken und Rechneroperationen ist es notwendig, zufällige binäre Signale zu verwenden. Sie lassen sich z.B. mittels Begrenzung aus Rauschsignalen ableiten. Dies ist jedoch meist umständlich und z. T. unzuverlässig. So entstanden Methoden, pseudozufällige Folgen zu erzeugen. Eine solche Folge zeigt beispielhaft Bild 8.5.19a). Sie wiederholt sich wie alle Pseudozufallsfolgen periodisch. Systeme, die Pseudozufallssignale empfangen, können diese von echt zufälligen Signalen dann nicht unterscheiden, wenn ihre Speicherkapazität kleiner als die für die Erzeugung bzw. Erkennung einer Pseudozufallsfolge notwendige ist.

Binäre Pseudozufallsfolgen werden durch speziell rückgekoppelte Schieberegister realisiert. Für ein 4-bit-Schieberegister zeigt die Schaltung Bild 8.5.19b). Die damit generierte Folge an den vier Ausgängen enthält Tab. 8.5.1. Sie ist an den vier Ausgängen gleich, jedoch um einen oder mehrere Takte verschoben. Es fällt auf, daß der Zustand ØØØØ fehlt. Wenn er auftritt, blockiert das Register. Er muß daher beim Einschalten vermieden werden. Dies kann z.B. durch spezielle Antiblockierschaltungen erreicht werden. Auffällig ist, daß die Verknüpfung der einzelnen Flipflops über die Exclusiv-Oder- (XOR-) Schaltung erfolgt. Sie tritt bei allen Schieberegistern auf, die Pseudozufallsfolgen erzeugen.

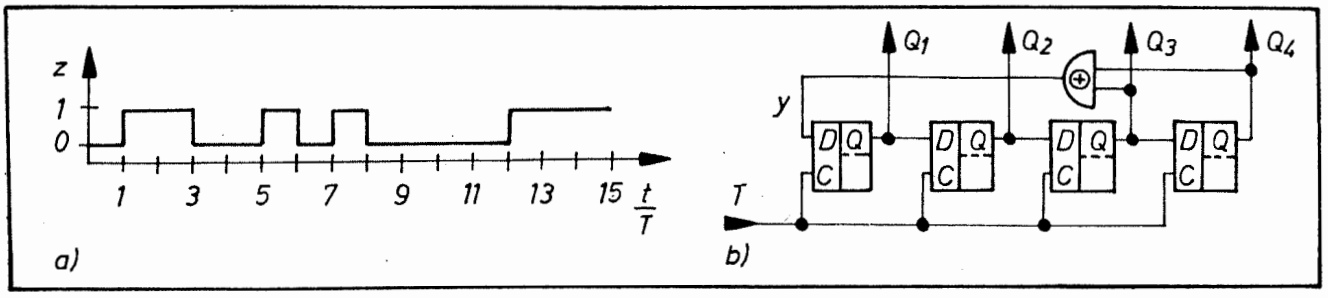


Bild 8.5.19. Zur Erzeugung von Pseudozufallsfolgen

a) Beispiel einer binären Zufallsfolge

b) Beispiel eines Zufallsgenerators mit einem Schieberegister aus vier FF

Tabelle 8.5.1. Zustandstabelle des Pseudozufallsgenerators mit 4 bit

T	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Q_1	1	Ø	Ø	1	1	Ø	1	Ø	1	1	1	1	Ø	Ø	Ø	1
Q_2	Ø	1	Ø	Ø	1	1	Ø	1	Ø	1	1	1	1	Ø	Ø	Ø
Q_3	Ø	Ø	1	Ø	Ø	1	1	Ø	1	Ø	1	1	1	1	Ø	Ø
Q_4	Ø	Ø	Ø	1	Ø	Ø	1	1	Ø	1	Ø	1	1	1	1	Ø
y	Ø	Ø	1	1	Ø	1	Ø	1	1	1	1	Ø	Ø	Ø	1	Ø

Tabelle 8.5.2. Tabelle der Rückkopplungsanschlüsse für Pseudozufallsgeneratoren

n	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Anschlüsse	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	2	3	3	5	4	7	5	7	9	11	10	13	14	14	14	11	18	17
						5				8	6	8		13			17	
						3				6	4	4		11			14	

Für Schieberegister mit 3 bis 20 FF zeigt Tab. 8.5.2 die Rückkopplung über ein XOR zum ersten FF. In einigen Fällen sind dabei vier Eingänge erforderlich. Die Periode der Pseudozufallsfolge ist bei n FF genau $2^n - 1$ Takte lang. Neben der in Tab. 8.5.2 gezeigten Rückkopplung gibt es noch andere Möglichkeiten der Rückkopplung. Sie betreffen nicht den Anschluß zum letzten und ersten FF.

8.5.6. Einfache Rechnerschaltungen

8.5.6.1. Vielfalt der Schaltungen für eine Funktion

Einfache Rechnerschaltungen wurden schon mehrfach behandelt. Besonders häufig wurde dabei bewußt die Addition berücksichtigt. Sie beginnt mit Halbaddern als kombinatorische Schaltungen in Bild 8.2.24 und dem daraus abgeleiteten Aufbau zu Volladdern. In Bild 8.2.24d) ist dann gezeigt, wie durch Reihenschaltung von Volladdern ein Adder für mehrere bit breite Zahlen entsteht. Den Aufbau einer Additionstafel für dezimale Zahlen zeigt Bild 8.2.24d). Die ALU in Bild 8.2.25 realisiert unter anderen Funktionen die Addition mit und ohne Übertrag. Aus Bild 8.2.26 kann leicht abgeleitet werden, wie mit einer PLA ein Adder zu realisieren ist. Schließlich zeigt Bild 8.3.17 ein ROM, welches ein Volladder realisiert. Im Abschnitt 8.5.6.3. wird gezeigt werden, wie ein Mehrbitadder aus Schieberegistern und einem Volladder mit FF für den Übertrag zu realisieren ist.

Diese Aufzählung führt zum späteren Abschnitt 8.9.1., wo allgemein gezeigt werden wird, daß eine Funktion durch viele Strukturen zu realisieren ist.

8.5.6.2. Register für Rechenschaltungen

Bei Rechenschaltungen besitzen die Register eine große Bedeutung: Sie nehmen die Daten für die Rechnung und für das Ergebnis auf und geben sie getaktet weiter. Entsprechend den Aufgaben können mehrere verschiedene Registerschaltungen in bezug auf die Schaltung, welche die zu berechnende Funktion ausführt, unterschieden werden. Besonders häufig sind Zwei- und Dreiregisterschaltungen, wie sie beispielhaft Bild 8.5.20 zeigt. Die Teilschaltung a) ist dann wichtig, wenn *Funktionen einer*

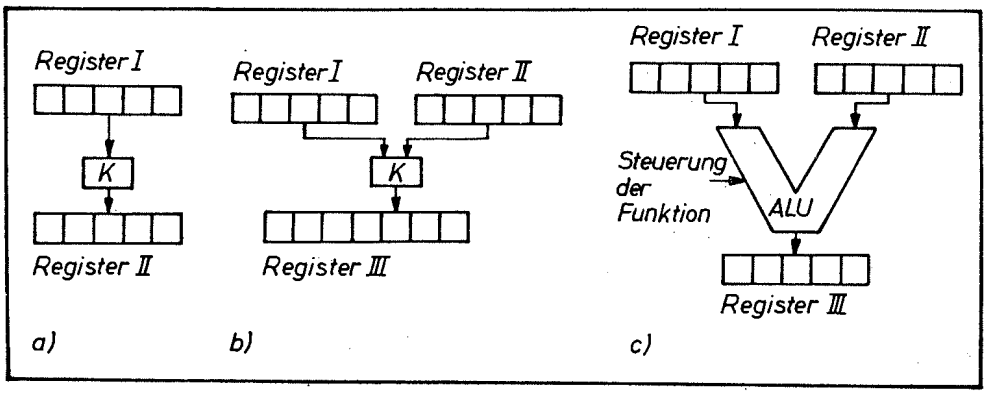


Bild 8.5.20. Anordnung von Registern bei Rechenschaltungen

Variablen, wie z. B.

$$x^2;\quad \sqrt{x};\quad \sin x;\quad \cos x;\quad \log x;\quad x!;\quad x=0;\quad x>y \quad \text{usw.},$$

zu berechnen sind. Die hier eingefügte Schaltung ist vereinfacht als kombinatorische Schaltung K angenommen. Wie im Abschnitt 8.6. gezeigt werden wird, sind die Verhältnisse meist komplizierter, denn es werden Zwischenergebnisse, Entscheidungen usw. benötigt, die zusätzliche Register benötigen.

Eine andere wichtige Klasse von Funktionen geht von *zwei Eingangswerten* aus, z. B.

$$x \pm y;\quad x \cdot y;\quad x/y;\quad x^y;\quad \binom{x}{y};\quad x>y;\quad x=y \quad \text{usw.}$$

Sie benötigt deshalb zwei Eingangs- und ein Ausgangsregister, also drei Register, und entspricht so dem Prinzip von Bild 8.5.20b), wobei oftmals wieder Register für Zwischenwerte bei der Rechnung erforderlich sind. Die Verallgemeinerung dieser Schaltung verwendet statt der kombinatorischen Schaltung eine ALU, welche mittels Steuerung mehrere Funktionen realisieren läßt. Diese Anordnung heißt zuweilen Register-ALU (RALU).

8.5.6.3. Addition mit Schieberegistern

Eine sequentiell arbeitende Additionsschaltung für binäre Zahlen aus mehreren bit, die wesentlich in ihrer Funktion durch Schieberegister bestimmt ist, zeigt Bild 8.5.21a). Sie enthält drei Schieberegister, die mit dem gleichen Takt zur Rechtsverschiebung versehen sind, einen Volladder für 1 bit und ein MS-FF zur Zwischenspeicherung des Übertrags. Die Funktion dieser Schaltung sei am Beispiel der binär durchgeführten dezimalen Addition 13 + 15 = 28 erklärt. Es werde mit dem Zustand begonnen, in dem schon die beiden zu addierenden Zahlen in den beiden Ein-

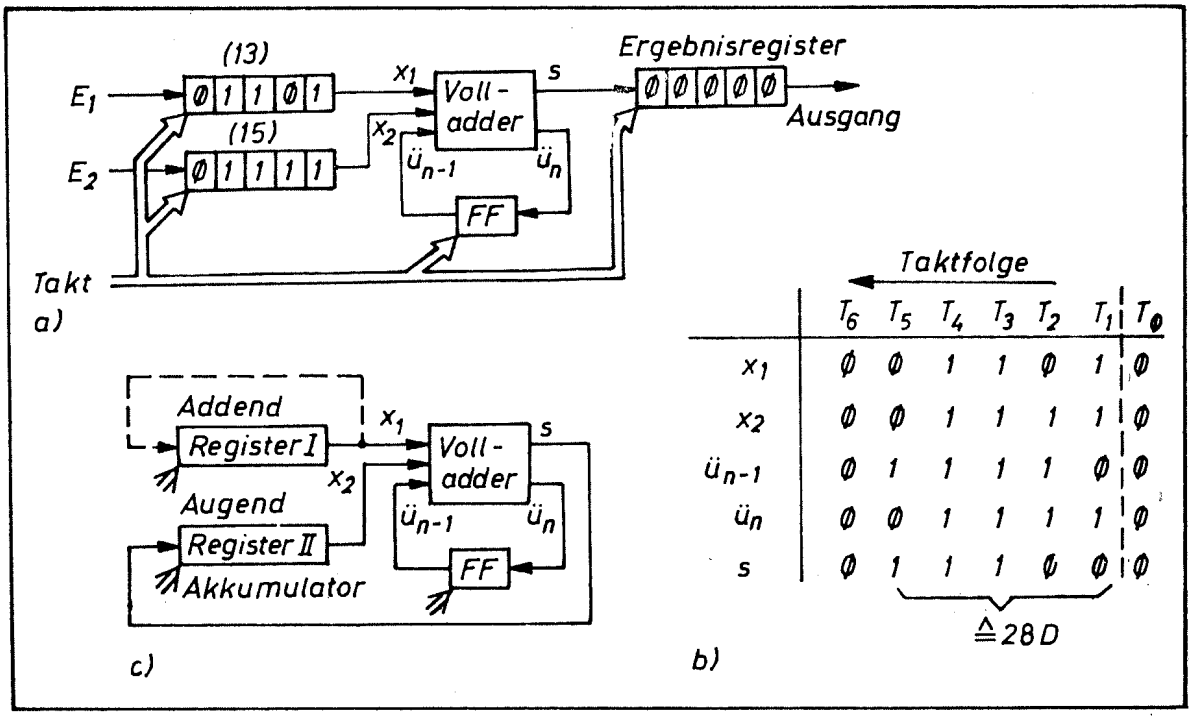


Bild 8.5.21. Addition mit Schieberegistern

a) Dreiregisterschaltung
b) Ablauf der Belegung bei den einzelnen Takten
c) Zweiregisterschaltung

gangsregistern stellengetreu stehen. Das Ausgangsregister sei Ø gesetzt, ebenso das FF. Dies ist der Zustand, wie er auch in der Tabelle gemäß Teilbild b) als T_0 dargestellt ist. Dort sind die Variablen am Eingang des Volladders VA x_1; x_2 und $\ddot{u}_{n-1}$ (Ausgang vom FF, Slave) sowie die am Ausgang s und $\ddot{u}_n$ (Eingang vom FF) als Zeilen aufgetragen. Im Takt T_1 sind am Eingang des VA die niederwertigsten bit der beiden Eingangsregister, und aus $1 \wedge 1$ entsteht $\ddot{u}_n = 1$ (FF) und $s = Ø$ (Ergebnisregister). Mit dem Takt T_2 werden die Registerinhalte um Eins nach rechts verschoben. Am Eingang des VA stehen daher dann $x_1 = Ø$; $x_2 = 1$ und $\ddot{u}_{n-1} = 1$ und erbringen das Ergebnis $\ddot{u}_n = 1$ und $s = Ø$. Mit fünf Takten bis T_5 werden die bit der Eingangsregister vollständig aufaddiert, der Takt T_6 ist notwendig für einen eventuell noch vorhandenen Übertrag. Nach T_6 steht im Ergebnisregister die binäre Summe 11ØØ B = 28 D. Aus der Darstellung dieser Zahl in der Tabelle von Bild 8.5.21b) ist ersichtlich, warum dort die Takte entgegen der Zeitrichtung aufgetragen wurden.

Aus dem *Ablauf der Addition* ist mit Bild 8.5.21a) ersichtlich:

- Die Register des Eingangs werden mit den Takten schrittweise geleert.
- Das Ergebnisregister wird schrittweise gefüllt.

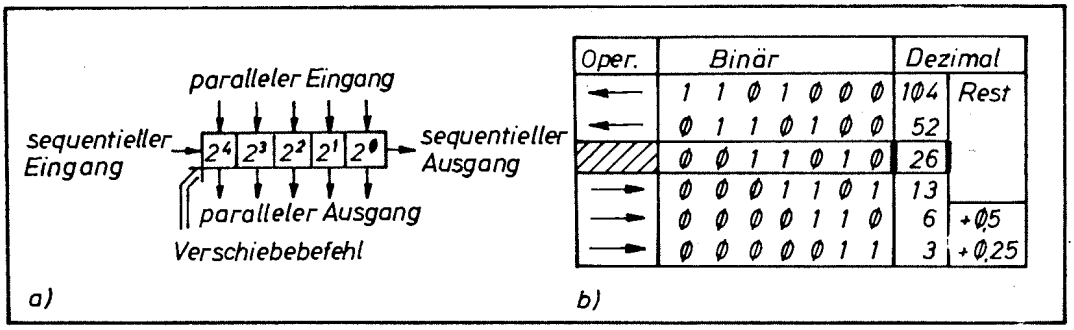
Dies hat Vor- und Nachteile. Zunächst könnte es sinnvoll sein, in beiden oder in einem Eingangsregister den Wert für neue Rechnungen erneut zur Verfügung zu haben. Dann ist dieses Register als Umlaufregister zu schalten. Dies ist in Bild 8.5.21c) für das Addendenregister gestrichelt angedeutet.

Weiter läßt sich unter Ausnutzung der Verschiebungen das Ergebnisregister mit dem Augendenregister als Einheit gestalten: Dort wo das Register leer wird, gelangt sofort das Ergebnisbit hinein. So entsteht der *Akkumulator*, wie ihn ebenfalls Bild 8.5.21c) zeigt, und aus der Dreiregisteraddition ist eine Zweiregisteraddition geworden. Der Inhalt des Akkumulatorregisters wird bei Taschenrechnern immer angezeigt. Er enthält vor der Addition den Augenden und nach der Addition das Ergebnis.

Schließlich sei noch einmal erwähnt, daß die Subtraktion ebenfalls in einem Addierwerk vorgenommen wird. Zuvor ist jedoch von der zu subtrahierenden Zahl das Komplement bzw. die entsprechende negative Zahl zu bilden.

8.5.6.4. Multiplikation und Division mit 2

Im Schieberegister sind die einzelnen FF geordnet verschaltet. Dies ermöglicht es, den einzelnen FF einen Stellenwert für eine Zahl der Basis 2 zuzuordnen. Das entsprechende Prinzip deutet Bild 8.5.22a) an. Wird nun in ein solches Schieberegister eine binäre Zahl eingeschrieben, so entstehen mit den Schiebeoperationen neue Zahlwerte. Dies ist in Bild 8.5.22b) gezeigt. In der hervorgehobenen Zeile ist binär die



Oper.	Binär	Dezimal	
←	1 1 Ø 1 Ø Ø Ø	104	Rest
←	Ø 1 1 Ø 1 Ø Ø	52	
▨	Ø Ø 1 1 Ø 1 Ø	26	
→	Ø Ø Ø 1 1 Ø 1	13	
→	Ø Ø Ø Ø 1 1 Ø	6	+Ø,5
→	Ø Ø Ø Ø Ø 1 1	3	+Ø,25

Bild 8.5.22. Schiebeoperationen und Multiplikation bzw. Division mit 2

a) Wertigkeit der Registerplätze beim 5-bit-Register

b) Folge von Schiebeoperationen bei einer Ausgangszahl 26 in einem 7-bit-Register

dezimale Zahl 26 eingeschrieben. Bei einer Schiebeoperation nach links wird daraus 52, bei einer zweiten 104. Bei einer Schiebeoperation nach rechts 13. Eine Schiebeoperation bedeutet also Multiplikation (links) bzw. Division (rechts) mit 2. Dabei werden Grenzen erreicht, wenn bei den Schiebeoperationen eine 1 das Register verläßt. Dies tritt in Bild 8.5.22 für die zweite Schiebeoperation nach rechts auf. Auf diesen Fehler ist bei derartigen, insbesondere mehrfachen Operationen zu achten.

8.5.6.5. Multiplikation: Beispiel

Ähnlich wie bei der Addition, läßt sich eine Multiplikation auch mittels verschiedener Schaltungen realisieren. Hier sei vom Beispiel der schriftlichen Multiplikation bei dezimalen Zahlen ausgegangen. Es ist für den Fall 33 mal 13 gleich 429 in Bild 8.5.23a) gezeigt. Völlig analog erfolgt auch die Multiplikation binär dargestellter Zahlen. Sie ist in Teilbild b) für dieselben Zahlenwerte dargestellt. Dabei fällt auf, daß sich hier die *Multiplikation auf drei Operationen zurückführen* läßt:

- die Verschiebung des Multiplikanden 100001 B = 33 D entsprechend dem Stellenwert des Multiplikatorteils,
- Nullsetzen des verschobenen Wertes, wenn der Multiplikatorteil gleich Null,
- Addition der verschobenen Werte.

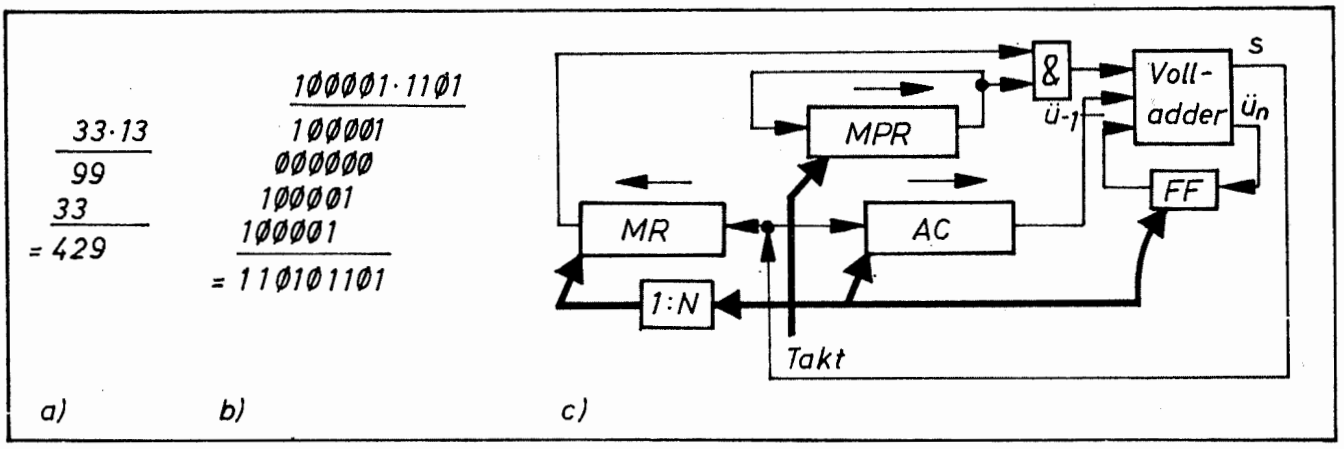


Bild 8.5.23. Zur Multiplikation von Zahlen

a) Schriftliche Multiplikation im Dezimalsystem,

b) im Binärsystem. Bei der Stellenverschiebung wiederholt sich der Multiplikand, oder es ist der Wert Null

c) Schaltbild einer binären Multiplikation mit Registern. Es werden zwei Takte im Verhältnis 1:N benötigt, wobei N die Länge der Register ist. Das Ergebnis steht nach der Multiplikation als $2N$-stellige Zahl im Multiplikatorregister MR und im Akkumulator AC

Die drei Operationen sind nach dem Vorangegangenen sehr einfach zu realisieren, und so entsteht die Schaltung von Bild 8.5.23c), die mit einem einfachen Volladder arbeitet. Zu Beginn stehen im Multiplikandenregister MPR 100001 und im Multiplikatorregister MR 001101. Der Akkumulator AC und das FF sind gelöscht. In das MR gelangt nur jeder N-te (hier 6te) Takt. Die Binärzahlen stehen richtig von links nach rechts angeordnet. Die Schieberichtungen müssen unterschiedlich gewählt werden (Pfeile). Im Gegensatz zur schriftlichen Multiplikation beginnt hier die Multiplikation mit der höchstwertigen Ziffer in MR (also in Teilbild b) unten), jedoch wegen der Stellenlänge zunächst mit zweimal Null. Das MPR ist ein Umlaufregister.

Im ersten Takt wird die niederwertigste Ziffer aus MPR und die höchste aus MR über das AND-Glied verknüpft. Der Ausgang ist Null, falls der Wert aus MR Null ist, sonst wird der Wert aus MPR angenommen. Dies wird zum Wert mit dem niedrigsten Wert aus AC und dem Übertrag aus den FF dem Volladder zugeführt. Taktweise

wird für $N = 6$ Takte so genau der Inhalt des MPR, multipliziert mit höchster Ziffer aus MR, in den Akkumulator addiert. Danach schiebt der Frequenzteiler $1:N$ den Inhalt des MR um eine Stelle nach links, und am AND-Glied wird jetzt der zweithöchste Wert vom Multiplikator wirksam. Gleichzeitig wird der höchste Wert aus dem AC ins MR übernommen. Nun wiederholt sich die Addition der nächsten Stelle in N Schritten. Nach N^2 Schritten ist die Multiplikation ausgeführt und das Ergebnis steht in der richtigen Reihenfolge nacheinander in MR und AC.

Es ist offensichtlich, daß die *Ablaufsteuerung*, d. h. auch das Programm und das damit beeinflußte Steuerwerk, für eine solche Multiplikation schon beachtlich sind. Auf dieses Problem wird im Abschnitt 8.6.1.4. eingegangen.

8.5.7. Zur Datenübertragung

Übertragung ist Aufgabe der Nachrichtentechnik, die nicht Inhalt dieses Buches ist. Dennoch müssen hier einige ausgewählte Probleme berücksichtigt werden. Generell besteht nämlich in der Elektronik die Notwendigkeit, Signale (im folgenden wird zur Vereinfachung Daten verwendet) von einem Geräteabschnitt bzw. Gerät zu einem anderen zu übermitteln. Hierbei handelt es sich dann in der Regel um sehr kurze Wege, so daß z.B. die Probleme der Dämpfung, der Phasenverschiebung und der zeitlichen Differenz zwischen sendendem und empfangendem Gerät vernachlässigt werden können. Hier geht es also vor allem um die Kopplungsprobleme zwischen den Geräten. Dabei werden im folgenden zwei spezielle Probleme ausgewählt:

Handshake-Prinzip
Busprinzip.

Aus beiden gehen die später im Abschnitt 8.7. genauer zu behandelnden Standardsysteme bzw. Standardinterfaces hervor.

8.5.7.1. Handshake-Prinzip

Soll ein Gerät, der Sender, an ein zweites Gerät, den Empfänger, Daten übermitteln, so genügt es nicht, einfach den Ausgang des Senders mit dem Eingang des Empfängers zu verbinden und den Sender die Daten abgeben zu lassen. Dann könnten nämlich folgende *Störungen wirksam* werden.:

- Infolge von Reflexionen und anderen Störungen sind die Daten erst nach einer bestimmten Zeit auf der Leitung stabil. Der Empfänger darf deshalb die Daten nicht zu früh übernehmen.
- Der Empfänger kann im Vergleich zum Sender zu langsam sein. Er würde dann einige Daten verlieren.
- Der Empfänger kann schließlich prinzipiell nicht imstande sein, Daten zu übernehmen, weil er noch mit anderen Aufgaben (z.B. auch Verarbeitung vorangegangener Daten) beschäftigt ist.

Die ersten beiden Fakten ließen sich durch einen synchronen Betrieb beider Geräte vermeiden, der mit der richtigen Taktfrequenz erfolgt. Dies würde eine Taktleitung zwischen Sender und Empfänger erfordern, wobei der Takt vom Sender oder Empfänger bestimmt werden kann. Das dritte Problem ist im wesentlichen unabhängig von den beiden ersten. Es kann nur vermieden werden, wenn der Empfänger dem Sender über eine Leitung mitteilt, daß er zur Übernahme von Daten bereit ist.

Der typische Fall für die Datenübertragung zwischen Geräten und Gerätesystemen ist der asynchrone. Um dabei alle oben genannten Probleme auszuschalten, ist ein spezielles *Handshake* entstanden, das im Prinzip Bild 8.5.24 zeigt. Es enthält zwei zusätzliche Leitungen (*DAV* und *DAC*): Es werde davon ausgegangen, daß der

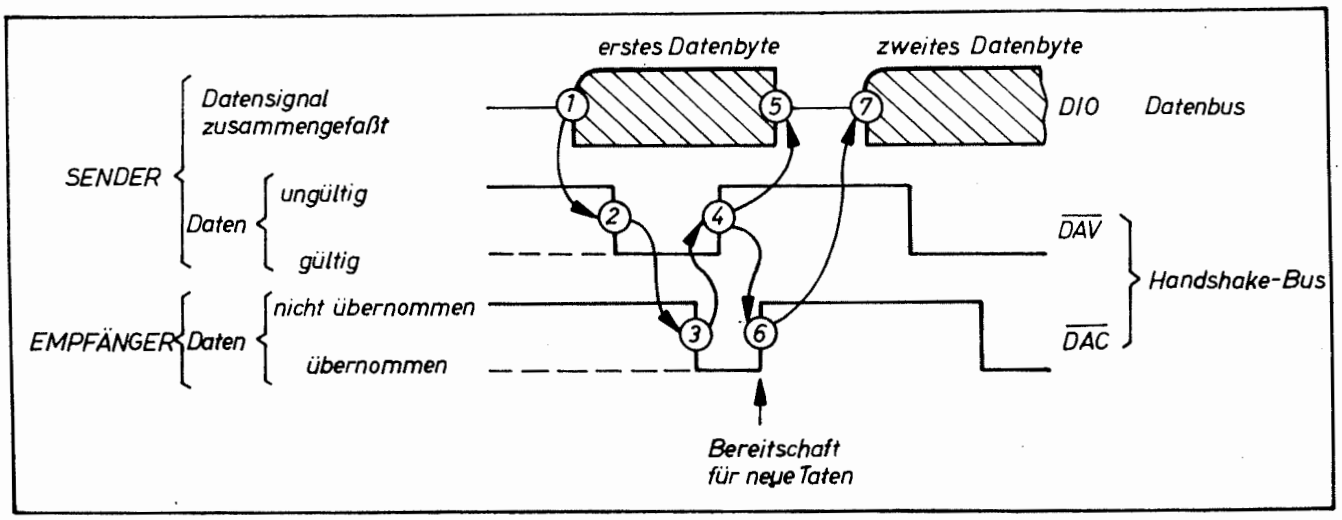


Bild 8.5.24. Prinzip des Zweileitungs-Handshake

Es bedeuten

DIO data-in-out = Datenleitung, eventuell mit mehreren parallelen Leitungen für ein Datenwort
DAV data valid = Daten gültig
DAC data accepted = Daten übernommen
Die Negierung von DAV und DAC wurde im Vorgriff auf den IEC-Bus (Abschnitt 8.7.4.1.) gewählt

Empfänger zur Übernahme von Daten bereit ist. Dies ist dadurch belegt, daß die Leitung *DAC* auf H liegt. Dann legt der Sender die Daten auf den Datenbus ①, wartet die Zeit ab, bis die Daten stabil sind, und legt dann die Leitung DAV auf L ②. Damit kann der Empfänger die Daten übernehmen. Ist dies erfolgt, so legt er die Leitung *DAC* auf L. Hierdurch weiß der Sender, daß er seine Daten zurücknehmen kann. Er erklärt deshalb die Daten zunächst für ungültig (*DAV* auf H) ④ und nimmt dann die Daten weg ⑤. Nach einer gewissen Zeit ist der Empfänger mit dem Übernehmen der Daten und deren Verarbeitung fertig. Er ist wieder empfangsbereit, deshalb legt er *DAC* auf H ⑥. Dies erfährt der Sender und kann mit neuen Daten beginnen ⑦.

8.5.7.2. Bussysteme

In vielen Fällen sind nicht zwei Geräte, sondern viele, z. B. informationell, zu koppeln. Sie können dabei auf die unterschiedlichste Weise zusammengeschaltet werden. Hierauf wird u. a. bei den Standardinterfaces im Abschnitt 8.7. eingegangen werden. Besonders vorteilhaft hat sich jedoch eine Struktur gezeigt, bei der alle Sender und/oder Empfänger gemäß Bild 8.5.25 an eine gemeinsame Sammelleitung (Sammelschiene) geschaltet sind. Für solche Sammelleitungen hat sich der Begriff Bus durchgesetzt. Er geht auf das lateinische omnis bzw. omnia (= jeder, alle, alles, ganz) zurück, kam mit dem Omnibus (Verkehrsmittel für alle) über das Französiche in fast alle Sprachen und wurde dann bzw. dabei zu Bus verkürzt. Über so einen Bus (Leitungssystem) kann also im Prinzip jedes angeschlossene Gerät an jedes andere angeschlossene Gerät

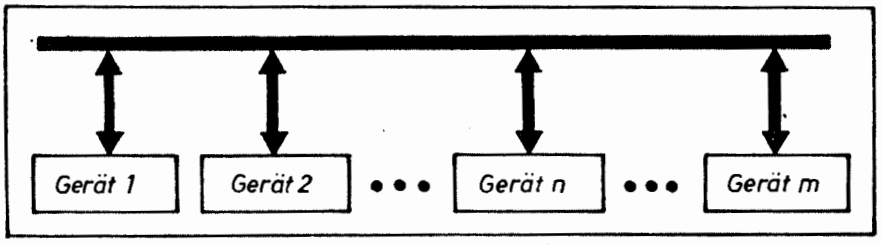


Bild 8.5.25. Prinzip der Anschaltung mehrerer Geräte an eine gemeinsame Busleitung

Information übermitteln. Hierzu werden die einzelnen Geräte mit Adressen — ähnlich, wie die Speicherzellen in einem Speicher — versehen. Über diese Adressen wird mitgeteilt, welches Gerät (Nr. = Adresse) an welches bzw. welche die Information übertragen will. Weiter müssen auf diesem Bussystem dann auch zur zuverlässigen Datenübertragung die zuvor geschilderten Handshake-Signale übertragen werden. Ein Bus enthält daher in der Regel *drei Signalgruppen*:

- zu übertragende Daten = Datenbus,
- zu übertragende Adressen = Adressenbus,
- zu übertragende Hilfsinformation, z.B. Handshake, = Steuerbus.

Diese drei Busteilsysteme können in der Praxis sehr unterschiedlich gestaltet sein. Die einzelnen Übertragungen können nacheinander oder parallel erfolgen. Unabhängig hiervon kann dies auch für die Signale der drei Teilbusse erfolgen. Es gibt folglich sehr vielfältige Bussysteme. Jeder im folgenden zu behandelnde Mikrorechner hat sein spezielles Bussystem nach außen. Intern existiert bei ihm ebenfalls ein spezielles Bussystem. Darüber hinaus gibt es Standardbusse. Alle diese Möglichkeiten werden im folgenden noch weiter detailliert.

8.6. Mikrorechentechnik

8.6.1. Einführung und Abgrenzung

8.6.1.1. Mikroelektronik, -rechner, -prozessor

Drei Begriffe werden fälschlicherweise in der letzten Zeit z. T. fast wie Synonyme verwendet. Es ist sinnvoll, sie gleich zu Beginn dieses Abschnitts gegeneinander abzugrenzen:

Mikroelektronik beinhaltet alle jenen Gebiete, die sich mit der Herstellung und Anwendung von integrierten Schaltkreisen befassen. Hierzu gehören u. a. TTL- und MOS-Schaltkreise, also kombinatorische, sequentielle und Speicherschaltkreise, Schaltkreise für verschiedene Anwendungen, z.B. Rundfunk, Fernsehen, Tonbandgeräte und industrielle Anwendungen, also analoge und digitale Schaltkreise. Dieses Buch betrifft also zu mehr als der Hälfte die Mikroelektronik.

Mikrorechner oder Mikrocomputer sind ein Teilgebiet der Rechentechnik, also der elektronischen Datenverarbeitung (EDV). Ihre Besonderheit besteht darin, daß sie zum überwiegenden Teil mit digitalen Bauelementen der Mikroelektronik aufgebaut sind. In diesem Sinne gehört der Mikrorechner sowohl zur Mikroelektronik als auch zur Rechentechnik. Den Mikrorechner betrifft folglich insbesondere nur dieser Abschnitt 8.6. sowie im erweiterten Sinne Kapitel 8. und z. T. Kapitel 9.

Mikroprozessor oder etwa synonym CPU (central processor unit) ist ein spezieller Schaltkreis der Mikroelektronik und ein besonders wichtiges Bauelement der Mikrorechner. Das Entstehen des Mikroprozessors war u. a. entscheidend für die Möglichkeit des Aufbaus von Mikrorechnern. Er stellt nicht das komplexeste (z.B. bezüglich der auf dem Chip enthaltenen Transistorfunktionen) Bauelement der Mikroelektronik dar, sondern ist vielmehr dadurch gekennzeichnet, daß er das „intelligenteste" Bauelement ist. In diesem Sinne wird er in diesem Abschnitt auch im Zentrum der Betrachtungen stehen. Er kann aber nur gekoppelt mit anderen Bauelementen wirksam werden. Wenn die bewußt in Anführungsstriche gesetzte „intelligente Eigenschaft" des Mikroprozessors zum Analogschluß (der nur bedingt richtig ist) auf den Menschen verwendet wird, kann er mit unserem Gehirn verglichen werden. Allein für sich ist es

sinnlos, doch in Verbindung mit allem anderen, was zum Menschen (anatomisch funktionell gesehen) gehört, ist es extrem wichtig und bestimmt in vielen Punkten seine Leistungsfähigkeit. Dem Mikroprozessor selbst sind nur die Abschnitte 8.6.2. bis 8.6.4. gewidmet.

Mit den gegebenen Erklärungen lassen sich die vier Gebiete mengentheoretisch gegeneinander abgrenzen. So entsteht Bild 8.6.1. Auffällig ist, daß — vom Mikroprozessor abgesehen — jeder Begriff Einheiten, Baugruppen usw. besitzt, die nur zu ihm gehören. Für die Mikrorechentechnik sind dies z.B. die Floppy discs. Ferner werden Mikroprozessoren aber auch außerhalb der Mikrorechentechnik verwendet.

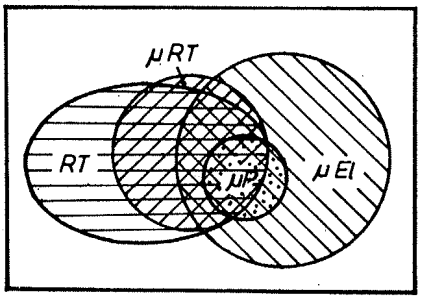


Bild 8.6.1. Zur Abgrenzung von Rechentechnik RT, Mikrorechentechnik μRT, Mikroelektronik μEl und Mikroprozessor μP

8.6.1.2. Anwendungen der Rechentechnik

Rechentechnik wird oft — z. T. noch verstärkt durch den Taschenrechner bedingt — mit der Zahlenrechnung oder mit mathematischen Problemen identifiziert. Hierzu trägt weiter der früher stark verbreitete Begriff Ziffernrechenautomat bei. In der Tat operiert die digitale Rechentechnik ausschließlich mit Zahlen, meist sogar nur mit binären Zahlen. Auf diesem Wege ist sie aber imstande, viele andere Probleme zu lösen. Dies liegt an mehreren Fakten, auf die z. T. schon eingegangen wurde.

Physikalisch gesehen kennt der Rechner nur die beiden Signalzustände L und H. Über die positive Logik (vgl. 8.2.3.3.) werden ihnen die binären Ziffern Ø und 1 zugeordnet. Wie im Abschnitt 8.1.4.9. gezeigt wurde, können einer Kombination von mehreren binären Zuständen (die auch als Zahlen interpretiert werden können) Buchstaben, Befehle usw. zugeordnet werden. Über solche Zuordnungen wird der digitale Rechner zum Universalrechner. Wie universal hier einzugrenzen ist, wird im Abschnitt 8.6.1.7. gezeigt werden. Hier seien jetzt einige wichtige Gebiete der Anwendung der Rechentechnik aufgezählt. Dabei kann keine Vollständigkeit erreicht, sondern es können nur Beispiele angegeben werden:

Die *Numerik* erfolgt heute auf der Basis der Rechentechnik. Es werden also mathematische Ausdrücke bezüglich ihrer Zahlenwerte behandelt, Formeln ausgewertet, Integrale gelöst, Funktionen approximiert usw. Auf einige Probleme der Numerik im Zusammenhang mit der Rechentechnik wird noch im Abschnitt 8.8. eingegangen.

Mit der Numerik in engem Zusammenhang steht die Fähigkeit des Rechners, *logische Entscheidungen* zu treffen. Dies wurde für einfache Fälle bereits im Zusammenhang mit den Bildern 8.2.23 und 8.2.25 gezeigt, läßt sich aber mittels des Rechners auf sehr komplexe Aufgabenstellungen ausdehnen.

Ein großes Gebiet ist die *Textverarbeitung*, in der mit alphanumerischen Daten operiert wird. Dieses Aufgabengebiet ermöglicht u. a. das Schreiben auf einen Bildschirm, wo leicht Korrekturen möglich sind. So kann ein fehlerfreies Manuskript erstellt werden, wofür noch viele Hilfsmittel bereitstehen, wie leichtes Einfügen von Texten, Umordnen von Abschnitten usw. Erst danach wird das Manuskript geschrieben. Dabei können u. a. noch Formataufgaben erledigt werden, wie z.B. Vorgabe der

Zeilenlänge, automatische Numerierung von Seiten. Mit diesen Methoden wird u. a. eine beachtliche Rationalisierung der Büroarbeit erreicht. Zur Textverarbeitung gehören aber auch Prozesse wie Ordnen, Sortieren, Auswählen. So können u. a. mit der Textverarbeitung Sachregister, Datenbanken usw. erstellt werden. Selbst ein interaktiver Betrieb mit einer Sprache, die der natürlichen ähnlich ist, könnte hier Bedeutung erlangen.

Mit der Bildschirmtechnik wurde es auch möglich, eine *automatische Bildverarbeitung* zu realisieren. Sie ermöglicht u. a. ein automatisches Auswerten von Bildern durch Konturenverschärfung, Anwendung von Falschfarben zur Hervorhebung von Details usw. Sie eröffnet aber auch für die Konstruktion und Entwicklung neue Möglichkeiten, z. B. den rechnergestützten Entwurf von Schaltungen, Leiterplatten oder Layouts für die Mikroelektronik. Dieses Gebiet befindet sich in starker Entwicklung.

Die *automatische Meßtechnik* ist insbesondere für die experimentellen Arbeiten wichtig, z. T. schon unverzichtbar. Mit ihr werden Meßwerte automatisch in den Rechner übernommen und verarbeitet. Dabei können u. a. Mittelwerte, Tendenzen, Korrelationen usw. erkannt werden. Für die automatische Meßtechnik sind die Standardinterfaces (8.7.) meist notwendig.

Im erweiterten Sinn kann auch die *Zeichenerkennung* zur Meßtechnik gerechnet werden. Hier werden von den Objekten (z. B. Schrift, Teile, Szenen) Meßwerte abgenommen und daraus durch Rechenprinzipien bestimmt, was für ein Objekt vorliegt.

Die *Automatisierung* der Produktion benutzt im großen Umfang Rechentechnik. Dies beginnt bei dem Anschalten von Mikrorechnern an kleine Prozesse und reicht bis zu vernetzten Prozeßrechnern in Großanlagen. Im Prinzip werden dabei immer an bestimmten Stellen des Prozesses Meßinformationen, z. B. über Temperatur, Druck, Durchfluß usw., abgeleitet. Diese Information wird im Rechner verarbeitet, und dabei werden Stellsignale gebildet, die den Prozeß beeinflussen, z. B. Ventile öffnen oder schließen, Heizungen anschalten usw.

Das soeben beschriebene Prinzip gilt in erster Linie für kontinuierliche Prozesse. Bei Stückgut- und Montageprozessen usw. gewinnt dagegen der *Roboter* an Bedeutung. Auch für ihn gilt die sensorische und stelltechnische Seite und sein Zentrum ist ein Rechner.

8.6.1.3. Klassischer Rechenautomat

Die Operationen mittels Rechenautomaten seien zur Vereinfachung jetzt Rechnungen genannt, schließen dann also das numerische Rechnen ein. Sie lassen sich heute auf vielfältige Weise durchführen, z. B. rein im Gedächtnis des Menschen, unter Zuhilfenahme von Papier und Bleistift, mittels eines Tisch- oder Taschenrechners und schließlich mittels eines Mikro-, Mini- oder Großrechners. Allen Methoden liegt dasselbe Prinzip zugrunde: Aus vorhandenen Daten werden mittels bekannter Methoden und Regeln neue Daten gewonnen. Aus dieser Sicht sind *zwei Arten von Informationen* zu unterscheiden:

- *Daten*, also Zahlen, alphanumerische, geometrische oder sonstige Größen;
- *Programme*, Algorithmen, Regeln, Abläufe, nach denen die Daten verarbeitet werden.

Beide Informationen sind für den Rechner zunächst zu speichern. In den alten Rechnern steht das Programm ein für alle Male gespeichert im Rechner fest und ist nicht veränderbar. Sie sind daher auch nur für einige wenige Spezialaufgaben geeignet. Der einfache Taschenrechner z. B. für die vier Grundrechenarten. Die entscheidende

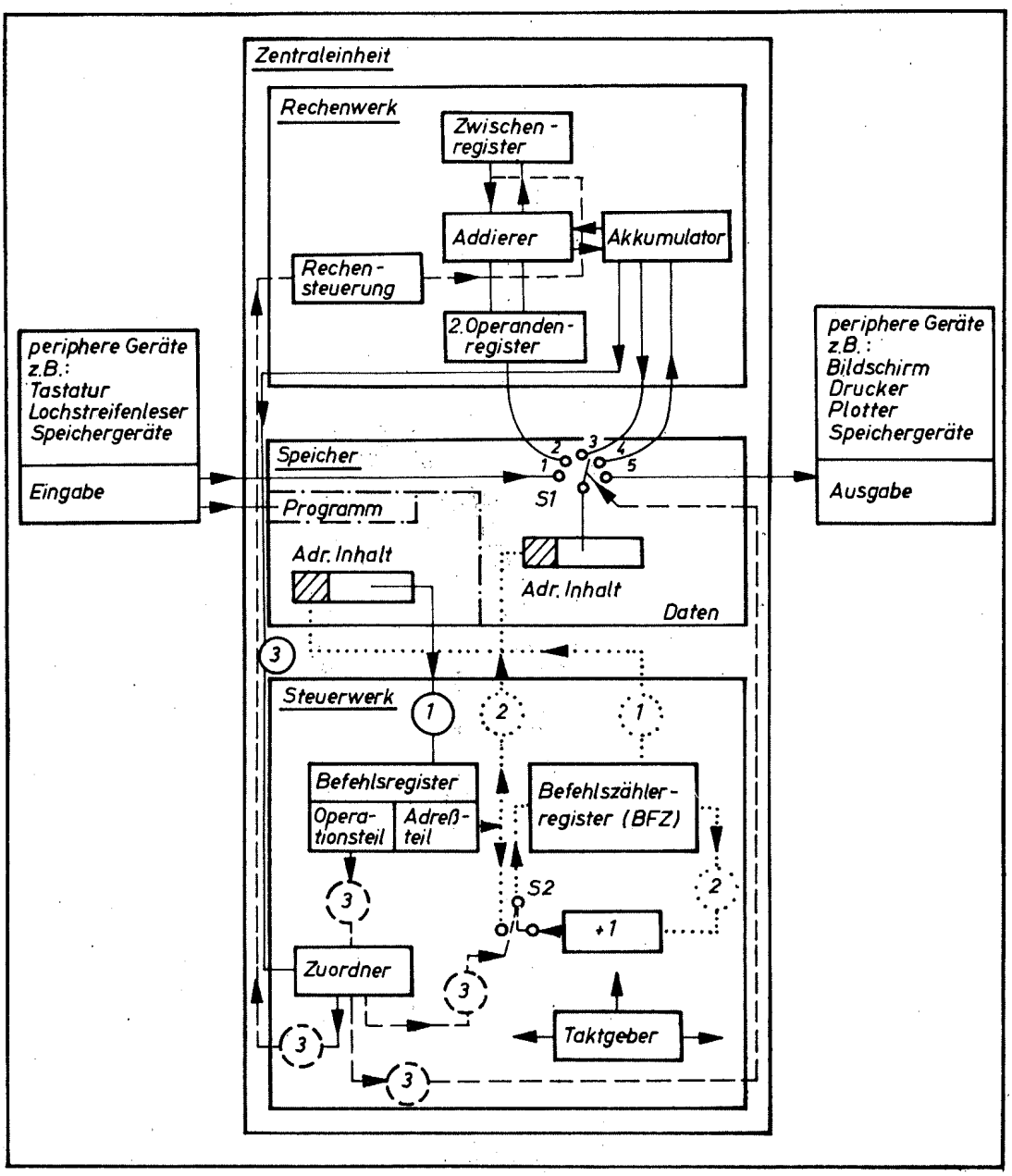


Bild 8.6.2. Prinzip des Aufbaus eines klassischen Rechenautomaten

Idee des heute als klassisch zu bezeichnenden Rechners geht auf J. v. NEUMANN und das Jahr 1945 zurück. Er schlug vor, den Programm- und Datenspeicher als Einheit aufzufassen und aufzubauen. Damit war es möglich (aber auch notwendig), dem Rechner vor Beginn der Arbeit das richtige Programm einzugeben. Dieser Nachteil wird dadurch aufgewogen, daß so jedes nur erdenkliche Programm eingegeben werden kann und der Rechner zum Universalrechner wird (Grenzen: 8.6.1.7.). Einen nach diesem *Prinzip* aufgebauten *Rechner* zeigt schematisch Bild 8.6.2. Er besteht aus *fünf Grundeinheiten*:

- *Rechenwerk* mit ALU und Rechensteuerung;
- *Steuerwerk*, zuweilen auch Leitwerk genannt;
- *Speicherwerk*, auch Speicher oder Arbetisspeicher genannt, wobei insbesondere eine Abgrenzung gegenüber peripheren Speichern notwendig ist;
- *Eingabeeinrichtungen*, mit denen Daten und Programme eingegeben werden;.
- *Ausgabeeinrichtungen*, mit denen Daten (Programme) ausgegeben werden.

Vielfach werden die drei Gruppen Rechen-, Steuer- und Speicherwerk als Zentraleinheit (oft auch CPU = central processor unit genannt) zusammengefaßt. Im Rahmen der Mikrorechentechnik konnte diese Einheit auf einem Chip untergebracht werden. Dabei wurde aber eine erhebliche Einschränkung der Speicherkapazität, meist auf nur wenige Register, notwendig.

Das *Rechenwerk* ist im wesentlichen eine ALU, eventuell mit Registern (RALU) und einer kleinen zusätzlichen Rechensteuerung für einfachste Programmabläufe der ALU. Es ist damit ausreichend im Zusammenhang mit Bild 8.2.25 und dem Abschnitt 8.5.6. behandelt. In der Regel ist ein Rechenwerk jedoch erheblich komplexer. Dies wird allein aus der Anzahl der verarbeitbaren Befehle, die meist über hundert liegt, deutlich. Ferner unterscheiden sich die verschiedenen Rechenwerke dadurch, wie schnell sie sind. Die Schnelligkeit wird mittels Operationen je Sekunde oder mit der zulässigen Taktfrequenz angegeben.

Das *Speicherwerk* ist ein mehr oder weniger umfangreicher Speicher. Je nach Größe und Typ des Rechners reicht die unmittelbar zur CPU gehörende Kapazität von wenigen Registern bis zu vielen Kbyte. Beim Mikrorechner ist im Gegensatz zum Klein- bzw. Großrechner zur Kapazität der CPU (die wenigen Register) noch die direkt adressierbare externe Kapazität zu addieren. Dann liegt auch der Mikrorechner heute bei Kapazitäten von wenigen bis zu ca. 100 Kbyte. Beim Speicherwerk ist wichtig, daß alle Speicherzellen zur Wortorganisation in einer bestimmten Wortbreite, meist 8, 16 oder 32 bit, organisiert und ihr Inhalt über eine Adresse zugänglich ist.

Das *Steuerwerk* ist in den bisherigen Betrachtungen höchstens indirekt behandelt worden. Es enthält folgende Teile:

Der *Taktgeber* schaltet die einzelnen Teile in richtiger Reihenfolge zusammen. In diesem Beispiel werden drei Taktteile angenommen, die im Bild durch ①, ② und ③ gekennzeichnet sind. In der Praxis kommen, wie beim Mikroprozessor noch zu behandeln ist, durchaus mehr Taktteile vor. Vereinfacht können sie wie die beiden Teiltakte beim MS-FF angesehen werden.

Der *Befehlszähler* wird beim Start eines Programms auf Null gesetzt und in jedem Takt beim Taktteil ② um Eins erhöht. Er zeigt damit der Reihe nach auf die Adressen der Speicherplätze, in welchen die Befehle gespeichert sind, die für das Programm in dieser Reihenfolge durch das Rechenwerk bzw. zur Ein- und/oder Ausgabe hin abgearbeitet werden.

In der Regel ist ein Programm nicht linear angeordnet, sondern aufgrund von Entscheidungen oder zur Vereinfachung (Unterprogramme) können Sprünge in Programmteile erfolgen, die auf anderen Adressen stehen. Dies ist möglich, indem der Schalter $S2$ (natürlich ebenfalls durch das Programm gesteuert) den Befehlszähler auf den Adreßteil des Befehlsregisters anschaltet.

Im Taktteil ① sendet der Befehlszähler die Adresse des jetzt benötigten Speicherplatzes des Programms aus. Der Inhalt dieser Speicherzelle gelangt dann zum *Befehlsregister*. Es zerlegt diesen Inhalt in zwei Teile (Taktteil ③):

Operationsteil und Adreßteil.

Der *Operationsteil* entscheidet über einen Zuordner (Dekoder), ob der Befehl der Rechensteuerung für die ALU, den Schalter $S1$ oder den Schalter $S2$ gilt. Gleichzeitig wird dann entschieden, was dort zu tun ist: im Rechner z.B. zu addieren oder eine logische Entscheidung zu treffen, den Schalter $S1$ in welche Stellung zu schalten, dito den Schalter $S2$.

Der *Adreßteil* aktiviert eine Speicherzelle bezüglich des Datenteils und bringt ihren Inhalt je nach der Stellung des Schalters $S2$ zur Ausgabe, zum Akkumulator oder zum

2. Operandenregister. In Stellung 1 bzw. 3 wird in die Speicherzelle ein Wert vom Eingang bzw. vom Akkumulator eingeschrieben.

Zur besseren Übersicht sind die *Informationen* in Bild 8.6.2 *unterschiedlich gekennzeichnet*:

- Speicherinhalte (Daten und Programme) sind ausgezogene Linien,
- Adreßinformationen sind gepunktet,
- Befehlsinformationen sind gestrichelt.

Die *Ein- und Ausgabegeräte* eines Rechners stellen seine Peripherie dar, ohne die er nicht arbeitsfähig ist und deren Umfang z. T. wesentlich seine Größe, seinen Preis und seine Leistungsfähigkeit bestimmt. Zum Teil sind dabei Ein- und Ausgabegeräte gleichartig. Dies gilt insbesondere für die verschiedenen peripheren Speicher, wie Floppy disc (vor allem Mikrorechner), Plattenspeicher der verschiedensten Art, z.B. Festplattenspeicher, Wechselplattenspeicher, WINCHESTER-Typ (Festplatte mit Luftfilter), früher Trommelspeicher, dann verschiedene Magnetbandgeräte und andere Massenspeicher, wie z.B. Bandbibliothek.

Andere periphere Geräte dienen nur der *Eingabe*. Hierzu zählen vor allem die Tastaturen, Lesegeräte, Bildabtaster, Lochkarten- und Lochbandeingabe.

Spezifisch für die *Ausgabe* sind dagegen der Bildschirm, Drucker der verschiedensten Art, z.B. Schnelldrucker, Paralleldrucker, Mosaikdrucker, Schreibwerk, Schreibmaschine, Plotter und Lochkarten- bzw. Lochbandstanzer.

Eine wichtige Kombination von Ein- und Ausgabe stellt das Display dar. Es besteht im wesentlichen aus einer Tastatur, die mit einer Bildschirmeinheit gekoppelt ist und über Leitungen, meist in größerer Anzahl, an einen Rechner angeschlossen sind.

8.6.1.4. Ablauf einer Rechnung

Als Beispiel für eine Rechnung sei die Sinusfunktion gewählt. Vorausgesetzt werde dabei, daß der vorhandene *Rechner bereits die folgenden Befehle besitze*:

- Transportieren eines Zahlenwertes auf und von einem Speicherplatz,
- Addition von Zahlenwerten,
- Multiplikation und Division von Zahlenwerten,
- Entscheidung, ob eine Zahl größer als eine andere ist,
- Vorzeichenwechsel,
- Betragsbildung.

Von der Multiplikation (und Division) abgesehen, besitzt bereits jeder Mikrorechner diese Befehle. Bei ihm müßte also ein zusätzliches Programm im Sinne von Bild 8.5.23 realisiert werden. Unter diesen Voraussetzungen ist es möglich, den Wert einer Sinusfunktion mittels der TAYLOR-Reihe

$$\sin x = \frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} \pm \dots \tag{1}$$

darzustellen. Sie hat aber in dieser Form einen erheblichen Nachteil:

- Für jedes neue Glied sind immer kompliziertere Multiplikationen notwendig.

Dies läßt sich durch eine einfache Umschreibung der Reihe verändern:

$$\sin x = \frac{x}{1!} - \frac{x}{1!}\cdot\frac{x^2}{2\cdot 3} + \frac{x^3}{3!}\cdot\frac{x^2}{4\cdot 5} - \frac{x^5}{5!}\cdot\frac{x^2}{6\cdot 7} \pm \dots . \tag{2}$$

0. 1. 2. 3. Glied

Es ist ersichtlich, daß sich das $(i + 1)$te Glied aus dem i-ten wie folgt ableitet:

$$t_{i+1} = (-1) \cdot t_i \cdot \frac{x^2}{2i(2i + 1)}. \tag{3}$$

Damit besteht die Möglichkeit einer rekursiven Programmierung, die meist erhebliche Vorteile bringt. Hierzu ist es notwendig, folgende Speicherplätze bereitzustellen. Sie werden hier wie in einer Programmiersprache durch Symbole als Variable bezeichnet. Für diese direkte Programmierung müßte statt ihrer die Adresse des Speicherplatzes stehen.

X Hier wird zu Beginn die Variable der Funktion eingegeben und bleibt bis zum Ende der Rechnung so.

J wird zu Beginn Null gesetzt und für jedes neue Glied um Zwei erhöht. Der Wert entspricht also $2i$ in Gl. (3) (Laufparameter).

T wird zu Beginn mit X belegt, also entsprechend dem ersten Glied der TAYLOR-Reihe, und nimmt dann gliedweise die Werte gemäß Gl. (3) als t_{i+1} an (Teilglied).

S Hier soll am Ende der Rechnung das Ergebnis stehen. Zu Beginn wird dieser Speicher entsprechend dem ersten Glied ebenfalls mit dem Wert von X gelegt. Gliedweise werden dann die nächsten Glieder gemäß

$$s_{i+1} = s_i + t_{i+1}\,. \tag{4}$$

aufaddiert.

$X2$ Hierin wird x^2, einmal gebildet, gespeichert, um es in der wiederkehrenden Gl. (3) nicht immer neu zu bilden und so Zeit zu sparen.

Damit besteht aber noch ein Problem, nämlich wann diese Reihe abgebrochen werden kann oder soll. Hierzu muß ein Fehler (Error) vorgegeben werden, für den ein weiterer Speicherplatz vorzusehen ist:

E Hier wird zu Beginn der maximal zulässige Fehler eingegeben.

Der Abbruch der Rechnung erfolgt dann, wenn das Teilglied kleiner als der vorgegebene Fehler ist:

$$t_{i+1} < E; \qquad (T < E)\,. \tag{5}$$

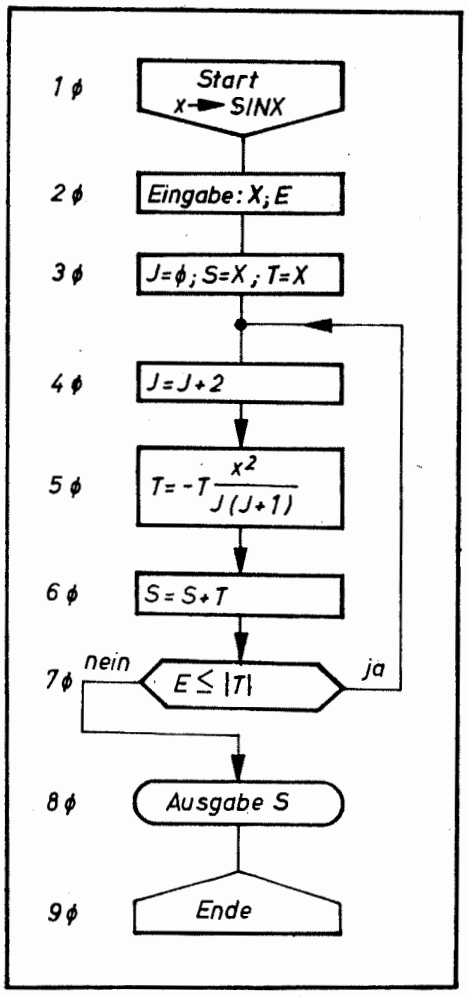
Eine formale Beschreibung dieses bisher verbal geschilderten Rechenablaufs ist in Bild 8.6.3 in zwei Varianten gegeben, und zwar als Programmablaufplan (PAP) und in der Sprache BASIC. Beide dürften nach dem hier Geschilderten unmittelbar lesbar sein. Dabei ist lediglich zu beachten, daß das Gleichzeichen (=) in diesem Fall als Ergibtzeichen zu interpretieren ist. Es wird z. T. daher auch als := bzw. → geschrieben. Der Ausdruck $A = B$; $A := B$; $A \rightarrow B$ bedeutet dann, daß der Speicher A den Wert des Speichers B übernimmt. REM bedeutet Remark (= Bemerkung/Kommentar).

Da der Ablaufplan und das BASIC-Programm mit der gleichen Zeilennummer gekennzeichnet sind, ist es jetzt relativ leicht, den Bezug zum Bild 8.6.2 herzustellen. Das Programm befindet sich dann in einer entsprechenden Reihenfolge, wie es in die Speicherzellen des Speichers geschreiben ist, und die Zeilen-Nr. dienen als grober Anhaltspunkt für mehrere Adressen, die in jeder Zeile abzufragen sind. Die Speicherzellen X; J; T; S; $X2$ und E sind Datenspeicher mit entsprechenden Adressen.

Die Zeile 10 ist nur ein Hinweis zur besseren Lesbarkeit des Programms. Sie wird im Programm selbst nicht weiter verwendet. Dies drückt der Befehl REM aus.

Mit Zeile 20 wird $S1$ auf Stellung 1 gebracht und zunächst auf Adresse X der X-Wert und dann (nächster Befehl) auf Adresse E der zulässige Fehler geschrieben.

In Zeile 30 wird aus dem Rechenwerk $J = \emptyset$ gesetzt, dem Akkumulator X zugeführt und dann von ihm in S und T X überschrieben. Weiter wird im Rechenwerk mittels



BASIC-Programm

```
1Ø REM: Berechnung von SIN(X)
2Ø INPUT „Variable, Fehler"; X, E
3Ø J=Ø: S=X: T=X: X2=X*X: REM:
   Startwerte
4Ø J=J+2
5Ø T=-T*X2/J/(J+1)
6Ø S=S+T
7Ø IF E<=ABS(T) GOTO 4Ø
8Ø PRINT S
9Ø END
```

Bild 8.6.3. Programmablaufplan zur Berechnung von sin (x) und zugehöriges BASIC-Programm

Zwischenregister X^2 gebildet und danach vom Akkumulator in $X2$ geschrieben. REM wird nicht bearbeitet.

In Zeile 40 wird J durch zweimaliges Inkrement auf $J + 2$ erhöht.

In Zeile 50 wird durch mehrfache Operationen zwischen Speicher und Rechenwerk die Gl. (3) realisiert.

In Zeile 60 wird zu S der Wert von T addiert.

In Zeile 70 wird zunächst T in das Rechenwerk gegeben, dann Betrag gebildet und im Zwischenregister gespeichert. Dann gelangt E in den Akkumulator, und es wird die logische Gültigkeit von Gl. (5) gebildet. Das Ergebnis beeinflußt den Zuordner und schaltet $S2$ um oder nicht um. Dementsprechend wird zur Zeile 40 zurückgesprungen (Befehlszähler auf entsprechenden Wert gesetzt) oder das Programm mit Erhöhung des Befehlszählers weiter fortgesetzt.

Zeile 80 druckt das Ergebnis über die Ausgabe aus.

Zeile 90 hält das Programm an.

8.6.1.5. Hard- und Software

Beim Rechner — allgemeiner bei allen steuerbaren Geräten — sind zwei Gruppen zu unterscheiden:

- *Hardware* als jener Teil, der materiell-stofflich-energetisch das System ausmacht. Er wird durch stoffwandelnde Prozesse produziert. Insbesondere sind also hardware alle Baugruppen, die in Bild 8.6.2 dargestellt sind, darüber hinaus auch Leiterplatten, Bedienknöpfe, Tasten, Gehäuse usw.

- *Software* ist dagegen jener Teil, der hierzu nicht gehört, aber zum Betrieb des Rechners notwendig ist, also in erster Linie die Programme und Daten, welche im Rechner beim Betrieb in den Speichern abgelegt werden.

Nicht eindeutig kann damit ein ROM, welches ebenfalls solche Programme und Daten enthält, der Hard- oder Software zugeordnet werden. Hier ist ursprüngliche Software ein für alle Male fixiert und als materielle Struktur beständig vorhanden. Deswegen wird hier zuweilen auch von als Hardware fixierte Software gesprochen. Sofern dies firmenspezifisch erfolgt und so einen Rechner verbessert, wird dann auch von Firmware gesprochen.

Hard- und Software haben viele spezifische Eigenschaften, auf die hier des Umfangs wegen nicht detaillierter eingegangen werden kann. Für weitere Details sei daher auf [V 12, Nd. 1] verwiesen. Wichtig ist hier jedoch der Fakt, daß für eine bestimmte Leistung eines Rechners Hard- und Software weitgehend gegenseitig ersetzbar sind. Dies wurde für die unterste Ebene bereits am Beispiel des Volladders, insbesondere im Abschnitt 8.5.6.3., beispielhaft demonstriert, gilt aber viel allgemeiner. Eine Grenze ist die minimale Hardware bei maximaler Software. Dieser Fall liegt beim TURING-Automaten vor, der aber nur ein Gedankenmodell ist und dessen Bedeutung vor allem bei generellen mathematischen Problemen zur Beweisführung verwendet wird (s. auch 8.6.1.7.). *Anwendbare Minimalkonfigurationen* benötigen, wie die Geschichte sowohl bei den ersten Rechnern als auch bei Mikroprozessoren zeigt, ganz grob tausend Transistorfunktionen. Es ist auffallend, daß dieser Größenordnung eventuell eine allgemeine Bedeutung zukommt. Es gibt z.B. keine Lebewesen unter entsprechend vielen DNS-Sequenzen, die kleinsten Gehirne liegen ebenfalls in dieser Größe.

Das andere Extrem ist der Rechner, der *vollständig als Hardware* realisiert ist und daher keine Software mehr benötigt. Dies ist ein Spezialrechner für konkrete Aufgaben, ein Beispiel ist der nichtprogrammierbare Taschenrechner.

Mit dem Verhältnis der möglichen Software zur existierenden Hardware ist also folglich das Verhältnis von Spezialisierung zur möglichen (universellen) Flexibilität gegeben.

Die jeweils wirklich *vorhandene Software* bestimmt und begrenzt daher gegenüber der *potentiell möglichen Software* den Einsatz der gegebenen Hardware für konkrete Anwendungen. Ein von der Hardware her sehr leistungsfähiger Rechner kann bei nur vorhandener Software für Numerik, selbst wenn diese noch so gut ist, z.B. nicht für Textverarbeitung eingesetzt werden. Hierzu muß dann erst eine neue, passende Software erarbeitet werden.

Ferner kann eine Software, die für einen anderen Rechnertyp für dieses Problem erarbeitet wurde, auch meist nur mit großem Arbeitsaufwand auf den vorhandenen Rechner übernommen werden. Beides weist die große Bedeutung der Software für einen Rechner aus. Sie muß für das jeweilige Anwendungsgebiet und genau passend zu dem Rechner vorliegen bzw. dafür erarbeitet werden. Da die Anzahl der Rechnertypen und die zu bearbeitenden Problemstellungen ständig zugenommen haben, ist auch ständig neue und umfangreiche Software zu schaffen. Dieses Problem verschärft sich noch weiter dadurch, daß die Problemstellungen ständig komplexer werden und so zusätzlich stets umfangreichere Programme, also mehr Software, benötigen. Dies ist der Grund, weshalb heute oft von *Softwarekrise* gesprochen wird. Abschätzungen zeigen, daß insbesondere im Zusammenhang mit der Anwendungsverbreitung durch die Mikrorechner bei der jetzigen Softwaretechnologie um ca. 1990 die ganze Weltbevölkerung nur programmieren müßte. Dies wird bestimmt nicht der Fall sein, sondern vorher werden *neue Softwaretechnologien* entstehen. Ein erster Schritt hierzu, der bereits in den sechziger Jahren erfolgte und daher schon in dieser Trend-

betrachtung voll berücksichtigt wurde, ist die Anwendung genormter höherer Programmiersprachen, wie PL 1, COBOL, ALGOL usw. Alle Rechner, die ein Softwaresystem bezüglich dieser Sprachen besitzen, können dann mit allen Programmen betrieben werden, die in dieser Sprache geschrieben und getestet wurden.

8.6.1.6. Größeneinteilung von Rechnern

Heute gibt es eine Vielzahl von „Größen" bei den Rechnern. Sie überstreicht die Palette von Superrechner über Großrechner, mittlere EDVA, Kleinrechner, Minirechner, Mikrorechner, Personalcomputer bis zu einer Vielzahl von Größen bei den Taschenrechnern. Dabei wird die Größe sowohl bezüglich des Volumens, des Preises als auch der Leistungsfähigkeit (meist in recht komplexer Weise und nicht nur in Operationen je Sekunde oder/und Speicherkapazität) erfaßt. Im Grunde liegen die Extreme schon immer an den beiden Grenzen: Was ist groß und leistungsfähig noch ökonomisch herzustellen und was ist gerade noch ein brauchbarer Rechner. Dieses Spektrum ist natürlich erst im Laufe der Zeit entstanden. Die ersten Rechner waren zugleich die größten und einzigen, also auch kleinsten. Wie diese Entwicklung vorangegangen ist, mag ein Vergleich mit der zwischen 1944 und 1946 entwickelten ENIAC und einem heutigen typischen Mikrorechner mit sogar größerer Leistungsfähigkeit und

Tabelle 8.6.1. Kennwerte der ENIAC (1944—46) und grobe Richtwerte heutiger typischer Mikrorechner und leistungsfähiger programmierbarer Taschenrechner

Daten	ENIAC	Mikrorechner	Taschenrechner
Aktive Bauelemente	18000 Röhren	100000 Transistorfunktionen	5000 Transistorfunktionen
Masse	$3 \cdot 10^6$ g	2000 g	100 g
Leistungsaufnahme	175 kW	20 W	0,1 W
Taktfrequenz	100 kHz	5 MHz	2 MHz
Volumen	80 m³	2000 cm³	100 cm³
Dauer einer Multiplikation	2 ms	10 µs	10 ms
Speicherkapazität	120 Worte	64 Kbyte	2 Kbyte

einem leistungsfähigen Taschenrechner gemäß Tab. 8.6.1 zeigen. Die Leistungsfähigkeit der Großrechner hat dabei ständig beträchtlich zugenommen, wobei ihr Preis jedoch immer in etwa der gleichen Größenordnung, um eine Million Mark, lag. Ein Grund für ihre Entwicklung war das von GROSCH (IBM) *aufgestellte Gesetz*. Es sagt aus, daß die Kosten eines Rechners etwa mit der Quadratwurzel seiner Leistung ansteigen. Das bedeutet: Der größte Rechner ist der effektivste. Da für dieses Gesetz keine genauen Beweise, sondern nur wenige empirische Anhaltspunkte vorgelegt wurden, ist es eigentlich *nur* ein *Postulat*, welches aber über lange Zeit bei IBM und anderen Firmen die Entwicklungslinie bestimmte. Mitte der siebziger Jahre wurden gegenüber diesem „Gesetz" Fallstudien publiziert, die zeigten, daß z.B. rechenintensive Programme mit nicht zu hohem Speicherbedarf (ca. 100 Kbyte) optimal auf Kleinrechnern bearbeitet werden können. Ein weiteres *Argument für Großrechner* ist die Universalität. Hierbei ist aber nicht die Universalität im mathematischen Sinn gemeint, daß nämlich der Befehlsvorrat ausreichen muß, um alle berechenbaren Funktionen auszurechnen. Es ist vielmehr eine praktische Universalität gemeint. Der Rechner soll für alle möglichen Anwendungsfälle, wie Stapel- und Teilnehmerbetrieb, digitale Simulation und Textverarbeitung, wissenschaftliche und kommerzielle

Aufgaben geeignet sein. Er soll auf alle Peripheriegeräte, Programmiersprachen usw. ausbaubar sein.

Die *Kosten eines Rechners* bestimmen wesentlich zwei gegenläufige Einflüsse, nämlich die Rechenkomplexität (Universalität im obigen Sinn) und die verbesserte Anpassung an bestimmte Klassen von Aufgaben. Hinzu kommt, daß sich immer mehr zeigt, daß auch die Dezentralisierung von Rechentechnik Vorteile bringt. Dies wird besonders deutlich beim Einsatz von Prozeßrechnern. Sie werden in der Industrie zumeist für fünfzig bis einige hundert Einzelregelungen eingesetzt. Wenn so ein Rechner ausfällt, besteht große Gefahr für einen falschen Ablauf der entsprechenden Prozesse. Deshalb sind dort entweder ein weiterer Bereitschaftsrechner (Doppelrechenanlage) oder analoge Bereitschaftsregler (back-up) vorgesehen.

Die *Tendenz der Dezentralisierung* hat sich durch die Mikrorechner sehr verstärkt. Sie führte zum persönlichen Computer am Arbeitsplatz bis zum Heimcomputer. Hierbei gewinnt die unmittelbare Verfügbarkeit der Rechner, d. h. die sofortige Zugreifbarkeit, große Bedeutung. Wenn dann weiter beachtet wird, daß diese Rechner heute die Leistungsfähigkeit von Großrechnern vor ca. 10 Jahren besitzen, entsteht die Frage, ob der Großrechner heute noch aktuell ist. Infolge der ständig anwachsenden Komplexität der größten Probleme und der stärker werdenden Tendenz zur Rechnerverkopplung in Rechnernetzen wird er weiterhin im gewissen Umfang benötigt. Die Tendenzen der Rechentechnik können folglich in *vier Gruppen* erfaßt werden:

- einzelne Groß- und Größtrechner für große Probleme und zur Steuerung der Rechnernetze (um 1982 waren ca. 5000 weltweit im Einsatz);
- Leistungsfähige Kleinrechner für Einrichtungen und Institutionen zur kollektiven Nutzung und zur Steuerung industrieller Prozesse;
- Mikrorechner für persönliche Arbeitsplätze und als Heimcomputer (um 1982 ca. 5 Millionen im Einsatz);
- persönliche Taschenrechner (um 1982 ca. 10^8 Stück).

In allen Klassen wird die Leistungsfähigkeit weiter steigen und damit auch die Vielfalt der Typen und Preisklassen.

8.6.1.7. Grenzen der Rechentechnik

Es gibt *objektive Grenzen* eines konkreten Rechners. Sie sind u. a. durch die *Operationsgeschwindigkeit* und die *Speicherkapazität* gegeben. Sie werden erreicht, wenn z.B. die gleichzeitig zu bearbeitenden Daten und Programme nicht mehr in dem Speicher unterzubringen sind. Sie werden auch dann erreicht, wenn die Aufgabe so viele Operationen verlangt, daß die *zur Verfügung stehende Zeit nicht mehr ausreicht*, um sie durchzuführen. Dieser Fall tritt insbesondere bei *kombinatorischen Aufgaben* auf. Ein Beispiel ist die Schnittstellenoptimierung. Eine Schaltung mit z.B. 1000 Bauelementen mit im Mittel je drei Anschlüssen soll so auf vier Leiterplatten aufgeteilt werden, daß zwischen ihnen minimale Verbindungen notwendig sind. Hierbei wären in etwa 10^{30} mögliche Anordnungen zu vergleichen. Selbst wenn jeder Vergleich nur 10^{-6} Sekunden dauert, würden über 10^{11} Jahre benötigt, also eine Zeit, die viel länger ist als die Welt existiert.

Im Grunde können die bisher genannten Probleme als technische betrachtet werden. Es gibt aber weitaus einschneidendere, nämlich *prinzipielle Grenzen*. Sie werden verständlich, wenn genauer allgemein der Begriff Programm, Rechenvorschrift oder exakter Algorithmus behandelt wird, z.B. [V 12, Bd. 1].

Jeder *Algorithmus* muß so gegeben sein, daß nach jedem Befehl (Programmierschritt) exakt zu entscheiden ist, *welcher Befehl als nächster auszuführen* ist. In den linearen Programmteilen, wo der Befehlszähler jeweils um Eins weiterschaltet,

bestehen keine Probleme. Kritischer sind die Entscheidungsstellen, wo das zuvor berechnete Ergebnis, welches ja beim Entwurf des Algorithmus nicht bekannt ist, in den Fortgang der Rechnung eingeht. Hier muß zwischen zwei oder mehreren weiteren Wegen eindeutig entschieden werden. Zur Vereinfachung sei hier zunächst der Fall mit zwei Varianten der Fortsetzung gewählt. Er läßt sich auch so beschreiben, daß die Frage bezüglich der ersten Weiterrechnung gestellt wird und eindeutig vom Rechner selbst mit Ja oder Nein beantwortet werden muß. Bei mehreren Fortsetzungsmöglichkeiten müßten die einzelnen der Reihe nach abgeprüft werden und für eine müßte die Antwort genau Ja lauten.

Es genügt also, die *Ja-Nein-Entscheidung als zentrales Problem* herauszustellen. Bei genauer Analyse dieses Problems stellt sich nun heraus, daß es viele Fragen gibt, die prinzipiell nicht mit Ja oder Nein zu beantworten sind. Im Zentrum derartiger Betrachtungen stehen die schon bei den alten Griechen bekannten *Antinomien*.

Hier sei eine mehr anschauliche Erklärung dieses Problems gegeben:

Ein Richter ärgert sich über die komplizierten Ausführungen eines Angeklagten und sagt darauf zu ihm: „Herr Angeklagter, bitte antworten sie künftig auf alle Fragen kurz und bündig. Auf jede sinnvolle Frage kann schließlich mit Ja oder Nein geantwortet werden" Darauf der Angeklagte: „Herr Richter, darf ich Ihnen eine Frage stellen?" „Ja!" antwortete der Richter stolz. „Herr Richter, haben Sie aufgehört zu stehlen?"

Gewiß, hier liegt eine Unterstellung vor. Unserer Problemstellung näher liegt der Fall:

Ein Kompaniechef erteilt einem Friseur den Befehl, alle die zu rasieren, die sich nicht selbst rasieren. Es ist leicht zu entscheiden, wer sich nicht selbst rasiert und wen damit der Friseur zu rasieren hat. Doch was macht er mit sich selbst? Rasiert er sich, so rasiert er einen, der sich selbst rasiert. Das ist gegen den Befehl. Rasiert er sich nicht, so rasiert er einen nicht, der sich selbst nicht rasiert. Auch das ist gegen den Befehl.

Dieses Problem liegt viel tiefer als zunächst zu vermuten ist. Es konnte in vielen Varianten über eine sehr lange Zeit nicht gelöst werden. Erst GOEDEL wies 1932 mathematisch exakt nach, daß es in der Arithmetik *unentscheidbare Probleme* gibt. Heute teilt die Mathematik ihre Fachgebiete in entscheidbare und unentscheidbare ein. Wir müssen uns also damit abfinden, daß nicht alles per Automat berechenbar ist. Dies gilt ganz besonders für ethische und ästhetische Fragen. Damit ist aber nicht gesagt, daß solche Fragen nicht beantwortbar sind. Doch die Antwort ist dann anderer Art als berechenbar gemäß Ja-Nein-Entscheidungen. Dies wies ja GOEDEL bezüglich der Antinomien nach. Ethische und ästhetische Probleme sind dagegen typisch menschliche und nicht logische Fragen.

Die entscheidende Aussage bezüglich dessen, was *berechenbar* ist, gab 1936 TURING mit seinem Modell, dem TURING-*Automaten*. Er besitzt unendliche Speicherkapazität (beidseitig unendliches Speicherband) und ist besonders einfach aufgebaut. Zeitfragen gehen bei ihm nicht ein. Es besteht nur die Frage, *ob der Automat nach dem Start irgendwann anhält und das Ergebnis vorliegt*. Wenn dies beweisbar ist, ist das Problem berechenbar. Nun sollte man annehmen, daß es einen TURING-Automaten mit Programm gibt, der alle Programme, also auch sein eigenes Programm, prüfen kann, ob sie anhalten werden. Dies ist wieder genau eine im GOEDELschen Sinne unentscheidbare Frage. Solchen Algorithmus (Automaten) kann es also nicht geben.

Dieser Einschub soll nur davor warnen, alle Probleme als berechenbar anzusehen. Dennoch gibt es viele Aufgaben, die auf Rechnern gelöst werden. Und ihre Anzahl und Komplexität wächst ständig. Diese *Dialektik zwischen* dem nach GOEDEL algorithmisch nicht Entscheidbaren und dem nach TURING, d. h. mit Algorithmen Berechenbaren ist nicht ganz einfach. Sie ist aber für den Rechner wichtig.

Ein Problem der Rechentechnik besteht darin, daß es bis heute keine eigentliche Theorie gibt. Dies sei wie folgt gemeint: Bei Wärmekraftmaschinen gibt der CARNOTsche Kreisprozeß die Grenze des theoretisch maximal möglichen Wirkungsgrades an. In der Nachrichtenübertragung leistet die SHANNONsche Informationstheorie ähnliches (vgl. 9.6.). Für einen Algorithmus bzw. für ein Problem, das algorithmisch zu lösen ist, gibt es nichts derartiges. So, wie für die Hardware eines Adders (8.5.6.3.) viele Lösungen existieren, gibt es für ein Problem bei jedem konkreten Rechner viele Algorithmen, und nur durch den unmittelbaren Vergleich ist entscheidbar, welcher leistungsfähiger ist. Es gibt nicht einmal Abschätzungen darüber, wie umfangreich z.B. in bezug auf Speicherplätze, Befehle und Zeit ein minimaler Algorithmus sein könnte. Dies wird noch im Abschnitt zur Software (8.8.) etwas ausführlicher behandelt. Trotz vieler Kenntnisse in der Automatentheorie und zur Programmierung *fehlt also insgesamt eine allgemeine Theorie der Rechentechnik*, ja selbst Ansätze hierzu sind kaum vorhanden. Dies könnte ein Grund für die bereits im Abschnitt 8.6.1.5. behandelte Softwarekrise sein.

8.6.1.8. Zur Geschichte der Rechentechnik

Die Geschichte der Rechentechnik ist bei allgemeiner Auffassung und Berücksichtigung der Entstehung ihrer Grundlagen bis weit vor die Zeitrechnung zurückzuverfolgen. Einen Überblick zu wichtigen Entwicklungen gibt Tab. 8.6.2.

In der Frühzeit ging es vor allem darum, überhaupt rechnen zu können und hierbei sowohl Rechenmethoden als auch einfache Rechenhilfen wie Tabellen, Rechenbretter usw. zu entwickeln. So wurde das Umgehen mit Zahlen zum Allgemeingut des Menschen. Erst als die Rechenmethoden hinreichend durchsichtig waren, konnten technische Lösungen mit festen und später veränderbaren Programmen erdacht und noch später technisch verwirklicht werden.

1832 entwarf in diesem Sinne CHARLES BABBAGE einen mechanischen Automaten, der genau all jene Baugruppen enthielt, die heute für einen programmgesteuerten Automaten als wesentlich angesehen werden. Er wurde aber nie gebaut. Um 1940 erreicht die Elektronik einen solchen Stand, daß erste Rechner auf Relais- und später auf Röhrenbasis gebaut werden können. Die Pioniere sind hier ZUSE und AIKEN. Von da ab geht die Entwicklung rasant voran, und heute werden allgemein vier technische Generationen von Rechnern unterschieden. Ihre wichtigsten Eigenschaften weist Tab. 8.6.3 aus. Seit etwa 1982 wird mit großem Aufwand in Japan eine fünfte Rechnergeneration vorbereitet. Es bleibt abzuwarten, wie dieses Konzept sich durchsetzt und wie andere Staaten darauf antworten. Auf alle Fälle scheint danach die fünfte Generation nicht mehr durch neue Bauelemente bzw. periphere Geräte, sondern durch neue Leistungsqualitäten bestimmt zu sein. Einen wahrscheinlich ebenso tiefgreifenden Einschnitt brachte die Entwicklung des Mikroprozessors (und z. T. auch des Taschenrechners) um 1970.

Diese geschichtliche Entwicklung anhand von Zeitdaten ist in Tab. 8.6.3. enthalten. Sie wird aber noch im Abschnitt 8.6. detaillierter beschrieben, weil sie dann besser verständlich ist.

Eine interessante Entwicklung betrifft noch die strukturelle Verkopplung der Teilgruppen eines Rechners, auch Rechnerarchitektur genannt. Sie ist in vier Stufen in Bild 8.6.4 zusammengestellt. Dabei ist zwischen Datenaustausch (Doppellinien) und Befehlsübermittlung (einfache Linien) unterschieden.

Im Teilbild a) ist der typische Automat mit festem Programm gezeigt. Ein Schreib-Lese-Speicher ist nur für Daten vorhanden. Er entspricht in seinem Aufbau etwa den Rechnern der nullten Generation. Dieses Konzept hat aber wieder bei dem Taschenrechner Bedeutung erlangt.

In der zweiten und dritten Generation herrschte die Konzeption von J. v. NEUMANN vor, bei der Programm und Daten in einem gemeinsamen Speicher abgelegt werden. Hierdurch wird eine wesentlich höhere Universalität erreicht.

Die Weiterentwicklung vermehrt dann die Verkopplungen zwischen den einzelnen Struktureinheiten. Dabei ist auch wichtig, daß der Rechner bezüglich der Eingabe und Ausgabe immer auf Geräte zurückgreifen muß. Eine neue Qualität stellt dann die Entwicklung der Mikroprogrammierung dar, bei der immer wiederkehrende Programmteile getrennt gespeichert werden und besonders schnell verfügbar gehalten werden. Obwohl dieses Konzept von WILKES bereits 1951

Tabelle 8.6.2. Zur Geschichte der Rechentechnik

vor unserer Zeitrechnung	
5000	Mensch „rechnet" mit den Fingern bzw. verwendet Steine oder ähnliches
4000	Erstes heute bekanntes Zahlensystem
1700	Rechenbuch in Ägypten
1100	Römer besitzen den Abakus (Rechenbrett)
1000	Inder kennen Zahlen bis etwa 10^5
300	EUKLID: Elemente der Geometrie
200	Werk „Arithmetik" in China
unserer Zeitrechnung	
500	Inder entwickeln aerabisches Zahlensystem mit Stellenwert
800	Maya kennen den dezimalen Stellenwert
825	BEN MUZA AL-CHAVARIZINI „Algebra"
1000	Null kommt nach Europa
1210	Arabische Zahlenschreibweise in Europa
1518	Erstes Rechenbuch von ADAM RIES
1600	Rechenkästen in Rußland in Gebrauch
1614	Lord NAPIER gibt erste Logarithmentafeln heraus
1620	Logarithmische Skalen als Vorläufer
1620	BRIGGSsche Logarithmen; Rechenschieber von E. GUNTER
1623	W. SCHICHARD baut für KEPLER Rechenuhr mit Spaltenübertragung
1632	Rechenschieber von W. OUGHTRED (auch PARTRIDGE 1650 zugeschrieben)
1641	B. PASCAL baut mit 19 Jahren seinem Vater sechsstellige Addiermaschine
1673	G. W. LEIBNIZ stellt seinen Vierspezies-Rechner vor
1703	G. W. LEIBNIZ gibt Schrift „Arithmetique binaire" heraus
1808	J. M. JACQUARD benutzt „Lochkarten" zur Steuerung von Webstühlen
1833	CH. BABBAGE entwickelt programmgesteuerte Rechenmaschine
1841	CH. WHEATSTONE nenutzt Lochstreifen
1855	Begründung der BOOLEschen Algebra
1890	H. HOLLERITH entwickelt Tabellier- und Sortiermaschinen mit Lochkarten
1895	SCHÄFFLER erhält Patent auf das Programmieren mit Telefonschnüren
1919	Der ECCLES-JORDAN-Trigger wird patentiert
1932	GÖDEL beweist Unentscheidbarkeit der Arithmetik
1933	G. TAUSCHEK (Wien) Patent für Trommelspeicher
1936	A. M. TURING schafft Modell für die Berechenbarkeit
1938	Vorführung der Z1 von K. ZUSE (Beginn 1936)
1938	SHANNON publiziert zur Schaltalgebra
1941	Relaisrechner Z3 von K. ZUSE fertiggestellt
1943	DIRKS entwickelt Magnettrommelspeicher: 1955 in IBM 704
1946	J. v. NEUMANN entwickelt Prinzip der speicherprogrammierten Rechenmaschine
1946	Elektronenröhrenrechner ENIAC und Z22
1947	H. BILLING entwickelt Trommelspeicher für Rechner
1947	F. W. VIENE Patent für Kernspeicher
1948	Rechteckkerne bekannt
1948	Erste Theorie des Programmierens, Klassifizierung von Adressenmanipulationen, Programmschleifen
1948	N. WIENER definiert Kybernetik
1950	Magnetbandgeräte für Recheneinsatz: 1951 in UNIVAC
1951	M. V. WILKES: Konzept für Mikroprogramm
1951	FORRESTER publiziert über Ferritkernspeicher
1953	Erste Ferritkernspeicher im Einsatz
1954	Erste Paralleldrucker
1954	Erste Sprachübersetzung russisch — deutsch
1956	Festplattenspeicher
1957	FORTRAN-Compiler
1957	Betriebsprogramme etwa 10^4 Zeilen

Fortsetzung Tabelle 8.6.2.

1957	Teilnehmerrechensystem
1957	Rechner mit Transistoren (2. Generation)
1958	Betriebssystem für Mehrprogrammbetrieb (breiter Einsatz erst ab 1965)
1958	ALGOL (1963 revidiert)
1958	Konzept eines ausfallsicheren Mehrrechnersystems
1958	Erste Belegleser und Mosaikdrucker
1959	COBOL
1960	McCarthy entwickelt LISP
1960	Erste Großrechner mit Mikroprogramm
1961	Virtuelle Speicherung bei Atlas
1961	Elektronischer Tischrechner
1961	Wechselplattenspieler
1964	PL/I
1964	Verwendung von integrierten Schaltkreisen (3. Generation)
1965	IBM 360 mit Halbleiterspeicher
1965	BASIC, entwickelt von J. G. Kemeny und T. E. Kurtz, Dartmouth Coll.
1966	PDP8 Prozeßrechner
1967	BESM 6
1968	Teilnehmerrechensystem
1968	Mikroprogrammierbare Rechner
1969	Für bemannten Mondflug von Apollo 11 sind 600 Rechner im Einsatz
1970	Betriebsprogramme mit 10^7 Zeilen ($6 \cdot 10^6$ IBM 370)
1970	Erster Taschenrechner von Canon auf dem Markt
1970/71	4-bit-Mikrorechnersatz aus 4 Chips mit der CPU 4004 (Intel)
1972	Wissenschaftlich-technischer Taschenrechner HP 35, 5 Chips, $2 \cdot 10^4$ Transistorfunktionen
1972	Vielfachprozessorrechner ILLIAC 4
1973	Entstehung der Rechnerbusstrukturen
1973	Winchester-Plattenspeicher
1973	PASCAL entwickelt
1973	INTEL 8080 entsteht, über viele Jahre Industriestandard
1974	CP/M Betriebssystem für Mikrorechner von G. A. Kindall
1974	Floppy-disc-Speicher
1974	Programmierbarer Taschenrechner HP 65
1975	20 Hersteller und 40 Typen Mikroprozessoren
1975	Mikroprozessor mit nur 5 V Betriebsspannung (Motoriola M 6800)
1975	Beginn der Produktion der persönlichen Rechner
1976	6502 von MOS-Technologie als Prozessor mit größter produzierter Stückzahl
1976	Z 80 von Zilog
1976	Erster 16-bit-Mikroprozessor (erlangte keine Bedeutung)
1978	Über 100 Mikroprozessortypen
1978	Minimal-BASIC definiert
1979	16-bit-Typen Z 8000 (Zilog) und 8086 (Intel)
1979	Superrechner Cray I liefert täglich Wetterkarten mittels $6 \cdot 10^9$ Operationen
1981	UNIX Betriebssystem für Mikrorechner
1981	32-bit-Mikroprozessoren iAPX32 (Intel) und CPU von Hewlett u. Packard
1982	Über 5 Millionen persönliche Rechner im Einsatz

ausgesprochen wurde, kam es erstmalig 1960 bei der IBM 360 zur Anwendung. Es wird generell erst bei der dritten Rechnergeneration wirksam, denn zur Erzeugung, Speicherung und schnellen Auswechslung von Mikroprogrammen sind LSI-Schaltkreise besonders geeignet.

Anfang der 70er Jahre bildete sich infolge der immer zahlreicher werdenden Verkopplungen zwischen den Teilen der Rechenanlage die Busstruktur (vgl. 8.5.7.) heraus. Sie zwingt die vielfältigen Informationsströme auf einen einheitlichen Weg und schafft dadurch eine große Über-

Tabelle 8.6.3. Mögliche Generationseinteilung bei Großrechnern, vor allem auf Basis der verwendeten Bauelemente

Generation	Ungefährer Zeitraum	Typische Bauelemente	Sonstige neue Eigenschaften und Techniken
0	bis 1947	Relais	Vorstufe, festes Programm
1	1946—1958	Elektronenröhren	• sequentielle Programmierung in Maschinensprache • Programm im Speicher • Trommelspeicher
2	1958—1965	Transistoren Dioden	• symbolische Sprachen (Assembler) • Unterscheidung von wissenschaftlich-technischen und kommerziell-ökonomischen Rechnern • Kernspeicher
3	ab 1965	Hybrid- oder Dünnschichtschaltungen MSI-Schaltungen	• Betriebssysteme • Problemsprache: ALGOL, FORTRAN, COBOL • Familienkonzepte • Time-sharing • Datenfernverarbeitung • Mikroprogramm- und Steuerspeicher • Bildschirmgeräte • Magnetbandgeräte • Plotter • Belegleser
4	ab 1975	LSI-VLSI-Schaltungen	• Modularstruktur • Parallele Datenverarbeitung • Busse • komplexe Speicherhierarchie • Rechnernetze

sichtlichkeit. Die Struktur eines Automaten ist dadurch erstmalig nicht mehr fest gegeben, sondern infolge der Adressen der einzelnen Geräte selbst durch Information steuerbar. Inzwischen hat die Busstruktur eine generelle Bedeutung (weit über die Rechentechnik hinaus) erlangt. Sie wird deshalb im Abschnitt 8.7.

noch ausführlicher behandelt.

Insgesamt hat sich nach diesen Betrachtungen folgende Entwicklung bezüglich der Struktur von Rechenanlagen ergeben, wobei die in Klammern gesetzten Buchstaben auf das Teilbild im Bild 8.6.4 verweisen:

— informatorisch gespeichertes Verhalten (a)
— informatorisch gespeicherte Programme (b)
— informatorisch gespeicherte Funktion (c)
— informatorisch gespeicherte Struktur (d).

Es gibt Andeutungen dafür, daß die Entwicklung zur informatorisch gespeicherten Konfiguration weitergeht. Für den Rechenvorgang werden dann nur jene Teile einbezogen, die er benötigt; die anderen können für getrennte Aufgaben parallel verwendet werden.

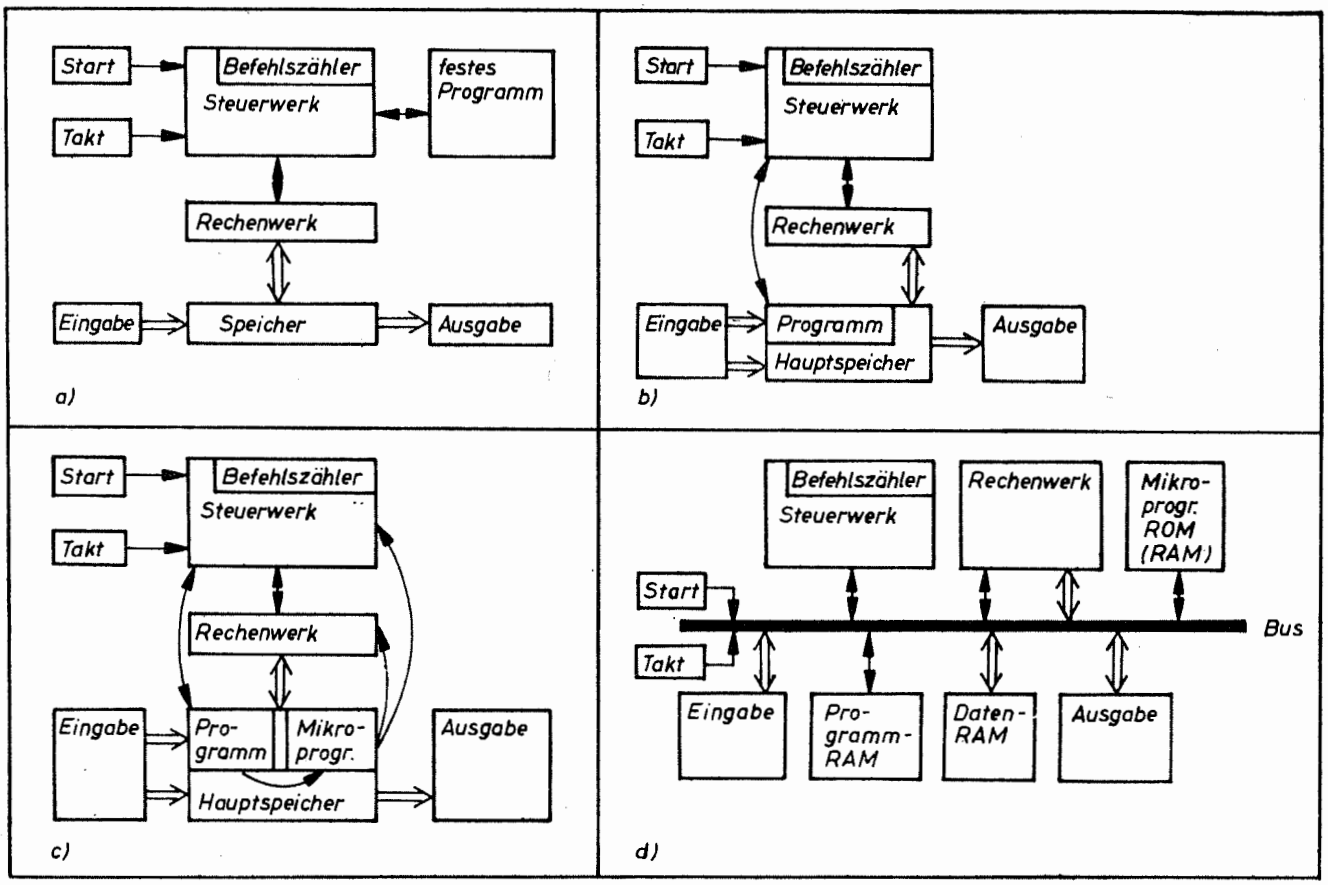


Bild 8.6.4. Zur Struktur von Rechnern. Doppellinien für Daten, Einfachlinien für Befehle

a) Rechner mit festem, wenn auch eventuell hardwaremäßig auswechselbarem Programm

b) J.-v.-NEUMANN-Automat, bei dem das Programm im Hauptspeicher abgelegt wird und daher schnell und leicht wechselbar ist

c) Rechner mit Mikroprogrammierung, die oft wiederkehrende Programmteile schnell abrufbar bereithält und eventuell ebenfalls schnell auswechselbar ist (Konzept von M. V. WILKES)

d) Rechner mit Busstruktur, bei der Informationsaustausch über Adressierung der Geräte erfolgt und so eine hohe Flexibilität und Übersichtlichkeit erreicht wird

8.6.2. Mikroprozessoren

8.6.2.1. Zusammenwirken mit anderen Baugruppen

Die Zusammenschaltung einer CPU (Mikroprozessor mit weiteren wichtigen Bausteinen) zeigt Bild 8.6.5a). Von der CPU gehen drei Busse aus:

Mit dem *Adreßbus* spricht die CPU die einzelnen Speicherplätze bzw. Ein- und Ausgabetore an. Die Unterscheidung zwischen beiden erfolgt durch ein spezielles Signal auf dem Steuerbus (siehe dort). Die Speicher können Daten übernehmen bzw. abgeben. Die Tore stellen die Verbindung zu peripheren Geräten, wie z.B. Tastatur, Drucker, Bildschirm usw., her. Über sie können also sowohl Daten für die CPU eingegeben als auch von der CPU nach außen weitergeleitet werden. Diese Unterscheidung über die Richtung zum und vom Speicher bzw. Tor wird ebenfalls durch Signale des Steuerbusses festgelegt.

Die Anzahl n der Adreßleitungen bestimmen den Adreßraum gemäß 2^n. Mit 8 bit Breite sind also 256 Speicherplätze bzw. In-out-parts adressierbar, mit 16 bit Breite sind es 65536 (64 K), mit 20 bit Breite 1048576 (1 M).

Der *Datenbus* ist im Gegensatz zum Adreßbus in beiden Richtungen von und zur CPU beschreibbar. Dies geht schon aus der Beschreibung des Adreßbusses hervor. Gehen Daten von der CPU zum Bus, wird von Schreiben (write: WR) gesprochen. Die

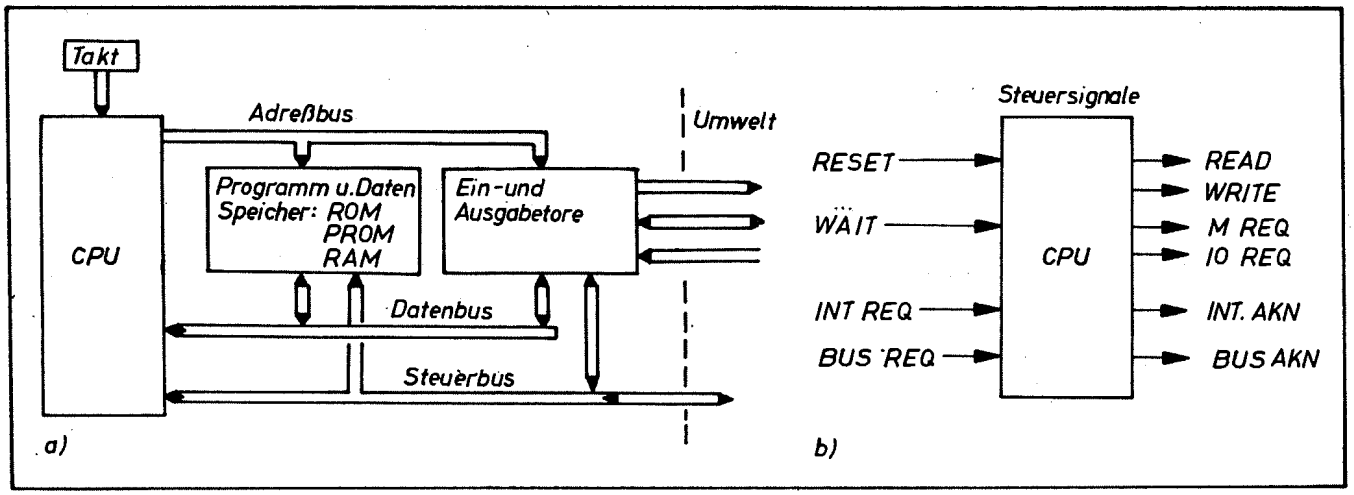


Bild 8.6.5. a) Zusammenschaltung eines Mikroprozessors (CPU) mit anderen Bausteinen und b) wesentliche Signale des Steuerbusses

Datenrichtung zur CPU heißt Lesen (read: RD). Die Datenbusbreite ist ein typisches Kennzeichen der Mikrorechner. Hiernach werden u. a. 1-, 4-, 8-, 16- und 32-bit-Prozessoren unterschieden. Die größte Anwendungsbreite besitzen die 8-bit-Prozessoren. Die 16-bit-Prozessoren erlangen zur Zeit für komplexere Aufgaben an Bedeutung. Die 32-bit-Prozessoren gibt es erst seit kurzer Zeit. Sie kommen nur für anspruchsvolle Anwendungen in Betracht. Die 4-bit-Prozessoren haben fast nur noch geschichtliche Bedeutung. Eine gewisse Anwendung besitzen sie zusammen mit den 1-bit-Prozessoren bei den Taschenrechnern. Der 1-bit-Prozessor hat zuweilen darüber hinaus Vorteile bei einigen Steueraufgaben. Viele Aussagen sprechen dafür, daß auch künftig der 8-bit-Prozessor ca. 60% des Bedarfs abdecken wird, dem 16-bit-Prozessor werden künftig etwa 30% zugesprochen. Die restlichen 10% bleiben maximal für die anderen Typen.

Der *Steuerbus* (Bild 8.6.5b)) (Control-bus) hat einige Leitungen, die zur CPU führen und sie beeinflussen. Andere Leitungen zeigen Zustände der CPU an bzw. beeinflussen die anderen Bausteine. Mit ihm wird das Zusammenarbeiten von CPU, Speichern, Toren und peripheren Signalen erreicht. Der Steuerbus ist bei den einzelnen Prozessortypen unterschiedlich ausgelegt, und die Leitungen sind z. T. unterschiedlich bezeichnet. Hier werden bevorzugt die Bezeichnungen des U 880/Z 80 verwendet. In der Regel enthält er um zehn Leitungen. Sie sind durch die etwa gleichen Aufgabenklassen aller CPU gegeben. Das schon behandelte Zusammenwirken mit den Speichern und den Toren geschieht über vier Signale gemäß der folgenden Matrix:

	MREQ	IOREQ
RD	Speicher werden gelesen	Daten vom Tor zur CPU
WR	Einschreiben in Speicher	CPU gibt Daten ans Tor

Die Abkürzungen bedeuten:

MREQ = memory request: Speicheraufruf
IOREQ = in-out-request: Toraufruf
RD = read: Lesen (Daten an CPU)
WR = write: Schreiben (Daten von der CPU).

Wenn ein Speicher oder ein peripheres Gerät langsamer als der Takt der CPU ist, kann das Signal **WAIT** die CPU auffordern zu warten.

Ein wichtiges Signal für den Start der CPU ist **RESET**, mit ihm wird der Programmzähler auf die Startadresse Ø gesetzt.

Bei den verschiedenen Anwendungen muß der Rechner bei seiner Arbeit auf unterschiedliche Weise unterbrochen werden, weil äußere Anlässe wichtiger sind. Hierbei sind im allgemeinen zwei Steuerbefehle vorgesehen:

Im ersten Fall wird das Programm unterbrochen, weil ein peripheres Gerät bedient, z.B. ein Meßgerät abgelesen, oder auf eine aufgetretene Veränderung in der Umgebung des Rechners reagiert werden muß. Dieses erfolgt durch Beeinflussung von **INT** = interrupt (Unterbrechung, eigentlich IRQ = interrupt requested = Unterbrechung gewünscht). Hierbei handelt es sich um eine Forderung an den Rechner: Unterbrich deine Arbeit, ich brauche dich! Um dies zu tun, muß der Rechner einiges sicherstellen, damit später die Arbeit fortgesetzt werden kann, z.B. Daten retten, Programmstelle notieren usw. Wenn dies erfolgt ist, meldet er sich mit
INAK = interrupt acknowledge (Unterbrechung vorgenommen). Damit kann dann das periphere Gerät von der CPU bedient werden. Dieses Signal existiert beim Z 80 nicht. Es wird aus
M1 = Maschinenzyklus 1 (holt Adresse = Fetch-Zyklus) und IOREQ gemäß $\overline{M1} \wedge$ IOREQ gebildet. Beim Z 80 gibt es außerdem noch eine zweite Unterbrechungsmöglichkeit:
NMI = nicht maskiertes Interrupt.

In einigen Fällen kann es vorkommen, daß ein peripheres Gerät direkt auf Speicher zugreifen möchte. Dann bittet es die CPU, den Adreß- und Datenbus freizumachen, und gibt das Signal
BUSREQ = bus requested. Wenn die CPU ihre Treiber in den Tristate-Zustand gebracht hat, antwortet sie
BUSAK = bus acknowledge (Busleitungen sind frei).
Dieser Befehlsablauf ist z.B. dann wichtig, wenn ein DMA = direct memory access (direkter Speicherzugriff) erfolgen soll.
Schließlich gibt die CPU Z 80 noch ein weiteres Signal
RFSH = refresh periodisch im Zusammenhang mit Adressen ab, um für dynamische Speicher den Refresh-Vorgang automatisch zu realisieren.

8.6.2.2. Innerer Aufbau

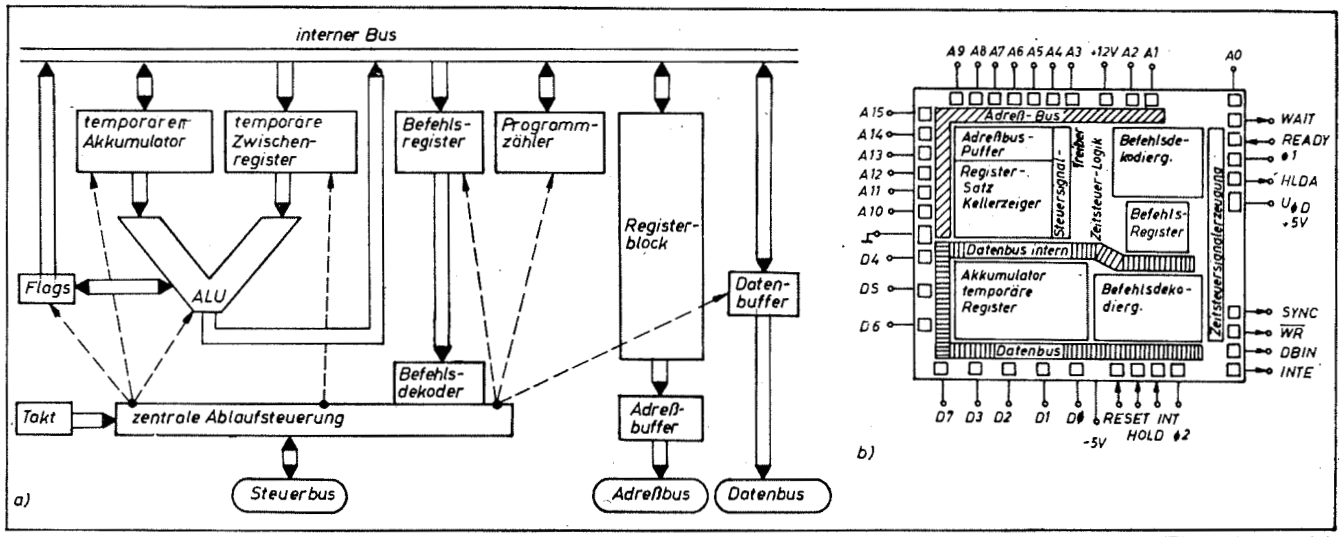
Bild 8.6.6a) zeigt das typische Blockschaltbild eines typischen Mikroprozessors. Es weicht nur wenig von der Struktur des Rechners von Bild 8.6.2 ab. Natürlich ist die Zentraleinheit beim Großrechner umfangreicher als beim Mikroprozessor ausgestattet. Die Unterschiede in den Bezeichnungen sind z. T. nur historisch bedingt.

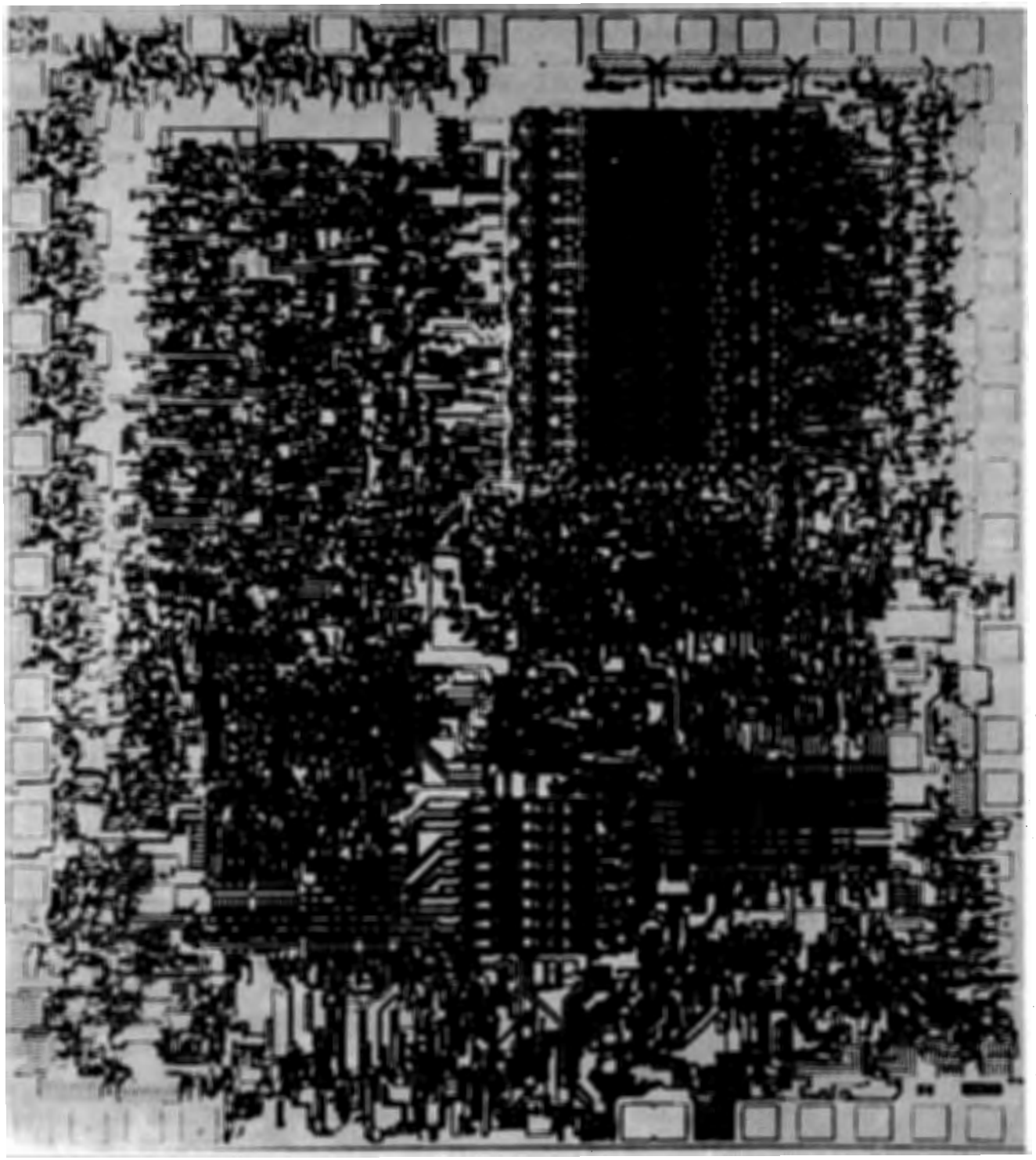
Beim *Rechenwerk* aus ALU und temporären Registern werden beim Mikrorechner immer Flags hervorgehoben. Das sind Flipflops, die eine Statusinformation der ALU speichern, z.B. Vorzeichen (sign), Überlauf (carry) usw.

Beim *Steuerwerk* ist im Mikroprozessor meist die Einteilung mit Befehlszähler (Programmzähler), Befehlsregister, Befehlsdekoder und zentraler Ablaufsteuerung (die auch Rechensteuerung des Rechenwerks mit beinhaltet) etwas anders gewählt.

Bild 8.6.6. Innerer Aufbau eines Mikroprozessors ▶

a) Blockschaltbild
b) Grobtopologie des Typs Intel 8080 (K 580/K 80)
c) Chipfoto des Typs Intel 8080 (rechte Seite in b) entspricht unten in c))


interner Bus
temporärer Akkumulator
temporäre Zwischenregister
Befehlsregister
Programmzähler
Registerblock
Datenbuffer
Flags
ALU
Befehlsdekoder
Takt
zentrale Ablaufsteuerung
Adreßbuffer
Steuerbus
Adreßbus
Datenbus
a)
A9 A8 A7 A6 A5 A4 A3 +12V A2 A1
A0
A15 A14 A13 A12 A11 A10
D4 D5 D6
Adreß-Bus
Adreßbus-Puffer
Register-Satz Kellerzeiger
Steuersignal-treiber
Zeitsteuer-Logik
Befehlsdekodierg.
Befehls-Register
Datenbus intern
Akkumulator temporäre Register
Befehlsdekodierg.
Datenbus
Zeitsteuersignalerzeugung
WAIT
READY
Φ1
HLDA
U Φ D +5V
SYNC
WR
DBIN
INTE
D7 D3 D2 D1 D0
-5V
RESET INT HOLD Φ2
b)



Das *Speicherwerk* ist im Mikroprozessor selbst auf einen kleinen Registersatz reduziert. Dafür ist aber über Adreß- und Datenbus unter Einbeziehung des Steuerbusses ein umfangreicher getrennter Speicherraum ansprechbar.

Die *Ein- und Ausgabe* erfolgt ebenfalls mittels der drei Busse.

Es gibt auch Großrechner, deren Strukturen der des Mikrorechners ähneln. Der hier betont hervorgehobene Unterschied muß also nicht immer bestehen.

Die verschiedenen *Typen von Mikroprozessoren unterscheiden sich* bei allgemeiner Gültigkeit von Bild 8.6.6a) vor allem in folgenden Punkten:

- Wortbreite der Daten
- Breite des Adreßbusses (Adressierungsbereich)
- Umfang des Registersatzes (Anzahl und Wortbreite)
- Leistungsfähigkeit der ALU, einschließlich Flags
- Befehlssatz der zentralen Ablaufsteuerung
- Signale des Steuerbusses.

Andere Unterscheidungsmerkmale betreffen

- die Technologie (*p*-MOS, *n*-MOS, ECL, I^2L, CMOS usw.)
- die Versorgungsspannungen (eine oder mehrere)
- die Verarbeitungsgeschwindigkeit (Taktfrequenz).

Eine weitere Untersetzung und Beschreibung dieser Unterschiede erfolgt in späteren Abschnitten.

In Bild 8.6.6b) ist die Grobtopologie eines speziellen Mikroprozessors gezeigt. Es handelt sich um den Typ Intel 8080. Er galt lange Zeit als Industriestandard und zählt zu den wenigen Prozessoren, von denen auch Chipfotografien (Bild 8.6.6c)) publiziert wurden. Er wurde hier auch deshalb ausgewählt, da er der Vorläufer des Z 80 (U880) ist, der in diesem Buch bevorzugt behandelt wird.

Tab. 8.6.4 gibt einen Überblick zur *Anzahl von Transistoren*, die sich auf einem Mikroprozessoren-Chip befinden, und über die für den Chip benötigte Fläche. Daraus geht deutlich hervor, wie ständig mit der Weiterentwicklung sowohl die Transistorzahl als auch die Bauelementedichte auf dem Chip zunimmt.

Die Chipfotografie vom Intel 8080 (Bild 8.6.6c)) zeigt wenig Regelmäßigkeit in der Feinstruktur. Auch hier ist durch den rechnergestützten Entwurf ein beachtlicher Fortschritt erreicht worden. Moderne Mikroprozessoren, soweit sie überhaupt publi-

Tabelle 8.6.4. Vergleich einiger Mikrorechner bezüglich Transistoranzahl und Chipfläche in mm^2

Typ	Trans.	mm^2
4004	2300	20
8008	3500	20
8080	4800	22
Z 80	8200	25
Z 8000	17500	39
8086	29000	33
MC 68000	70000	44
i APX 32	250000	
HP 32 bit	450000	
Zum Vergleich Taschenrechnerchip		
Mostek (1. Rechner)	2100	20
U 820/821	6000	25

ziert werden, zeigen erheblich deutlicher *Wiederholstrukturen*. Für die Regelmäßigkeit R wurde ein Quotient eingeführt, der die Zahl der Funktionen je Chip zur Zahl der unterschiedlich zu entwerfenden Funktionen berücksichtigt. Für einige Prozessoren gilt dabei:

Typ	8080	8085	9086	i APX 32
R	1,1	3,1	4,4	16

8.6.2.3. Gewinnung und Aufteilung des Befehlswortes: Befehlstypen

Die meisten Mikroprozessoren besitzen eine Datenwortbreite von 8 bit. Damit können $2^8 = 256$ Befehle kodiert werden. Dies erscheint auf den ersten Blick sehr viel, reicht aber meist nicht aus, denn es müssen z. B. bei 64 Kbyte Speicherbereich auch alle Speicherplätze einzeln ansprechbar sein. Deshalb ist es notwendig, ein Befehlswort aus mehreren Byte zusammenzusetzen. Hier bestehen mehrere Möglichkeiten: Entweder werden alle Befehlsworte gleich lang gemacht oder, wie beim Morsealphabet, wird die *Wortlänge nach der Häufigkeit des voraussichtlichen Gebrauchs* des Befehls bestimmt. Bei den Mikroprozessoren wird fast immer der zweite Weg gewählt. Er hat den Vorteil der höheren Geschwindigkeit häufiger Befehle und den Nachteil, daß der Mikro-

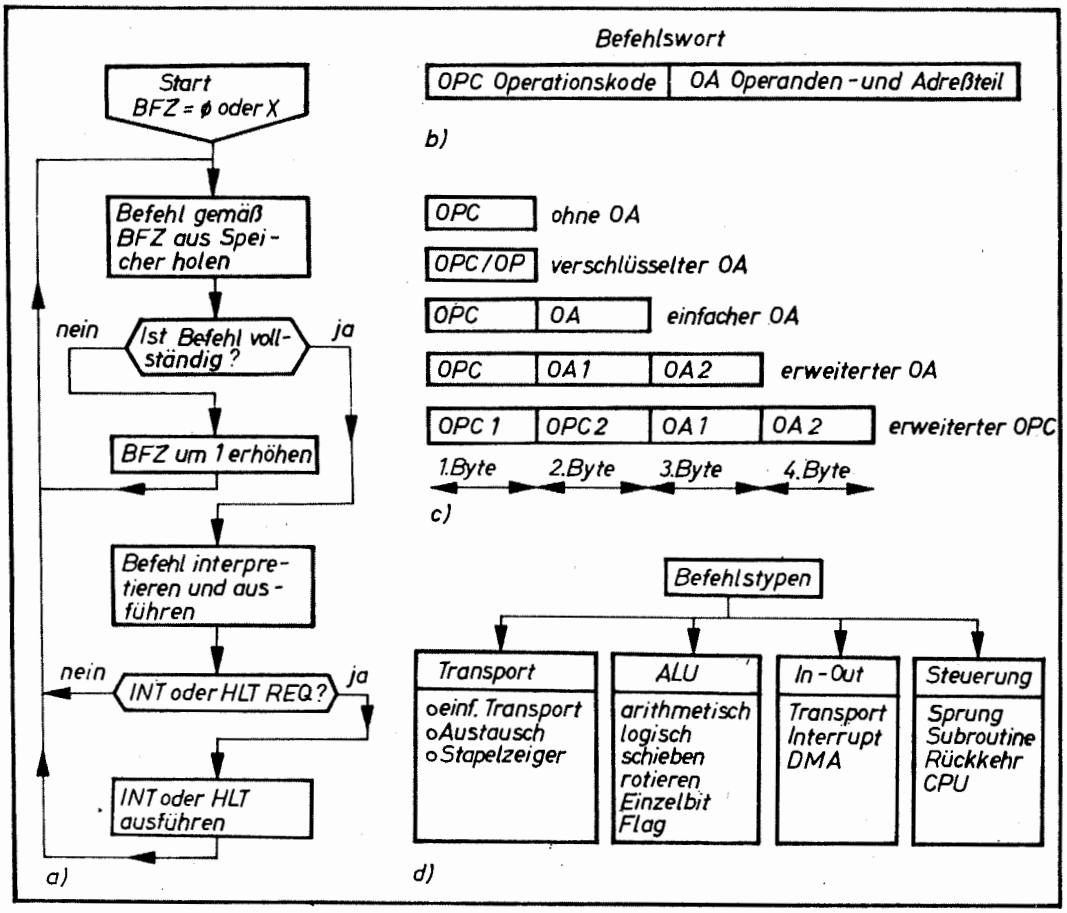


Bild 8.6.7. Struktur von Befehlen

a) Gewinnung des Befehlswortes (kurz: Befehl) beim Mikroprozessor

b) Einteilung des Befehlswortes in die zwei der Funktion nach unterschiedlichen Teile. Der Operationskode sagt, was getan werden soll (Befehlstyp: Teilbild d)). Der Operanden- bzw. Adreßteil sorgt dafür, daß die Daten direkt oder über verschiedene Adressierungsarten aus Speichern bereitstehen

c) Ein Befehlswort für Mikroprozessoren kann aus mehreren Byte unterschiedlich zusammengesetzt sein

d) Einteilung der Befehlstypen (In der Literatur gibt es auch andere Klassifikationen.)

prozessor entscheiden muß, ob der Befehl schon aus allen Byte zusammengesetzt ist. So entstehen die Möglichkeiten für das Befehlswort gemäß Bild 8.6.7b) und c).

Der *Operationskode* sagt in kodierter Form aus, *was getan werden soll*, also z.B. Inhalt eines Speichers in den Akkumulator holen, Zahl ausgeben an ein peripheres Gerät (vgl. auch Bild 8.6.2, Operationsteil). Er führte zu den verschiedenen Befehlstypen (Ende dieses Abschnitts und 8.6.2.4. usw.).

Der *Operanden- oder Adreßteil* muß Daten (als Zahlenwert, auch wenn er etwas anderes, z.B. Buchstaben, bedeutet) für die Operation herbeischaffen. Dies kann sehr unterschiedlich geschehen, z.B.:

- als Zahlenwert, der direkt angehängt ist;
- als Adresse einer Speicherzelle, in der Daten stehen;
- als Adresse für eine Speicherzelle, in der die Adresse des Speichers steht, der den Zahlenwert enthält.

Besonders der letzte Fall scheint unnötig kompliziert zu sein. Der praktische Umgang mit dem Rechner zeigt aber, daß gerade die Vielfalt der Beschaffungsoperationen von Daten entscheidend für die Leistungsfähigkeit eines Rechners ist. Aus diesem Grunde haben sich viele *Adressierungstypen* herausgebildet. Sie werden im Abschnitt 8.6.2.8. behandelt.

In Bild 8.6.7a) ist ein Ablaufplan dargestellt, wie die meisten Mikroprozessoren ihre Befehle gewinnen und abarbeiten: Beim Start wird der Befehlszähler BFZ auf Ø oder einen anderen Startwert X gesetzt. Dieser Wert wird als Adresse aufgefaßt und der Inhalt (das Byte) dieser Speicherzelle in das Befehlsregister geholt (Fetch-Zyklus). Dann wird geprüft, ob der Befehl vollständig ist. Bei negativem Ergebnis wird der BFZ um 1 erhöht und das Befehlswort um das auf dem nächsthöheren Speicherplatz stehende Byte ergänzt. Je nach der notwendigen Länge des Befehlswortes wiederholt sich dieser Schritt. Wenn das Befehlswort vollständig ist, wird der Befehl interpretiert und ausgeführt. Hierbei wird die Adresse des BFZ um 1 erhöht oder entsprechend dem Befehl auf einen neuen Wert gesetzt (mittels Schalter S 2 in Bild 8.6.2). Die Mehrzahl der Mikroprozessoren prüft dann, ob eine Anforderung für ein Halt oder Interrupt vorliegt. Ist dies nicht der Fall, wird entsprechend dem Stand des BFZ der neue Befehl ins Befehlsregister geholt. Anderenfalls wird die Abarbeitung der Befehlsfolge unterbrochen, der Stand des Befehlszählern notiert und erst zwischendurch Halt bzw. Interrupt ausgeführt. Mit dieser Methode kann ein Mikrorechner schnell auf Anforderungen von peripheren Geräten reagieren.

Die Einteilung der *Befehlstypen* wird in der Literatur recht unterschiedlich gehandhabt. Hier werden gemäß Bild 8.6.7d) viei Typen unterschieden. Sie betreffen *Teile des Rechners*:

- Transportbefehle vermitteln Informationen zwischen Registern und/oder Speicher.
- ALU-Befehle beziehen sich auf die ALU und sind damit arithmetischer oder logischer Natur.
- In-out-Befehle betreffen die Ports (Tore) und damit Beziehungen zur Umwelt.
- Steuerungsbefehle (Organisationsbefehle) betreffen die CPU selbst, also in erster Linie Befehlsregister, Befehlszähler und Steuerbus.

Diese Gruppen sollen im folgenden detaillierter behandelt werden. Dabei sollen einmal möglichst allgemeingültige Aussagen (also für alle Mikroprozessoren) und zugleich eine weitgehende Anpassung an die hier bevorzugt behandelte CPU Z 80 (U 880), insbesondere was mnemotechnische Bezeichnungen betrifft, erreicht werden. Die konkreten Z 80-Befehle werden im Abschnitt 8.6.3. aus den Betrachtungen dieses Abschnitts weiter detailliert.

In diesem Abschnitt werden noch nicht berücksichtigt die verschiedenen Adressierungsmöglichkeiten. Die Änderungen der einzelnen Befehlsgruppen behandelt der nächste Abschnitt. Von Ausnahmen abgesehen, werden der Übersichtlichkeit halber nur direkte Speicheradressen in den Beispielbildern verwendet. Anstelle dieser Speicheradressen können auch Register oder indirekte Adressen stehen oder z. T. sogar direkte Zahlenwerte. (Wenn jemand den Umgang mit programmierbaren Taschenrechnern gewohnt ist, kann es vorteilhaft sein, vorher den Abschnitt 8.9. zu lesen.)

8.6.2.4. Transportbefehle

Die *Transportbefehle* faßt Bild 8.6.8 zusammen. Bei ihnen muß eine Information, d. h. eine Zahl, Buchstabe usw., von einer Quelle zum Ziel transportiert werden. Dies kann auf mehrfache Weise beschrieben werden:

Information von Quelle zum Ziel
LD Ziel, Quelle
Ziel := Quelle (Kurzform: Ziel = Quelle)
Quelle → Ziel

Ziel und Quelle können dabei Register und/oder Speicher sein, Quelle kann auch eine direkt angegebene Zahl sein, z. B. LD A, 38 H, d. h.: Lade das Register A der CPU mit dem Zahlenwert 38 Hex = 56 D. Dagegen bedeutet LD A, 38: Lade Re-

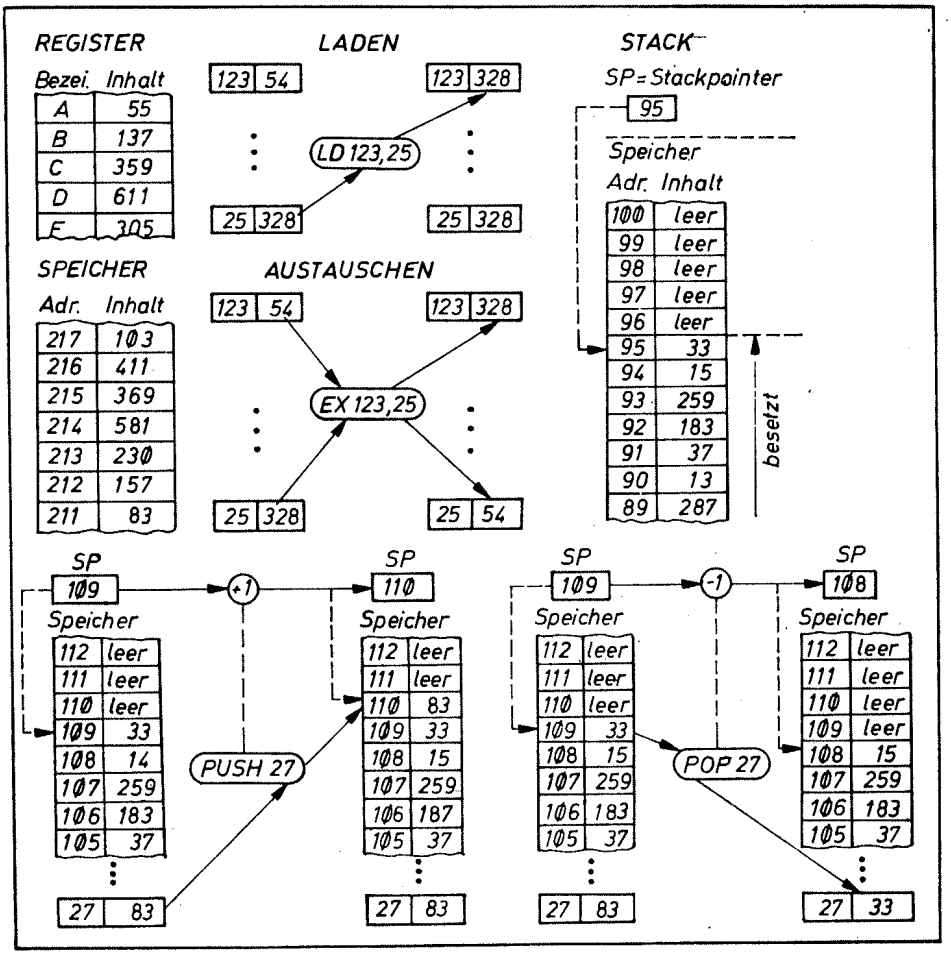


Bild 8.6.8. Zur Veranschaulichunt der Transportbefehle

Erklärt wird der Fall der direkten Adressierung, wobei die Register mit den Bezeichnungen A, B, C usw. genauso wie die Adressen, z. B. 25, 123 usw., behandelt werden. Die Darstellung erfolgt hier mit dezimalen Zahlen, in Wirklichkeit sind die Adressen und Inhalte binär

gister A mit dem Wert aus dem Register 38. Hieraus sind dann noch viele weitere Befehle abzuleiten.

Ein besonderer Ladebefehl ist der Austausch.

EX 123, 25

bedeutet, daß die Speicherinhalte der Speicherplätze 123 und 25 vertauscht werden. Für diesen Befehl kann kein Zahlenwert verwendet werden.

Noch komplizierter sind die Stackbefehle PUSH und POP (zuweilen auch PULL). Die Funktion eines Stacks wurde in Bild 8.3.26 eingeführt und ist vereinfacht noch einmal in Bild 8.6.8, rechts oben, erklärt. Das Stackregister funktioniert hier als Zeiger (SP = Stackpointer) und weist mit seinem Inhalt auf eine Adresse im Speicherblock, im Beispiel auf die Adresse 95, wo der Inhalt 33 D steht. Der Befehl

PUSH 27

bewirkt nun (links unten im Bild), daß der Inhalt des Speicherplatzes 27 (hier 83 D) auf den Speicherplatz übertragen wird, der um Eins höher als die Stackadresse ist (hier also 11Ø). Dieser Speicherplatz war zuvor leer. Anschließend wird auch der Inhalt des SP um Eins, auf 11Ø, erhöht. Mit dem PUSH-Befehl werden also Speicherinhalte auf dem Stack (Stapelregister) oben abgelegt. Der Befehl

POP 27

bewirkt die Umkehrung von PUSH. Hier wird ein Wert vom Stack heruntergenommen, in die Speicherzelle mit der Adresse 27 transportiert und der Stackpointer um Eins erniedrigt.

8.6.2.5. ALU-Befehle

Die ALU-Befehle faßt Bild 8.6.9 zusammen. Als rein

arithmetische Befehle

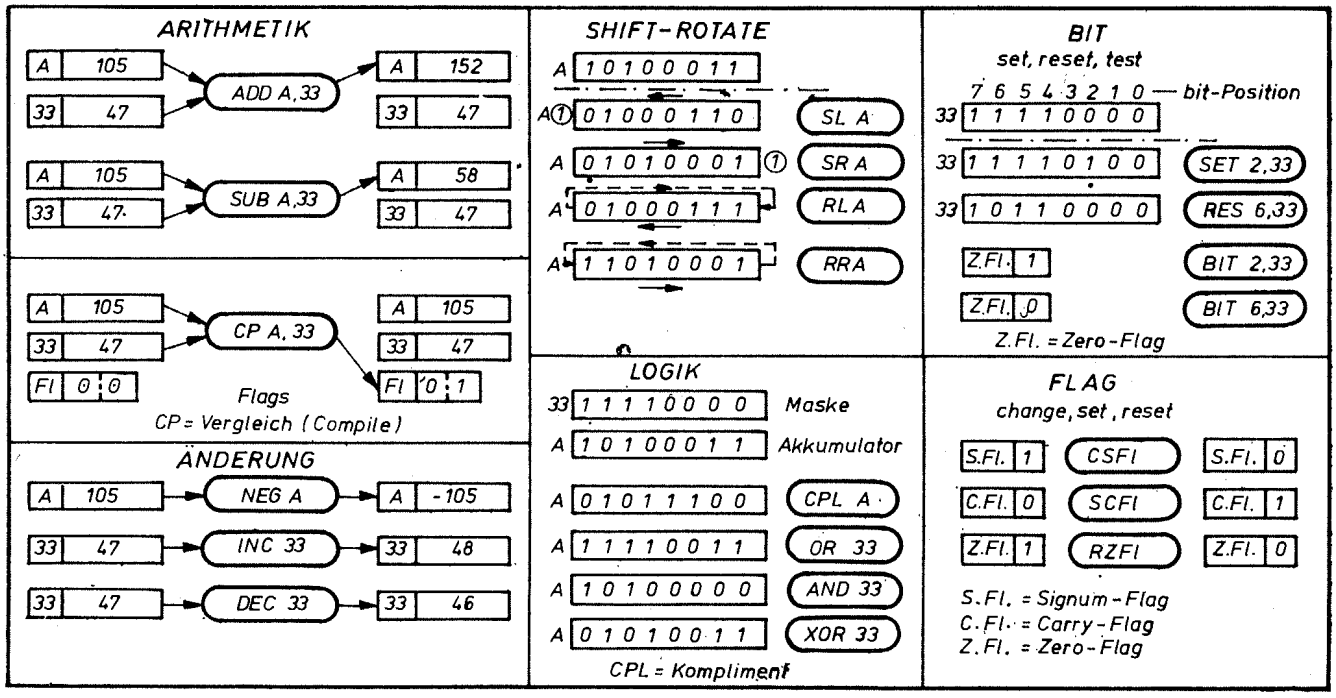


Bild 8.6.9. Zur Veranschaulichung der ALU-Befehle

Wie es vorwiegend der Fall ist, wurden sie auf den Akkumulator A und einen Register- bzw. Speicherplatz, z.B. 33, bezogen. Bei Arithmetik-, Vergleichs- und Änderungsbefehlen ist der Inhalt in dezimalen Zahlen dargestellt. Bei den anderen sind die einzelnen bit eingetragen

existieren in der Regel nur Addition und Subtraktion, jeweils in der Form mit und ohne Übertrag bzw. Borgen. Sie werden in der Form

ADD Akkumulator, 2. Operand
ADD A, 33
A = A + 33

geschrieben, d. h. im Akkumulator steht das Ergebnis aus der Summe des Akkumulatorinhalts vor der Operation und des Inhalts vom Register 33. Auch hier kann wieder hinter dem Komma ein Zahlenwert stehen: ADD A, 33 H bedeutet, daß zum Akkumulatorinhalt Hex 33 addiert wird. Statt A kann natürlich auch wieder ein Register stehen, also z.B. ADD 57, 22 H bedeutet, daß zum Inhalt des Registers 57 der Wert Hex 22 addiert wird. Dieselben Aussagen gelten für die Subtraktion SUB. Bei einigen Mikroprozessoren existiert neben der immer vorhandenen binären Arithmetik auch eine dezimale. Andere enthalten einen Befehl, der eine Umrechnung in BCD erlaubt. Es gibt schließlich auch Prozessoren, welche bereits Multiplikation und Division implementiert haben. Nur bei den Spezialprozessoren der wissenschaftlichen Taschenrechner existieren spezielle Funktionen wie LOG, SIN usw. Sie werden bei den Mikrorechnern per Programm oder mittels spezieller Arithmetikprozessoren (8.6.4.) ergänzt.

Ein weiterer wichtiger Befehl ist der

Vergleich (Compile).

Seine Form lautet z.B.: Setze ein Flag, wenn Register A und Platz 33 den gleichen Inhalt besitzen,

CP A, 33.

Schließlich gibt es arithmetische Befehle, die ein Register bzw. einen Speicherplatz betreffen.

NEG A

erzeugt eine Zahl mit dem entgegengesetzten Vorzeichen im Akkumulator.

INC 33

erhöht den Inhalt des Speicherplatzes 33 um die Zahl 1.

DEC 33

erniedrigt ihn um Eins. Diese beiden Operationen erfolgten für den Stackpointer bereits bei PUSH und POP.

Im Abschnitt 8.5.6.4. wurde gezeigt, wie eine Verschiebeoperation für die Multiplikation bzw. Division mit 2 genutzt werden kann. Deshalb sind die Funktionen

Shift (Schieben)
SLA = Shift links Akkumulator
SRA = Shift rechts Akkumulator

so wichtig. Auch sie können wieder das Überlaufflag zusätzlich in diese Operationen einbeziehen.

Bei der Multiplikation gemäß 8.5.6.5. mußte im MPR (in Bild 8.5.23c)) die Information ständig rotieren. Dies leisten die Funktionen

Rotate (Rotieren)
RLA = rotiere links um den Akkumulator
RRA = rotiere rechts um den Akkumulator.

Auch diese Funktionen existieren mit und ohne Übertragsflag. Die

Logik-Operationen

existieren meist in vier Arten:

Komplementieren
CPLA (bezüglich Akkumulator).

Hierbei werden Ø-Werte in 1 und 1-Werte in Ø umgewandelt. Dieser Befehl entspricht also bezüglich der einzelnen bit dem NOT. Die anderen drei Befehle beziehen sich in der Regel auf den Akkumulator und ein zweites Register bzw. einen Speicherplatz. Bei

OR 35 $\triangleq$ [A] = [A] $\vee$ [35]

wird für jedes bit die OR-Operation zwischen Akkumulator und Speicherplatz 35 gebildet und im Akkumulator gespeichert. Analog gilt

AND 35 $\triangleq$ [A] = [A] $\wedge$ [35]

und

XOR 35 $\triangleq$ [A] = [A] XOR [35],

wobei XOR das Exklusiv-Oder ist.
Einige Prozessoren besitzen auch

Bit-Operationen.

Sie beziehen sich auf ein bestimmtes Bit (die Numerierung hierzu enthält Bild 8.6.9, recht oben) eines Registers. Bei

SET 2,33

wird das Bit 2 des Speicherplatzes 33 auf 1 gesetzt; wenn dort schon eine 1 steht, bleibt sie erhalten. Bei

RES 6,33

ist das Bit 6 des Speicherplatzes auf Ø zurückgesetzt (reset), unabhängig davon, was zuvor dort stand. Die Operation

BIT 2,33 (oft auch TEST 2,33 genannt)

stellt fest, welchen Wert das Bit 2 im Speicherplatz besitzt, und setzt ein Flag auf den entgegengesetzten Wert. Im Speicher selbst erfolgt keine Änderung.

Zuweilen besitzen Mikroprozessoren auch noch die Möglichkeit, direkt auf Flags einzuwirken bzw. sie zu testen. In Bild 8.6.9 ist neben den Beispielen, die schon bei den Bitoperationen beschrieben sind, auch noch der Fall

CSFL = change signum flag,

d. h., das Flag wird auf den entgegengesetzten Wert gesetzt, behandelt.

8.6.2.6. Ein-, Ausgabe

Die Möglichkeiten und nicht nur die Befehle zur Ein- und Ausgabe faßt Bild 8.6.10 zusammen. Die Befehlsausstattung bezüglich der Ein- und Ausgabe ist bei den Mikroprozessoren im allgemeinen nicht sehr umfangreich. Sie existiert meist nur in der Form

IN A, 27 $\triangleq$ in Ziel, Quelle
OUT 27, A $\triangleq$ out Ziel, Quelle.

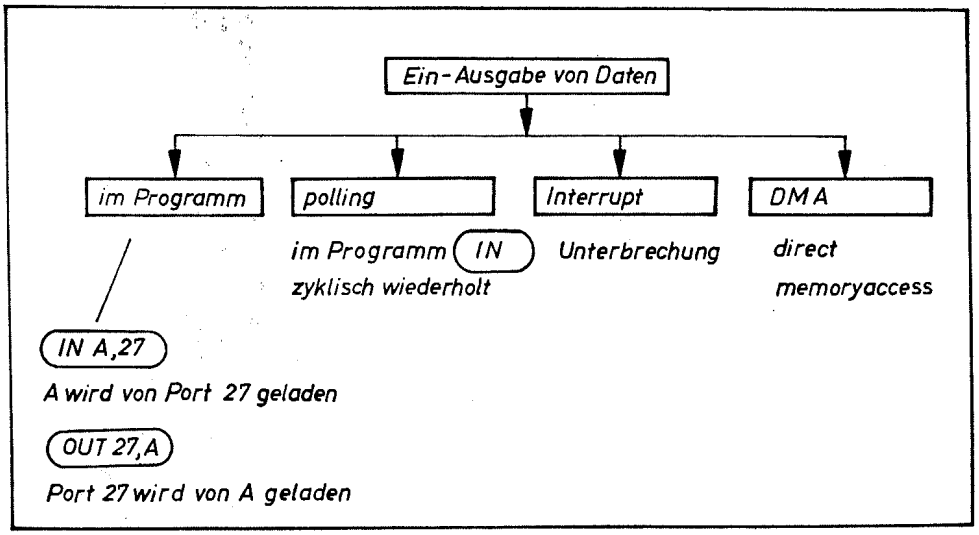


Bild 8.6.10. Einteilung der Ein- und Ausgabemöglichkeiten
Neben den Befehlen IN und OUT gibt es für die Eingabe zusätzliche Möglichkeiten beim Rechner

Die Anzahl der adressierbaren Ports (hier Nr. 27 verwendet) ist oft beschränkt (beim Z 80 z.B. auf je 256). Ähnliches gilt dabei für die zugänglichen Register und/oder Speicherplätze. Der Grund hierfür ist vielfältig, meist jedoch der, daß die drei anderen Methoden noch möglich sind.

Beim

Polling

wird in einer zyklischen Folge immer wieder eine IN-Operation realisiert. Beim

Interrupt

wird die Ein- bzw. Ausgabe über eine von außen gesteuerte Interrupt-Forderung eingeleitet (vgl. 8.6.4.). Beim

DMA = direct memory access,

dem direkten Speicherzugriff, wird unter Ausschaltung des Mikroprozessors, durch einen speziellen DMA-Baustein, eine schnelle Ein- bzw. Ausgabe von Daten unmittelbar auf die Speicherplätze realisiert.

Es sei erwähnt, daß einige Prozessoren auch über Ein-Ausgabebefehle verfügen, die ein schnelles, blockweises Ein- und Ausgeben von Daten ermöglichen.

8.6.2.7. Steuerbefehle

Die *Steuerbefehle* betreffen die CPU. Sie können in zwei Gruppen geteilt werden:

- Befehle zur Änderung im Programmablauf
- Befehle, die den Zustand der CPU betreffen.

Die letzteren sind schon beim Steuerbus im Abschnitt 8.6.2.1. behandelt:

NOP, WAIT, HALT und Interrupt-Befehle.

Die andere Gruppe sind im weitesten Sinne

Sprungbefehle.

Sie faßt schematisch Bild 8.6.11 zusammen. Hiervon gibt es *drei Gruppen*:

JUMP, CALL, RETURN.

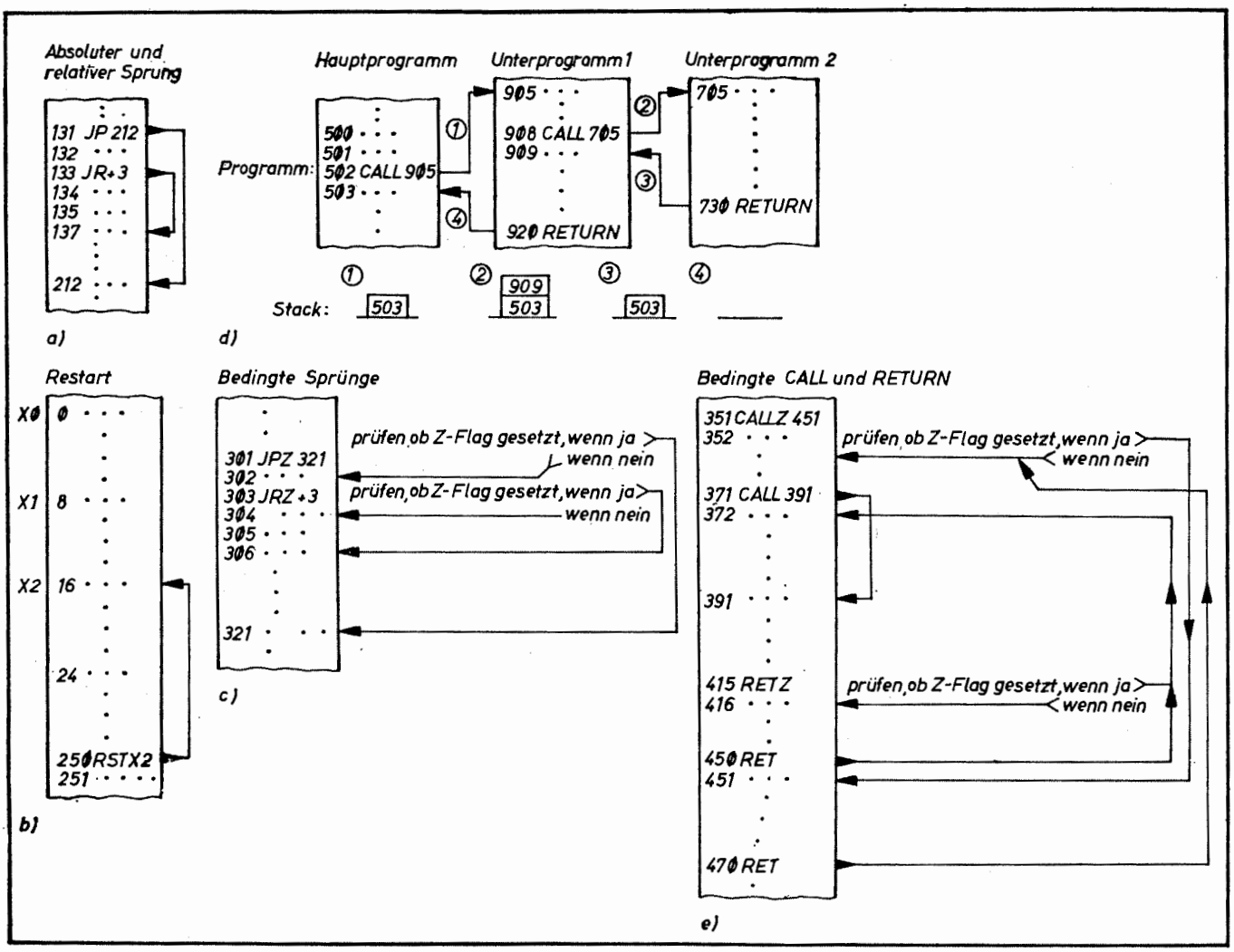


Bild 8.6.11. Zu Sprungbefehlen und Unterprogrammen

Alle drei können

unmittelbar bzw. bezüglich Bedingungen

erfolgen. Darüber hinaus besteht bei JUMP und CALL meist noch die Möglichkeit, daß sie sich auf

absolute oder relative Adressen

beziehen. Schließlich können die Adressen z. T. noch, gemäß dem nächsten Abschnitt, indirekt gewonnen werden. Die Vielfalt dieser Befehlsgruppe ist also beachtlich.

In Teilbild a) ist die Wirkung des absoluten und des relativen Sprungbefehls

JP 212 $\triangleq$ JUMP auf Speicherplatz 212
JR + 3 $\triangleq$ JUMP 3 Speicherplätze höher

dargestellt. Einige Prozessoren enthalten noch einen speziellen Befehl (Teilbild b))

RST = Restart X 1,

der etwa bedeutet: Beginne ein Unterprogramm (CALL) beim Speicherplatz X 1, wobei X 1, X 2 usw. einige speziell vom Prozessor ein für alle Male festgelegte Speicherplätze sind.

Der bedingte Sprungbefehl (Teilbild c))

JPZ 321

bedeutet etwa: Setze den Befehlszähler auf (springe auf Speicherplatz) 321, sofern das Z-Flag (zero) gesetzt ist. Anderenfalls gehe zum nächstfolgenden Befehl über. Das Z-Flag ist nur eine der möglichen Bedingungen.

Ein entsprechender relativer Befehl lautet

JRZ +3.

Hier wird bei gesetztem Z-Flag der Befehlszähler um 3, sonst um 1, erhöht.

Die verschiedenen

CALL-Befehle

bedeuten den Übergang zu einem Unterprogramm. Sie sind daher nur dann effektiv, wenn dieses Unterprogramm vom Hauptprogramm mehrfach aufgerufen wird (s. 8.8., Software). Bei diesen Befehlen ist der bereits behandelte Stackpointer (8.6.2.5.) notwendig. Er ist deshalb in Teilbild d) ergänzend mit dargestellt. Beim Aufruf

CALL 905

wird die nächsthöhere Befehlsadresse 503 im Stack abgelegt. Dann werden die Befehle ab Speicherplatz 905 abgearbeitet. Kommt der Rechner dabei zum Befehl Nr. 908 mit

CALL 705,

wird die Befehlsnummer 909 auf dem Stapelspeicher über 503 angeordnet, und dann werden die Befehle ab 705 abgearbeitet. Bei

RETURN = RET

wird die Adresse 909 vom Stack mit einer POP-Operation in den Befehlszähler geholt, und die Befehlsabarbeitung wird ab Adresse 909 fortgesetzt. Beim nächsten RETURN wiederholt sich dieser Prozeß, und damit ist der Stack leer und der Rechner wieder im Hauptprogramm.

Sowohl für CALL als auch für RETURN besteht nun noch die Möglichkeit, sie an Bedingungen zu binden. Dies zeigt Teilbild e). Weiter besteht die Möglichkeit, in Analogie zu Teilbild c) auch die CALL-Befehle relativ zu gestalten.

Der Pfeil in Bild 8.6.11e) ist eingefügt, da hier das Programm unterbrochen werden mußte, weil es sonst kein Ende finden würde. Dies kann u. a. durch einen Sprungbefehl auf eine Adresse kleiner als 351 oder größer als 470 erfolgen.

8.6.2.8. Adressierungstypen

Gemäß Bild 8.6.7b) besteht ein Befehlswort aus dem Operationskode und dem Operanden-Adreßteil. Der Operationskode ist durch die Befehlstypen gemäß den vorangegangenen Abschnitten bestimmt. Der Operanden-Adreßteil ist immer — oft nach mehrfachen Speicherabfragen — eine oder zwei binäre Zahlen. Die Prinzipien, wie diese Zahlen gewonnen werden, machen die verschiedenen Adressierungstypen aus. Sie sind in der Literatur z. T. unterschiedlich klassifiziert und z. T. bei gleicher Klassifizierung unterschiedlich bezeichnet. Dies erschwert oft — neben der Vielfalt an Möglichkeiten — die Übersicht.

Die *Problematik der Adressierung* ergibt sich aus den folgenden *vier Fakten*:

- Der Speicherumfang soll möglichst groß sein. Dies bedeutet einen langen Adreßteil. Bei 64 Kbyte sind z. B. bereits 2 Byte (16 bit) notwendig. Bei zwei Operanden, z. B. Addition, müßte dann der Operandenteil immer 4 Byte lang sein. Verkürzte Befehle können also wesentliche Vorteile bringen.
- Der Speicherumfang ist durch Zusatzgeräte, periphere Speicher, z. B. Floppy disc, Plattenspeicher, aber auch Halbleiterspeicher, zuweilen viel größer, als der

Mikrorechner adressieren kann. Hierdurch verschärft sich nochmals die eben aufgeführte Adressierungsproblematik.

- Bei vielen Programmen werden oft dicht aufeinanderfolgende Adressen benötigt. Beispiele sind die lineare Befehlsabarbeitung, d. h. ohne Sprung und Unterprogramm bzw. Zahlenwerte, die immer wieder zusammenhängend benötigt werden. Hierzu gehören z. B. die Koeffizienten eines Polynoms oder von Matrizen. Solche Fälle fordern von sich aus direkt verkürzte Befehle.
- Jeder Rechner enthält wenige Register (bzw. auch Flags) die viel einfacher behandelt werden können, da sie durch kurze, aber spezielle Adressen erreicht werden können.

Einen Überblick zu den Adressierungstypen gibt Bild 8.6.12. Das unterschiedlich lange Befehlswort zerfällt in den Operationskode und den Operanden-Adreßteil. Wird beides in einem Byte untergebracht (Bild 8.6.7c), zweite Zeile), so wird vom verschlüsselten Operanden-Adreßteil oder auch von impliziter Adressierung gesprochen.

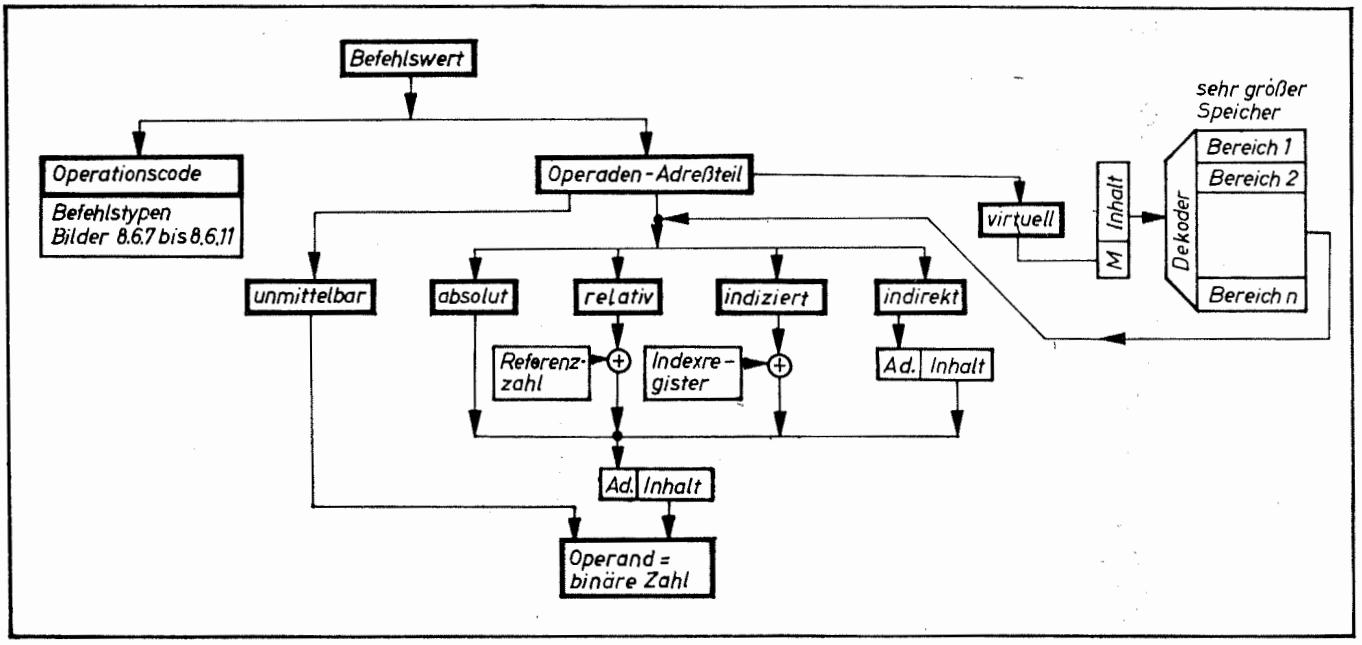


Bild 8.6.12. Systematik von Adressierungen

Viel häufiger werden aber für den Operanden-Adreßteil ein oder zwei Extrabytes verwendet. Hierfür wurde der allerdings wenig gebräuchliche Begriff konsekutive Adressierung verwendet. Die Unterteilung implizit — konsekutiv bezieht sich also darauf, wieweit der Operaranden-Adreßteil deutlich vom Operationskode getrennt existiert. Aus diesem Grunde wurde diese Unterteilung der besseren Übersicht wegen in Bild 8.6.12 weggelassen, und es werden nur die *Unterteilungen*

unmittelbar (direkt, immediate)
absolut (vollständig, direkt, erweitert direkt)
relativ (verkürzt, Nullseitenadressierung)
indiziert (indexiert)
indirekt (Ersetzung, impliziert, implied)
virtuell (Kachel, Cache, Page, Bankselect)

verwendet. Diese Aufzählung enthält zugleich in Klammern anderweitig verwendete bzw. nochmals untersetzte Begriffe. Die Begriffe seien im folgenden besprochen. Vorher ist es jedoch notwendig, auf die möglichen *Speicherbezüge* einzugehen. Sie

können vor allem sein

Flag
einzelne bit (Bitadressierung)
Register der CPU (Registeradressierung)
Tore für Ein-Ausgabe
Speicherzellen
Speicherbereiche.

Sie alle sind mittels binärer Zahlen (Operanden, Adressen) erreichbar. Deshalb sei im folgenden zwischen ihnen nur in Sonderfällen unterschieden. Ansonsten wird immer von Speicherzellen gesprochen, da sie besonders wichtig sind.

Die *unmittelbare* Adressierung beinhaltet sofort die binäre Zahl und ist daher im strengen Sinne keine Adressierung. Beispiele hierfür sind JP 335, LD A, 33 H. Insbesondere bei JUMP und CALL existiert auch die unmittelbare relative Adressierung, z.B. JR +3. Hier wird der Stand des Befehlszählers als Referenz verwendet. Es sei erwähnt, daß die nach JUMP und CALL folgende Zahl oft wie die Adresse einer Speicherzelle behandelt wird. Sie ist aber der eigentliche Operand dieser Befehle, welche ja vor allem den Befehlszähler auf diese Zahl setzen. Die damit adressierte Speicherzelle enthält dann ein neues Befehlswort.

Die *absolute* Adressierung verweist auf eine Speicherzelle, deren Inhalt als binäre Zahl verwendet wird. Beispiele sind JP [IX], LD A, B. Im ersten Fall wird der Inhalt des Registers IX verwendet.

Die *relative* Adressierung ist eine verkürzte Adressierung. Es wird eine Referenzzahl und eine positive oder negative Distanzzahl (meist von −127 bis +128) verwendet. Ein Sonderfall liegt bei der Zero-page- (Null-Seiten-) Adressierung vor. Hierbei wird z.B. der Speicherbereich von Ø bis 255 verwendet, und es genügen 8 bit zur Adressierung. Die Referenz ist dann also Null.

Bei der *indizierten* Adressierung wird als Referenz der Inhalt eines speziellen Indexregisters verwendet. Dieser Fall ist besonders wertvoll für die Koeffizienten von Polynomen oder anderen Feldern. Für Matrizen ist es günstig, wenn zwei Indexregister existieren. Sie heißen dann IX und IY und können für die Bezeichnung der Zeilen und Spalten verwendet werden, bzw. sie sind daher zusätzlich besonders leicht zu inkrementieren und zu dekrementieren.

Die *indirekte* Adressierung hat für komplexe Programme große Bedeutung. Hier wird über ein Register oder eine Speicherzelle jene Speicherzelle (oder Register) adressiert, dessen Inhalt den Operanden ergibt. In verschiedenen Fällen existiert auch noch eine Kombination von indirekter und indizierter Adressierung.

Die *virtuelle* Adressierung wurde hier als ein Übergriff verwendet, der jene Fälle zusammenfaßt, wo der Speicher größer ist, als es der Adressierungsbereich des Mikroprozessors zuläßt.

Ein Spezialfall, der besonders bei den Mikroprozessoren angewendet wird, ist das *Bank-select* (Bankumschaltung). Das Prinzip geht aus Bild 8.6.12, rechts, hervor. Es existieren dabei Banken (Speicherbereiche), z.B. zu je 64 Kbyte. Jede Bank ist mit einer 16-bit- (2-Byte-) Adresse voll zu selektieren. Zwischen ihnen wird mittels eines Multiplexers MUX oder eines Dekoders umgeschaltet. Die Ansteuerung erfolgt oft über mehrere Flipflops, deren Zusammenschaltung dann als Bar bezeichnet wird. In dem ausgewählten Speicherbereich wird wie in dem sonst fest gegebenen gearbeitet. Die Umschaltung zwischen den Speicherbereichen erfolgt in der Regel sehr selten.

Der Begriff *virtuell* stammt eigentlich aus der Großrechentechnik, wo z.B. der gesamte Inhalt eines Plattenspeichers adressiert werden soll. Hierbei werden statt des Bank-select immer entsprechende Teilauszüge (Seiten) des Plattenspeichers in den Arbeitsbereich geholt. So entsteht der virtuelle Eindruck, daß direkt mit einem

Speicher von der Größe des Plattenspeichers gearbeitet wird. Da in englisch die Seite page heißt, wird in diesem Zusammenhang auch von *paging* gesprochen.

In Großrechnern wird oft aus Preisgründen nur ein Teil des Arbeitsspeichers besonders schnell ausgeführt, dies ist der sogenannte *Cache-Speicher*. Ähnlich wie beim Bank-select oder beim virtuellen Speicherprinzip wird nur versucht, den am häufigsten benötigten Adreßraum im Cache abzulegen. Hiermit in engem Zusammenhang steht die *Seitenadressierung* bzw. die relative bzw. verkürzte Adressierung. Dabei muß jedoch nicht der schnelle Cache-Speicher vorhanden sein.

8.6.2.9. Registeranordnungen und -typen

Ein Mikroprozessor muß eine *bestimmte Anzahl von Registern* unbedingt besitzen. Dieses *Minimum* ist in etwa aus Bild 8.6.6a) erkennbar:

- Befehlszähler BFZ (Programmcounter PC)
- Befehlsregister
- temporäre Register der ALU.

Auf sie ist hier nicht mehr einzugehen, da ihre Funktion und Wirkungsweise schon beschrieben wurde.

Es bringt gewisse Vorteile, wenn darüber hinaus zusätzliche Register vorhanden sind. Sie sind in Bild 8.6.6a) als Registerblock zusammengefaßt. Diese Register sollen hier behandelt werden. Da die Register für einen speziellen Mikroprozressor z. T. wesentlich typischer als der Befehlssatz sind, müssen sie hier an einigen Beispielen behandelt werden (Bild 8.6.13). Den wohl einfachsten Registersatz enthält der Mikroprozessor **6502**, der zugleich einer der weitverbreitetsten Mikroprozessortypen ist. Er enthält nur drei 8-bit-Register, A (Akkumulator), X, Y, einen 8-bit Stackpointer SP, ein Flag-Register und den Programmzähler PC. Hervorgegangen ist der 6502 aus dem **6800**, der sich im Registersatz nur durch größere Register (und etwas abweichende Bezeichnungen) unterscheidet. Diese beiden Mikroprozessoren sind also deutlich auf bevorzugte Arbeit mit dem externen Speicher orientiert. Sie enthalten dabei zur schnelleren Abarbeitung der Befehle ein Pipeline-Prinzip (8.9.7.2.). Weiter besitzen die Typen eine sonst nicht übliche einfache Umschaltung zwischen binärer und dezimaler Arithmetik. Umfangreicher ist der Registersatz bei der Typenfolge

8008 (U 808), 8080/8085, Z 80 (U 880).

Der Intel **8008** war der zweite realisierte Mikroprozessor. Er verfügt neben dem 8-bit-Akkumulator noch über sechs 8-bit-Register, acht 14-bit-Register und ein 4-bit-Flag-Register. Von den 14-bit-Registern wird eins als Programmzähler und die restlichen sieben als Adressenverwaltung im Sinne eines Stack verwendet. Damit sind sieben verschachtelte Unterprogrammroutinen möglich. Es gibt bei ihm daher nur einen Stackpointer SP von 3 bit. Die Registerpaare BC, DE und HL sind so gestaltet, daß sie auch als 16-bit-Register verwendet werden können. Sie sind dadurch gut für die Adressierung in einem 16-bit-Raum (64 Kbyte) geeignet. So ergibt sich auch die Bezeichnung HL mit dem Bezug auf L = Low- und H = High-Anteil für die Adressen.

Bei der Weiterentwicklung des 8008 zum **8080** wurde einmal das Doppelregisterprinzip beibehalten. Das Flag-Register wurde auf 8 bit (die nicht alle genutzt werden) erweitert. Der Stackspeicher jedoch wurde, wie es seit diesem Schaltkreis allgemein üblich ist, in den äußeren Speicher verlegt, so daß nur noch der Stackpointer SP als speziel-

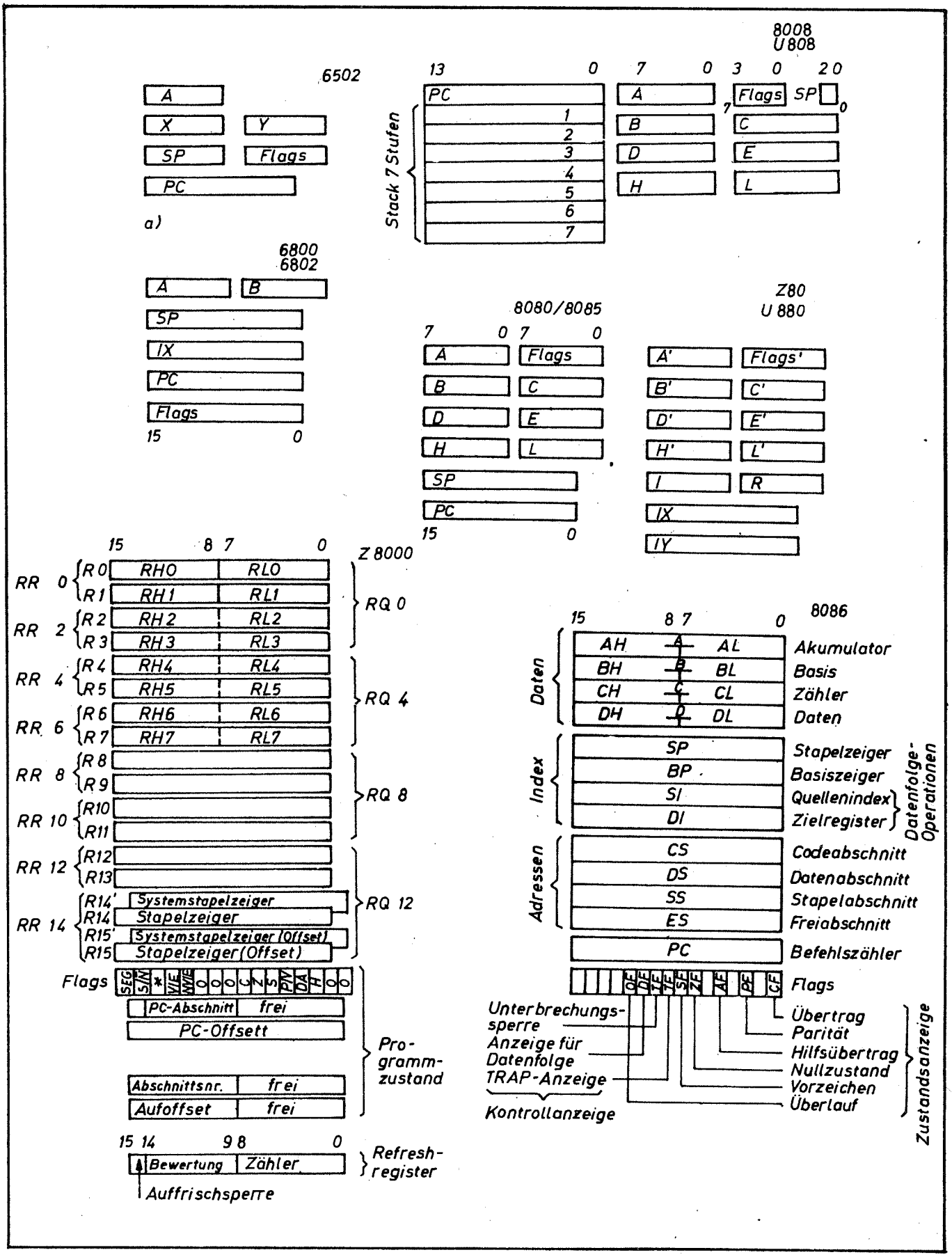


Bild 8.6.13. Registerstruktur einiger viel angewendeter 8- und 16-bit-Mikroprozessoren

Es bedeuten

PC Programmcounter (Befehlszähler)
SP Stackpointer (Stapelzeiger)
IX, IY Indexregister
A im allgemeinen Akkumulator
R Refresh-Register

les Register im Mikroprozessor liegt. Dadurch kann eine beliebige Tiefe des Stacks realisiert werden.

Mit dem **Z 80** wurden alle Register verdoppelt. Dies bringt einen sehr großen Vorteil bei Interrupt-Behandlungen und anderen Operationen. Es ist eine einfache „Bankumschaltung" zwischen den beiden Registersätzen mittels des Befehls EXX schnell möglich. Weiter sind zwei Indexregister IX, IY, ein Interruptregister (für eine Interrupthierarchie, s. 8.6.4.) sowie ein Refresh-Register für dynamische Speicher vorgesehen. Ein solcher Mikroprozessor stellt also allein durch die Register eine hohe Komplexität für die Anwendung bereit.

Vergleicht man die beiden wohl wichtigsten Entwicklungsreihen der „Standard"-8-bit-Mikroprozessoren, so zeigt sich:
6800 ... 6502 sind speicherorientiert und ermöglichen durch ihren einfachen Registeraufbau spezielle Eigenschaften in den Befehlstypen.
8008 — 8080 — Z 80 sind stärker registerorientiert und ermöglichen dadurch einen flexibleren Einsatz mit komplexeren Befehlen.

Die Weiterentwicklung bei den Prozessoren folgte etwas bevorzugt dem zweiten Weg. Hier sei nur kurz auf drei Typen von 16-bit-Prozessoren und einen Ein-Chip-Prozessor eingegangen.

Der **Z 8000** (Zilog) ist eine Weiterentwicklung des Z 80 zu einem 16-bit-Prozessor. Wie das Bild 8.6.13 zeigt, ist der Registersatz hier noch umfangreicher geworden. Auch die Teilung und Zusammenfassung von Registern wurde weiter verfeinert. Die ersten acht Register lassen sich auf je 2×8 bit teilen. Andererseits lassen sich je zwei Register zu einem 32-bit-Register zusammenfassen. Der Z 8000 wird deshalb zuweilen als „verkappter" 32-bit-Prozessor bezeichnet. Hierzu trägt weiter bei, daß auch vier Register zusammengefaßt werden können und dann eine 64-bit-Verarbeitung möglich ist. Alle Register sind nahezu universell, d. h gleichberechtigt anwendbar. Indexregister sind daher speziell nicht mehr vorgesehen.. Hierfür kann je nach Bedarf jedes Register oder auch die oberen Registerteile verwendet werden. Die Vielfalt der Kontrollregister (Programmzustand) ist als erhebliche Erweiterung des sonst vorhandenen Flag-Registers zu betrachten.

Eine andere Entwicklung legte Intel als Weiterentwicklung aus dem 8080 vor, den **8086.** Er ist neben der möglichen Teilung der vier obersten Register auf 8 bit von konsequenter 16-bit-Registerstruktur. Die Register sind hier auf spezielle Anwendungen orientiert:

4 Datenregister
4 Indexregister
4 Adreßregister
2 Kontrollregister.

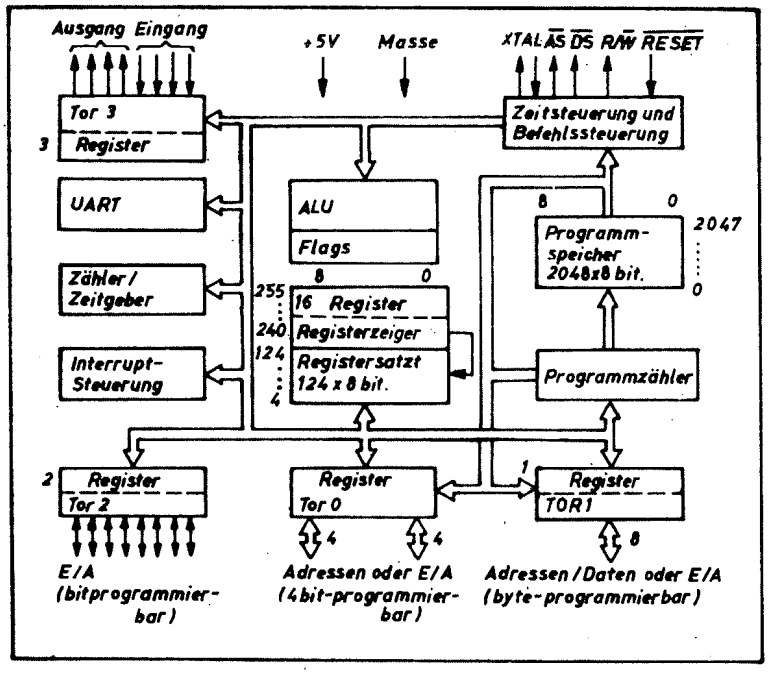
Das Flag-Register (Status) ist bei ihm recht übersichtlich angelegt. Er besitzt also die vergleichsweise einfachere Registerstruktur.

Die extreme Einfachheit bei den 16-bit-Prozessoren besitzt der viel ältere **TMS 9900** von Texasinstrument. Noch einfacher als beim 6502 existieren nur drei interne Register:

Arbeitsbereichszeiger
Programmzähler
Flag (Zustandsregister)

von je 16 bit Breite. Er konnte wahrscheinlich infolge dieser Einfachheit keine größere Anwendungsbreite erreichen.

Ein-Chip-Prozessoren zeichnen sich dadurch aus, daß sie im Prinzip zum Aufbau eines Systems ohne Zusatzschaltkreise (vgl. 8.6.7.) auskommen können. Sie sind



a)

Interne Adressen		Bezeichnung
255	Stackpointer (Bit 0...7)	SPL
254	Stackpointer (Bit 8...15)	SPH
253	Registerzeiger	RP
252	Programmsteuerflags	FLAGS
251	Interruptmaskenregister	IMR
250	Interruptanforderungsregister	IRQ
249	Interruptprioritäts-register	IPR
248	Tore 0 und 1 Betriebsart	P01M
247	Tor 3 Betriebsart	P3M
246	Tor 2 Betriebsart	P2M
245	T0 Voruntersetz. laden	PRE0
244	Zähler/Zeitgeber 0 laden	T0
243	T1 Voruntersetz. laden	PRE1
242	Zähler 1 Zeitgeber 1 laden	T1
241	Zeitgeberbetrieb	TMR
240	Serielle E/A	SIO
	nicht vorhanden	
127 ... 4	Register für allgemeine Anwendungen	
3	Tor 3	P3
2	Tor 2	P2
1	Tor 1	P1
0	Tor 0	P0

b)

Bild 8.6.14. a) Aufbau und b) interne Adressen der internen 128 + 16 Register des Ein-Chip-Prozessors Z 8, U 8810, U 8820

dadurch in der Regel vor allem für einfache Aufgaben geeignet, können aber natürlich trotzdem extern, insbesondere durch Speicherbausteine, erweitert werden. Ein Beispiel hierfür ist der Z 8, dessen Prinzipaufbau und Registerstruktur Bild 8.6.14 zeigt. Als zusätzliche Baugruppen gegenüber einem Standardmikroprozessor besitzt er

- Programmspeicher (ROM), der die peripheren Adressen 13 bis 2047 belegt;
- vier Ein-Ausgangstore (interne Adressen Ø bis 3 und 245 bis 248);
- Zähler/Zeitgeber (interne Adressen 241 bis 244);
- serielle Ein-Ausgabe (UART; interne Adresse 240);
- Interrupt-Steuerung (interne Adressen 249 bis 251).

Der Stackpointer liegt auf den internen Adressen 254, 255. Der interne Registersatz ist mit 128 × 8 bit sehr umfangreich, liegt auf den internen Adressen Ø bis 127 und wird vom Registerzeiger auf der internen Adresse 253 verwaltet.

Die internen Adressen 240 bis 255 sind also als 16 Steuer- und Statusregister ausgebildet. Die Register Ø bis 3 stehen für die Tore und die Adressen 4 bis 127 zur allgemeinen Anwendung bereit. Alle Register sind 8 bit breit.

Nach außen kann folglich über Tore Ø bis 3 Verbindung aufgenommen werden. Dabei kann u. a. über ein Tor auch ein Speicherbereich bis 64 Kbyte adressiert werden, wobei allerdings der Bereich bis 2047 durch das ROM einschließlich der unteren 12 Adressen für den Interrupt-Vektor fest belegt ist.

8.6.3. U 880 — Z 80

8.6.3.1. Aufbau der Hardware

Die interne Struktur des Z 80 zeigt Bild 8.6.15. Eine Zusammenfassung der Anschluß- und Steuersignale sowie ihre kurze Erklärung enthält Tab. 8.6.5. Es ist dabei auffällig, daß für alle Steuersignale Low aktiv (negative Logik) gilt. Dies hat Vorteile bei der Geschwindigkeit (HL-Flanken), größere Störsicherheit bez. Kapazitäten und

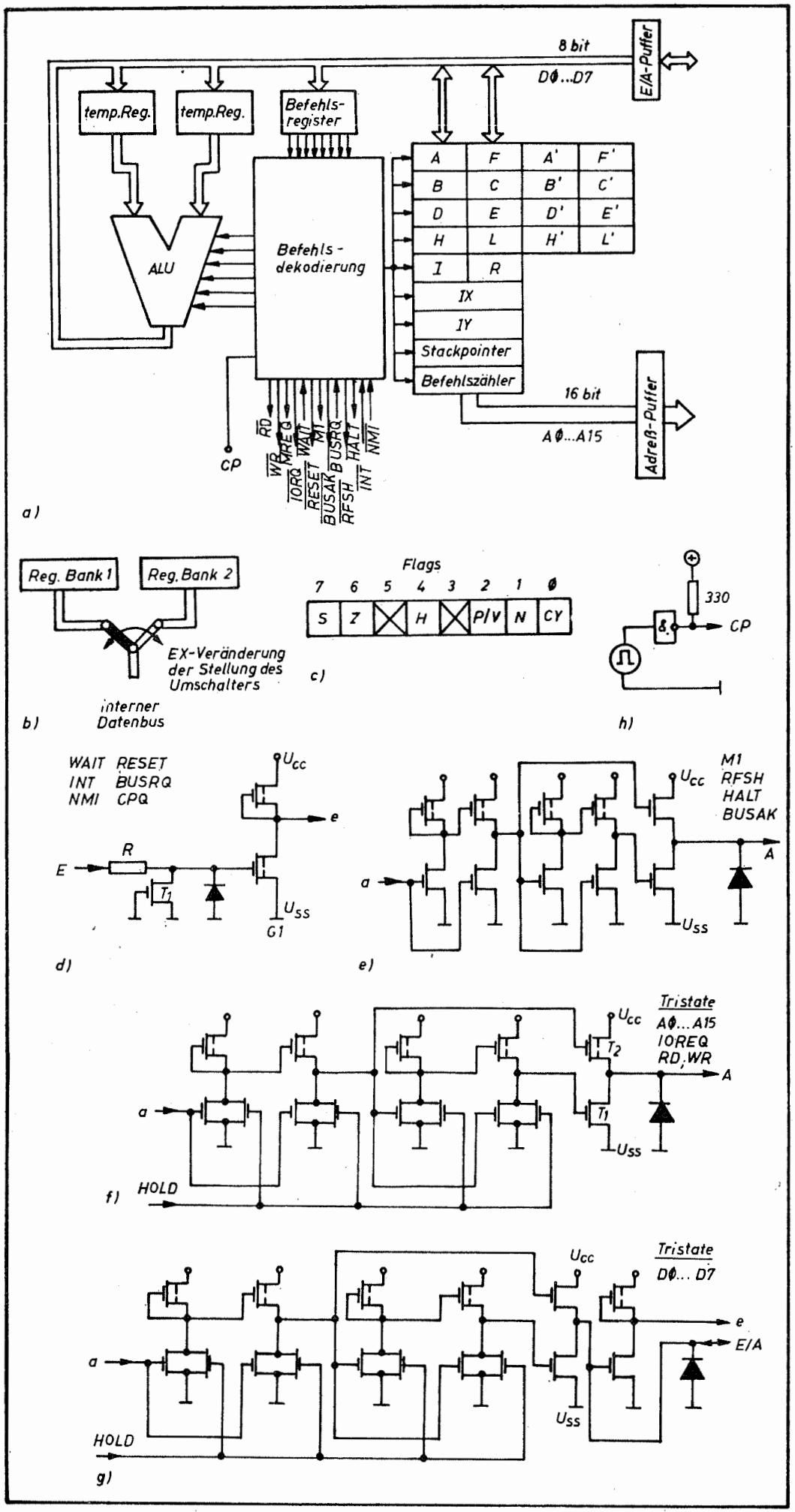


Bild 8.6.15. a) Innerer Aufbau vom Z 80, b) Prinzip der Umschaltung der Registersätze, c) Anordnung der Flags und d) bis g) Schaltung der verschiedenen Ein- und Ausgänge. Hold bewirkt den hochohmigen Tristate-Zustand. h) zeigt die Anschaltung für das Taktsignal

Tabelle 8.6.5. Beschreibung der Anschlüsse des Z80 (U880)

A0 ... A15	16-bit-Adreßbus	— Tristate-Ausgänge — liefern Adressen für externe Speicher (max. 64 Kbyte) und E/A-Bausteine
D0 ... D7	8-bit-bidirektionaler Datenbus	— Tristate-Ein-/Ausgänge — Datenaustausch zwischen CPU und Speicher bzw. CPU und E/A-Bausteinen
$\overline{M1}$	Maschinenzyklus 1	— Tristate-Ausgang, Low aktiv — bei M1 aktiv ist der laufende Maschinenzyklus der Operations kode-Lesezyklus der momentan auszuführenden Anweisung — Bestätigung der Interrupt-Anforderung (nur zusammen mit $\overline{IORQ}$ aktiv)
$\overline{MREQ}$	Speicheranforderung (memory request)	— Tristate-Ausgang, Low aktiv — Speicherzugriff (Lesen oder Schreiben)
$\overline{IORQ}$	E/A-Anforderung, (input/output-request)	— Tristate-Ausgang, Low aktiv — niederwertige 8 bit des Adreßbusses liefern Adresse zur E/A-Torauswahl (Ein- oder Ausgabe) — Bestätigung der Interruptanforderung (nur zusammen mit $\overline{M1}$ aktiv)
$\overline{RD}$	Lesen (read)	— Tristate-Ausgang, Low aktiv — Lesen der Daten vom Speicher oder von einem E/A-Baustein
$\overline{WR}$	Schreiben (write)	— Tristate-Ausgang, Low aktiv — CPU hält Daten für Speicher oder E/A-Baustein auf Datenbus bereit
$\overline{RFSH}$	Auffrischen der Information (refresh)	— Tristate-Ausgang, Low aktiv — niederwertige 7 bit des Adreßbusses führen Refresh-Adresse für dynamische Speicher — Verwendung der laufenden $\overline{MREQ}$-Signale zur Einleitung eines Refresh-Zyklus für alle angeschlossenen Speicher
$\overline{HALT}$	Halt-Zustand (Halt)	— Ausgang, Low aktiv — CPU im (Software-) Halt-Zustand — zur weiteren Programmabarbeitung ist Interrupt-Signal nötig — CPU führt zur Sicherstellung des Refresh-Vorganges Leerbefehle (NOP) aus
$\overline{WAIT}$	Warte-Anforderung (wait)	— Eingang, Low aktiv — Warten der CPU, bis Speicher oder E/A-Schaltung bereit
$\overline{INT}$	maskierte Interrupt-Anforderung (interrupt-request)	— Eingang, Low aktiv — Unterbrechungsanforderung an die CPU (wird erst nach Abarbeitung des in Ausführung befindlichen Befehls berücksichtigt)
$\overline{NMI}$	nichtmaskierbare Interrupt-Anforderung (nonmaskable interrupt)	— Eingang, Low aktiv — Interrupt höherer Priorität — bei Anlegen eines Low-Signals an $\overline{NMI}$ wird entsprechend einer RESTART-Instruktion der Programmablauf bei Speicheradresse 0066 H fortgesetzt
$\overline{RESET}$	Rücksetzen (reset)	— Eingang, Low aktiv — Rücksetzen von Interrupt-Freigabe-Flipflop, Befehlszähler, Register I und R

Fortsetzung Tabelle 8.6.5.

		— CPU in Interrupt-Betriebsart 0 während $\overline{\text{RESET}}$ aktiv — Daten- und Adreßbus hochohmig, die übrigen Ausgänge inaktiv
$\overline{\text{BUSRQ}}$	Bus-Anforderung (bus-request)	— Eingang, Low aktiv — Adreß-, Daten- und Steuerbus hochohmig, so daß diese Leitungen von externen Schaltungen benutzt werden können
$\overline{\text{BUSAK}}$	Bus-Anforderungsbestätigung (bus-acknowledgement)	— Ausgang, Low aktiv — Bestätigung, daß Adreß-, Daten- und Steuerbus in hochohmigem Zustand sind
$\overline{\text{CP}}$	Systemtakt (clock pulse)	— Takteingang für einphasigen TTL-Takt

die Möglichkeit eines verdrahteten AND. Für die Adreß- und Datenleitungen dagegen gilt High aktiv. Für die Ein- und Ausgänge gibt es vier verschiedene Schaltungen. Sie zeigt Bild 8.6.15d) bis g). Eine Beschaltung des Eingangs CP zeigt Bild 8.6.15h). Die Standardfrequenz ist 2,5 MHz und wird in der Regel von einem Quarz gewonnen.

Die *Registersätze* A bis F und H, L sowie A′ bis F′ und H′, L′ sind mittels Bankumschaltung austauschbar. Das Prinzip deutet Bild 8.6.15b) an. Daraus wird auch deutlich, weshalb stets nur ein Registersatz zugänglich ist. Der jeweils nicht zugängliche wird mit dem hochgestellten Strich versehen. Dabei sind getrennt austauschbar die Register AF ↔ A′F′ zum einen und zum anderen gemeinsam BC ↔ B′C′; DE ↔ D′E′; HL ↔ H′L′.

Das Register A wird als Akkumulator und die Register B, C, D, E, H, L als universelle 8-bit-Register verwendet. Das F-Register enthält die *Statusinformationen* der

Tabelle 8.6.6. Bedeutung der Flags beim Z 80

CY	FØ	Carry-Flag: Übertrag: wird gesetzt bei Überlauf von Additionen und beim Borgen bei Subtraktionen. Es wird z. T. beim Schieben und Rotieren verwendet.
N	F1	Additions-Subtraktions-Flag: Es dient zur Steuerung des DAA-Befehls für BCD-Zahlen.
P/V	F2	Parity-overflow-Flag: Parität—Überlauf. Dieses Flag hat zwei Aufgaben: bei logischen Operationen Anzeige der Parität, bei arithmetischen Operationen Anzeige eines Überlaufs bei vorzeichenbehafteten Zweierkomplimenten. Es ist also gesetzt bei Zahlen kleiner −128 oder größer +127.
—	F3	nicht vorhanden
H	F4	Halbbyte-Übertrags-Flag: gilt bezüglich BCD-Zahlen, wenn ein Überlauf des niederwertigen Byte auftritt, wichtig für den Befehl DAA.
—	F5	nicht vorhanden
Z	F6	Zero-Flag: Null: wird gesetzt, wenn Ergebnis im Akkumulator Null ist.
S	F7	Signum-Flag: Vorzeichen: dient der Verarbeitung von vorzeichenbehafteten Zahlen, speichert daher den letzten Wert von Bit 7 im Akkumulator.

Bei *Programmverzweigungen* werden folgende Entscheidungen bezüglich der Flags verwendet:

Wert = 1		Wert = Ø		Flag	
C	Übertrag	NC	kein Übertrag	F1	CY
PE	gerade (even) bzw. Überlauf	PO	ungerade (odd) bzw. kein Überlauf	F2	P/V
Z	Null	NZ	nicht Null	F6	Z
M	minus (negativ)	P	positiv	F7	S

ALU gemäß Tab. 8.6.6 und Bild 8.6.15c). Die 8-bit-Register I werden für den Interruptvektor und R für die jeweilige Refresh-Adresse verwendet. IX und IY sind Index-Register mit jeweils 16 bit.

Die **ALU** verfügt über folgende Funktionsarten, z. T. in jeweils mehreren Varianten, auf die noch genauer bei den Befehlen im Abschnitt 8.6.3.5. eingegangen wird:

- Addition
- Subtraktion
- Inkrementieren
- Dekrementieren
- Negieren
- Dezimalkorrektur
- AND
- OR
- XOR
- Bit setzen
- Bit rücksetzen
- Bit testen
- links und rechts schieben, arithmetisch und logisch
- links und rechts rotieren, bit- und 4-bit-weise

8.6.3.2. Zeitverhalten

Alle Zeitabläufe der CPU werden vom externen Takt CP gesteuert und lassen sich mittels sieben *einfacher Zeitabläufe* beschreiben (Bild 8.6.16):

- Aufruf des Operationskodes des Befehls (M1-Zyklus),
- Speicherdatenschreib- oder -lesezyklus,
- E/A-Lese- oder -Schreibzyklus,
- Bus-Anforderungs- und -Bestätigungszyklus,
- Interrupt-Anforderungs- und -Annahmezyklus,
- nicht maskierbarer Interrupt: Anforderungs- und -Annahmezyklus,
- Ausführung eines HALT-Befehls.

Jeder Zyklus ist dabei ein Maschinen- (M-) Zyklus, der auf 3 bis 6 Takte (T-Zyklen) fällt und der zur Synchronisation mit langsamen Speichern oder peripheren Geräten um eine oder mehrere WAIT-Zeiten verlängert werden kann. Hierzu wird das $\overline{\text{WAIT}}$-Signal im zweiten Takt jedes Maschinenzyklus bei Ausführung des WAIT bei jeder abfallenden Taktflanke abgelegt.

Beim *Aufrufzyklus* (M1-Zyklus) wird der Inhalt des Programmzählers auf den Adreßbus gegeben. Mit $\overline{\text{MREQ}}$ wird eine halbe Taktperiode später der entsprechende Programmspeicherplatz aufgerufen und am Ende von T2 mit $\overline{\text{RD}}$ der Inhalt als Befehl übernommen (IN). Der M1-Zyklus wird mit einem $\overline{\text{RFSH}}$ zum Auffrischen von dynamischen Speichern (T4, T5) abgeschlossen.

Der *Speicherlese-* und *-schreibzyklus* weist einen ähnlichen Verlauf wie der M1-Zyklus auf. Statt des Inhalts des Programmzählers werden jedoch die gewünschten Speicheradressen für Daten auf den Adreßbus gegeben. Bei den einzuschreibenden Daten muß statt RD das Signal $\overline{\text{WR}}$ auf Low gesetzt werden. Etwas zuvor und auch noch etwas nachher müssen jedoch die einzuschreibenden Daten auf dem Datenbus stabil liegen.

Die *Ein-* und *Ausgabezyklen* unterscheiden sich von den Speicherzyklen nur dadurch, daß nicht $\overline{\text{MREQ}}$, sondern $\overline{\text{IORQ}}$ auf Low gesetzt wird.

Gemäß Abschnitt 8.6.2.2. wird es insbesondere bei DMA-Betrieb notwendig, daß die CPU verschiedene Leitungen in den hochohmigen Zustand versetzt. Dann ist nämlich u. a. ein schneller Speicherzugriff von außen möglich. Dies erfolgt im *Bus-Anforderungs-* und *-Bestätigungszyklus*. Mit der Anstiegsflanke des letzten Taktes eines jeden Maschinenzyklus wird das BUSRQ-Signal abgetastet. Bei Bus-Anforderung liegt es dann auf Low, sonst auf High. Liegt es auf Low, so werden alle Tristate-Ausgänge der CPU in den hochohmigen Zustand überführt und gleichzeitig das

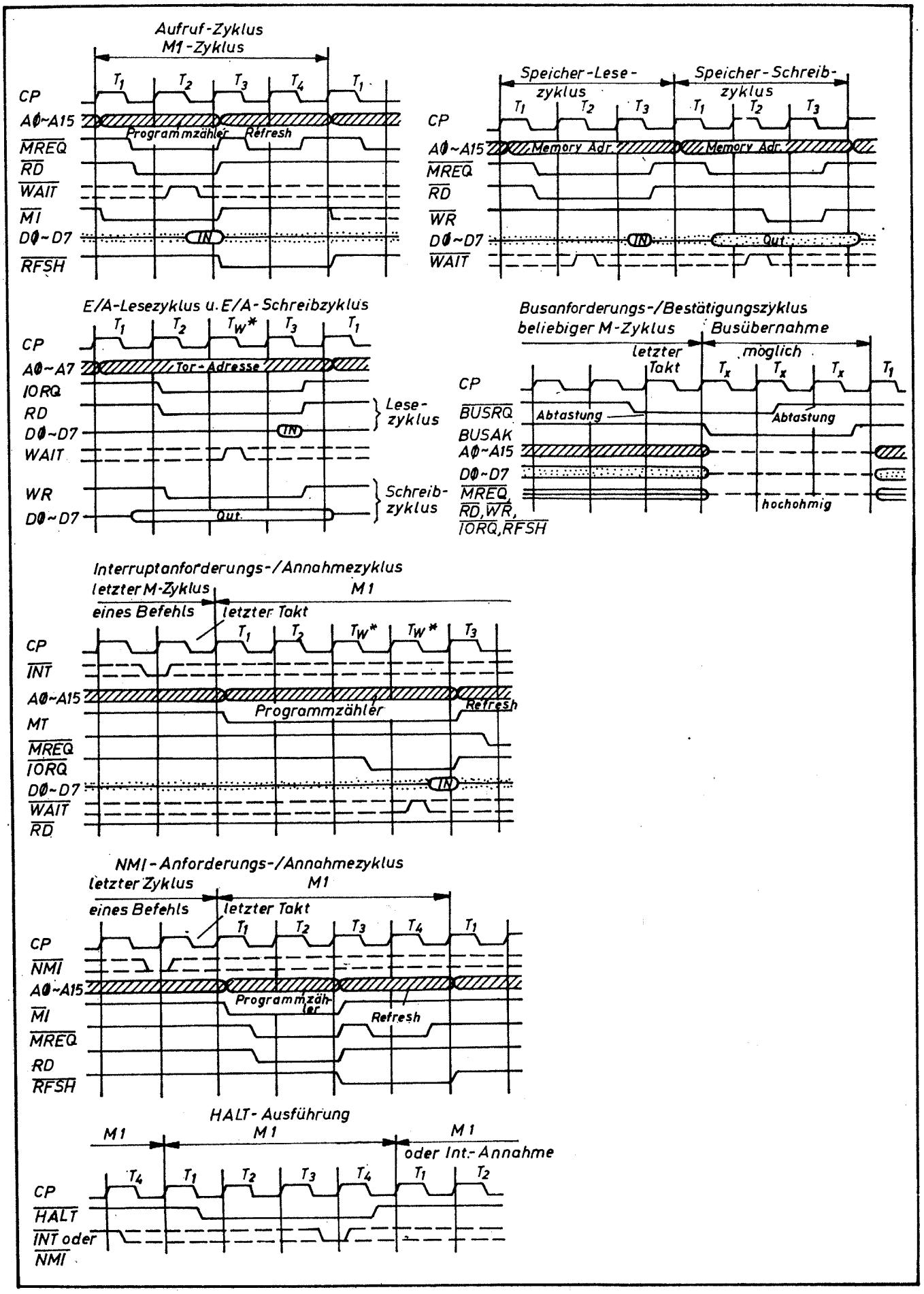


Bild 8.6.16. Zusammenstellung der verschiedenen Maschinenzyklen eines U 880

Signal BUSAK auf Low gesetzt. Hierbei ist zu beachten, daß dann auch kein $\overline{\text{RFSH}}$ von der CPU abgegeben werden kann. Erst nachdem $\overline{\text{BUSRQ}}$ wieder auf High geht, kann die CPU in den normalen Betriebszustand zurückkehren und ihre Arbeit fortsetzen.

Auch der $\overline{\text{INT}}$-Eingang (Interrupt) der CPU wird bei der steigenden Flanke des letzten Taktes jedes Befehls abgefragt. Das Signal wird nicht berücksichtigt, wenn das Interrupt-Annahme-Flipflop der CPU abgeschaltet ist oder wenn das $\overline{\text{BUSRQ}}$-Signal aktiv ist. Bei Annahme des $\overline{\text{INT}}$-Signals wird ein besonderer M1-Zyklus mit $\overline{\text{IORQ}}$ aktiv und M1 aktiv erzeugt. Die anfordernde Schaltung kann nun ihren 8-bit-Interruptvektor auf den Datenbus geben. Zur Sicherstellung der Durchschaltzeit der Prioritätssignale in der Interruptkette werden von der CPU automatisch zwei zusätzliche WAIT-Zustände eingefügt.

Ein *nichtmaskierbarer Interrupt* wird durch das NMI-Signal angezeigt. Es wird ebenfalls bei der steigenden Flanke des letzten Taktes abgefragt, kann aber nicht softwaremäßig abgewiesen werden. Übergeordnet ist nur das $\overline{\text{BUSRQ}}$-Signal. Die CPU reagiert ähnlich wie bei einer normalen Speicherleseoperation. Es wird jedoch der Inhalt des Datenbusses ignoriert und ein RESTART bei ØØ66 H eingeführt.

Während des *HALT-Signals* führt die CPU Leerbefehle aus. Dadurch werden alle Speicher wieder aufgefrischt. Die zwei Interrupt-Eingänge werden dabei an jeder ansteigenden Taktflanke abgefragt. So ist es möglich, auch beim HALT-Signal Interrupt auszuführen.

8.6.3.3. Verschiedene Darstellungen des Befehlssatzes

Die Vielzahl der Befehle des U 880 macht eine übersichtliche Darstellung schwierig. Insgesamt existieren 696 unterschiedliche Operationskodes, wobei die numerisch einzusetzenden Zahlenwerte, wie direkte Adressen, relative Adresssen, Eingabe fixer Zahlen usw., als jeweils nur ein Operationskode gezählt werden. In dieser Komplexität arbeitet die CPU andererseits auch nicht die Befehle ab. Dies erfolgt hierarchisch geordnet, wobei sich Teilzweige an verschiedenen Stellen gleichartig wiederholen.

In diesem Sinne wird meist davon gesprochen, daß die CPU 150 Befehlsgruppen abarbeiten kann. Aber selbst diese Zahl ist für eine übersichtliche Darstellung immer noch zu groß. Durch geschickte weitere Zusammenfassung (vgl. Tab. 8.6.7) entstehen 36 Befehlstypen, von denen einige eine Vielzahl von Varianten besitzen (8.6.3.4.).

Die *Darstellung aller Befehle* kann auf *drei Arten* erfolgen, als

- Bit-Kombination ØØ11 1ØØØ B
- Hex-Zahlen 78 H
- Mnemonics LD, A, B.

Die Bit-Kombination ist ähnlich wie die Mnemonics zur Verallgemeinerung geeignet. Ø1 r r′ entspricht LD r, r′, wobei r und r′ Register A, B, C, D, E, H, L bzw. 3-bit-Zahlen sind.

Bereits aus diesem Beispiel ist zu erkennen, daß die Hex-Zahlen die einfachste Darstellung sind. Hier liegt ihre große Anwendungsbreite begründet.

Die Bit-Kombinationen werden zwar in der CPU verarbeitet, sind aber für die Kontrolle durch den Menschen sehr ungeeignet. Deshalb geht ihre Anwendung auch immer weiter zurück. Wichtig ist aber zu wissen, daß hier leicht die Zusammenfassung von Befehlen wie oben mit r, r′ möglich ist. Schließlich arbeitet nach diesem Prinzip die CPU auch die Befehle (verallgemeinert) ab.

Tabelle 8.6.7. Bedeutung und Werte von Kurzzeichen bei Befehlen des U 880. Die eingeklammerten Bezeichnungen werden zuweilen in der Literatur benutzt. In der Spalte Bit sind die Positionen angegeben, wo die zugehörigen Bit in einem Byte stehen, in der Spalte Byte die Lage in einem (Mehrbyte-) Befehl. Die unteren Tabellen weisen die Bedeutung je nach Belegung der Bit aus. Für d gilt ØØ H bis 7F H $\triangleq$ Ø D bis 127 D; 8Ø H bis ØF FH $\triangleq$ — 128 D bis — 1 D.

Zeichen	Beschreibung	Wert	Bit	Byte
b	Bit-Position im Bereich von Ø—7	Ø—7	3 ÷ 5	letztes
c	Bedingung für Sprünge (2 bit)		3 + 4	I
cc	Bedingung für Sprünge und Subroutinen		3 ÷ 5	I
d	Distanz im Speicherregister IX, IY	—128 ÷ 127	Ø ÷ 7	III
dd	Registerpaare BC, DE, HL, SP		4 ÷ 5	I/II
e	relative Sprungweite (d + 2)	—126 ÷ 129	Ø ÷ 7	II
(f)	siehe m			
ii	Indexregister IX oder IY		5	I
M	= (HL)			
m	r, (HL), (IX + d), (IY + d)			
n	Zahlenwert	Ø ÷ 255	Ø ÷ 7	letztes
nn	Zahlenwert	Ø ÷ 65553	Ø ÷ 7	beide letzten
p	wie t gebraucht			
pp	Registerpaare BC, DE, IX oder IY, SP		4 + 5	II
qq	Registerpaare BC, DE, HL, AF		4 + 5	I
r	Register A, B, C, D, E, H, L (n bzw. (HL)		3 ÷ 5	I/II
r′	wie r, jedoch speziell als Quelle		Ø ÷ 2	I/II
(rr)	wie pp, jedoch nur IY, dann pp nur IX		4 + 5	II
s	r, n, (HL), (IX + d), (IY + d)			
(ss)	wie dd (Source, Drain)			
t	Sprungadressen		3 ÷ 5	I
tt	Registerpaare BC, DE		4	I

( ) bedeutet: Inhalt des Registers bestimmt die Adresse der Speicherzelle

Bit	ii	tt
Ø	IX	BC
1	IY	DE

bit	c	dd	pp	qq
ØØ	NZ	BC	BC	BC
Ø1	Z	DE	DE	DE
1Ø	NC	HL	IX/IY	HL
11	C	SP	SP	AF

Bit	cc	b	r	t
ØØØ	NZ	Ø	B	ØØH = ØD
ØØ1	Z	1	C	Ø8H = 8D
Ø1Ø	NC	2	D	1ØH = 16D
Ø11	C	3	E	18H = 24D
1ØØ	PO	4	H	2ØH = 32D
1Ø1	PE	5	L	28H = 4ØD
11Ø	P	6	n/(HL)	3ØH = 48D
111	M	7	A	38H = 56D

Die *Mnemonics* sind die für den Menschen besonders übersichtliche Darstellung, da sie nach einiger Übung schnell erlernt werden können, insbesondere dann, wenn eine Gruppenzusammenfassung gemäß z.B. LD r, r′ erfolgt. Ihre Umsetzung in den für die CPU notwendigen binären Kode verlangt jedoch ein spezielles, nicht ganz einfaches Programm (mindestens 8 Kbyte). Hierzu wird mehr im Abschnitt 8.8.3. ausgeführt.

Die *Darstellung der Befehle* als Tabelle erfolgt meist auf *dreifache Weise*:

- alphabetisch nach den Mnemonics
- geordnet nach Befehlsgruppen
- geordnet nach Hex-Zahlen.

Die alphabetische Darstellung gibt unter Anwendung der Kürzungen den schnellsten Gesamtüberblick zu den Befehlen. Sie wird hier im folgenden zur Kurzbeschreibung der Befehle herangezogen.

Die nach Befehlsgruppen geordnete Darstellung ist beim Schreiben eines Programms vorteilhaft, da sie am schnellsten alle Möglichkeiten in der Befehlsgruppe offenbart und unmittelbar zum Hex-Kode führt. Sie wird hier in der einfachsten Form als Tabellen vorgestellt. In handlichen Sammlungen ist jeder Befehlstyp in einer Befehlsgruppe genau nach Funktion, Setzen und Rücksetzen von Flags usw. beschrieben.

Die Ordnung nach Hex-Zahlen hat beim Testen und Erproben eines Programms große Bedeutung. Hier sind in der Regel nur die Hex-Zahlen bekannt, und ihre Bedeutung muß aus den Tabellen entnommen werden. Auch sie sollen hier vorgestellt werden.

Über die genannten Darstellungen hinaus existieren noch weitere. Es seien hier nur die Referenzen zu anderen Prozessortypen beim U 880, also vor allem zum 8080, 8085, U 808, 8008 und die Gliederung nach Takt- bzw. Maschinenzyklen genannt.

8.6.3.4. Alphabetische Ordnung der Befehlsgruppen

ADC: Inhalt eines Quellregisters bzw. ein Zahlenwert wird zum Zielregister mit Berücksichtigung des Carry-Flag addiert. Es existieren 8-bit-Operationen ADC A, s und 16-bit-Operationen ADC HL, dd.

ADD: Addition ohne Carry-Flag. Es existieren 8-bit-Operationen ADD A, s und 16-bit-Operationen ADD HL, dd sowie ADD ii, rr.

AND s: Der Inhalt des Akkumulators (A) wird mit dem durch s bestimmten Speicherinhalt durch die AND-Operation bitweise verknüpft, und danach steht das Ergebnis in A.

BIT b, m: Es wird das bit b des durch m bestimmten Speicherinhalts getestet und das Komplement in das Z-Flag geladen.

CALL: Aufruf einer Subroutine mit der Adresse nn. Der Aufruf kann bedingt (CALL cc, nn) und unbedingt (CALL nn) erfolgen (Bild 8.6.11).

CCF: Es wird das C-Flag komplementiert.

CMP s (manchmal auch CP): Der durch s bestimmte Inhalt wird mit dem Inhalt von A ohne Änderung der Inhalte verglichen. Spezialisierungen der CMP-Befehle sind die Blockvergleichsbefehle CPI, CPD, CPIR, CPDR. Sie sind für den Vergleich in Speicherblöcken vorgesehen. Der gegenüber A zu vergleichende Speicherblock reicht von den Adressen HL bis HL $\pm$ BC. + gilt für Inkrement (CPI, CPIR) und — für Dekrement (CPD, CPDR). Bei den Befehlen mit R (run) am Ende wird dies fortlaufend (sonst einmalig) getan, bis zum Ende (BC = 0) oder bis A = (HL) gilt. Schematisch gilt für diese Befehle also Bild 8.6.17. Nach jedem Vergleich werden die Flags wie bei CMP s gesetzt.

CPL (Complement): Der Inhalt von A wird bitweise negiert.

DAA: Es wird der Inhalt des Akkumulators A nach einer Addition bzw. Subtraktion von BCD-Zahlen so verändert, daß die richtige BCD-Zahl entsteht.

DEC: Es wird der Inhalt eines Registers DEC m bzw. von Registerpaaren (16 bit) DEC ss, DEC ii um 1 erniedrigt.

DI (disable interrupt): Das maskierte Interrupt wird durch Rücksetzen des Interrupt-Freigabe-Flipflops gesperrt.

DJNZ e: Von Register B wird der Wert um 1 vermindert (dekrementiert). Wenn dabei ein Wert ungleich Null (non zero) erhalten wird, erfolgt ein relativer Sprung (jump) um den relativen Wert e — 2 (Bild 8.6.17d)).

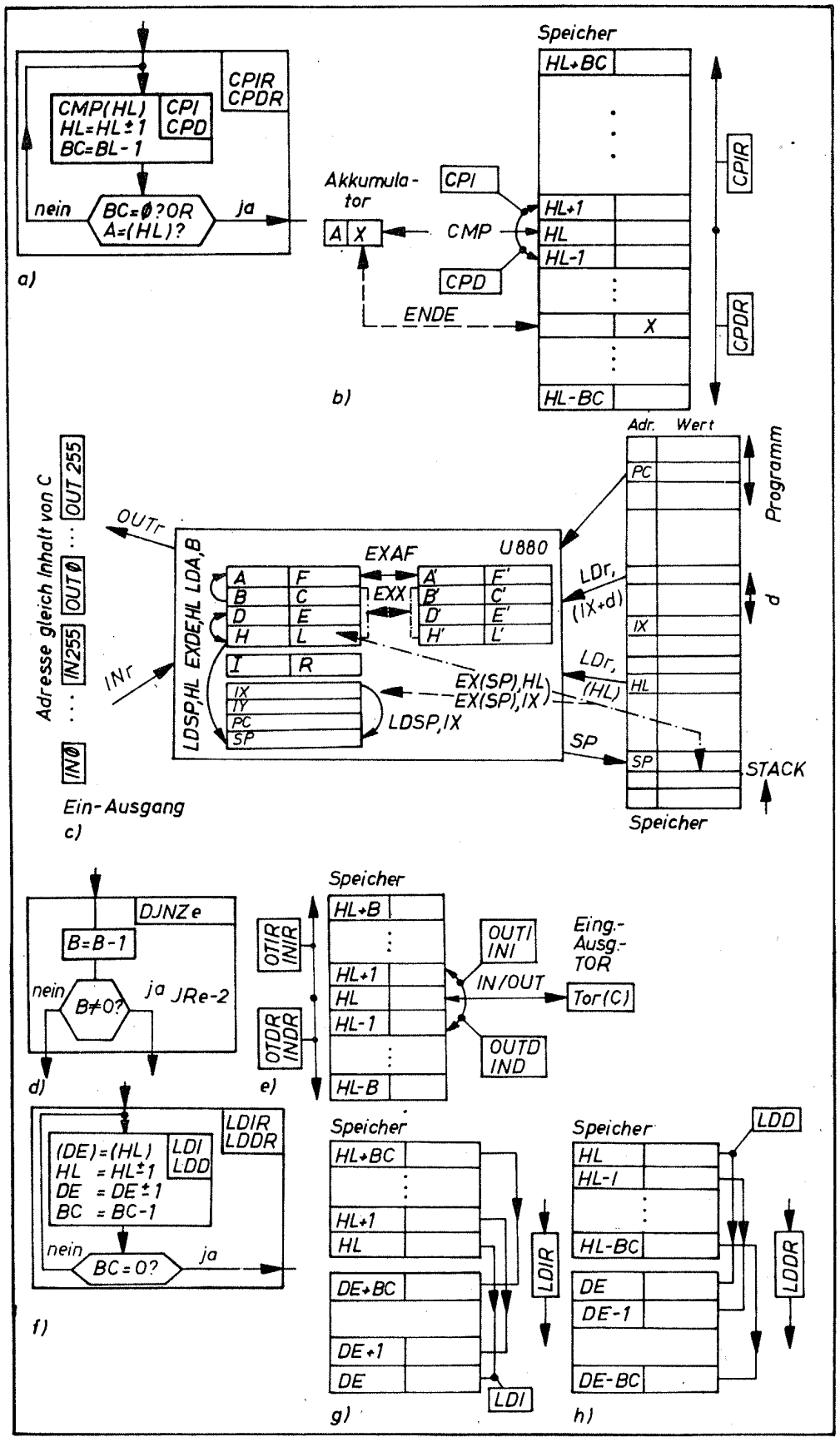


Bild 8.6.17. Schematische Beschreibung einiger Befehlsgruppen

a) und b) CPI, CPD, CPIR, CPDR
c) EX und LD
d) DJNZe
e) INI, IND, OUTI, OUTD, INIR, INDR, OTIR, OTDR
f) bis h) LDI, LDD, LDIR, LDDR

EI (enable interrupt): Das maskierte Interrupt wird durch Setzen des Interrupt-Freigabe-Flipflops freigegeben.

EX: Austausch der Inhalte von Registern (Bild 8.6.17c)). Es existieren die Operationen EXX (Austausch BC DE HL ↔ B′C′ D′E′ H′L′); EX AF (AF → A′F′); EX DE, HL; EX (SP), HL; EX (SP), ii. Bei den letzten beiden Operationen geht der Inhalt der durch SP gekennzeichneten Speicherzelle in den L- (Low-) und der von SP + 1 in den H- (High-) Teil. SP behält seinen Wert.

HALT: Es werden NOP- (no operation) Befehle mit Refresh ausgeführt, bis ein Interrupt oder RESET aktiv wird.

IM: Festlegung der Interruptart: IMØ: Restart-Adresse wird von peripherer Einheit gegeben. IM1: Restart bei ØØ38 H. IM2: legt in direktem Sprung auf Interrupt-Routine fest.

IN: Eingabe von Daten in ein Register:
IN n: A wird mit dem Inhalt des Tores n geladen.
IN r: r wird mit dem Inhalt des Tores (C) geladen.
INF: Flag-Register wird mit dem Inhalt des Tores gemäß Adresse aus (C) geladen.
Weiter existieren Eingaben in einen Speicherblock, der bei HL beginnt und die Länge B hat. Mit jeder Eingabe kann ein Dekrementieren von HL (IND; INDR) bzw. ein Inkrementieren von HL (INI; INIR) erfolgen. Bei den Befehlen mit R (run) am Ende geschieht dies automatisch, während sonst der Befehl wiederholt werden mnß (Bild 8.6.17e)).

INC: Erhöhen des Inhalts von Registern um 1. Es existieren die Operationen INC m, INC ii, INC ss.

JP; JR: Absolute und relative (JUMP) Sprünge mit und ohne Bedingung (vgl. Bild 8.6.11). Es existieren die Operationen JP (HL); JP (ii); JP cc, nn; JP nn; JR e; JR c, e (vgl. auch DJNZ e).

LD: Sehr umfangreiche, ansonsten aber übersichtliche Gruppe von Ladebefehlen mit einem oder zwei Byte. Für Details siehe Bild 8.6.19 und Bild 8.6.8 bzw. 8.6.12c).
Zu dieser Gruppe gehören die Blocktransferbefehle. Hierbei wird ein Speicherbereich HL ± BC in einen Speicherbereich DE ± BC übertragen, von HL und DE aus erfolgt also gemeinsam ein Inkrementieren (LDI; LDIR) bzw. Dekrementieren (LDD; LDDR). Bei den Befehlen mit R (run) am Ende erfolgt dies fortlaufend, sonst einzeln mit jedem Befehl (Bild 8.6.17f) bis h)).

NEG: Im Akkumulator wird das Vorzeichen gewechselt, A = Ø − A.

NOP: Die CPU führt vier Taktzyklen, keine Operation außer Refresh aus.

OR s: Es wird der Inhalt des Akkumulators (A) mit dem durch s bestimmten Speicherinhalt mittels der OR-Operation bitweise verknüpft. Das Ergebnis steht danach in A.

OUT: Ausgabe von Daten eines Registers:
OUT n: Akkumulatorinhalt wird nach Tor n geladen.
OUT r: Akkumulatorinhalt wird zu dem Tor geladen, das durch Register C gegeben ist.
Weiter existiert die Ausgabe eines Speicherblocks, der bei HL beginnt und die Länge B hat. Mit jeder Ausgabe erfolgt ein Dekrementieren von HL (OUTD, OTDR) bzw. ein Inkrementieren von HL (OUTI, OTIR). Bei den Befehlen mit R (run) am Ende erfolgt die automatische Ausgabe des Blockes, sonst nur die einer Speicherzelle je Befehl.

POP: Herausholen von 16-bit-Daten bzw. Adressen aus dem Stack. Da der Stack beim U 880 abwärts zählt, gilt Bild 8.6.18a). Der Stackpointer ist nach dem Befehl um 2 erhöht.

PUSH: Retten von 16-bit-Daten bzw. Adressen in den Stack (Bild 8.6.18b)). Der Stackpointer ist nach dem Befehl um 2 erniedrigt. (Beim U 880 zählt Stack abwärts).

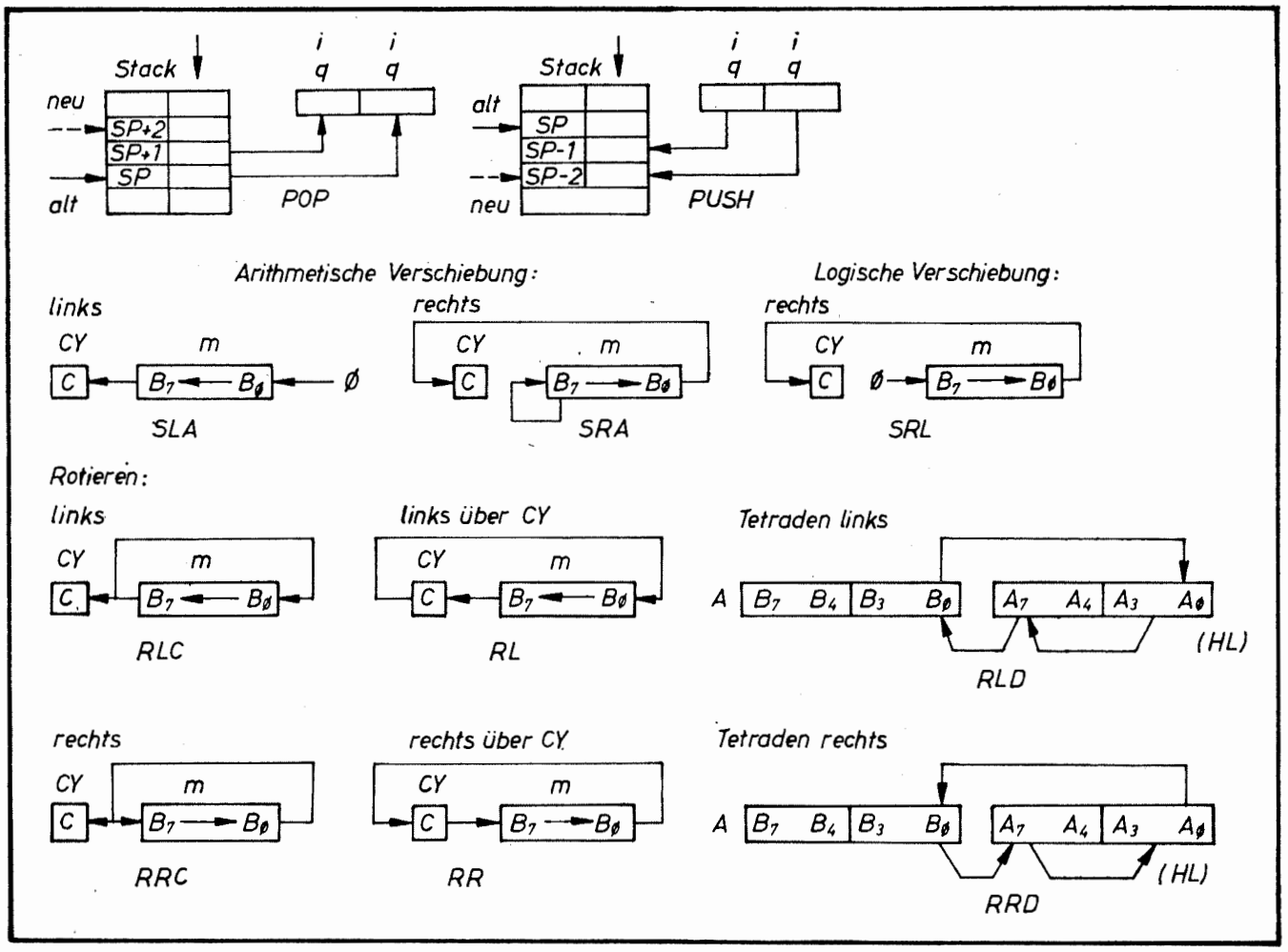


Bild 8.6.18. Bildliche Darstellung der Befehle POP, PUSH, Schieben und Rotieren

RES b, m: Das bit b des Registers r bzw. der Speicherzelle (HL); (IX + d); (IY + d) wird Null gesetzt.
RET: Rücksprungbefehl (return), der unbedingt oder bedingt, RET cc (= R cc), erfolgen kann. Der Programmzähler wird dabei aus dem Stack auf die Adresse zu Beginn des entsprechenden CALL geladen. Rücksprünge aus einem Interrupt erfolgen mit RETI, bei nichtmaskiertem Interrupt mit RETN.
RL + RR (rotate left or right): Links bzw. rechts rotieren: Gemäß Bild 8.6.18 sind mehrere Varianten zu unterscheiden:

- einfaches bitweises Rotieren über bzw. ohne das C-Flag bei den Registern r und den Speichern (HL), (IX + d) und (IY + d),
- 4-bitweises Rotieren zwischen A und (HL).

RST (restart): Spezieller CALL-Befehl zu festen Adressen (vgl. Tab. 8.6.7).
SBC (substract): Inhalt eines Quellregisters bzw. ein Zahlenwert wird zum Zielregister mit Berücksichtigung des Carry-Flag (Borgen) subtrahiert. Es existieren 8-bit-Operationen SBC A, s und 16-bit-Operationen SBC HL, ss.
SCF: Das Carry-Flag wird auf 1 gesetzt.
SET b, m: Das bit b des Registers r bzw. der Speicherzelle (HL); (IX + d); (IY + d) wird auf 1 gesetzt.
SL + SR (shift left or right): Links bzw. rechts schieben: Gemäß Bild 8.6.18 sind arithmetisches (SLA, SRA) und logisches (SRL) Schieben zu unterscheiden.
SUB s: Der Inhalt des Registers r, die Zahl bzw. der Inhalt des durch (HL); (IX + d); (IY + d) bestimmten Speichers wird von Akkumulator A subtrahiert (nur 8-bit-Operationen).

Tabelle 8.6.8. Zusammenfassung der Flag-Operationen

↕ : Flag wird entsprechend dem Ergebnis verändert
. : Flag wird nicht verändert
0: Flag wird gelöscht
1: Flag wird gesetzt
X: Flag ist unbestimmt

Befehl	Flags						Bemerkungen
	C	Z	P/V	S	N	H	
ADD s, ADC s	↕	↕	V	↕	0	↕	8-bit-Addition oder Addition mit Übertrag
SUB s, SBC s, CMP s, NEG	↕	↕	V	↕	1	↕	8-bit-Subtraktion, Subtraktion mit Übertrag, Vergleich und Negation des Akkumulators
AND s	0	↕	P	↕	0	1	
OR s, XOR s	0	↕	P	↕	0	0	logische Operationen
INC d	.	↕	V	↕	0	↕	8-bit-Erhöhung
DEC d	.	↕	V	↕	1	↕	8-bit-Erniedrigung
ADD HL, ss	↕	.	.	.	0	X	16-bit-Addition
ADC HL, ss	↕	↕	V	↕	0	X	16-bit-Addition mit Übertrag
SBC HL, ss	↕	↕	V	↕	1	X	16-bit-Subtraktion mit Übertrag
RLA, RLCA, RRA, RRCA	↕	.	.	.	0	0	zyklische Verschiebung — Akkumulator
RL s, RLC s, RR s, RRC s	↕	↕	P	↕	0	0	zyklische Verschiebung — Speicherplatz s
SLA s, SRA s, SRL s	↕	↕	P	↕	0	0	Verschiebung — Speicherplatz s
RLD, RRD	.	↕	P	↕	0	0	zyklische Verschiebung — Zahl links und rechts
DAA	↕	↕	P	↕	.	↕	Dezimalkorrektur — Akkumulator
CPL	.	.	.	.	1	1	Komplement des Akkumulators
SCF	1	.	.	.	0	0	Setzen des Übertrags
CCF	↕	.	.	.	0	X	Komplement des Übertrags
IN r, INF	.	↕	P	↕	0	0	Eingabe, indirekte Registeradresse
INI, IND, OUTI, OUTD	.	↕	X	X	1	X	Blockein- und -ausgabe, Z = 0, wenn B = 0, sonst Z = 1
INIR, INDR, OTIR, OTDR	.	1	X	X	1	X	Z = 0, wenn B = 0, sonst Z = 1
LDI, LDD	.	X	↕	X	0	0	Blocktransferbefehle
LDIR, LDDR	.	X	0	X	0	0	P/V = 1, wenn BC = 0, sonst P/V = 0
CPI, CPIR, CPD, CPDR	.	↕	↕	X	1	X	Blocksuchbefehle, Z = 1, wenn A = (HL), sonst Z = 0; P/V = 1, wenn BC ≠ 0, sonst P/V ≠ 0
LD A,I; LD A,R	.	↕	IFF2	↕	0	0	Inhalt des Interrupt-Annahmeflipflops 2 (IFF2) ins P/V-Flag überführt
BIT b,s	.	↕	X	X	0	1	Zustand des Bits b im Speicherplatz s ins Z-Flag überführt

XOR s: Der Inhalt des Akkumulators (A) wird mit dem durch s bestimmten Wert durch die XOR-Operation (Exclusiv-OR) bitweise verknüpft, und danach steht das Ergebnis im Akkumulator.
FLAG: Tab. 8.6.8 faßt die Beeinflussung der Flags durch die verschiedenen Operationen zusammen.

8.6.3.5. Ordnung nach Befehlsgruppen

Die vorangegangenen Betrachtungen haben gezeigt, daß der U 880 eine große Fülle von Befehlen besitzt. Beim Programmieren ist deshalb dreierlei wichtig:

ein guter Überblick zu den Befehlstypen,
eine Kontrolle über die zugelassenen Parameter eines Befehls und der bequeme Zugang zum einzugebenden Hex-Kode.

Dies leisten recht gut spezielle Tabellen, die die Befehle nach bestimmten Gruppen ordnen, die Parameter matrixförmig erfassen und in den Feldern den Hex-Kode enthalten. Hier wurde die übliche Form noch einmal so verdichtet, daß alle Möglichkeiten *auf drei Druckseiten* geordnet *nach folgenden Gruppen* und Befehlstypen stehen:

- Transportbefehle (Bild 8.6.19): 8-bit-LD; 16-bit-LD; Austausch und Blocktransfer;
- Arithmetik- und Logikbefehl (Bild 8.6.20): 8-bit; 16-bit; Shift, Rotate, Blockvergleich und DAA, CPL, NEG, CCF, SCF;
- restliche Befehle (Bild 8.6.21): BIT, JP, CALL, RET, RST, IN, OUT (auch Block), CPU.

Die Erklärung der Befehle konnte dabei gekürzt werden, da sie im vorigen Abschnitt gegeben ist.

8.6.3.6. Ordnung nach Hex-Kode

Die Ordnung der Befehle nach dem Hex-Kode muß hierarchisch erfolgen, und zwar je nach der Anzahl der Bytes, die zu einem Befehl notwendig ist. Das erste Byte ist in Bild 8.6.22 oben dargestellt. Die jeweils links unten stehende Anzahl von Rastern zeigt an, wieviel Bytes (0, 1 oder 2) noch folgen können. Dies sind stets feste Zahlenwerte n bzw. nn. Eine Besonderheit stellen die Kodeanfänge CB, DD, ED, FD dar. Für sie existieren untersetzte Befehlskodes, die in den weiteren Bildern erläutert sind.

Für den Kode CB existiert nur jeweils ein fortsetzendes Byte, welches dann Rotations-, Schiebe- bzw. Bit-Befehl bewirkt. Auf die Besonderheit der Kodes CB 3Ø bis CB 37 wird im folgenden Abschnitt eingegangen.

Die möglichen Fortsetzungen des Kodebeginns ED zeigt Bild 8.6.23 oben. Hier fällt auf, daß nur wenige Felder belegt sind. Wie sich die CPU bei den unbelegten Feldern verhält, ist unbekannt. Recht vielfältig sind die Kodeanfänge DD und FD, welche die Register IX und IY betreffen. (Alle hier mit ☐ bezeichneten Felder siehe nächster Abschnitt.) Für den Anfang DD CB d bzw. FD CB d gibt es wiederum eine weitere Untersetzung, welche Rotations-, Schiebe- und Bit-Befehle dieser Register betrifft (Bild 8.6.24). Die wenigen offiziellen Befehle sind hier dick umrandet.

8.6.3.7. Zusätzliche Befehle

Die offiziell nicht belegten Kodewörter waren Anlaß für einige Programmierer, herauszufinden, was hier geschieht. Dabei wurden vier neue Befehlstypen gefunden, die im Bild 8.6.25 zusammengestellt sind.

Im Teilbild a) ist das IX- bzw. IY- in zwei 8-*bit-Register iØ bzw. Øi*, d. h. HIX, HIY bzw. LIX, LIY, geteilt. Für diese Registerteile existieren mehrere 8-bit-Operationen. Sie sind in Bild 8.6.23 unten eingetragen. Es ist ersichtlich, daß dadurch diese beiden Registerpaare gleichwertig mit den Registerpaaren BC, ED und HL werden.

▶

Bild 8.6.19. Transportbefehle

8-bit-LD			Quelle															
			implizit		Register							Register indirekt		indiziert		erweiterte Adresse	Direkt-wert	
			I	R	A	B	C	D	E	H	L	M	(BC) (DE)	(IX + d)	(IY + d)	(nn)	n	
Ziel	Register	A	ED 57	ED 5F	7F	78	79	7A	7B	7C	7D	7E	0A 1A	DD 7E d	FD 7E d	3A n n	3E n	
		B			47	40	41	42	43	44	45	46		DD 46 d	FD 46 d		06 n	
		C			4F	48	49	4A	4B	4C	4D	4E		DD 4E d	FD 4E d		0E n	
		D			57	50	51	52	53	54	55	56		DD 56 d	FD 56 d		16 n	
		E			5F	58	59	5A	5B	5C	5D	5E		DD 5E d	FD 5E d		1E n	
		H			67	60	61	62	63	64	65	66		DD 66 d	FD 66 d		26 n	
		L			6F	68	69	6A	6B	6C	6D	6E		DD 6E d	FD 6E d		2E n	
	Register indirekt	M			77	70	71	72	73	74	75						36 n	
		(BC)			02													
		(DE)			12													
	indiziert	(IX + d)			DD 77 d	DD 70 d	DD 71 d	DD 72 d	DD 73 d	DD 74 d	DD 75 d						DD 36 d n	
		(IY + d)			FD 77 d	FD 70 d	FD 71 d	FD 72 d	FD 73 d	FD 74 d	FD 75 d						FD 36 d n	
	erweiterte Adresse	(nn)			32 n n													
	implizit	I			ED 47													
		R			ED 4F													

Austauschbefehle	
EXX	D9
EXAF	08
EXDE	EB
EX(SP),HL	E3
EX(SP),IX	DD E3
EX(SP),IY	FD E3

Blocktransfer	
LDI	ED A0
LDIR	ED B0
LDD	ED A8
LDDR	ED B8

16 bit LD			Quelle									
			Register							erweit. Direkt-wert	erweit. Adresse	indir. Reg. Adr.
			AF	BC	DE	HL	SP	IX	IY	nn	(nn)	(SP)
Ziel	Register	AF										F1
		BC								01 n n	ED 4B n n	C1
		DE								11 n n	ED 5B n n	D1
		HL								21 n n	2A n n	E1
		SP				F9		DD F9	FD F9	31 n n	ED 7B n n	
		IX								DD 21 n n	DD 2A n n	DD E1
		IY								FD 21 n n	FD 2A n n	FD E1
	erweit. Adr.	(nn)		ED 43 n n	ED 53 n n	22 n n	ED 73 n n	DD 22 n n	FD 22 n n			POP-Befehle
PUSH-Befehle	indir. Reg. Adr.	(SP)	F5	C5	D5	E5		DD E5	FD E5			

Achtung! Die PUSH- u. POP-Befehle verändern den SP-Inhalt nach jeder Ausführung

8-bit-arithm. logisch	Quelle										
	Register-Adressierung							indir. Reg.-Adr.	indizierte Adressierung		Direkt-wert
	A	B	C	D	E	H	L	M	(IX + d)	(IY + d)	n
Addition ADD	87	80	81	82	83	84	85	86	DD 86 d	FD 86 d	C6 n
Addition mit Übertrag ADC	8F	88	89	8A	8B	8C	8D	8E	DD 8E d	FD 8E d	CE n
Subtraktion SUB	97	90	91	92	93	94	95	96	DD 96 d	FD 96 d	D6 n
Subtraktion mit Übertrag SBC	9F	98	99	9A	9B	9C	9D	9E	DD 9E d	FD 9E d	DE n
„UND" AND	A7	A0	A1	A2	A3	A4	A5	A6	DD A6 d	FD A6 d	E6 n
exklusives oder XOR	AF	A8	A9	AA	AB	AC	AD	AE	DD AE d	FD AE d	EE n
oder OR	B7	B0	B1	B2	B3	B4	B5	B6	DD B6 d	FD B6 d	F6 n
Vergleich CP	BF	B8	B9	BA	BB	BC	BD	BE	DD BE d	FD BE d	FE n
Erhöhen um 1 INC	3C	04	0C	14	1C	24	2C	34	DD 34 d	FD 34 d	
Erniedr. um 1 DEC	3D	05	0D	15	1D	25	2D	35	DD 35 d	FD 35 d	

DAA	27
CPL	2F
NEG	ED 44
CCF	3F
SCF	37

Blockvergleich	
CPI	ED A1
CPIR	ED B1
CPD	ED A9
CPDR	ED B9

16-bit-arithm.-logisch			Quelle					
			BC	DE	HL	SP	IX	IY
Ziel	ADD	HL	09	19	29	39		
		IX	DD 09	DD 19		DD 39	DD 29	
		IY	FD 09	FD 19		FD 39		FD 29
	Addition mit Übertrag und Flagsetzer ADC	HL	ED 4A	ED 5A	ED 6A	ED 7A		
	Subtraktion mit Übertrag und Flagsetzer SBC	HL	ED 42	ED 52	ED 62	ED 72		
	Erhöhen um 1 INC		03	13	23	33	DD 23	FD 23
	Erniedrigen um 1 DEC		0B	1B	2B	3B	DD 2B	FD 2B

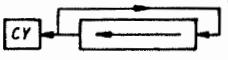
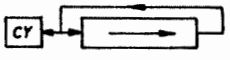
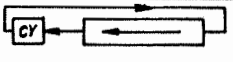
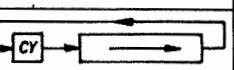
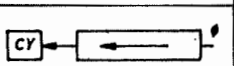
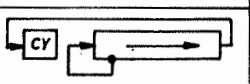
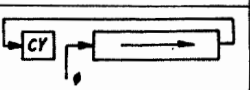
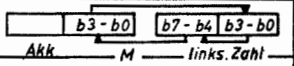
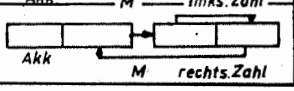
Quelle und Ziel														
	A	B	C	D	E	H	L	M	(IX + d)	(IY + d)		A		
RLC	CB 07	CB 00	CB 01	CB 02	CB 03	CB 04	CB 05	CB 06	DD CB d 06	FD CB d 06	RLCA	07		zykl. Verschiebung links herum RLC
RRC	CB 0F	CB 08	CB 09	CB 0A	CB 0B	CB 0C	CB 0D	CB 0E	DD CB d 0E	FD CB d 0E	RRCA	0F		zykl. Verschiebung rechts herum RRC
RL	CB 17	CB 10	CB 11	CB 12	CB 13	CB 14	CB 15	CB 16	DD CB d 16	FD CB d 16	RLA	17		zykl. Verschiebung links RL
RR	CB 1F	CB 18	CB 19	CB 1A	CB 1B	CB 1C	CB 1D	CB 1E	DD CB d 1E	FD CB d 1E	RRA	1F		zykl. Verschiebung rechts RR
SLA	CB 27	CB 20	CB 21	CB 22	CB 23	CB 24	CB 25	CB 26	DD CB d 26	FD CB d 26				Verschiebung links, arithm. SLA
SRA	CB 2F	CB 28	CB 29	CB 2A	CB 2B	CB 2C	CB 2D	CB 2E	DD CB d 2E	FD CB d 2E				Verschiebung rechts, arithm. SRA
SRL	CB 3F	CB 38	CB 39	CB 3A	CB 3B	CB 3C	CB 3D	CB 3E	DD CB d 3E	FD CB d 3E				Verschiebung rechts logisch SRL
RLD								ED 6F						zykl. Verschiebung RLD
RRD								ED 67						zykl. Verschiebung RRD

Bild 8.6.20. Verarbeitungsbefehle

	BIT	Registeradressierung A	B	C	D	E	H	L	indir. Reg. (HL)	indiziert (IX + d)	(IY + d)
TEST „BIT"	0	CB 47	CB 40	CB 41	CB 42	CB 43	CB 44	CB 45	CB 46	DD CB d 46	FD CB d 46
	1	CB 4F	CB 48	CB 49	CB 4A	CB 4B	CB 4C	CB 4D	CB 4E	DD CB d 4E	FD CB d 4E
	2	CB 57	CB 50	CB 51	CB 52	CB 53	CB 54	CB 55	CB 56	DD CB d 56	FD CB d 56
	3	CB 5F	CB 58	CB 59	CB 5A	CB 5B	CB 5C	CB 5D	CB 5E	DD CB d 5E	FD CB d 5E
	4	CB 67	CB 60	CB 61	CB 62	CB 63	CB 64	CB 65	CB 66	DD CB d 66	FD CB d 66
	5	CB 6F	CB 68	CB 69	CB 6A	CB 6B	CB 6C	CB 6D	CB 6E	DD CB d 6E	FD CB d 6E
	6	CB 77	CB 70	CB 71	CB 72	CB 73	CB 74	CB 75	CB 76	DD CB d 76	FD CB d 76
	7	CB 7F	CB 78	CB 79	CB 7A	CB 7B	CB 7C	CB 7D	CB 7E	DD CB d 7E	FD CB d 7E
LÖSCHEN „RES"	0	CB 87	CB 80	CB 81	CB 82	CB 83	CB 84	CB 85	CB 86	DD CB d 86	FD CB d 86
	1	CB 8F	CB 88	CB 89	CB 8A	CB 8B	CB 8C	CB 8D	CB 8E	DD CB d 8E	FD CB d 8E
	2	CB 97	CB 90	CB 91	CB 92	CB 93	CB 94	CB 95	CB 96	DD CB d 96	FD CB d 96
	3	CB 9F	CB 98	CB 99	CB 9A	CB 9B	CB 9C	CB 9D	CB 9E	DD CB d 9E	FD CB d 9E
	4	CB A7	CB A0	CB A1	CB A2	CB A3	CB A4	CB A5	CB A6	DD CB d A6	FD CB d A6
	5	CB AF	CB A8	CB A9	CB AA	CB AB	CB AC	CB AD	CB AE	DD CB d AE	FD CB d AE
	6	CB B7	CB B0	CB B1	CB B2	CB B3	CB B4	CB B5	CB B6	DD CB d B6	FD CB d B6
	7	CB BF	CB B8	CB B9	CB BA	CB BB	CB BC	CB BD	CB BE	DD CB d BE	FD CB d BE
BITSETZEN „SET"	0	CB C7	CB C0	CB C1	CB C2	CB C3	CB C4	CB C5	CB C6	DD CB d C6	FD CB d C6
	1	CB CF	CB C8	CB C9	CB CA	CB CB	CB CC	CB CD	CB CE	DD CB d CE	FD CB d CE
	2	CB D7	CB D0	CB D1	CB D2	CB D3	CB D4	CB D5	CB D6	DD CB d D6	FD CB d D6
	3	CB DF	CB D8	CB D9	CB DA	CB DB	CB DC	CB DD	CB DE	DD CB d DE	FD CB d DE
	4	CB E7	CB E0	CB E1	CB E2	CB E3	CB E4	CB E5	CB E6	DD CB d E6	FD CB d E6
	5	CB EF	CB E8	CB E9	CB EA	CB EB	CB EC	CB ED	CB EE	DD CB d EE	FD CB d EE
	6	CB F7	CB F0	CB F1	CB F2	CB F3	CB F4	CB F5	CB F6	DD CB d F6	FD CB d F6
	7	CB FF	CB F8	CB F9	CB FA	CB FB	CB FC	CB FD	CB FE	DD CB d FE	FD CB d FE

Ein-Ausgabe

	IN	OUT
n	DB n	D3 n
A	ED 78	ED 79
B	ED 40	ED 41
C	ED 48	ED 49
D	ED 50	ED 51
E	ED 58	ED 59
H	ED 60	ED 61
L	ED 68	ED 69
F	ED 70	

Restart

RST	HEX	CALL-Adr
0	C7	0000H
8	CF	0008H
16	D7	0010H
24	DF	0018H
32	E7	0020H
40	EF	0028H
48	F7	0030H
56	FF	0038H

Block-Ein-Ausgabe

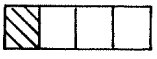
	IN	OUT
I	ED A2	ED A3
IR	ED B2	ED B3
D	ED AA	ED AB
DR	ED BA	ED BB

NOP	00
HALT	76
DI	F3
EI	FB
IM0	ED 46
IM1	ED 56
IM2	ED 5E

Sprung				Bedingung	unbedingt	Übertrag C	kein Übertr. NC	≠ 0 Z	= 0 NZ	P/V = 1 gerade PE	P/V = 0 ungerade PO	S = 1 M	S = 0 P	Reg. B ≠ 0
	„JR" „J"	erweiterter Direktwert	nn	PC + e	C3 n n	DA n n	D2 n n	CA n n	C2 n n	EA n n	E2 n n	FA n n	F2 n n	10 e − 2
	„JR"	relativ	PC + e		18 e − 2	38 e − 2	30 e − 2	28 e − 2	20 e − 2					
	„JM" „J"		M	E9						EC n n	E4 n n	FC n n	F4 n n	
	„JM" „J"	indirekt Register	(IX)	DD E9										
	„JM"		(IY)	FD E9										
	„CALL" „CA"	erweiterter Direktwert	nn	PC + e	CD n n	DC n n	D4 n n	CC n n	C4 n n					
Ernietrigen B, Sp. bei ≠ 0 DJNZ		relative	PC + e											10 e − 2
Rücksprung „RET" „R"		indirekt Register	(SP) (SP + 1)		C9	D8	D0	C8	C0	E8	E0	F8	F0	
Rücksprung v. Interrupt RETI		indirekt Register	(SP) (SP + 1)	ED 4D										
Rücksp. v. nicht mask. Interrupt RETN		indirekt Register	(SP) (SP + 1)	ED 45										

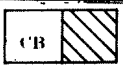
Bild 8.6.21. Restbefehle

1. Tetrade \ 2. Tetrade	0	1	2	3	4	5	6		8	9	A	B	C	D	E	F
0	NOP	LD BC, ,nn	LD(BC) ,A	INC BC	INC B	DEC B	LD B, n	RLCA	EX AF,AF'	ADD HL,BC	LDA, (BC)	DEC BC	INC C	DEC C	LD C n	RRCA
1	DJNZ e	LD DE, ,nn	LD(DE) ,A	INC DE	INC D	DEC D	LD D, n	RLA	JR, e	ADD HL,DE	LDA, (DE)	DEC DE	INC E	DEC E	LD E, n	RRA
2	JR NZ, e	LD HL, nn	LD(nn) ,HL	INC HL	INC H	DEC H	LD H, n	DAA	JR Z, e	ADD HL,HL	LD HL ,(nn)	DEC HL	INC L	DEC L	LD L, n	CPL
3	JRNC, e	LD SP, nn	LD(nn) ,A	INC SP	INC (HL)	DEC (HL)	LD(HL) ,n	SCF	JR C, .e	ADD HL,SP	LD A (n,n)	DEC SP	INC A	DEC A	LD A, n	CCF
4	LD B,B	LD B,C	LD B,D	LD B,E	LD B,H	LD B,L	LD B,(HL)	LD B,A	LD C,B	LD C,C	LD C,D	LD C,E	LD C,H	LD C,L	LD C,(HL)	LD C,A
5	LD D,B	LD D,C	LD D,D	LD D,E	LD D,H	LD D,L	LD D,(HL)	LD D,A	LD E,B	LD E,C	LD E,D	LD E,E	LD E,H	LD E,L	LD E,(HL)	LD E,A
6	LD H,B	LD H,C	LD H,D	LD H,E	LD H,H	LD H,L	LD H,(HL)	LD H,A	LD L,B	LD L.C	LD L,D	LD L,E	LD L,H	LD L,L	LD L,(HL)	LD L,A
7	LD (HL),B	LD (HL),C	LD (HL),D	LD (HL),E	LD (HL),H	LD (HL),L	HALT	LD (HL),A	LD A,B	LD A,C	LD A,D	LD A,E	LD A,H	LD A,L	LD A,(HL)	LD A,A
8	ADD B	ADD C	ADD D	ADD E	ADD H	ADD L	ADD (HL)	ADD A	ADC B	ADC C	ADC D	ADC E	ADC H	ADC L	ADC (HL)	ADC A
9	SUB B	SUB C	SUB D	SUB E	SUB H	SUB L	SUB (HL)	SUB A	SBC B	SBC C	SBC D	SBC E	SBC H	SBC L	SBC (HL)	SBC A
A	AND B	AND C	AND D	AND E	AND H	AND L	AND (HL)	AND A	XOR B	XOR C	XOR D	XOR E	XOR H	XOR L	XOR (HL)	XOR A
B	OR B	OR C	OR D	OR E	OR H	OR L	OR (HL)	OR A	CP B	CP C	CP D	CP E	CP H	CP L	CP (HL)	CP A
C	RET NZ	POP BC	JP NZ,nn	JP nn	CALL NZ,nn	PUSH BC	ADD A,n	RST 0	RET Z	RET	JP Z,nn	Bild 8.6.22a	CALL Z,nn	CALL nn	ADC A,n	RST 8
D	RET NC	POP DE	JP NC,nn	OUT n	CALL NC,nn	PUSH DE	SUB n	RST 10H	RET C	EXX	JP C,nn	IN n	CALL C,nn	Bild 8.6.23	SBC A,n	RST 10H
E	RET PO	POP HL	JP PO,nn	EX (SP),HL	CALL PO,nn	PUSH HL	AND n	RST 20H	RET PE	JP (HL)	JP PE,nn	EX DE,HL	CALL PE,nn		XOR n	RST 28H
F	RET P	POP AF	JP P,nn	DI	CALL P,nn	PUSH AF	OR n	RST 30H	RET M	LD (SP),HL	JP M,nn	EI	CALL M,nn		CP n	RST 38H



1. 2. 3. 4.

1. Tetrade \ 2. Tetrade	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	RLC B	RLC C	RLC D	RLC E	RLC H	RLC L	RLC (HL)	RLC A	RRC B	RRC C	RRC D	RRC E	RRC H	RRC L	RRC (HL)	RRC A
1	RL B	RL C	RL D	RL E	RL H	RL L	RL (HL)	RL A	RR B	RR C	RR D	RR E	RR H	RR L	RR (HL)	RR A
2	SLA B	SLA C	SLA D	SLA E	SLA H	SLA L	SLA (HL)	SLA A	SRA B	SRA C	SRA D	SRA E	SRA H	SRA L	SRA (HL)	SRA A
3	SLI B	SLI C	SLI D	SLI E	SLI H	SLI L	SLI (HL)	SLI A	SRL B	SRL C	SRL D	SRL E	SRL H	SRL L	SRL (HL)	SRL A
4	BIT 0,B	BIT 0,C	BIT 0,D	BIT 0,E	BIT 0,H	BIT 0,L	BIT 0,(HL)	BIT 0,A	BIT 1,B	BIT 1,C	BIT 1,D	BIT 1,E	BIT 1,H	BIT 1,L	BIT 1,(HL)	BIT 1,A
5	BIT 2,B	BIT 2,C	BIT 2,D	BIT 2,E	BIT 2,H	BIT 2,L	BIT 2,(HL)	BIT 2,A	BIT 3,B	BIT 3,C	BIT 3,D	BIT 3,E	BIT 3,H	BIT 3,L	BIT 3,(HL)	BIT 3,A
6	BIT 4,B	BIT 4,C	BIT 4,D	BIT 4,E	BIT 4,H	BIT 4,L	BIT 4,(HL)	BIT 4,A	BIT 5,B	BIT 5,C	BIT 5,D	BIT 5,E	BIT 5,H	BIT 5,L	BIT 5,(HL)	BIT 5,A
7	BIT 6,B	BIT 6,C	BIT 6,D	BIT 6,E	BIT 6,H	BIT 6,L	BIT 6,(HL)	BIT 6,A	BIT 7,B	BIT 7,C	BIT 7,D	BIT 7,E	BIT 7,H	BIT 7,L	BIT 7,(HL)	BIT 7,A
8	RES 0,B	RES 0,C	RES 0,D	RES 0,E	RES 0,H	RES 0,L	RES 0,(HL)	RES 0,A	RES 1,B	RES 1,C	RES 1,D	RES 1,E	RES 1,H	RES 1,L	RES 1,(HL)	RES 1,A
9	RES 2,B	RES 2,C	RES 2,D	RES 2,E	RES 2,H	RES 2,L	RES 2,(HL)	RES 2,A	RES 3,B	RES 3,C	RES 3,D	RES 3,E	RES 3,H	RES 3,L	RES 3,(HL)	RES 3,A
A	RES 4,B	RES 4,C	RES 4,D	RES 4,E	RES 4,H	RES 4,L	RES 4,(HL)	RES 4,A	RES 5,B	RES 5,C	RES 5,D	RES 5,E	RES 5,H	RES 5,L	RES 5,(HL)	RES 5,A
B	RES 6,B	RES 6,C	RES 6,D	RES 6,E	RES 6,H	RES 6,L	RES 6,(HL)	RES 6,A	RES 7,B	RES 7,C	RES 7,D	RES 7,E	RES 7,H	RES 7,L	RES 7,(HL)	RES 7,A
C	SET 0,B	SET 0,C	SET 0,D	SET 0,E	SET 0,H	SET 0,L	SET 0,(HL)	SET 0,A	SET 1,B	SET 1,C	SET 1,D	SET 1,E	SET 1,H	SET 1,L	SET 1,(HL)	SET 1,A
D	SET 2,B	SET 2,C	SET 2,D	SET 2,E	SET 2,H	SET 2,L	SET 2,(HL)	SET 2,A	SET 3,B	SET 3,C	SET 3,D	SET 3,E	SET 3,H	SET 3,L	SET 3,(HL)	SET 3,A
E	SET 4,B	SET 4,C	SET 4,D	SET 4,E	SET 4,H	SET 4,L	SET 4,(HL)	SET 4,A	SET 5,B	SET 5,C	SET 5,D	SET 5,E	SET 5,H	SET 5,L	SET 5,(HL)	SET 5,A
F	SET 6,B	SET 6,C	SET 6,D	SET 6,E	SET 6,H	SET 6,L	SET 6,(HL)	SET 6,A	SET 7,B	SET 7,C	SET 7,D	SET 7,E	SET 7,H	SET 7,L	SET 7,(HL)	SET 7,A



1. 2.

Bild 8.6.22 bis **8.2.24.** Ordnung der Hex-Kodes vom U 880; die mit ☐ versehenen bzw. in Bild 8.6.24 nicht dick umrandeten Befehle sind später gefundene, die in keinem Assembler stehen (vgl. 8.6.3.7.)

1. Tetrade \ 2. Tetrade	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0																
1																
2																
3																
4	IN B	OUT B	SBC BC	LD (nn),BC	NEG	RETN	IM ø	LD I,A	IN C	OUT C	ADC BC	LD BC,(nn)		RETI		LD R,A
5	IN D	OUT D	SBC DE	LD (nn),DE			IM 1	LD A,I	IN E	OUT E	ADC DE	LD DE,(nn)			IM 2	LD A,R
6	IN H	OUT H	SBC HL					RRD	IN L	OUT L	ADC HL					RLD
7			SBC SP	LD (nn),SP					IN A	OUT A	ADC SP	LD SP,(nn)				
8																
9																
A	LDI	CPI	INI	OUTI					LDD	CPD	IND	OUTD				
B	LDIR	CPIR	INIR	OTIR					LDDR	CPDR	INDR	OTDR				
C																
D																
E																
F																

ED			
1.	2.	3.	4.

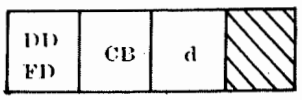
1. Tetrade \ 2. Tetrade	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0										ADD ii,BC						
1										ADD ii,DE						
2		LDii, nn	LD (nn),ii	INC ii	INC iø	DEC iø	LD iø,n			ADD ii,ii	LD ii, (nn)	DEC ii	INC øi	DEC øi	LD øi,n	
3					INC (ii)	DEC (ii)	LD (ii),n			ADD ii,SP						
4					LD B,iø	LD B,øi	LD B,(ii)						LD C,iø	LD C,øi	LD C,(ii)	
5					LD D,iø	LD D,øi	LD D,(ii)						LD E,iø	LD E,øi	LD E,(ii)	
6	LD iø,B	LD iø,C	LD iø,D	LD iø,E	LD iø,ii	LD iø,øi	LD H,(ii)	LD iø,A	LD øi,B	LD øi,C	LD øi,D	LD øi,E	LD øi,iø	LD øi,øi	LD L,(ii)	LD øi,A
7	LD (ii),B	LD (ii),C	LD (ii),D	LD (ii),D	LD (ii),H	LD (ii),L		LD (ii),A					LD A,iø	LD A,øi	LD A,(ii)	
8					ADD A,iø	ADD A,øi	ADD (ii)						ADC A,iø	ADC A,øi	ADC (ii)	
9					SUB iø	SUB øi	SUB (ii)						SBC iø	SBC øi	SBC (ii)	
A					AND iø	AND øi	AND (ii)						XOR iø	XOR øi	XOR (ii)	
B					OR iø	OR øi	OR (ii)						CP iø	CP øi	CP (ii)	
C												Bild 8.6.24				
D																
E		POP ii		EX (SP),ii		PUSH ii				JP (ii)						
F										LD SP,ii						

IX IY		ii =
iø	øi	(IX + d) (IY + d)

DD FD			
1.	2.	3.	4.

Bild 8.6.23

1. Tetrade \ 2. Tetrade	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	rRLC B	rRLC C	rRLC D	rRLC E	rRLC H	rRLC L	RLC ii	rRLC A	rRRC B	rRRC C	rRRC D	rRRC E	rRRC H	rRRC L	RRC ii	rRRC A
1	rRL B	rRL C	rRL D	rRL E	rRL H	rRL L	RL ii	rRL A	rRR B	rRR C	rRR D	rRR E	rRR H	rRR L	RR ii	rRR A
2	rSLA B	rSLA C	rSLA D	rSLA E	rSLA H	rSLA L	SLA ii	rSLA A	rSRA B	rSRA C	rSRA D	rSRA E	rSRA H	rSRA L	SRA ii	rSRA A
3	rSLI B	rSLI C	rSLI D	rSLI E	rSLI H	rSLI L	SLI ii	rSLI A	rSRL B	rSRL C	rSRL D	rSRL E	rSRL H	rSRL L	SRL ii	rSRL A
4							BIT 0,ii								BIT 1,ii	
5							BIT 2,ii								BIT 3,ii	
6							BIT 4,ii								BIT 5,ii	
7							BIT 6,ii								BIT 7,ii	
8	rRES B,0	rRES C,0	rRES D,0	rRES E,0	rRES H,0	rRES L,0	RES 0,ii	rRES A,0	rRES B,1	rRES C,1	rRES D,1	rRES E,1	rRES H,1	rRES L,1	RES 1,ii	rRES A,1
9	rRES B,2	rRES C,2	rRES D,2	rRES E,2	rRES H,2	rRES L,2	RES 2,ii	rRES A,2	rRES B,3	rRES C,3	rRES D,3	rRES E,3	rRES H,3	rRES L,3	RES 3,ii	rRES A,3
A	rRES B,4	rRES C,4	rRES D,4	rRES E,4	rRES H,4	rRES L,4	RES 4,ii	rRES A,4	rRES B,5	rRES C,5	rRES D,5	rRES E,5	rRES H,5	4RES L,5	RES 5,ii	rRES A,5
B	rRES B,6	rRES C,6	rRES D,6	rRES E,6	rRES H,6	rRES L,6	RES 6,ii	rRES A,6	rRES B,7	rRES C,7	rRES D,7	rRES E,7	rRES H,7	rRES L,7	RES 7,ii	rRES A,7
C	rSET B,0	rSET C,0	rSET D,0	rSET E,0	rSET H,0	rSET L,0	SET 0,ii	rSET A,0	rSET B,1	rSET C,1	rSET D,1	rSET E,1	rSET H,1	rSET L,1	SET 1,ii	rSET A,1
D	rSET B,2	rSET C,2	rSET D,2	rSET E,2	rSET H,2	rSET L,2	SET 2,ii	rSET A,2	rSET B,3	rSET C,3	rSET D,3	rSET C,3	rSET H,3	rSET L,3	SET 3,ii	rSET A,3
E	rSET B,4	rSET C,4	rSET D,4	rSET E,4	rSET H,4	rSET L,4	SET 4,ii	rSET A,4	rSET B,5	rSET C,5	rSET D,5	rSET E,5	rSET H,5	rSET L,5	SET 5,ii	rSET A,5
F	rSET B,6	rSET C,6	rSET D,6	rSET E,6	rSET H,6	rSET L,6	SET 6,ii	rSET A,6	rSET B,7	rSET C,7	rSET D,7	rSET E,7	rSET H,7	rSET L,7	SET 7,ii	rSET A,7



ii =
(IX + d)
(IY + d)

Bild 8.6.24

In Teilbild b) ist eine spezielle Linksschiebeoperation dargestellt. Von der Operation SLA unterscheidet sie sich nur dadurch, daß nicht eine Ø, sondern eine 1 in das unterste Bit $B_Ø$ geschoben wird. Deshalb erhielt diese Operation den Namen **SLI** (inverted). Auf die Grundregister und (HL) angewendet, stehen diese Befehle bei CB 3Ø bis CB 37 in Bild 8.6.22 unten und bei DD CB d 36 bzw. FD CB d 36 für (IX + d) bzw. (IY + d) in Bild 8.6.24.

Zwei weitere komplexe Befehle enthält Bild 8.6.25 c) und d). Sie sind dadurch gekennzeichnet, daß in den durch (IX + d) bzw. (IY + d) definierten Speicherzellen zunächst eine Operation (Schieben, Rotieren, Bit-Setzen oder Bit-Rücksetzen) erfolgt und danach ein Transport in eines der Register A, B, C, D, E H, L, durchgeführt wird. So entstehen die Befehle

rSLA, rSRA, rSLI, rSRL
rRLC, rRRC, rRL, rRR
rRES, b und rSET, b.

Diese Befehle können auch so interpretiert werden, daß die auf die Speicherzelle erfolgte Operation in dem jeweiligen Register erfolgte, aber der Rücktransport des dort ursprünglich stehenden Registerinhalts ausblieb.

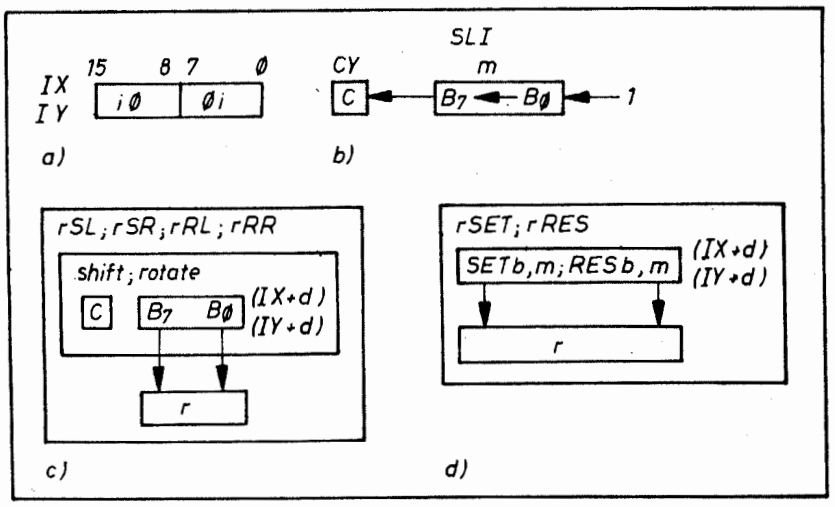


Bild 8.6.25. Zusätzlich beim Z 80 gefundene Befehlstypen

Es sei betont, daß die hier zusätzlich genannten Befehle an erprobten Z 80 funktionierten. Damit besteht aber keine Gewähr dafür, daß sie bei allen funktionieren. Darüber hinaus existiert für diese Befehle z. Z. kein Assembler. Sie können aber, als Hex-Kode eingegeben, zuweilen durchaus vorteilhaft sein.

8.6.4. Interrupt-Prinzipien

In verschiedenen Fällen ist es notwendig, daß ein Mikrorechner auf äußere Einflüsse mehr oder weniger schnell reagiert. Beispiele hierfür sind u. a. Eingabe von Daten bei Anfall, Reagieren auf Überschreiten von Grenzwerten bei Peripheriegeräten, z. B. Temperaturen oder Spannungen, Alarmauslösung usw. Für diese *Forderungen* bestehen

- unterschiedlicher zulässiger Verzug (Dringlichkeit)
- unterschiedliche Wichtigkeit (Priorität)
- unterschiedliche Bedienungsdauer (Bearbeitungszeit).

Deshalb existieren vielfältige Methoden in der Abarbeitung solcher (zufällig) anfallenden Arbeiten. Gemäß Bild 8.6.10 können vier Prinzipien, Eingabe per Programm, periodische Abfrage (polling), Interrupt — als allgemeine Unterbrechung des Hauptprogramms von außen — und DMA — als spezielle Unterbrechung zur Dateneinund -ausgabe —, unterschieden werden. Hier wird nur der Fall der Interrupt-Prinzipien behandelt.

8.6.4.1. Einfaches Interrupt

Es soll nur ein Interrupt möglich sein, dann erfolgt die Abarbeitung gemäß Bild 8.6.26a). Nach dem Eintreffen des Interrupt-Signals wird das laufende Hauptprogramm sobald als möglich unterbrochen und zum Interrupt-Programm übergegangen. Hier müssen die Daten (Register und Flags) zunächst (meist in den Stack) gerettet werden. Erst dann kann das Bedienprogramm für jene Peripherieeinheit begonnen werden, die das Interrupt verlangt. Nach der Bedienung werden die zuvor geretteten Daten wieder zur CPU gebracht, und es erfolgt ein Rücksprung in das Hauptprogramm. Damit dies möglich ist, muß vor Anlauf des Interrupt-Programms der Programmzählerstand gerettet werden. Dies ist die minimale Forderung gemäß

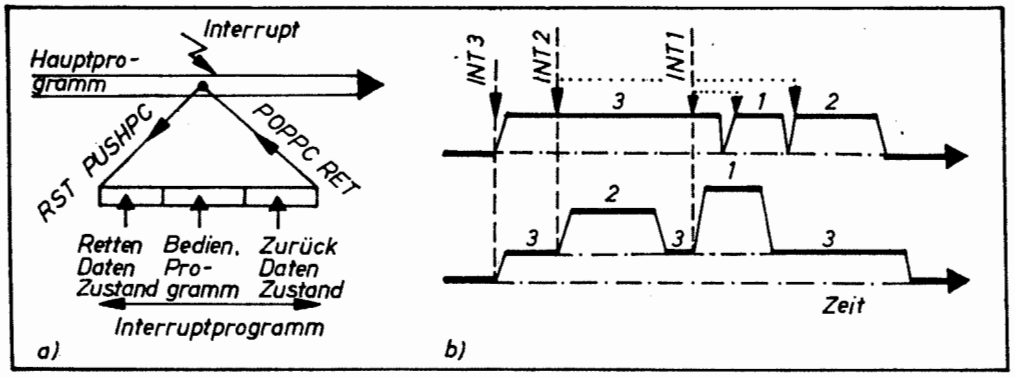


Bild 8.6.26. Prinzipien von Interrupt-Routinen

a) Aufbau und Wirkung eines Interrupt-Programms

b) Abarbeitung von mehreren Interrupt-Forderungen unterschiedlicher Priorität: oben wird jede begonnene zu Ende bearbeitet und dann folgt die mit der jeweils höchsten Priorität. Unten existieren Interrupt-Ebenen entsprechend den Prioritäten

den oben genannten. Es kann aber auch zuvor noch ein bestimmter Teil des Hauptprogramms abgearbeitet werden oder ein Teil des Rettungsprogramms z.B. hardwaremäßig abgewickelt werden. Bezüglich der Programmabwicklung lassen sich z.B. zeitkritische Programmteile dadurch schützen, daß die Interrupt-Möglichkeit solange unterbunden (maskiert) wird, bis ihre Bearbeitung beendet ist. Befehle hierfür sind: Disable und Enable Interrupt (DI und EI). Programmteile, die von diesen Befehlen eingeschlossen werden, sind also vor Unterbrechung (Interrupt) geschützt.

8.6.4.2. Mehrfache Interrupts

In vielen Fällen müssen Interrupts von mehreren Geräten bearbeitet werden. Sie müssen dann in einer Prioritätenfolge angeordnet werden. Sie sei hier mit INT 1 (höchste), INT 2, INT 3 (niedrigste Priorität) bezeichnet. Dann bestehen zwei verschiedene Methoden der Abarbeitung:

die sequentielle (oben in Bild 8.6.26 b))
die verschachtelte oder hierarchische (unten im Bild).

Bei der *sequentiellen Bearbeitung* wird die zuerst vorhandene Interruptquelle sofort bis zu Ende bedient. Dann wird gefragt, welche Interrupt fordernde Quelle die höchste Priorität hat. Sie wird als nächste wieder bis zu Ende bedient usw. Die Methode hat den Vorteil, daß für den Rechner nur zwei Ebenen, die des Hauptprogramms und die des gerade bedienten Interrupts, existieren. Sie hat den Nachteil, daß eine Forderung höherer Priorität zuweilen recht lange warten muß.

Die *verschachtelte Bearbeitung* geschieht in mehreren Ebenen. Die gerade vorhandene höchste Priorität unterbricht alle anderen Programme. Dies hat ein häufigeres Wechseln der Bearbeitung, viel Speicherraum (da bei allen unterbrochenen Routinen die Daten zu retten sind) und meist etwas längere Bearbeitungszeit zur Folge. Dabei wird aber die jeweils wichtigste Interrupt-Quelle sofort bearbeitet.

8.6.4.3. Realisierungstechniken

Wenn mehrere Interrupt-Forderungen vorliegen, muß die CPU entscheiden, welche Geräte Interrupt fordern und welches von ihnen die höchste Priorität hat. Für jedes Interrupt fordernde Gerät muß eine spezielle Interrupt-Routine angesprungen werden, und mittels der Priorität muß entschieden werden, welches Gerät als nächstes bedient wird. Für diesen Prozeß bestehen mehrere Prinzipien.

Tabelle 8.6.9. Beeinflussung der Interrupt-Freigabeflipflops

CPU-Aktionen	Flipflop	Bemerkungen
RESET	IFF1 = IFF2 = Ø	IMØ; I-REG = Ø
DI	IFF1 = IFF2 = Ø	
EI	IFF1 = IFF2 = 1	
LDA, I und LDA, R	unverändert	P/V = IFF2
INT	IFF1 = IFF2 = Ø	
NMI	IFF2 = IFF1; IFF1 = Ø	
RETN und RETI	IFF2 bleibt; IFF1 = IFF2	

Beim *Software-Polling* fragt die CPU alle Geräte in der Reihenfolge ihrer Priorität ab und bedient jenes, welches zuerst mit Interrupt-Forderung gefunden wird.

Beim Hardware-Polling leistet eine spezielle Schaltung, die *Daisy-chain* dasselbe. Hierbei besitzt jedes Gerät ein Flipflop, das bei Interrupt-Forderung gesetzt wird. Die Flipflops werden so über kombinatorische Schaltung verbunden, daß die Geräte entsprechend der Priorität der CPU am „nächsten" sind und so das der CPU zunächst erscheinende Flipflop zur Gerätebedienung herangezogen wird.

Es gibt auch Mikroprozessoren mit mehreren Interrupt-Eingängen, die nach Prioritäten geordnet sind. In diesem Fall spricht man von *Multilevel-Interrupt.*

Ein besonders elegantes Verfahren zur Interrupt-Behandlung liegt dann vor, wenn das nach Priorität ausgewählte Gerät seine für die Bedienung erforderliche Sprungadresse (Interrupt-Vektor) gleichzeitig mit dem Interrupt-Beginn auf den Datenbus gibt. Die CPU kann dann diesen Wert sofort für das Bedienprogramm verwenden.

8.6.4.4. Interrupt-System des U 880

Der Prozessor U 880 kann auf dreierlei Weise bei der Abarbeitung eines Hauptprogramms unterbrochen werden:

- Busanforderung infolge von $\overline{\text{BUSRQ}}$
- Interrupt-Meldung infolge von $\overline{\text{NMI}}$
- Interrupt-Meldung infolge von $\overline{\text{INT}}$.

Die Anforderungen sind bereits in der richtigen Priorität angeordnet. Für jede Anforderung existiert ein Auffangflipflop, so daß ein kurzer Impuls von mindestens 80 ns zum Speichern ausreicht. Lediglich während des Befehls EI wird $\overline{\text{INT}}$ nicht registriert.

Die *Bus-Anforderung* wird nach jedem Maschinenzyklus abgefragt, die Interrupt-Anforderungen nur jeweils am Ende des letzten Maschinenzyklus eines Befehls.

Bei $\overline{\text{BUSRQ}}$ führt die CPU den gerade laufenden Befehl noch zu Ende aus und geht dann in den HOLD-Zustand und den hochohmigen Zustand für die BUS-Leitungen über. Dann wird das Signal $\overline{\text{BUSAK}}$ zur Bestätigung ausgegeben. Erst wenn $\overline{\text{BUSRQ}}$ zurückgenommen wird, kann die CPU wieder wirksam werden und eventuelle Interrupt-Forderungen bearbeiten.

Der *nichtmaskierte Interrupt-Eingang* $\overline{\text{NMI}}$ kann nicht durch die CPU gesperrt werden. Für den Eingang des maskierten Interrupts $\overline{\text{INT}}$ existiert dagegen ein Interrupt-Freigabeflipflop IFF1. Wenn es auf Null zurückgesetzt ist, kann das Interrupt nicht wirksam werden. Dies erfolgt u. a. durch den Befehl DI, während es durch den Befehl EI auf 1 gesetzt wird und so das Interrupt von $\overline{\text{INT}}$ wirksam werden

kann. Da $\overline{\text{NMI}}$ höhere Priorität hat, kann es ebenfalls das IFF1 Null setzen und damit $\overline{\text{INT}}$ sperren. Damit aber gespeichert bleibt, ob $\overline{\text{INT}}$ wirksam wurde, wird mit $\overline{\text{NMI}}$ der Inhalt von IFF1 in ein zweites Flipflop IFF2 zwischengespeichert.

Bei einem $\overline{\text{NMI}}$ geschieht somit folgendes:

IFF2 = IFF1:IFF1 = Ø,
Retten des PC im Stack, höherer und niederer Teil getrennt,
RESTART bei 66 H, dort steht der Beginn der Interrupt-Routine,
Ausführung der Interrupt-Routine, die mit RETN endet,
IFF1 = IFF2,
PC vom Stack zurückholen,
Fortsetzung des Programms.

Das *maskierte Interrupt* $\overline{\text{INT}}$ kann nur wirksam werden, wenn IFF1 gesetzt ist, also weder vorher DI gegeben wurde, noch ein $\overline{\text{NMI}}$ oder $\overline{\text{BUSRQ}}$ vorliegt. Bei Wirksamwerden von $\overline{\text{INT}}$ werden der Programmzähler in den Stack gerettet und beide Flipflops IFF1 und IFF2 rückgesetzt. Dadurch sind weitere $\overline{\text{INT}}$ unzulässig, und es erfolgt der Übergang zur Interrupt-Routine, für die es drei weiter unten zu besprechende Moden gibt. Danach wird das entsprechend (nach Mode und Gerät) erforderliche Bedienungsprogramm angesprungen und abgearbeitet. An dessen Ende steht der Befehl RETI. Durch ihn wird IFF1 = IFF2 (also Null) und der Inhalt des PC vom Stack zurückgeholt, so daß mit der Abarbeitung des Hauptprogramms fortgefahren werden kann.

Folgende drei mögliche Interrupt-Betriebsarten bei INT können vor Eintreffen des Interrupts durch die Befehle IMØ, IM1 bzw. IM2 festgelegt werden.

In der *Betriebsart IMØ* können die 8 RST-Befehle interpretiert werden. Je einer ist einem Interrupt fordernden Gerät zugeordnet und wird unter Beachtung der Priorität gegenüber den anderen mit dem Signal $\overline{\text{INT}}$, insbesondere bezüglich der Bits 3 bis 5, auf den Datenbus gegeben und bei Annahme durch die CPU als Befehl interpretiert. Dadurch gelangt die CPU in die richtige von acht möglichen Serviceroutinen.

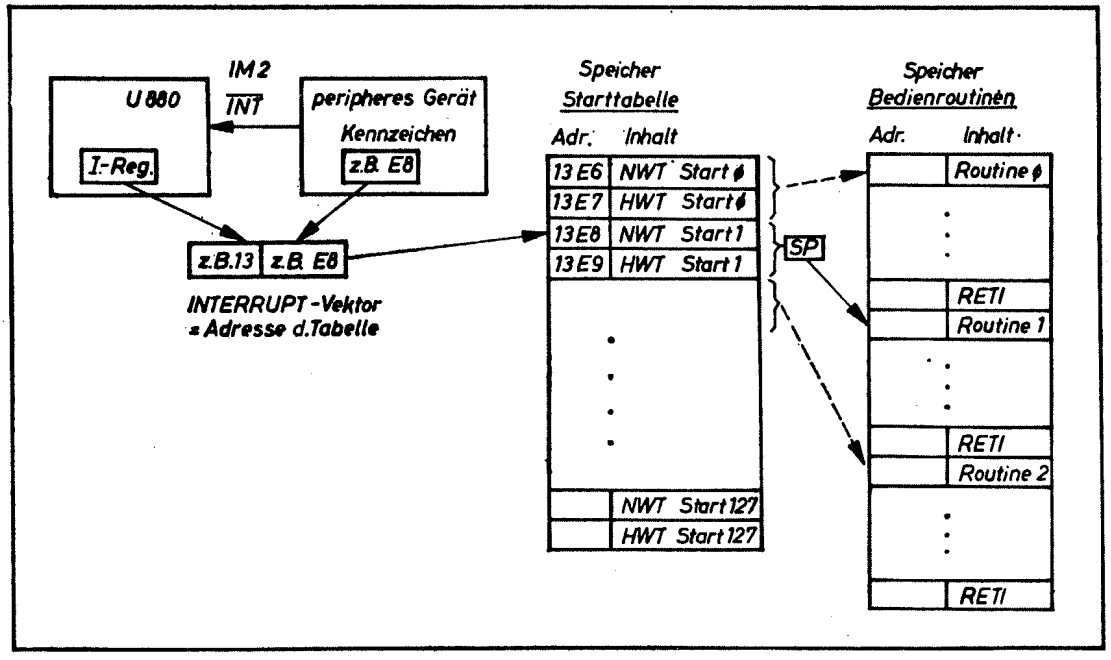


Bild 8.6.27. Prinzip des maskierten Interrupts beim Mode IM2 zur Verwirklichung von 128 anschließbaren Geräten

Die *Betriebsart IM1* ist für ganz einfache Interrupt-Bedingungen vorgesehen. Bei ihr wird unmittelbar in der CPU RST 38 H gebildet. Die *Betriebsart IM2* ist höchst flexibel und ermöglicht viele Varianten. Die Interrupt-Adresse wird hierbei aus zwei Teilen gebildet (Bild 8.6.27):

ADR = I-Register + Interruptvektor.

Mit dem I-Register wird also das höherwertige Adressenbyte für eine Tabelle erzeugt, und das niederwertige kommt vom Interrupt fordernden Gerät auf die Datenleitung. Mit diesem Interrupt-Vektor wird eine Tabelle angesprungen, welche die Startadressen der Bedienroutinen enthält. Wenn alle Möglichkeiten ausgeschöpft werden sollen, können mit dem Byte aus dem Gerät nicht 256, sondern nur 128 Geräte bedient werden, denn die Adressen der Bedienroutinen benötigen 2 Byte für den niederwertigen NWT und den höherwertigen HWT der Adresse. Bei weniger Geräten können im Prinzip kurze Bedienroutinen direkt im Raum der Tabelle von 256 Byte angeordnet werden.

8.6.5. Peripherieschaltungen

Periphere Schaltungen wird als Oberbegriff für alle Schaltungen und Schaltkreise verwendet, welche den Mikroprozessor mit der Umgebung, d. h. mit peripheren Geräten und Einrichtungen, also z.B. Tastatur, Anzeigen, Bildschirm, Speichern (z.B. Bandgeräten, Floppy disc, Lochbandeinheiten), Spezialprozessoren, Meßgeräten, Stelleinrichtungen usw., verbinden. In diesem weiten Sinne sind also durch sie die internen digitalen Daten mit einer entsprechenden Wortbreite zur Eingabe und Ausgabe in bzw. aus entsprechenden analogen und digitalen peripheren Daten zu wandeln. Dabei sind u. a. *folgende Signaländerungen notwendig*:

- Pegel- und/oder Kodeumwandlungen,
- Adressierung des jeweiligen peripheren Geräts,
- Synchronisation der Datenströme über Pufferung, Multiplexer, gemeinsamen Takt, Handshake und/oder Interrupt,
- Umwandlung zwischen parallelen und seriellen Datenströmen,
- Digital-Analog- bzw. Analog-Digital-Wandlung (s. 9.1. und 9.2.).

Die entsprechende Anpassungselektronik heißt *Interface* (Schnittstellenschaltung). Bei Interface im engeren Sinne des Wortes werden analoge Daten nicht zugelassen. Das Grundprinzip einer Interface-Schaltung zeigt Bild 8.6.28. In vielen Fällen erfolgt mit einer Interface-Schaltung der Übergang zwischen den Signalen des Mikroprozessors und einem *Standard-Interface* (Standard-Bus), welche im Abschnitt 8.8. behandelt werden. Hier werden infolgedessen nur die grundsätzlichen Prinzipien und einige Spezialschaltkreise, insbesondere jene, die zur CPU U 880 (Z 80) gehören, behandelt.

Generell werden folgende Peripherieschaltkreise unterschieden:

PIO	paralleles IN-OUT-Port:	parallele Ein-Ausgabe
SIO	serielles IN-OUT-Port:	serielle Ein-Ausgabe
CTC	count and timing-circuit:	Zähler-Zeitgeber-Baustein
DMA	direct-memory-access:	Schaltkreis für direkten Speicherzugriff
USART	universal synchronous asynchronous transmitter:	Bausteine für synchrone und asynchrone Übertragung

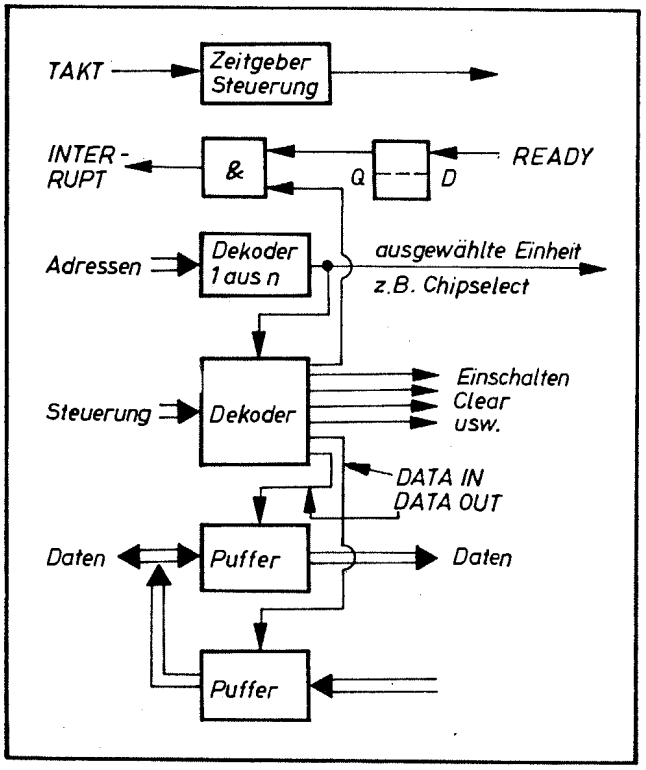


Bild 8.6.28. Blockschaltung eines Interfaces

Über den Dekoder 1 aus *n* wird entschieden, ob die Einheit mit ihrer Adresse überhaupt angewählt ist. Der zweite Dekoder wird erst dann wirksam, wenn die Einheit angewählt ist. Er setzt die Steuersignale angepaßt um. In der Regel werden die Daten über Puffer nach Ein- und Ausgabe getrennt. Außerdem sind Teilschaltungen für z.B. Synchronisation (Zeitgeber) und Interruptsteuerung vorhanden

Weiter werden u. a. Steuerbausteine verwendet für

Video-Interface (Bildschirm-Controller)
Interrupt
Floppy disc
Bubble-Speicher
Standard-Bus-Systeme
dynamische Speicher
Programmierung von PROM
Mathematikprozessoren.

Die meisten dieser Bausteine sind programmierbar, d. h., mittels spezieller Befehle können sie unterschiedliche Funktionen realisieren: z.B. Funktionen, die die Ein- oder Ausgabe, die Synchronisation, die Übertragungsgeschwindigkeit, die Betriebsart (MODE) beispielsweise bezüglich Interrupt oder Handshake betreffen. Bei diesen Funktionen müssen sie aber für die CPU transparent sein. Dies bedeutet, daß sie durch spezielle Signale der CPU mitteilen müssen, was sie gerade tun.

Die Peripheriebausteine werden ähnlich wie Speicherschaltkreise, d. h. per Adresse, betrieben. Bei vielen Prozessoren (auch U 880) existiert jedoch ein spezielles Signal (IORQ), das Speicheradressen und Adressen für periphere Bausteine unterscheiden läßt. Es besteht aber auch die Möglichkeit, die Speicheradressen direkt für die peripheren Bausteine zu verwenden. Dann wird von Speicher-Ein-Ausgabe (memory mapped IN-OUT) gesprochen.

8.6.5.1. PIO U 855

Die PIO U 855 ist den Eigenschaften des U 880 angepaßt. Ihren grundsätzlichen Aufbau zeigt Bild 8.6.29a), die Bedeutung der zugehörigen Signale Tab. 8.6.10. Sie

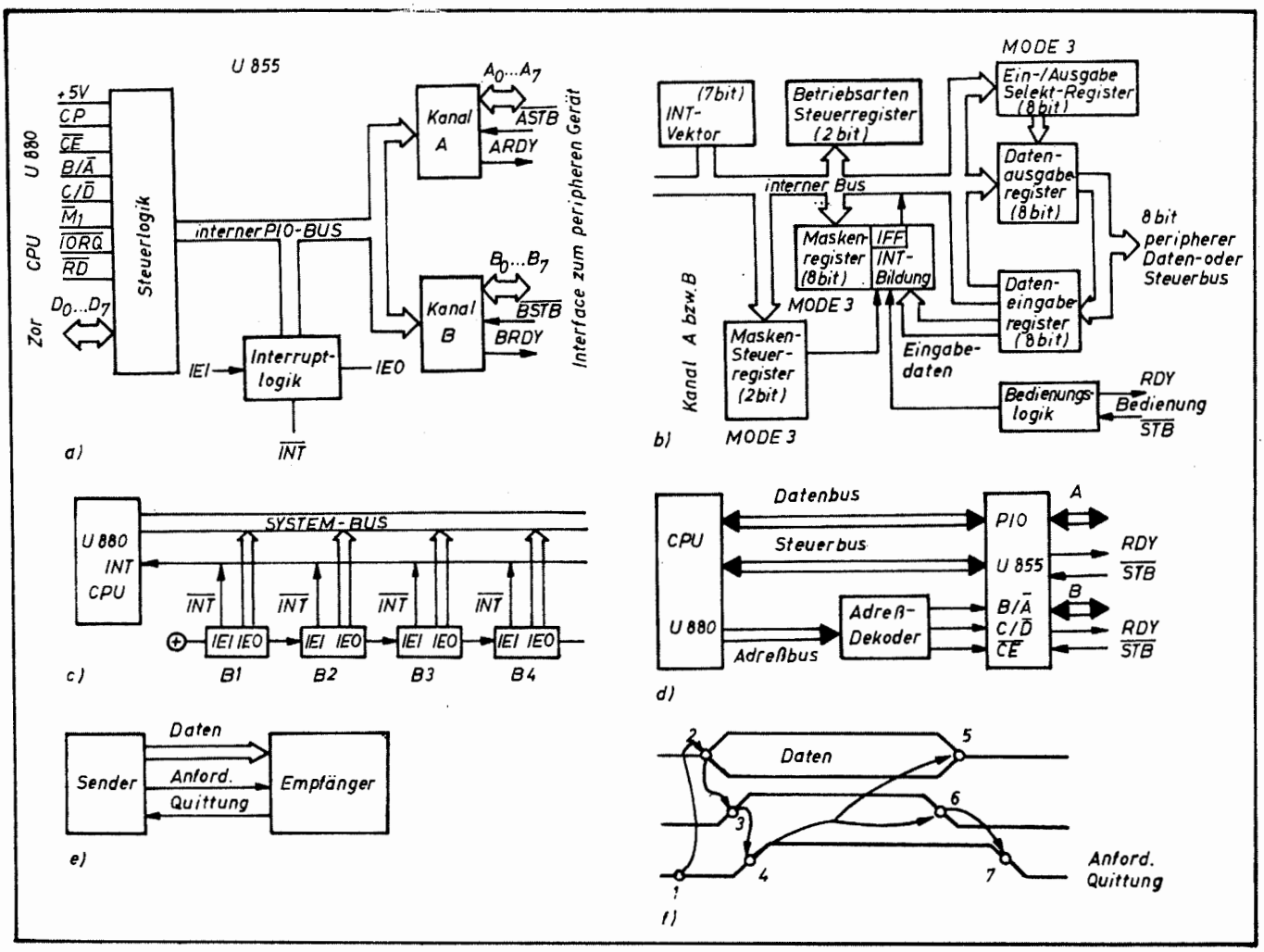


Bild 8.6.29. Aufbau und Eigenschaften der PIO U 855

a) Grundstruktur
b) Struktur eines Kanals
c) Einordnung in die Interrupt-Kette (daisy chain)
d) Zusammenschaltung mit U 880 (Z 80)
e) Byte-Übertragungsschema
f) Signalverläufe bei e)

verfügt über eine Steuereinheit, welche direkt mit gleichlautenden Signalen der CPU U 880 verbunden wird.

Die *Interrupt-Logik* ermöglicht in Zusammenhang mit der Interrupt-MODE IM2 des U 880 den Aufbau einer Prioritätskette gemäß Bild 8.6.29c), wobei die Prioritätsfolge B1, B2, B3, B4, mit B1 als höchster Priorität, gilt. Die Wirkung der Ein- bzw. Ausgänge IEI und IEO ist unten in Tab. 8.6.10 dargestellt.

Die innere Struktur einer der beiden gleichwertigen Kanäle A und B zeigt Bild 8.6.29b). Sie ist durch sieben Register und ein Flipflop gekennzeichnet. Im Betriebsfall gilt dabei:

Das *Interrupt-Register* (INT-Vektor) enthält die niederwertige Adresse für das anzuspringende Interrupt-Programm (vgl. Bild 8.6.27: Kennzeichen).

Das *Betriebsartensteuerregister* legt die Betriebsart der PIO fest (MODE Ø bis 3). Sie werden weiter unten besprochen.

Das *Datenausgaberegister* speichert Daten, die von der CPU an die Peripherie übergeben werden sollen, zwischen.

Das *Dateneingaberegister* speichert Daten zwischen, die von der Peripherie kommen.

Das *Interrupt-Freigabeflipflop* IFF wirkt genauso wie IFF1 beim U 880.

Tabelle 8.6.10. Bedeutung und Funktion der Anschlußleitungen der PIO U 855

Folgende Anschlüsse werden mit den gleichnamigen der CPU U 880 verbunden:

$\overline{\text{INT}}$; $\overline{\text{IORQ}}$; $\overline{\text{M1}}$; CP; $\overline{\text{RD}}$; DØ bis D7

Zusätzliche Bedeutung besitzen dabei:

$\overline{\text{RD}}$	hieraus wird auch ein Schreibsignal intern generiert $\overline{\text{WR}} = \text{RD} \wedge \overline{\text{CE}} \wedge \overline{\text{IORQ}}$
DØ bis D7	übertragen Daten- und Steuerwörter
IEI und **IEO**	(interrupt enable in bzw. out) für Prioritätskette (daisy chain) siehe unten
$\overline{\text{CE}}$	(chip enable) aktiviert bei Signal Low die PIO; wird meist aus den Adressen A2 bis A7 logisch abgeleitet
B/$\overline{\text{A}}$	Auswahl von Kanal (Port) A (Low) bzw. B (High); wird meist von AØ abgeleitet (siehe unten)
C/$\overline{\text{D}}$	(control-/data-select) legt fest, ob auf Datenleitungen ein Datenwort (Low) oder ein Steuerwort (High) liegt; wird meist von A1 abgeleitet (siehe unten)
AØ bis A7 **BØ bis B7**	bidirektionale Tristate-Busleitungen der Kanäle A bzw. B

Für den Handshake- bzw. Interrupt-Betrieb existieren je Kanal zwei Leitungen, deren Funktion von der Betriebsart (Mode Ø bis 4) abhängt.

A STB **B STB**	strobe pulse = Quittierungssignal des peripheren Geräts der Kanäle A bzw. B
A RDY **B RDY**	ready = Bereitschaftssignal der Kanäle A bzw. B
Rücksetzen	erfolgt beim Einschalten der PIO oder durch $\overline{\text{M1}} \wedge \text{RD} \wedge \text{IORQ}$ (dabei bleibt INT-Vektor erhalten)

IEI	H	Interrupt-Forderung	möglich
	L		nicht möglich
IEO	H	Interrupt-Programm	läuft nicht
	L		läuft

Wert	A1	A0	Bedeutung
	C/$\overline{\text{D}}$	B/$\overline{\text{A}}$	
0	0	0	Datenwort für Kanal A
1	0	1	Datenwort für Kanal B
2	1	0	Steuerwort für Kanal A
3	1	1	Steuerwort für Kanal B

Die weiteren drei Register haben nur in MODE 3 Bedeutung und werden dort besprochen.

Die *Zusammenschaltung* mit der CPU zeigt Bild 8.6.29d). Bedeutsam ist, daß vom Adreßbus ein Adreßdekoder notwendig ist. Er hat folgende Entscheidungen zu treffen:

- Festlegung der Adresse der PIO durch 1 aus n Dekodierungen zum Anschluß $\overline{\text{CE}}$. Nur wenn eine entsprechende Adresse vorliegt, wird die PIO aktiviert.
- Auswahl des Kanals über Leitung B/$\overline{\text{A}}$
- Auswahl, ob über die Datenleitung eine Datenübertragung oder Programmierung C/$\overline{\text{D}}$ erfolgt.

Tabelle 8.6.11. Programmierung der PIO U 855. Es gilt die positive Logik 1 ≙ H, Ø ≙ L.

Voraussetzungen: CE = Ø; C/$\overline{D}$ = 1

Kanalwahl: B/$\overline{A}$ = Ø: Kanal A; B/$\overline{A}$ = 1: Kanal B

1. *Eingabe des Interrupt-Vektors*
   DØ = Ø bestimmt Adresse: Interrupt-Register (INT-Vektor)
   D7 bis D1 bestimmen den Low-Teil der Sprungadresse

2. *Wahl der Betriebsart*

   D3 ... DØ = 1111 =ØFH bestimmt die Adresse: Betriebsartenregister
   D7, D6 bestimmen Betriebsart = MODE; D5, D4 ohne Bedeutung

MODE	D7	D6	Betriebsart
Ø	Ø	Ø	Byteausgabe
1	Ø	1	Byteeingabe
2	1	Ø	bidirektional nur für A möglich (dann B nur noch Bitbetrieb)
3	1	1	Bitbetrieb. Es muß unmittelbar 2a folgen

2a. *Ein-Ausgabe-Wahlwort*: Adresse: Ein-Ausgabe-Selektregister
   D7 ... DØ: Bitwert = 1 Eingabe; Bitwert = Ø Ausgabe

3. *Interrupt-Kontrollwort*

   D3 ... DØ= Ø111 = 7H bestimmt Adressen: Interrupt-Flipflop und Maskensteuerregister
   D7 = 1 Interrupt möglich

nur bei Mode 3:

D6, D5 Interruptart:

D6	D5 = Ø	D5 = 1
Ø	OR; Low	AND; Low
1	OR; High	AND; High

D4 bestimmt, ob folgt: ↓

3a. *Maskierungswort* (bei zuvor D4 = 1): Adresse: Maskenregister

   D7 ... DØ: Bitwert = 1 Wert wird für Interrupt genutzt
   Bitwert = Ø Wert wird nicht für Interrupt genutzt

4. *Interrupt-Kontrollwort*

   D3 ... DØ = ØØ11 = 3H bestimmt Adresse: Interrupt-Flipflop
   D7 = 1 Interrupt möglich; D4 bis D6 ohne Bedeutung

In den meisten Fällen wird B/$\overline{A}$ mit AØ und C/$\overline{D}$ mit A1 direkt verbunden. Dann belegt die PIO je 4 aufeinanderfolgende Ausgangs- und Eingangsadressen. Dieser Zusammenhang ist unten in Tab. 8.6.10 dargestellt.

Die *Programmierung* der PIO erfolgt gemäß Tab. 8.6.11. Es sind zwei Arten von Programmierungsworten zu unterscheiden. Die Worte 1 bis 4, die bestimmte Register betreffen und jederzeit eingebbar sind, und andererseits die Worte 2a und 3a, die zeitlich unmittelbar auf die Worte 2 bzw. 3 folgen müssen, wenn diese Worte bestimmten Typs sind. Dies ist auch daraus ersichtlich, daß die jederzeit eingebbaren Worte einen Adreßteil und einen Informationsteil besitzen, die beiden anschließbaren Worte dagegen die 8 bit voll als Informationsteil verwenden.

Mit der Aufstellung der Programmierungsmöglichkeiten in der Mitte von Tab. 8.6.11 sind auch die vier Moden erwähnt. Sie sind weitgehend (nicht vollständig: MODE 2 nur für A) für jeden Kanal vorgebbar.

Bei **MODE Ø** Byteausgabe wirkt der jeweilige Kanal als Sender (Bild 8.6.29e) und f)). Das RDY-Signal ist ein Anforderungssignal (≈ data valid, Bild 8.5.24), und das $\overline{\text{STB}}$-Signal ist ein Quittierungssignal (≈ data accepted).

Bei **MODE 1** Byteeingabe wirkt der jeweilige Kanal als Empfänger. Jetzt ist das $\overline{\text{STB}}$-Signal ein Anforderungssignal und RDY ein Quittierungssignal.

MODE 2 kann nur für Kanal A realisiert werden (B ist dann nur noch in MODE 3 nutzbar). Hierbei werden die Signale $\overline{\text{STB}}$ und RDY von A für die Ausgabe und die Signale $\overline{\text{STB}}$ und RDY von B für die Eingabe in beiden Fällen des Kanals A verwendet.

Das Signal von $\overline{\text{STB}}$ vom Kanal A ist dabei für die weitere Verarbeitung führend. Nur bei A $\overline{\text{STB}}$ Low gehen vom Datenausgaberegister Daten auf die Ausgangsleitungen AØ bis A7. Bei A $\overline{\text{STB}}$ High übernimmt dagegen das Dateneingaberegister Daten von den Leitungen AØ bis A7. Die CPU übernimmt beim Lesen Daten aus einem der beiden Register: Bei A $\overline{\text{STB}}$ H aus dem Dateneingaberegister, bei A $\overline{\text{STB}}$ L aus dem Datenausgaberegister.

Bei **MODE 3** existiert kein Handshake-Betrieb. Hier können für jeden Kanal die vorhandenen acht Leitungen als Ein- bzw. Ausgabe wunschgemäß programmiert werden. Diese Programmierung erfolgt unmittelbar nach der Modewahl und belegt das Eingabe-Ausgabe-Select-Register. Hierbei ist zur Synchronisation die Interrupt-Möglichkeit äußerst wichtig. Sie wird durch Eingabe des Interrupt-Kontrollwortes in das Maskensteuerregister (3 in Tab. 8.6.11) festgelegt. Über Bit 5 wird dabei festgelegt, ob auszuwählende Leitungen über AND oder OR, und über Bit 6, ob der Low- oder High-Pegel gültig sein soll. Welche Leitungen ausgewählt werden, bestimmt dann die in das Maskenregister einzugebende Maske 3a. Weiter kann noch mit einem Interrupt-Kontrollwort nach 3 oder 4 das Interrupt-Freigabeflipflop IFF gesetzt bzw. gelöscht werden. Schließlich gibt es ein spezielles Löschsignal (RESET) nicht. Das Löschen der Register und der Flipflops des Bausteins geschieht entweder beim Anlegen der Betriebsspannung oder durch M1, wenn nicht gleichzeitig eines der Signale $\overline{\text{RD}}$ bzw. $\overline{\text{IORQ}}$ aktiv ist. Dieses Löschen erzeugt folgenden Zustand:

- die Maskenregister sind gelöscht;
- beide Interrupt-Freigabeflipflops sind gelöscht;
- beide Ausgaberegister sind gelöscht;
- das Register zur Bereitstellung des Interruptvektors wird nicht gelöscht;
- Tor A und B werden hochohmig, die Handshake-Signale $\overline{\text{ASTB}}$, ARDY, $\overline{\text{BSTB}}$, BRDY werden inaktiv.

Der RESET-Zustand bleibt so lange bestehen, bis die PIO ein Steuerwort vom Prozessor erhält.

8.6.5.2. SIO U 856

Das Grundprinzip einer seriellen Ein-Ausgabe-Einheit (= serial in-out) besteht darin, daß die bitparallelen Byteworte nacheinander in den einzelnen Bit übertragen werden bzw. eingehende Bitströme wieder zu Byteworten zusammenzufügen. Diese Aufgabe existiert bei vielen peripheren Geräten, z. B. Lochstreifengeräten, Magnetbandgeräten, Plattenspeichern, Floppy-disc-Speichern, Fernschreibern, Geräten zur Datenkommunikation (Telefon), Rechnerkopplungen, und wird durch Schieberegister

realisiert. Die Fülle dieser Geräte und die Besonderheiten der sequentiellen Übertragung erfordern auch eine entsprechende Vielfalt der Gestaltung der seriellen Datenströme, z.B. mittels spezieller Kodierungen, Einfügen von Start- und Stopsowie Synchronisationszeichen. Darüber hinaus ist es sinnvoll (notwendig), fehlererkennende Verfahren zu realisieren. Dies alles soll mit einer SIO realisiert werden, die so aufgebaut ist, daß sie vom Rechner auf den speziellen Anwendungsfall programmierbar ist. Dies führt dazu, daß eine gute SIO hinsichtlich der Komplexität ihrer Hardware und der Möglichkeiten ihrer Programmierbarkeit mit zu den kompliziertesten Bauelementen der Mikroelektronik gehört. Dies ist der Grund, weshalb sie hier nur relativ oberflächlich behandelt werden kann; weitere Details können z.B. aus [C 4 und K 39] entnommen werden.

Den Grundaufbau der SIO U 856 zeigt Bild 8.6.30a). Sie enthält — wie die PIO U 855 — zwei gleichartige Kanäle. Den Aufbau eines Kanals zeigt Teilbild b). Neben den Schieberegistern sind ein Synchronisationszeichenregister, Taktzähler für Sender und Empfänger sowie drei Pufferregister zu 8 bit für den Empfang und je ein Pufferregister für den Sender enthalten. Die CRC (vgl. 8.6.5.8.) dienen der Feststellung von Übertragungsfehlern mittels spezieller rückgekoppelter Register. Weiter fällt sofort die Vielzahl der programmierbaren Schreib- und Leseregister WRØ bis WR7 und RRØ bis RR2 auf. Sie werden weiter unten behandelt.

Folgende Anschlüsse der SIO werden — ähnlich wie bei der PIO — mit den gleichnamigen der CPU verbunden:

DØ bis D7, $\overline{\text{M1}}$, $\overline{\text{RD}}$, $\overline{\text{IORQ}}$ und $\overline{\text{INT}}$.

Die Anschlüsse

CP und RESET

sind ähnlich zu handhaben. Analog zur PIO werden für die Kanalauswahl, Datenübertragung — Programmierung, Adressierung und Interrupt die Anschlüsse (s. Tab. 8.6.10)

B/$\overline{\text{A}}$, C/$\overline{\text{D}}$, CE, IEI und IEO

gehandhabt. Damit sind nur noch die speziellen Anschlüsse der Tab. 8.6.12 zu behandeln.

Bei der seriellen Übertragung können asynchrone und synchrone Verfahren gemäß Bild 8.6.30c) und d) unterschieden werden. Die asynchronen besitzen eine definierte Länge (5 bis 8 bit), ein Start- und ein Stopbit und können zusätzlich ein Parity-Bit zur Fehlerprüfung besitzen. Zwischen einem Stop- und einem Startsignal bleibt der Pegel konstant auf High (Mark). Programmierbar ist hierbei folglich

Anzahl der Bits (WR3; D6 + D7; WR5; D5 + D6)
Stopbitlänge (WR4; D2 + D3)
Parität (WR4; DØ + D1)
Frequenzverhältnis (WR4; D6 + D7).

Bei der synchronen Übertragung wird ein Datenfeld unterschiedlich nach Synchronisationszeichen, CRC-Zeichen (cyclic redundancy chec = spezielle Methode der Fehlererkennung, vgl. 8.6.5.8.), Adressen und/oder Flag eingerahmt, und der Takt der Übertragung wird dabei über längere Zeit aufrechterhalten. Die Vielfalt dieser Möglichkeiten verlangt auch eine Vielfalt der Programmierung in den Registern.

Die Programmierung der Register WRØ bis WR7 erfolgt bei C/$\overline{\text{D}}$ = 1. Die Register haben dabei eine 3-bit-Adresse, z.B. 5D = 1Ø1B, die immer zuerst in die drei letzten Bits des WRØ geschrieben werden müssen. Das danach folgende Byte gelangt in das zugehörige Register. Der Interrupt-Vektor befindet sich im WR2, das nur in Kanal B

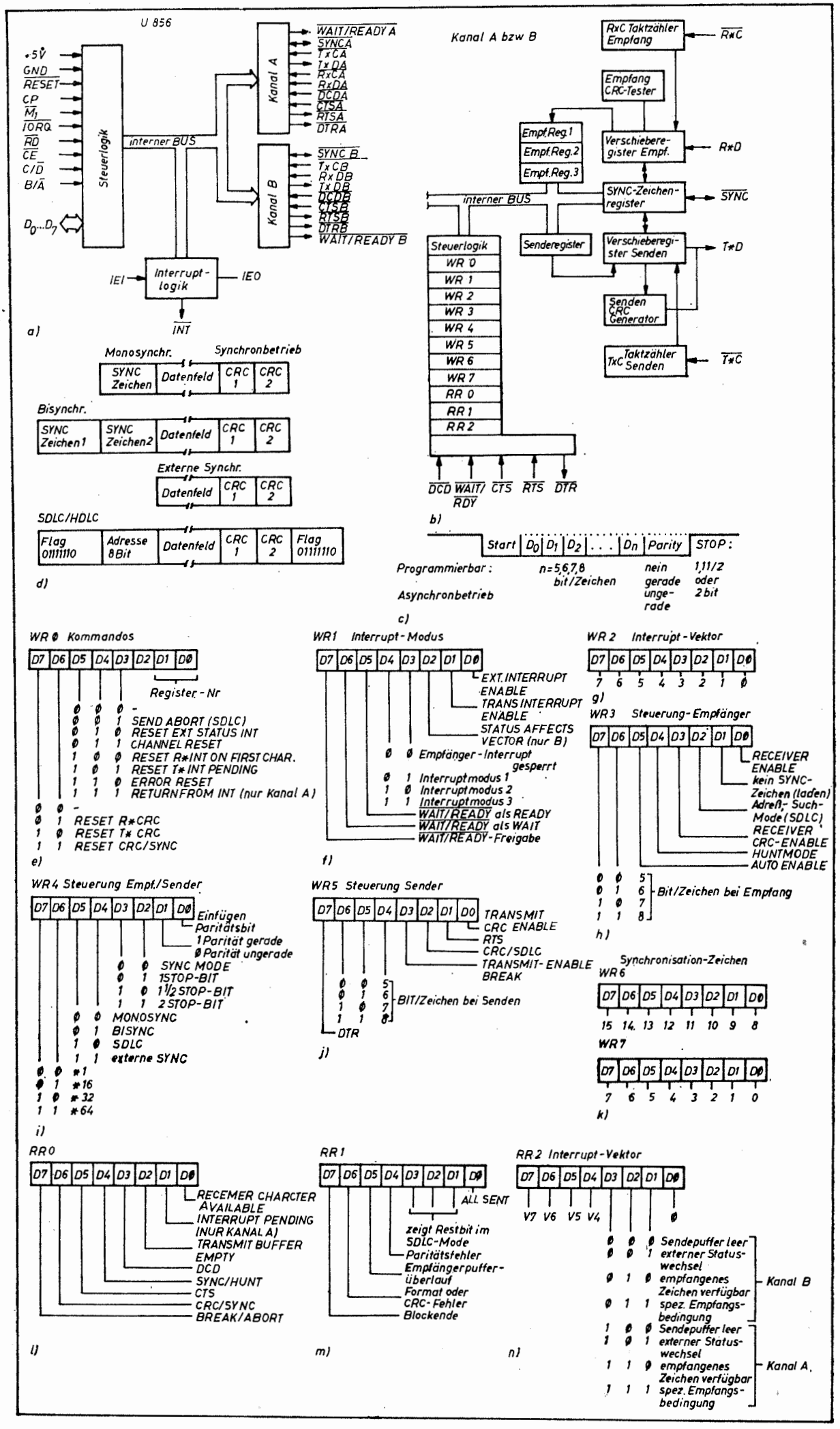

U 856
+5V
GND
RESET
CP
M1
IORQ
RD
CE
C/D
B/A
D0...D7
Steuerlogik
interner BUS
Kanal A
Kanal B
WAIT/READY A
SYNC A
TxCA
TxDA
RxCA
RxDA
DCDA
CTSA
RTSA
DTRA
SYNC B
TxCB
RxDB
TxDB
DCDB
CTSB
RTSB
DTRB
WAIT/READY B
IEI
Interrupt-logik
IEO
INT
a)
Monosynchr.
Synchronbetrieb
SYNC Zeichen
Datenfeld
CRC 1
CRC 2
Bisynchr.
SYNC Zeichen1
SYNC Zeichen2
Externe Synchr.
SDLC/HDLC
Flag 01111110
Adresse 8Bit
Flag 01111110
d)
Kanal A bzw B
RxC Taktzähler Empfang
R*C
Empfang CRC-Tester
Empf.Reg.1
Empf.Reg.2
Empf.Reg.3
Verschiebe-register Empf.
R*D
SYNC-Zeichen-register
SYNC
Verschiebe-register Senden
T*D
Senden CRC Generator
TxC Taktzähler Senden
T*C
Steuerlogik
WR 0
WR 1
WR 2
WR 3
WR 4
WR 5
WR 6
WR 7
RR 0
RR 1
RR 2
Senderegister
DCD
WAIT/RDY
CTS
RTS
DTR
b)
Start
D0 D1 D2 . . . Dn
Parity
STOP:
Programmierbar:
n=5,6,7,8 bit/Zeichen
nein gerade ungerade
1,11/2 oder 2 bit
Asynchronbetrieb
c)
WR 0 Kommandos
D7 D6 D5 D4 D3 D2 D1 D0
Register-Nr
0 0 0 -
0 0 1 SEND ABORT (SDLC)
0 1 0 RESET EXT STATUS INT
0 1 1 CHANNEL RESET
1 0 0 RESET R*INT ON FIRST CHAR.
1 0 1 RESET T* INT PENDING
1 1 0 ERROR RESET
1 1 1 RETURN FROM INT (nur Kanal A)
0 0 -
0 1 RESET R*CRC
1 0 RESET T* CRC
1 1 RESET CRC/SYNC
e)
WR1 Interrupt-Modus
EXT. INTERRUPT ENABLE
TRANS INTERRUPT ENABLE
STATUS AFFECTS VECTOR (nur B)
0 0 Empfänger-Interrupt gesperrt
0 1 Interruptmodus 1
1 0 Interruptmodus 2
1 1 Interruptmodus 3
WAIT/READY als READY
WAIT/READY als WAIT
WAIT/READY-Freigabe
f)
WR 2 Interrupt-Vektor
7 6 5 4 3 2 1 0
g)
WR3 Steuerung-Empfänger
RECEIVER ENABLE
kein SYNC-Zeichen (laden)
Adreß-Such-Mode (SDLC)
RECEIVER CRC-ENABLE
HUNTMODE
AUTO ENABLE
0 0 5
0 1 6
1 0 7
1 1 8
Bit/Zeichen bei Empfang
h)
WR4 Steuerung Empf./Sender
Einfügen Paritätsbit
1 Parität gerade
0 Parität ungerade
0 0 SYNC MODE
0 1 1STOP-BIT
1 0 1 1/2 STOP-BIT
1 1 2 STOP-BIT
0 0 MONOSYNC
0 1 BISYNC
1 0 SDLC
1 1 externe SYNC
0 0 *1
0 1 *16
1 0 *32
1 1 *64
i)
WR5 Steuerung Sender
TRANSMIT
CRC ENABLE
RTS
CRC/SDLC
TRANSMIT-ENABLE
BREAK
0 0 5
0 1 6
1 0 7
1 1 8
BIT/Zeichen bei Senden
DTR
j)
Synchronisation-Zeichen
WR 6
15 14 13 12 11 10 9 8
WR 7
7 6 5 4 3 2 1 0
k)
RR 0
RECEIVER CHARCTER AVAILABLE
INTERRUPT PENDING (NUR KANAL A)
TRANSMIT BUFFER EMPTY
DCD
SYNC/HUNT
CTS
CRC/SYNC
BREAK/ABORT
l)
RR 1
ALL SENT
zeigt Restbit im SDLC-Mode
Paritätsfehler
Empfängerpuffer-überlauf
Format oder CRC-Fehler
Blockende
m)
RR 2 Interrupt-Vektor
V7 V6 V5 V4
0 0 0 Sendepuffer leer
0 0 1 externer Statuswechsel
0 1 0 empfangenes Zeichen verfügbar
0 1 1 spez. Empfangsbedingung
Kanal B
1 0 0 Sendepuffer leer
1 0 1 externer Statuswechsel
1 1 0 empfangenes Zeichen verfügbar
1 1 1 spez. Empfangsbedingung
Kanal A
n)

Tabelle 8.6.12. Spezifische Anschlüsse der SIO U 856. Die ergänzenden Bezeichnungen für den Kanal A bzw. B sind hinzuzufügen.

Auf Empfänger bezogen		
R × D	receive data	Dateneingang
R × C	receiver clock	Takteingang Programmierbar ist, ob die Taktfrequenz 1-, 16-, 32- oder 64mal so hoch wie die Baudrate sein soll. Dies gilt insbesondere bei synchronem Betrieb. Die Datenrate darf höchstens $^1/_{4,5}$ des Systemtaktes CP sein. Bei einer Taktfrequenz von 2,5 MHz sind folglich Daten bis zu 550 Kbit/s verarbeitbar.
DCD	data carrier detect	Empfänger sperren (handshake)
DTR	data terminal ready	Empfänger bereit (handshake)
Auf Sender bezogen		
T × D	transmit data	Datenausgang
T × C	transmitter clock	analog zu R × C, mit gleichem Faktor bei gleichem Kanal
CTS	clear to send	Sender sperren (handshake)
RTS	request to send	Sendegesuch (handshake)
Allgemeine Anschlüsse		
WAIT/RDY	wait/ready	Generierung von Wartezuständen bzw. DMA-Steuerung
SYNC	synchronization	Ein- und Ausgang für spezielle externe Synchronisationen

existiert. In den Registern (WR6 und) WR7 stehen die Werte für das ausgewählte Synchronisationszeichen von 8 bzw. 16 bit Länge.

Die Leseregister enthalten im wesentlichen Informationen über den Status des jeweiligen Kanals. Sie werden daher auch nicht von der CPU programmiert, sondern von ihr gelesen. In sie gelangt die Information beim Ablauf der Datenübertragung ähnlich wie in das Flag-Register bei der CPU. Sie werden wie folgt gelesen: Bei C/D = 1 wird in WRØ die Adresse Ø, 1 oder 2 eingeschrieben, und dann wird mit R/D das entsprechende Register gelesen. Dabei enthält das Register RR2 genau den Interrupt-Vektor, der in das Register WR2 eingeschrieben ist, sofern D2 = Ø in WR1 steht. Anderenfalls werden D1 bis D3 in RR2 gemäß dem Bild 8.6.30n) verändert. RR2 existiert dementsprechend wie WR2 nur in Kanal B.

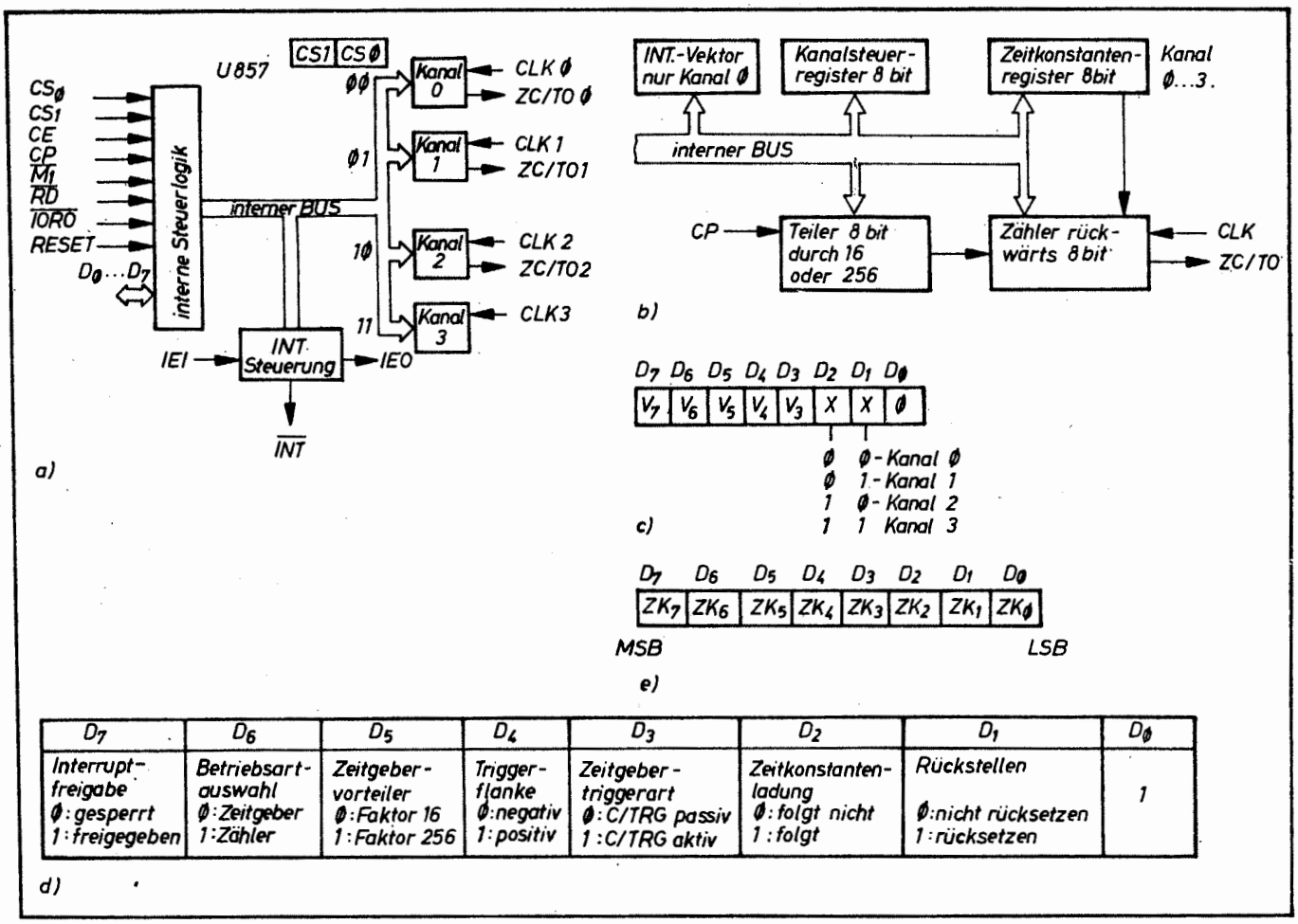
8.6.5.3. CTC U 857

Der Schaltkreis CTC realisiert zwei recht universell anwendbare Funktionen: Zählen und Frequenzteilen. Hiermit kann infolge seiner Programmierung u. a. eine Echtzeituhr realisiert werden. Auch die Erzeugung von frei vorgebbaren Frequenzen, das

◀

Bild 8.6.30. Aufbau und Eigenschaften der SIO U 856

a) Grundstruktur
b) Struktur eines Kanals
c) Prinzip der asynchronen Übertragung
d) Möglichkeiten der synchronen Übertragung (SDLC = synchronous data link control)
e) bis k) Inhalte der 8 Schreibregister für die Programmierung
l) bis n) Inhalte der 3 Leseregister für die Programmierung



D_7	D_6	D_5	D_4	D_3	D_2	D_1	$D_\emptyset$
Interrupt-freigabe Ø: gesperrt 1: freigegeben	Betriebsart-auswahl Ø: Zeitgeber 1: Zähler	Zeitgeber-vorteiler Ø: Faktor 16 1: Faktor 256	Trigger-flanke Ø: negativ 1: positiv	Zeitgeber-triggerart Ø: C/TRG passiv 1: C/TRG aktiv	Zeitkonstanten-ladung Ø: folgt nicht 1: folgt	Rückstellen Ø: nicht rücksetzen 1: rücksetzen	1

d)

Bild 8.6.31. Aufbau und Eigenschaften der CTC U 857

a) Grundstruktur
b) Struktur eines Kanals
c) Aufbau des Interrupt-Vektors für die vier Kanäle. Kanal 0 hat die höchste Interrupt-Priorität
d) Bedeutungen des Kanalsteuerregisters
e) Aufbau des Zeitkonstantenregisters

Messen analoger Größen usw. ist möglich. Der Aufbau des Schaltkreises (vgl. Bild 8.6.31) ist relativ übersichtlich. Er kann — ähnlich wie die PIO und SIO — mit dem U 880 unmittelbar gekoppelt werden und bezüglich Interrupt-Verhalten betrieben werden. Er enthält jedoch vier gleiche Kanäle gemäß Teilbild b). Sie werden durch die Chip-Select-Leitungen CSØ und CS1 angewählt. In der Regel werden CSØ mit AØ und CS1 mit A1 verbunden. Dadurch werden die vier Kanäle mit $\overline{CE}$ an vier aufeinander folgende Adressen gelegt.

Die Programmierung erfolgt ausschließlich über das Signal $\overline{WR}$ (s. U 880). Das Bit $D\emptyset = \emptyset$ legt damit die Adresse für den Interrupt-Vektor (Teilbild c)) nur in Kanal Ø fest, wobei D1 und D2 intern mittels der Kanäle Ø bis 3 gemäß dem Bild ergänzt werden. Ist dagegen $D\emptyset = 1$, so geht der Wert automatisch in das Steuerregister (Teilbild d)). Ist hier $D2 = 1$, so erfolgt ein Wert für das Zeitkonstantenregister. Die wichtigste funktionelle Programmierung ist also mit dem Kanalsteuerregister gegeben. Insbesondere D6 legt dabei fest, ob der Kanal als Zeitgeber oder Zähler benutzt wird.

Im *Zähler-Mode* ist der Teiler ohne Bedeutung. Bei Beginn wird der Inhalt des Zeitkonstantenregisters in den Rückwärtszähler übernommen und mit der Flanke (nach Definition D4 clock/trigger) um Eins dekrementiert. Die maximale Zählfrequenz ist 1,25 MHz (CP/2). Es wird nämlich auf die steigende Flanke von CP synchronisiert. Wenn der Zählerwert Null erreicht wird, wird der Ausgang ZC/TO

(zero count/time out) auf High geschaltet. Wenn eine Zählung von Ø bis 255 nicht ausreicht, kann dieser Ausgang neben der Anzeige des Nulldurchganges auch zur Kaskadierung mehrerer Zähler verwendet werden. Er wird dazu mit dem CLK-Eingang des nächsten Zählers verbunden. Nach dem Nulldurchgang wird jeder Zähler automatisch wieder vom Zeitkonstantenregister geladen. Dadurch verhält sich der Zähler so, als ob ein Ringzähler der Länge vorliegt, die in das Zeitkonstantenregister programmiert wurde.

Im *Zeitgeber-Mode* wird zunächst der Takt über den Vorteiler programmierbar um 16 bzw. 256 (D5) geteilt, bei 2,5 MHz also auf 156250 bzw. 975, 625 Hz. Mit dem Rückwärtszähler ist je nach dem Wert im Zeitkonstantenregister dann nochmals eine Teilung zwischen 1 und 256 (Wert ØØ im Register) möglich. Diese Frequenz erscheint am Zeitgeberausgang mit kurzen Impulsen. Je nach dem Wert von D3 kann dieser Teilervorgang durch den Ladezeitpunkt des Zeitkonstantenregisters oder durch eine Flanke (D4) am Eingang CLK ausgelöscht werden. Zeitgeberkanäle sind selbst nicht kaskadierbar, sie können nur durch Zählerkanäle weiter geteilt werden.

8.6.5.4. U 880-DMA

Die DMA ist ein spezieller Baustein für schnelle Datentransporte und/oder Datenprüfung. Den Aufbau der DMA zeigt Bild 8.6.32. Sie besitzt wiederum eine Interface-Schaltung zur CPU, mit der sie allerdings vielfältiger als die bisher behandelten Bausteine umgehen kann. Sie muß nämlich in den DMA-Zyklen die CPU vollständig (mit Ausnahme der Refresh-Zyklen) ersetzen können. Das heißt, in ihren Arbeitszyklen sind die Busleitungen des U 880 hochohmig, und die DMA bestimmt die Adressen der anzusprechenden Speicher bzw. Ports. Zu diesem Zweck muß sie Daten empfangen und abgeben können, Adressen ausgeben und ebenso die Signale

$\overline{WR}$, $\overline{MREQ}$, $\overline{IORQ}$, $\overline{BUSRQ}$

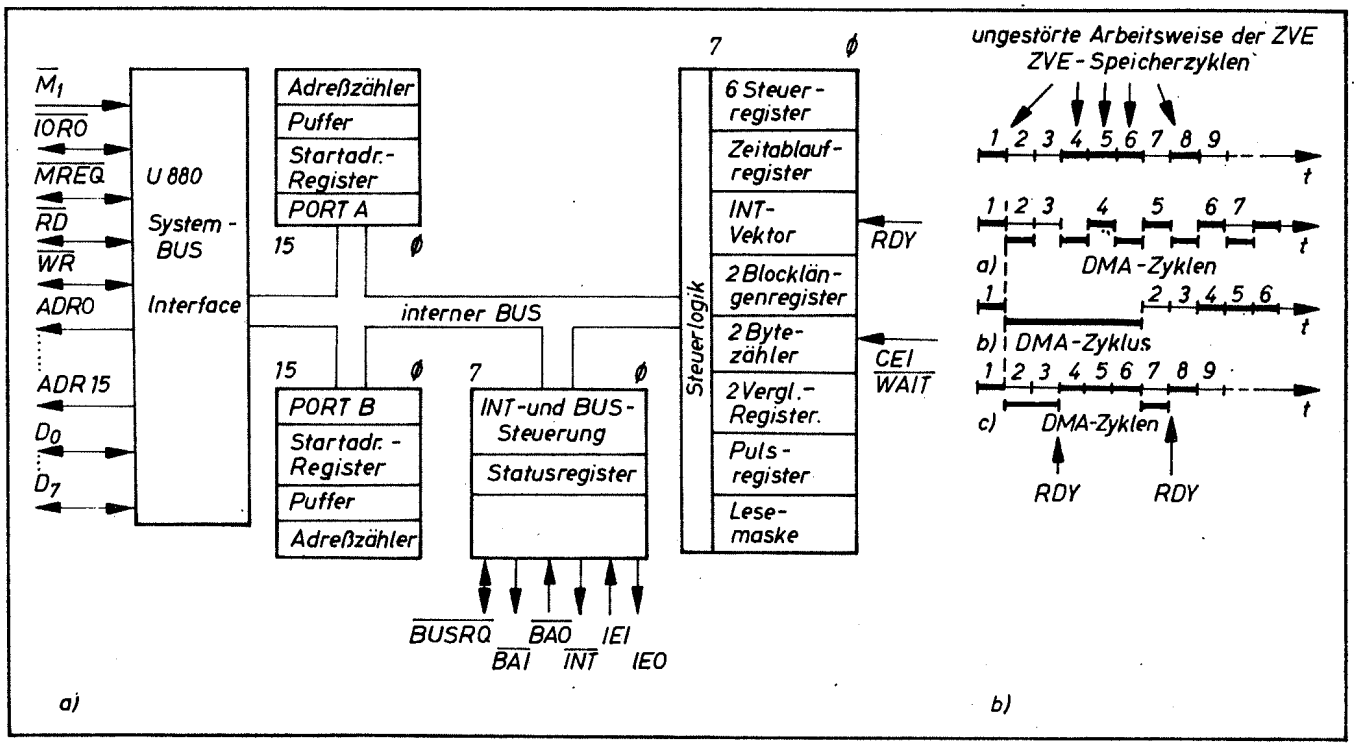


Bild 8.6.32. Aufbau und Eigenschaften der U 880-DMA

a) Grundstruktur
b) Zyklenaufteilungen zwischen DMA und CPU

abgeben und empfangen können, weiter besitzt sie dazu die beiden Ports A und B sowie die schon bekannten Interrupt-Signale

$\overline{\text{IEI}}$, $\overline{\text{IEO}}$, $\overline{\text{INT}}$

und die Sondersignale, welche den Wechsel der Arbeit mit der CPU bestimmen:

$\overline{\text{BAI}}$ bus acknowledge in
$\overline{\text{BAO}}$ bus acknowledge out
CE/$\overline{\text{WAIT}}$ chip enable oder wait
RDY ready

Je nach der Programmierung ist entweder Port A Eingang; Port B Ausgang oder Port B Eingang; Port A Ausgang. Die Verwendung der Busse in Arbeitsteilung mit der CPU kann in drei Betriebsweisen erfolgen

- *Byte-Mode* (Zyklus stehlen = cycle stealing): Hierbei wird immer genau ein Byte durch die DMA übertragen und dann die Steuerung bis zum nächsten Byte an die CPU zurückgegeben.
- *Continuous-Mode*: Hierbei wird die gesamte programmierte Anzahl von Bytes durch die DMA in unmittelbarer Folge übertragen und erst dann die Steuerung an die CPU zurückgegeben (Teilbild b)).
- *Burst-Mode*: Sie entspricht der Continuous-Mode bis auf den einen Umstand, daß die DMA-Arbeitsweise vor Blockende über RDY unterbrochen werden kann (z.B. Teilbild c)).

Bei den beiden letzten Moden ist zu beachten, daß bei dynamischen Speichern die DMA nicht länger als ca. 2 ms ohne Pause arbeitet. Dann würden die Speicher nämlich den Inhalt verlieren. Da die DMA aber mit ca. 1,2 MHz arbeiten kann, sind das immerhin ca. 50 Speichertransporte. Besonders effektiv ist andererseits die Byte-Mode, da sie sich im Schatten der CPU ohne wesentliche Verzögerung vollzieht.

Je nach der Programmierung erfolgt weiter der Datentransport in folgende Richtungen:

Speicher → Speicher
Speicher → Ausgabeeinheit
Eingabeeinheit → Speicher
Eingabeeinheit → Ausgabeeinheit.

Bei der *Datensuche* findet kein Datentransport statt, sondern es werden nur ein Adressenbereich von Speichern bzw. ein bestimmter Adressenbereich von Toren auf ein bestimmtes Datenbyte geprüft, wobei auch ein oder mehrere Bits ignoriert (maskiert) werden können. Sobald das gesuchte Byte (oder Byteteil) gefunden wird, ist das Ziel erreicht, und die Steuerung geht an die CPU endgültig zurück.

Bei der *Datensuche mit Datentransfer* sind die beiden eben geschilderten Prinzipien gekoppelt.

Um die vielfältigen Aufgaben erfüllen zu können, besitzt die DMA wieder eine größere Anzahl von programmierbaren und abfragbaren (lesbaren) Registern. Sie sind in Bild 8.6.32a) z. T. nur angedeutet. Die programmierbaren Register sind primär 1A, 1B, 2A bis 2D. Sie werden durch die Bytes DØ, D1 und D7 adressiert. Bei D7 = Ø sind die 1- und bei D7 = 1 die 2-Register angesprochen.

Bei allen Registern wird durch Belegung bestimmter Bits festgelegt, welches der folgenden Register sofort zeitlich danach programmiert wird:

- Port A bzw. B nieder- und höherwertiges Adreßbyte (Adreßzähler)
- Blocklänge nieder- und höherwertiges Byte (Bytezähler)

- Zeitsteuerbyte (Zeitablaufregister)
- Maskenbyte
- Vergleichsbyte } (Vergleichsregister)
- Interrupt-Vektor
- Lesemaske (die bezeichneten Bytes können danach gelesen werden).

Weiter kann nach Programmierung von 2D mit BB das

Statusregister

gelesen werden.

Ausführliche Aussagen zum DMA-Schaltkreis sind u. a. in [B 4, Bd. 3, und O 4] enthalten.

8.6.5.5. Arithmetikprozessoren

Arithmetikprozessoren, zuweilen auch Coprozessoren genannt, erfüllen eine wichtige Aufgabe, wenn bei Mikroprozessoren eine solide Arithmetik implementiert werden soll. Dies kann auf zwei Wegen erfolgen:

- durch einen entsprechenden Befehlssatz von einigen Kbytes
- durch die zu behandelnden Prozessoren.

Beim Befehlssatz werden für viele Operationen neben der erforderlichen Speicherkapazität Rechenzeiten von mehreren bis vielen ms benötigt. Bei Arithmetikprozessoren ist die Bearbeitung der Zahlen dagegen vorwiegend parallel organisiert, und es werden nur Zeiten von wenigen ms oder einigen ...zig µs benötigt.

Hier sollen zwei typische Generationen von Arithmetikprozessoren beschrieben werden (vgl. Bild 8.6.33).

Der 8231 ist speziell für die Mikroprozessoren 8080, 8085, Z 80 und 6800 entwickelt worden. Die Zusammenschaltung mit dem Z 80 zeigt Bild 8.6.33c). Er arbeitet daher extern mit dem 8-bit-Datenbus und kann intern mit 16-bit- bzw. 32-bit-Zahlen operieren, wobei die 32-bit-Zahlen sowohl als Integer als auch als Gleitkommazahlen behandelt werden. Daher besitzt er die drei intern verarbeitbaren Formate:

16 bit ganze Zahlen	± 32767
32 bit ganze Zahlen	± 2147483647
32 bit Gleitkomma	$\pm 8{,}388608 \pm 10^{64}$

Der interne Stack kann dementsprechend auf zwei Arten gemäß Teilbild b) organisiert, also 8 oder 4 Stufen tief sein. Als Funktionen realisiert er

Grundfunktionen, wie:	Stackoperationen, Dateneinn- und -ausgabe, Austausch von Daten, Vorzeichenwechsel, Umwandlungen zwischen den Zahlenformaten;
Grundoperationen:	Addition, Subtraktion, Multiplikation, Division;
Funktionen:	Quadratwurzel; SIN, COS, TAN (und Umkehrungen); LOG, LN, EXP, X hoch Y.

Die Funktionen sind dabei über TSCHEBYSCHEFF-Polynome approximiert und besitzen einen Fehler, der meist kleiner als 10^{-7} ist.

Der Typ 8087 ist für den Mikroprozessor 8086 entwickelt, besitzt also 16-bit-Schnittstellen (Bild 8.6.33d) und e)). Er enthält 65000 Transistoren. Bei ihm wurden alle Fortschritte der Numerik, insbesondere der IEEE-Gleitkommastandard, berücksichtigt. Dabei wird auch sinnvoll weitergerechnet, wenn ein Überlauf in Richtung

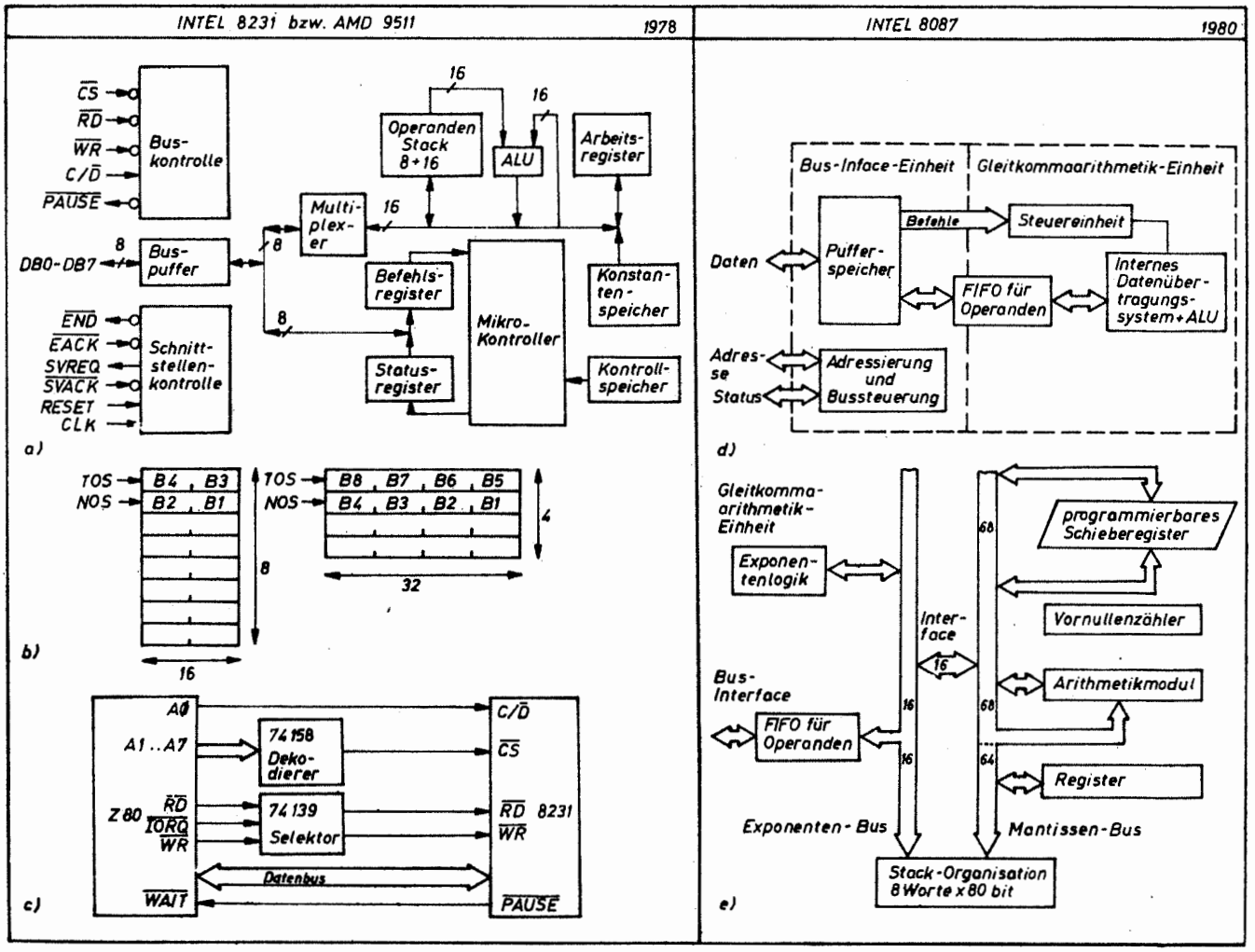


Bild 8.6.33. Arithmetikprozessoren

a) Blockschaltbild des 8231
b) Stackvarianten des 8231
c) Kopplung von U 880 und 8231
d) Blockschaltbild des 8087
e) Detaillierte Struktur der Gleitkommaarithmetik des 8087

Neben den 8 Registern zu 40 bit existieren 4 Register zu 16 bit für Status, Steuerung, Befehlszähler und Datenzeiger

$\pm\infty$ oder eine Division durch Null erfolgt. Das heißt, daß Überlaufzahlen wie $\pm\infty$ behandelt werden, und eine Zahl $+X$ dividiert durch Null ergibt dann z.B. $+\infty$.

Der nächste Fortschritt liegt darin, daß intern mit einem 80-bit-Wort aus 16-bit-Exponenten und 64- (68-) bit-Mantisse in 8 Registern binär gerechnet wird. Der intern verarbeitete Zahlenbereich liegt daher zwischen $3{,}4 \cdot 10^{-4932}$ und $1{,}2 \cdot 10^{4932}$, ist also auf 19 gültige Ziffern der Mantisse ausgelegt. Es existiert eine getrennte Operationstechnik für den Exponenten und die Mantisse. Nach außen wird mit 6 Zahlenformaten operiert, dabei erfolgt immer eine exakte Rundung:

	bit	gültige Ziffern
3 × Ganzzahlen (Integer)	16, 32 oder 64	5, 10 oder 19
Gepacktes Dezimal	80	18
Gleitkomma $8{,}43 \cdot 10^{-37}$ bis $3{,}37 \cdot 10^{38}$	32	7
Gleitkomma $4{,}19 \cdot 10^{-307}$ bis $1{,}67 \cdot 10^{308}$	64	16

Neben den Standardoperationen wie Laden, Austauschen, Ausgeben, Addition, Subtraktion, Multiplikation, Division, Inkrementieren, Dekrementieren, Vorzeichenwechsel, Absolutwert, getrennte Ausgabe von Exponent und Mantisse verfügt er nur über wenige Funktionen,

$$\sqrt{x},\ \tan,\ \arctan,\ 2^x - 1,\ y \cdot \mathrm{ld}\, x,\ y \cdot (\mathrm{ld}\, x + 1)\,,$$

und die Konstanten

$$\emptyset,\ 1,\ \pi,\ \log_{10}(2),\ \log_e(2),\ \log_2(e),\ \log_2(10)\,.$$

Für die Funktionen existieren einmal die sogenannten Cordic-Algorithmen, die mit optimalen und beliebig genau zu machenden Drehoperationen von Vektoren im Komplexen operieren und so zwei Komponenten ausgeben, mit denen besonders gut weitergerechnet werden kann. Dies sei am Beispiel $Y/X = \tan\varphi$ gezeigt. Aus den beiden Werten X und Y folgt z.B. $\sin\varphi = Y/\sqrt{X^2 + Y^2}$ bzw. $\cos\varphi = X/\sqrt{X^2 + Y^2}$.

In ähnlicher Weise lassen sich aus den implementierten Funktionen alle weiteren Standardfunktionen sehr leicht und wiederum genau gewinnen.

8.6.5.6. Tastatur- und Anzeigenanschlüsse

Damit ein Mikrorechner betrieben werden kann, benötigt er Ein- und Ausgabeeinheiten. Die einfachsten zur Verfügung stehenden Mittel sind Tasten und Anzeigeelemente, z.B. Siebensegmentanzeigen. Den Betrieb beider zeigt Bild 8.6.34.

Die übliche Methode des Tastaturbetriebes geht über die Abfrage. Dazu wird die Tastatur in einer Matrix angeordnet, im Bild 64 Tasten als 8 × 8 Matrix. An jedem Kreuzungspunkt befindet sich ein Kontakt, der bei Tastenbestätigung geschlossen wird. Über das Out-Tor werden an die Spalten Ø bis 7 in einem 8fach-Zyklus Bytes eingegeben: Ø1, Ø2, Ø4, Ø8, 1Ø, 2Ø, 4Ø, 8Ø. Der Reihe nach wird also je einer Spaltenleitung ein H-Signal aufgeprägt. Nach jeder Eingabe wird das In-Tor abgefragt, ob es einen Wert ungleich Ø besitzt. Wenn dies der Fall ist, wurde eine Taste in dem Zyklus betätigt. Aus dem eingegebenen Byte und dem erhaltenen Byte kann dann die gedrückte Taste berechnet werden. Es kann auch bestimmt werden, ob fehlerhafterweise zwei Tasten betätigt wurden. Durch Wiederholung der Abfrage im etwa

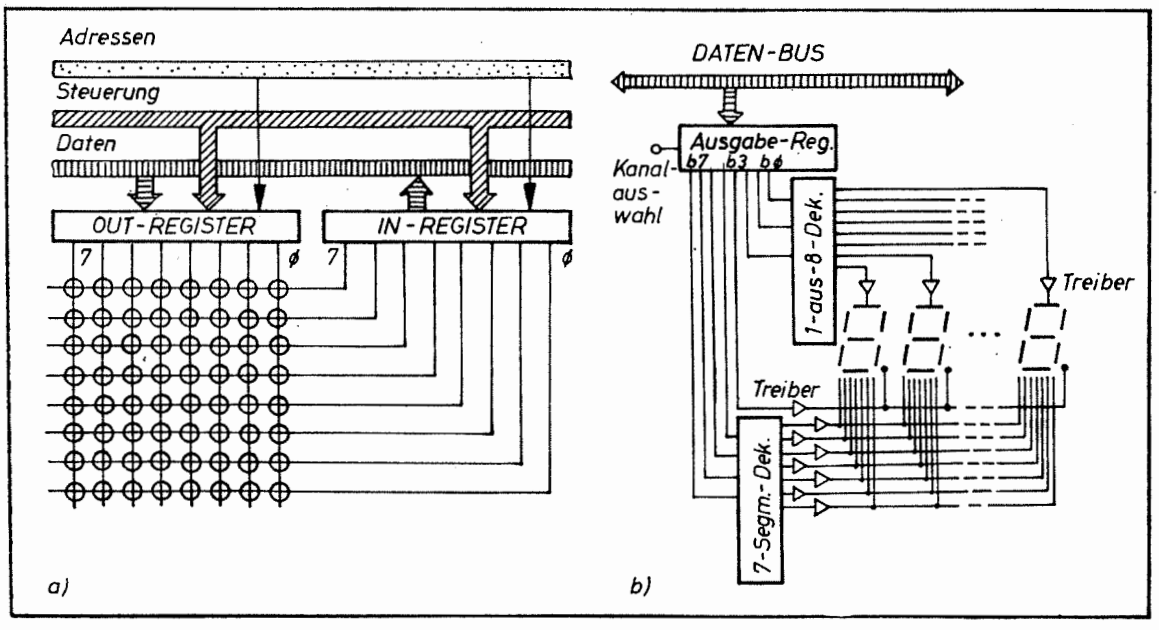


Bild 8.6.34. a) Eingabe per Tastatur und b) Ausgabe per Siebensegmentanzeigen

20-ms-Rhythmus kann weiter ein Prellen der Kontakte unterdrückt werden. Es muß dazu eben nur in zwei aufeinanderfolgenden Zyklen die gleiche Taste ermittelt werden.

Neben der hier geschilderten einfachen Tastenabfrage über je ein Ausgabe- und Eingabetor kann mittels Kodierung und Dekodierung auch über ein einziges Ein-Ausgabetor derselbe Effekt erreicht werden. Dies wird im Prinzip bei der folgenden Anzeige gezeigt. Weiter kann auch ohne Abfrage — einfach über Pegeltastung und Kodierung sowie Interrupt — eine Tastatureingabe erfolgen. In der Praxis existieren also vielfältige Methoden, von denen die in Bild 8.6.34a) gezeigte nur besonders übersichtlich ist.

Eine Anzeige mit bis zu acht Siebensegmentanzeigen zeigt Bild 8.6.34. Über ein Ausgabetor werden über 3 Bits, einen 1-aus-8-Dekoder und Treiberstufen die Katoden der einzelnen Ziffern angesteuert. Bit 4 geht direkt zu den Dezimalpunkten und die 4 höherwertigen Bits betreiben einen Siebensegmenttreiber, der die zusammengeschalteten Anoden der Anzeigen ansteuert. Dieses einfache Interface verlangt allerdings ein komplexes Programm zur Umwandlung der Zahlen, um sie dann umkodiert auf den Datenbus zur Ausgabe zu geben. Mittels der Bits bØ bis b2 wird zyklisch in acht Stufen der Reihe nach eine der Ziffern angesteuert. Auf die Bits 4 bis 7 ist dann kodiert der an der ausgewählten Stelle anzuzeigende Wert zu geben und auf b3, je nachdem ob ein Punkt erscheinen soll oder nicht, ebenfalls ein H oder L.

8.6.5.7. Fernseh-Interface

Den Aufbau eines FS-Interfaces zeigt Bild 8.6.35a). Für viele Anwendungen werden Fernsehempfänger zur Anzeige (display) benutzt. Dann sind Ablenkgenerator und Endstufe für die Helligkeitssteuerung sowie die Stromversorgung usw. unmittelbar gegeben. Die drei Signale Helligkeit, Bild- und Zeilenimpuls werden dem Fernsehempfänger dann kombiniert als BAS- (Bild-Austast-Synchron-) Signal zugeführt. Gegebenenfalls kann das BAS-Signal auch im Rechner noch auf einen FS-Kanal aufmoduliert und dann der HF-Eingangsbuchse zugeführt werden. Von diesen Details sei hier abgesehen. Weiter wird auch nur ein spezielles Prinzip des Fernseh-

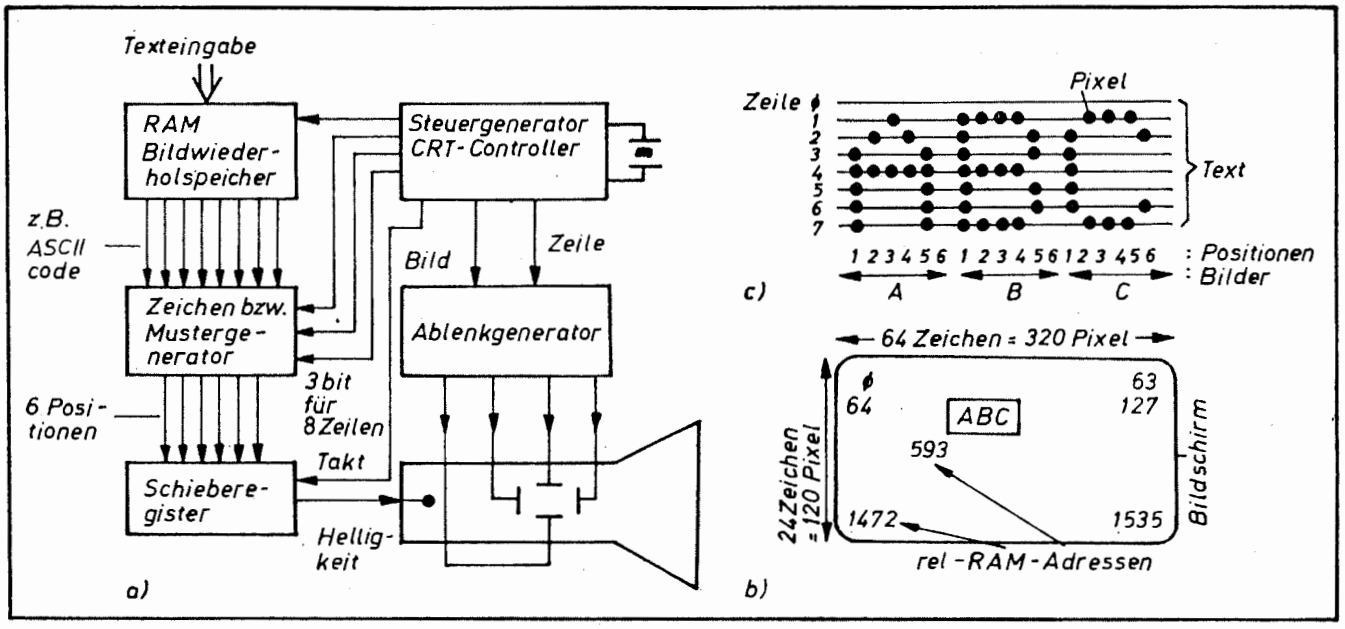


Bild 8.6.35. Prinzip des Videointerfaces

a) Schaltungsstruktur

b) Zuordnung von Bildschirmzeichen (-flächen) und relativen RAM-Adressen

c) Prinzip der Bilderzeugung über Pixel mittels Muster- (Zeichen-) Generatoren

displays, nämlich das quasigrafische — und zunächst nur zur Textdarstellung — behandelt.

Der Fernsehschirm ist zu diesem Zweck in Teilflächen für einzelne Zeichen eingeteilt. Im Teilbild b) sind dies 64 je Zeile in 24 Zeilen. Jeder dieser Teilflächen ist ein Speicherplatz im Bildwiederholspeicher von 1536 byte zugeteilt. Hier wird die darzustellende Information entsprechend den Positionen des Bildschirms eingelesen. Sie möge als ASCII- bzw. SIF-Kode gemäß Tab. 8.1.5 existieren. Der Steuergenerator ist vielfach ein wesentlicher Teil eines Spezialschaltkreises des CRT-Controllers (CRT = cathode-ray-tube = Katodenstrahlröhre = Bildröhre). Er leitet von einem Quarztakt mit einigen MHz einen zyklischen Adressendurchlauf für den RAM ab. Am Ausgang des RAM stehen so in der genau richtigen Zeitfolge die 7 oder 8 Signalwerte des Kodes entsprechend den gerade benötigten Zeichen zur Verfügung. Sie werden einem Mustergenerator (meist ROM für die Bildpunktkodierung) zugeleitet. Er wird seinerseits über drei Leitungen mit einem 1-aus-8-Dekoder angesteuert. Die 1-aus-8 entspricht dabei gemäß Teilbild c) der gerade durchlaufenen Zeile auf dem Bildschirm. Es wird hierbei vorausgesetzt, daß je Zeichen 8 Zeilen und 6 Bit-Positionen, also eine 5×7 Matrix plus je ein zusätzlicher Abstand gegenüber den anderen Zeichen verwendet wird. Vom Bildschirmwiederholspeicher kommt also der für das Zeichen nötige Kode und vom Steuergenerator die entsprechende Zeile. Dadurch werden die 5 (6) Werte für die Positionen parallel in ein Schieberegister übertragen und mit dem Takt sequentiell zur Helligkeitssteuerung der Bildröhre zugeführt. Mit der komplexen Steuerung entsteht so also das Bild auf dem Schirm.

Das geschilderte Verfahren ist viel universeller als nur zur Textdarstellung anwendbar. Sofern der Mustergenerator neben ROM- auch RAM-Zellen enthält, können in diese beliebige Zeichen über die 6×8 Pixel definiert (programmiert) werden. Dadurch ist dann auf dem Bildschirm eine beliebige Grafikdarstellung, im Beispielfall mit $320 \times 120 = 38400$ Punkten, möglich.

8.6.5.8. Speicher-Interface

Die Speicherung umfangreicher Daten erfolgt außerhalb der Mikrorechner, vor allem auf Bandgeräten (oft einfache Tonband- bzw. Kassettengeräte), auf Floppy discs und seltener auf Plattenspeichern (vgl. 7.7.5.). Hierbei können die Daten nur in sequentieller Folge längs einer Spur untergebracht werden. Es kommen deshalb die im Abschnitt 8.6.5.2. behandelten SIO für die Datenausgabe in Betracht. Darüber hinaus sind aber noch einige zusätzliche Besonderheiten zu berücksichtigen. Sie betreffen vor allem

Kodierverfahren, Adressierung und Fehlersicherung.

Die *Kodierverfahren* sind notwendig, da vom Magnetband- oder Floppy-disc-Speicher infolge der induktiven Magnetköpfe nur differenzierte Signale, also kein Gleichwert, übertragen werden kann. Dadurch ist es nicht möglich, eine beliebige Ø1-Folge als NRZ (non return to zero) direkt zu übertragen (Bild 8.6.36a)). Es würden dann nämlich nur Impulse an den dargestellten Flanken wirksam. Aus einem solchen Signal ist praktisch nicht die Ø1-Folge — allein schon infolge auftretender Zeitfehler — wiederzugewinnen. Es muß vielmehr ein Taktimpuls T gemäß der FM-Aufzeichnung (auch Biphase oder Wechseltaktschrift genannt, Teilbild b)) eingefügt werden. Dies ist ein besonders störsicheres Verfahren, welches aber infolge der vielen Flußwechsel auf dem Band nur eine relativ geringe Speicherdichte zuläßt. Daher wird es bei Floppy-disc-Speichern als Single density bezeichnet. Läßt man jeden zweiten Impuls weg, so entsteht die MFM (modified FM, Teilbild c)), auch MILLER-Kode

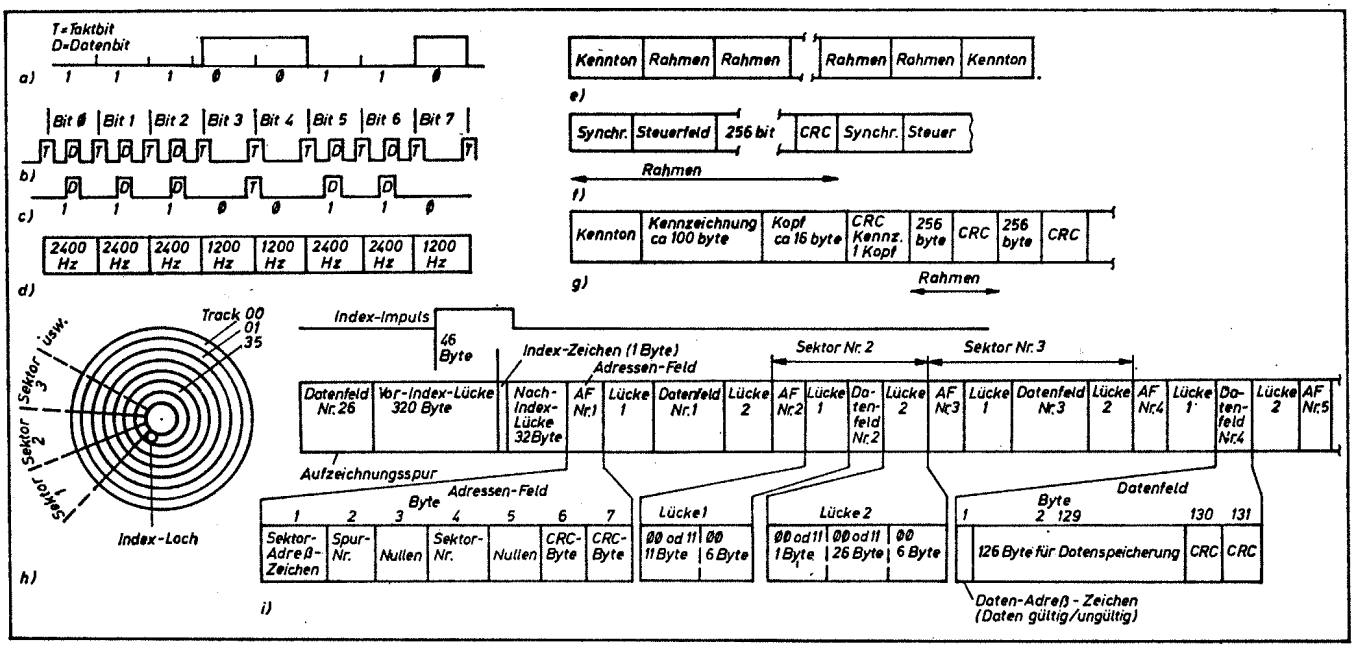


Bild 8.6.36. Prinzipien der Speicherung auf magnetischen Medien

a) NRZ-Signal eines Bytes
b) Zugehöriges Biphase-Signal (FM-Signal)
c) Dito als MFM (modifiziertes FM)
d) als Signalfolge bei Kassettenrecordern
e) Signalfolge auf Audiorecordern
f) Aufbau eines speziellen Rahmens
g) Ein anderes Prinzip für Audiorecorder
h) Formatierung eines Floppy disc in Spuren (Track) und Sektoren
i) Detaillierte Sektoreinteilung beim Standardformat IBM 3740

genannt. Sie führt zur Double density mit höherer Speicherdichte, aber höherem Kodier- und Dekodieraufwand.

Es gibt noch eine Vielzahl weiterer Kodierverfahren. Sie besitzen aber bei den Mikrorechnern nur geringe Bedeutung. Es sei deshalb nur noch der Gruppenkode erwähnt. Dabei werden jeweils vier Signalbits um ein redundantes fünftes so ergänzt, daß eine bestimmte Anzahl von Flußwechseln nicht über- und unterschritten wird und so sowohl Takt als auch Signal relativ gut bei höchster Speicherdichte zu gewinnen sind.

Bei der einfachsten Aufzeichnung auf billigen Kassettenrecordern wird oft ein viel einfacheres Verfahren angewendet. Es kodiert z.B. die 1 durch 10 Schwingungen von 2400 Hz und die Ø durch 5 Schwingungen von 1200 Hz (Teilbild d)). Die Aufzeichnung auf dem Kassettenband (Teilbild e)) beginnt mit einem Kennton, dem sich Rahmen anschließen. Sie können z.B. gemäß Teilbild f) aus einem Synchronisationssignal, einem Steuerfeld (welches Hinweise bezüglich der folgenden Daten enthält), 256 bit Daten und einem Kontrollkode CRC (siehe unten) bestehen. Bei höheren Ansprüchen folgt dem Kennton zunächst ein umfangreiches Kennzeichnungsfeld, dann ein Kopf mit speziellen Daten, ein CRC für die beiden Eingangsfelder, und dann schließen sich Rahmen an, die z.B. aus 256 byte Daten mit dem zugehörigen CRC bestehen.

Bei Floppy-disc- und Plattenspeichern kommt für die Aufzeichnung ein weiterer Service hinzu. Die Daten können definiert nach Adressen abgelegt und wiedergefunden werden. Dies sei am Beispiel des Standardformates für Floppy disc erklärt. Gemäß Bild 8.6.36h) existieren hier 35 genau definierte Spuren, von denen wieder jede in 26 Sektoren geteilt wird. In jedem Sektor existiert ein Datenfeld für 126 byte. Die Strukturierung der Sektoren in einer Spur zeigt detaillierter Teilbild i). Für die verschiedenen Floppy discs und Plattenspeicher existieren auch hier viele ab-

weichende Formate. Die Aufgabe eines Floppy-disc-Controllers ist es, sowohl die Signale zu erzeugen als auch zu erkennen und zu verwalten. Er muß also aus einfachen Befehlen ableiten können, wo die gewünschten Daten stehen bzw. hingeschrieben werden. Zu diesem Zweck verfügt er über eine Sonderspur, in der er sich den jeweils aktuellen Stand der gespeicherten Information notiert und aus der er die Positionierung des Schreib-Lese-Kopfes auf die entsprechende Spur und die Aufteilung bzw. Wiedergabe in dem gewünschten Sektor realisiert. Für diese Aufgaben existieren ebenfalls spezielle Schaltkreise.

Die *Fehlersicherung* geschieht in der Regel über den schon mehrfach erwähnten CRC (= cyclic-redundancy-check = zyklischr Redundanztest). Es existiere eine Bitfolge $b_7 b_6 \ldots b_1 b_0$. Sie wird mit einer BOOLEschen Variablen x, die die Werte Ø und 1 annehmen kann, als Polynom interpretiert:

$$b_7 x^7 + b_6 x^6 + b_5 x^5 + b_4 x^4 + b_3 x^3 + b_2 x^2 + b_1 x + b_0 \,. \tag{1}$$

Diese Variable x wird beim späteren Test durch das gerade gelesene Bit ersetzt. Ein Binärwort 1ØØØ ØØ11 führt damit zu einem Polynom

$$x^7 + x + 1 \,. \tag{2}$$

Derartige Polynome werden auch als Generatorpolynome zur Erzeugung des CRC-Wertes benutzt, indem sie mit den aktuellen Signalwerten beschickt werden und die entstehenden Werte als Bytes, also modulo 256, aufaddiert und gespeichert werden. Beim Test werden die eingehenden Bytes durch dasselbe Generatorpolynom dividiert. Der dabei eventuell entstehende Rest zeigt einen Fehler an.

Bei der SIO U 856 sind folgende Generatorpolynome implementiert:

$$x^{16} + x^{15} + x^2 + 1 \text{ (SDLC)} \tag{3a}$$

$$x^{16} + x^{12} + x^5 + 1 \,. \tag{3b}$$

8.6.6. Bedeutung der Informationen

Im Mikroprozessor und im Rechner überhaupt existieren nur binäre Signale (Bits), die auf die beiden Pegel H und L bzw. die ihnen zugeordneten Werte Ø und 1 zurückgeführt werden können. Diese Bits können unterschiedlich strukturiert sein, sowohl was ihre zeitliche Abfolge als auch was ihre Parallelität in Bytes bzw. Worten betrifft. Sie werden schließlich an unterschiedlichen Geräte- bzw. Schaltungsteilen wirksam, zu diesem Zweck werden sie z. T. über spezielle Busse, z.B. Adreß-, Daten- oder Steuerbus, geführt. Dieser Kontext gibt den einzelnen Bits bzw. Bytes einen Sinn, eine Bedeutung, und hiernach werden sie eingeteilt. Infolge der Komplexität des Zusammenwirkens und des zeitlichen Auftretens dieser Werte ist ihre Klassifikation nicht ganz einfach. Bild 8.6.37 versucht einen allgemeinen Überblick. Dabei werden die Informationen zunächst in vier Hauptklassen eingeteilt:

Adressen, welche in etwa auf Orte von Schaltungsteilen, z.B. Register, Speicher oder Tore, hinweisen.

Befehle, welche eine bestimmte komplexe Reaktion auslösen. Sie bestehen in der Regel aus einem *Befehlskode*, der auf die Art und Weise des Tuns hinweist und damit ein spezielles Steuersignal, z.B. für die ALU oder das Steuerwerk eines Mikroprozessors, darstellt. Ihr zweiter Teil ist der oder die Operanden, mit denen etwas geschehen soll. Er kann daher eine Adresse oder ein Datenteil sein.

Daten sind Werte, mit denen gearbeitet wird. Dies können u. a. binäre oder BCD-Zahlen (auch einzelne Bits oder Flags), alphanumerische Werte, z.B. gemäß Fern-

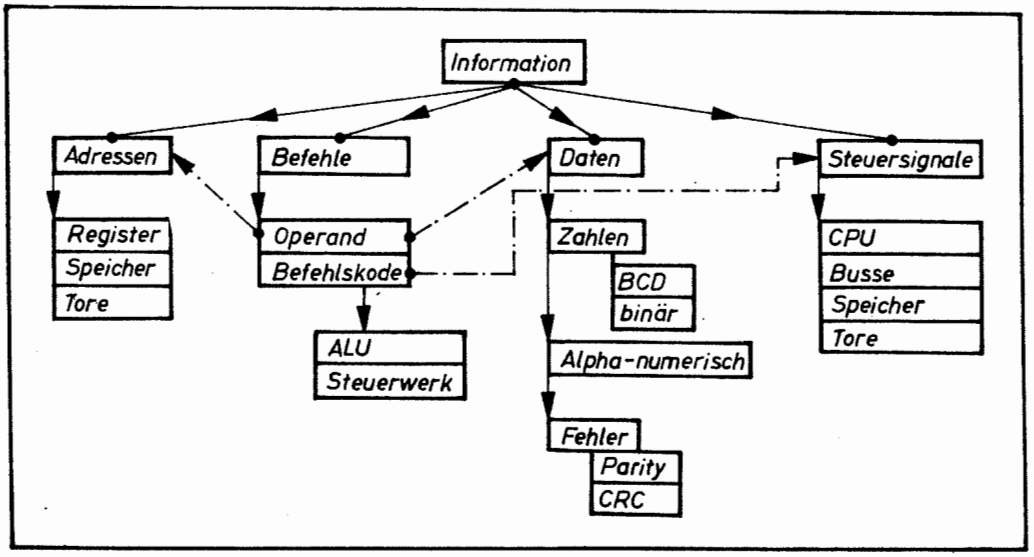


Bild 8.6.37. Einteilung der Informationen, die in einem Mikrorechnersystem auftreten

schreib- oder ASCII-Kode, bzw. Fehlerwerte, z.B. als Party-Bit oder CRC-Byte, sein.

Steuersignale sind den Befehlen sehr ähnlich. Während Befehle meist als komplex gegebener Bytesatz existieren, sind Steuersignale meist als Einzelbit, wie z.B. Write, Read, Interrupt usw. vorhanden. Wenige Steuersignale, z.B. beim Handshake, existieren als Doppelbit, das dann aber in unterschiedliche Richtungen weist. Steuersignale können u. a. danach klassifiziert werden, auf welche Baugruppen sie einwirken, z.B. CPU, Busse, Speicher, Tore.

8.6.7. Geschichte der Mikroprozessoren

8.6.7.1. Vom 4004 zum Z 80 und 6502

Die Entwicklung des ersten Mikroprozessors geht auf einen Auftrag der japanischen Firma Busicom zurück, die programmierbare Tischrechner herstellt. Für einen neuen Typ gab sie Intel den Auftrag, spezielle Rechnerchips zu entwickeln. MARCIAN E. „TED" HOFF hielt die von Busicom vorgeschlagene Lösung für die damalige Technologie für zu komplex und kostenintensiv. Es wären 3000 bis 5000 Transistoren je Chip mit 36 bis 40 Anschüssen notwendig gewesen. Aufgrund seiner Erfahrungen bei DEC mit der Struktur des Rechners PDP-8 versuchte er, statt einer reinen Hardwarelösung eine programmgesteuerte Variante zu finden. So entstand die CPU 4004 mit ca. 2300 Transistoren. Die Entwicklung der weiteren notwendigen Bausteine ging jedoch zunächst wegen der Unterbesetzung bei Intel schlecht voran. Einen wesentlicher Fortschritt entstand, als FEDERICO FAGGIN (der spätere Gründer und Präsident von Zilog) von Fairchild zu Intel 1970 überwechselte. So wurde 1970 mit Busicom der Rechner vollendet. Es waren statt der ursprünglich geplanten 12 nur 4 Chips (der Satz HCS-4) notwendig. Diese waren: CPU, ROM-Chip. RAM-Chip und Schieberegister-Chip. Für eine effektive Produktion war der Bedarf an Schaltkreisen zu gering. So entschieden sich beide Firmen für breitere Anwendung auch für andere Zwecke, und am 15. 11. 1971 erschien die erste Werbung in "Electronics News". Bis Februar 1972 hatte dann Intel 85000 Chipsätze verkauft.

Entsprechend dem Auftrag von Japan war der erste Mikrorechner eine 4-bit-Variante im BCD-Format. Doch schon bei dieser Entwicklung erkannten die Entwickler, daß für viele Anwendungen 8-bit-Prozessoren vorteilhafter wären. Hierzu gab es auch Aufträge, die jedoch aus verschiedenen Gründen nie praxiswirksam wurden. Dennoch wurde bis zum April 1972 der Intel 8008 fertiggestellt. Bei ihm zeigte sich sehr bald, daß er für viele Anwendungen zu langsam war. Als Konsequenz brachte Intel im April 1974 den von FAGGIN entwickelten 8080 der zweiten Generation heraus. Er galt für viele Jahre als Industriestandard und besitzt sogar heute noch vielfältige Anwendungen.

Diese Entwicklung bei Intel löste viele Aktivitäten bei anderen Halbleiterfirmen aus, und so entstanden andere Neuerungen bei anderen Firmen, z.B. der 6800 von Motoriola (1974) und bei National Semiconductor "Pace", der erste 16-bit-Prozessor, der allerdings keine Bedeutung erlangte.

Die dritte Generation der Mikroprozessoren beginnt um 1976. Hier brachte Zilog den wiederum von FAGGIN entwickelten Z 80 und die ehemaligen Motoriola-Entwickler bei MOS-Technologie den aus dem 6800 abgeleiteten 6502 heraus. Sie zählen noch heute zu den am häufigsten verwendeten 8-bit-Prozessoren.

Das schnelle Anwachsen des Mikroprozessormarktes zeigt allein eine Zahl: 1980 wurden rund 150 Millionen Stück verkauft.

8.6.7.2. Marktanteile und Tendenzen

Für 1980 galten bei den Mikroprozessoren folgende Werte:

Bitbreite	Stückzahl/%	Geldwert/%	1985 Geldwert/%
4	66	18	17
8	33	65	48
16	1	17	28
32	—	—	7

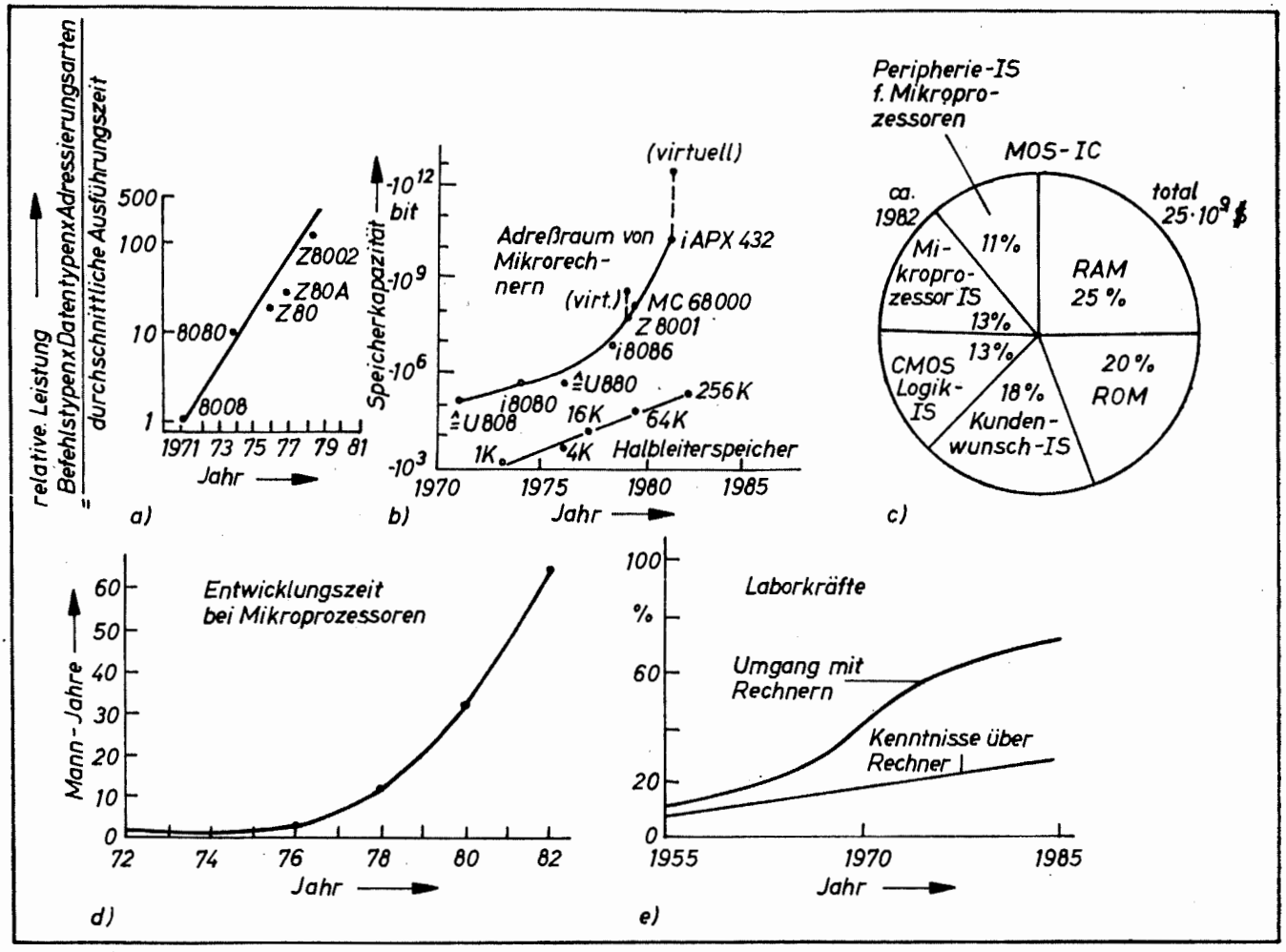


Bild 8.6.38. Tendenzen bezüglich Rechnern und Mikrorechnern

a) Entwicklung der Leistungsfähigkeit
b) Entwicklung von Adreßraum bei Mikroprozessoren und Kapazität bei Speichern
c) Relatives Marktvolumen der verschiedenen MOS-IC
d) Steigender Entwicklungsaufwand bei Mikroprozessoren
e) Kenntnisse und Umgang von Laborkräften bezüglich Rechnern

Dabei wird die Gesamtstückzahl für 1980 auf 150 Millionen, der Geldwert auf 200 Millionen Dollar und für 1985 auf rund 1 Milliarde Dollar geschätzt. Den Hauptanteil bei den 8-bit-Prozessoren besaßen 1980 folgende Typen:

8080/8085 (6); Z 80 (5); 6800 (4); 6502 (3).

Die Klammerwerte sind verkaufte Millionen Stück.

Bei den Zahlen von Mikroprozessoren ist zu beachten, daß dazu ein noch weitaus größerer Markt für periphere Bausteine, also Speicher, I/O-Ports usw. hinzukommt. Das Verhältnis in Marktanteil CPU: ergänzende Bausteine beträgt 1982 32:68 und wird für 1987 auf 24:76 geschätzt.

Einen weiteren Überblick zu Tendenzen gibt Bild 8.6.38. Einmal ist in Teilbild a) die schnelle Zunahme der relativen Leistungsfähigkeit von Mikroprozessoren aufgezeigt. Teilbild b) zeigt die Zunahme des Adreßraumes von Prozessoren und der Kapazität von Speicherchips.

Auffällig sind die ab 1976 enorm zunehmenden Entwicklungsaufwendungen für Mikroprozessoren (Teilbild d)). Sie erreichen von da ab eine Komplexität, die nur noch iterativ mit Rechnern durch ein beachtliches Kollektiv zu erreichen ist. In der Pionierzeit reichten Aufwendungen um 1 Mannjahr, um 1982 waren dagegen bereits über 60 Mannjahre notwendig.

Den Anteil der Mikroprozessoren, die zu über 90% in MOS-Technik gefertigt werden, am Gesamt-MOS-Aufkommen zeigt für das Jahr 1982 Teilbild c). Schließlich zeigt Teilbild e) noch die Kenntnisse von Laborkräften bezüglich Rechnern und die weitaus höhere Anwendung von Rechnern im Labor. Besonders auffällig ist der steile Anstieg der Anwendung mit der Einführung von Mikrorechnern um das Jahr 1975.

8.7. Standard-Busse; Standard-Interface

8.7.1. Überblick

Interface gehört zu einer Schnittstelle zwischen zwei oder mehreren Systemen. Hier sind verschiedene Festlegungen zu treffen, damit die Geräte zusammenarbeiten können. In vielen Fällen erfolgt diese Zusammenarbeit über Bussysteme. Deshalb werden Begriffe hier z. T. synonym verwendet (vgl. auch Bild 8.7.1 d)).

Bereits im Zusammenhang mit Bild 8.6.4 wurde die Bedeutung von Bussystemen hervorgehoben. Bei Mikrorechnern zeigte sich dann immer deutlicher, wie wichtig die richtige Gestaltung von Bussystemen ist. Diese Entwicklung hat sich in dem letzten Jahrzehnt konsequent auf allen Gebieten vollzogen, wo Daten zwischen mehreren Geräten oder Einrichtungen in veränderbarer Weise ausgetauscht oder übertragen werden müssen. Zunehmend existiert heute eine Vielzahl von Bussystemen, die teils auf unterschiedliche Entwicklungen zurückzuführen sind, teils unterschiedliche Anforderungen erfüllen müssen. Dies hat dazu geführt, daß einerseits die Vielzahl nahezu unübersehbar ist und andererseits sich die verschiedenen internationalen Normungsausschüsse bemüht haben, generelle Festlegungen zu treffen. Dabei haben sich bisher nur wenige Bussysteme allgemein durchgesetzt. Es hat den Anschein, daß das Gebiet der Standardisierung von Bussystemen noch sehr im Fluß ist. Dies erschwert hier eine allgemeingültige Darstellung. Es wird folglich eine Klassifikation versucht, und einige ausgewählte Bussysteme werden genauer dargestellt. Für einige insbesondere auf die Meßtechnik bezogenen Busse — nicht alle hier behandelten — gibt das Buch [N 1] eine ausführliche Darstellung. Der älteste Standardbus ist CAMAC (von 1968), die meisten Bussysteme entstanden Mitte bis Ende der 70er Jahre.

Generell können Bussysteme nach folgenden Kriterien unterschieden werden:
Je nach der Art der Daten existieren *analoge und digitale Busse*. Die analogen Busse haben vor allem in der Konsumgüterindustrie, z.B. bei der Studio- und Fernsehtechnik, Bedeutung. Sie interessieren daher hier nicht. Weiter werden folglich nur digitale Busse behandelt.
Die Übertragung von Daten kann *parallel oder seriell* erfolgen. Sie wurde schon bei den PIO und SIO im Abschnitt 8.6.5. behandelt und gilt natürlich auch für Standard-

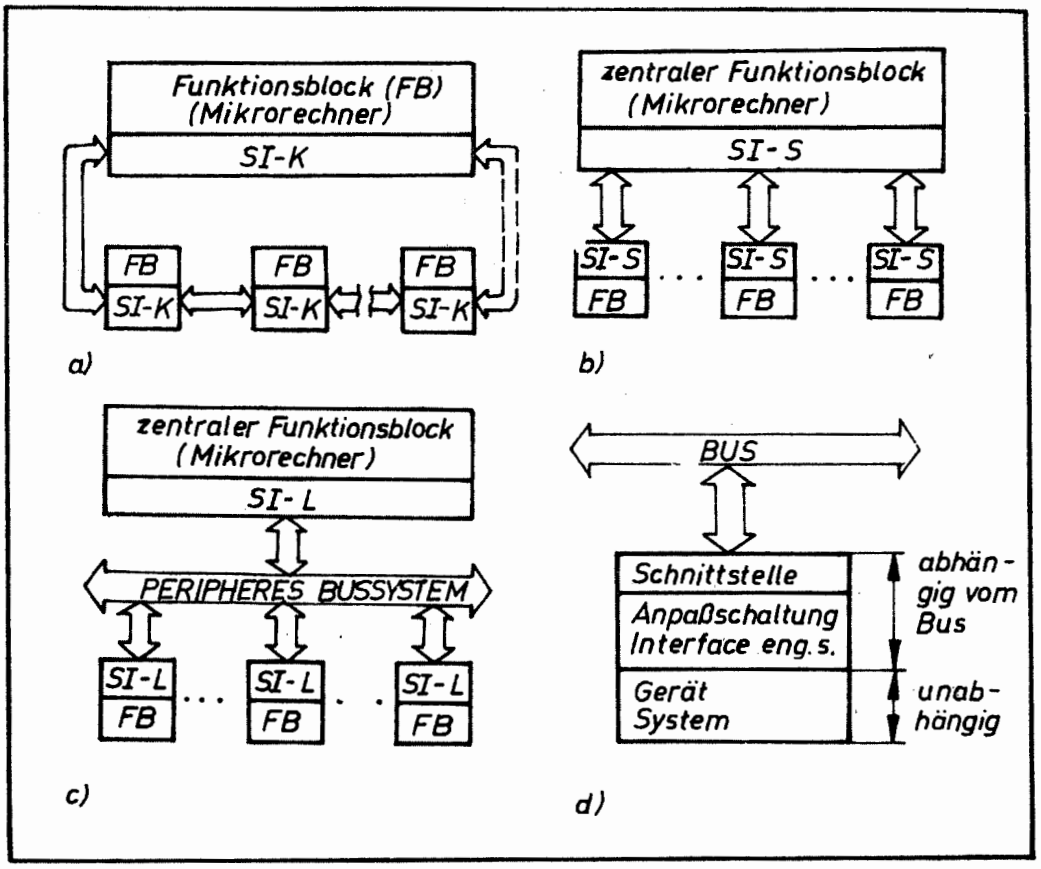


Bild 8.7.1. a) Kettenförmiger, b) sternförmiger, c) linienförmiger Bus
d) Zur Abgrenzung von Bus, Schnittstelle und Interface im engeren Sinne (als Schaltung, die Gerät und Busschnittstelle verbindet)

Busse. Aus den dortigen Ausführungen geht auch hervor, daß es verschiedene Methoden der *Datensicherung* gibt, z.B. mittels Handshake oder durch Synchron-, Takt-, Parity- oder CRC-Signale.
Bezüglich des Datenaufbaus besitzen *Zahlenwerte* wie Breite des Datenwortes, Übertragungsgeschwindigkeit, Interrupt-Ebenen, Anzahl der anschaltbaren Geräte, zulässige Länge der Busleitungen usw. Bedeutung. Hier liegt der eigentliche Grund, weshalb verschiedene Bussysteme benötigt werden.

Ein weiterer Grund für die Vielfalt der Bussysteme liegt in den *Anwendungsgebieten*. Hier können unterschieden werden

- *integrierte Busse*, Bussysteme auf dem Halbleiterchip;
- *Leiterplattenbusse*, sie existieren auf einer Leiterplatte zwischen den Bausteinen;
- *Rechner-System-Busse*, sie verbinden die verschiedenen Baugruppen eines Rechners und wurden für den Fall der Mikrorechner im vorigen Abschnitt behandelt:
- *Instrumentationsbusse*, sie sind in der Regel genormt bzw. standardisiert und dienen zum Datenaustausch zwischen sonst auch einzeln verwendbaren Meßgeräten. Daher auch die Bezeichnung Meßgerätebusse.
- *Prozeßbusse*, sie stellen im wesentlichen die Spezialform der Instrumentationsbusse mit serieller Übertragung dar;
- *Netzwerkbusse*, auch Bürobusse, Rechnernetzwerke (local area network) usw. genannt. Sie verbinden verschiedene Rechner und verschiedene Gerätesysteme eines größeren Gebietes zur gemeinsamen Kommunikation.

Wenn die Abfolge genauer betrachtet wird, ergibt sich auch eine *Hierarchie* der Bussysteme, z.B. im Sinne von Chip, Leiterplatte, Geräte, Netzwerk.

Tabelle 8.7.1. Einige Merkmale von Interfaceklassen. FE bedeutet Funktionseinheit ≈ Gerät

Merkmal	Kettenschaltung	Sternschaltung	Busschaltung
Informations-austausch	unmittelbar von FE zu FE	unmittelbar von FE an zentrale FE und von dieser evtl. an andere FE	von FE über Bus zur zentralen FE und evtl. an andere FE
Steuerung	FE steuern sich gegenseitig	erfolgt unmittelbar durch zentrale FE	erfolgt über Bus in der Regel von zentraler FE aus
Adressierung	nicht erforderlich	nicht erforderlich	erfolgt über Bus von zentraler FE aus
Programmierung (wenn erforderlich)	FE gegenseitig und/ oder durch eine Programmier-FE	unmittelbar durch zentrale FE	durch zentrale FE über Bus
Erforderliche Signalarten	Informations-, Steuer-, Programmsignale	Informations-, Steuer-, Programmsignale	Informations-, Steuer-, Programm-, Adressen-, Sondersignale

Die *Verschaltung der Geräte* und Systeme kann auf verschiedene Weise erfolgen. Dementsprechend haben sich drei Gruppen herausgebildet. Sie sind in Bild 8.7.1 dargestellt, während Tab. 8.7.1 die Eigenschaft der Schaltung hervorhebt:

- Beim *Ketten-Interface* (Standard-Interface-Kette = SI-K) werden die Geräte hintereinandergeschaltet. Das erste Gerät ist der Master, die anderen folgen als Slave. Zuweilen geht vom letzten Gerät auch wieder eine Leitung zum ersten zurück. Dann wird auch vom ringförmigen Interface gesprochen.
- Beim *Stern-Interface* (SI-S) hat ein Gerät die absolut dominierende Funktion. Nur von ihm aus können alle anderen Geräte angesteuert werden, und alle Geräte können auch nur hierüber untereinander wechselwirken. Dieses Beispiel gilt für den im vorigen Abschnitt behandelten Mikrorechner, wenn vom DMA-Betrieb abgesehen wird.
- Das *Linien-Interface* (SI-L) ermöglicht es, daß im Prinzip jedes Gerät mit jedem direkt kommuniziert. Damit dabei nicht gleichzeitig mehrere Geräte ihre Daten auf den Bus geben können und so ein Chaos entsteht, legt das zentrale Gerät (der Rechner) fest, welches Gerät wann Daten senden darf.

Da das Ketten-Interface heute nur noch geringe Bedeutung hat, seien nur die Vor- und Nachteile der beiden anderen Systeme gegenübergestellt:

Sternverkehr: Vorteile:

- jedes Gerät kann individuell optimal angepaßt werden;
- minimaler Grundaufwand für die Anschlußbildung;
- keine Störungen der Geräte untereinander möglich.

Nachteile:

- nur relativ wenige Geräte anschließbar;
- großer Aufwand an Kabeln und Steckern.

Linienverkehr: Vorteile:

- Aufbau aus Gerätemoduln, daher sehr übersichtliche Gestaltung;
- minimaler Verkabelungsaufwand;
- einheitliche Gestaltung des Informationsaustausches, daher für Rechner vorteilhaft.

Nachteile:

- alle Geräte bezüglich Logik und Signalpegel identisch (hohe Standardisierung);
- Leitungsreflexionen können Störungen bewirken;
- hoher Grundaufwand auch bei wenigen Geräten.

Als letzter Aspekt sei behandelt, welche *Kopplung* an einem Bus *standardisiert* wird:

- Die *informationelle Kopplung* muß in gewissem Umfang immer standardisiert sein. Sie betrifft die Festlegung der Signale, ihrer Pegel und logischen Bedeutung (z. B. positive und negative Logik), die Anzahl der Leitungen für die einzelnen Signale und z. T. die Belegung der Stecker.
- Die *energetische Kopplung* wird dann notwendig, wenn mit dem Bus auch Stromversorgungen realisiert werden. Dann sind die Spannungen, Belastungen und ihre Toleranzen sowie eventuell geregelte oder ungeregelte Spannungen festzulegen.
- Eine Festlegung der *stofflichen oder konstruktiven Kopplung* erfolgt nur selten. Sie betrifft dann Gehäusegrößen, Einschubgrößen usw.

8.7.2. Kettenförmiges System SI 1.2

Im RGW ist das Informationsmeßsystem IMS-1 standardisiert. In der Pegelkategorie II (TTL-Pegel) ist es gleichwertig mit dem in der DDR standardisierten Standard-Interface SI 1.2. Dies ist eines der wenigen Ketten-Interfaces. Bei ihm werden neben den Programmsignalen (P) *drei Signalarten* unterschieden (die wichtigsten Signale sind benannt):

- *I-Signale* (Information; Daten): bevorzugt ist hier 28 bit Wortbreite und die Bedeutungen BCD und ISO-7-bit (Tab. 8.1.5)
- *B-Signale* (Befehle). Es existieren:

BØ	Lösch-Start-Signal
B1, B2	Auslösungssignale
B4	Fehlererkennungssignal
B5	Gültigkeit für Programmsignale
B6	Fernsteuerung

- *M-Signale* (Meldesignale):

M1	Bewertung von I- und P-Signalen
M2	Gültigkeit für I- und P-Signale
M3	Bereitschaftsignal des Gerätes
M4	Fehlersignal
M5	Bewertung der P-Signale

8.7.3. Sternförmige Systeme

8.7.3.1. Sif 1000

Das Standardinterface daro 1000 wurde zum Datenverkehr zwischen Prozeßrechnern und der Standard-Peripherie in der DDR entwickelt. Es weist *folgende Merkmale* auf (Bild 8.7.2a)):

- sternförmige Struktur,
- getrennte 8 bit breite Datenkanäle für Ein- und Ausgabe DAT-E und DAT-A,
- Zweidraht-Handshake mit den Signalbezeichnungen RUF und END,
- je 3 Leitungen für Status- und Kommandosignale STA, KOM,
- Möglichkeit für ungerades Paritätsbit der Daten PB-A, PB-E,
- Möglichkeit eines Interrupts IR-A, IR-E.

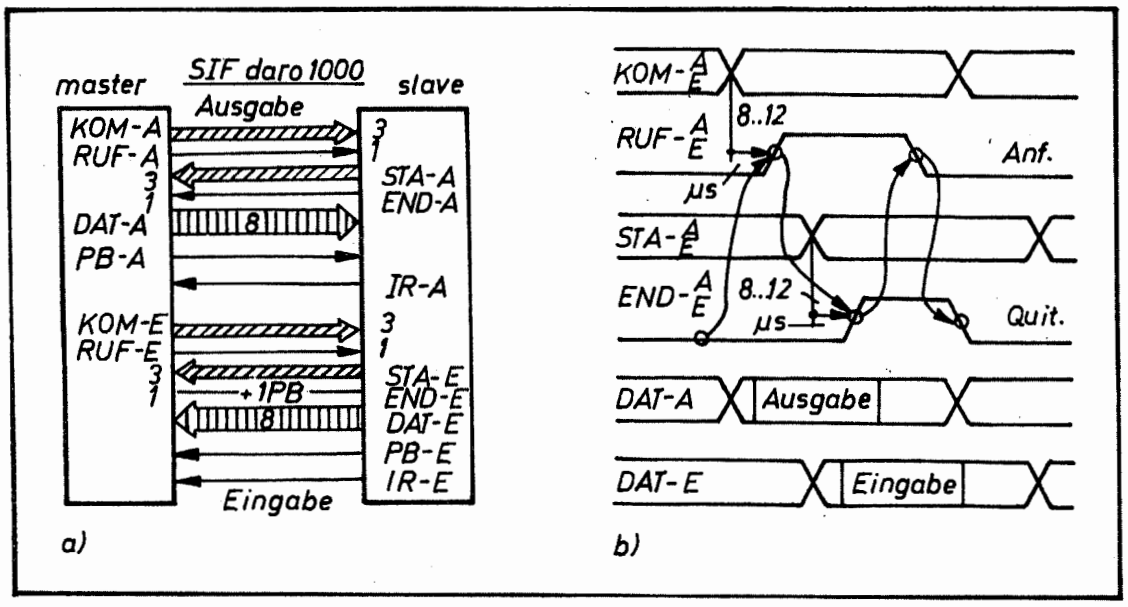


Bild 8.7.2. a) Signale und b) Zeitabläufe beim Sif daro 1000

Bevorzugt verwendet wird der ISO-7-bit-Kode (Tabelle 8.1.5). Es sind zwei Pegel zugelassen: für MOS 1: −11 bis −13 V und Ø: 0 bis −2 V sowie TTL. Die maximale Übertragungsrate liegt bei 900 Zeichen/s. Den zeitlichen Signalablauf zeigt Bild 8.7.2b).

Die Statussignale werden u. a. für Fehleranzeige verwendet. Die Kommandosignale werden meist vor der Datenübertragung verwendet und stellen bestimmte Zustände in dem Empfangs- bzw. Sendesystem her.

8.7.3.2. S-100-Bus: IEEE 696

Mit dem Mikroprozessor 8080 brachte die Firma Altair 1975 den ersten Mikrorechnerbausatz so preiswert auf den Markt, daß sich hieraus die Personalcomputer entwickelten. Gleichzeitig wurde hierbei ein Bussystem entwickelt, das sich breit einführte und erst um 1980 zum IEEE Standard 696 wurde. Die Bezeichnung S-100 rührt daher, daß 100 Leitungen verwendet werden. Sie sind der jeweiligen Leiterplatte oben und unten aufkaschiert. Es werden die folgenden Signale verwendet:

16 Adreßleitungen
8 Dateneingangsleitungen,

8 Datenausgangsleitungen,
44 Leitungen zur Steuerung, u. a.
Takt, Interrupt, Status, Schreiben, Lesen,
6 Leitungen zur Stromversorgung,
+8 V, +16 V, −16 V ungeregelt,
18 Reserveleitungen.

Die Fülle dieser Leitungen erscheint sehr redundant. Sie hat aber den Vorteil, daß eine sehr große Variabilität besteht. Auf dem internationalen Markt existieren für diesen Bus mehrere hundert Leiterplatten, die direkt verwendet werden können. Dies war nur möglich, weil die Spannungsversorgung und die Leiterplattenabmessungen mit standardisiert wurden. Ebenso wurde vorgeschrieben, daß jede Leitung mit 10 Standard-TTL-Lasten belastbar sein soll. Ferner trug die große Anwendungsbreite des 8080 und später des Z 80 zu dieser Entwicklung bei.

8.7.3.3. Weitere Mikrorechnerbusse

Neben dem S-100-Bus, der eigentlich durch keine Firma vertreten wird, erlangten folgende Mikrorechnerbusse größere Bedeutung

Unibus bzw. LSI-11-Bus von DEC (Digital Equipment Corporation),
Multibus von Intel.

Die ersten beiden Busse verwenden 72, der Multibus 86 Leitungen. Die Adreßbreite beträgt 18 bzw. 16 bit, die Datenbreite 16 bit. Standardisiert wurde mit IEEE 796 in etwa der Multibus. Auch hier wurden wieder Leiterplattenabmessungen und Betriebsspannungen (allerdings geregelt +5, −12, +12 V) festgelegt. Von kleineren Systemen haben eine gewisse Bedeutung der SS-50-Bus, der Apple-Bus und der TRS-80-Bus.

8.7.4. Linienförmige Systeme

Die linienförmigen Bussysteme sind dadurch gegenüber den anderen ausgezeichnet, daß die peripheren Geräte mit Adressen versehen werden müssen und daß ohne Einbeziehung des Rechnersystems bzw. der CPU ein direkter Datenaustausch zwischen diesen peripheren Geräten möglich ist. Während bei den sternförmigen Systemen die Adressen nur dazu dienen, die Geräte bzw. Systeme vom Rechner zu erreichen, ihnen Daten zuzuführen bzw. von ihnen zum Rechner Daten zu vermitteln, geschieht bei den linienförmigen Systemen der Datentransport nach der Programmierung durch den Rechner von einem peripheren Gerät zu einem oder mehreren anderen. Die drei Hauptsysteme sind der IEC-Bus, SI 2.2 und CAMAC. Hiervon wird der IEC-Bus wegen seiner großen Bedeutung ausführlicher besprochen.

8.7.4.1. IEC-Bus

Der IEC-Bus wurde in nur wenig anderer Form zuerst von Hewlett-Packert (HP) verwendet und wird deshalb auch oft als HP-Bus bezeichnet. Von der International Electrotechnical Commission (IEC) wurde er von der Untergruppe TC 66 (Technical Comitee) international genormt. In diesem System ist ein HP-Patent bezüglich des Dreileiter-Handshake enthalten. Es wurde bei der Normung erreicht, daß HP jedem Nutzer unter gleichen Bedingungen die Lizenznahme zusichert. Vom IEEE (Institute of Electrical Electronic Engineers) erfolgte die Normung als IEEE 488. Entsprechend dem Ablauf auf den Datenleitungen wird auch oft von byteseriellen-bitparallelen Interfaces gesprochen. Im RGW wird die Bezeichnung IMS-2 verwendet. Zwischen den verschiedenen Normen gibt es z. T. geringfügige Abweichungen.

Der IEC-Bus ist durch *folgende Fakten* gekennzeichnet:

- byteserielle-bitparallele asynchrone Datenübertragung mit einem speziellen Dreileiter-Handshake,
- TTL-Pegel mit z. T. positiver oder negativer Logik mit Fan-out 30, mit Verlusten also 15 Lasten,
- direkt sind 31 Datensender (Talker = Sprecher) und 31 Datenempfänger (Listener = Hörer) adressierbar,
- maximale Datenrate von 1 Mbyte/s,
- maximale Länge des Übertragungsweges 20 m mit Verkabelung über Huckepackstecker.

Tabelle 8.7.2. Überblick über die Busleitungen des IEC-Bus

Leitungsgruppe		Zahl der Leitungen	Bezeichnung der Leitung	Aussage der L-Pegel auf der Leitung	gesendet von
Data Bus	Datenbus	8	DIO 1 . . . . . . DIO 8 (data input output	Datenbit ist log. 1	Steuerer oder Sprecher
Data Byte Transfer Control Bus	Steuerbus	3	DAV (data valid) NRFD (not ready for data NDAC (not dat accepted)	Die Nachricht auf DIO 1 ... DIO 8 — ist seitens des Steuerers/Sprechers gültig — kann von den FB noch nicht angenommen werden — wird von den FB angenommen (Annahme noch nicht beendet)	— Steuerer oder Sprecher Hörer und Sprecher Hörer und Sprecher
General Interface Management Bus	Bus für allgemeine Interface-Arbeitsorganisation	5	IFC (interface clear) ATN (attention) REN (remate enable) EOI (end or IDENTIFY) SRQ (service request)	Interface löschen Achtung (Steuerer sendet Interfacemehrleitungsnachricht oder IDENTIFY) Fernsteuerung erlaubt Ende (sofern ATN H-Pegel hat) oder identifiziere die Bedienanforderung (sofern ATN L-Pegel hat) eines oder mehrere Geräte am System fordern Bedienung an	Steuerer Steuerer Steuerer Steuerer oder Sprecher Steuerer Sprecher

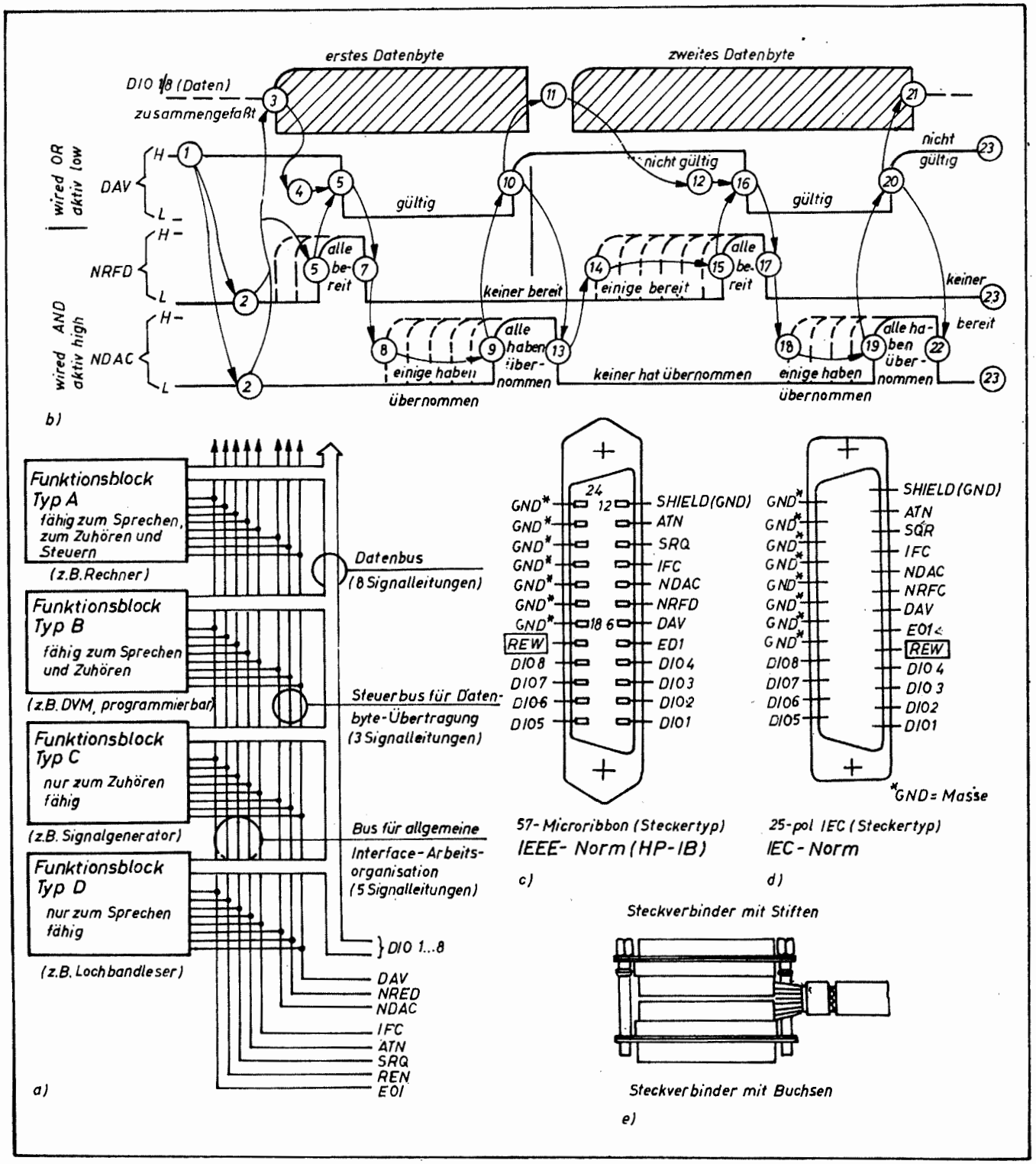


Bild 8.7.3. Zum IEC-Bus

a) Einteilung der Gerätegruppen (Funktionsblöcke) sowie ihr Anschluß an den Bus
b) Beschreibung des Dreileiter-Handshakes
c) Anschlußstecker bei IEEE 488,
d) bei IEC
e) Huckepackstecker des IEC-Kabels

Es existieren *drei Busleitungen*, deren Eigenschaften und Bedeutungen Tab. 8.7.2 zeigt. Gemäß Bild 8.7.3a) existieren *drei Gerätetypen* (Funktionsblöcke):

A. Steuerer, Controller. Er kann Steuersignale abgeben, Daten senden und empfangen.
B. Talker/Listener (Sprecher/Hörer). Er kann Daten senden und empfangen.

C. Listener (Hörer). Er kann nur Daten empfangen.
D. Talker (Sprecher). Er kann nur Daten senden.

Durch den dominaten Controller ist zu sichern, daß zu jeder Zeit nur ein Gerät Daten, Adressen oder Befehle sendet. Die Unterscheidung, ob *Daten oder Adressen* über den DIO-Bus übertragen werden, wird durch das Control-Signal

ATN, Attention, Achtung

festgelegt. Beim Pegel H erfolgt Datenübertragung. Die Adressen und Befehle beim L sind bestimmten Kodegruppen zugeteilt. Dies zeigt Tab. 8.7.3. Dabei fällt auf, daß nicht alle Adressen genutzt werden. Die Ursache liegt darin, daß dann die Bustreiberstufen leistungsfähiger ausgelegt sein müßten. Es werden für die DIO-Leitungen Tristate- oder Open-Collector-Stufen empfohlen. In Zusammenhang mit Bild 8.5.24 ist der Zweileiter-Handshake behandelt. Entscheidend sind dabei *drei Flanken*:

2 Daten sind gültig
3 Daten sind übernommen
6 Bereitschaft für Übernahme neuer Daten

Tabelle 8.7.3. Auszug aus den im IEC-Bus genormten Adressen

Lfd. Nr.	Kode								Bedeutung
	8	7	6	5	4	3	2	1	
0		H	H	H	H	H	H	H	adressierte Befehle
.		.	.	.	.	.	.	.	
15		H	H	H	L	L	L	L	
16		H	H	L	H	H	H	H	Universalbefehle
.		.	.	.	.	.	.	.	
31		H	H	L	L	L	L	L	
32		H	L	H	H	H	H	H	Empfängeradressen
.		.	.	.	.	.	.	.	Listener-adress-group
62		H	L	L	L	L	L	H	
63		H	L	L	L	L	L	L	Unlisten-adress
64		L	H	H	H	H	H	H	Senderadressen
.		.	.	.	.	.	.	.	Talker-adress-group
94		L	H	L	L	L	L	H	
95		L	H	L	L	L	L	L	Untalk-adress

Alle anderen Flanken sind eine Folge hiervon. Werden jetzt mehrere Geräte zusammengeschaltet, so müssen hierbei auch die Verkopplungen entsprechend den TTL-Gattern mit offenem Kollektor wirksam werden. Bei der Leitung für gültige Daten (DAV) scheint es zunächst keine Probleme zu geben, da hier nur ein Sender wirkt. Dennoch ist zu bedenken, daß andere, jetzt nicht aktive Sender an der Leitung hängen. Sie müssen einen Zustand annehmen, der dem einen Sender das Senden auch ermöglicht. Hier ist also nur ein verdrahtetes ODER zulässig. Der aktive Pegel des Senders muß also *aktiv Low*, d. h. die negative Logik, sein.

Dieselbe Aussage gilt sofort auch für den Datenbus (DIO).

Das Signal Daten übernommen (DAC) darf erst dann auftauschen, wenn alle Empfänger die Daten übernommen haben. Dies ist eine UND-Verknüpfung, folglich müßte hier aktiv High, positive Logik verwendet werden. Auch die Flanke „Bereit-

schaft für neue Daten" darf erst wirksam werden, wenn alle Empfänger die Bereitschaft besitzen. Sie müßte daher im Gegensatz zu Bild 8.5.24 die Richtung High-Low genauso die Flanke „Daten übernommen" besitzen. Dies ist offensichtlich nicht realisierbar. Deshalb ist für sie eine dritte Leitung „ready for data" (RFD) erforderlich. So entsteht das Dreileitungs-Handshake-Verfahren. Da nun aber alle Leitungen den gleichen Pegel haben sollen, wird für das „Daten übernommen" und „Bereitschaft für Daten" das negierte Signal bezeichnet. So enthält der Übergabesteuerbus (Handshakebus, Quittierungsbus) folgende drei Leitungen:

DAV (data valid)	Daten sind gültig
NDAC (not data accepted)	Daten sind nicht übernommen
NRFD (not ready for data)	Nicht bereit zur Übernahme neuer Daten

Unter diesen Bedingungen dürfte die Funktion des Dreileiter-Handshakes gemäß Bild 8.7.3b) unmittelbar einsichtig sein.

Der Schnittstellensteuerbus verfügt über *fünf Leitungen*, mit denen die Gesamtbedingungen des Gerätesystems beeinflußt werden können. Er heißt deswegen auch zuweilen Steuerbus oder im englischen Original general interface management bus.

Die Leitung ATN wurde schon bezüglich ihrer Funktion mit dem Datenbus, nämlich der durch sie erfolgenden Umschaltung auf Adressen behandelt. Sie erhält jedoch mit dem EOI-Signal zusammen eine umfangreichere Bedeutung. Dies ist in Tab. 8.7.4 dargestellt. Während das ATN-Signal also nur vom Steuergerät gegeben werden kann, besteht für das EOI-Signal die Möglichkeit, daß es auch der jeweilige Sender abgibt. Dann bedeutet es im Zusammenhang mit ATN passiv High, das Ende der Datenübertragung.

Das IFC-Signal gibt das Steuergerät bei Inbetriebnahme des Systems und vor jeder neuen Adressierung. Hiermit wird der ganze Bus in die Ruhestellung gebracht, bei der kein Gerät Sender oder Empfänger ist.

Meßgeräte in einem Bussystem sind oft sowohl von der Frontplatte als auch durch Fernsteuerung vom Steuergerät aus bedienbar. Mit dem REN-Signal schaltet das Steuergerät die Frontbedienung aus.

In gewissen Systemen treten unregelmäßig Daten auf. Sie müssen dann möglichst umgehend eingearbeitet werden. Vom Anfall dieser Daten muß das Steuergerät in-

Tabelle 8.7.4. Zusammenwirken von ATN und EOI im Schnittstellensteuerbus

Dabei bedeuten

H hohes, also passives Signal
L tiefes, also aktives Signal

Die am L tiefgestellten Indizes bedeuten

C Controller (Steuergerät)
T Talker (Sender)

ATN	EOI	Bedeutung
H	H	Es werden Daten vom Sender zu den Empfängern übertragen (Normalbetrieb).
H	L_T	Das EOI Signal, in diesem Falle Endsignal, wird vom Sender gegeben und bedeutet, daß die letzten Daten eines Blockes gesendet wurden. Es bedeutet für das Steuergerät, daß es entsprechend seinem Programm einen neuen Sender und neue Empfänger festzulegen hat.
L_C	H	Mit dieser Signalkombination legt das Steuergerät die Adressen und Funktionen der Geräte für die nächste Übertragung fest.
L_C	L_C	Diese Signalkombination erfolgt nach einem SRQ-Ruf. Es wird vom Steuergerät gegeben, damit anschließend jenes Gerät, das die Bedienung wünscht, identifiziert werden kann.

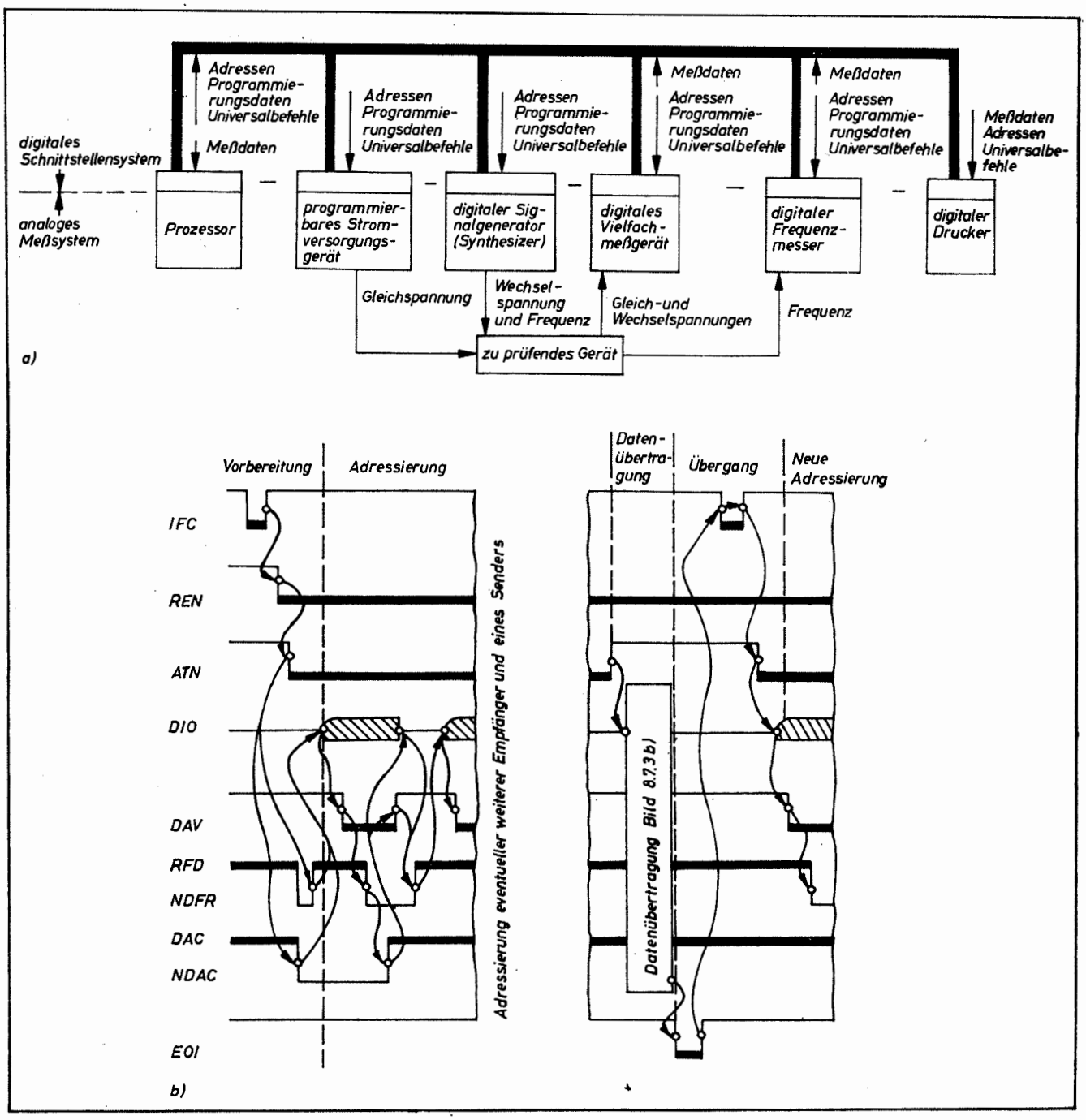


Bild 8.7.4. Beispiel eines automatisierten Meßprozesses

a) Meßgerätezusammenstellung für ein zu prüfendes Gerät
b) Ablauf der Adressierung und des Meßvorganges

formiert werden, um den mehr oder weniger periodischen Betriebsablauf zu unterbrechen. Hierfür steht die SRQ-Leitung zur Verfügung. Das Gerät, wo die Daten anfallen, zieht sie auf aktiv Low. Damit weiß des Steuergerät, daß zufällige Daten anfallen. Es weiß aber noch nicht, wo. Es unterbricht daraufhin zu gegebener Zeit seinen Routinelauf und senkt ATN und EOI auf Low. Dies ist der Hinweis für das Gerät, wo die Daten anfallen, zur Identifikation, und es wird daraufhin zum Sender erklärt.

Im Bild 8.7.4a) ist ein *typisches Beispiel für eine Gerätezusammenstellung* mit Benutzung des IEC-Bus gezeigt. Ein zu prüfendes Gerät wird mit einstellbaren Gleich- und Wechselspannungen mit veränderbarer Frequenz versorgt. Die dabei entstehenden

Spannungen und Frequenzen werden gemessen und ausgedruckt. Der ganze Meßablauf ist im Steuergerät (Processor) als Programm niedergelegt und wird daher von ihm automatisch abgewickelt. Am Ende des Vorganges liegt ein gedrucktes Meßprotokoll vor. Dabei ist vorgesehen, daß auch das Steuergerät Meßdaten empfängt und eventuell danach den Meßablauf verändert.

Einen Ausschnitt aus dem Meßablauf zeigt Bild 8.7.4b). Zu Anfang gibt das Steuergerät das Signal IFC, damit werden alle Geräte in den Ruhezustand versetzt. Dann schaltet es für die Dauer des automatischen Meßvorganges das Signal REN ein. Nun können Geräte adressiert werden. Dazu legt das Steuergerät ATN auf L. Darauf melden sich nach einer gewissen Zeit alle Geräte mit RFD. Außerdem zeigen alle Geräte an, daß sie noch keine Daten übernommen haben, NDAC. Dann kann das Steuergerät eine Empfängeradresse aussenden. Nach der Einschwingzeit meldet es über DAV die Adresse als gültig. Daraufhin zieht der angesprochene Empfänger die NRFD-Leitung auf L. Sobald er die Adresse verarbeitet hat, d. h. mittels eines Flipflops gespeichert hat, daß er Empfänger ist, meldet er DAC. Daraufhin nimmt das Steuergerät das Signal DAV und die Adressen weg. Der Empfänger ist damit festgelegt. Anschließend kann in ähnlicher Weise eventuell noch ein zweiter, dritter Empfänger adressiert werden. Dann erfolgt die Adressierung eines Senders nach demselben Prinzip.

Nachdem so Empfänger und Sender festliegen, nimmt das Steuergerät das ATN-Signal weg, und die Datenübertragung kann beginnen. Sie ist bereits im Bild 8.7.3b) dargestellt. Am Ende der Übertragung zieht der Sender das EOI-Signal auf Null. Damit weiß das Steuergerät, daß es den nächsten Zyklus beginnen muß. Mit IFC oder anderen Signalen hebt es die alte Adressierung auf und beginnt mit ATN die neue Adressierung.

Die Beschreibung des IEC-Busses ist hier nur in den wesentlichen Teilen erfolgt. Im Laufe der Entwicklung sind darüber hinaus verschiedene Änderungen wirksam geworden. So hat es sich bei vielen Meßgeräten eingebürgert, mit extended adress zu arbeiten. Hierbei werden zwei Bytes als Adresse verwendet. Dadurch sind statt 31 Sendern bzw. Empfängern 961 zu adressieren.

Die breite Anwendung des IEC-Busses auf allen Gebieten ließ insbesondere für Nichttechniker die Programmierung zum Problem werden. Dies führt zu der Tendenz, statt mit Bytes mit Mnemonics zu arbeiten. Diese Entwicklung erreichte um 1981 einen Stand, der eine erweiterte Normung für den IEC-Bus erforderlich machte. So wird zur Zeit der Standard IEEE 728 als übergeordneter Komplex ausgearbeitet.

8.7.4.2. SI 2.2; SIAL

Im RGW gibt es für Meß-, Steuer- und Regelungstechnik seit etwa 1979 das Linieninterface SIAL. Bis auf Festlegungen zu den Steckern, Steckerbelegungen und Kabeln ist es mit dem in der DDR standardisierten System SI 2.2 gleich. Es ist u. a. für die Klein- bzw. Prozeßrechner KRS 4200, PRS 4000 und Geräte des Systems Ursadat 4000 bzw. 4010 genutzt. Den Aufbau des Bussystems zeigt Bild 8.7.5. Der Bus A führt 12 Adreßleitungen für 4096 adressierbare Kanäle und 4 Leitungen für Programminformationen. Die Busse DA und DE sind für 16 bit breite Datenaus- und -eingabe reserviert. Der Handshake wird ähnlich wie beim Interface SI 1.2 mit den Leitungen B2, M1 und B1 realisiert. Die Ausgangsdaten DA dürfen sich nur bei B2 = Ø und die Eingangsdaten DE bei M1 = Ø ändern. Für die Unterbrechungen (Interrupt) existieren 16 Leitungen. Für die Leiterplatten wird die Fläche 90 mal 170 mm² mit 58poligem Stecker genutzt.

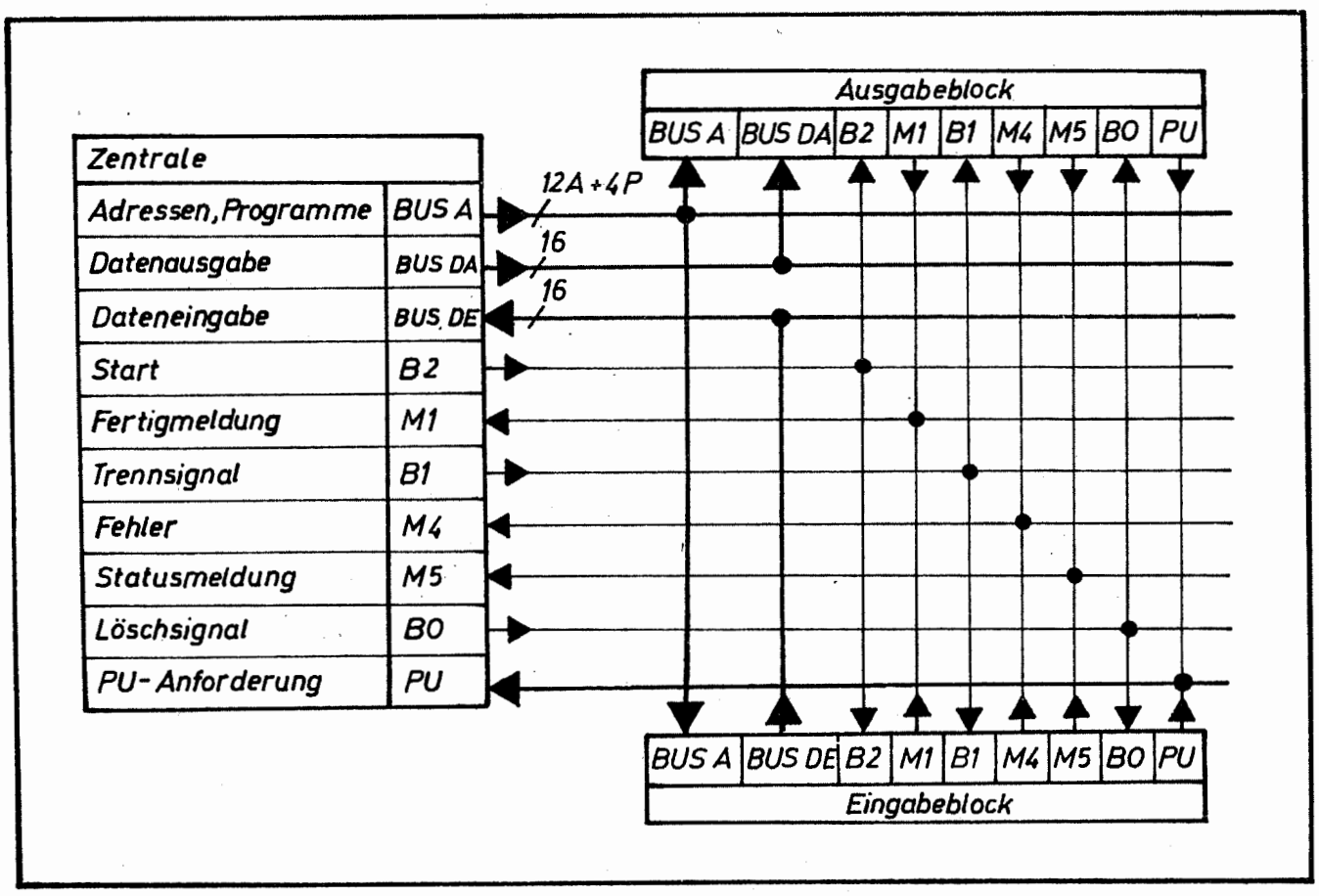


Bild 8.7.5. Signale und Zusammenschaltung beim SI 2.2

8.7.4.3. CAMAC

Die Arbeiten zum CAMAC-System begannen Ende der fünfziger Jahre im ESONE-Komitee (European Standard Organization of Nuclear Electronics). Im Herbst 1968 lag der erste Entwurf vor. Ursprünglich war CAMAC ein Kunstwort, das durch seine Symmetrie das einheitliche Bindeglied zwischen Prozeßrechner und Peripherie symbolisieren sollte. Später, als es über die Kerntechnik hinaus Bedeutung erlangte, wurde es als computer application to measurement and control interpretiert. Anstelle von application wird auch automated bzw. aided verwendet.

Das CAMAC-System ist in erster Linie auf Einschübe gut geschützter Leiterplatten ausgerichtet. Jede einzelne Leiterplatte mit den Abmessungen 222 × 17,2 × 306 mm³ stellt einen funktionsfähigen Modul dar. Derartige Module werden hierarchisch zusammengefaßt, einmal bis zu 25 horizontal nebeneinander als Crate und dann bis zu 7 solcher Crates übereinander. Jede horizontale Zusammenfassung enthält ein ihr zugeordnetes Steuergerät, den Crate-Controler CC. Sie werden vertikal wiederum durch ein zentrales Steuergerät, den System-Controler SC vereinigt (Bild 8.7.6a)). Bei der Steuerung dieses Systems treten ähnliche Verhältnisse wie beim IEC-Bus auf. Auch hier gilt der TTL-Pegel. Es wird allerdings nur der Zweileiter-Handshake mit einigen Ergänzungen angewendet. Das Leitungssystem ist umfangreicher als beim IEC-Bus. Jeder Modul enthält einen 86-poligen Steckverbinder. Die jeweils rechts liegende Rahmensteuerung besitzt sogar einen 132poligen Steckverbinder. Die entsprechende Verschaltung deutet Bild 8.7.6b) an. Über die N-Leitungen werden die Module adressiert. Dabei dienen zur weiteren Adreßaufschlüsselung die A-Leitungen. Mit den F-Leitungen werden bestimmte Funktionen auf den Modulen gesteuert. Als Datenbus stehen je 24 Schreib- und Leseleitungen zur Verfügung. Die Signale S_1 und S_2 wirken ähnlich wie DAV beim IEC-Bus. Die L-Leitungen entsprechen SRQ. Die Leitungen

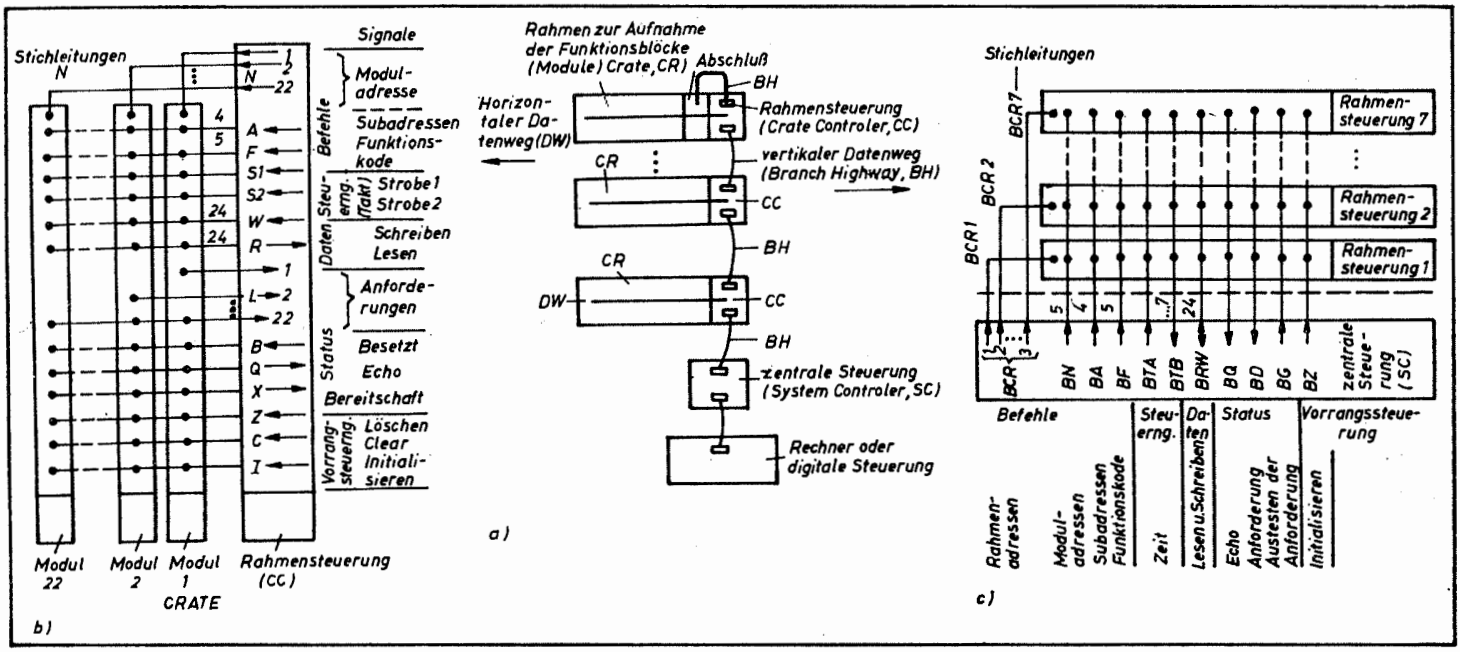


Bild 8.7.6. CAMAC-Interface

a) Im horizontalen Datenweg sind bis zu 25 Module zusammengeschaltet, und durch Rahmensteuerung wird ihr Betrieb organisiert. Wiederum 7 solcher Rahmen (Crates) sind dann vertikal zusammengefaßt. Sie werden in ihrem Ablauf durch die zentrale Steuerung eventuell unter Zuhilfenahme eines Rechners gesteuert

b) Leitungsbelegung beim horizontalen Datenweg

c) Leitungsbelegung beim vertikalen Datenweg

BGX dienen dem Handshake-System. Die letzten drei Leitungen sind genereller Natur. Z entspricht etwa IFC. Mit C werden bestimmte Register gelöscht, und I ist ein gezieltes Löschen, z. B. für speziell adressierte Register.

Der vertikale Datenweg heißt Branch highway BH (Teilbild a)). Bei ihm bewirkt die zentrale Steuerung SC, eventuell unterstützt durch einen zusätzlichen Rechner, den Ablauf. Hierzu erfolgt die Verkopplung entsprechend Bild 8.7.6c). Alle Leitungen sind also durch den Zusatz B von branch gekennzeichnet. Die Leitungen BTA und BTB stellen einen Zweileitungs-Handshake dar. Im Gegensatz zum horizontalen Teilsystem enthält das vertikale nur einen für Schreiben und Lesen gemeinsamen Datenbus.

Ähnlich wie jetzt beim IEC-Bus wurde für CAMAC schon früher eine spezielle Programmiersprache entwickelt. Sie ist ähnlich wie BASIC mit Zeilen-Nr. organisiert (vgl. 8.8.5.). Eine spezielle CAMAC-Sprache entstand 1972.

8.7.5. Serielle Systeme

Die Beschreibung der parallelen Busse läßt zwei Fakten erkennen: Sie benötigen viele Leitungen und sind nur für relativ kurze Leitungslängen bis ca. 50 m sinnvoll anwendbar. In diesem Bereich sind sie aber zugleich die schnellstmöglichen Systeme. Es gibt andererseits aber auch eine Vielzahl langsamer Geräte, wie z. B. den Fernschreiber. In diesen und ähnlichen Fällen ist es sinnvoll, serielle Busse zu verwenden. Wesentliche Grundlagen hierzu wurden bereits im Zusammenhang mit der SIO U 856 im Abschnitt 8.6.5.2. behandelt. Die wichtigsten Signale dazu zeigt Bild 8.6.30c) für asynchronen und d) für synchronen Betrieb. Zur Übertragung kommen u. a. folgende Netze in Betracht: Telexnetz mit etwa 50 bit/s; Datennetz von etwa 200 bis 9600 bit/s; Fernsprechnetz 300 bis 1200 bit/s und Sonderleitungen, z. B. Lichtleiter, Koaxkabel usw. bis zu mehreren Mbit/s.

Tabelle 8.7.5. Leitungen der Schnittstelle V. 24 sowie anderer äquivalenter Normen. Es bedeuten

CCITT V24 Comité Consultatif International Télégraphique et Téléfonique
EIA RS232C Electronic Industries Association
ISO 2110 International Standard Organization
DIN 66020 (West-)Deutsches Institut für Normung

DB25-Stift Vgl. Bild 8.7.3	Signal-richtung	Kode			Techn. Abkür-zung	Bezeichnung	
		DIN	V24	RS232		deutsch	englisch
1		E1	101	AA	GND	Schutzerde	protective ground; earth ground
2	←	D1	103	BA	XMIT	Sendedaten	transmitted data
3	→	D2	104	BB	RCVR	Empfangs-daten	received data
4	←	S2	105	CA	RTS	Sendeteil einschalten	request to send
		M2	106	CB	CTS	Sende-bereitschaft	ready for sending clear to send
6	→	M1	107	CC	DSR	Betriebs-bereitschaft	data set ready
7	→	E2	102	AB	GND	Betriebserde	signal ground; common return; logic ground
8	→	M5	109	CF	DCD	Empfangs-signalpegel	data channel received line signal detector; data carrier detector
11	←	S5	126	—		hohe Sende-frequenzlage einschalten	select transmit frequency
12	→	HM5	122	—	SCF	Hilfskanal-empfangs-signalpegel	backward channel received line signal detector; supervisory (secondary) channel carrier detector
13	→	HM2	121	—		Hilfskanal-sendebereit-schaft	backward channel ready; supervisory (secondary channel ready
14	←	HD1	118	—		Hilfskanal-sendedaten	transmitted back ward channel data; transmitted supervisory (secondary) channel data

Tabelle 8.7.5. (Fortsetzung)

15	→	T2	114	DB	—	Sendeschritttakt von der DÜE	transmitted signal element timing; transmitted clock internal
16	→	HD2	119	—		Hilfskanalempfangsdaten	received backward channel data; received supervisory (secondary) channel data
17	→	T4	115	DD	—	Empfangsschrittakt	received signal element timing; received bit clock
19	←	HS2	120	—	—	Hilfskanalsendeteil einschalten	transmit backward channel data; transmit supervisory (secondary) channel data
20	←	S1.1	108/1	—	—	Übertragungsleitungen einschalten	connect data set to line
		oder					
	←	S1.2	108/2	CD	DTR	Datenendeinrichtung betriebsbereit	data terminal ready
22	→	M3	125	CE	—	ankommender Ruf	calling indicator; ring indicator
23	←	S4	111	CH	—	hohe Übertragungsgeschwindigkeit einschalten	data signalling rate selector
24	←	T1	113	DA	—	Sendeschritttakt zur DÜE	transmitter signal element timing; transmitted bit clock, external

8.7.5.1. V. 24 und RS 232

Bei den seriellen Bussystemen V. 24 bzw. RS 232 usw. sind zwei Schnittstellen zu unterscheiden (vgl. Bild 8.7.7a)). Dies sind die beiden Seiten des Modems (von **Mo**dulator — **Dem**odulator), auch als DÜE = Datenübertragungseinrichtung bezeichnet,

- zum Datenendgerät (DEE) als dem Rechner bzw. dem Peripheriegerät,
- zur Leitung, die in der Regel zwei Leitungen (evtl. auch vier) besitzt.

Die definierte Schnittstelle bezieht sich auf die Stelle DEE ↔ DÜE. Hierfür gelten die Daten von Tab. 8.7.5. Die Anschlüsse sind meist an den 25poligen Stecker gemäß

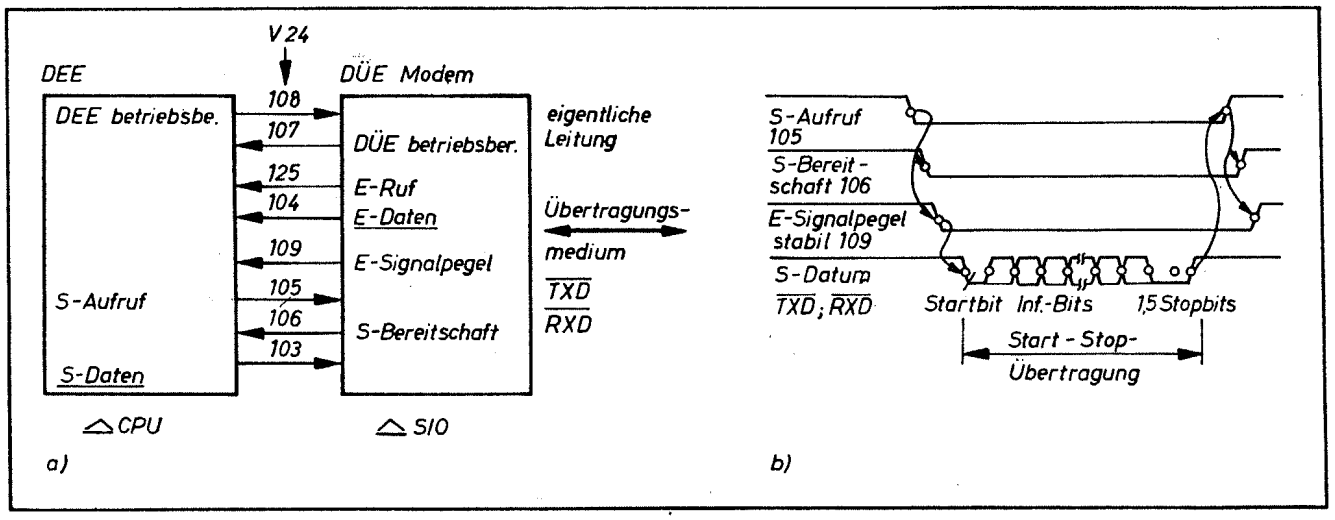


Bild 8.7.7. Schnittstelle V. 24 zwischen Datenendeinrichtung DEE und Datenübertragungseinrichtung DÜE (Modem) und Übertragungsweg (a) sowie die zugehörigen Signalverläufe

Bild 8.7.3 gebunden. Daher auch die Bezeichnung V. 24, weil sie sich ursprünglich auf 24 Kontakte bezog.

Für die Pegel auf der Leitung ist anderseits die negative Logik festgelegt:

Ø H +(3 ... 25) V
1 L −(3 ... 25) V

Der Innenwiderstand soll größer als 300 Ω sein, damit kein Kurzschlußstrom größer als 0,5 A auftreten kann. Die anzuschließenden Impedanzen sollen im Bereich 3 bis 7 kΩ liegen, die Slew-Rate soll kleiner als 30 V/µs sein. Die angegebene Bitrate in bit/s oder, fast immer gleichwertig, in Bd (Baud) bezieht sich immer auf den Übertragungsweg. Für V.24 bzw. RS 232 gelten maximal 19 Kbit/s. Die Pegel auf dem Übertragungsweg bzw. die Bd-Werte weichen bei ungünstigen Normen von den oben genannten Werten ab, z.B. RS 232 C: ± (3 ... 25) V, RS 232 B: ± (3 ... 15) V.

Nicht genormt und dennoch oft angewendet wird die sogenannte Stromschleife (current-loop-interface). Sie arbeitet mit den Stromwerten 1 = 10 ... 100 mA; Ø = 0 ... 0,1 mA. Meist werden die 1-Werte auf 20 mA festgelegt.

Ein Fernschreiber (TTY = Teletype) arbeitet bis zu etwa 300 bit/s. Dabei wird der Kode entsprechend Tab. 8.7.6 verwendet.

In den USA existiert noch die Norm RS 422 bzw. 423. Sie ist der Norm RS 232 sehr ähnlich und weicht mit folgenden Daten ab:

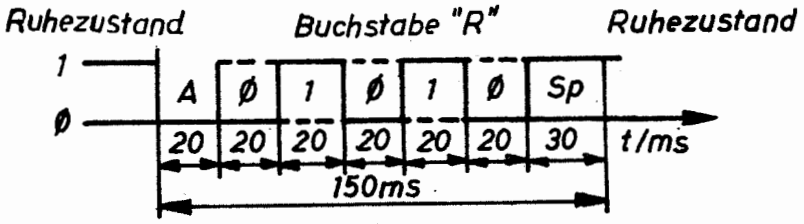
RS 422 ± (2 ... 6) V 10 Mbit/s
RS 423 ± (4 ... 6) V 100 Kbit/s

Schließlich existiert noch das RS 449, welches gegenüber dem RS 232 zusätzliche Kontrollsignale verwendet.

8.7.5.2. PDV-Bus

Der PDV-Bus wurde international für die Prozeßdatenverarbeitung entwickelt. Er erlaubt Werte von 120 Kbit/s bis zu 3 km Leitungslänge und Reaktionszeiten um 10 ms. Von der Sendestation, die den Ablauf steuert, aber auch durch alarmberechtigte Geräte unterbrochen werden kann, gehen die Signale gemäß Bild 8.7.8a) auf der Nachrichtenleitung aus, während die angeschlossenen Geräte auf der Antwortleitung Informationen zurücksenden. Dabei entsteht ein Handshake auf höherem Niveau, wie

Tabelle 8.7.6. Fernschreibkode CCITT Nr. 2



A Anlaufschritt, Sp Sperrschritt

Übertragungsgeschwindigkeit: 50 Bd = 6,67 Zeichen/s = 150 ms/Zeichen

				Hex mit Start- und Stopbit	
CCITT Nr.	Kode	Buchst.	Buchst.	Ziff./Zeich.	Linksbündig invertiert
1	11000	A	—	62	9C
2	10011	B	?	4E	BØ
3	Ø1110	C	:	3A	C4
4	10010	D	✢	4A	B4
5	10000	E	3	42	BC
6	10110	F	☐	5A	A4
7	Ø1011	G	☐	2E	DØ
8	ØØ1Ø1	H	☐	16	E8
9	Ø1100	I	8	32	CC
10	11010	J	Bell	6A	94
11	11110	K	(	7A	84
12	Ø1ØØ1	L	)	26	D8
13	ØØ111	M	.	1E	EØ
14	ØØ110	N	,	1A	E4
15	ØØØ11	O	9	ØE	FØ
16	Ø1101	P	Ø	36	C8
17	11101	Q	1	76	88
18	Ø1Ø10	R	4	2A	D4
19	10100	S	’	52	AC
20	ØØØØ1	T	5	Ø6	F8
21	11100	U	7	72	8C
22	Ø1111	V	=	3E	CØ
23	11ØØ1	W	2	66	98
24	10111	X	/	5E	AØ
25	10101	Y	6	56	A8
26	10001	Z	+	46	B8
27	ØØØ1Ø	↑ <CR	↑	ØA	F4
28	Ø1ØØØ	≡ LF		22	DC
29	11111	<A ...		7E	8Ø
30	11Ø11	1 ...>		6E	9Ø
31	ØØ100	Zwischenraum		12	EC
32	ØØØØØ	☐	☐	Ø2	FC

✢ = wer da? ☐ frei Beispiel: „R“ ≙ = EQR

0010101	Ø		1101010	Ø
2	A		D	4

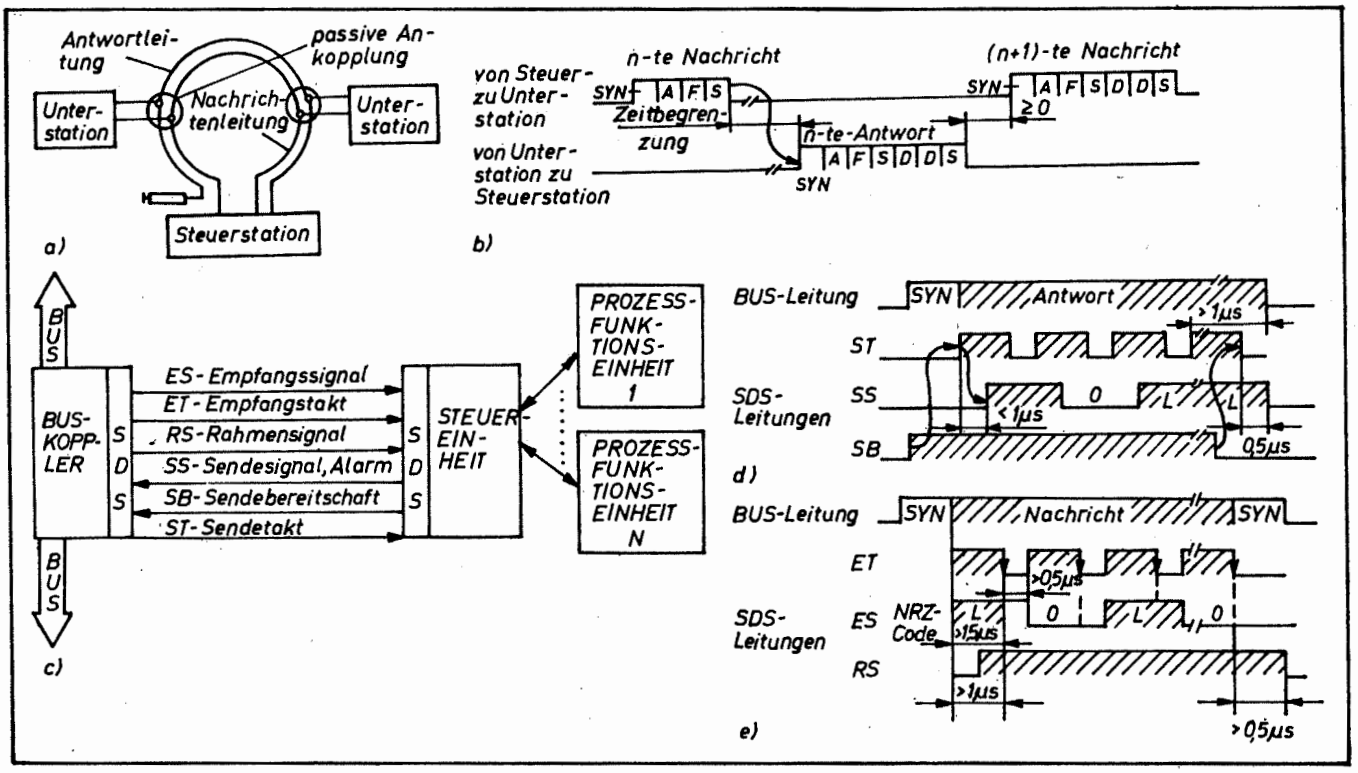


Bild 8.7.8. Der PDV-Bus

a) Doppelringleitung des Busses
b) Signale auf der Doppelringleitung
c) Prinzipieller Aufbau einer Unterstation mit den SDS-Leitungen
d) SDS-Verlauf beim Senden,
e) beim Empfangen

es Teilbild b) zeigt. Die einzelnen Anteile bedeuten dabei

SYN	Synchronisationszeichen
A	Adreßteil (8 bit) = 256 Geräte
F	Funktionsteil (8 bit)
S	Sicherungsteil (8 bit CRC)
DD	Datenteil (16 bit).

Die Buskopplung erfolgt wiederum über ein Modem, hier Buskoppler genannt. Er erzeugt die Signale der seriellen digitalen Schnittstelle SDS (Bild 8.7.8c)). Die beim Senden und Empfangen dort auftretenden Signale zeigen die Teilbilder d) und e). Diese Signale entsprechen der Norm RS 422 (s. 8.7.5.1., RS 232). Die Sendesignale SS besitzen das NRZ-Format, deshalb ist zusätzlich der Sendetakt ST notwendig. Dasselbe gilt für den Empfänger ES und ST. Die Signale RS und SB sind die zugehörigen eigentlichen Handshake-Signale.

8.7.5.3. HP-IL-Bus

Für kleine Systeme hat Hewlett Packard einen sehr leistungsarmen seriellen Bus entwickelt, der auch von Taschenrechnern und anderen batteriebetriebenen Geräten gesteuert werden kann. Die Bezeichnung IL bedeutet interface loop, also Interface-Schleife. Er verwendet zwei Leitungen, die vom Steuergerät ausgehen (Bild 8.7.9a)), beim ersten Gerät einmünden, es durchlaufen und danach weiter zum nächsten Gerät gehen usw., um so schließlich wieder beim Steuergerät anzukommen. Hierdurch wird ein spezieller Handshake erreicht: Das Steuergerät wartet so lange, bis die Antwortsignale bei ihm ankommen. Für die Signale werden drei Pegel, −1,5 V; 0 V und

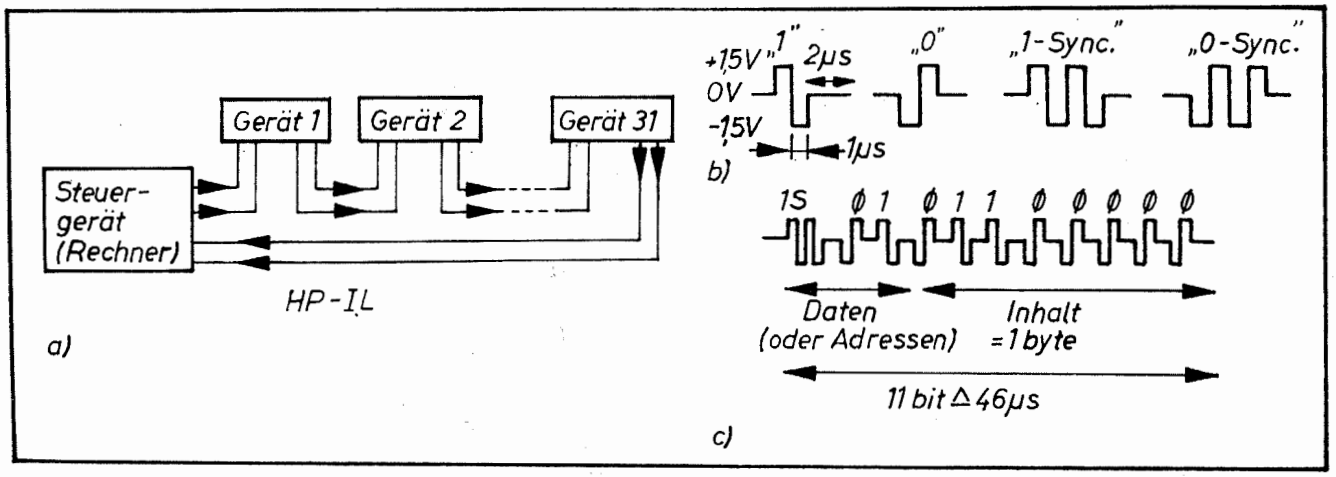


Bild 8.7.9. Zum HP-IL-Bus

a) Anschaltung der Geräte an das Steuergerät über Doppelleitungen in Ringform
b) Die vier Signale mit drei Pegeln −1,5; 0 und +1,5 V
c) Aufbau eines Rahmens (frame) aus 11 bit, wovon die ersten drei festlegen, ob Daten oder Adressen gesendet werden bzw. ob das Ende eines Datenblocks vorliegt (Bit 2) oder Service gewünscht wird (Bit 3)

+1,5 V, verwendet. In Ruhe fließt also kein Strom. Die Impedanz der Doppelleitung beträgt 100 Ω. Die positiven und negativen Pegel werden für je 1 µs gesendet, und zwar so, daß kein Gleichstromanteil existiert. Dadurch können die Signale auch über Transformatoren weitergegeben werden. Die Geräte können so gleichstrommäßig entkoppelt werden. Bei den Signalen sind vier Typen zu unterscheiden (Bild 8.7.9 b)): Je zwei Signale, die mit einem positiven Impuls beginnen, und zwei normale Signale für Bits bzw. zur Synchronisation. Jedes dieser vier Signale wird mit einer Nullinie von 2 µs abgeschlossen. Jedes zu übertragende Byte besteht aus 11 bit und hat eine Länge von 46 µs. Das erste Bit ist ein Synchronisationsbit. 1 bedeutet hier Daten, Ø Adressen. Das zweite Bit wird im letzten Byte einer Daten- bzw. Adressenfolge gesetzt. Das dritte Bit ist für Interrupt vorgesehen. Die weiteren 8 bit sind das Byte. Für Adressen werden fünf davon verwendet, so daß maximal 31 Geräte adressierbar sind. Nur das adressierte Gerät übernimmt die Information. Nichtadressierte Geräte geben sie einfach zum Ausgang weiter. Nach diesem Prinzip sind theoretisch ca. 20 Kbyte/s zu übertragen. Infolge von Verzögerungen in den Geräten werden meist nur 3 bis 5 Kbyte/s realisiert.

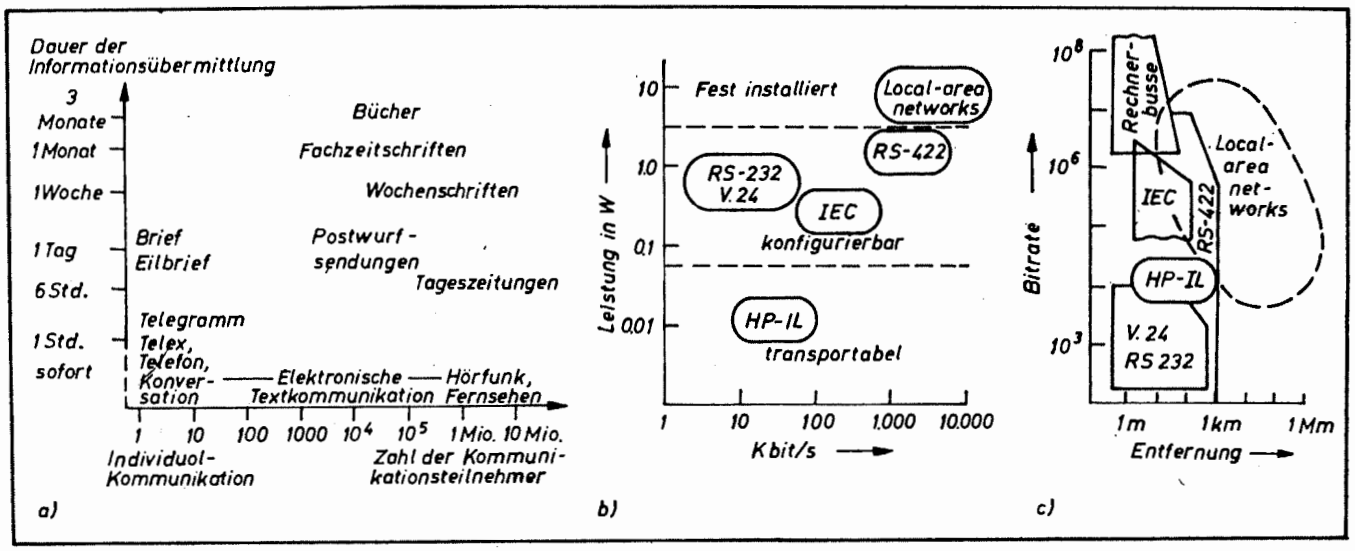


Bild 8.7.10. Zum Vergleich von Bussystemen

a) Kommunikationsformen
b) Leistungs-Datenraten-Diagramm
c) Datenrate-Entfernungs-Diagramm

8.7.6. Vergleich von Bussen

Bussysteme haben die Aufgabe, Daten zwischen Kommunikationspartnern zu vermitteln. Dies führt zu einem Vergleich mit Systemen, die nicht zu den Bussen gehören. Diesen Vergleich zeigt Bild 8.7.10a). Die Eigenschaften der Bussysteme wurden im Abschnitt 8.7.1. behandelt. Hier sind ergänzend die Zusammenhänge bezüglich der notwendigen Leitung und der Übertragungsrate (Teilbild b)) sowie Übertragungsrate und Entfernung (Teilbild c)) zusammengestellt. Dabei fällt deutlich auf, daß die verschiedenen Busse jeweils speziellen Bedingungen optimal genügen.

8.8. Software

Die Definition und wesentliche Eigenschaften der Software wurden bereits im Abschnitt 8.6.1. behandelt. Im strengen Sinne gehört die Software nicht zur Elektronik, aber wie in den vorangegangenen Abschnitten immer wieder gezeigt wurde, benötigen alle steuerbaren Hardwarestrukturen eine Software zu ihrem Betrieb. Besonders wichtig ist die Software bei den Mikroprozessoren und den Mikrorechnern. Dies ist der Grund, weshalb im folgenden einige Grundlagen der Softwareherstellung behandelt werden sollen. Darüber hinaus ist es bedeutsam, daß gerade aufgrund der Existenz des Mikroprozessors es erstmalig einem breiten Kreis deutlich wurde, wie Soft- und Hardware zusammenhängen und gegeneinander ausgetauscht werden können. Der vorher häufig vorhandenen Mystifizierung der Software ist dadurch effektiv zu begegnen.

8.8.1. Allgemeine Grundlagen der Programmierung

8.8.1.1. Effizienz und Top-down

Die Programmierung hat die Aufgabe, eine Kodefolge von Befehlen so zu schreiben, daß sie in dem Programmspeicher abgelegt den Rechner in die Lage versetzt, die zu bearbeitende Aufgabe richtig durchzuführen. In dieser Aussage ist bereits enthalten, daß es für eine Aufgabe eine Vielzahl von Kodefolgen geben kann, die dies ermöglichen. Es gibt z. Z. keine Aussage dazu, wie die bestmögliche Kodefolge aussieht. Dies ist mit den jetzt bekannten theoretischen Grundlagen nicht entscheidbar. In der Praxis gibt es immer *mehrere Lösungen* (Programme), *die das Problem auf unterschiedliche Weise realisieren*. Sie können sich selbst bei einem Rechnertyp u. a. unterscheiden bezüglich

- Länge der Kodefolge = Speicherplatz für das Programm,
- Zeitdauer, in der das Programm abgearbeitet ist,
- Genauigkeit und Zuverlässigkeit (Fehlersicherheit) der Lösung,
- Umfang des benötigten Speichers bei der Bearbeitung.

Solche Probleme zählen zur Effizienz des Programms und können bisher nur im Vergleich zwischen den realisierten Kodes bewertet werden. Im Gegensatz hierzu stehen *zwei Methoden*, mit denen der Kode erarbeitet werden kann:

- Top-down
- Bottom-up.

In der Anfangszeit der Rechentechnik war es üblich, ein Problem wie folgt zu bearbeiten: Intuitiv wurden Kodes für vermeintliche Teilaufgaben geschrieben und erprobt, und dann wurden diese Teilprogramme zum Gesamtprogramm zusammengefaßt. Diese von unten (Maschinenkode) nach oben zum Gesamtprogramm führende Methode heißt folgerichtig *Bottom-up*. Sie bringt meist vielfältige Schwierigkeiten

beim Zusammenfügen der Teilprogramme mit sich, denn in der Regel sind hierbei noch Änderungen in den Teilprogrammen notwendig. Deshalb gilt diese Methode heute als ungeeignet für größere Programme.

Bei der *Top-down*-Methode wird genau umgekehrt vorgegangen. Das Gesamtproblem wird in Teilprobleme zerlegt, und dabei werden die Schnittstellen zwischen den Teilproblemen und zum Gesamtproblem hin genau definiert. Dieser Prozeß wird solange fortgesetzt, bis überschaubare Teile entstanden sind. Erst dann wird mit der Kodierung dieser Teile begonnen, und nun lassen sich in der Regel die Teilprogramme leicht zum Gesamtprogramm zusammenfassen. Beim Top-down wird also erst nach völliger Klärung und Zerlegung der Gesamtaufgabe in überschaubare Teilprogramme mit der Kodierung begonnen. Beim Bottom-up steht die Kodierung von Anfang an im Vordergrund. Das Top-down gilt heute als der einzige effektive Weg für die Gewinnung effizienter Programme.

8.8.1.2. Strukturierte Programmierung und deren Grundelemente

Die Top-down-Methode schafft vor allem Übersichtlichkeit bei der Programmierung. Die einzelnen Programmteile sind notwendig weitgehend entkoppelt. Dadurch treten zuweilen gleiche Programmabschnitte in den Teilprogrammen auf. Die Methode führt aber nicht immer zum „effektivsten" Kode. Da dieser bisher ohnehin nicht theoretisch berechnet werden kann, ist die Übersichtlichkeit ein besonders wichtiges Kriterium geworden. Sie hat außerdem noch den Vorteil, daß Änderungen am Programm infolge von Änderungen der Aufgabenstellung im Laufe der Zeit (bei Benutzung des Programms) leichter und damit wieder effektiver durchzuführen sind. Diese grundsätzlichen Überlegungen führten zur strukturierten Programmierung, die im wesentlichen durch *drei Prinzipien* zu kennzeichnen ist:

- Top-down-Entwurf
- beschränkter Vorrat an Strukturelementen
- Strukturierung der Daten.

Die *Strukturierung der Daten* wird im Abschnitt 8.8.1.5. behandelt.

Der *beschränkte Vorrat* an Strukturen zur Steuerung der Programme ist heute weitgehend auch theoretisch abgeklärt. Es werden *drei Grundelemente* unterschieden:

Folge, Verzweigung und Schleife.

Sie sind in Bild 8.8.1 sowohl mit den anderen gebräuchlichen Namen als auch in verschiedenen Darstellungen gezeigt und werden im Abschnitt 8.8.1.3. und 8.8.1.4. genauer behandelt. Sie lassen sich wiederum aus *zwei unterschiedlichen Elementen* aufbauen:

- Verarbeitungsschritt (A1 bis A6 in Bild 8.8.1)
- Entscheidung (c in Bild 8.8.1).

Der *Verarbeitungsschritt* kann im einfachsten Fall ein Befehl sein, der aber keine Bedingung enthält, also z. B. Addieren oder Transportieren.

Die *Entscheidung* kann als Schalter interpretiert werden, der je nach irgendwelchen Bedingungen in zwei verschiedene Stellungen gestellt werden kann. Das Hardware-Analogon ist der Multiplexer. Die Bedingungen können z. B. durch ein Flag (gesetzt — nicht gesetzt), durch logische Werte Ø, 1 oder durch Vergleiche realisiert werden.

Der einfachste Verarbeitungsschritt bzw. die Entscheidung knüpfen also unmittelbar an die Hardware an. Dementsprechend werden im unteren Teil von Bild 8.8.1 auch gleich Bezug zu dem Programmspeicher und den in ihm realisierten Möglich-

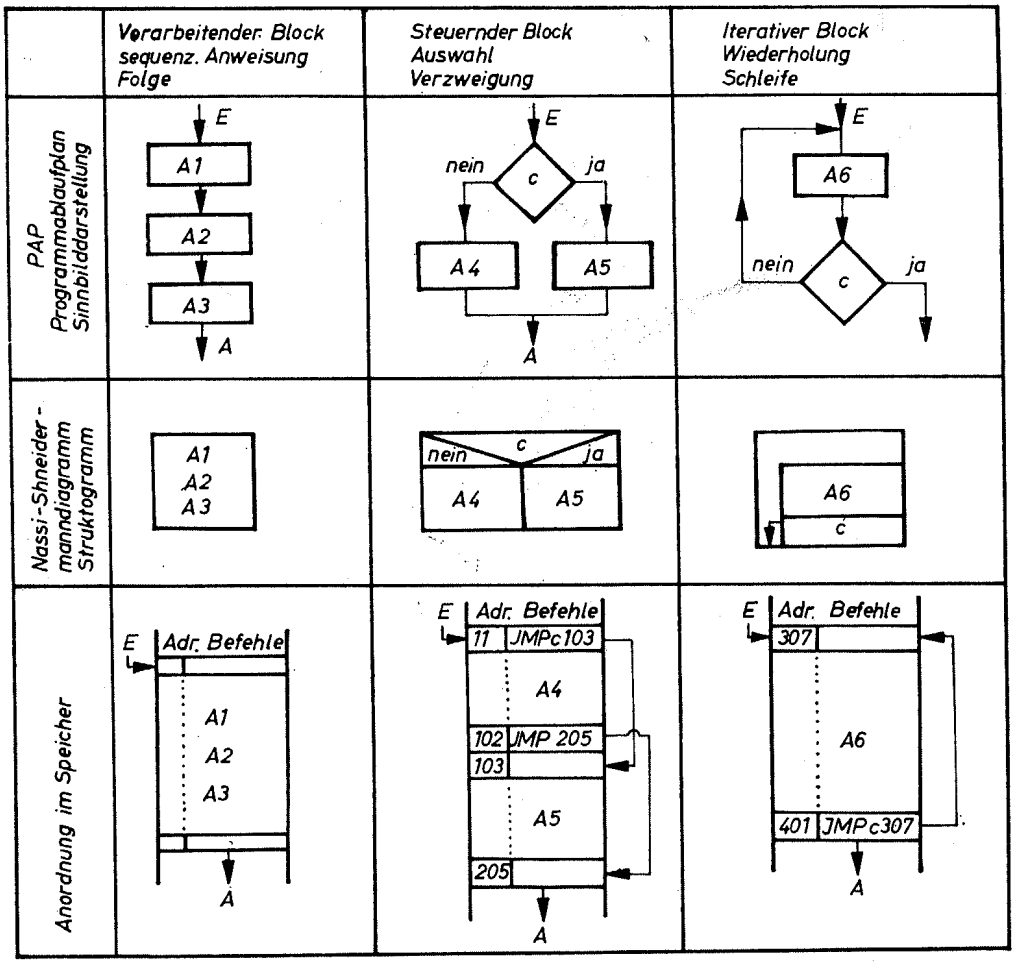


Bild 8.8.1. Grundstrukturen der Programmierung in verschiedenen Darstellungen

A1 bis A6 sind einzelne Verarbeitungsschritte, Operationen, z.B. Befehle oder eine Zusammensetzung aus ihnen. c sind Entscheidungen, z.B. mittels Flag

keiten der drei Grundstrukturen genommen. Für die mehr softwarebezogene Darstellung gibt es heute noch die beiden anderen *Darstellungsarten*:

PAP = Programmablaufplan
Struktogramm.

Sie haben beide ihre Vor- und Nachteile. Wenngleich aus theoretischen Gründen heute das Struktogramm, welches auf NASSI und SHNEIDERMAN zurückgeht, bevorzugt werden sollte, so ist oft der ältere PAP, auch Sinnbilddarstellung oder Flußbilddarstellung genannt, im Vorteil, insbesondere was die Übersichtlichkeit angeht. Das Struktogramm kommt dagegen den Grundlagen der strukturierten Programmierung mehr entgegen, da es deutlicher den noch zu behandelnden Ein- und Ausgang eines Blocks einer Prozedur betont und außerdem sich auch selbst mit einem Rechner generieren läßt.

Von einem Block, einer Prozedur bzw. einem Teilprogramm wird immer dann gesprochen, wenn es aus den drei Grundelementen von Bild 8.8.1 so zusammengesetzt wird, daß nur ein einziger Eingang und ein einziger Ausgang besteht. Im PAP ist dies durch den Anfangs- und Endpfeil ausgedrückt, im Struktogramm durch das rechteckige Kästchen.

Ein Block, der nur aus linear angeordneten einzelnen Anweisungen besteht, bereitet keine Probleme. Sind in ihm aber Verzweigungen bzw. Schleifen enthalten, so müssen gewisse Randbedingungen eingehalten werden. Deshalb sollen diese Teile weiter detailliert behandelt werden.

8.8.1.3. Verzweigungen

Bild 8.8.2 zeigt verschiedene Arten von Verzweigungen, dabei stehen jeweils untereinander PAP, Struktogramm und Ausdrücke aus höheren Sprachen (vgl. 8.8.5.).

Die linke Struktur ist die Wiederholung aus Bild 8.8.1. Bei der mittleren Struktur entfällt in einem Abschnitt die weitere Verarbeitung. Sie erscheint auf den ersten Blick einfacher, ist aber oft bei der Programmierung schwieriger zu handhaben.

Eine besondere, verallgemeinerte Verzweigung zeigt die rechte Seite als Fallabfrage. Sie läßt sich aus der einfachen Bedingung durch eine Hierarchie von Bedingungen ableiten. An dem Nein-Ausgang einer Raute (Bedingungen) wird wiederum eine Entscheidung (Raute) angeschlossen. Dies wird bereits durch das Struktogramm angedeutet. Der letzte Fall wird bei dieser Fallabfrage oft als Fehlerzweig verwendet.

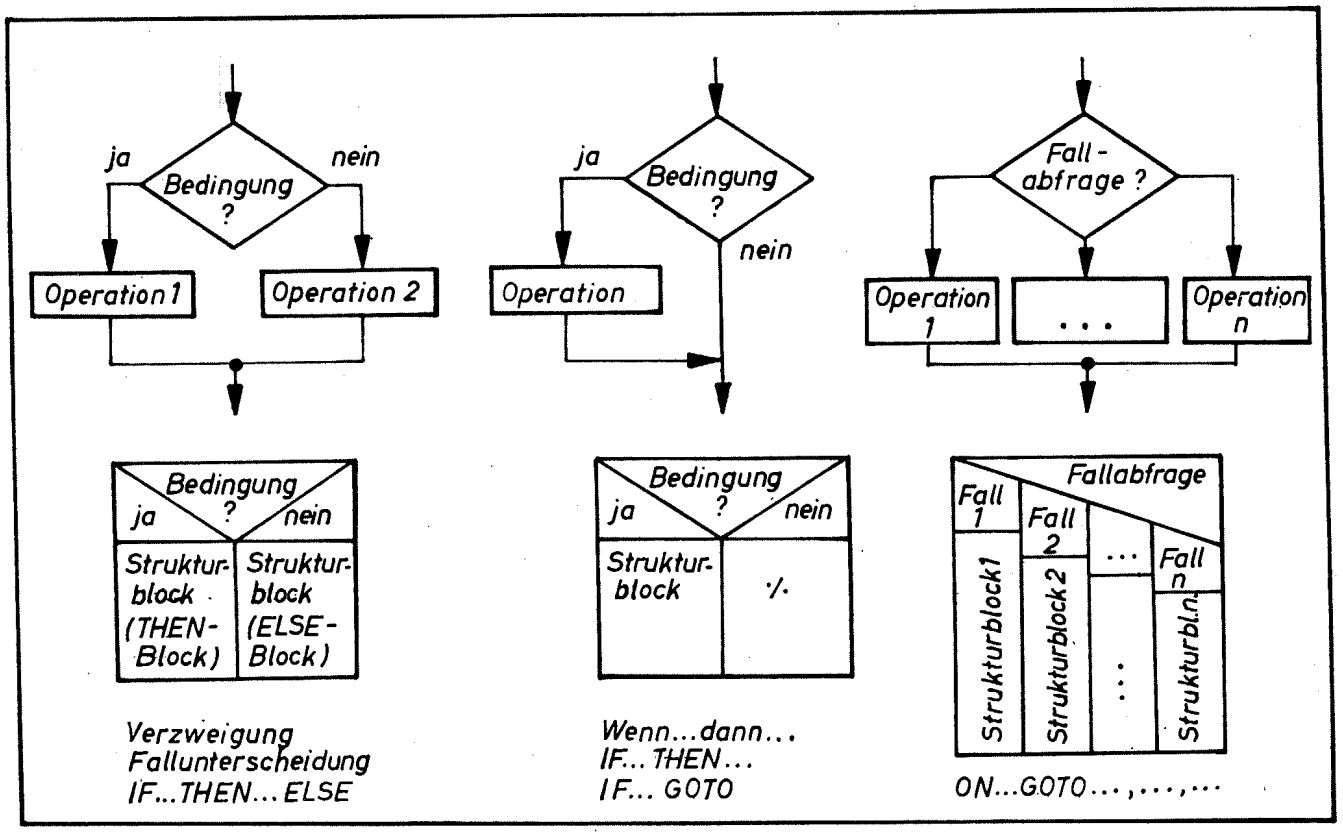


Bild 8.8.2. Verschiedene Möglichkeiten von Verzweigungen

8.8.1.4. Schleifen

Schleifen werden zuweilen auch als geschlossene Strukturen bezeichnet: Ein Programmabschnitt wird hier wiederholt ausgeführt. In den verschiedenen Programmiersprachen existieren hierfür Ausdrücke wie: WHILE ... DO; FOR ... DO; WITH ... ... DO; REPEAT ... UNTIL; FOR ... TO ... STEP ... NEXT.

Schleifen lassen sich nach zwei Prinzipien bezüglich der Abbruchbedingungen unterscheiden:

Art und Lage.

Die *Art* der Abbruchbedingung führt zu

- *Zählschleifen* bzw. sukzessiven Schleifen. Bei Eintritt in die Schleife steht fest, wie oft sie durchlaufen wird.

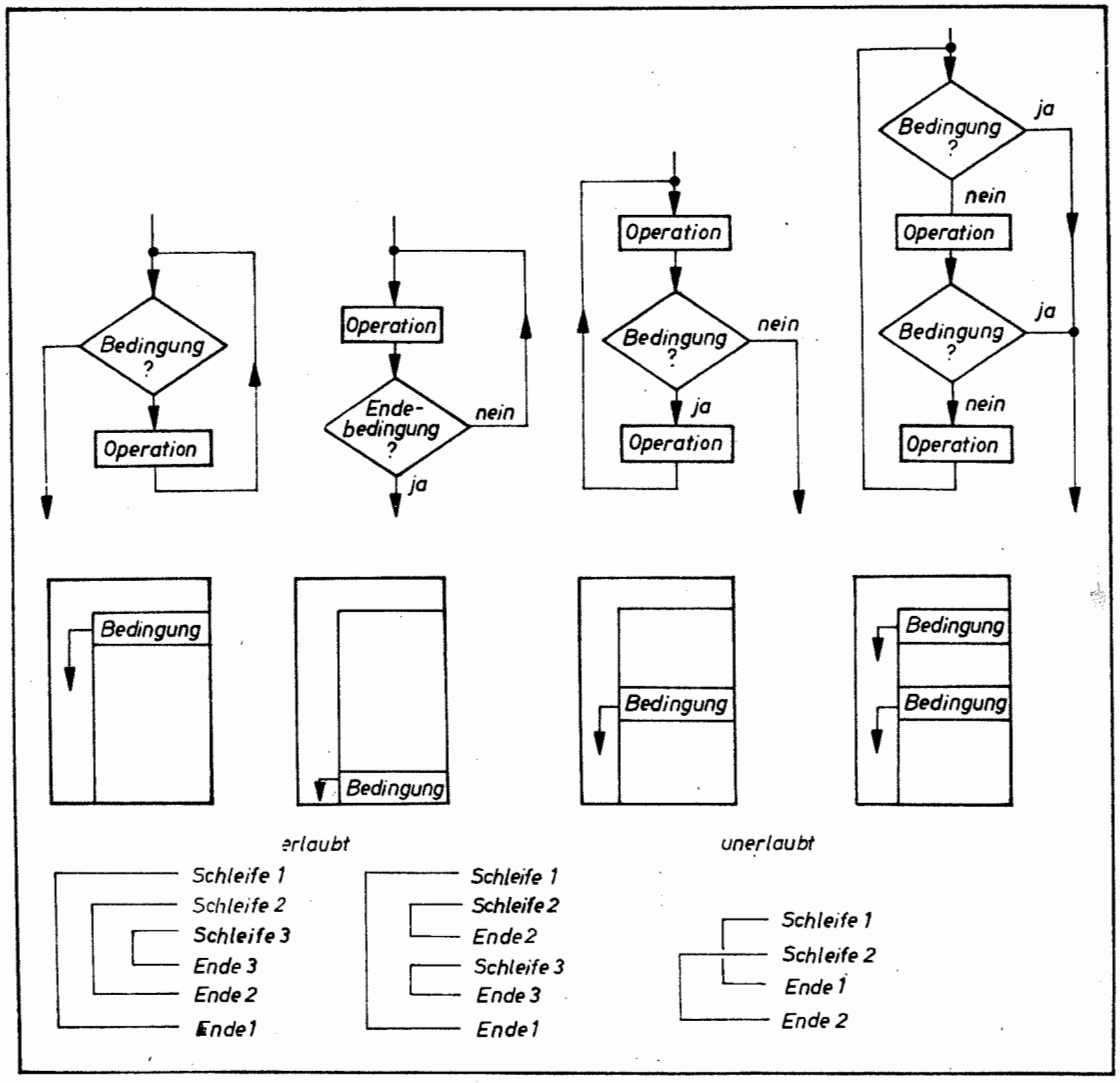


Bild 8.8.3. Verschiedene Möglichkeiten bei Schleifen sowie erlaubte und unerlaubte Schachtelung von Schleifen

- *Bedingungsschleifen* bzw. datenabhängigen oder iterativen Schleifen. Sie werden so oft durchlaufen, bis die Abbruchbedingung als Funktion der berechneten Daten erreicht ist. Beispiele sind: Unterschreiten eines Fehlers oder Finden eines gesuchten Wortes.

Bezüglich der *Lage* der Abbruchbedingung zum verarbeitenden Teil gibt Bild 8.8.3 einen Überblick. Die Bedingungen können also am Anfang, in der Mitte oder am Ende des Verarbeitungsteiles der Schleife liegen. Je nachdem wird der Verarbeitungsteil minimal keinmal, z. T. keinmal und z. T. einmal oder generell einmal durchlaufen. Es sind aber auch Schleifen möglich, die mehrere Entscheidungen enthalten.

Ein zusätzliches Problem der Schleifen besteht darin, daß mehrere Schleifen ineinander geschachtelt werden können. Hierbei müssen die Anordnungen besondere Relationen berücksichtigen. Nicht erlaubt ist es, daß Schleifen ineinander verweisen.

8.8.1.5. Datenstrukturierung

Die Datenstrukturierung betrifft zwei Unterscheidungsmerkmale:

- Verarbeitungs- und Steuerungsdaten
- lokale und globale Daten.

Steuerungsdaten betreffen vor allem Entscheidungen bzw. Bedingungen. Sie können also sowohl in den Bedingungen vorkommen als auch in den Sprüngen, d. h. als Adresse.
Verarbeitungsdaten sind jene Daten, die z.B. bei Berechnungen als Ausgangswerte, Zwischenwerte und Ergebnisse auftreten. Alphanumerische oder grafische Daten sind fast ausschließlich Verarbeitungsdaten.

Die Trennung von Verarbeitungs- und Steuerungsdaten ist bei Bedingungen schwierig, die vom Verarbeitungsergebnis abhängen. Hier ist daher besonders kritisch auf die Bedeutung der Daten zu achten.

Die Unterscheidung in *lokale und globale Daten* ist notwendig, wenn zusammengesetzte Blöcke verwendet werden. Sie läßt sich nur bei einigen Programmiersprachen generell realisieren, ist aber in jeder Sprache teilweise durch Wiederverwendung von Variablen für Zwischenergebnisse möglich. Jeder Strukturblock hat dann seine speziellen lokalen Daten, die nur für ihn verbindlich sind und vor allem für Zwischenergebnisse, Laufparameter usw. verwendet werden: Daten also, die außerhalb des Blocks nicht mehr benötigt werden. Die globalen Daten besitzen dagegen für (fast) alle Strukturblöcke Bedeutung und werden deshalb von Block zu Block übergeben. Anders ausgedrückt bedeutet dies, daß zu den globalen Daten jederzeit von jedem Block zugegriffen werden kann. Die lokalen Daten müssen dagegen zu Beginn jedes Blocks neu definiert werden und verlieren am Ende ihren Sinn.

Zur Datenstrukturierung gehört schließlich noch, daß die *Namen der Daten* so gewählt werden sollen, daß sie klar erkennen lassen, zu welchem Objekt sie gehören.

8.8.1.6. Unterprogramme

Unterprogramme sind Programmabschnitte, die mehrfach in einem Programm verwendet werden können. Sie sind in der strukturierten Programmierung nicht vorgesehen, da sie das Prinzip durchbrechen, daß jeder Teilblock (gleich welcher Hierarchiestufe) nur genau einen Ein- und einen Ausgang besitzen soll. Sie besitzen aber oft so große Vorteile, insbesondere was Einsparung von Speicherplatz für das Programm und vereinfachte Schreibweise des Programms selbst betrifft, daß sie in allen Sprachen und Sprachebenen vorgesehen sind. Typische Befehle hierfür sind CALL und GOSUB. Aus diesem Grunde wird hier zur Erklärung auch die Anordnung im Programmspeicher gemäß Bild 8.8.4 verwendet.

Die Grundstruktur des Unterprogramms zeigt Teilbild a). An verschiedenen Stellen des Hauptprogramms wird mittels GOSUB 300 (CALL 300) in das Unterprogramm gesprungen, und jedesmal erfolgt die Rückkehr ins Hauptprogramm, sobald der Befehl RETURN erreicht wird. Das Unterprogramm benötige einschließlich des Befehls RETURN n Speicherplätze, und m-mal werde das Unterprogramm im Hauptprogramm verwendet. Dann werden

$$x = m(n - 2) - m + n \tag{1}$$

Speicherplätze eingespart gegenüber einer Anordnung, bei der das Unterprogramm (dann ohne RETURN) direkt an der Stelle von GOSUB stehen würde. Die Einsparung ist folglich um so größer, je öfter das Unterprogramm verwendet wird. Um so öfter sind aber auch die Adressenänderungen und die Stapelung der Rückkehradresse notwendig. Dies bedeutet, daß ein häufiger Unterprogrammaufruf, der viel Speicherplatz einspart, anderseits auch mehr Zeit im Programmablauf erfordert. Hier gilt es also, ein Optimum unter verschiedenen Bedingungen zu finden.

Im Teilbild b) ist eine abgewandelte Variante des Unterprogramms dargestellt. Es entspricht dem IF-THEN-Teil in Bild 8.8.2. Hiermit soll gezeigt werden, daß ein

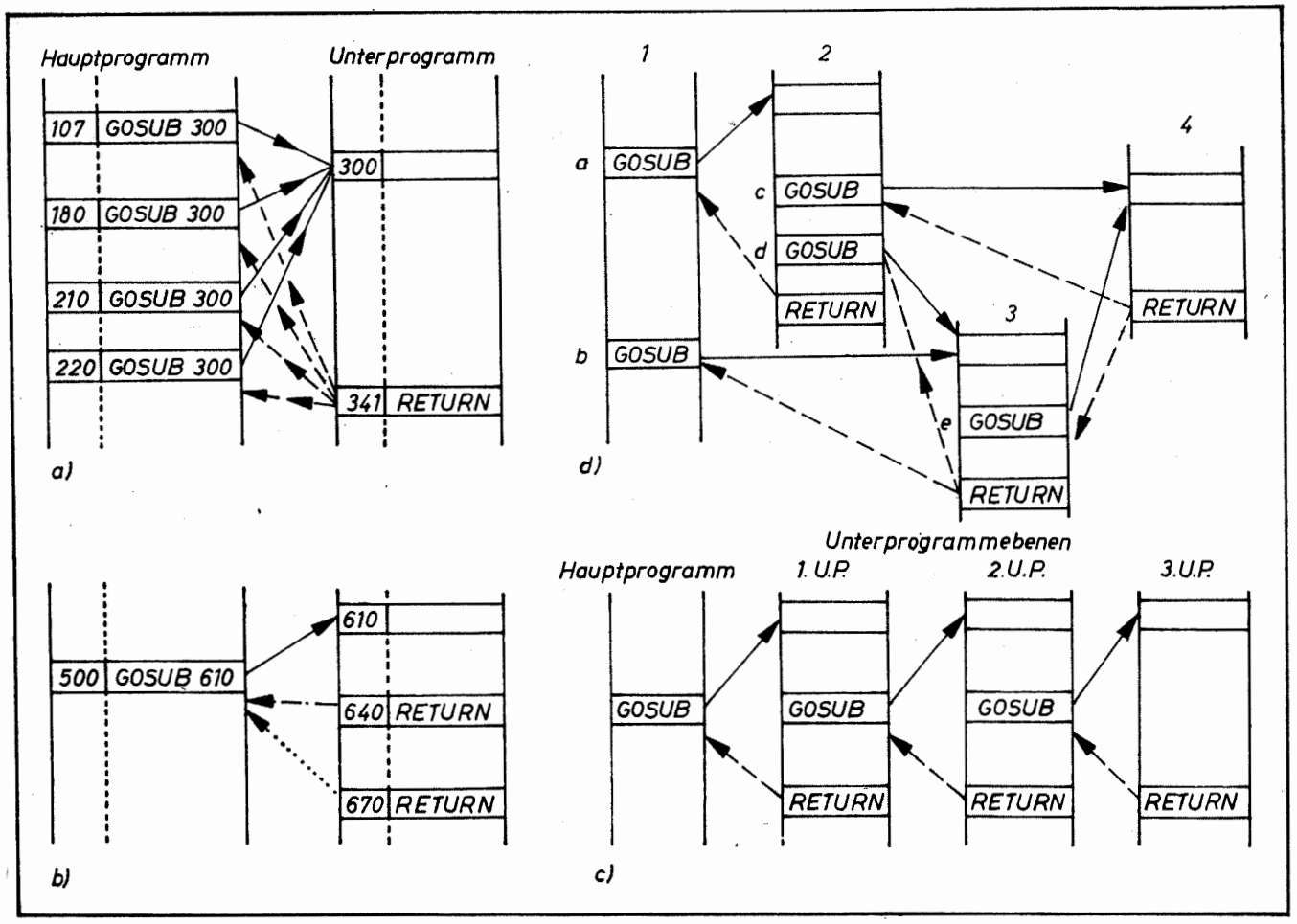


Bild 8.8.4. Verschiedene Arten von Unterprogrammen

Unterprogramm auf verschiedene Weisen beendet werden kann, aber jeder Zweig des Unterprogramms mit RETURN oder einem äquivalenten Befehl abschließen muß.

In Teilbild c) ist gezeigt, wie mit mehreren Unterprogrammebenen 1. UP bis n. UP gearbeitet wird. Hier wurde eine bewußt übersichtliche Struktur verwendet, um diese Ebenen deutlich hervorzuheben. Sie sind wegen Gl. (1) aber erst dann nützlich, wenn einzelne Unterprogramme mehrfach angesprungen werden. Hierbei können sehr komplexe Zusammenhänge realisiert werden, und es kann auch eine hohe Unübersichtlichkeit entstehen. Dies sei an einem noch relativ einfachen Beispiel mit Teilbild d) demonstriert. Es bestehen die Aufrufe

$$1 \nearrow a \to 2 \nearrow c \to 4 \quad 2 \searrow d \to 3 \to e \to 4 \,.$$
$$1 \searrow b \to 3 \to e \to 4$$

Das Unterprogramm 3 taucht folglich je einmal in der 2. und 3. Unterprogrammebene auf, das Unterprogramm 4 in der 3. und 4. Ebene.

8.8.2. Einteilung der Software

8.8.2.1. Beispielprogramm Addition

Es sei folgende Aufgabenstellung zu programmieren:

2 n-stellige Zahlen sind zu addieren.

Bei den Befehlen des Z 80 sind nur 8- bzw. 16-bit-Additionsbefehle vorhanden. Für größere n wird es also sinnvoll sein, die beiden Zahlen in je einem Speicherbereich unterzubringen und einen dritten für das Ergebnis zu reservieren. Eine günstige Anordnung einer 16stelligen BCD-Zahl in einem Speicherbereich zeigt Bild 8.8.5c) und d). Es ist in diesem Beispiel günstig, neben den Zahlen auch das Programm anzuordnen, so ergibt sich ein Speicherplan, wie ihn Bild 8.8.5b) zeigt.

Der Adreßbereich von 3C5Ø bis 3C9F ist zufällig ausgewählt, er könnte auch an beliebiger anderer Stelle liegen. Die Anordnung der Zahlen ist bereits so gewählt, daß bei den niedrigsten Adressen auch die wertniedrigsten BCD-Ziffern stehen. Deshalb muß bei 3C7Ø und 3C8Ø die Addition begonnen werden. Das Ergebnis (ohne eventuellen Überlauf, der im Flag CY steht) ist nach 3C9Ø zu transportieren. Danach sind die drei Adressen zu erhöhen und erneut die zwei Zahlen unter Berücksichtigung

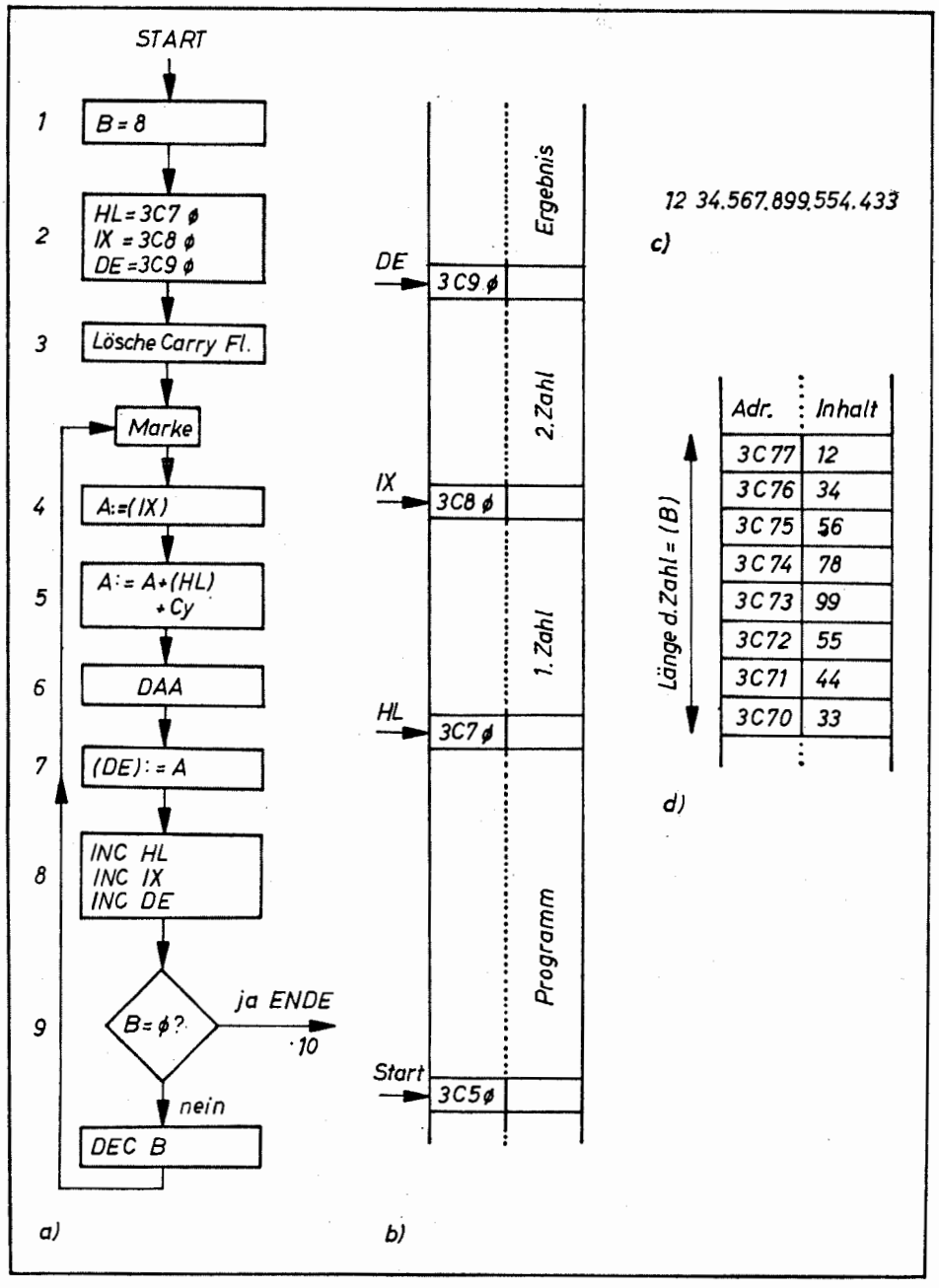


Bild 8.8.5. Zur Addition von zwei mehrstelligen BCD-Zahlen

a) Programmablauf
b) Anordnung von Programm und Zahlen im Speicher
c) Beispiel einer 16stelligen Zahl
d) Anordnung der Zahl im Speicher

des CY-Flag zu addieren und zu speichern. Dann sind wieder die drei Adressen um 1 zu erhöhen usw., bis alle Speicherstellen abgearbeitet sind. Auf diese Weise ergibt sich der Programmablaufplan von Bild 8.8.5a). B = 8 bedeutet, daß im Register B die Länge, also eine 16stellige Zahl, steht. Vom Speicherplatz (s. Teilbild b)) her könnte B zwischen 1 und 16 frei gewählt werden. Mit einem anderen Wert für B lassen sich also bei dieser Speicheranordnung bis zu 32stellige Dezimalzahlen im BCD-Format addieren. Als Zeiger für die Speicherplätze werden die Register HL, IX und DE verwendet (2). Zu Beginn der Rechnung ist das Carry-Flag zu löschen (3). Die Addition erfolgt im Akkumulator A (4, 5). Der Befehl DAA ist notwendig, weil BCD-Zahlen und keine Binärzahlen verwendet werden sollen (6). Dann wird das Ergebnis in den Speicher transportiert (7), die drei Zeigerregister incrementiert (8) und entschieden, ob B bereits Null ist (9).

Das zugehörige Programm zeigt Tab. 8.8.1 in einer durch Mnemonics und Kommentare lesbaren Form. Es benötigt 27 byte, hat also in dem vorgesehenen Speicherraum Platz.

Tabelle 8.8.1. Programm und Hex-Dump der Addition von zwei 16stelligen BCD-Zahlen mit dem Z 80. Die Numerierung bezieht sich auf den Programmablaufplan in Bild 8.8.5a).

Programm:

Marke	Nr.	Adr.	Kode	Mnemonic	Kommentar
	1	3C5Ø	Ø6 Ø8	LD B,Ø8	;8 byte für 16 Stellen
	2	3C52	21 7Ø 3C	LD HL,3C7Ø H	;Zeiger für ersten Summanden
		3C55	DD 21 8Ø 3C	LD IX,3C8Ø H	;Zeiger für zweiten Summanden
		3C59	11 9Ø 3C	LD DE,3C9Ø H	;Ergebnis ab 3C9Ø
	3	3C5C	37	SCF	;Setzen des Carry-Flag
		3C5D	3F	CCF	;Complementieren des Carry-Flag
Addi:	4	3C5E	DD 7E ØØ	LD A,(IX+d)	;lade A mit Summanden
	5	3C61	8E	ADC A,(HL)	;addiere zweiten Summanden
	6	3C62	27	DAA	;Korrektur auf Tetraden
	7	3C63	12	LD (DE),A	;lade Speicherplatz (DE) mit Ergebnis
	8	3C64	23	INC HL	;erhöhe alle Zeiger um 1
		3C65	13	INC DE	
		3C66	DD 23	INC IX	
	9	3C68	10 F4	DJNZ Addi	;wenn nicht fertig, dann zur Marke
		3C68			
	10	3C6A	DF	RST 18H	;Abbruch z. B. zur Anzeige

Hex-Dump:

	Ø	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
3C5Ø	Ø6	Ø8	21	7Ø	3C	DD	21	8Ø	3C	11	9Ø	3C	37	3F	DD	7E
3C6Ø	ØØ	8E	27	12	23	13	DD	23	10	F4	DF					

Einen Hex-Dump des Programms zeigt ebenfalls Tab. 8.8.1. Er enthält in dicht gepackter Form die auf den Speicherplätzen vorhandenen hexadezimalen Werte der Befehle. Dies ist die kürzeste, übersichtlichste Form für die Dokumentation von Programmen. Diese Form ist gut für die Tastatureingabe kleiner Programme geeignet. Ein Hex-Dump ist aber sehr schwer zu verstehen. Um den Ablauf eines Programms erkennbar zu machen, wird meist die obere Form von Tab. 8.8.1 gewählt. Sie ist aber offensichtlich sehr redundant.

8.8.2.2. Arten von Programmen

Mit dem zuvor beschriebenen Programm ist die Addition zweier vielstelliger Zahlen realisiert. Dies ist jedoch nur ein Teilproblem der Aufgaben für den Rechner, wenn er Zahlen addieren soll. Denn damit er dies vollzieht, müssen die beiden Zahlen zunächst per Tastatur auf die beiden Speicherplätze gebracht werden. Dies erfordert ein spezielles Eingabeprogramm. Letztendlich muß das Ergebnis angezeigt werden können. Dies erfordert ein spezielles Ausgabeprogramm. Weiter muß das Programm selbst mit einer Eingaberoutine eingegeben werden. Es ist also für diese und ähnliche Aufgaben zu jedem Rechner eine *Minimalsoftware* für seinen Betrieb notwendig. Sie wird in der Regel auf ROM gespeichert und damit fest eingebaut. Derartige Grundprogramme heißen Monitor. Da man sich bei einer solchen Eingabe leicht vertippt, gehört zu einem Monitor meist auch ein kleiner Editor, mit dem die eingegebenen Hex-Zahlen angezeigt und eventuell korrigiert werden können.

Die Programmentwicklung für das einfache Programm erfolgte auf intuitivem Wege. Große Programme lassen sich so in der Regel nicht entwickeln. Hierzu benötigt man *Entwicklungssysteme*. Sie bestehen oft aus dem *Assembler*, der aus Mnemonics die Bitkombinationen generiert. Im Zusammenhang mit dem Assembler steht oft auch seine Umkehrung, der Disassembler. Er generiert aus den Hex-Kodes z. B. im Sinne eines Hex-Dump ein beschreibendes Programm mit Mnemonics, das wesentlich leichter lesbar ist.

Zu einem Entwicklungssystem zählen auch weitere Teilprogramme, auf die im Abschnitt 8.8.3. eingegangen wird.

Für die Anwendung eines Mikrorechners benötigt man spezielle *Anwenderprogramme*. Dies kann z. B. ein Programmsystem für Numerik, Textverarbeitung, Bildverarbeitung, Prozeßsteuerung und ähnliches sein.

Universelle Programmsysteme sind die *höheren Programmiersprachen*, wie z. B. BASIC, PASCAL usw. Sie erleichtern dem Programmierer ganz wesentlich die Arbeit.

8.8.3. Maschinennahe Programmierung

Die Erzeugung und Eingabe eines Programms für den Rechner kann auf *verschiedenen Niveaus* erfolgen:

- direkte Eingabe der Bitkombinationen
- Eingabe der entsprechenden Hex-Zahlen
- Benutzung eines Assemblers
- Verwendung eines anderen Rechners
- Benutzung einer höheren Sprache.

Die beiden ersten Varianten sind nahezu trivial, aber auch zugleich sehr mühselig und nur unter zwei Gesichtspunkten vertretbar: Bei Entwicklung eines kurzen Programms (bis ca. 100 Byte) oder, wenn ein bereits ausgearbeitetes Programm vorliegt, das nur eingegeben und eventuell an wenigen Stellen angepaßt werden soll.

8.8.3.1. Assembler

In diesem Fall ist der Assembler ein fertiges Maschinenprogramm, welches bereits für jenes System existiert, auf dem weitere Programme entwickelt werden sollen. Es benötigt zwischen 5 und 10 Kbyte Speicherraum. Ist dieses System besonders kom-

fortabel für die Programmentwicklung ausgelegt, so wird von einem Entwicklungssystem gesprochen.

Der Assembler besteht meist aus mehreren Teilprogrammen, die im folgenden nacheinander besprochen werden.

In den Assembler gibt man ein Quellprogramm (Source-Kode) ein, das eine gewisse Ähnlichkeit mit den Spalten von Tab. 8.8.1 hat. Es zeichnet sich durch *zwei Eigenschaften* aus:

- Verwendung von Mnemonics für Befehle
- Verwendung von Symbolen für Daten und Adressen.

Für eine speziell formatierte Eingabe existieren *vier Felder*:

Markenfeld	Operationsfeld	Operandenfeld	Kommentarfeld
Addi:	LD	A, (IX + d)	; lade A mit Summanden

Das *Markenfeld* kann leer bleiben. Es wird immer dann mit einem frei wählbaren Namen (bestimmte Länge) belegt, wenn diese Zeile einmal aus späteren oder früheren Programmteilen durch Sprungbefehle erreicht werden muß. Der Assembler weiß dann über diese Marke den zugehörigen Befehlszählerzustand. Es liegt also über die Marke eine symbolische Adressierung der Zeile vor. Das Markenfeld endet oft mit einem Doppelpunkt.

Im *Operationsfeld* steht einer der 74 für den U 880 möglichen Maschinenbefehlsmnemonics (oder ein Markenaufruf oder eine Pseudo- bzw. Steueranweisung). Das Operationsfeld muß an seinem Ende mindestens eine Leerstelle besitzen.

Das *Operandenfeld* enthält eine Hex-Zahl oder Adressen. Mehrere Operanden werden durch Komma getrennt. Als eine spezielle symbolische Adresse, z. B. für einen Sprung, kann auch auf die Marke einer Zeile verwiesen werden. Bei einigen Befehlen bleibt das Operandenfeld leer.

Das *Kommentarfeld* beginnt (wenn es nicht leer bleibt) mit einem Semikolon. Der danach folgende Text wird vom Assembler ignoriert. Er dient nur dem Programmierer bzw. Anwender zur besseren Verständlichkeit des Programms.

Eine vollständige Zeile dieser Art heißt *Quellanweisung* (source statement). Sie enthält also weder eine Zeilennummer noch eine Speicheradresse, wohin das Programm später geladen werden soll. Auch fehlt hier der binäre oder hexadezimale Kode für den Befehl. Hierdurch ist es leicht möglich, ein Quellprogramm zu editieren. Es können Zeilen gelöscht und eingefügt werden, Label a s Marken lassen sich zusätzlich anbringen usw.

Die Konsequenz dieser symbolischen Behandlung von Adressen ist, daß einige *Pseudooperatoren* verwendet werden müssen. Üblich sind hier u. a.:

ORGnn (Origin) gibt an, daß auf dem Speicherplatz nn das Programm beginnen soll.

EQUnn (Equivalent) gibt an, welchen Zahlenwert eine Variable ständig haben soll: ABC EQU 3FH = Der Variablen ABC wird der Wert 3F Hex zugewiesen.

DEFnn (Definition) definiert bis zum nächsten Befehl dieser Art, welchen Wert eine Variable besitzen soll: AQ DEF 7FH = Der Variablen AQ wird vorübergehend der Wert 7F Hex zugewiesen.

GLOBAL definiert einen Operanden als global für das gesamte Programm, d. h., er behält seine Bedeutung nur einmal im Programm, z. B. GLOBAL ACD.

Weiter können zur Vereinfachung *Macros* benutzt werden. Sie entsprechen in etwa den Unterprogrammen mit CALL, jedoch mit dem Unterschied, daß der Assembler sie bei jedem Aufruf vollständig ins Programm einfügt. Dabei kön en zusätzlich formale Parameter verwendet werden.

Das *Quellprogramm* wird vom Assembler meist in *zwei Durchläufen* übersetzt. Im ersten Durchlauf legt er Symboltabellen für die Marken, Adressen und Daten an.

Im Zweiten Durchlauf übersetzt er Zeile für Zeile nacheinander in den Maschinenkode. Nach jedem Durchlauf weist er Fehler aus. Auf dieser Basis kann das Quellprogramm editiert werden. *Bei fehlerfreien Durchläufen existieren keine Syntaxfehler mehr, die der Assembler erkennt.* Dies bedeutet aber noch lange nicht, daß ein solches Programm dann fehlerfrei arbeitet. Dies kann erst in der *Testphase* festgestellt werden. Der Assembler weiß nicht, was das Programm tut.

Das Objektprogramm fügt zu dem Quellprogramm die Speicheradressen und die Befehlskodes hinzu und ersetzt die symbolischen Variablen. Auf Wunsch druckt der Assembler auch die Werte und Zeilen aus, in denen Marken verwendet werden. Diese Liste entsteht im ersten Durchlauf und heißt *Cross-Reference*. Die Ausgabe der Dokumentation wird oft als dritter Durchlauf des Assemblers bezeichnet.

Sowohl das *Quell-* als auch das *Objektprogramm* kann mittels Assembler extern, z.B. auf Band oder Floppy disc, *gespeichert* werden. Beide können weiter ausgedruckt werden. Die Speicherung des Quellprogramms ist, obwohl fast hundertmal so viel Speicherplatz wie für das Objektprogramm benötigt wird, aus zwei Gründen nützlich: Es ist eine *gute Dokumentation*, und durch Änderung der Pseudooperatoren kann ein Objektprogramm erzeugt werden, das an anderen Stellen des Speicherbereichs arbeitet bzw. andere Konstanten verwendet.

Das Objektprogramm kann schließlich direkt in den richtigen Speicherbereich geladen werden. Hierzu existiert meist ein spezielles Programm, der *Lader*. Mit Hilfe eines *Programmiergerätes* kann es auch in PROM übertragen werden.

Bei vielen Assemblern ist es auch möglich, Programmteile, die vorhanden sind oder entwickelt werden, miteinander zu verbinden. Das dafür notwendige Maschinenprogramm ist der *Binder* bzw. Linker.

8.8.3.2. Testhilfen

Ein Programm, insbesondere ein komplexes, ist so gut wie nie fehlerfrei. Bei der ersten Erprobung weist es — trotz Fehlerfreiheit im zweiten Assemblerlauf — meist noch Fehler auf, die schrittweise zu beheben sind. Hierzu existiert ein *Debugging*-Programm (bug = engl. Wanze, also ein Programm zum Entwanzen). Es besteht aus drei Teilen:

- Anzeige der Register und Speicherinhalte
- Befehlsweises Abarbeiten des Programms
- Methoden zum Ändern des Programms.

Für das befehlsweise Abarbeiten existieren zwei Varianten:

- *Single step*: Es wird immer nur ein Befehl abgearbeitet, und danach werden die Registerinhalte angezeigt
- *Breakpoints* = Haltepunkte, die vorübergehend an beliebiger Stelle im Programm angeordnet werden können, dann einen genau definierten Abschnitt durchlaufen, und danach werden die Register angezeigt.

Beim Durchlaufen von Programmen bis zu Breakpoints wird auch von *Patching* gesprochen. Das durchlaufene Programmstück wird Trace genannt.

Schließlich gibt es noch die Möglichkeit, für jedes Programm spezielle *Testprogramme* zu schreiben bzw. genau bekannte Beispiele für den Testdurchlauf zu benutzen.

Zu den Testhilfen spezieller Art kann auch der Disassembler gerechnet werden. Er übersetzt ein Maschinenprogramm zurück in eine Form, die dem Quellprogramm ähnlich ist, und macht dadurch das Maschinenprogramm wieder für den Menschen lesbar.

8.8.4. Grundsoftware

Ein Rechner ist ohne eine Grundsoftware nicht betriebsfähig. Mit der Grundsoftware *muß mindestens folgendes möglich sein*:

- Eingabe von Daten und Programmen in den Speicher
- Anzeige von Speicher- und Registerwerten
- Starten eines eingegebenen Programms.

Diese Grundsoftware ist ein spezielles Maschinenprogramm, das in der Regel in einem ROM festgelegt ist und den Namen *Monitor* trägt. Es gibt natürlich Monitore unterschiedlicher Leistungsfähigkeit. Auf die Monitorroutinen (Teilprogramme) greifen alle übergeordneten Programmsysteme zurück. Das gilt auch für die zuvor beschriebene maschinennahe Software, also u. a. für Assembler, Debugger usw. Deshalb sind bei der Grundsoftware schon Teilprogramme vorhanden, die bereits auf dem speziellen und damit übergeordneten Assembler-Niveau genannt wurden.

Dadurch, daß ein Monitor als ROM in einem bestimmten Speicherbereich liegt, ist er noch nicht in Betrieb. Beim Starten bzw. RESET des Rechners muß vielmehr erreicht werden, daß das Monitorprogramm an einer genau definierten Stelle (Programmadresse) gestartet wird. Hierfür sorgt eine spezielle Hard- (und Soft-) Ware. Sie heißt *Urlader* bzw. Bootstrap-Lader. Dieser Begriff wird auch dann verwendet, wenn das Monitorprogramm nicht in einem ROM (residenter Monitor) bei einer bestimmten Adresse, sondern auf einem Floppy disc existiert. Dann sorgt der Urlader dafür, daß der Floppy gestartet wird, daß das Monitorprogramm in den richtigen Speicherbereich gelesen und dort gestartet wird.

Mit dem Starten des Monitorprogramms werden gleichzeitig verschiedene Parameter, periphere Geräte usw. in genau definierter Weise als *Default*-Werte eingestellt. So wird z.B. eine bestimmte Ausgabe (der Bildschirm) und Eingabe (die Tastatur) adressiert. Ferner werden den Speichern bestimmte Werte, z.B. Null, zugeordnet. Danach ist das Monitorprogramm in Betrieb und meldet sich mit einem definierten Zeichen, dem Prompt. Jetzt können Befehle, u. a. zur Eingabe, Ausgabe, Setzen von Parametern, Start usw., erteilt werden. Ein Monitor besitzt in der Regel auch einen kleinen *Editor*, mit dem Text- bzw. Programmzeilen ein- und ausgegeben werden können. Er ermöglicht es, Zeichen zu ändern, Zeilen und Zeichen einzufügen bzw. zu löschen. Bei Rechnern mit *Bildschirm* kann er dies auf vielfältige Weise steuern. Der *Cursor* (welcher anzeigt, wo ein neues Zeichen hingeschrieben wird) kann in allen Richtungen über den Schirm bewegt werden, der Bildschirm kann gelöscht werden usw.

Jeder Monitor besitzt verschiedene *Fehleranzeigen*, die angeben, daß ein unzulässiger Befehl oder eine unzulässige Eingabe vorliegt. Viele Monitore besitzen auch Test- und Diagnoseprogramme, z.B. um zu prüfen, ob alle Speicherzellen einwandfrei arbeiten, ob die peripheren Geräte funktionstüchtig sind usw.

Im Zusammenhang mit den Floppy discs sind spezielle leistungsfähige Betriebssysteme entstanden. Das wohl erste System dieser Art entstand 1973 bei INTEL für den 8080 von G. A. KINDALL. Es ist das **CP/M**, welches heute als Standardsoftware für die meisten 8-bit-Prozessoren gilt. Seit 1976 besteht es *aus drei Teilroutinen*:

CCP control commond prozessor (Bedienungsroutine)

BDOS basic disk operating system (Steuerung des Lesens und Schreibens auf dem Floppy)
BIOS basic input output system (Steuerung der Ein- und Ausgabe)

Jede dieser Routinen hat mehrere komfortable Befehle, und nur der Teil CCP ist für jeweils einen speziellen Mikroprozessor geschrieben.

Ein anderes wichtiges Betriebssystem, das RIO, stammt von Zilog. Für die 16-bit-Prozessoren hat das Betriebssystem UNIX Bedeutung erlangt.

8.8.5. Höhere Programmiersprachen

8.8.5.1. Compiler und Interpreter

Höhere Programmiersprachen dienen der einfacheren Programmierung. Hierbei werden spezielle feststehende Wörter als Befehle, Variable, Adressen usw. verwendet. Sie sind in der Regel der englischen Sprache so entlehnt, daß sie relativ leicht der Bedeutung nach erlernbar sind. Während die Mnemonics eines Assemblers genau einem Maschinenbefehl entsprechen, fassen die Wörter einer höheren Sprache eine ganze Folge von Befehlen zusammen. Dadurch ist ein Programm in höherer Programmiersprache sowohl kürzer als auch übersichtlicher, z.B. entspricht im Mittel eine Programmierzeile von Programmen in einfachem BASIC ca. 40 Befehlen in Maschinensprache.

Von der Wirkungsweise her lassen sich die höheren Programmiersprachen in zwei Gruppen, nämlich Compiler und Interpreter, einteilen. Beide sind wiederum selbst Maschinenprogramme. In beiden Fällen wird, ähnlich wie beim Assembler, das Quellprogramm und das Objektprogramm unterschieden. Das Quellprogramm wird mittels eines Editors in den Speicher eingegeben. Unterschiedlich wird jedoch der Objektkode erzeugt.

Beim *Compiler* wird der Objektkode wie beim Assembler komplett aus dem Quellprogramm erzeugt. Es existieren dabei auch meist zwei Durchläufe. Wenn der Objektkode erzeugt wurde, ist der Compiler im Prinzip überflüssig, denn das fertige Maschinenprogramm benötigt ihn nicht mehr. Es ist — falls fehlerfrei — ohne Compiler und Quellprogramm unmittelbar auf dem Rechner ablauffähig. Der Compiler wird also nur zum Übersetzen und zum Korrigieren des Quellprogramms benötigt. Es benötigt auch nur für diese Zeit Speicherplatz.

Ein *Interpreter* wird dagegen ständig im Speicher des Rechners benötigt. Er übersetzt das Quellprogramm zeilenweise nur während der Anwendung. Der Objektkode wird also nie vollständig aus dem Quellprogramm erzeugt, sondern nur zeilenweise. Dies bedeutet, daß sowohl das Quellprogramm als auch der Interpreter beim Rechnerlauf ständig anwesend sein müssen. Der Objektkode entsteht also nur für den Augenblick, wo er benötigt wird, und kann damit nicht vollständig gewonnen werden. Dies hat den Nachteil, daß Interpreter meist erheblich langsamer arbeiten. Sie haben aber den Vorteil, daß sie besonders gut interaktiv arbeiten. Beim Programmablauf wird sofort deutlich, wann ein Fehler im Programm ist. Dies gilt sogar für kurze Teile von Programmen. Wenn der Fehler offensichtlich ist, wird er direkt im Quellprogramm korrigiert. Das ist der Grund, weshalb Interpreter besonders günstig beim Programmieren mit wenigen oder gar keinen Kenntnissen bezüglich der Maschinensprache sind.

Bezüglich des Vergleichs von Compiler, Interpreter und Assembler gelten die Aussagen von Bild 8.8.6. Die Kosten bezüglich einer Programmentwicklung (Teilbild a)) sind also beim Interpreter generell am niedrigsten. Das liegt daran, daß ein Interpreter selbst relativ einfach ist und wenig kostet und die interaktive Programment-

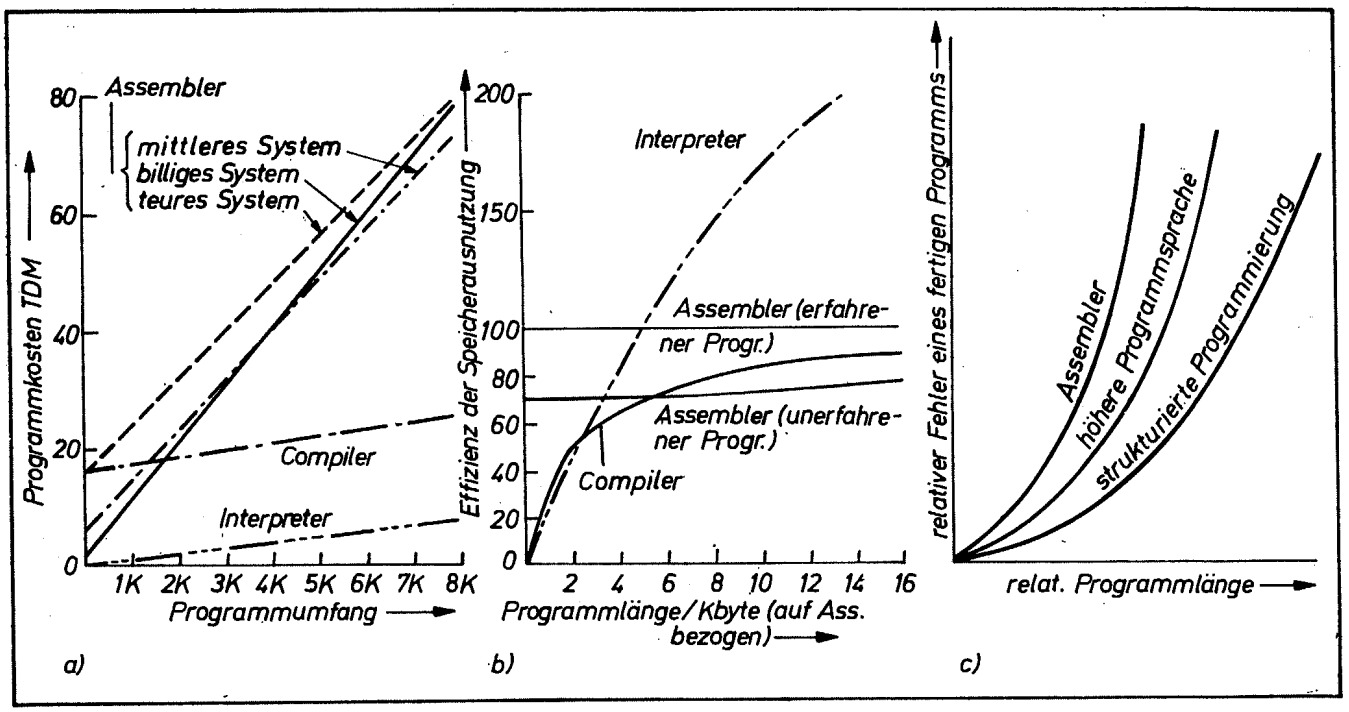


Bild 8.8.6. Zum Vergleich von Assembler, Interpreter und Compiler bezüglich
a) Programmkosten, b) Programmnutzung
c) auftretender Fehler nach Inbetriebnahme eines Programms

wicklung sehr effektiv ist. Ein Assembler benötigt je nach seiner Qualität höhere Anschaffungskosten. Außerdem ist hier die Programmierung kompliziert und daher je Befehl teuer. Ein Compiler erfordert die höchsten Anschaffungskosten. Infolge der Programmierung in der höheren Sprache verlaufen dann aber die auf das Programm bezogenen Kosten ähnlich flach wie beim Interpreter.

Bezüglich der Rechenzeit gilt die entgegengesetzte Aussage als zu den Kosten. Ein mittels Assembler erzeugtes Programm läuft generell am schnellsten. Ein über Compiler erzeugtes Programm ist deutlich langsamer, während ein mittels Interpreter ablaufendes Quellprogramm um etwa einen Faktor 10 langsamer ist.

Auch bezüglich der Ausnutzung der Speicherkapazität des Rechners gelten interessante und typische Aussagen. Ein compiliertes Programm ist hier in der Regel (besonders bei kleinen Programmen) recht uneffektiv. Ein Interpreter erreicht bei großen Programmen sogar eine Effektivität, die über einem assemblierten Programm liegt. Die Ursache dafür liegt darin, daß nie das vollständige Maschinenprogramm, sondern nur das kürzere Quellprogramm vorliegt.

In Teilbild c) ist qualitativ die relative Fehlerhäufigkeit als Funktion der Programmgröße gezeigt. Hier wird auch der Vorteil der strukturierten Programmierung (8.8.1.2.) deutlich.

8.8.5.2. Verschiedene Programmiersprachen

Es gibt heute mehrere hundert Programmiersprachen und über zwanzig für Mikrorechner. Dies hat vielfältige Gründe, sowohl die geschichtlich als auch durch die zahlreichen Anwendungen bedingt sind. Die erste Programmiersprache **FORTRAN** entstand 1953, als die Großrechner etwa so leistungsfähig wie heute ein Taschenrechner waren, und lag 1955 als erste Alternative zur Assemblerprogrammierung bei IBM vor. Ihrem Namen **for**mula **trans**later entsprechend, war sie gezielt für die naturwissenschaftlich-technischen Anwendungen konzipiert. Wie Bild 8.8.7 zeigt, gehört FORTRAN trotz vielfältiger Mängel bei einigen Weiterentwicklungen heute noch zu den am weitesten verbreiteten Sprachen. Seine Vorteile liegen bei der komplexen

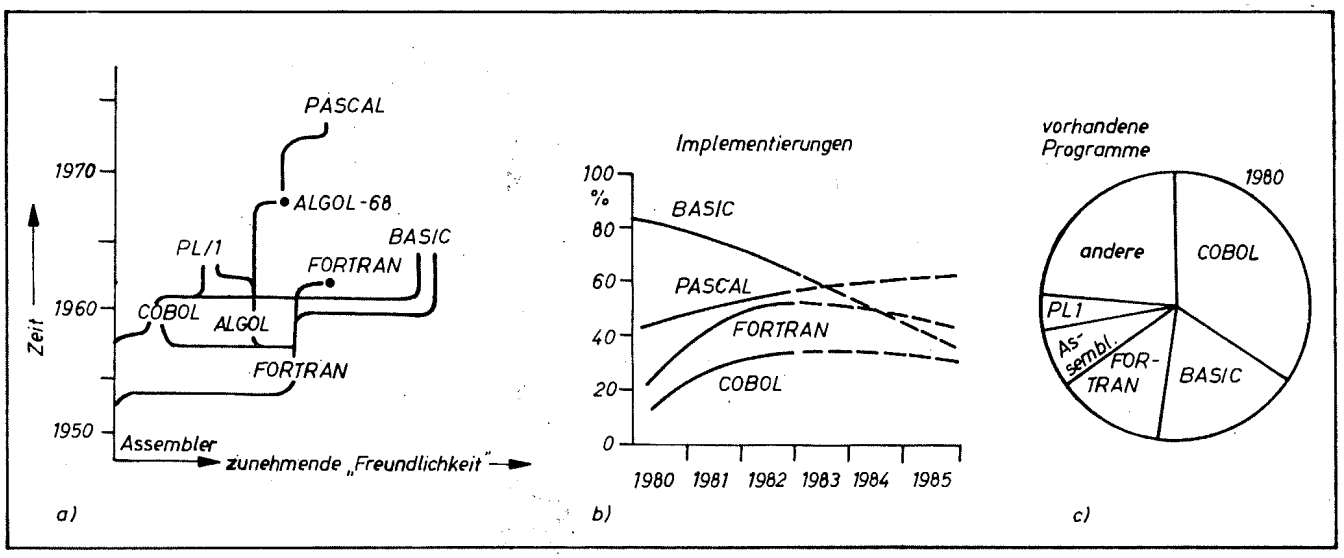


Bild 8.8.7. Zum Sprachvergleich

a) Zeitliches Entstehen einiger wichtiger Sprachen und ihre relative Anwenderfreundlichkeit

b) Relativer Anteil von Sprachimplementierungen. Ein Gesamtwert von über 100% ergibt sich daraus, daß auf einem Rechner oft mehrere Sprachen laufen. In diesem Bild scheint — bezogen auf andere Quellen — PASCAL überbewertet zu sein

c) Relativer Sprachanteil bei Programmen um 1980 (PASCAL wird in einer zu den Arbeiten gehörenden ausführlichen Tabelle nicht einmal ausgewiesen)

Numerik und einer sehr guten Normung (seit 1972 von der ISO); seine Nachteile bei der Textverarbeitung und keiner Möglichkeit zur strukturierten Programmierung.

Die zweite Sprache (mit den heute generell größten Programmsammlungen) ist **COBOL** (= **co**mmon **bu**sinell **o**rientated **l**anguage). Sie entstand auf Drängen von Kaufleuten, Managern und des USA-Verteidigungsministers in neun Monaten bei IBM und stand 1960 zur Verfügung. COBOL ist die erste Sprache, welche standardisiert wurde (bereits 1965). Obwohl sich heute alle Rechentechniker darüber einig sind, daß sie den Ansprüchen seit langem nicht mehr genügt, besteht keine Aussicht, sie zu ersetzen. Allein in den USA wurden bisher mehrere hundert Milliarden Dollar für COBOL-Programme investiert, und dieser Bestand ist wohl kaum jemals zu ersetzen. So besteht COBOL weiter und erfordert unvorstellbaren Programmieraufwand. Für Mikrorechner hat COBOL so gut wie keine Bedeutung.

Aus den Erfahrungen von FORTRAN bemühten sich die Theoretiker, eine „saubere" Sprache zu schaffen. Sie legten als Lösung 1960 **ALGOL** (= **algo**rithmic **l**anguage) vor. Sie hat keine große Anwendungsbreite erlangt, wurde aber 1968 weiter zu ALGOL 68 verbessert.

Auf der Basis von COBOL, FORTRAN und ALGOL 60 entwickelte IBM für die Generation 360 eine neue universelle Sprache **PL/1** (= **p**rogramming **l**anguage one). Sie sollte COBOL und FORTRAN ersetzen, erlangte aber keine größere Bedeutung. Hieraus kann gefolgert werden, daß sich eine zu universelle Sprache nicht einmal mit der Potenz von IBM durchsetzen läßt. Für Mikrorechner wurde hieraus eine kleine Untermenge als PL/M (Mikrorechner) abgeleitet. Auch sie fand keine große Verbreitung.

Für die Mikrorechner hat sich eine Sprache mit vielen Nachteilen zum Erstaunen ihrer Entwickler durchgesetzt, das **BASIC** (= **b**eginner's **a**ll-purpose **s**ymbolic **i**nstruction **c**ode). Sie wurde 1964 im Dartmouth Collage (Hanover, N. H.) unter Leitung von J. G. KEMENY und T. E. KURTZ für die Lehre entwickelt und stellt so etwas wie ein „abgemagertes" FORTRAN unter zusätzlicher Einbeziehung von Textverarbeitung dar. Ihr Vorteil ist die leichte Erlernbarkeit und die fast universelle Anwendbarkeit. Sehr schwach ausgebildet ist die Unterprogrammtechnik und die Dateibildung.

Ein weiterer Nachteil ist, daß sie praktisch nicht genormt wurde und daher heute eine Vielzahl von Dialekten mit sehr unterschiedlichem Sprachumfang vorliegt. Sie ist aber jene Programmiersprache, die heute am weitesten verbreitet ist. Sie wird von mehr als 10 Millionen Menschen „gesprochen". Mit COBOL und FORTRAN wird sie am häufigsten benutzt. Dies ist der Grund, weshalb umfangreiche Bemühungen zur Standardisierung laufen. Dabei wird auch versucht, den Nachteil, daß sie schlecht für strukturierte Programmierung geeignet ist, abzubauen. BASIC ist jene Sprache, die überwiegend in Interpretern verwendet wird.

Eine Sprache, der heute in vielen Arbeiten bezüglich Mikrorechner eine große Zukunft vorausgesagt wird, ist **PASCAL**. Sie wurde von JENSEN und WIRTH (ETH Zürich), auf ALGOL aufbauend, 1973 vorgestellt. Ihr Namen wurde zu Ehren von BLAISE PASCAL gewählt. Auch sie entstand vor allem für den Bereich der Ausbildung, berücksichtigt aber im Gegensatz zu BASIC die wesentlichen neuen Erkenntnisse bei der Programmierung. Sie ist also eine moderne Sprache mit guter theoretischer Fundierung. So kann mit ihr gut die strukturierte Programmierung angewendet werden. Sie ist stark selbstdokumentierend angelegt. Nachteilig bei PASCAL ist der schwierige Zugriff auf die Hardware des Rechners, u. a. bei Ein- und Ausgabe. Durch den bewußt stark eingeschränkten Sprachumfang entstanden auch bei PASCAL verschiedene Dialekte. Neben den euphoristischen Einschätzungen zu PASCAL gibt es auch kritische Stimmen, die annehmen, daß sich PASCAL genausowenig wie ALGOL gegen FORTRAN und COBOL gegenüber BASIC wird durchsetzen können.

Einen Vergleich einiger Sprachen zeigt neben Bild 8.8.7 auch Tab. 8.8.2.

Tabelle 8.8.2. Vergleich verschiedener Sprachen bezüglich der Anwendung von Interpretern und Compilern in Systemen, des mittleren Befehlsumfanges für einen Interpreter bzw. Compiler und des Arbeitsaufwandes zur Schaffung eines Interpreters bzw. Compilers. Die Gesamtsumme von über 100% ergibt sich dadurch, daß oft mehrere Sprachen auf einem System laufen.

	Interpreter	Compiler	Befehle	Mannjahre
BASIC	60%	20%	6000	1 bis 2
PASCAL	24%	30%	10000	2 bis 3
ALGOL 60	—	—	30000	5 bis 8
FORTRAN	5%	31%	50000	8 bis 10
COBOL	10%	15%	120000	15 bis 20
PL/1	2%	11%	350000	20 bis 30
Andere	37%	15%	—	—

8.8.6. Programmieren auf anderen Rechnern

Bisher wurde davon ausgegangen, daß die Programme auf dem Rechner entwickelt werden, auf dem sie Anwendung finden. Es gibt insbesondere aber für Mikrorechner einen anderen Weg, nämlich die Software auf einem anderen, meist einem Großrechner, zu entwickeln. Diese Großrechner werden dann Wirts- bzw. Host-Computer genannt. Wird auf diesem Rechner mittels eines Programms der Befehlssatz eines anderen Rechners nachgebildet, so spricht man von *Emulation*. Wird der Rechner dagegen nur zur Entwickling der Software für den anderen Rechner verwendet, so spricht man von *Cross-Assemblern* bzw. *Cross-Compilern*, wobei der Cross-Compiler höhere Programmiersprachen, z. B. PL/M, benutzt. Über beide entsteht dann genauso wie beim üblichen Assembler meist in zwei Durchläufen der Objektkode für den Mikrorechner. Er wird dabei in einem ROM fixiert oder auf einem Datenträger abgelegt, der später in den Mikrorechner eingelesen werden kann. Wird die Software auf dem

Großrechner auch getestet, z.B. am Emulationsprogramm, so spricht man von einem *Simulator*. Es ist verständlich, daß für alle diese Prozesse ein Großrechner weitaus mehr Komfort bietet als der Mikrorechner.

8.8.7. Zusammenfassung

Die Programmierung von Mikrorechnern ist ein komplizierter und auf vielfältige Weise realisierbarer Prozeß. Dabei müssen *fünf Phasen* unterschieden werden:

Konzipierung des Programms. Sie sollte heute immer in Top-down erfolgen. Für sie wird je nach Größe des Programms ca. 30 bis 50% der Zeit benötigt. Für sie gibt es keine automatisierbaren Methoden, sondern nur eine hohe Disziplin des Programmierers.

Erzeugung des Maschinenkodes, also die Programmierung im engeren Sinne. Sie kann heute auf vielfältige Weise erfolgen und ist stark durch Rechner unterstützbar. Den grundsätzlichen Aufwand zeigt Bild 8.8.8a), je nach der Länge des Programms und den vorhandenen Hilfsmitteln werden ca. 10 bis 30% der Zeit für diese Phase benötigt.

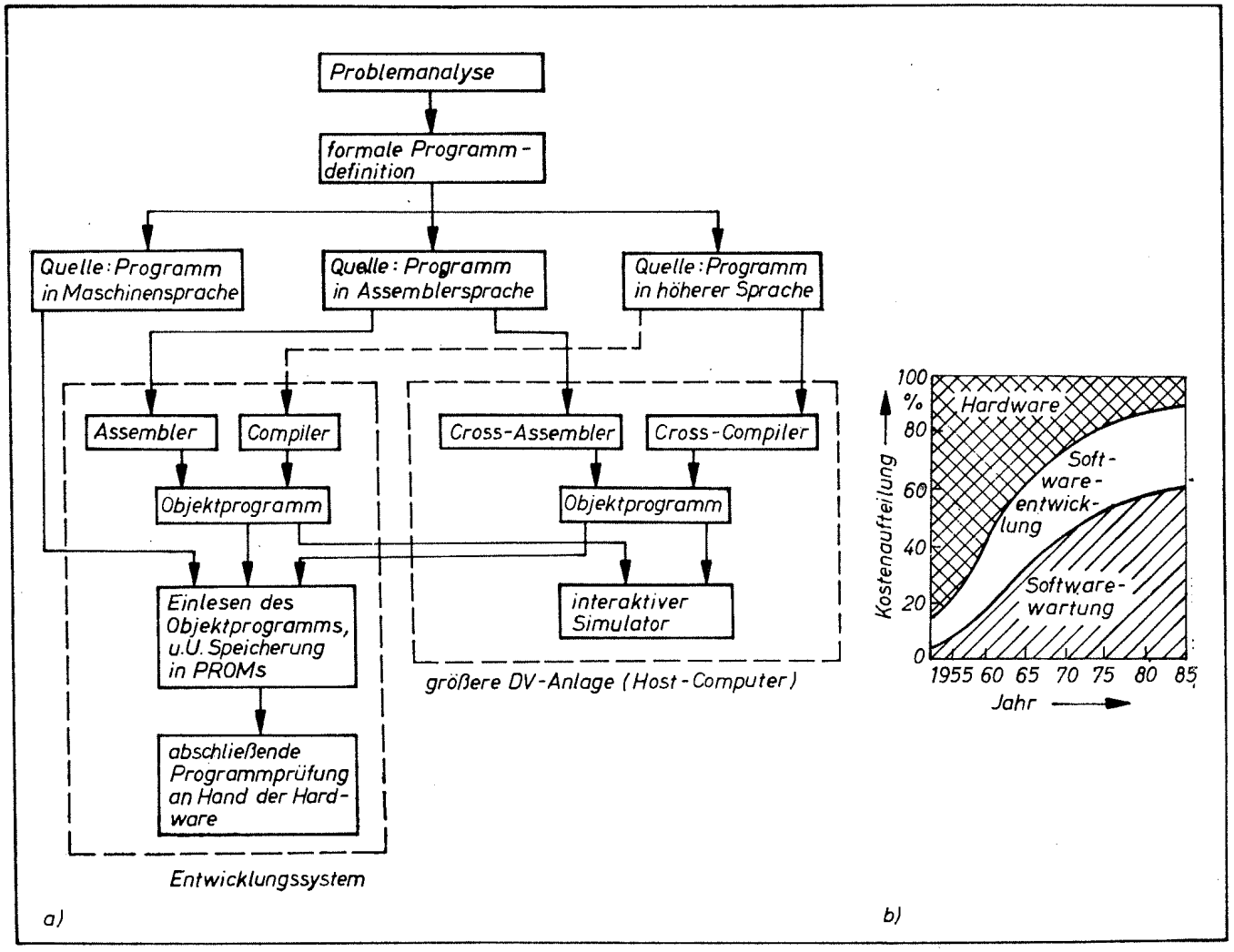


Bild 8.8.8. a) Überblick zu den Methoden der Programmentwicklung in Maschinensprache. Neben diesen Wegen gibt es für höhere Programmiersprachen noch den Interpreter, der nicht zu einem kompletten Maschinenkode führt.
b) Verhältnis von Kosten der Hardware-, der Softwareentwicklung und -wartung im Laufe der Zeit

Austesten der Programme. Hierbei wird das Programm von Fehlern befreit. Dabei treten größere Probleme auf, wenn die erste Phase nicht gründlich genug durchgeführt wurde. Unter normalen Bedingungen wird der Zeitaufwand für diese Phase um 30% geschätzt.

Dokumentation der Programme. Diese Phase wird ebenfalls häufig unterschätzt, der für sie notwendige Aufwand liegt bei ca. 10%. Sie ist vor allem deshalb so wichtig, weil auch andere Nutzer das Programm anwenden und verstehen sollen. Zum anderen vergißt jeder Programmierer recht schnell, wie eine Lösung gestaltet wurde, und schließlich ist die Dokumentation für die letzte Phase unentbehrlich.

Softwarepflege. Programme können oft nicht bezüglich aller Wege und variablen Größen ausgetestet werden. Weiter ändert sich im Lauf der Zeit die Problemstellung. Beides bedeutet, daß an eigentlich fertigen Programmen immer wieder gewisse Änderungen notwendig sind. Im Laufe der Zeit hat sich der hierfür notwendige Aufwand ständig vergrößert (Bild 8.8.8b)). Um 1985 werden in weiterer Detaillierung folgende Anteile erwartet:

3% Anforderungsanalyse		8% Modultest
3% Spezifikation	7% Codierung	7% Integrationstest
5% Entwurf		67% Wartung.

Das Verhältnis des Gehalts eines qualifizierten Mitarbeiters zu den Kosten einer kleinen CPU lag 1960 bei ca. 1:3000, erreichte 1975 etwa 1:1 und dürfte zur Zeit bei 300:1 liegen. Einen ähnlichen Trend weist das Verhältnis der Anzahl der Programmierer zu den vorhandenen Rechenanlagen auf. Um 1955 lag es bei 10:1, erreichte um 1978 etwa 1:1 und dürfte um 1985 bei 1:3 liegen. Um 1985 wird erwartet, daß der Marktumsatz der Mikrorechner den der Großrechner erreicht.

8.9. Speicherschaltungen und Sonderrechner

8.9.1. Systematik digitaler Schaltungen; Anwendungen des Mikroprozessers

Die vorangegangenen Abschnitte haben die große Vielfalt digitaler Schaltungen aufgezeigt. Sie entsteht im Prinzip aus kombinatorischen Schaltungen und Speichern. Diese vielfältigen Zusammenhänge versucht Bild 8.9.1 zu systematisieren. Es beginnt oben mit den Unterschieden der Technologien und Schaltkreisfamilien, um dann zu den drei Hauptgruppen der Schaltungen überzugehen. Sie sind dann wiederum nach verschiedenen Gesichtspunkten weiter untergliedert. Besonders umfangreich sind die Arten von Speicherschaltungen, die einerseits schon, z.B. in der Kodierung, kombinatorische Schaltungen enthalten und zu denen im weiteren Sinne auch die nochmals zusätzlich gegliederten Register und Flipflops hinzugezählt werden müßten. da bei ihnen der Speicheranteil überwiegt.

Eine Einteilung der Anwendung von Mikroprozessoren zeigt Bild 8.9.2 unter zwei Gesichtspunkten: In welchen Gebieten werden sie eingesetzt (Teilbild a)), und welche hergebrachten Schaltstrukturen ersetzen sie zum Teil. Besonders auffällig ist, daß sie eine große Anzahl neuer Produkte bewirken.

8.9.2. Mikroprogrammierung

Die Mikroprogrammierung von Rechnern erfolgt mit speziellen Einrichtungen, die selbst einen Automaten darstellen. Das Grundprinzip geht von Bild 8.9.3a) aus. Es existiert also ein ROM mit M Eingängen und N Ausgängen. Davon werden nur

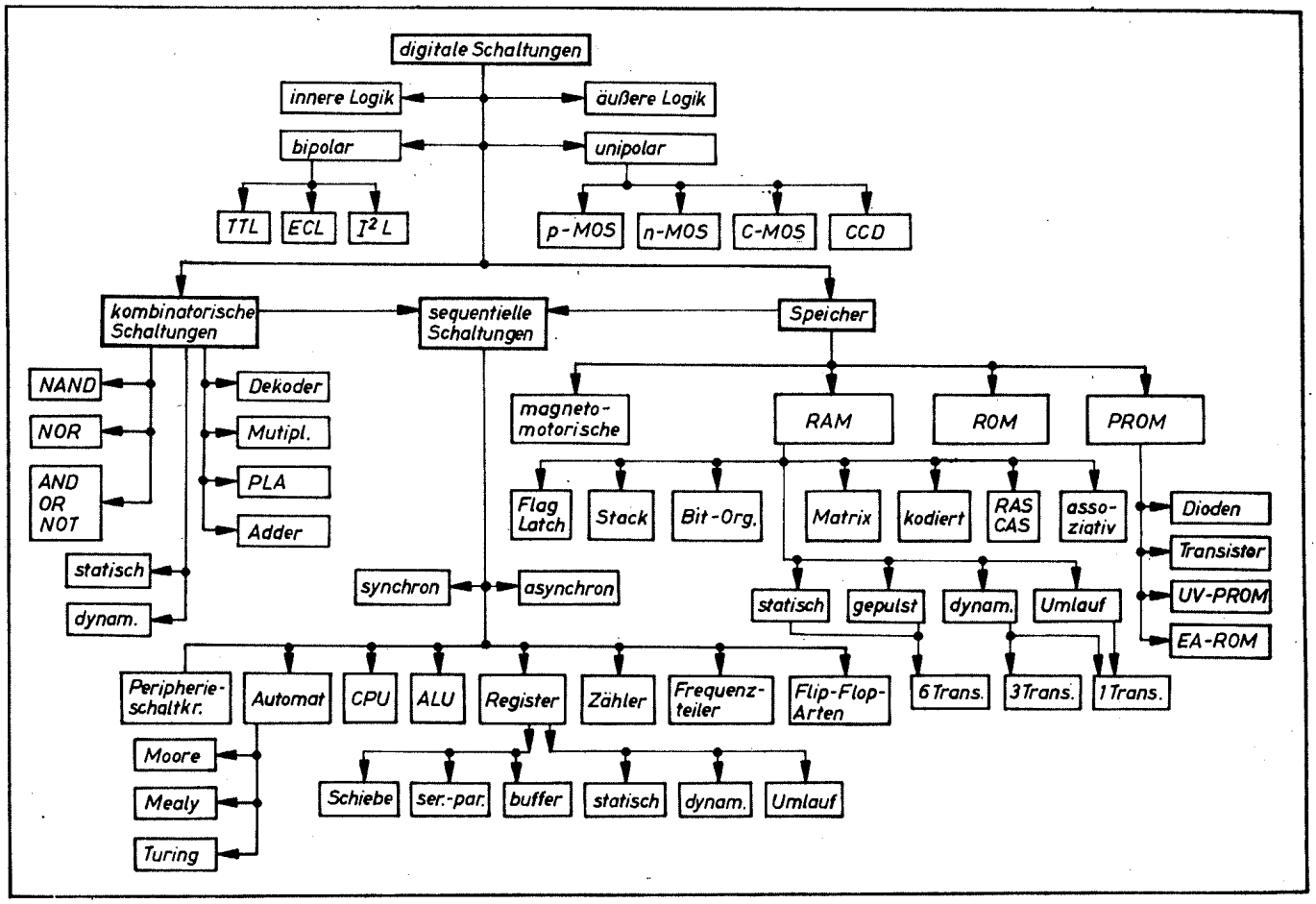


Bild 8.9.1. Versuch einer Systematisierung aller digitalen Schaltungen

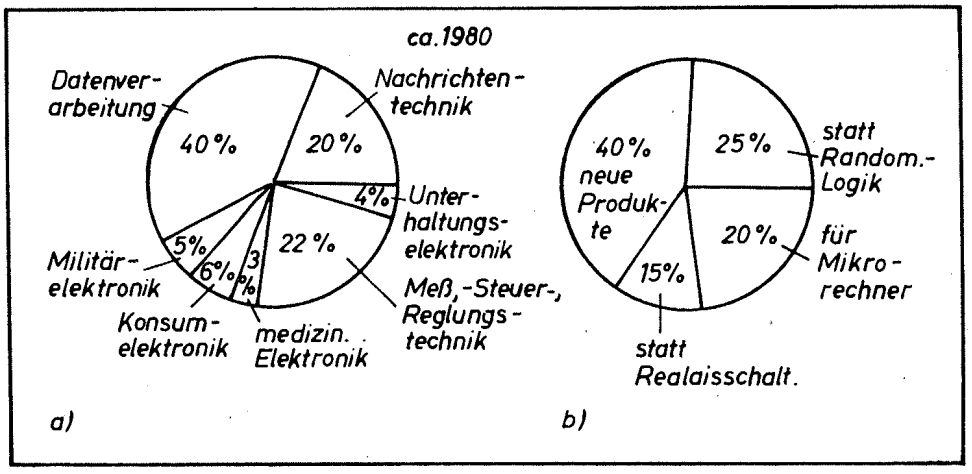


Bild 8.9.2. a) Anwendungsgebiete und b) Verwendungszwecke von Mikroprozessoren

$m < M$ und $n < N$ als eigentliche Ein- bzw. Ausgänge benutzt. Die anderen $N - n = M - m$ werden über ein getaktetes Register rückgekoppelt. Es werde angenommen, daß zu Beginn alle N Ausgänge und die m Eingänge den Wert Ø besitzen. Beim Start wird eine ausgewählte Signalkombination auf die m Eingänge gegeben. Dadurch ändern sich einige Ausgänge, darunter auch welche im Bereich $n + 1$ bis N. Diese Änderung wird vom Register aufgenommen. Wenn jetzt der Taktimpuls kommt, werden die entsprechenden Werte auf die Eingänge $m + 1$ bis M weitergeleitet, und dadurch entstehen neue Ausgangssignale, die z. T. auch wieder auf das Register gelangen. Mit dem Takt wird also eine Signalfolge an den Ausgängen 1 bis n generiert. Sie kann u. a. als Steuerprogramm für einen Rechner verwendet werden. Die Periodizität dieser Folge kann dabei durch die Anzahl der Taktimpulse unterbunden werden.

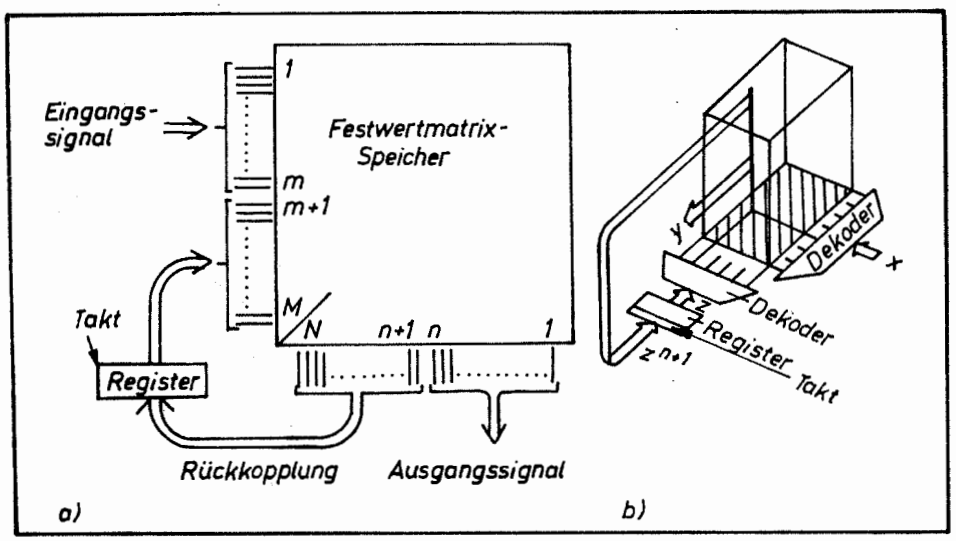


Bild 8.9.3. Prinzip des Mikroprogrammspeichers
a) nach WILKES mit einem ROM realisiert
b) nach GÜNTSCH-HÄNDLER

Wichtig ist weiter, daß für jede Eingangssignalkombination an 1 bis m mit dem Takt eine andere Ausgangssignalfolge, also ein anderes Steuerprogramm, realisiert werden kann. Diese unterschiedlichen Programme heißen Mikroprogramme. Wird statt des ROM ein RAM bzw. PROM verwendet, so sind die Mikroprogramme selbst programmierbar.

Einen anderen Aufbau eines in der Funktion gleichartigen Mikroprogrammspeichers zeigt Bild 8.9.3b). Aus dieser Darstellung wird im Zusammenhang mit Abschnitt 8.4. deutlich, daß dieses Prinzip ein Automat mit den Eingangsgrößen X, den Ausgangsgrößen Y und den inneren Zuständen Z ist. Umgekehrt bedeutet dies, daß jeder Automat auf diese Weise auf einen Speicher mit Rückkopplung zurückführbar ist.

8.9.3. Gegenseitige Ersetzbarkeit binärer Schaltungen

Die Einteilung der digitalen Schaltungen gemäß Bild 8.9.1 ist teilweise nach der geschichtlichen Entwicklung und teilweise nach den Funktionen der Schaltungen vorgenommen. Sie ist aber nicht zwingend. Dies soll am Bild 8.9.4 besprochen werden. Generell ist dabei zu bemerken, daß bei den Schaltungen für die Anwendung immer die Funktion, welche sie ausüben sollen, interessiert. Wird dagegen die Schaltung hergestellt zum Beispiel als integrierter Schaltkreis, so muß letztlich ein Maskensatz hierfür, also eine Struktur, entstehen. Aus der Kybernetik ist aber bekannt, daß eine Funktion durch z. T. sehr unterschiedliche Strukturen realisiert werden kann. Genau um dieses Problem geht es hier.

Es sei von den kombinatorischen Schaltungen ausgegangen. Sie stellen im wesentlichen Verstärkerschaltungen mit mehreren verknüpften Eingängen dar. Die Speicher entstehen hieraus durch Rückkopplung. Dies wurde im Abschnitt 8.3., z. B. Bild 8.3.5, gezeigt. Werden mehrere kombinatorische Schaltungen zusammengefaßt, so entsteht eine formale Schaltung, wie sie die Mitte des Bildes 8.9.4 zeigt. Beispiele hierfür sind Kodewandler (Bild 8.2.2), Dekoder (Bild 8.2.20 und 8.2.21), Multiplexer (Bild 8.2.22), Vergleicher (Bild 8.2.23), Adder (Bild 8.2.24) und ALU (Bild 8.2.25).

Aus dem Speicher läßt sich der Festwertspeicher (Bild 8.3.1) entwickeln. Dabei sind die eingeschriebenen Informationen durch feste Verknüpfungen gegeben. Werden

mehrere Festwertspeicher zusammengeschaltet, so entsteht wieder zunächst formal die Struktur in der Mitte des Bildes 8.9.4. Aber bereits in Bild 8.3.17c) wurde gezeigt, daß hiermit die verschiedensten Funktionen der kombinatorischen Schaltungen realisiert werden können. Eine komplexe Art des Festwertspeichers ist die PLA gemäß Bild 8.2.26. Sie besteht aus zwei in Reihe geschalteten Festwertspeichern, und bereits dort wurde gezeigt, daß sie auf ideale Weise alle Funktionen kombinatorischer Schaltungen erfüllen kann. Damit ist die Relativität bzw. Austauschbarkeit zwischen kombinatorischer Schaltung und Speicherschaltung vollständig gegeben. Somit ist auch verständlich, daß die sequentiellen Schaltungen, die auf beiden aufbauen, je einmal vollständig aus kombinatorischen Schaltungen bzw. Speicherschaltungen aufgebaut werden können.

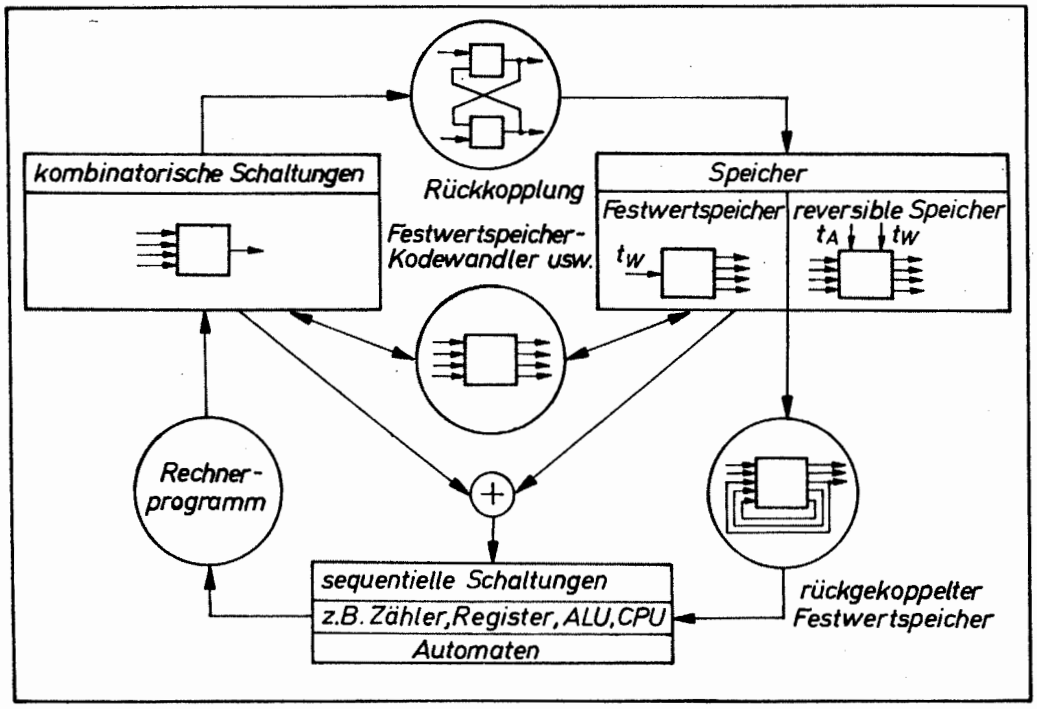


Bild 8.9.4. Zur gegenseitigen Ersetzbarkeit digitaler Schaltungen

Im vorigen Abschnitt wurde gezeigt, daß ein mehrfach rückgekoppelter Festwertspeicher eine Grundlage der Mikroprogramme ist, also einen Automaten darstellt. Nicht weiter zu beweisen ist der Fakt, daß eine sequentielle Schaltung, also ein Mikrorechner, sowohl als Speicher als auch als kombinatorische Schaltung zu verwenden ist. Als Speicher wirkt er, indem er über IN-OUT-Operationen einen Zugriff zu seinem internen Speicher realisiert. Als kombinatorische Schaltung wirkt er, indem er die verschiedenen Eingänge abfragt, die dort vorhandenen Pegel nach den Gesetzen der BOOLEschen Algebra abfragt und das Ergebnis an einen Ausgang legt. Dies erscheint auf den ersten Blick sehr umständlich, ist aber wegen der Universalität dieser Prinzipien und der leichten Änderbarkeit der gewünschten Funktionen gemäß Programm oft vorteilhaft. Schließlich wird damit ergänzend zu den Ausführungen dieses Abschnitts auch die Universalität des Mikroprozessors bzw. seiner Schaltungen gezeigt. Da er auf diese und ähnliche Weise Steuerungen realisiert, wird dann von speicherprogrammierten Steuerungen gesprochen. Dies drückt aus, daß die Steuerung auf einem im Speicher des Mikrorechners stehenden Programm beruht. Bei Steuerungen der klassischen Art, die auf speziell ausgelegten kombinatorischen (eventuell sequentiellen) Schaltungen beruhen, wird dagegen von verbindungsprogrammierter Steuerung oder allgemeiner von Random-Logik gesprochen.

8.9.4. Kundenwunschschaltungen

Soll eine komplexe digitale Schaltung aufgebaut werden, so können heute drei Wege unterschieden werden:

- Aufbau aus diskreten Bauelementen und Standardschaltkreisen auf einer *Leiterplatte*,
- Aufbau als *Mikrorechnervariante*,
- Entwicklung eines *Spezialschaltkreises* (Kundenwunschschaltung).

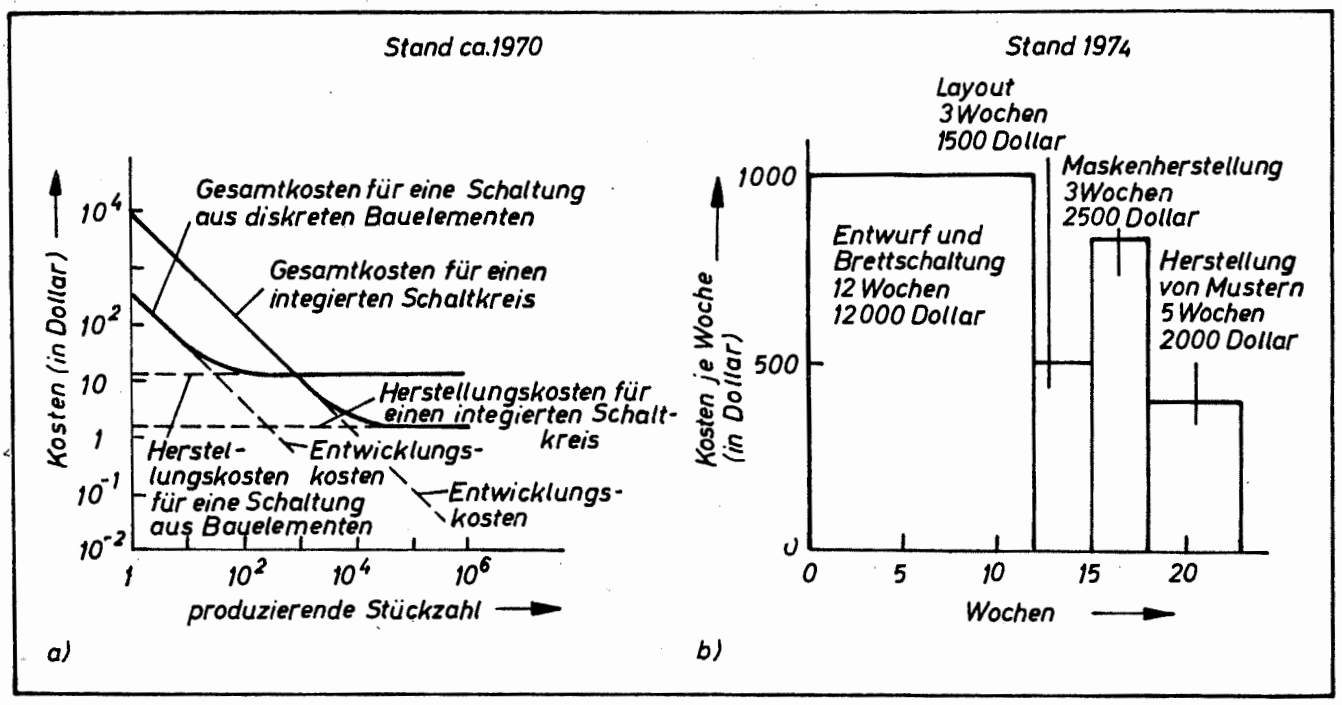


Bild 8.9.5. a) Vergleich der Kosten beim Aufbau einer Schaltung aus Bauelementen bzw. für einen komplexen integrierten Schaltkreis bei gleicher Aufgabe
b) Entwicklungsablauf und Kosten für den Entwurf eines integrierten Schaltkreises

Einen wenn auch relativ alten Vergleich zwischen einem Entwurf mit Standardbauelementen und einem integrierten Schaltkreis zeigt Bild 8.9.5a). Beim Schaltkreis ist stets der Grundaufwand, d. h. die Entwicklungskosten, höher. Die Herstellung des einen Schaltkreises ist aber billiger als der Aufbau der Schaltung und die Kosten der Standardbauelemente. Aus diesem Grund ist der IC für große Serien und die Leiterplattenlösung für kleine Serien vorteilhaft.

Den zeitlichen Entwicklungsablauf für einen integrierten Schaltkreis zeigt Bild 8.9.5b). Da die Schaltungen inzwischen viel komplexer geworden sind, liegen heute die Kosten höher, und die Zeiten sind länger. Hierzu gibt Tab. 8.9.1 einen Überblick, während Bild 8.9.7b) auf S. 933 den Anstieg des Entwicklungsaufwandes ausweist.

In diesem Abschnitt sollen nur Prinzipien für den Fall des Spezialschaltkreises betrachtet werden. Es existieren wiederum drei Prinzipien:

- *Maskenprogrammierung* bei einem relativ universellen Schaltkreis,
- *Zellenentwurf*,
- voll *optimierter Schaltungsentwurf*.

Die *Maskenprogrammierung* wurde bereits im Abschnitt 8.2.8.7. behandelt. Sie beruht allgemein darauf, daß ein Grundraster von Bauelementen bzw. Grund-

Tabelle 8.9.1. Zeitablauf der Entwicklung eines optimierten Kundenschaltkreises um 1980

Entwicklungsetappe	Entwicklungszeit Wochen
Entwicklung der Spezifikationen und Vertragsabschluß	4
Logischer Entwurf	6
Schaltungsanordnung	3
Vorgabe der Parameter und der Modellierung	1
Bestätigung der Parameter	3
Rechnergestützte Modellierung	1
Erarbeitung der Topologie	8
Bestätigung der Topologie	2
Maskenherstellung	4
Herstellung einer Kontrollcharge	6
Beurteilung der Kontrollwafer	6
Korrektur der Masken	6
Herstellung der ersten Muster	8
Verpackung	2
Nachprüfung der Spezifikationen	5
Lebensdauertest	8
Gesamt	73

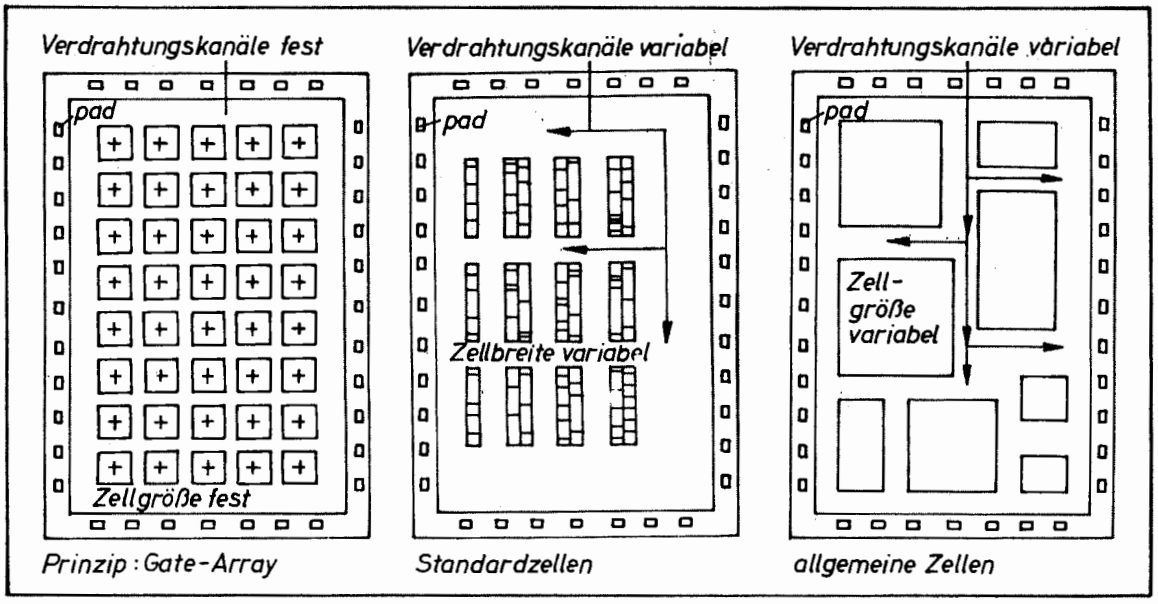


Bild 8.9.6. Vergleich eines Grob-Lay-outs bei den drei semispezifischen Kundenwunschschaltungen

schaltungen vorliegt, welches durch eine einzige spezielle Maske für die konkrete Anwendung „verdrahtet" wird (Bild 8.9.6a)).

Für den *Zellenentwurf* sind zwei Varianten zu unterscheiden: Standardzellen und allgemeine Zellen.

Standardzellen sind berechnete, optimierte und topologisch ausgelegte Strukturen. Ihre Schaltung, Funktion, Topologie und technischen Daten sind als Katalog vorhanden. Die Funktionen betreffen z.B. Inverter, Ausgangsverstärker, verschiedene Gatter, Flipflops, Schieberegister und Volladder. Die Zellen haben im wesentlichen gleiche Höhe und genormte Anschlüsse für die Versorgungsspannungen. Solche Zellen lassen sich also leicht aneinanderreihen. Es entfallen dann allerdings 60% der Kristallfläche auf die Verdrahtung und nur 40% auf die Zellen, also noch weniger auf die eigentlich aktiven Strukturen (Teilbild 8.9.6b). Die Methode hat aber zwei wesent-

liche Vorteile: Sie erleichtert den Übergang zu komplexen Strukturen und ist besonders gut für den rechnergestützten Entwurf geeignet.

Bei den *allgemeinen Zellen* wird gemäß Bild 8.9.6c) auch noch das Zellengrundraster verlassen. Hierdurch können die verschiedensten erprobten Schaltungen aneinandergefügt werden. Dadurch ist aber ein bereits komplizierter und damit teurer Entwurf erforderlich. Dieses Prinzip hat vor allem bei komplexeren Schaltungen, die ohnehin einen höheren Entwurfsaufwand erfordern, Bedeutung.

Beim *optimierten Entwurf* braucht auf keine Randbedingungen Rücksicht genommen zu werden. Er führt daher zur höchsten Packungsdichte, zum kleinsten Leitungsverbrauch, zur bestangepaßten Schaltzeit usw., erfordert dafür aber den höchsten Entwurfsaufwand.

Tabelle 8.9.2. Vergleich von Entwurfprinzipien für Spezialschaltkreise

	Gate-Array	Standard-zellen	Allgemeine Zellen	Optimierter Entwurf
Chipgröße	fest	variabel	variabel	variabel
Anzahl d. Zellen	fest	variabel	variabel	variabel
Bondinseln	fest	variabel	variabel	variabel
Zellenhöhe	fest	fest	variabel	nicht vorhanden
Zellenbreite	fest	variabel	variabel	nicht vorhanden
Verdrahtungswege	fest	z. T. fest	variabel	beliebig

Einen *Vergleich der vier Methoden* bezüglich einiger Parameter zeigt Tab. 8.9.2. Daraus geht deutlich die wachsende Freiheit in der Strukturierung von Gate-Array über Standardzellen, allgemeine Zellen bis zum optimierten Entwurf hervor. Diese wachsende Freiheit bedeutet immer Vorteile bezüglich der Optimierung, aber dafür auch erhöhten Aufwand. Damit wird deutlich, daß es für jede Variante einen optimalen Stückzahlbereich gibt. Es sei denn, daß aus anderen Gründen (nicht ökonomischen) nur eine bestimmte Lösung in Betracht kommt. Ein Beispiel hierfür ist z.B. eine elektronische Schaltung für einen Fotoapparat oder eine Armbanduhr. Sie muß vor allem klein sein und energiearm arbeiten. Deshalb kommt hier nur eine entsprechend optimierte Schaltung in Betracht, und die erforderliche Stückzahl (Ökonomie) steht erst an zweiter Stelle.

8.9.5. Vergleich von Anwendungen

Ein Vergleich und eine Optimierung zwischen den verschiedenen Methoden kann nach unterschiedlichen Gesichtspunkten vorgenommen werden. Beispiele dazu zeigt Bild 8.9.7.

Teilbild a) gibt den gemittelten Aufwand nach verschiedenen Teilarbeiten an. Hieraus kann z.B. abgeschätzt werden, welcher Mitarbeiterstab für eine Lösung sinnvoll ist.

In Teilbild c) sind die Gesamtkosten als Funktion einer bestimmten Anzahl zu fertigender Schaltungen aufgetragen. Dabei gibt es vier typische Gesichtspunkte, bei denen der Übergang zu einer anderen Technologie sinnvoll ist. Die dafür notwendigen Stückzahlen hängen stark von der Zeit, von den jeweiligen Gegebenheiten und vom Schaltkreis selbst (z.B. Komplexität) sowie vom Anwendungsgebiet ab.

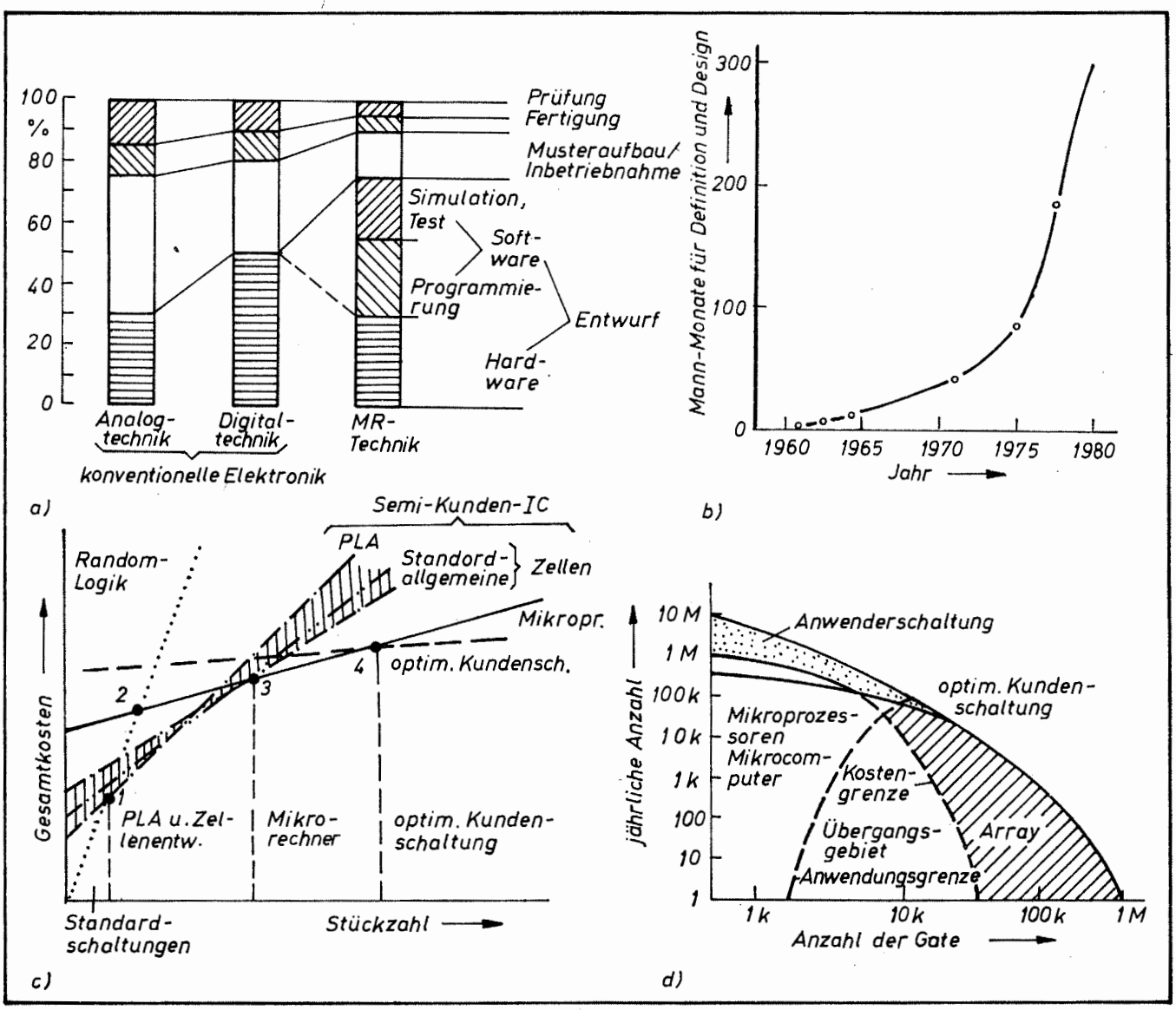


Bild 8.9.7. Zum Vergleich von Entwurfsmethoden

a) Konventionelle Technik (bei digitaler Random-Logik)

b) Anstieg des Entwicklungsaufwandes für digitale Schaltungen. Parallel damit, aber bei weitem nicht so steil, steigt die Komplexität

c) Zusammenhang von Entwicklungs- und Fertigungsaufwand für Systeme bei unterschiedlicher Stückzahl. Alle Kurven beginnen mit dem Grundaufwand, der nur einmal zu leisten ist, und steigen mit den Kosten, die je Stück zusätzlich aufzubringen sind. Die Übergangspunkte 1 bis 4 sind im Text erklärt

d) Zusammenhang zwischen Komplexität und jährlich zu produzierender Stückzahl sowie sich daraus ergebende optimale Gebiete

Die groben Richtwerte von 1980 lagen bei etwa:

1 500 bis 3000 Stück
2 1000 bis 5000 Stück (50 Gatter)
3 3000 bis 100000 Stück
4 50000 bis 300000 Stück.

Teilbild d) weist schließlich einen Zusammenhang zwischen der Komplexität einer Schaltung und der zu fertigenden Stückzahl aus. Die hier gezeigten Grenzen sind in der Regel noch unschärfer als bei Bild 8.9.7c). Weitere Merkmale zur Abgrenzung, die noch stärker qualitative Kennzeichen berücksichtigen, enthält Tab. 8.9.3.

Tabelle 8.9.3. Merkmale des wirtschaftlichsten Einsatzes festverdrahteter und programmierbarer Schaltkreise

Standard-schaltkreis	PLA Zellenentwurf	ROM, PROM	Mikrorechner-anwendung	Kundenwunsch-schaltkreis
— Kleine Stückzahl und/oder — Unterkritischer Platz- oder Leistungsverbrauch — Geringe Komplexität der Informationsverarbeitung (viel Logik, viel Arithmetik) — Keine Änderung der Aufgabenstellung — Geringe Softwarekenntnis, aber große Erfahrung in Schaltungsentwicklung — Kein Optimum an Verarbeitungsgeschwindigkeit — Mittelmäßige Forderung an Entwicklungszeit	— kleine Stückzahl — rasch realisiert — Prüfung einfach — schlechte Flächennutzung — lange Verbindungsleitungen — mittlere Komplexität	— mittlere Stückzahl — keine Änderung der Aufgabenstellung — geringere Entwicklungszeit — geringer Verdrahtungsaufwand — fast nur Programmentwicklung erforderlich — Platzbedarf größer — nicht laufzeitoptimiert	— beliebige Stückzahl und/oder — große Komplexität (viel Logik, viel Arithmetik) — Änderung der Aufgabenstellung und Eingangsdaten zu erwarten — große Anforderung an Speicherfähigkeit — höhere Anfangsinvestition — geringe bis mittelmäßige Verarbeitungsgeschwindigkeit — kurze Entwicklungszeit	— große Stückzahl und/oder — geringster Platz- oder Leistungsverbrauch — keine Änderung der Aufgabenstellung — Optimum an Verarbeitungsgeschwindigkeit — keine spezifischen Anforderungen an Entwicklungszeit
Zunahme der Programmierbarkeit →				— feste Aufgabe

8.9.6. Taschenrechner

Die Taschenrechner gehen — ähnlich wie die Mikroprozessoren — auf Entwicklungen in Japan zurück. Der erste Taschenrechner wurde von CANNON 1970 produziert. Erst auf dieser Basis, insbesondere durch den persönlichen Eindruck von BILL HEWLETT auf einer Japanreise, erfolgte in den USA bei Hewlett-Packard die Entwicklung des ersten wissenschaftlich-technischen Taschenrechners HP-35, der ab 1972 produziert wurde. 1974 folgte der erste voll programmierbare Taschenrechner HP-65. Heute gibt es eine Fülle von preiswerten Taschenrechnern verschiedener Preis- und Leistungsklassen. Sie reichen bis zum in BASIC programmierbaren Taschencomputer mit einer Vielzahl anschließbarer Peripheriegeräte. Die zu diesem Taschenrechner gehörenden CPU und Speicher sind häufig in CMOS-Technik aufgebaut, so daß Daten und Programme nahezu ohne Stromverbrauch, auch bei sonst abgeschaltetem Rechner, erhalten bleiben.

Die in diesen Rechnern verwendeten Schaltkreise sind stark optimierte Kundenwunschschaltkreise, welche durch die meist sehr großen Stückzahlen sehr preiswert produziert werden können.

Im Vergleich zu den bisher behandelten universellen Mikroprozessoren zeichnet sich die CPU der Taschenrechner durch folgende Besonderheiten aus:

Taschenrechner sind fast ausschließlich für *numerische Berechnungen* ausgelegt.

Die Verarbeitung erfolgt meist in *Bitbreite* im BCD-Format.

Die eigentliche *Wortbreite* ist relativ groß, z.B. 8 × 4 bit oder 12 × 4-bit-Mantisse plus 2 × 4-bit-Exponent.

Es existieren nur *wenige Register* für die Datenspeicherung.

Die *ALU* ist auf Addition und Subtraktion, eventuell noch auf logische Entscheidungen spezialisiert.

Alle realisierten Funktionen sind in einem speziellen ROM gespeichert und werden *sequentiell abgearbeitet.*

Die Bedienung erfolgt nur über *Tastensteuerung.*

Die *programmierbaren Taschenrechner* besitzen ein relativ kleines RAM, sind also nur in gewissem Umfang programmierbar. Dennoch können hochwertige Taschenrechner praktisch alle numerischen Probleme lösen.

Die *Ablaufsteuerung* eines Taschenrechners erfolgt im Sinne eines Interpreters für die Tastenbefehle bzw. für die diesbezüglich gespeicherten Werte beim programmierbaren Rechner. Dies hat zur Folge, daß zwar die einzelnen Mikroprogramme, z.B. für die Grundrechenarten und die Standardfunktionen, noch relativ schnell ablaufen, dagegen die Sprungbefehle usw. sehr viel Zeit beanspruchen.

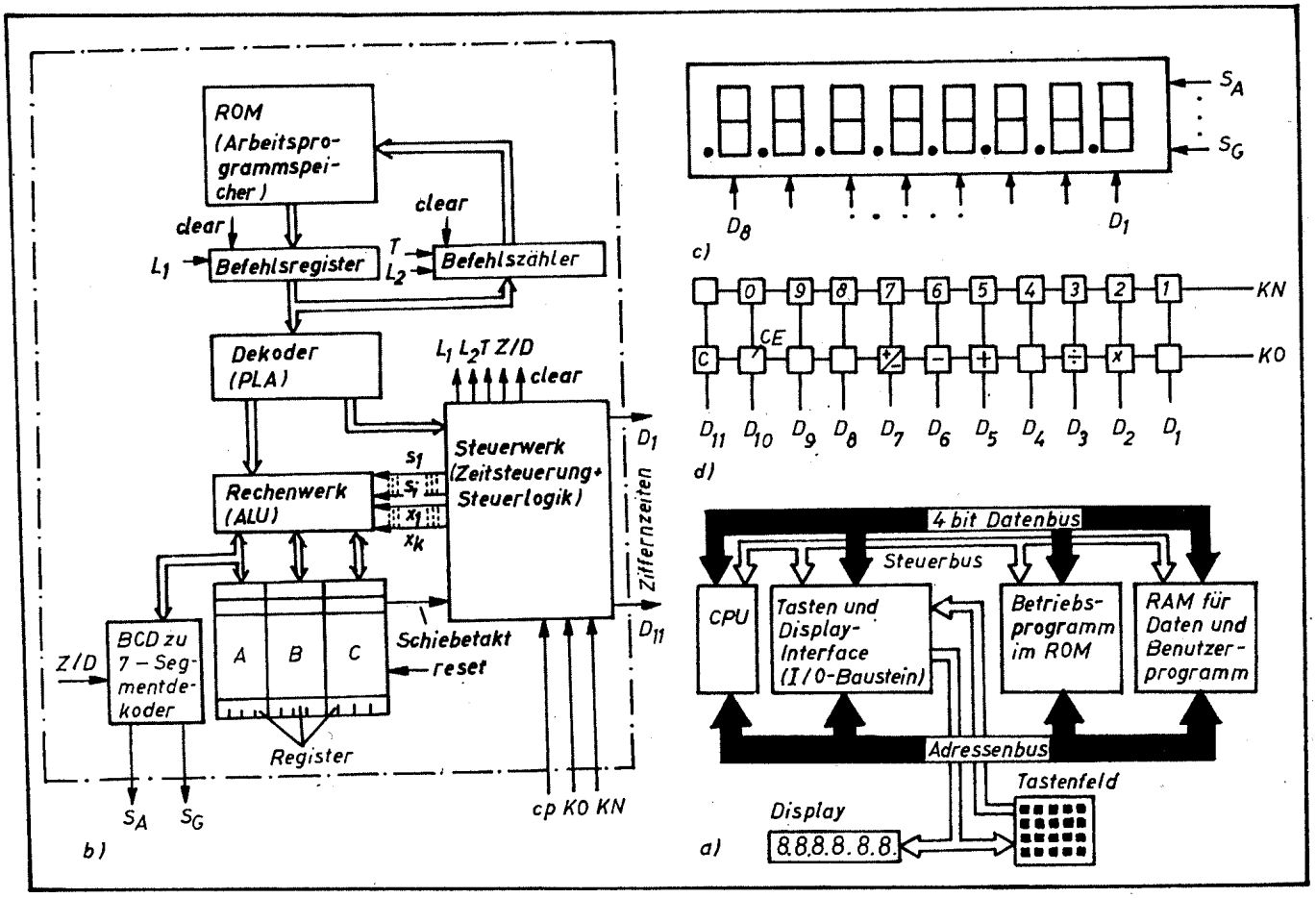


Bild 8.9.8. Zum Aufbau von Taschenrechnern

a) Grundprinzip

b) Teilschaltungen bei einem Rechner für die vier Grundrechenarten

c) Anschluß des Displays und d) der Tastatur

Beim Taschenrechner besteht kein Zugriff zur Maschinensprache der CPU.

Den prinzipiellen Aufbau eines Taschenrechners zeigt Bild 8.9.8a). Display und Tastenfeld sind die eigentlichen peripheren Einheiten. Nur in Sonderfällen existiert ein weiterer peripherer Zugriff. Den Aufbau eines einfachen Taschenrechners für die vier Grundrechnungsarten zeigt Teilbild b) und den Anschluß der Tastatur und des Displays c) und d). Typisch für alle Taschenrechner ist, daß gemeinsame Leitungen (hier D_1 bis D_8) für Display und Tastatur verwendet werden.

Im ROM befinden sich ca. 200 bis 500 Befehlsworte mit einer Breite um 12 bit. Das ROM wird vom Befehlszähler adressiert und gibt dann die Befehle ins Befehlsregister. Über eine PLA werden sie dekodiert und steuern die ALU und das Steuerwerk. Das Steuerwerk dekodiert darüber hinaus die Tasteneingaben und bewirkt über L_1 und L_2 die Steuerung des Befehlsregisters bzw. Befehlszählers. Außerdem gibt es über D_1 bis D_8 und S_A bis S_G den Inhalt des Akkumulators im Zeitmultiplex an das Display. Die Steuersignale S_1 bis S_i steuern die ALU und die Signale X_1 bis X_k die Kommunikation zwischen den drei Registern A, B und C sowie der ALU.

8.9.7. Spezialrechner

8.9.7.1. Problemstellung

Bei vielen mathematischen Problemen ist die Anzahl der Variablen n wichtig. Bei einigen, wie Matrizen, Determinanten, Gleichungssystemen, Differentialgleichungen, steigt die Rechenzeit etwa proportional n^2 bzw. n^3.

Bei anderen Problemen, wie Leiterplattenentwurf, Schnittstellenproblemen, Stundenplanerstellung usw., wächst die Rechenzeit dagegen mit etwa 2^n. Zwischen diesen Problemklassen besteht ein wesentlicher Unterschied. Dies sei am folgenden Beispiel gezeigt:

n	10	20	30	40	50	100	1000
n^2	100 µs	400 µs	1 ms	1,6 ms	2,5 ms	10 ms	1 s
n^3	1 ms	8 ms	27 ms	64 ms	125 ms	1 s	17 min
2^n	1 ms	1 s	18 min	13 d	35 a	10^{11} a	—

Dabei wurde die Dauer für einen Schritt $n = 1$ mit einer Mikrosekunde angesetzt. Es fällt auf, daß in diesem Beispiel für Probleme mit n^2 die Rechengrenzen bei $n \approx 10000$ liegen, für Probleme mit n^3 erfolgt bereits ein steilerer Übergang bei ca. $n \approx 1000$, während für 2^n sich zwischen 30 und 40 ganz steil eine Grenze auftut, die nicht zu überschreiten ist. Hieraus ergab sich immer wieder die Forderung, die Rechnerleistung zu steigern, und in der Vergangenheit wurden gewaltige Fortschritte erreicht. Seit etwa Ende der sechziger Jahre aber war der Fortschritt zu gering. Die Grenze guter Rechenanlagen liegt seitdem bei ca. 10 Millionen Operationen je Sekunde. Deshalb wird seit dieser Zeit ständig nach völlig anderen Methoden der Rechentechnik gesucht. Hierbei existieren auch schon eine Vielzahl von Lösungsansätzen für Spezialfälle, ohne daß sich ein Prinzip durchgesetzt hätte.

So existieren spezielle schnelle Prozessoren, u. a. für die schnelle (fast) FOURIER-Transformation (FFT), für Matrizenoperationen und Bildverarbeitungsalgorithmen. Das Grundprinzip schneller Arithmetikprozessoren mit hoher Parallelität (zur Reduzierung von n) wurde bereits im Zusammenhang mit Bild 8.6.33d) und e) beschrieben. Alle diese Lösungen beruhen also darauf, daß die Eigenschaften spezieller

Probleme ausgenutzt werden. Eine Beschleunigung im klassischen Sinne wird noch durch Rechner mit Bauelementen der Tieftemperaturphysik erwartet (ca. 100 Millionen Op/s). Neben diesen Prinzipien haben drei relativ allgemeingültige Prinzipien eine größere Bedeutung erlangt:

Pipelining (Fließbandarbeit)
Feldrechner
Baumstrukturen.

Sie seien im folgenden kurz erklärt.

8.9.7.2. Pipelining

Viele Prozesse bei Rechnern folgen zeitlich nacheinander zu verschiedenen Aufgaben. Ein Beispiel dieser Art ist der Speicherzugriff. Im Zusammenhang mit der DMA wurde dabei bezüglich Bild 8.6.32b) gezeigt, daß die CPU nur zu bestimmten Zeiten zu den Speichern zugreifen kann bzw. muß. In den Zwischenzeiten ist es daher der DMA möglich, zu den Speichern zuzugreifen. So kann ohne Zeitverlust für die CPU — sozusagen im Schatten der CPU — die DMA zusätzlich Speicheroperationen realisieren. Dieses Prinzip ist dem Pipelining bereits sehr nahe. Das Prinzip einer Befehlsausführung sei hierfür als Beispiel gewählt. Dieser Vorgang kann in vier unabhängige Teiletappen eingeteilt werden (Bild 8.9.9a)):

IF::	instruction fetch	= Befehl aus Speicher holen
ID:	instruction decoding	= Dekodieren des Befehls
OF:	operand fetch	= Holen des Operanden
EX:	execution	= Ausführen des Befehls.

In der Regel werden die vier Etappen nacheinander und ohne Überlappung von Befehlszähler, Befehlsdekoder und Steuereinheit ausgeführt (Bild 8.9.9b)). Die Struktur dieser Einheiten wird nun so ausgeführt, daß immer die vier Prozesse in voll unabhängigen Einheiten erfolgen können. Dann besteht die Möglichkeit, daß die vier Prozesse voll gleichzeitig, jedoch für vier aufeinanderfolgende Befehle ausgeführt werden. Dadurch verkürzt sich die Ausführungszeit eines Befehls im Mittel auf ein Viertel. Wie am Fließband werden die Informationsinhalte von Einheit zu Einheit weitergeschoben.

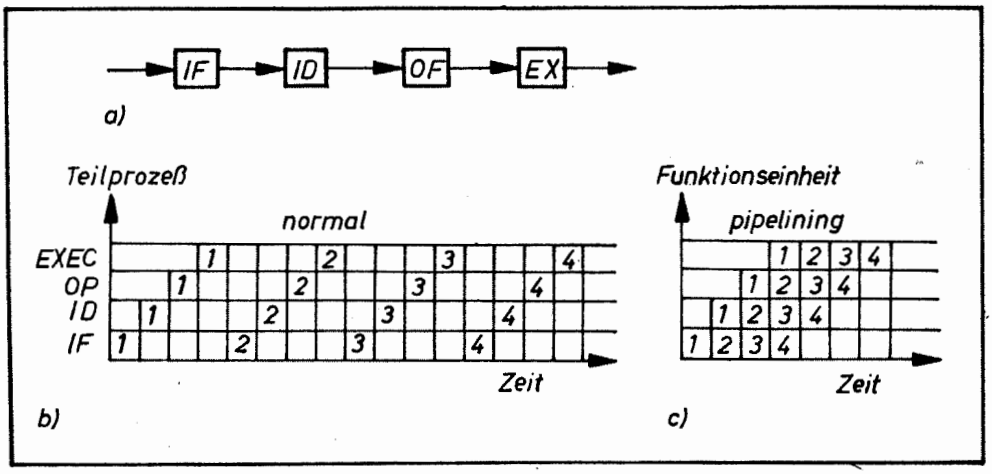


Bild 8.9.9. Prinzip des Pipelining am Beispiel der Abarbeitung eines Befehls

a) Phasen IF = instruction fetch; ID = decoding; OF = operand fetch; EX = execute

b) Normale Abarbeitung

c) Bei Pipelinestruktur, wo für jeden Teil eine besondere Hardwarestruktur existiert

Nach diesem Prinzip lassen sich verschiedene Prozesse verkürzen. Die Struktur des Rechnerteils muß dann eben nur speziell für diese Problematik ausgelegt sein. Dabei werden Faktoren der Geschwindigkeit zwischen 2 und 10 realisiert. Eine spezielle Art der Pipelinerechner sind die Vektorrechner. Hierbei existieren mehrere voneinander unabhängige, aber miteinander koppelbare Pipelines. Mit dem Rechner CYBER 205 werden so Grenzleistungen bis zu 400 Millionen 64-bit-Gleitkommaoperationen je Sekunde erreicht.

8.9.7.3. Feldrechner

Alle bisher behandelten Rechner arbeiten das Problem betont seriell ab, wenngleich auch, wie beim Pipelining oder beim Arithmetikprozessor, Strukturen teilweise parallelisiert sind. Es gibt aber auch seit geraumer Zeit stark betont parallele Verarbeitung. Sie wird in der Regel dadurch erreicht, daß ein Rechner mehrere Rechenwerke oder gar Zentraleinheiten besitzt. Solche Rechner heißen dann *Mehrprozessorenanlagen*. Zuweilen wird auch der Begriff Feldrechner, Array-Prozessor, Parallelprozessor, Zellrechner, Zellularautomat, Schachbrettmusterautomat oder Tessellation-Automat verwendet. Den Prinzipaufbau derartiger Automaten zeigt Bild 8.9.10a). Aufgrund ihrer Matrixstruktur sind diese Automaten besonders für folgende Anwendungen geeignet:

— partielle Differentialgleichungen
— Matrixoperationen
— Monte-Carlo-Methoden
— Mustererkennung.

Dabei läuft die Information, wie es z. B. Bild 8.9.10b) zeigt, vom linken oberen Rand durch die Struktur nach rechts und unten. Eine triviale Zelle, die hierfür geeignet ist,

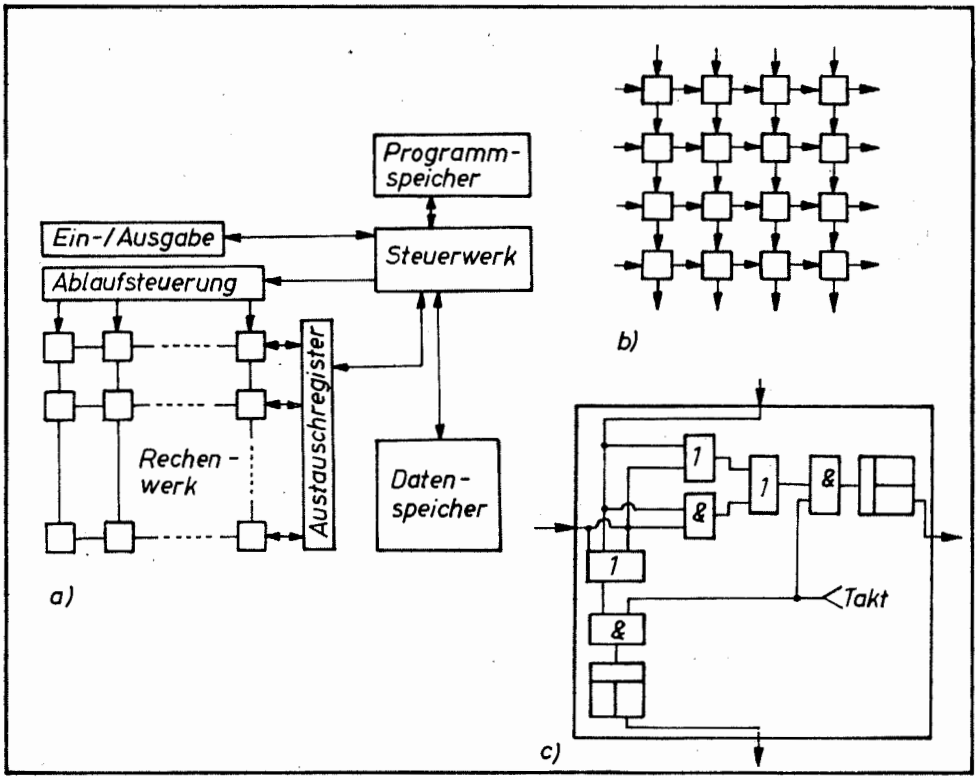


Bild 8.9.10. Zum Feldrechner

a) Prinzipieller Aufbau eines Rechners mit mehreren verkoppelten oder unter sich gleichen und einheitlich gesteuerten Rechenwerken

b) Prinzip des möglichen Informationsflusses zwischen den Rechenwerken

c) Beispiel einer möglichen, aber noch sehr einfachen Rechenzelle

zeigt Bild 8.9.10c). Eine gewisse Analogie besteht hierbei zu homogenen Strukturen bzw. zu den Nervennetzen. Ein Hauptproblem alle dieser Strukturen ist ihre Programmierung. Früher setzte die Ökonomie diesen Prinzipien schnell Grenzen. Durch die Mikroelektronik hat dieser Punkt kaum noch Bedeutung.

Die Idee dieser Automaten geht bereits auf ZUSE zurück, der von den oben genannten Anwendungen ausging. Seit Anfang der 70er Jahre arbeitet ILLIAC IV nach diesem Prinzip. Sie besitzt $16 \times 16 = 256$ Rechenwerke. Mit ihr wurde bereits damals eine Geschwindigkeit von 100 Millionen Op/s erreicht. Inzwischen sind einige weitere Rechner dieses Prinzips gebaut worden, wobei fast 10^9 Op/s erreicht werden. Ein Durchbruch für diese Technik konnte noch nicht erreicht werden.

8.9.7.4. Baumstruktur

Die Matrixstruktur ist nur eine von vielen Strukturen, die zu hoher Parallelisierung führen und damit die Geschwindigkeit erhöhen. Theoretisch sind heute viele Strukturen untersucht, einige sind auch erprobt. Die Struktur, die neben der Matrix die größte Bedeutung zu haben scheint, ist die Baumstruktur. Ihr Prinzip zeigt Bild 8.9.11. Sie hat z. B. große Vorteile beim Sortieren und bei ähnlichen Operationen. Die Eingabe und Ausgabe ist seriell, aber durch die Auffächerung wird von Stufe zu Stufe immer mehr Zeit zum Verarbeiten gewonnen. Gegen Ende werden die Informationen mit geringer werdender Bearbeitung wieder ineinandergeschachtelt. Den Teilaufbau eines solchen Baumes auf dem Chip zeigt beispielhaft Bild 8.9.11b). Es ist aber auch eine rückläufige Struktur gemäß Bild 8.9.11c) möglich.

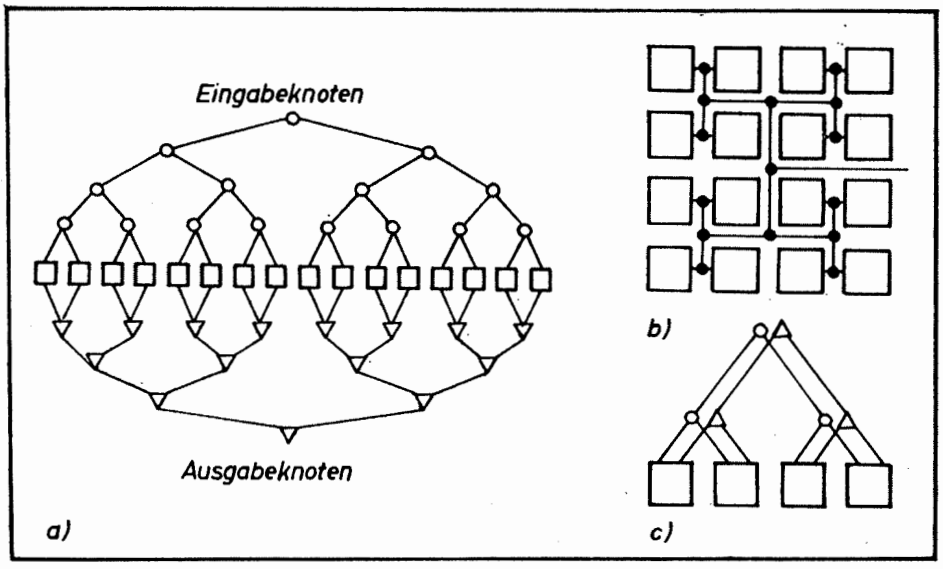


Bild 8.9.11. Zur Baumstruktur eines Rechners

a) Prinzip
b) Teilstruktur auf einem Chip
c) Gespiegelter Baum

8.9.7.5. Parallelität auf höherer Ebene

Eine vom bisher Geschilderten abweichende Variante entsteht dann, wenn ein Hauptrechner jede anfallende Aufgabe automatisch in kleinere Aufgaben zergliedert, die in getrennten und z. T. spezialisierten Rechnereinheiten verarbeitet werden. Hier wird also eine Parallelität nicht durch gleichartige, verkoppelte Rechnerstruktur, sondern durch spezialisierte Arbeitsteilung erreicht. Hierbei konnten in Sonderfällen sehr hohe Leistungen erreicht werden.

9. Spezielle Gebiete

Die Schaltungstechnik behandelte in den Kapiteln 7. bzw. 8. die wichtigsten analogen bzw. digitalen Prinzipien insbesondere auf der Basis von integrierten Schaltkreisen der monolithischen Technologien. Damit sind alle jene Prinzipien nicht behandelt, die an den Grenzen dieser Gebiete liegen. Dies gilt also insbesondere den Analog-Digital- bzw. Digital-Analog-Wandlern und den Hybridtechnologien, auch Dick- und Dünnschichttechnologie. Sie werden in diesem Kapitel ergänzt.

Ferner werden die daraus hervorgehenden Techniken der Digitalfilter bzw. des Analogprozessors sowie Gebiete des digitalen Messens behandelt. Weiter werden hier noch zwei Gebiete erfaßt, die von übergeordneter Bedeutung für alle Kapitel dieses Buches sind: Probleme der Zuverlässigkeit, Betriebssicherheit und ähnliches sowie die Informationstheorie.

9.1. Digital-Analog-Wandler

Von den Digital-Analog-Wandlern, auch Digital-Analog-Umsetzer — DAU oder DAC (= Converter) sowie Entschlüßler — und den Analog-Digital-Wandlern (ADU, ADC, Verschlüßler) ist der DAU die einfachste Schaltung. Sie wird in der Literatur meist nebenbei behandelt. Das Hauptaugenmerk gilt den ADU. Sie werden im Abschnitt 9.2. behandelt. Im Sinne von Abschnitt 8.1.1. erzeugt ein DAU keine analogen Ausgangssignale, sondern diskret gestufte. Sie können aber wie analoge Signale weiter verwendet werden. Hier werden bei den ADU drei Varianten erfaßt:

- Schalterprinzip
- Kondensatorentladung
- Impulsdauerprinzip.

Daneben hat noch ein Prinzip Bedeutung, das ausführlicher bei den ADU (9.2.) behandelt wird. Es beruht darauf, daß ein ADU verwendet wird und — in einen Gegenkopplungszweig gelegt — so wirkt, daß in der Gesamtheit ein DAU vorliegt.

Bei den drei oben genannten Verfahren wird im folgenden bezüglich der digitalen Signale jene Art gewählt, die sich für den DAU am einfachsten verarbeiten läßt. Dies bedeutet u. a. ein paralleles Anliegen aller Signale, seriell kann zuerst das MSB (most significant bit) oder das LSB (least SB) verlangt werden. Wenn primär andere digitale Signale vorliegen, werden sie durch eine digitale Schaltung vor dem DAU umgesetzt. In einigen Schaltungen treten durch die einzelnen Bits gesteuerte Schalter auf. Sie werden zur Vereinfachung als Kontakte gezeichnet, obwohl sie in der Praxis stets durch einen geschalteten Transistor bzw. eine Diode ersetzt sein dürften. Eine andere, häufige Variante besteht darin, Konstantstromquellen statt großer Widerstände (R) zu verwenden. Dann können die sehr kleinen Widerstände r durch realistische Werte ersetzt werden, und es entstehen an ihnen brauchbare Ausgangsspannungen.

Statt der in allen Bildern gezeigten Spannungsquelle hat sich in den integrierten Techniken weitgehend die Stromquelle durchgesetzt. Die Widerstände R; $2R$; $4R$ usw. im Bild 9.1.1 sind dann Stromspiegel (7.3.4.2.) mit den relativen Flächen 1; $^1/_2$; $^1/_4$ usw. Die Schalter sind häufig

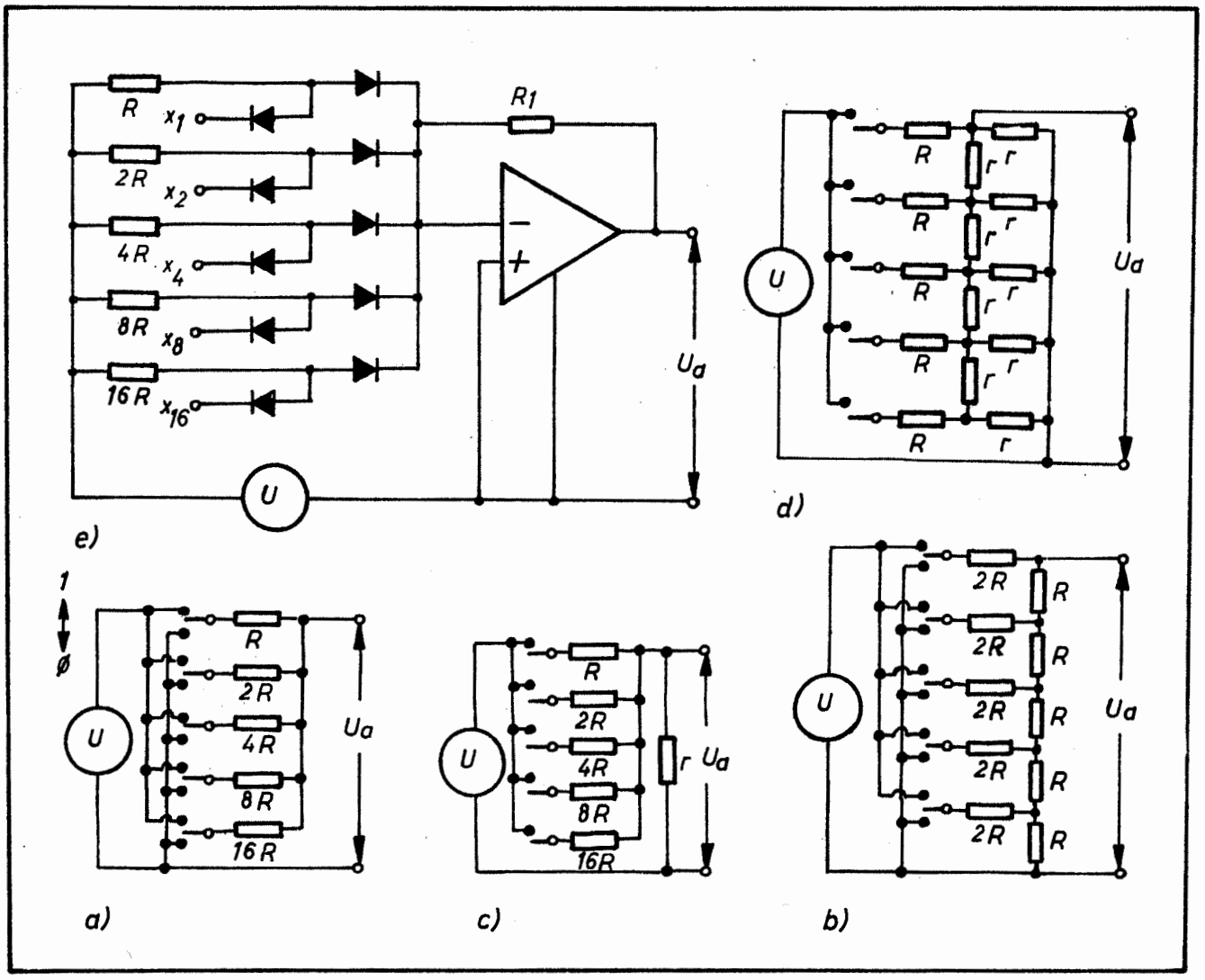


Bild 9.1.1. Einige Prinzipschaltungen des DA-Umsetzers nach dem Schalterprinzip

a) Spannungsgesteuert, b) spannungsgesteuert mit gleichen Widerständen, c) stromgesteuert, d) stromgesteuert mit gleichen Widerständen. Es muß $r \ll R$ sein. e) Zusammenschaltung von stromgesteuerten DA-Umsetzern (Bild c)) und einem Operationsverstärker mit Diodenschaltern

FET, aber auch bipolare Transistoren werden verwendet. Ein besonders übersichtlicher DAU entsteht, wenn an die Ausgänge einer PIO unmittelbar die Widerstände gemäß den Bildern 9.1.1a) bis d) angeschlossen werden. Dabei entstehen allerdings Fehler, wenn die einzelnen Ausgänge Spannungen abgeben, die unterschiedlich von Null bzw. vom Maximalwert, z. B. 5 V, sind. Sind alle Null- und Maximalwerte untereinander gleich, so kann eine Kompensation dieser Werte erfolgen. Insgesamt entsteht daher aus den Grundprinzipien eine große Vielfalt.

9.1.1. Schalterprinzip

Es sei angenommen, daß die Digitalwerte als Schalterstellungen vorliegen. Ein DA-Umsetzer kann dann spannungs- bzw. stromgesteuert ausgeführt werden. Der spannungsgesteuerte Umsetzer benötigt meist Umschaltkontakte und immer eine sehr konstante Spannungsquelle U. Im Beispiel von Bild 9.1.1a) tritt bei 1 1 1 1 1 am Ausgang genau der Wert U auf. Zu allen anderen Schalterkombinationen gehört ein entsprechend geteilter Betrag. Ein Nachteil dieser Schaltung sind die stark verschiedenen Widerstände. Er läßt sich weitgehend durch eine Schaltung gemäß Bild 9.1.1b) vermeiden. Eine Belastung beider DA-Umsetzer ändert, sofern sie konstant ist, nur

die Größe der Ausgangsspannung, jedoch nicht die Genauigkeit. Sie ist durch die Genauigkeit der Widerstände bestimmt.

Eine Vereinfachung der Schalter zu Ein-Aus-Schaltern läßt sich nach dem Prinzip der Stromsteuerung gemäß Bild 9.1.1c) und d) erreichen. Die Spannung U muß hier wiederum sehr konstant sein. Der bzw. die Widerstände r müssen jedoch einen sehr kleinen Wert gegenüber R besitzen. Dies ist deshalb notwendig, damit der Strom nur von den Widerständen R, $2R$, $4R$, $8R$ usw. bestimmt wird. Abweichungen infolge eines Spannungsabfalls an r bedeuten Fehler in der Umsetzung. Der Vorteil der einfachen Schalter wird also durch den Nachteil der kleinen Ausgangsspannung erkauft. Die kleine, gegen Null strebende Ausgangsspannung erfordert nun aber gerade der Operationsverstärker in der Umkehrschaltung (vgl. 7.5.). Werden die mechanischen Schalter durch Diodenschalter ersetzt, so geht Bild 9.1.1c) über in Bild 9.1.1e). Dabei entspricht positive Spannung an den Eingängen x_n einer 1 und negative Spannung einer Ø. Bei positiver Spannung sind nämlich die Dioden zum Eingang x_n gesperrt, und der Strom von U fließt in den Operationsverstärker.

Bei negativer Spannung ist die Diode vor dem Verstärker gesperrt und die andere geöffnet. Der Strom wird also nach x_n abgeleitet. Natürlich führt hier das Verhältnis von Durchlaß- und Sperrwiderstand zu Fehlern.

9.1.2. Prinzip der Kondensatorentladung

Für die folgende Methode muß das digitale Signal drei spezielle Forderungen erfüllen.

- Es ist die Dualzahl rückwärts, also bei den kleinen Werten beginnend, anzubieten. Die Zahl 22 muß somit in der Reihenfolge Ø11Ø1 auftreten.
- Die einzelnen Impulse (0 oder 1) müssen in genau gleichem und festgelegtem Abstand auftreten.
- Die 1-Impulse besitzen einen festgelegten Ladungswert; die Ø-Impulse verschwinden.

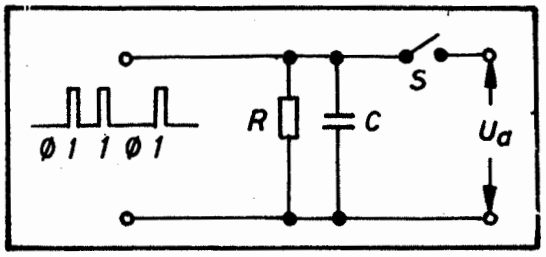


Bild 9.1.2. Prinzipschaltung des DA-Umsetzers nach der Kondensatorentladeschaltung

Gemäß Bild 9.1.2 wird die Impulsfolge dann auf eine Parallelschaltung aus R und C gegeben. Ihre Zeitkonstante $T = RC$ ist so gewählt, daß genau in der Zwischenzeit zwischen zwei möglichen Impulsen, also während der reziproken Impulsfrequenz T_i, die vorhandene Ladung auf die Hälfte absinkt. Es muß also gelten

$$e^{-T_i/RC} = 0{,}5\,. \tag{1}$$

Folglich ergibt sich

$$T_i/RC \approx 0{,}6932\,. \tag{2}$$

Der erste Impuls möge die gegebene Ladung auf C übertragen. Zum Zeitpunkt des zweiten Impulses ist dann nur noch die Hälfte, beim dritten Impuls $^1/_4$, beim vierten $^1/_8$ usw. vorhanden. Auf einem gleichen Endzeitpunkt, zu dem der Schalter S

eingeschaltet ist, verbleiben die nach dem ersten Impuls folgenden mit der doppelten, vierfachen usw. Ladung wirksam. Es liegt somit ein sehr einfacher DA-Umsetzer vor. Seine Probleme bezüglich der Genauigkeit liegen jedoch im Einhalten der drei zuvor geforderten Bedingungen und bei der Einhaltung des gleichen Schaltpunktes für S.

9.1.3. Impulsdauerprinzip

Schon mehrmals wurde auf das Prinzip der Pulslängenmodulation zur Übertragung und Verstärkung analoger Werte eingegangen, z. B. in Abschnitt 7.3.8.4. und 7.4.6.2. Auch für die Digital-Analog-Wandlung bildet es eine günstige Voraussetzung. Es sind ein Schalter, Vergleicher, Zähler und Oszillator nötig. Es bestehen die beiden Möglichkeiten, mit konstanter und variabler (stochastischer) Ausgangsfrequenz zu arbeiten. Beide Prinzipien zeigt Bild 9.1.3.

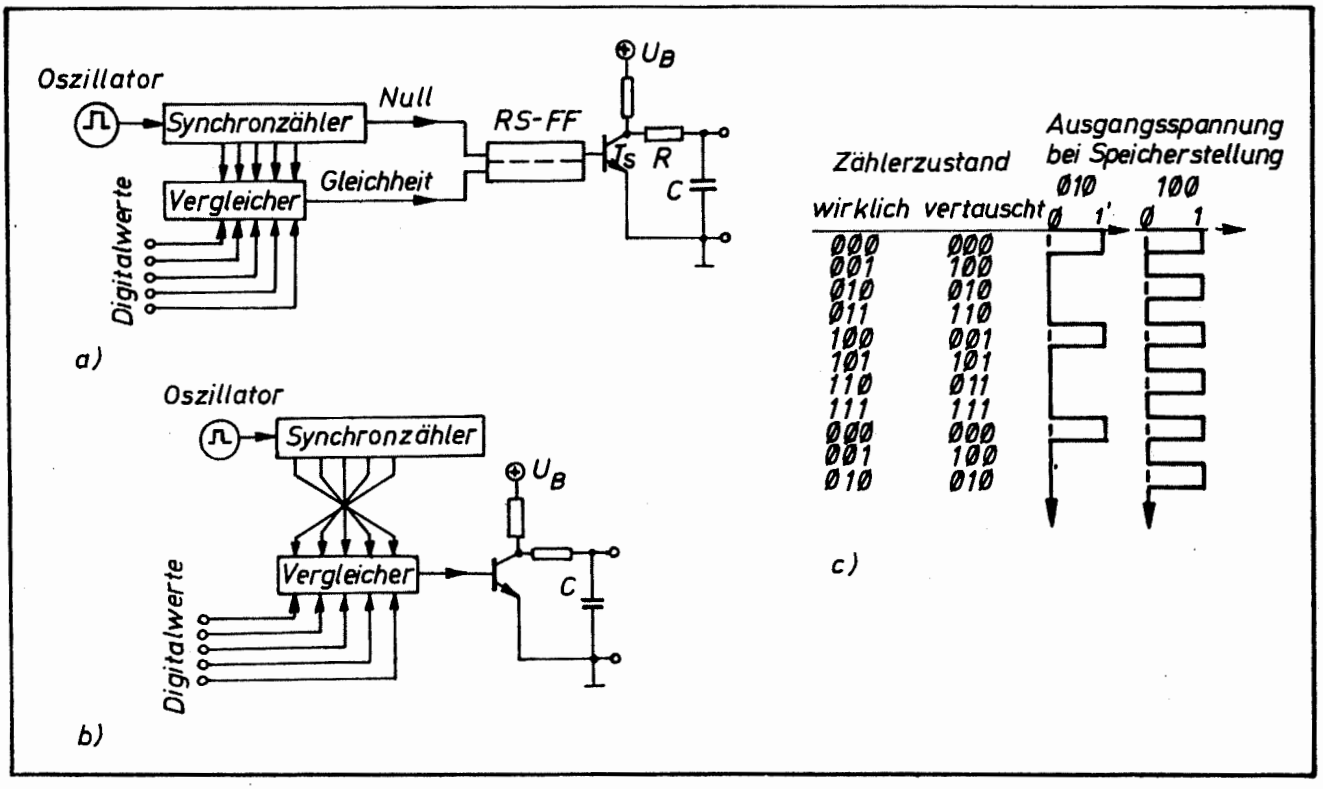


Bild 9.1.3. Prinzipien für DA-Wandler nach dem Prinzip der Pulslängenmodulation

a) Verfahren mit konstanter Ausgangsfrequenz
b) Verfahren mit sich verändernder Ausgangsfrequenz
c) Impulsbildung bei Verfahren mit variabler Frequenz

Beim Prinzip mit konstanter Frequenz (Teilbild a)) läuft der Zähler durch und setzt immer dann das RS-Flipflop, wenn er Null durchläuft. Dann wird der Schalttransistor geöffnet. Der Vergleicher setzt das Flipflop zurück, wenn Gleichheit zwischen den Eingangssignalen und dem Zählerstand besteht. Dabei wird der Schalttransistor gesperrt. Da der Zähler mit konstanter Frequenz betrieben wird, ist der Rechteckimpuls am Ausgang des Schalttransistors genau so lang, wie die Anzahl der Impulse von Null bis zum digitalen Eingangswert dauert. Die gesamte Impulsperiode ergibt sich andererseits aus der Anzahl der Speicherstellen des Zählers und der Oszillatorfrequenz.

Bei variabler Frequenz gilt Teilbild 9.1.3b). Hier ist der Vergleicher komplizierter. Er gibt ein Ausgangssignal dann ab, wenn der Wert des Eingangssignals größer als

der des Synchronzählerstandes ist. Ferner werden am Vergleicher die hoch- und niederwertigen Anschlüsse bezüglich Signal und Zähler vertauscht. So entsteht eine Ausgangssignalfolge, wie sie Bild 9.1.3c) aufzeigt. Es entstehen immer kurze Impulse von der Dauer eines Zählerstandes. Die Pausen dazwischen ergeben sich aus dem jeweiligen Eingangssignalwert. Für den Eingangswert 0 0 0 0 0 bzw. 1 1 1 1 1 treten offensichtlich die gleichen Impulsfolgefrequenzen wie beim ersten Verfahren auf. Für die anderen Eingangssignale sind sie kürzer. Bei beiden Prinzipien ist offensichtlich eine konstante Frequenz des Oszillators wesentlich. Aus ihr werden dann mittels Zählung die Ausgangsimpulse abgeleitet. Deshalb tragen derartige DA-Wandler auch den Namen Zählprinzip. Wird die Ausgangsspannung integriert, so entsteht ein unsymmetrischer Sägezahn. Daher kommt auch der Name Sägezahnprinzip. Ihm kommt noch besondere Bedeutung im Zusammenhang mit den AD-Wandlern zu.

Im Bild 9.1.3 wurde eine bipolare Schaltstufe ausgewählt. Es sind auch andere möglich, z.B. in MOS- oder CMOS- oder generell Gegentaktstufen. Da dieses Wandlerprinzip besonders gut integrierbar ist, wird die Schaltstufe sich meist nach der Technik der anderen Teile richten. Fehler entstehen vor allem durch die Flankensteilheit bei der Schaltstufe und den sie beeinflussenden Temperaturkoeffizienten sowie durch Nichtlinearitäten bei den Schaltern.

Nach diesem Prinzip sind relativ gut digital einstellbare Potentiometer realisierbar. Der Durchgangswiderstand des Schalttransistors wirkt bei entsprechend hoher Taktfrequenz ähnlich. So wie in der Schaltung aus der Impulsfolge mittels des *RC*-Glieds der Gleichwert ausgefiltert wird, ist er dann mit einem Kondensator zu shunten.

9.1.4. Multiplizierende DA-Wandler

Für alle DA-Wandler ist eine konstante Bezugsspannung bedeutsam. Eine Änderung ihrer Größe geht multiplikativ in die Ausgangsspannung ein. Wie unter anderem im Abschnitt 7.5.7.3. gezeigt wurde, gibt es bei der Multiplikation nur relativ komplizierte Prinzipien. Hier bietet sich nun eine neue Möglichkeit an, die sich vor allem durch hohe Genauigkeit auszeichnet. Vielfach ist auch ein Betrieb mit positiver und negativer Bezugsspannung möglich. Dann entsteht sofort ein Zweiquadrantenmultiplikator. Es sind so nur noch zwei, in den anderen Fällen vier solche Schaltungen zu einem Vierquadrantenmultiplikator zu vereinen.

9.2. Analog-Digital-Wandler

9.2.1. Grundlagen

Ein Analog-Digital-Wandler hat zwei Aufgaben zu erfüllen:

— Diskretisierung (Quantisierung) der analogen Signale
— Kodierung der diskreten Signale zu digitalen.

Bei beiden Teilen können Fehler auftreten. Bei der Diskretisierung handelt es sich um eine Intervallzuordnung entsprechend Bild 7.2.6. Es entsteht also auch der entsprechende Quantisierungsfehler. Für n bit beträgt er maximal

$$\Delta x = \frac{1}{2^{n+1}}. \tag{3}$$

Weitere Fehlermöglichkeiten ergeben sich aus der Verarbeitung des Eingangssignals. Dabei sind zwei Typen von AD-Wandlern zu unterscheiden:

— Momentanwertwandler
— integrierende AD-Wandler.

Die integrierenden AD-Wandler bilden das Digitalsignal aus dem Eingangsmittelwert während eines Taktes. Die Momentanwandler benötigen dagegen während des ganzen Taktes ein konstantes Eingangssignal. Ihnen wird deshalb entsprechend Abschnitt 7.3.6. eine Abtastschaltung vorgeschaltet. An sie werden für eine gleichbleibende Fehlerfreiheit sehr hohe Anforderungen gestellt. Soll z.B. der Fehler entsprechend Gl. (3) eingehalten werden, so muß ein vorgeschalteter Verstärker eine erheblich größere Bandbreite besitzen als das zu verarbeitende Signal. Gemäß Abschnitt 2.3. gilt für den Betrag des Übertragungsfaktors

$$f(\omega) = \frac{1}{1 + \left(\frac{\omega}{\omega_g}\right)^2}. \tag{4}$$

Darin bedeuten ω_g die Grenzfrequenz des Verstärkers und ω die zu übertragende Frequenz.

Soll dieser Wert den Fehler von Gl. (3) unterschreiten, so folgt nach einer Näherungsrechnung wegen $2^n > 1$ etwa

$$\omega_g \approx \omega \cdot 2^{n/2}. \tag{5}$$

Für 6 bit ist also bereits die 8fache und für 10 bit die 32fache Bandbreite notwendig. Andererseits führt die Bandbreitenerweiterung durch die Zunahme des Verstärkerrauschens mit der Bandbreite zu zusätzlichen statistischen Fehlern, so daß hier oft ein Kompromiß zu suchen ist.

Die Systematisierung der AD-Wandler ist schwierig. Es existieren eine Vielzahl von Prinzipien, die nur zwangsweise in ein System passen.

Hier sollen zwei Wege gegangen werden. Zunächst erfolgt eine Beschreibung unterschiedlicher Wandlerverfahren, die keineswegs vollständig ist. Es werden die folgenden Prinzipien behandelt:

— mit DA-Wandler
— direkte Wandlung
— Ausschlagprinzip
— Wandler auf Zeitbasis
— Wandler mit Spannungs-Frequenz-Umsetzern
— Wägeprinzip
— mit Ladungsübertragung.

Danach erfolgt eine Klassifizierung nach den Methoden des Vergleichs zwischen Analogwert und Normal bzw. der damit im Zusammenhang stehenden Anzahl von sequentiellen Vergleichsoperationen.

Schließlich werden abschließend noch einige Aussagen zum Stand der Technik, bezüglich der Anwendungen, Auflösung und Wandlerrate mitgeteilt.

9.2.2. AD- mit DA-Wandlern

Wie bereits mitgeteilt, sind in der Regel DA-Wandler einfacher als AD-Wandler. Deshalb ist es sinnvoll, mittels DA-Wandlern und Zusatzelektronik AD-Wandler aufzubauen. Zwei Prinzipien zeigt Bild 9.2.1. Beim Teilbild a) liegt eine Schaltung mit

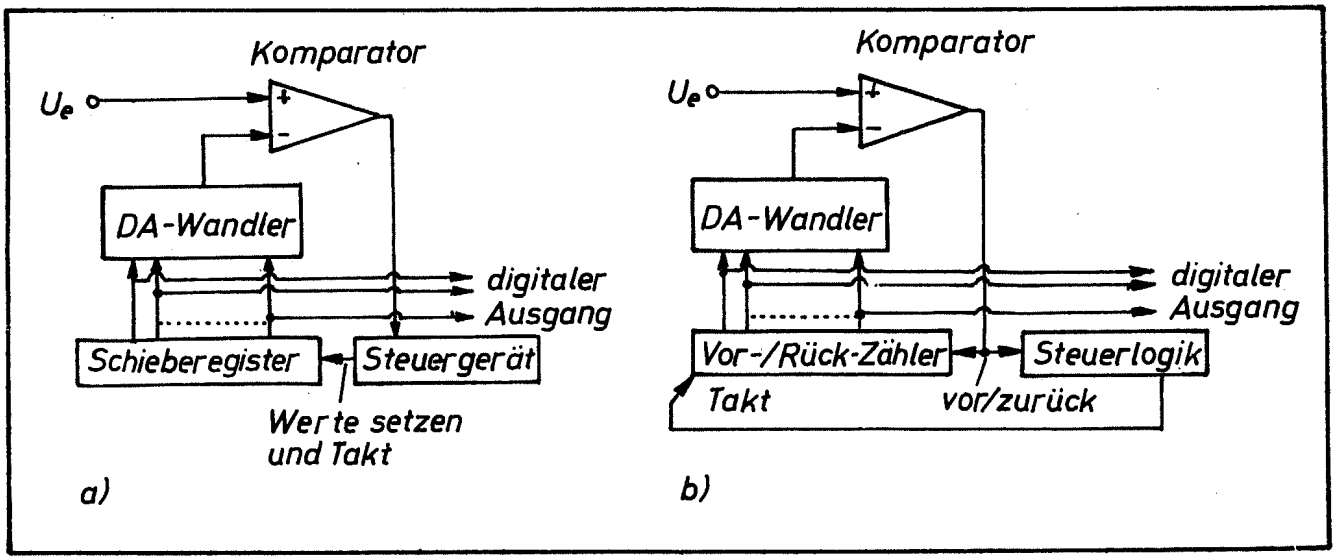


Bild 9.2.1. AD-Wandler unter Ausnutzung von DA-Wandlern
a) mit stufenweiser Annäherung,
b) mit Vor-Rückwärts-Zähler

stufenweiser Näherung an den richtigen Wert vor. Zu Beginn eines jeden Meßzyklus wird das Schieberegister Null gesetzt. Dann wird das hochsignifikante Bit gesetzt. Je nachdem, ob am Ausgang des Komparators dabei ein + oder — entsteht, bleibt es bestehen oder wird gelöscht. Dasselbe folgt der Reihe nach für alle weiteren Bits. Zum Schluß liegt der richtige Digitalwert vor und kann ausgegeben werden.

Wird statt des Schieberegisters ein Vor-Rückwärts-Zähler verwendet, so wird immer nur der DA-Wandler einem sich ändernden Eingangssignal nachgeführt (Teilbild b)). Deshalb heißt dieses Prinzip auch Nachlaufverschlüßler. Es arbeitet relativ schnell und ohne periodische Zyklen.

9.2.3. Direkte Wandlung

Das Prinzip der direkten Wandlung zeigt Bild 9.2.2. Für n bit sind hier zunächst $2^n - 1$ Komparatoren und eine beträchtliche Anzahl von logischen Gattern notwendig.

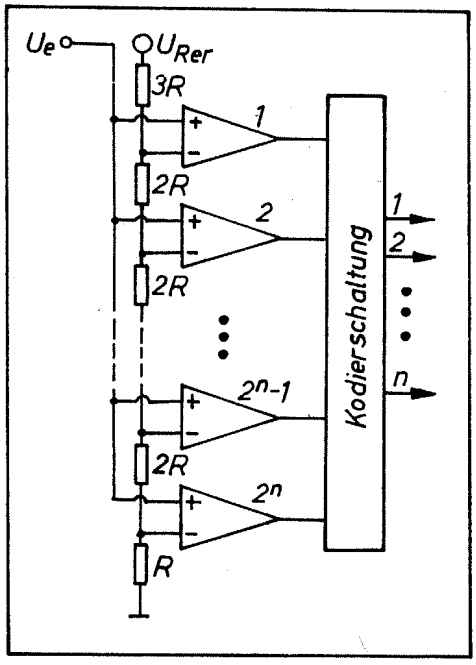


Bild 9.2.2. Prinzipschaltung des AD-Wandlers nach dem direkten Verfahren

Dieses Prinzip ist extrem schnell, da der Digitalwert praktisch sofort zur Verfügung steht, es benötigt aber einen großen Aufwand. Weiterhin ist von Nachteil, daß jeder Komparator eine Schaltschwelle besitzen muß, die klein gegenüber dem minimal signifikanten Bit ist. Die einzelnen Komparatorwerte dürfen sich auch nicht durch Drift oder Temperatureinflüsse gegeneinander verschieben. Aus diesem Grunde werden derartige Wandler nur für eine geringe Bitzahl gebaut.

9.2.4. Wegumsetzer

Das einfachste und älteste Prinzip für AD-Wandler ist der Wegumsetzer. Am übersichtlichsten ist sein Prinzip bei radial betriebenem Scheibenverschlüßler entsprechend Bild 9.2.3. Hier sind im üblichen Binärkode die Winkel in 32 Felder geteilt und werden mit einer Maske, die unter einem Winkel von +30° angedeutet ist, z.B. lichtoptisch abgetastet. So entstehen die im Bild angegebenen 1Ø-Werte. Bei diesem Prinzip gibt es Probleme, wenn sich gleichzeitig beim Verdrehen der Scheibe zu einem anderen Winkel zwei Werte ändern. Dies ist im Bild der Fall, wenn vom gezeigten Wert 10 zum Wert 9 übergegangen wird. Deshalb haben hier die einschrittigen Kodes besondere Bedeutung (vgl. 8.1.1.). Verwendet wurden derartige oder ähnlich gestaltete Kodescheiben unter anderem unmittelbar auf der Achse von Drehspulinstrumenten oder im Zusammenhang mit speziellen Oszillographenröhren. Heute werden sie vor allem noch bei der direkten Winkelkodierung benutzt.

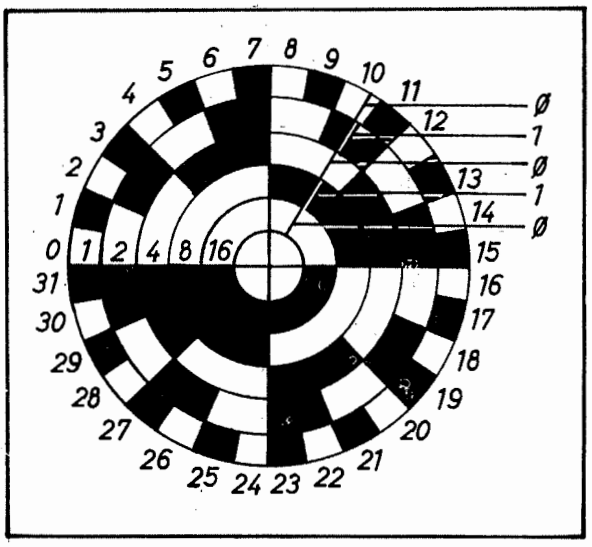


Bild 9.2.3. Prinzip des Scheiben- (Weg-) Verschlüßlers

9.2.5. Wandler auf Zeitbasis

AD-Wandler auf Zeitbasis gibt es in einer größeren Anzahl von Varianten. Hier sei vor allem der Sägezahnverschlüßler gemäß Bild 9.2.4 besprochen. Er benötigt einen Präzisionssägezahngenerator. Dessen Ausgangsspannung wird in einem Komparator mit der Eingangsspannung verglichen (single slope = Einflanken-ADU). Der Zähler kann vom Quarzsignal gesteuert werden. Dazu muß ein Tor auf- und zugesteuert werden. Der Start des Zählers beginnt, sobald der Sägezahn die Nullinie durchläuft. Er wird gestoppt, sobald Meßwert und Sägewert gleich groß sind. Diese Auslösung erfolgt durch den Komparator.

Eine Erhöhung der Meßgenauigkeit entsteht, wenn Start und Stop des Zählers durch den Komparator bewirkt werden. Dann muß das Tor aber während der Zeit aufgesteuert sein, in der die Sägezahnspannung größer als die Meßspannung ist. Damit

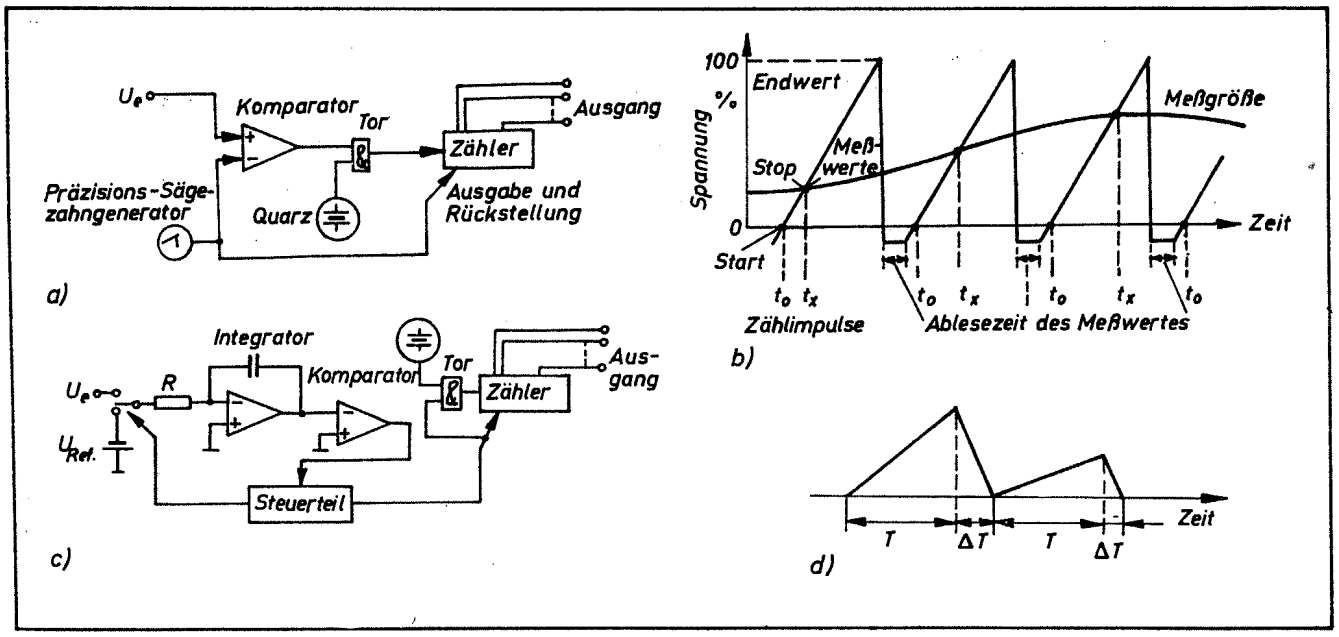


Bild 9.2.4. Prinzip des Sägezahnwandlers

a) Prinzipieller Aufbau
b) Wirkungsweise
c) Dual-Slope-Verfahren
d) Wirkungsweise

dann die Änderung der Meßspannung nicht eingeht, ist ein Halteglied für U_e vorzuschalten. Die Einschaltung des Tores erfolgt auf der langsamen Vorderflanke des Sägezahnes und die Ausschaltung auf der steilen Rückflanke. Hierdurch können Nichtlinearitäten des Sägezahnes kompensiert werden. Dies führt in der Weiterentwicklung zum Dual-Slope-Prinzip (Zweiflankenwandler) gemäß Bild 9.2.4c). Es verwendet meist einen Integrator zur Erzeugung des Sägezahnes. Zunächst wird für eine bestimmte Zeit, die vom Steuerteil festgelegt wird, die Eingangsspannung integriert. Dann wird auf eine größere, entgegengesetzt gepolte Bezugsspannung U_{Ref} umgeschaltet. Gleichzeitig wird die Quarzfrequenz vom Steuerteil an den Zähler geschaltet. Dabei wird die Integrationskapazität C über R mit derselben Zeitkonstante, mit der sie vorher aufgeladen wurde, entladen. Sobald dabei der Nullwert erreicht wird, beeinflußt der Komparator das Steuerteil. Es schaltet den Eingangsschalter auf die Eingangsspannung, öffnet das Tor zum Zähler und bewirkt die Ausgabe des Zählerstandes. Nach einer bestimmten Zeit wird wieder der Eingangsschalter umgeschaltet und das Tor für die Quarzfrequenz geöffnet. Das Ausgangssignal wird offensichtlich nur durch die Länge $\Delta T/T$ bestimmt. Daher werden langsame Änderungen der RC-Zeitkonstante und der Quarzfrequenz nicht wirksam. Ebenso kompensieren sich im gewissen Umfang Nichtlinearitäten des Sägezahnes.

9.2.6. Wandler mittels Spannungs-Frequenz-Wandler

Spannungs-Frequenz-Wandler wurden u. a. schon im Unterabschnitt 7.3.2.3. und 7.3.7.3. erwähnt. Sie gestatten einen relativ einfachen Aufbau von AD-Wandlern. Das Prinzip zeigt Bild 9.2.5. Vom Steuergerät werden mittels eines Frequenznormals der Zähler und das Tor geschaltet. Das Tor wird für eine bestimmte Zeit geöffnet, die dann gezählten Impulse entsprechen der Eingangsspannung. In der folgenden Pause wird der Zählerwert ausgegeben und der Zähler auf Null zurückgestellt. Im Bild 9.2.5 b) ist ein sehr gut linearer Frequenzwandler nach dem im vorigen Unterabschnitt be-

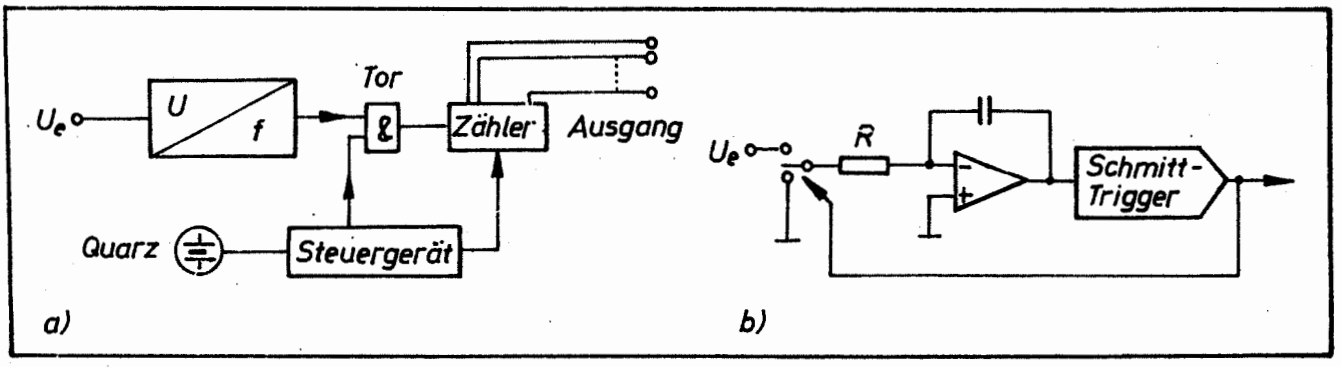


Bild 9.2.5. a) Prinzip eines AD-Wandlers mit einem Spannungs-Frequenz-Wandler und b) Anwendung des Dual-Slope-Prinzips als Spannungs-Frequenz-Wandler

sprochenen Dual-Slope-Prinzip dargestellt. Er schaltet ständig zwischen den beiden Schwellwerten des SCHMITT-Triggers (vgl. 7.5.7.6.) um. Die Frequenz hängt dabei von der Spannung und der Integrationszeitkonstanten ab.

9.2.7. Wägeprinzip

Das Wägeprinzip wird besonders viel angewendet. Es kann im gewissen Sinne als reziprok zum Schalterprinzip des DA-Wandlers angesprochen werden. Es trägt auch die Namen Wandler mit Widerstandsnetzwerk, mit sukzessiver Näherung, Stufenverschlüßler. Sein Prinzip zeigt Bild 9.2.6. Er benötigt Widerstandsnetzwerke und Schalter, wie sie z.B. schon im Bild 9.1.1 gezeigt sind.

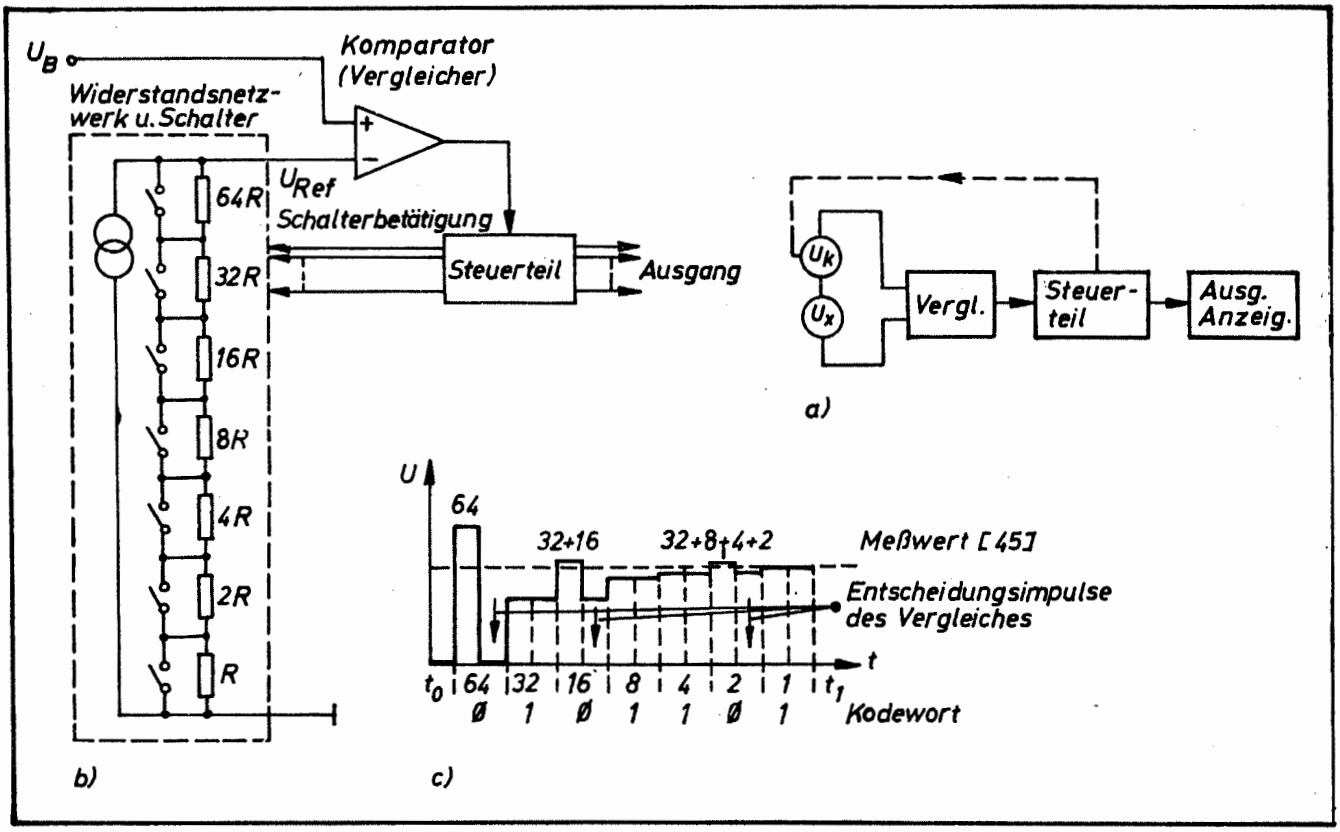


Bild 9.2.6. AD-Wandler nach dem Wägeprinzip

a) Prinzipschaltung
b) Detaillierte Schaltung mit Spannungsteiler
c) Abgleichvorgang

Die Ausgangsspannung des Widerstandsnetzwerkes wird mit der Eingangsspannung einem Komparator zugeführt, der seinerseits ein relativ kompliziertes Steuerteil beeinflußt. Es setzt zu Beginn alle Schalter auf geschlossen und betätigt sie dann in der Reihenfolge von höchster nach niedrigster Wertigkeit. Entsprechend Bild 9.2.6c) möge ein Meßwert 45 vorliegen. Wird dann der Schalter für $64R$ geöffnet, so ist die Bezugsspannung U_{Ref} zu groß. Dies bewirkt einen negativen Ausgangsimpuls des Komparators an das Steuerteil. Dadurch wird dieser Schalter wieder geschlossen und bleibt es auch in diesem Zyklus. Dann wird der Schalter für 32 R geöffnet. Da die Ausgangsspannung U_{Ref} jetzt kleiner als U_{e} ist, bleibt er geöffnet. Als nächster folgt der Schalter für $16R$. Es entsteht eine Spannung $32 + 16 = 48$. Sie ist zu groß. Folglich wird mittels eines negativen Impulses des Komparators der Schalter wieder geschlossen. So werden alle Schalter betätigt, und am Ende kann das binäre Signal entsprechend der eingestellten Schalter ausgegeben werden. Auch eine sequentielle Ausgabe während des Abgleichvorganges (Wägeprinzip) ist möglich.

9.2.8. Prinzip der Ladungsübertragung

Dieses Prinzip ist relativ wenig bekannt. Es zeigt Bild 9.2.7. Wesentlich sind zwei Fakten:

Wird C_2 bzw. C_4 an den Eingang des Integrators geschaltet, so übernimmt C_3 die gesamte Ladung. Dies ist darin begründet, daß zwischen den Eingangsklemmen eines Operationsverstärkers die Spannungsdifferenz Null angestrebt wird.

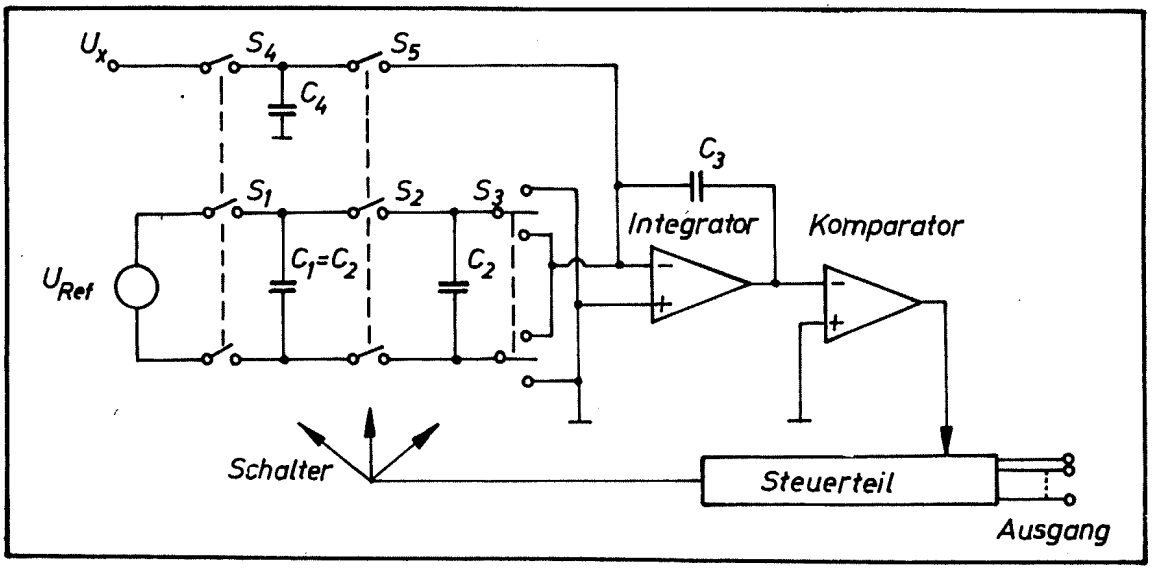


Bild 9.2.7. Prinzip eines AD-Wandlers nach der Ladungsübertragung

C_1 sei von der Normalspannungsquelle U_{Ref} aufgeladen. Wird jetzt mittels Schalter S_2 die ungeladene Kapazität C_2 zugeschaltet und gilt $C_1 = C_2$, so teilt sich die Ladung exakt auf beide Kapazitäten auf. Dieser Vorgang kann beliebig oft wiederholt werden. Es können also damit binär gestaffelte Ladungsmengen erreicht werden.

Zu Beginn eines Meßzyklus werden S_1 und S_4 kurzzeitig geschlossen. Dann trägt C_4 eine Ladungsmenge, die der Eingangsspannung, und C_1 eine, die U_{Ref} entspricht. Mittels S_5 wird im nächsten Taktteil die Ladung von C_4 auf C_3 übertragen. Gleichzeitig wird mittels S_2 die Ladung von C_1 und C_2 geteilt. Dann folgt mit S_3 der Abbau der Meßladung auf C_3. Der Schalter wird dazu nach unten betätigt. Je nachdem, ob die Ausgangsspannung am Integrator weiterhin positiv oder negativ ist, gibt der Kom-

parator eine 1 oder Ø zum Steuerteil als höchst signifikantes Bit. Danach wird die halbe Ladung auf C_1 mittels Betätigung von S_2 nochmals halbiert. Je nachdem, ob zuvor der Komparator 1 oder Ø ausgegeben hat, wird S_3 in derselben Richtung zur weiteren Ladungssubtraktion oder entgegengesetzt zur Ladungsaddition betätigt. Die dann entstehende Ausgangsspannung des Komparators steuert in entsprechender Weise die weiteren Schaltschritte bis zum niedrigstelligsten Bit. Dann wird C_3 und C_2 völlig entladen, und ein neuer Meßzyklus kann mit Betätigung von S_1 und S_4 beginnen.

Für die Wirkungsweise der Schaltung ist die Qualität der Kondensatoren und die hinreichende Kleinheit der Leckströme von großer Bedeutung. Es ist günsitg, wenn nicht nur $C_1 = C_2$ gilt, sondern C_4 genauso groß ist. Dann entfallen nämlich alle Maßstabsfaktoren.

9.2.9. Vergleich von ADU

Der Vergleich von ADU kann nach unterschiedlichen Gesichtspunkten erfolgen. Bezüglich der Anzahl der Normale und der notwendigen *Abgleichschritte* sind zunächst zwei Extreme zu erkennen.

Bei der *direkten Methode* gemäß Bild 9.2.2 sind für n bit

$$h = 2^n - 1 \tag{6}$$

Normalquellen (Spannungspegel) und genauso viele Komparatoren notwendig. Dies bedeutet z.B.

Auflösung	4	6	8	10	12	bit
Komparatoren	15	63	255	1023	4095	

Parallele ADU werden also *kaum über 8 bit hinaus* zu bauen sein. Dafür stehen die einzelnen Bits aber auch sofort, d. h. ohne Abgleichschritte zur Verfügung. Daher auch die Bezeichnung parallel bzw. englisch *word at a time*. Die Normale werden in der Regel über das Widerstandsnetzwerk realisiert, wodurch für akustische Anwendungen auch eine nichtlineare Staffelung möglich ist.

Das andere Extrem stellen die *Zählmethoden*, z.B. gemäß Bild 9.2.4 und 9.2.5, dar. Hier sind

$$i = 2^n - 1 \tag{7}$$

Abgleichschritte, aber nur ein Normal notwendig. Sie benötigen also die längste Zeit, da sie die einzelnen Impulse zählen (Zählmethode oder englisch *level at a time*). Sie sind besonders gut *integrierbar*.

Zwischen den beiden Extremen liegen *Iterationsverfahren*, meist Wägeverfahren (digit at a time), wie sie z.B. Bild 9.2.6 und 9.2.7 zeigen. Im typischen Fall sind bei ihnen die Anzahlen der Normale und Abgleichschritte gleich, also gilt für sie dann

$$h = i = n\,. \tag{8}$$

Zwischen diesen drei Typen gibt es Übergänge, z.B. durch *Kaskadierung* einzelner Verfahren. Einen typischen Zusammenhang hierfür zeigt Bild 9.2.8a).

Eine andere Einteilung der ADU folgt aus den *Anwendungsgebieten* gemäß Bild 9.2.8b). Die Digitalmultimeter (digitale Meßinstrumente) verlangen die höchste Meßgenauigkeit. Die Anzahl der angezeigten *Ziffern* (digit) hängt mit der *Bitauflösung* wie folgt zusammen:

$$\text{digit} = n \log 2 \approx 0{,}3n\,. \tag{9}$$

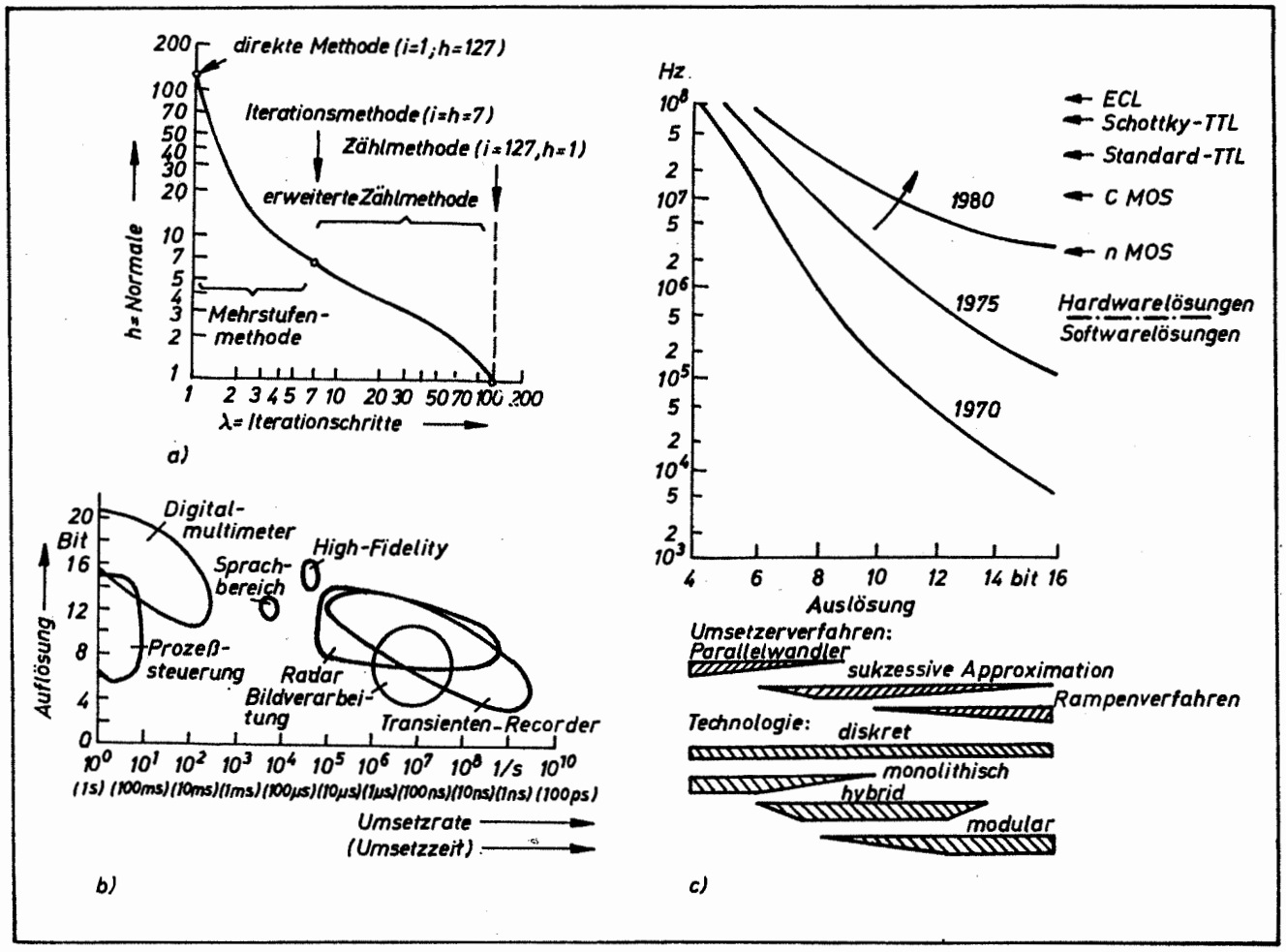


Bild 9.2.8. Zum Vergleich von ADU

a) Einteilung nach Zahl der Normale und Iterationsschritte

b) Einteilung nach Anwendungsbereichen

c) Bei drei Jahreszahlen in etwa erreichte Grenze bez. Wandlerfrequenz und Bitauflösung

Mit 20 bit sind also 6 gültige Ziffern erreichbar. Im akustischen Bereich hängt mit der Bitzahl die erreichbare *Dynamik* (D) zusammen:

$$D/\mathrm{dB} = 20n \log 2 \approx 6n\,. \tag{10}$$

Mit 16 bit wird also eine Dynamik von 100 dB möglich. Dieser Wert wird heute für digitale Audiosignale angestrebt.

Eine besonders hohe Zeitauflösung wird beim Radar, bei der Bildverarbeitung (u. a. digitales Fernsehen) und bei Transientenrecordern (digitalen Oszillographen) notwendig. Hier werden z. Z. gerade die unbedingt notwendigen Forderungen erreichbar. Um dies deutlicher zu zeigen, ist in Bild 9.2.8c) die Grenze der Geschwindigkeit als Funktion der Bitauflösung für die Jahre 1970, 1975 und 1980 dargestellt. Bezüglich der Arbeitsfrequenzen sind die typischen digitalen Techniken angedeutet, während an der Abszisse typische Verfahren der ADU aufgezeigt sind.

9.3. Digitalfilter

9.3.1. Prinzip

Im Abschnitt 7.9.8. wurde ein Überblick zu den verschiedenen Filtern gegeben. Behandelt sind dort vor allem die nichtklassischen Filter, also Abtastfilter, CCD- und SC-Filter. Die Abtastfilter sind in den Abschnitten 7.9.4. und 7.9.5. so ausführlich behandelt, daß jetzt für die Digitalfilter

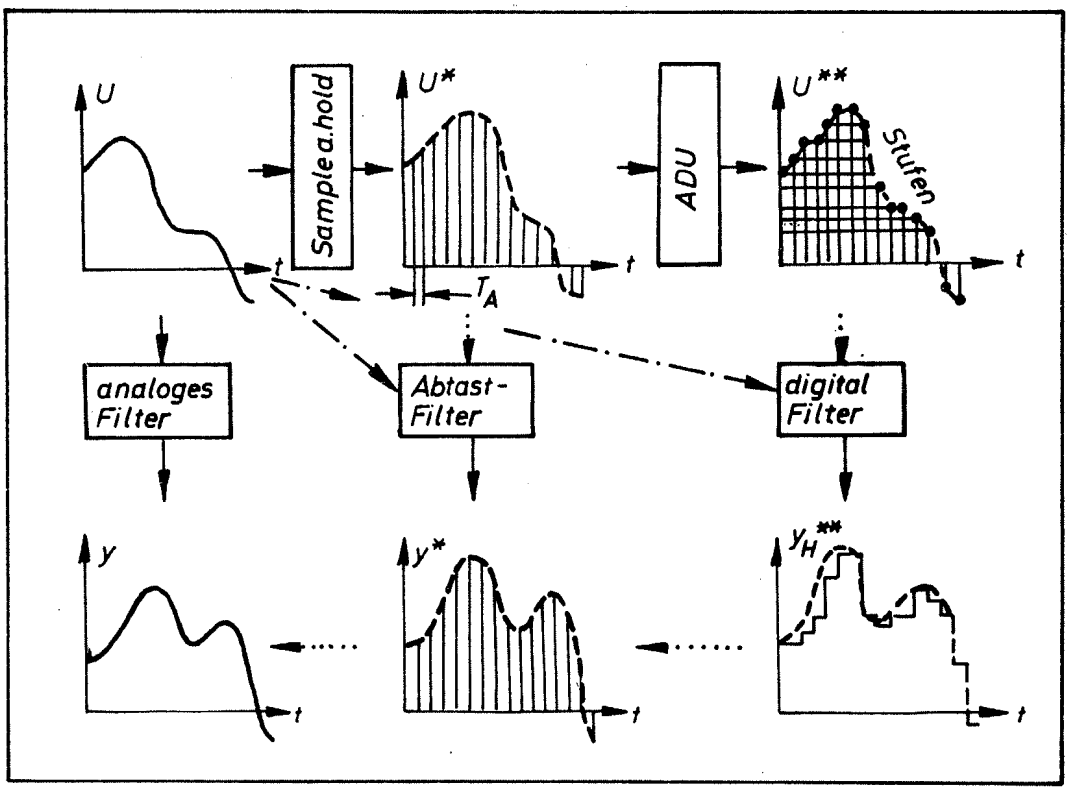


Bild 9.3.1. Zum Vergleich der Signale bei Analog-, Abtast- und Digitalfiltern

nur darauf zu verweisen ist. Das wesentlich Neue sei mittels Bild 9.3.1 beschreiben. Dort ist ganz links ein beispielhafter Zeitverlauf für ein analoges Filter gezeigt. Den Übergang zum Abtastfilter (ohne auf dessen andersartige Struktur und Theorie einzugehen) zeigt die Mitte: Mittels Sample-and-hold werden zeitdiskrete Abtastwerte gewonnen und weiterverarbeitet. Aus dem kontinuierlichen Signal wird eine Impulsfolge. Sie kann gemäß den Darstellungen in Bild 7.9.3 durch eine Gewichtsfunktion, z. B. Tiefpaß, wieder in ein kontinuierliches Signal zurückgewandelt werden.

Beim Digitalfilter folgt nach dem Sample-and-hold ein Analog-Digital-Wandler (ADU). Er legt auch die Amplituden der Impulsfolge auf diskrete Werte fest (im Bilde rechts). Damit liegt im Sinne von Bild 8.1.1 jetzt ein Signal vor, das in den Amplituden- und Zeitwerten diskretisiert ist. Dieses Signal ist mittels eines Digitalrechners bearbeitbar. Wenn dies geschieht, liegt ein Digitalfilter vor. Sein Ausgangssignal ist wiederum in den Amplituden- und Zeitwerten diskretisiert (s. im Bild unten rechts). Es wird durch ein Sample-and-hold (Speicherkondensator) und ein Tiefpaßfilter am Ausgang wieder auf eine möglichst gute, analog verarbeitbare Signalform geglättet.

9.3.2. Beispiel

In Bild 9.3.2a) ist ein analoges Filter, und zwar als Beispiel ein Doppel-T-*RC*-Glied (vgl. 2.3.4.), dargestellt. Seinen Frequenzgang zeigt Teilbild c) als ausgezogene Kurve. Ein einfaches Abtastfilter mit sehr ähnlichem Frequenzgang (gepunktet) zeigt Teilbild b). Bezüglich der Eingangs- und Ausgangsfolgen gilt

$$y(n) = x(n) - 1{,}46x(n-1) + x(n-2)\,. \tag{1}$$

Die Übertragungsfunktion folgt damit zu

$$H(z) = 1 - 1{,}46z^{-1} + z^{-2}\,. \tag{2}$$

Für die gewählten 60 Hz wären 500 Abtastungen je Sekunde notwendig.

Liegen nun bereits die Abtastwerte $x(n)$ vor, so müssen sie für das Digitalfilter nur noch zu Zahlenwerten (in einem ADU) umgewandelt werden. Da nur drei aufeinander-

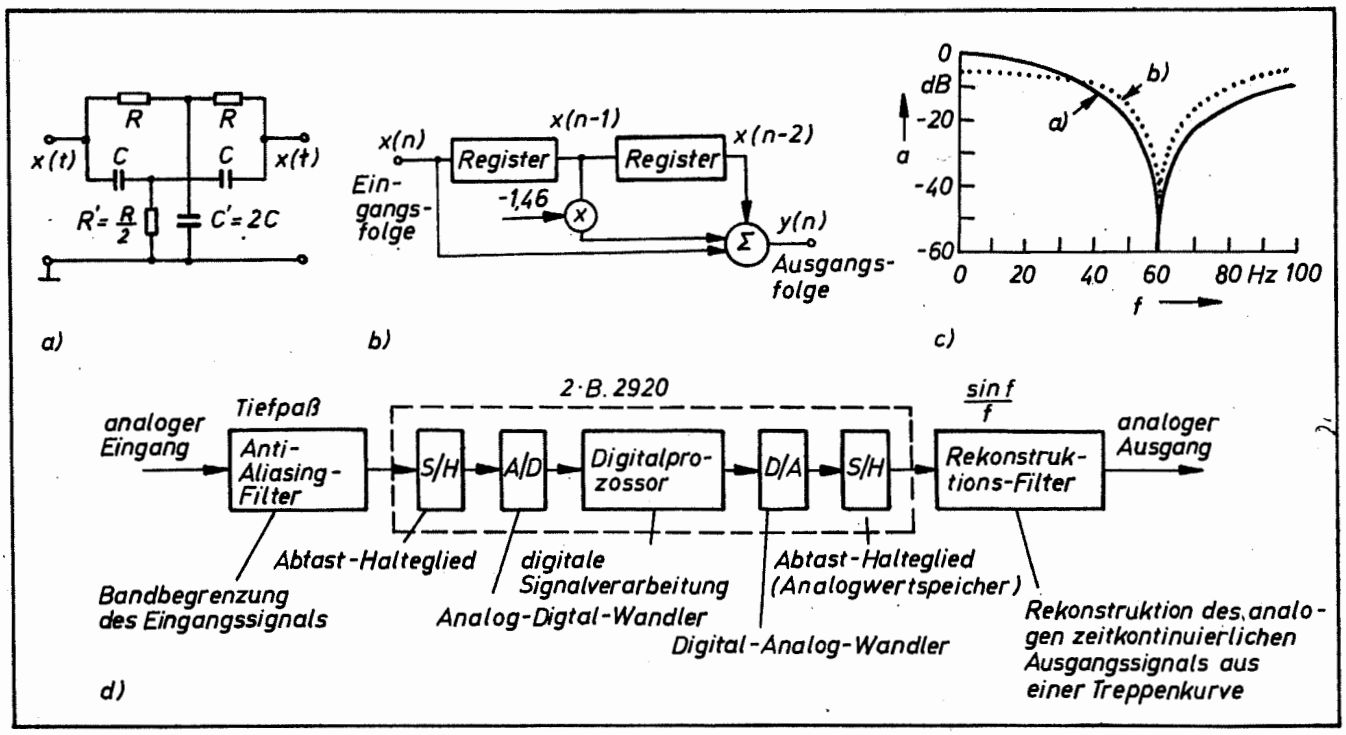


Bild 9.3.2. Beispielvergleich

a) Analoges *RC*-Filter
b) Abtast- bzw. Digitalstruktur
c) Frequenzgänge der beiden Fälle
d) Prinzipielle Struktur eines Digitalfilters

folgende Werte benötigt werden, genügt also ein zweistufiges Schieberegister. Der Wert nach der ersten Verzögerung wird mit einem Faktor — 1,46 digital multipliziert. Digital wird er dann zu dem Eingangswert addiert. Hierzu wird noch der Ausgangswert des Schieberegisters addiert. Im digitalen Bereich sind also nur notwendig

- Addition
- Multiplikation
- Zeitverzögerungen (Register).

Alle drei Operationen sind mit einem Mikroprozessor relativ einfach über ein Programm zu realisieren. Dabei ist zu beachten, daß die Multiplikationen besonders einfach auf der Basis Zwei möglich sind. Gemäß Bild 8.5.22 genügen dann nämlich bitweise Schiebeoperationen. Für den Faktor 1,46 gilt approximativ

$$2^0 + 2^{-1} - 2^{-4} + 2^{-5} = 1 + 0{,}5 - 0{,}0625 + 0{,}0312 = 1{,}4687$$

Er wird also um so genauer, je länger das digitale Wort ist.

Insgesamt zeigt sich die Struktur eines Digitalfilters damit gemäß Bild 9.3.2d), wobei der innere Teil durch spezielle Prozessoren (9.3.4.) realisiert werden kann.

9.3.3. Vor- und Nachteile

Der Nachteil der digitalen Filter war bis vor einiger Zeit der große technische Aufwand. Dieser Nachteil ist durch die VLSI-Technik nicht nur behoben, sondern gerade ins Gegenteil verkehrt. Jetzt sind digitale Filter gerade dadurch mit *extrem kleinen Abmessungen, Verbrauch an elektrischer Leistung* und *Kosten* mittels eines programmierbaren Chips realisierbar. Sie besitzen jedoch Nachteile bezüglich der höchsten *ver-*

arbeitbaren Frequenz. Entsprechend dem Samplingstheorem der Informationstheorie muß die Abtastrate mindestens doppelt so hoch sein, wie die höchste zu übertragende Frequenz. Da beim digitalen Filter zu jedem Abtastwert noch mehrere Multiplikationen und Additionen erfolgen müssen, benötigt der zugehörige Mikroprozessor eine Taktfrequenz, die rund das Hundertfache der höchsten Signalfrequenz beträgt.

Ein zweites Problem betrifft die schon im vorigen Abschnitt erwähnte *Wortlänge* der Zahlen. Sie bestimmt einmal die Amplitudenauflösung, also das Quantisierungsrauschen. Zum anderen besteht die Gefahr von Rundungs- und Überlauffehlern bei der Rechnung. Dies bedeutet eine Skalierung der Eingangssignale.

Aufgrund der Rechnungsfehler müssen die Fehler bei rekursiven und nichtrekursiven Filterstrukturen unterschieden werden. Bei den *nichtrekursiven* Filtern existieren im wesentlichen

- Quantisierungs- und Rundungsrauschen sowie Überlaufverzerrungen.

Bei den *rekursiven* Filtern, die ja leistungsfähiger sind, kommen noch hinzu

- Grenzzyklen und Überlaufschwingungen.

Sie entstehen dadurch, daß die rückgekoppelten Signale zu periodischen Erscheinungen aufgrund von zeitweiligem Überlauf bzw. Rundungseffekten führen.

Den genannten Nachteilen stehen viele *Vorteile* gegenüber. Durch Vergrößern der Wortlänge kann eine fast beliebige Erhöhung der Genauigkeit von Filtern erreicht werden. Es tritt keine störende Alterung von Bauelementen, kein störender Temperatureffekt auf. Dadurch entfällt auch jede Nacheichung.

Da die Eigenschaften von Digitalfiltern vor allem durch das Programm und die Taktrate bestimmt werden, sind sie einmal in der Serienproduktion stabil und reproduzierbar herstellbar. Darüber hinaus sind auch ihre Eigenschaften durch ein anderes Programm schnell veränderbar. Dieser Punkt sei am Beispiel von Bild 9.3.2 besprochen. Soll dort die Nullfrequenz von 60 auf 50 Hz verlegt werden, so müssen beim *RC*-Filter alle drei Kapazitäten (oder Widerstände) neu festgelegt werden. Beim Digitalfilter genügt es, den Koeffizienten auf $-1{,}62$ zu ändern. Eine andere Möglichkeit wäre, die Taktfrequenz für die Abtastrate von 500 auf 417 Hz zu ändern.

9.3.4. Signalprozessor 2920

Die Architektur des 1979 von Intel herausgebrachten Signalprozessors zeigt Bild 9.3.3a). Er kann grob in die drei Anteile zergliedert werden:

Der *Analogteil* enthält vier Signaleingänge, die im Multiplexbetrieb auf eine Sample-and-hold-Schaltung einwirken. Die Haltekapazität kann entsprechend dem Frequenzbereich von außen angeschaltet werden. Die Analog-Digital-Wandlung erfolgt über einen DAU (9.2.2.) von 8 bit Breite mit Spannungsteiler gemäß Bild 9.1.1 und FET-Schaltern. Das Bit 9 ist ein Vorzeichenbit. Mittels der Steuerlogik erfolgt der Abgleich, und das Ergebnis wird ins Datenregister überführt. Die Ausgabe erfolgt nach der Rechnung aus dem Datenregister. Mittels des DUA erfolgt die Rückwandlung in analoge Werte. Sie werden geglättet und können über den Multiplexer auf acht Ausgänge geleitet werden.

Der *Digitalteil* enthält eine 28 bit breite ALU, ein Schieberegister und ein RAM für 40 Worte zu je 25 bit. Jeder Wert kann ein Folgenwert sein. Dann sind Filter mit 40 Polen bzw. Nullstellen realisierbar. Infolge der Multiplex-Ein- und -Ausgänge kann der RAM-Block aber auch parallel mehrere Signalfolgen abarbeiten bzw. erzeugen. Die ALU ist auf Addition und Subtraktion sowie logische Operationen spe-

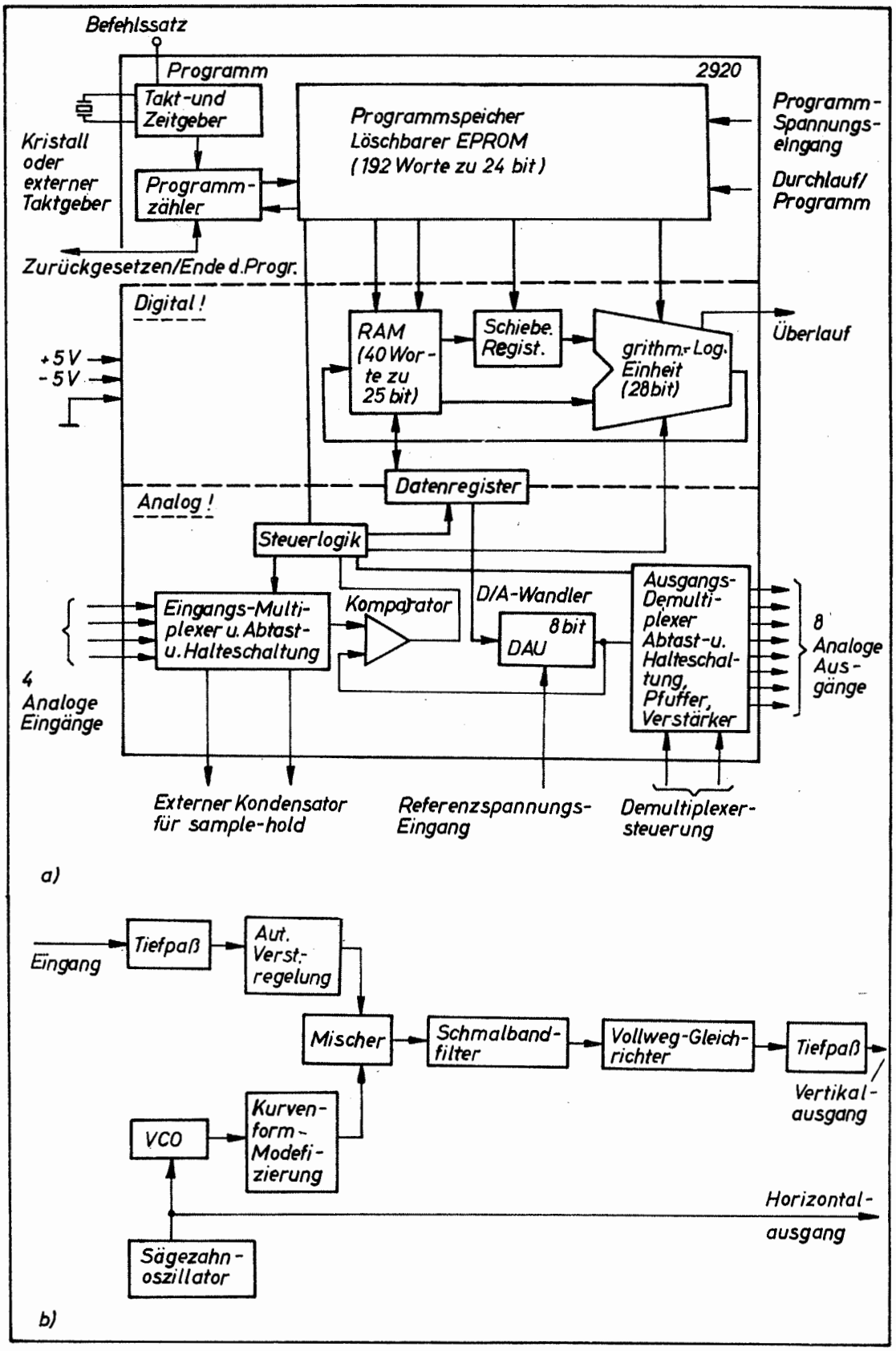


Bild 9.3.3. Signalprozessor 2920 von Intel

a) Architektur

b) Anordnung für einen kompletten Spektrumanalysator

zialisert. Mittels Schieberegister und ALU werden die Konstantenmultiplikationen realisiert. Zur schnellen Verarbeitung existiert eine 4-fach-Pipelinestruktur.

Der Digitalteil wird von dem *Programmteil* gesteuert. Sein Hauptteil ist der UV-löschbare PROM für 192 Befehlsworte der Länge 24 bit. Dieser PROM wird durch

eine besondere Eingabe und einen dafür entwickelten Assembler so gesetzt, wie der entsprechende Signalprozessor arbeiten soll. Weiter enthält dieser Teil den Programmzähler und den Takt- und Zeitgeber, der durch einen Quarz oder extern gesteuert werden kann. Die maximale Taktfrequenz liegt bei 10 MHz.

Die Abtastfrequenz und damit die *erreichbare Grenzfrequenz* des Signalprozessors hängt von zwei Faktoren ab:

Taktfrequenz und Programmlänge.

Jeder Befehl benötigt *vier Taktzyklen*, und ein Programm darf *keine bedingten Sprünge* enthalten, damit seine Dauer immer gleich lang ist. Deshalb bestimmt die Anzahl der verwendeten Befehle die Dauer für einer Abtastwert. Werden alle 192 Befehle und die 10 MHz ausgenutzt, so sind immerhin noch Abtastfrequenzen von 13 kHz, also Signalfrequenzen bis zu ca. 6 kHz möglich. Mit kurzen Programmen sind also auch höhere Signalfrequenzen zu verarbeiten, mit 77 Befehlen z.B. bis ca. 10 kHz und mit 39 Befehlen bis ca. 20 kHz.

Den Befehlsaufwand für einige typische Operationen zeigt die folgende Aufstellung:

4-Quadranten-Multiplikation	12
4-Quadranten-Division	14
Filter mit einem komplexen Pol (oder Nullstelle)	7 bis 12
Sägezahngenerator	3 bis 7
Begrenzer oder Zweiweggleichrichter	1

Ungewöhnlich groß ist die Befehlswortlänge des 2920. Sie setzt sich systematisch aus fünf Teilen zusammen:

Operationsteil
Zieladresse
Quelladresse
Schiebewert
Analogbezug.

Dies ist wegen der Pipelinestruktur notwendig.

Die Anwendung des Signalprozessors ist sehr vielseitig. Das *Digitalfilter* ist eine Möglichkeit davon. Mittels Programm können am Ausgang beliebige Signalformen generiert werden, z.B. Sinus, Rechteck, Sägezahn, Treppenfunktion usw. So entsteht ein universeller *Signalgenerator* mit einer Auflösung von 9 bit. Weitere Anwendungen sind *Modulation* und Demodulation und *Signalbearbeitung* mit beliebigen Funktionen, also auch Multiplikation, Logarithmus, Exponentialfunktion usw.

Ein komplexes Beispiel möge dies demonstrieren: Bild 9.3.3b) zeigt den prinzipiellen Aufbau eines *Spektralmeßplatzes* mit dem 2920. Das Eingangssignal wird intern mit einem Tiefpaß begrenzt und automatisch auf den richtigen Pegel eingestellt. Weiter wird intern ein Sägezahngenerator mit etwa 10 Hz erzeugt. Dieser steuert einmal über einen Ausgang die Horizontalablenkung eines Oszillographen und zum anderen einen VCO, der dabei Frequenzen im Bereich von 4,8 bis 7,8 kHz realisiert. Aufbereitetes Eingangssignal und VCO-Ausgang gelangen auf einen Mischer (Multiplikator). Das folgende Schmalbandfilter hat eine Mittenfrequenz von 4,5 kHz bei einer Abtastrate von 13 kHz. So werden aus dem Eingangssignal die Frequenzen zwischen 300 Hz und 3kHz je nach Frequenz des VCO ausgefiltert. Der Vollweggleichrichter ist über Betragsbildung realisiert, und der Tiefpaß erzeugt das leicht geglättete Signal, das dem Vertikaleingang des Oszillographen zugeführt wird.

9.4. Sondertechnologien

9.4.1. Überblick

Die bisherigen Betrachtungen dieses Buches waren auf Grundlagen, Eigenschaften und Technologien von Bauelementen und deren Zusammenfassung zu Schaltungen ausgerichtet. Dabei stand in den beiden letzten Kapiteln die monolithische Technologie, bevorzugt die Si-Epitaxieplanartechnik, im Vordergrund. Sie hat u. a. Vorteile bezüglich der z. Z. höchstmöglichen Packungsdichte von Bauelementen und der Herstellung von praktisch allen aktiven und wichtigen passiven Bauelementen. Sie hat aber dabei den Nachteil, daß für jeden integrierten Schaltkreis extrem *hohe Vorleistungen* notwendig sind. Dadurch benötigt die *Herstellung* derartiger *IC* eine sehr große Stückzahl, die für viele Anwendungen nicht benötigt wird. Aus diesem Grunde wurden auch *andere Verfahren* der Herstellung von elektronischen Bauelementen entwickelt. Hierzu gehören u. a.

- *diskreter Aufbau* aus Einzelbauelementen oder Standardschaltkreisen,
- *Hybridschaltungen* auf der Basis von Dünn- bzw. Dickschichttechnologien.

Daneben ist es aus den verschiedensten Gründen aber notwendig — auch bei hochintegrierten Schaltkreisen — eine komplexe Schaltung aus vielen Bauelementen aufzubauen. Diese Bauelemente müssen dabei

mechanisch fixiert und elektrisch verbunden

werden. Dieses leistet in der Regel die *Leiterplatte*. Weiter können aus Stabilitätsgründen die Leiterplatten nur eine bestimmte Größe erreichen. So müssen oft für ein Gerät wiederum viele Leiterplatten mechanisch gehalten und elektrisch verbunden werden. Dies kann wiederum auf vielfältige Weise erfolgen.

Allein die bisherigen Betrachtungen zeigen *zwei Typen von Technologien*, die im folgenden etwas genauer behandelt werden sollen:

- Verbindungstechnologien
- Hybridschaltungen und ähnliches.

Leider ist zur Zeit die *Geschichte* der auch hier in den letzten Jahren erfolgten bedeutsamen Entwicklungen *wenig erforscht*. Einen ungefähren Überblick gibt Tab. 9.4.1.

Tabelle 9.4.1. Kurzer Abriß einiger Verbindungs- und Hybridtechnologien

1920	Freie Verschalung
1927	Telefunken nietet Leiterbahnen auf
1933	Fa. Hescho: gedruckte Schaltungen auf keramischem Substrat
1940	Eisler (England): metall-kaschiertes Material für Ätzung
1945	Erste Anwendung für Hörgerät
1950	Breite Anwendung der Leiterplatte
1952	Idee der integrierten Schaltung
1955	Wrap-Technik (Wickel-Verdrahten)
1961	Erste integrierte Schaltung
1962	Dickschichttechnik
ca. 1965	Durchkontaktierte, zweiseitig kaschierte Leiterplatte
ca. 1970	Dünnschichttechnik
1975	Multiwire-Technik
ca. 1975	Mehrlagenleiterplatte
1981	Mehrlagenkeramikmodule

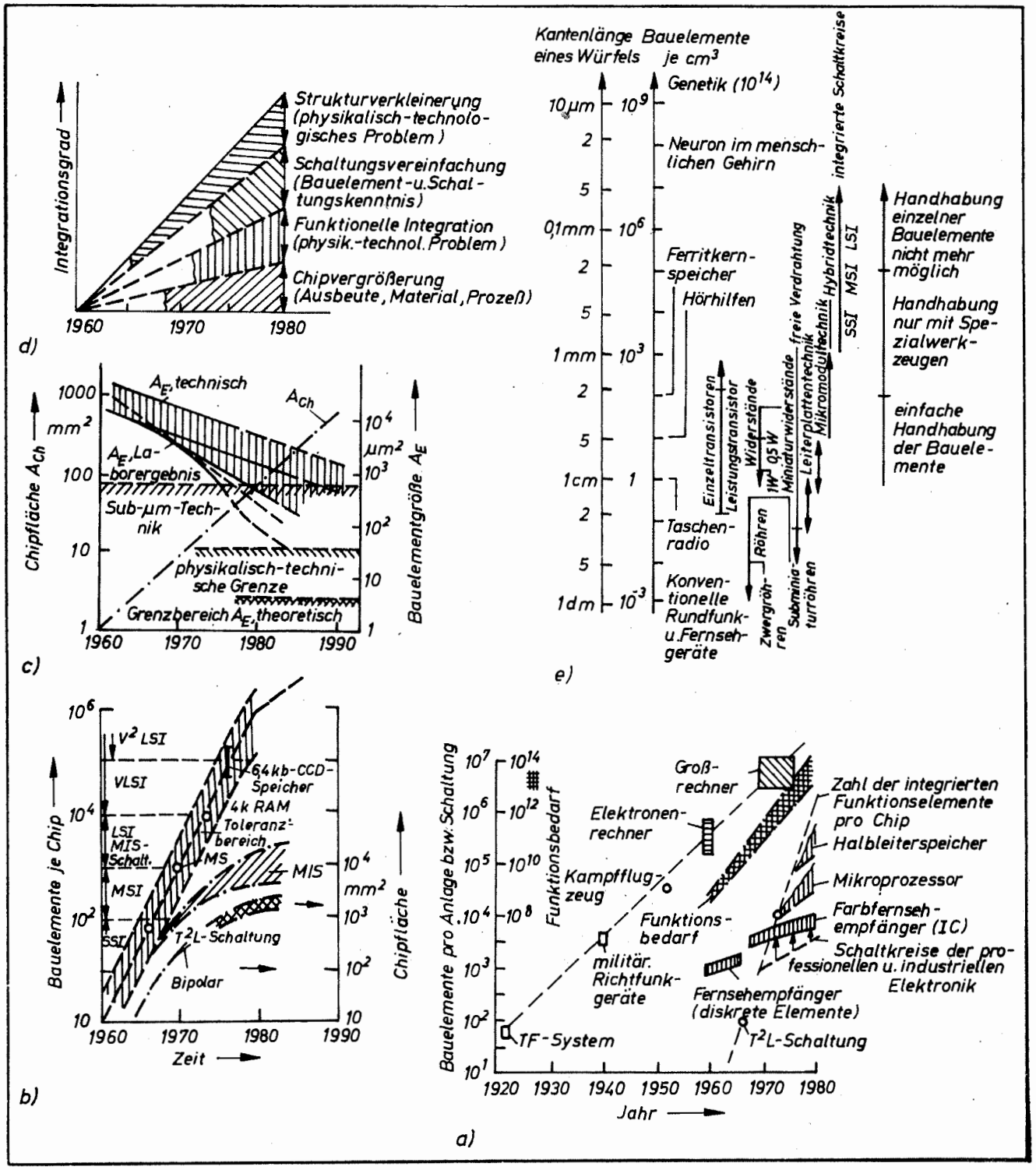


Bild 9.4.1. Zur Entwicklung der Bauelementeanzahl und -dichte nach verschiedenen Quellen

a) Bauelemente bzw. Funktionsbedarf für Anlagen, Schaltungen und Geräte
b) Bauelemente je Chip und Chipgröße
c) Chipfläche und Bauelementegröße
d) Unterschiedliche Anteile zur Steigerung des Integrationsgrades (schraffiert intensive Arbeiten zu den Teileinflüssen)
e) Bauelementedichten bei verschiedenen Technologien

Genauer ist dagegen die Entwicklung der Mikroelektronik untersucht. Hierzu faßt Bild 9.4.1 einige Aussagen zusammen. Da sie bewußt aus unterschiedlichen Quellen entnommen wurden, sind die Aussagen nicht kompatibel und zeigen so auch Streubereiche an. Im Teilbild e) sind auch Daten für im folgenden zu behandelnde Technologien vorhanden. Die *Packungsdichte* ist dabei stets geringer als bei der Mikroelektro-

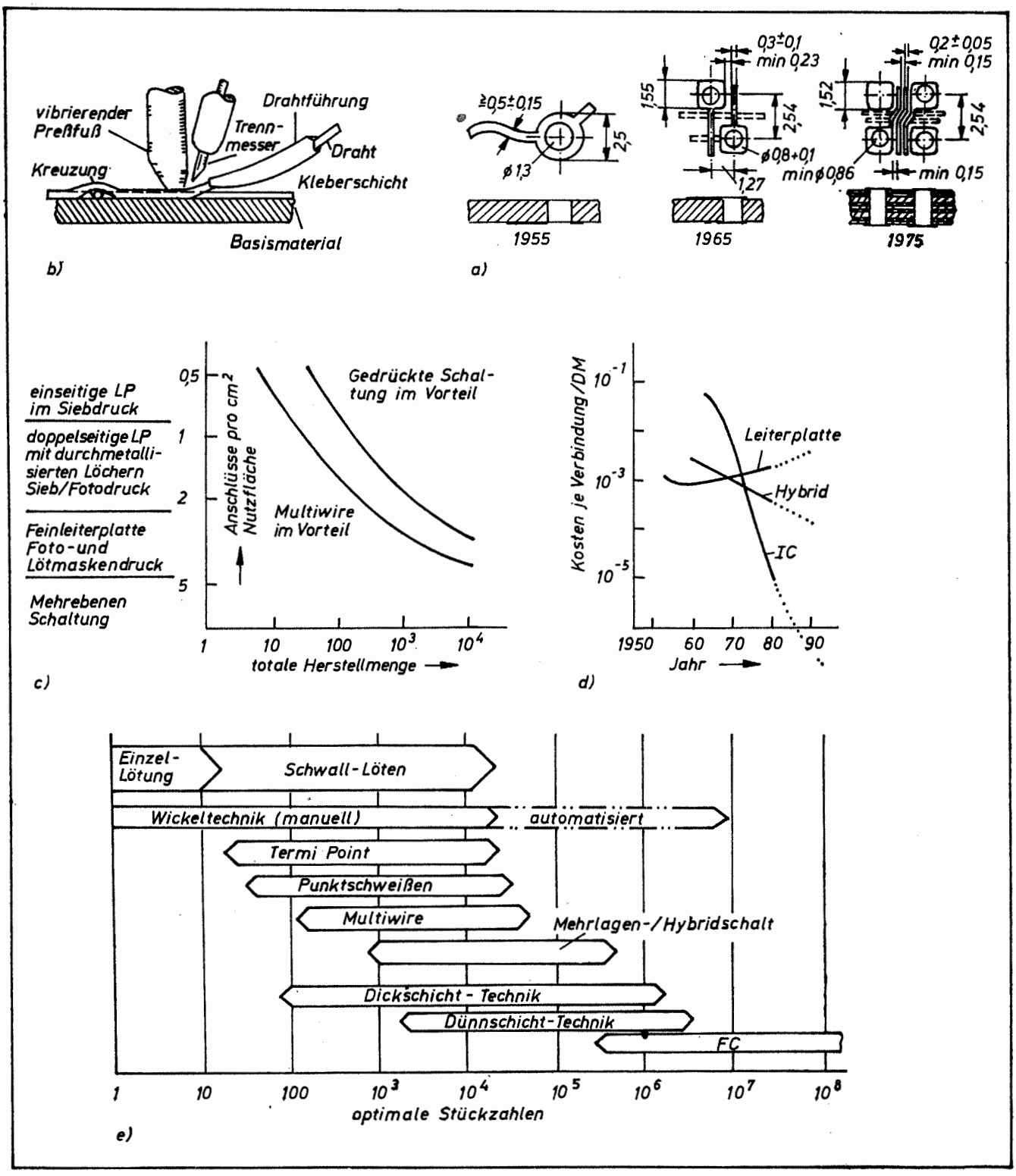


Bild 9.4.2. Zu den Verbindungstechniken

a) Ansteigen der Leiterdichte und -ebenenzahl bei Leiterplatten
b) Prinzip des Multiwiring
c) Die Vorteile des Multiwire liegen bei hoher Leiterdichte und geringer Stückzahl
d) Entwicklung der Kosten je Verbindung
e) Ungefähr optimale Stückzahlen bei verschiedenen Verbindungstechnologien (statt FC lies IC)

nik. Bei solchen Aussagen ist generell zu beachten, daß die Packungsdichte in der *Hierarchie des Aufbaus* stets erheblich abnimmt. Sind auf einem Chip ca. 10^7 Bauelemente je cm^3 möglich, so beträgt nach der Verkapselung die Dichte noch rund 10^5, bezüglich vieler solcher Bauelemente auf der Leiterplatte noch etwa 10^4 und in einem Gehäuse mit vielen Leiterplatten gar 10^3. Die Werte nehmen oft noch stärker ab.

Gründe hierfür sind u. a. die *notwendige Wärmeabfuhr*, die *Handhabbarkeit zur Reparatur* usw. Dabei ist zu beachten, daß in allen Hierarchieebenen im Laufe der Zeit die Packungsdichte erhöht werden konnte. Nicht immer, d. h. nicht für alle Anwendungen wird die höchste Packungsdichte verwendet. Hierbei sind vor allem ökonomische Faktoren, u. a. die produzierte Stückzahl, von Einfluß. Einige Ergänzungen hierzu gibt Bild 9.4.2. Im Teilbild a) ist zunächst die Entwicklung der *Leiterplattentechnik* in den letzten Jahrzehnten gezeigt. Um 1955 existierte die einseitig kaschierte Leiterplatte mit recht groben Leiterzügen und Lötaugen. Um 1965 entstand die doppelseitig kaschierte Leiterplatte mit wesentlich feineren Abmessungen und durchkontaktierten Lötaugen. Um 1975 wurde auch die Multiwire-Technik eingeführt. Hierbei werden gemäß Teilbild b) isolierte Drähte rechnergestützt in eine Klebeschicht eingebettet, wobei auch isolierte Kreuzungen in vielfältiger Hinsicht möglich sind. Ein grober Vorteil dieser Technologie ist die leichte Änderbarkeit und die effektive Fertigung von kleinen Stückzahlen bei hoher Leiterdichte. Dies weist deutlich Teilbild c) aus.

Die *Kosten einer Verbindung* hängen stark von der Technologie ab und ändern sich im Laufe der Zeit. Diese Entwicklungen weist Teilbild d) aus. Billiger werden vor allem die IC-Verbindungskosten, denn hier erfolgen die Verbindungen alle parallel mit oft nur einer Schablone. Weniger steil fällt die Hybridtechnik-Kurve, weil die Kosten bei der Leiterplatte infolge höherer Lohnkosten und Ansprüche zunehmen.

Den optimalen Einsatz verschiedener Verbindungstechniken als Funktion der zu produzierenden Stückzahlen zeigt Teilbild e). Hiervon sind drei Technologien bisher nicht genannt.

Die *Wickeltechnik* verwendet Wickelbolzen mit scharfen Kanten, um die in mehreren Windungen ein Draht mit so hohem Zug gelegt wird, daß Mikroverschweißungen auftreten. Dieses Verfahren hat gegenüber dem Löten viele Vorteile, wie z. B. einfachste Handhabung bei leichter Änderbarkeit und großer Freiheit in der Lage der Verbindung.

Das *Thermo Point* ist ein spezielles Bonden, während das *Punktschweißen* die Kontaktierung durch hohen Stromfluß bewirkt. Schließlich seien noch beispielhaft einige pauschal relative Daten zu drei Technologien gegeben:

	Gedruckte Schaltung	Schicht-schaltung	Monolith. Schaltung
Volumen	100	5	1
Fläche	100	10	1
Masse	100	20	1
Mindest-stückzahl	1	100	10000
EW-Kosten	1	5	50
EW-Dauer	1	2	3

Tabe. 9.4.2 ergänzt detailliertere Daten.

Eine der möglichen Einteilungen für die verschiedenen Technologien zur miniaturisierten Herstellung von Schaltungen zeigt Bild 9.4.3.

In der *Miniaturtechnik* werden ausschließlich übliche Bauelemente wie Widerstände, Kondensatoren, Induktivitäten und Halbleiter verwendet und zu speziellen Blocks zusammengefaßt. In ihrer speziellen Form als *Subminiaturtechnik* werden bis zu etwa 10 diskrete aktive und passive Bauelemente in dichter Packung zu einem Block von ca. 1 bis 2 cm Kantenlänge zusammengefaßt. Innerhalb des Blockes sind sie verschaltet. Zur mechanischen Stabilität und zum Schutz gegenüber äußeren Einflüssen sind sie in Kunstharz eingegossen. Der Block benötigt relativ weniger Fläche

Tabelle 9.4.2. Typische Herstellungsverfahren, Materialien und Abmessungen der Mikroeektronik

	Gedruckte Leiterplatte	Dickschichttechnik	Dünnschichttechnik	Halbleiterblocktechnik
Typisches Herstell-verfahren	Foto-, Siebdruck Ätztechnik	Siebdruck	Vakuumbedampfung oder Katoden-zerstäubung (Sputtern) mit Masken- oder Ätztechnik	Epitaxieplanartechnik
Grundmaterial	kaschiertes oder freies Basis-material	Keramik	Glas, Saphir, Keramik	Siliciumwafer
Gut herstellbare Bauelemente	Leitungen	Leitungen: 25 mΩ/□ Widerstände 100 Ω ... 500 kΩ bis 15 W/cm² 10 Ω ... 1 MΩ/□	Leitungen: 50 mΩ/□ Widerstände 10 Ω ... 0,1 MΩ 10 ... 1000 Ω/□ Kondensatoren ca. 0,1 μF/cm² 10 pF ... 50 μF	Leitungen uni- und bipolare Transistoren Dioden Widerstände 15 Ω ... 50 kΩ mit FET 10^3 ... 10^{10} Ω Kondensatoren 500 pF
Bedingt herstellbare Bauelemente	Spulen Kondensatoren	Spulen 10 μH Kondensatoren 10 pF ... 5 μF (20 μF 3,5 μF/cm² bei 20 V	Induktivitäten 1 μH	—
Abmessungen Materialstärke Schichtstärke	dm² 1 ... 2 mm 35 (70) μm	1 ... 30 cm² 0,5 ... 1 mm 15 ... 25 μm	0,5 .. 5 cm² 0,1 ... 1 mm 10 ... 50 nm	2 ... 30 mm² 0,1 ... 0,3 mm Leitbahnen 0,5 μm Diffusionstiefen 2 ... 8μm
Strukturbreite Randschärfe Bauelemente je cm² Leistungsdichte in mW/mm²	0,1 ... 2 mm ±10 μm < 1	0,1 ... 0,5 mm +5 μm ≦ 5 16 auf Keramik	10 ... 100 μm ±2 μm 10 ... 50 3 auf Glas	1 ... 25 μm ±0,1 μm > 1000 LSI ca. 10^5 30 auf Si

auf der üblichen Druckplatine und erspart zugleich einige Leitungszüge auf der Platine. Solche Blöcke sind meist immer teurer als die Summe der in ihnen enthaltenen Bauelemente. Auch die Zuverlässigkeit erhöht sich nicht gegenüber den Einzelbauelementen. Heute werden bis auf ganz wenige Ausnahmen derartige Blöcke nicht mehr produziert.

Die *Modul-* bzw. *Mikromodultechnik* entstand um 1950 bzw. Ende der fünfziger Jahre in den USA. Als Grundeinheit geht sie von quadratischen Keramikplättchen mit ca. 25 bzw. 8 mm Kantenlänge und einigen zehntel Millimetern Dicke aus. Jedes Plättchen trägt ein bis wenige, z. T. speziell hergestellte, z. T. übliche Bauelemente und hat an den Kanten ringsherum Nuten zur Verbindung. Die Plättchen werden übereinander gestapelt und durch senkrechte Verbindungsdrähte in den Nuten verschaltet und anschließend vergossen. Diese Technik konnte sich nur zeitweilig in militärischen Geräten behaupten. 1979 führte IBM in Weiterentwicklung dieser Techniken die *Mehrschichtkeramikmodule* für den Rechner 4300 und 1981 eine Weiterentwicklung für den Rechner 3801 ein. In beiden Fällen werden 23 bzw. 33 quadrati-

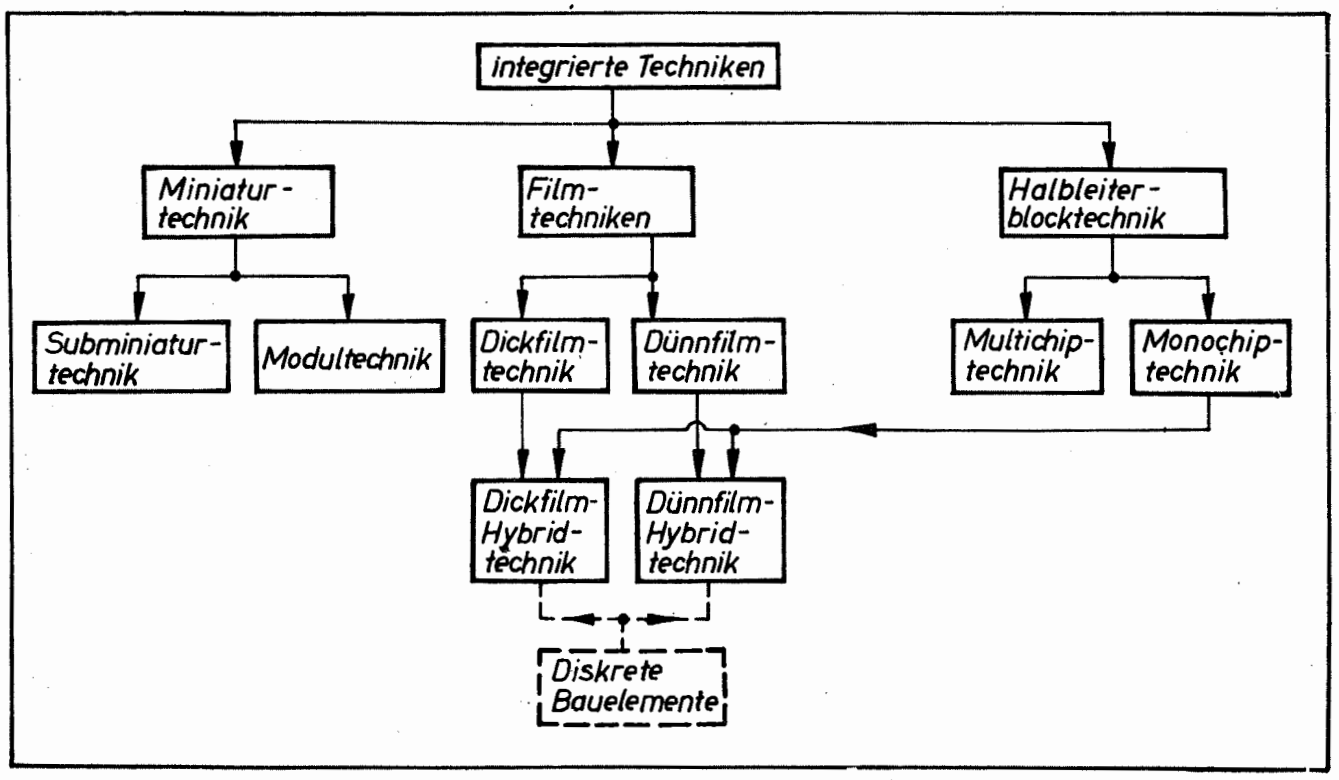


Bild 9.4.3. Einteilung der Haupttechnologiezweige für die miniaturisierte Herstellung von Schaltungen

sche Keramiklagen von 50 bzw. 90 mm Kantenlänge übereinander angeordnet. Dadurch können bis zu 118 Chips untergebracht werden, und die Wärme (bis zu 500 W) kann an eine in Wasser gehüllte Aluminiumplatte abgeführt werden (thermal conduction modul).

Die *Halbleiterblocktechnik* führt zu den integrierten Schaltkreisen, die bereits ausführlich besprochen sind. Wird nur ein Chip in ein Gehäuse gefügt, liegt die Standardtechnologie vor. In Sonderfällen können aber auch mehrere Chips in ein Gehäuse eingebaut und verdrahtet werden. Die e Multichiptechnik hat vor allem dann Bedeutung, wenn spezielle Schaltungen auf einfache Weise schnell und preiswert hergestellt werden sollen. Da sie gewisse Ähnlichkeit mit den noch zu behandelnden Hybridtechniken besitzt, sei sie hier nicht weiter behandelt.

Die *Filmtechniken* können auf der Basis der Dünn- bzw. Dickschichttechnik realisiert werden. Beide werden noch ausführlich behandlet, so daß hier ein grober Überblick genügt.

Bei der *Dickschichttechnik* (9.4.4.) werden spezielle Leit-, Widerstands- und Isolationspasten im Siebdruck auf eine Unterlage gedruckt und durch Brennen, d. h. bei stark erhöhter Temperatur, fixiert und gehärtet. Hierbei sind also Leitungen, Widerstände und z. T. Kapazitäten und Induktivitäten herstellbar. Weitere passive und aktive Bauelemente bis zu Schaltkreisen werden dann in diese gedruckten Schaltungen nachträglich eingefügt.

Die *Dünnschichttechnik* (9.4.5.) entspricht bis auf den ersten Schritt völlig der Dickschichttechnik. Hier werden Leiterbahnen, Widerstände und z. T. Kapazitäten im Vakuum durch Sputtern bzw. Aufdampfen hergestellt. Zur Strukturierung werden Masken bzw. über Fotolithographie Ätzungen angewendet. Erst in den letzten Jahren ist es teilweise gelungen, spezielle Dünnschichttransistoren bei diesen Prozessen zu realisieren. Sie besitzen aber immer noch schlechte Eigenschaften.

9.4.2. Substrate

Die Substrate der Dick- und Dünnschichttechnik haben vielfältigen, z. T. widersprüchlichen Anforderungen zu genügen. Unter anderem sind dies:

- Die Oberflächenrauhigkeit soll klein (kleiner als die Schichtdicke) sein.
- Die Wärmeleitfähigkeit soll hoch sein.
- Der Wärmeausdehnungskoeffizient soll den aufzubringenden Schichten möglichst nahe sein.
- Der Isolationswiderstand soll hoch sein.
- Die HF-Verluste sollen gering sein.
- Gegenüber Ätzmaterialien soll das Substrat resistent sein.
- Das Material soll billig sein.

Hauptsächlich werden folgende *Substrate* verwendet:

Gläser existieren in vielen Modifikationen. Sie zeichnen sich durch eine sehr glatte Oberfläche bis zu 5 nm (feuerpoliert) aus. Störend kann bei ihnen eventuell der Gehalt an Alkalimetallen, insbesondere Na, sein. Deswegen existieren Spezialgläser. Nachteilig bei Glas ist die geringe Wärmeleitfähigkeit. Für die Dickschichttechnik bestehen Probleme bezüglich der Hochtemperaturbrennzyklen. Deshalb ist heute Glas vor allem bei der Dünnschichttechnik das geeignete Substrat.

Einen wichtigen Platz für viele Anwendungen nimmt die *Al_2O_3-Keramik* (94 bis 99,6% Al) ein. Sie wird aus Al_2O_3- und Glaspulver bei ca. 1500 °C gesintert. Für die Beschaffenheit des Substrates ist folglich u. a. die Korngrößenverteilung wesentlich.

Tabelle 9.4.3. Kennwerte wichtiger Substrate für die Dünn- und Dickfilmtechnik

Material (Handelsname)	Zusammensetzung %	Thermische Leitfähigkeit 10^{-3} Wcm^{-1}K^{-1} bei 25 °C	Dielektrizitätskonstante	Erweichungstemperatur °C
Alkaliglas (Mikroskopgläser) (Corning 0080)	SiO_2 (72) CaO (9) Na_2O (16)	9,5	7,2	695
Borsilikatglas (Pyrex) (Corning 7740)	SiO_2 (80) B_2O_3 (13) Al_2O_3 (2) NaOH (3)	11	4,6	820
Aluminium-Silikatglas (Corning 7059)		12,5	5	872
Quarzglas (Quartex)	SiO_2 (100)	14	3,8	1670
Glas-Keramik (Fotoceram)		23	5,6	
Aluminiumoxid gesintert	Al_2O_3 (94)	190	8,9	≈1600
Berylliumoxid	BeO (99,5)	2300	6,4	1600
Magnesium-Silikatkeramik (Steatit)		25	6,3	
Aluminiumoxid kristallin (Saphir)	Al_2O_3 (100)	420	7,5 ... 10	1800

Die Standardrauhigkeit liegt bei 1 μm und kann bis zu 0,25 μm durch Polieren gesenkt werden. Hierdurch treten aber andere Oberflächenstörungen, z.B. Mikrorisse, auf, so daß dieser Weg selten begangen wird. Häufiger wird statt dessen eine z. T. partiale Glasierung (ca. 50 μm) durchgeführt. Sie löst zugleich zwei Probleme: Sie macht die Oberfläche glatt und die Keramik beständig gegenüber Ätzmaterialien.

Ein sehr teures, aber am besten wärmeleitendes Material ist *BeO*. Es hat vor allem für Leistungsschaltungen Bedeutung. Nachteilig ist sein sehr giftiger Staub, der bei mechanischen Bearbeitungen auftritt.

Hohen Ansprüchen bezüglich Oberfläche, Isolationswiderstand und HF-Verlusten genügt *Saphir*, also einkristallines Al_2O_3. Es wird vor allem im Mikrowellenbereich eingesetzt.

In der letzten Zeit erlangen *Folien* aus *Polyamid* gewisse Bedeutung.

Einen *Überblick zu den Eigenschaften* von Substraten gibt Tab. 9.4.3.

Vor der Beschichtung sind die Substrate zu reinigen. Besonders bei der Dünnschichttechnik ist hier ein hoher Aufwand zu treiben.

9.4.3. Spezialbauelemente

9.4.3.1. Überblick und Standardmethoden

Mit den Schichttechnologien können — wie noch genauer gezeigt werden wird — Leiter, Widerstände und in gewissem Umfang Kapazitäten, Induktivitäten und Dünnschichttransistoren hergestellt werden. Dies bedeutet, daß auch spezielle Bauelemente in diese Schaltungen eingefügt werden müssen. Sie haben dabei einige spezielle Forderungen zu erfüllen, u. a. sind dies:

- kleine Abmessungen,
- Lage der Anschlußelemente in einer Ebene,
- günstige Form für automatische Bestückung,
- Verträglichkeit mit Substrat-, Umhüll-, Fluß-, Wasch-, Klebe- und Passivierungsmitteln und -materialien,
- Beständigkeit gegen Ultraschallreinigung,
- Temperaturfestigkeit bezüglich Kontaktieren, Umhüllen und eventuellem Reparaturlöten.

Einzelwiderstände können aus folgenden Gründen notwendig sein: Extrem große Werte, Rauscharmut, Vermeidung von Leitungskreuzungen. Vor allem werden Dickschichtwiderstände verwendet, die an den Enden so metallisiert sind, daß sie direkt auf das bearbeitete Substrat aufgeklebt und dann kontaktiert werden können.

Kondensatoren sind meist effektiver, da in den Schichttechnologien nur kleine Kapazitätswerte hergestellt werden. Verwendet werden meist Multiplayer-Keramikkondensatoren (1.5.5.2.) und Tantalkondensatoren. Sie sind größtenteils quaderförmig und an den Enden kontaktiert.

Für *Dioden* existiert noch keine spezielle Form.

Für *Transistoren* und *IC* werden heute viele Varianten verwendet. Sie sind in Bild 9.4.4 zusammengestellt.

In Teilbild a) ist ein *nackter Chip* auf dem Substrat mit Leitkleber oder eutektisch befestigt und dann über die Pad zur Leiterstruktur verbunden. Hierdurch wird ein guter Wärmekontakt erreicht. Ferner ist die benötigte Fläche extrem klein. Nachteilig sind jedoch die beiden Bondstellen je Anschluß (vgl. Bild 4.5.3). Weiter ist nachteilig, daß Chips nur in wenigen Parametern gemessen vom Hersteller geliefert werden können und daß vereinzelte Chips doch recht empfindlich (z.B. Kratzer) und schwierig zu handhaben sind. Zuweilen werden die Wafer geritzt und auf eine elastische

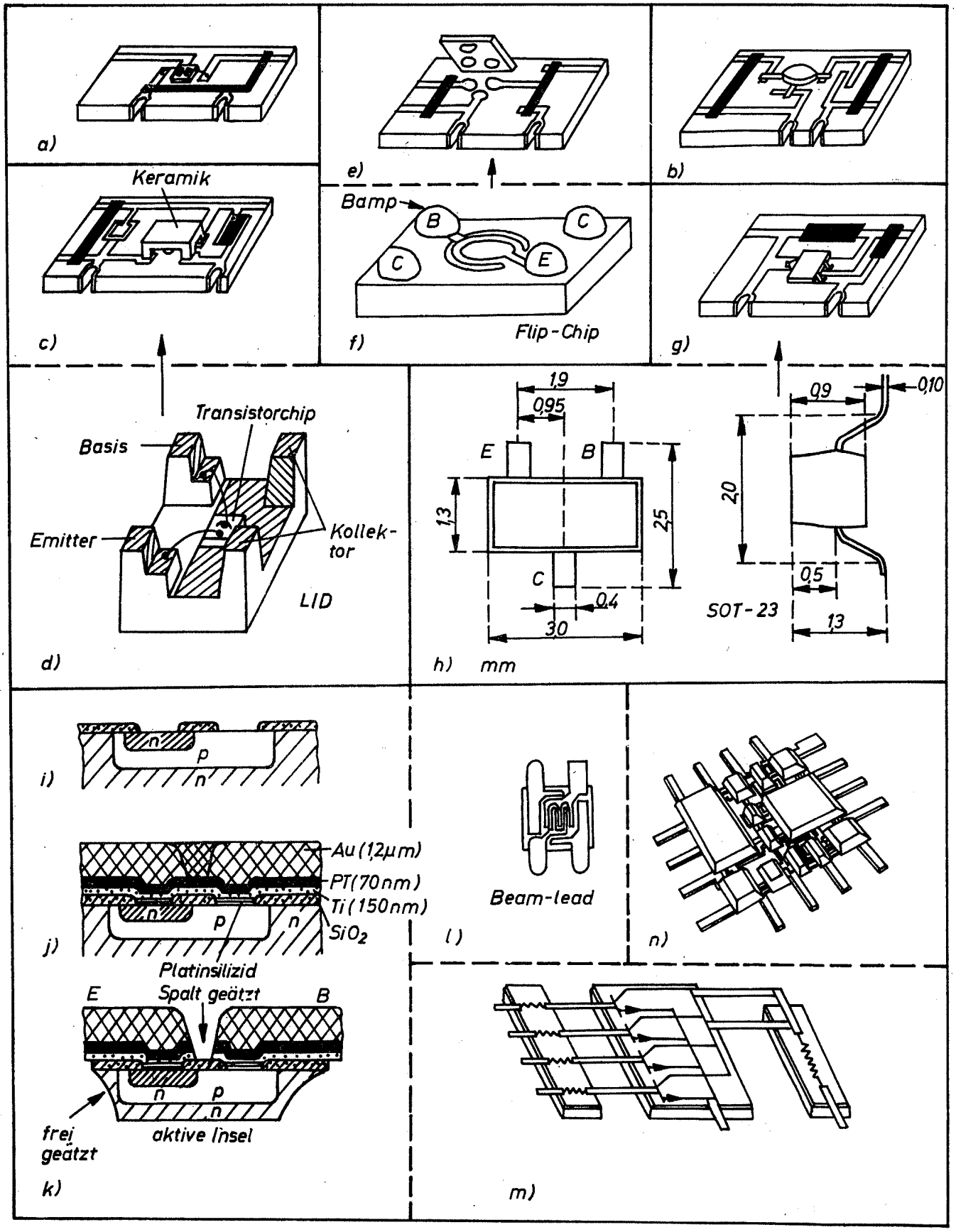


Bild 9.4.4. Aktive Bauelemente der Hybridtechnik

a) Prinzip des Einfügens von nackten Chips
b) Einfügen von Streifentransistoren
c) LID-Einbau und d) Aufbau eines LID
e) Flip-Chip-Einbau und f) Aussehen eines Flip-Chip
g) Einfügen von Mikrogehäusen
h) Abmessungen des SOT-23
i) bis k) Herstellung eines Beam-lead-Transistors
l) Aussehen eines Beam-lead-Transistors
m) und n) Komplexe Beam-lead-Strukturen

Folie geklebt geliefert. Dann ist es möglich, über die Nachbarschaft der einzelnen Chips möglichst gleichartige Bauelemente auszuwählen.

Im Teilbild b) werden die *Streifentransistoren* der HF-Transistoren eingefügt. Nach dieser Methode werden auch die weiter unten zu behandelnden Beam-lead-Bauelemente eingefügt.

Gemäß Teilbild d) wird der Transistor, die Diode oder andere Bauelemente in einen speziellen Keramikträger eingefügt, das *LID* (leadness inverted device), und dieses wird dann mit dem „Gesicht nach unten" (face down) in die Schaltung eingelötet (Teilbild c)). Der Vorteil ist, daß so das Bauelement in dieser Weise geschützt ist. Nachteilig ist die schlechte Wärmeableitung und die Schwierigkeit, zu erkennen, ob alle Kontakte beim Einlöten gut sind.

Ein häufig verwendetes Prinzip ist das *Flip-Chip* (Teilbild f)). Hierbei werden die Pads von Chips oder äquivalente Strukturen mit halbkugelförmigen Höckern, den Bumps, versehen. Sie besitzen einen Durchmesser von ca. 75 bis 160 µm und eine Höhe von 20 bis 100 µm. Sie bestehen je nach der Art des Aufbringens (Ultraschallkontaktieren, Löten usw.) aus verschiedenen Materialien, z.B. vernickelt oder aus Pb-Sn. Ein Vorteil dieser Methode besteht darin, daß alle Chips relativ einfach zu solchen Flip-Chip-Bauelementen aufbereitet werden können. Ungünstig ist wieder die Wärmeableitung infolge fa ce-down.

Als spezielles *Kunststoffgehäuse* hat sich das SOT-23 bzw. mit etwas anderen Abmessungen das Mikro-E weitgehend durchgesetzt. Nach Kontaktierung des Transistors wird er mit den drei speziell ausgerichteten Beinen in Kunststoff eingepreßt. Da diese Elemente vorher vollständig ausmeßbar sind, können die vorgegebenen Daten garantiert werden. Außerdem ist die Form gut für die Automatisierung geeignet. Nachteilig ist eine nicht sehr hohe mechanische Stabilität, so daß es zuweilen zu mechanischen Brüchen des Bauelements kommt.

Eine gewisse Anwendung besitzen auch die in *Flat-pack-Gehäuse* eingebauten integrierten Schaltkreise. Sie lassen sich z. T. recht gut in die Schichtschaltungen einfügen.

9.4.3.2. Beam-lead

Die Beam-lead-Technik ist sehr aufwendig und hat viele Vorteile. Auf einem Transistor (Bild 9.4.4i)) wird zur Kontaktierung eine Ti-Schicht aufgedampft. Sie gewährleistet sperrfreie Übergänge. Zur Verhinderung der Diffusion von z.B. Goldatomen in das Si folgt dann eine dünne Platinschicht. Dabei kann eventuell Pt in Si diffundieren und es entsteht Platinsilizid. Hierauf folgt eine etwa 10 bis 20 µm dicke Goldschicht. Sie bildet später den mechanisch stabilen Anschluß. Abschließend werden in die Leiterstruktur die notwendigen Spalte zur Isolierung der Anschlüsse geätzt. An den Seiten wird zur Freilegung der Kontaktbahnen auf einer Länge von einigen hundert µm ein Teil der Siliciumfläche — ähnlich wie beim Mesatransistor — freigeätzt. So entsteht schließlich ein Transistor, wie ihn Bild 9.4.4l) zeigt. Er wird, ähnlich wie in Teilbild b) und g), in die Schaltung eingefügt.

Mit der Beam-lead-Technik lassen sich aber auch kompliziertere Strukturen, wie sie Teilbild m) und n) zeigen, herstellen. Es kann also eine ideale Isolierung im Gegensatz zur üblichen *pn*-Insel bei den Standardtechnologien erreicht werden. In diesem Sinne war die Beam-lead-Technik zeitweilig eine Konkurrenz für die SOS-Technik.

9.4.3.3. TFT

Thin-film-transistors (TFT) sind Feldeffekt-Transistoren, die mit den Mitteln der Dünnschichttechnik direkt hergestellt werden können. Da auf dem Substrat, meist

Corning-Glas 7056, eventuell aber auch Keramik ausreichender Glätte, keine Halbleiterkristalle erzeugt werden können, muß mit polykristallinem Material gearbeitet werden. Dies hat Nachteile in den Kennlinien und Daten zur Folge. Weiter sind mit der Dünnschichttechnik auch nur gröbere Strukturen als bei der monolithischen Technik zu realisieren. Die Herstellung der TFT erfolgt mittels Drahtmasken (Bild 9.4.5) oder Fotolithographie. Als Halbleitermaterial kommen zur Anwendung CdSe, CdS, Te, CdTe, InAs, PbS, SnO_2, Ge oder Si. Als Gateisolation wird SiO_x bzw. Al_2O_3 verwendet. Die Dickte des Halbleiters liegt bei 300, in den Grenzen von 10 bis 1000 nm, die Kanallänge bei 50 (10 ... 100) µm, die Kanalbreite bei 0,5 (0,2 bis 2) mm und die Dicke des Gateoxids bei 100 (30 ... 400) nm. Zur Verbesserung der Daten werden die TFT zuweilen thermisch (und in spezieller Atmosphäre) nachbehandelt.

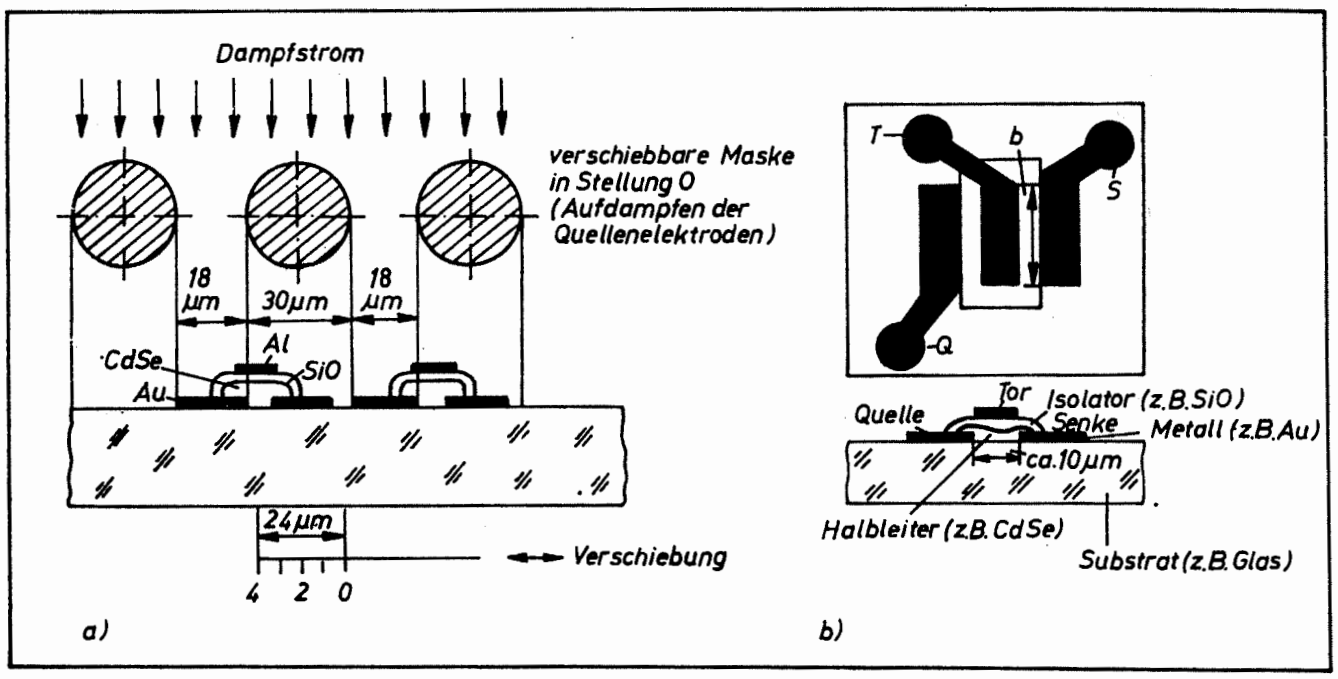


Bild 9.4.5. a) Prinzip der Herstellung eines TFT über Gittermasken und deren Verschiebung
b) Oberflächenaussehen eines TFT

9.4.4. Dickschichttechnik

Die Dickschichttechnik wird zuweilen auch als "screen and fire" gemäß ihrer Haupttechnologien, dem Siebdruck- und dem Brennverfahren bezeichnet. Im Prinzip wurde diese Technologie bereits in den dreißiger Jahren versucht. Aus Silber und Glaspulver unter Zusatz von Binde- und Lösungsmitteln entstand eine druck- oder streichfähige Masse. Sie konnte auf keramische Materialien aufgetragen und dort zu Leiterbahnen eingebrannt werden. Der entscheidende Fortschritt gelang um 1962, als spezifische Halbleiterbauelemente und neue Pasten für Widerstände vorlagen.

9.4.4.1. Pastensysteme

Für die Dickschichttechnik werden *drei Sorten von Pasten* benötigt, und zwar für

Leitbahnen
Widerstände
Dielektrika für Kondensatoren.

Jede dieser Pasten (auch zuweilen Cermet genannt) besteht aus *jeweils drei Komponenten*:

Metallen, Oxiden oder Dielektrika je nach Zweck
glasartigen Bindern
organischer Flüssigkeit.

Da die Pasten zum Absetzen neigen, müssen sie gut verrührt und in günstiger Viskosität verwendet werden.

Die ersten Pastensysteme (und viele heutige) verwenden Edelmetallsysteme. Infolge des hohen Preises geht die Entwicklung *zu unedlen Systemen* über (Tab. 9.4.4.) Sie müssen aber meist in einer nichtreaktiven Schutzatmosphäre gebrannt werden. Die Qualität aller Pasten hängt stark von der Korngröße und deren Verteilung sowie von den Brennbedingungen ab. Sie müssen deshalb z. T. sehr eng eingehalten werden. Die genaue Herstellung der Pasten und ihre Zusammensetzung ist Werksgeheimnis der Hersteller.

Tabelle 9.4.4. Die beiden Dickschichtpastensysteme

	Edelmetallsysteme			Unedelmetallsysteme		
	Leiterbahn	Widerstandsbahn	Dielektrikumsschicht	Leiterbahn	Widerstandsbahn	Dielektrikumsschicht
Funktionelle Bestandteile	Au, Pt/Au, Ag, Pd/Ag	Pd/Ag $Bi_2Ru_2O_7$, RuO_2	$BaTiO_3$, Glas, Glas-Keramik	Cu, Ni, Al	LaB_6, Ta (+ Zusätze)	Glas, Keramik
Permanente Bindemittel	*Gläser*: Borsilikatgläser, Aluminiumsilikatgläser *Oxide*: CuO, CdO, Glas-Oxidsysteme			*Gläser*: Calciumborsilikatgläser		*Oxide*: CeO_2, PrO_2, CrO_2
Organische Bindemittel	*leichtflüchtige Lösungsmittel*: Terpineol, Butylkarbitolacetat (BCA) *schwerflüchtige Bindemittelanteile*: Äthylzellulose			*leichtflüchtige Lösungsmittel*: Terpineol, Butylkarbitolacetat (BCA) *schwerflüchtige Bindemittelanteile*: Äthylzellulose		

9.4.4.2. Herstellungsprozeß

Ein zentraler Prozeß der Dickfilmtechnik ist das *Siebdruckverfahren*, wie es im grafischen Gewerbe seit langem verwendet wird. Als Siebe kommen Fäden aus Nylon, Polyester oder rostfreiem Stahl von etwa 0,03 bis 0,07 mm Durchmesser zur Anwendung. Die Maskenweite der Siebe liegt bei 80 Fäden/cm. Diese Siebe werden mit Fotolack geschichtet, dann mit den zu erzeugenden Strukturen belichtet. Nach der Entwicklung des Fotolacks sind im Sieb dann jene Stellen durchlässig, wo gedruckt werden soll.

Gemäß Bild 9.4.6a) befindet sich das zu bedruckende, gut gereinigte Substrat in einem Rahmen, über dem das Sieb ausgespannt ist. Darauf befindet sich die Paste, und ein Rakel drückt bei seiner Bewegung die Paste durch die offenen Stellen auf das Substrat. Teilbild b) zeigt, daß die aufgetragene Dicke von der Weite der Öffnungen abhängt. Nach dem Druckvorgang muß die Paste langsam trocknen, weil sonst beim Brennen Blasenbildungen die Schicht zerstören könnten. Das Brennen muß nach ei-

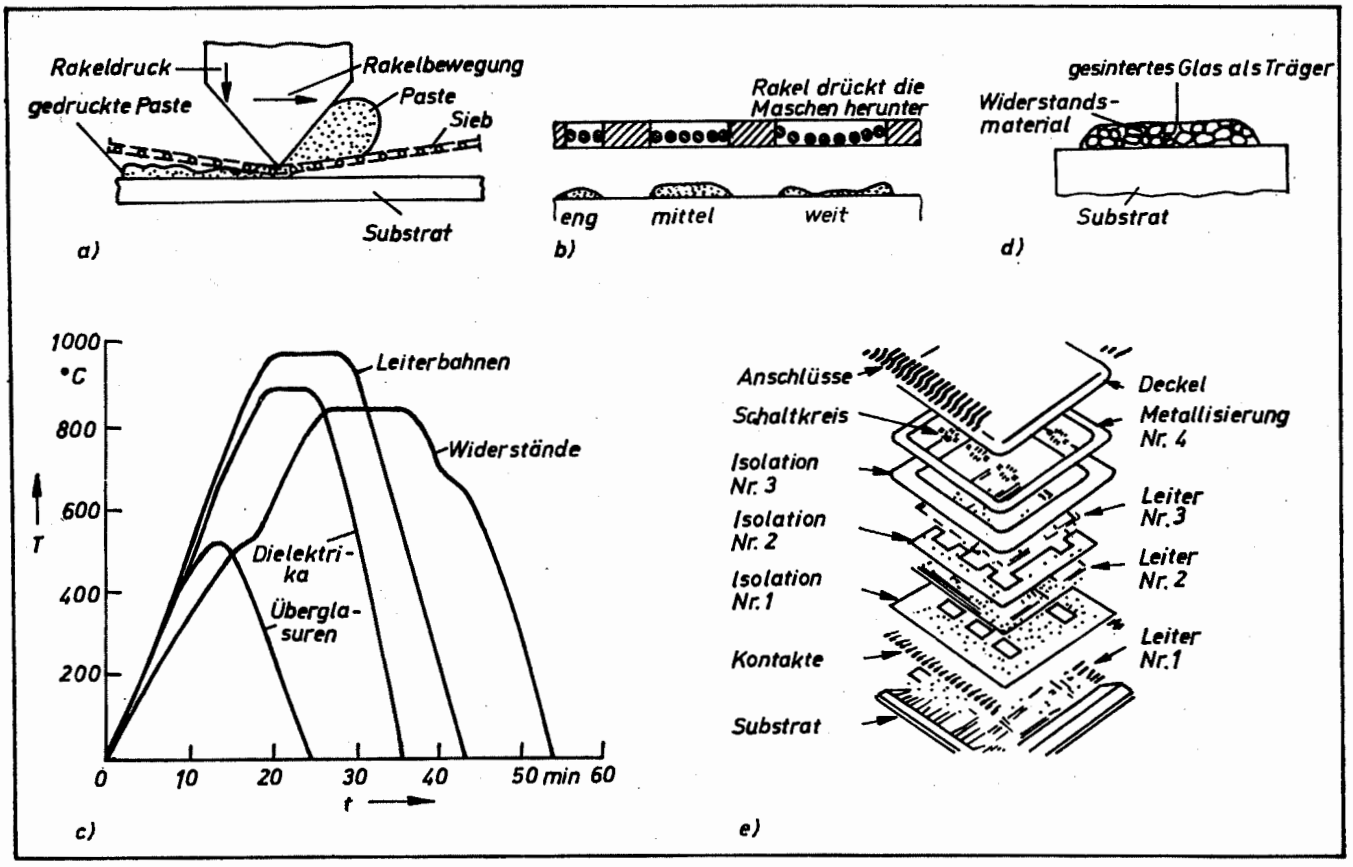


Bild 9.4.6. Zur Dickschichttechnik

a) Prinzip des Siebdrucks
b) Auftragsdicke als Funktion der Weite
c) Beispiele von Temperaturprofilen
d) Prinzipieller Aufbau eines Dickschichtwiderstandes
e) Maskenfolge bei Herstellung einer Schichtschaltung

nem streng einzuhaltenden Temperaturregime erfolgen. Bild 9.4.6c) zeigt dies beispielhaft für einige Pasten. Hieraus wird die Abstimmung der Pasten innerhalb eines Pastensystems deutlich. In der Regel sollte die jeweils als nächste folgende Schicht möglichst bei etwas niedrigerer Temperatur gebrannt werden. Bild 9.4.6d) zeigt schematisch, wie in etwa eine gebrannte Widerstandsschicht aussieht. Die einzelnen Körner des Widerstandsmaterials berühren sich und sind durch den Glasbildner mechanisch kompakt verbunden. Die fertigen Schichten besitzen eine *Dicke um 10 bis 30* µm. Den Ablauf der Herstellung einer Schichtschaltung zeigt Bild 9.4.6e). Die Prinzipien der Strukturierung enthalten u. a. die Bilder 5.1.1, 5.2.1 und 5.3.1.

9.4.4.3. Eigenschaften der hergestellten Bauelemente

Die verschiedenen *Widerstandspasten* erzeugen bei ca. 25 µm Schichtdicke je nach der Rezeptur Flächenwiderstände von 1 bis 10^6 Ω/□. Damit sind Widerstände von ca. 10 bis 10^8 Ω herstellbar. Der Temperaturkoeffizient liegt im Bereich von ±50 bis ±500 × 10^{-6}/K. Die Toleranz der unmittelbar hergestellten Widerstände liegt bei ca. 10%, kann aber durch Abgleich (9.4.6.) auf ±0,5% eingeengt werden. Die Belastbarkeit von Widerständen auf Keramik liegt bei rund 2 W/cm².

Die *Leitschichten* erreichen Flächenwiderstände von etwa 3 bis 100 Ω/□. Der Widerstand wird zuweilen durch Verzinnen weiter verringert.

Als *Isolierschichten* werden in der Regel zwei Klassen unterschieden: die NDK (1.5.5.2.), zu denen hier auch das Glas gezählt wird, und die HDK. Je nach diesen beiden Typen werden Flächenkapazitäten von 100 bis 400 pF/cm² bzw. 15 bis 40 nF/

cm^2 erreicht. Der herstellbare Kapazitätsbereich liegt daher bei 10 bis 1000 pF bzw. 0,1 bis 10 nF. Die Toleranz der Kapazitätswerte liegt bei etwa 10% und mit Abgleich, insbesondere bei den NDK, bei ca. 1%. Die Durchschlagsspannung beträgt einige hundert Volt.

Induktivitäten können bis zu einigen μH hergestellt werden. Ihre Güte ist schwer reproduzierbar zu halten und erreicht nur selten Werte bis zu 100, bei 30 MHz z.B. max. nur 15.

9.4.5. Dünnschichttechnik

Die Dünnschichttechnik erfolgt vorwiegend im Rahmen der Vakuumtechnologie nach vor allem *zwei Verfahren*:

- Aufdampfen
- Sputtern (Katodenzerstäuben).

Daneben existieren noch in gewissem Umfang chemische Technologien.

Unabhängig beherrschen *zwei Schichtsysteme*, insbesondere für die Widerstände, die Technik:

- NiCr-Schichten
- Tantalschichten.

Beide werden ergänzt durch Leitermaterialien, u. a. Cu, Al, Ag, Au, und Isolierschichten, insbesondere auf der Basis von SiO, SiO_2 und Ta_2O_5.

9.4.5.1. Aufdampftechnik

Bei der Aufdampftechnik (Bild 9.4.7a)) wird das Material, welches strukturiert auf das sehr gut gereinigte Substrat übertragen werden soll, im Abstand von einigen cm im Hochvakuum so stark erhitzt, daß es verdampft und auf diese Weise sich geradlinig auf das Substrat zubewegt und dort niederschlägt. Die Materialstärke ist dabei proportional dem cos Φ, so daß stets in der Mitte eine größere Schichtdicke auftritt. Dieser Effekt kann nur durch kleinere Substrate bzw. größere Abstände verringert werden. Dies bedeutet aber, daß sich viel verdampftes Material im Rezipienten niederschlägt. Vor dem Substrat wird weiter eine Maske angeordnet, die zur Strukturierung des niedergeschlagenen Materials dient. Der Verdampfer kann auf unterschiedliche Weise geheizt werden, z.B. durch eine Widerstandswendel mittels Stromwärme oder durch einen Elektronenstrahl, der auf den Verdampfer trifft. Die Elektronenstrahlerwärmung hat den Vorteil, daß hierbei die Entmischung der Mehrkomponentensysteme geringer ist.

Den Aufbau eines kompletten Vakuumbedampfungssystems mit Wechselmasken und mehreren nacheinander zu bedampfenden Substraten zeigt Bild 9.4.7c).

9.4.5.2. Sputtern

Das Prinzip des Sputterns zeigt Bild 9.4.7b). In einem mit Edelgas, z.B. Argon, gefüllten Vakuumgefäß wird bei einem Druck von etwa 0,1 bis 10 Pa zwischen zwei Elektroden eine *Gasentladung* (vgl. 6.2.4.) gezündet. Sie bewirkt, daß die an der Katode mit hoher Energie auftreffenden Gasionen Atome des Katodenmaterials herausschleudern. Sie fliegen in alle Richtungen, werden aber durch die Gasreste relativ stark gestreut. So gelangen sie aus vielen Richtungen auf das Substrat, welches

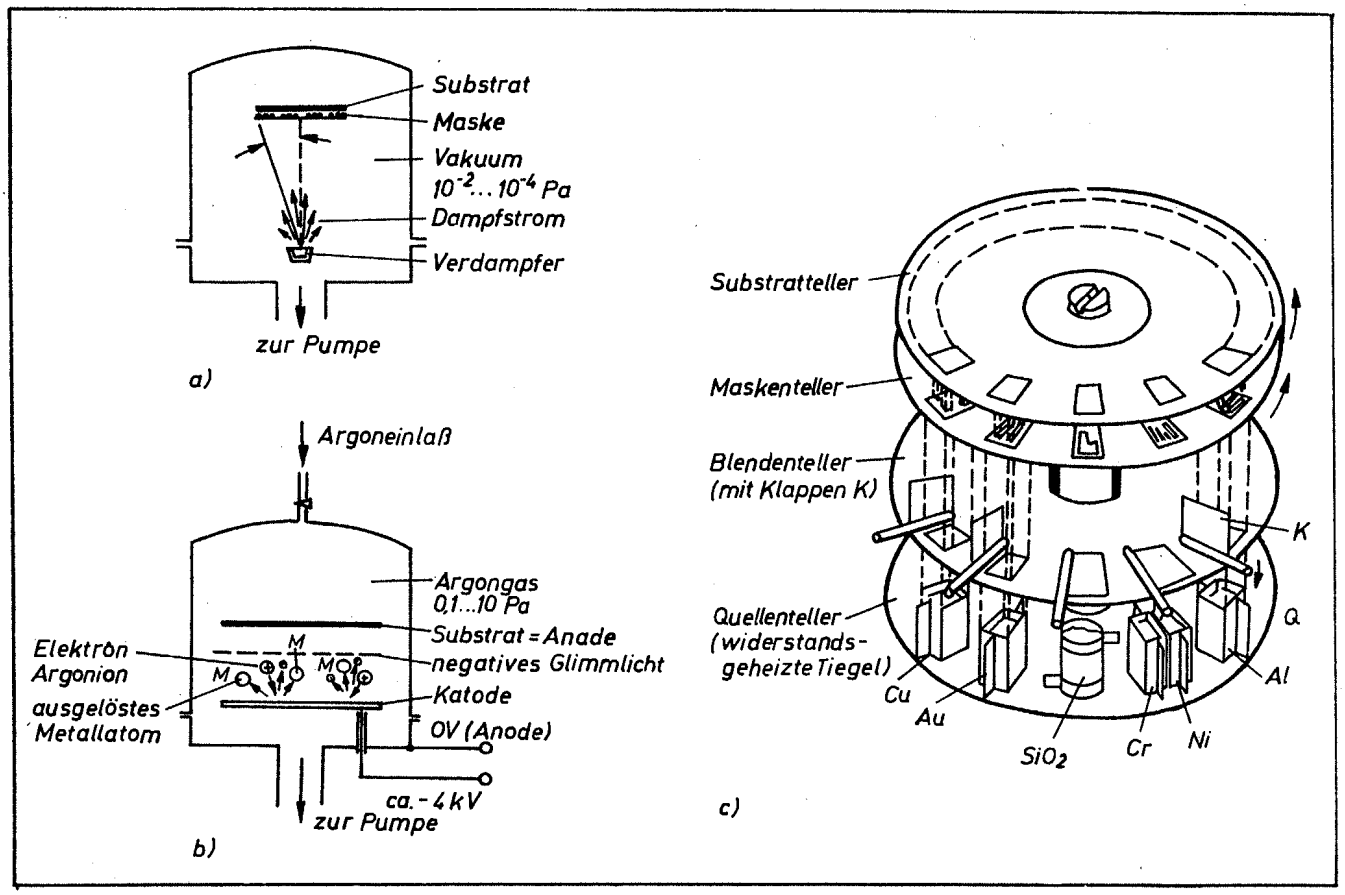


Bild 9.4.7. Zur Dünnschichttechnik

a) Prinzip der Vakuumbedampfung
b) Prinzip des Sputterns (Katodenzerstäubung)
c) Prinzipaufbau einer komplexen Bedampfungsapparatur für die Dünnschichttechnik

auf der Anode befestigt ist. Infolge des großen Auftreffwinkels können beim Sputtern kaum Masken verwendet werden. Die *Strukturierung* muß hier also durch *nachträgliches Ätzen* erfolgen. Das Sputtern hat vor allem *drei Vorteile*: Die Atome treffen einmal mit *hoher Energie* auf das Substrat und lagern sich dadurch u. a. haftfester ab. Das auf der Katode befindliche Material kann komplexer Zusammensetzung sein. Es erfolgt *keine Entmischung*. Schließlich kann der *Abstand* zwischen Katode und Anode mit einigen cm relativ gering gehalten werden, so daß vergleichsweise *mehr Material auf das Substrat* gelangt.

Beim Sputtern gibt es neben der gerade geschilderten Anordnung verschiedene Varianten, u. a. kann zusätzlich Hochfrequenz verwendet werden. Es kann auch bei niedrigem Gasdruck mittels zusätzlicher Elektroden oder Elektronenquellen indirekt die Gasentladung aufrechterhalten werden. Schließlich wird beim Plasmatron die Gasentladung mittels zusätzlicher Magnetfelder auf bestimmte Bereiche zusammengedrängt.

9.4.5.3. NiCr-System

Die Bezeichnung der beiden wichtigsten Systeme geht vom verwendeten Widerstandsmaterial aus. Das NiCr-System wird vor allem in der Bedampfungstechnologie realisiert. Die aufgebrachten Schichten liegen bei ca. 20 nm. Dabei werden Flächenwiderstände von etwa 30 bis 500 Ω/□ erreicht. Mittels eines Breiten-Längen-Verhältnisses bis zu 1000 (Mäander) können so Widerstände im Bereich von etwa 10 bis 100000 Ω

erzeugt werden. Infolge des Glassubstrates beträgt die spezifische Belastung ca. 0,7 W/cm².

Für die *Leiterbahnen* werden vor allem Cu, Al, aber auch Ag und Au verwendet. Dabei sind Flächenwiderstände zwischen 100 und 300 mΩ/□ erreichbar. Die Schichtdicke muß wegen der Haftfestigkeit unter 1 μm bleiben. Sie kann aber gegebenenfalls später z.B. elektrolytisch verstärkt werden. Die Leiterbreite liegt bei einigen zehn μm. Auf diesem Wege lassen sich auch Induktivitäten erzeugen. Da sich vakuumtechnisch auch Ferrit auftragen läßt, kann gegebenenfalls der Wert der Induktivität bis auf einige μH erhöht werden.

Als *Isolationsschichten* kommen vor allem SiO und SiO_2, eventuell auch Glas in Betracht. Die erreichbare spezifische Kapazität bei den üblichen Schichtdicken um 200 nm beträgt ca. 20 bis 100 pF/mm². Infolge des hohen Aufwandes wird in der Regel auf die Erzeugung von Kondensatoren verzichtet. Sie werden leichter später als diskrete Bauelemente eingefügt. In jedem Fall ist aber eine Isolation bei Leitungskreuzungen notwendig.

9.4.5.4. Tantalsystem

Das Tantalsystem hat infolge mehrerer chemischer Verbindungen des Tantals beachtliche Bedeutung erlangt. Es kann als Metall als Leiter, als TaN als Widerstandsmaterial und als Ta_2O_5 als Isolator verwendet werden. Die einzelnen Materialien können direkt gesputtert werden oder z. T. ineinander umgewandelt werden. So wird z.B. aus Ta durch anodische Oxydation Ta_2O_5 erzeugt. Hierbei können sehr dünne Schichten, also hohe spezifische Kapazitäten von 600 bis 3000 pF/mm² erzeugt werden. Die Dicke des Oxids beträgt dabei etwa 3 nm. Es handelt sich hierbei allerdings um einen polarisierten Elektrolytkondensator mit ca. 20 V Betriebsspannung. Weiter kann die anodische Oxydation zum Widerstandsabgleich der TaN-Widerstände herangezogen werden. Es wird ein dünner Schichtanteil in Oxid umgesetzt. Das TaN wird durch reaktives Sputtern von Ta in Stickstoffatmosphäre erzeugt. Sein Flächenwiderstand liegt bei 10 bis 100 Ω/□, so daß Widerstände im Bereich von 10 Ω bis zu einigen hundert kΩ erzeugt werden können. Der Vorteil des Ta-Systems liegt darin, daß alle Schaltungen durch anodische Oxydation passiviert, d. h. gegenüber späteren Einflüssen geschützt werden können.

9.4.6. Abgleich von *R* und *C*

Schon mehrfach wurde der Abgleich von Widerständen und Kondensatoren der Schichttechnologien erwähnt. Abgesehen von der anodischen Oxydation für Ta-Widerstände, erfolgt der Abgleich generell durch Abtragen von Material. Dies erfolgt vor allem mit den folgenden Methoden:

Laserstrahl
Elektronenstrahl (ca. 40 kV/100 ... 500 μA)
Funkenerosion
Sandstrahlen
Diamantritzen.

Alle Verfahren haben Vor- und Nachteile und werden vor allem je nach der verfügbaren Technik eingesetzt. Bei den Kondensatoren wird ein Teil der Elektrode abgetragen oder von der Zuleitung abgetrennt. Bei den Widerständen werden unterschiedliche Einschnitte, u. a. gemäß Bild 9.4.8, verwendet. Dabei kann Grob- und Feinabgleich unterschieden werden.

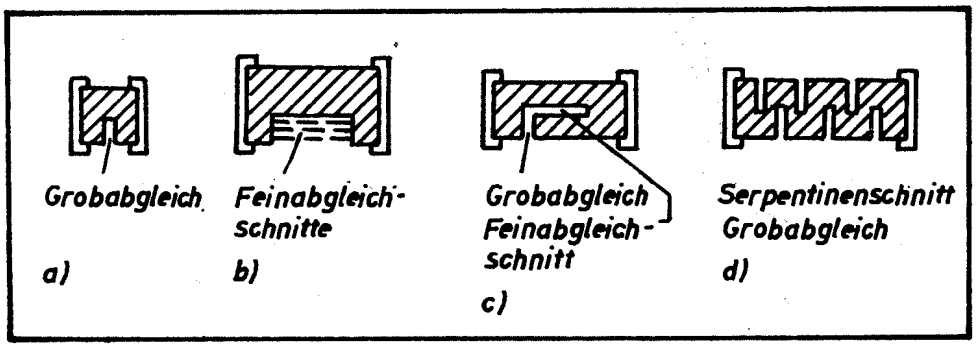


Bild 9.4.8. Einige Methoden des Abgleichs von Widerständen

9.4.7. Kapselung der Schaltungen

Zum Schutz gegen äußere Einflüsse werden alle Schichtschaltungen in Gehäuse gefügt (Bild 9.4.9). Die einfachste Variante, bei der allerdings keine nackten Chips verwendet werden dürfen, ist das Einhüllen in Kunstharz durch Tauchen, Umgießen usw. Für höhere Ansprüche hat sich das Metallgehäuse in zwei Varianten bewährt. Mittels Glasverschmelzungen befinden sich die Anschlüsse im Boden eingelassen, und verschlossen wird mit tiefgezogenem Deckel (Teilbild b)). Nachteilig ist diese Variante beim Einfügen in Leiterplatten, wenn die Stifte nicht genau ausgerichtet sind.

Die andere Variante zeigt Teilbild c). Sie führt die Anschlüsse zur Seite heraus und entspricht in etwa einem Flat-pack-Gehäuse.

Schließlich besteht für nicht extrem hohe Ansprüche die Möglichkeit, die Schichtschaltung gleich als Boden des Gehäuses zu verwenden (Teilbild d)). Dann ist nur eine Keramikkappe aufzukleben.

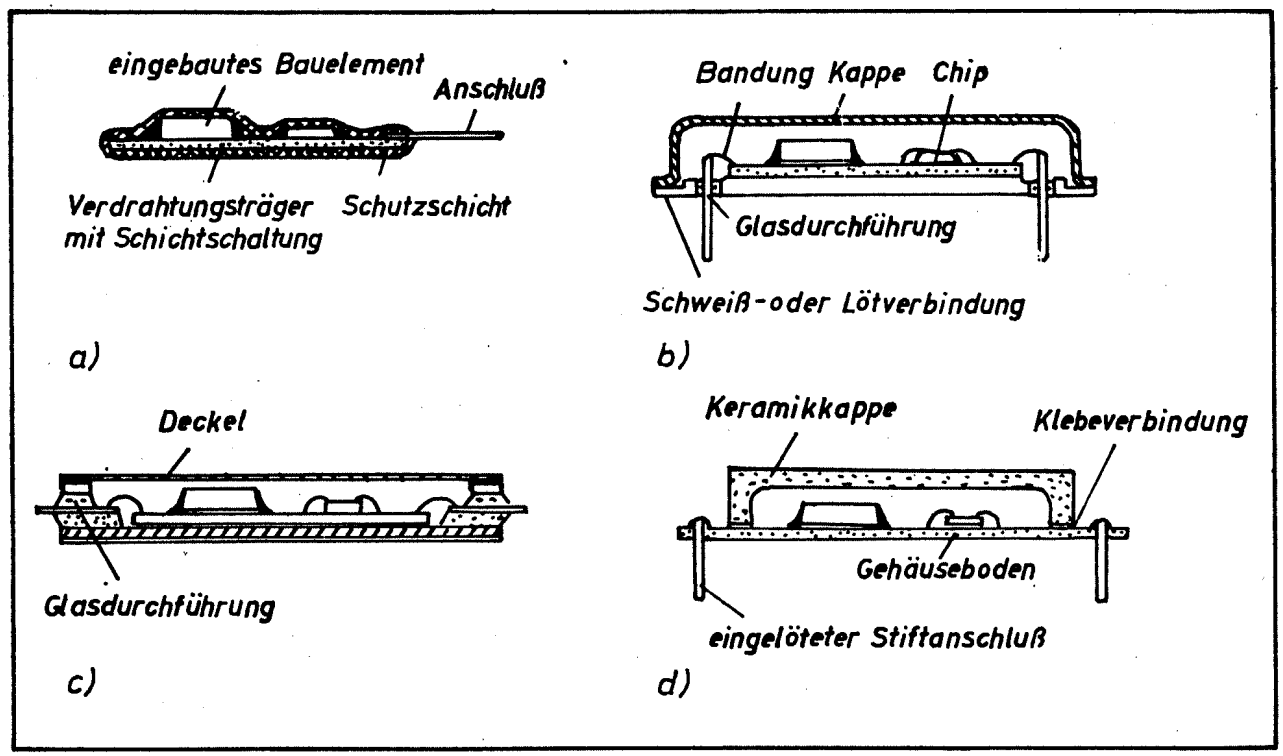


Bild 9.4.9. Varianten für die Verkapselung von Schaltungen

a) Vergießen
b) Metallgehäuse mit Anschlüssen im Boden
c) Metallgehäuse in Flat-pack-Variante
d) Keramikgehäuse, wobei das Substrat zugleich der Boden ist

9.4.8. Vergleich mit anderen Techniken

Im Abschnitt 8.9., insbesondere 8.9.5., erfolgte ein Vergleich verschiedener monolithischer Schaltungen. Dieser Vergleich ist jetzt auf die Hybridtechnologien mit Dünn- und Dickschichttechnik zu erweitern. Ein typisches Kennzeichen ist hierbei die ökonomisch optimale Stückzahl.

Den prinzipiellen Verlauf in Ergänzung zu Bild 8.9.7 und unter Berücksichtigung von Bild 9.4.2 zeigt Bild 9.4.10. Der Schnittpunkt A des Stückpreises hängt von vielen Fakten ab, u. a. von der Komplexität der Schaltung, der Variante der monolithischen bzw. hybriden Technologie, von der Zeit. Um 1980 lag er bei ca. 10000 bis 100000 Stück. Generell sind Hybridschaltungen ab etwa 1000 Stück effektiv herstellbar. Bei mehreren Anwendungen werden aber auch erhöhte Kosten in Kauf genommen, so daß zuweilen Losgrößen von 10 bis 100 Stück gefertigt werden.

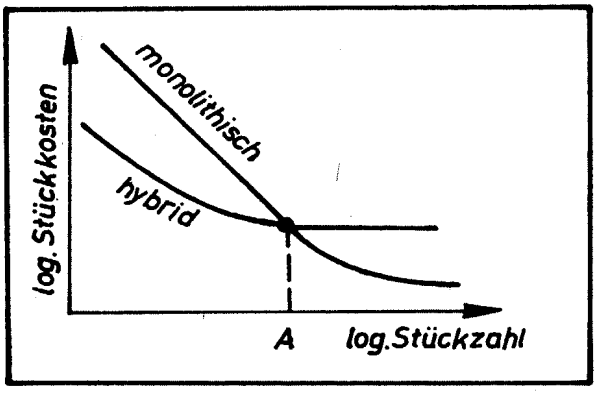


Bild 9.4.10. Prinzipieller Verlauf der Stückkosten von monolithischen und Hybridschaltungen als Funktion der Losgröße

Ein genereller Nachteil der Hybridschaltungen ist der weitaus höhere Materialeinsatz, insbesondere von Edelmetallen. Entscheidende Vorteile der Hybridschaltungen sind die sehr flexible Fertigung und die Abgleichbarkeit, was insbesondere sehr exakte Analogschaltungen zu fertigen erlaubt, sowie die schnelle Entwicklung neuer Schaltungen. Weiter lassen sich Hybridschaltungen sehr gut bis zu höchsten Frequenzen (GHz) realisieren. Ferner können schon vorhandene IC in die Schaltungen einbezogen werden. Dies führt letztlich zu der speziellen Variante der *Multichiptechnik*, bei der nur noch mehrere Chips zu einem neuen Schaltkreis verbunden werden. Eine andere Variante ist die *Kombinationstechnik*, bei der auf monolithische Chips zusätzlich u. a. Widerstandsnetzwerke ausgedampft werden. Schließlich sei noch der Vorteil der Hybridtechnik für größere Leistungen erwähnt. Insbesondere die keramischen Substrate lassen hier erhebliche Vorteile entstehen.

9.4.9. Bedeutung diskreter Bauelemente

Vergleicht man alle bisher behandelten Technologien, so entsteht die Frage, welche Bedeutung die diskreten Bauelemente in der Zukunft besitzen. Eine Teilantwort hierauf wurde bereits zu Beginn des Buches mit Bild 0.1 gegeben. Abgesehen von Röhren und diskreten Halbleitern nimmt ihr relativer Anteil eher etwas zu als ab. Dies wurde früher häufig falsch eingeschätzt und führte daher zu Engpässen. Insgesamt passen sie sich also durch Weiterentwicklung den neuen Anforderungen an.

Typische Beispiele für diese Aussage sind u. a. die Spezialisierung in der Röhrentechnik auf Sondergebiete und der wieder erhöhte Bedarf an Ge-Leistungstransistoren wegen der niedrigeren Schleusenspannung.

9.5. Zuverlässigkeit

9.5.1. Problembereich

Der Begriff Zuverlässigkeit wird unterschiedlich weit aufgefaßt bzw. definiert. Hier soll er zunächst im weitesten Sinne verstanden werden, nämlich im Sinne dessen, daß ein Bauelement, ein Gerät oder ein System (nicht nur ein technisches) fehlerfrei im Sinne seiner Aufgabenstellung funktioniert. Dabei sind u. a. Fragen bezüglich der Voraussetzungen für die Zuverlässigkeit (z. B. Betriebsbedingungen), für die Dauer der zuverlässigen Funktion (u. a. wirkliche bzw. zu erwartende Betriebszeit und Garantie) sowie für Voraussetzungen, um eine hohe Zuverlässigkeit zu erreichen (z. B. Fehlerfreiheit bei Entwicklung und Produktion, Prüfung und Testung, Redundanz, aber auch die Möglichkeiten der Wartung und Reparatur), zu unterscheiden.

Hier können nur einige dieser vielfältigen Probleme kurz behandelt werden. Dabei wird die mathematische Zuverlässigkeitstheorie fast vollständig ausgeklammert. Sie ist u. a. in [K 41 und P 19, Bd. 2] behandelt. Ein guter allgemeiner und auch stärker auf die Praxis bezogener Überblick ist u. a. in [A 2, D 6 und P 24] enthalten.

Probleme, die in engem Zusammenhang mit der Zuverlässigkeit stehen, wurden bereits in folgenden Abschnitten angeschnitten:

Messen, Prüfen, Abgleichen usw.	7.7.1.
fehlersichere Kodes	8.1.4.
Soft-Errors bei Speichern	8.3.4.8.
Hazards	8.2.9.2.
Race	8.4.8.
Software	8.8.1.

Wie stark der Mensch in vielfältiger Weise auf diese Zuverlässigkeit einwirkt, sei an drei Richtungen belegt:

- Als Entwickler und Produzent macht er Fehler.
- Er kann in fehlerhafter Weise das Gerät usw. anwenden.
- Die Zuverlässigkeit von Software wird nur durch ihn bestimmt.

Bei den technischen Geräten usw. kommt es darauf an, daß möglichst früh ein Fehler erkannt und vermieden wird. Je später dies erfolgt, desto höher ist der Aufwand. Dies sei am folgenden Beispiel erklärt: Wird ein defekter Transistor eingelötet oder ist die Lötstelle schlecht, so möge die sofortige Reparatur ca. 1 M kosten. Wird aber ein Fehler erst bei der fertigen Platine entdeckt, dann würde die Fehlersuche und Reparatur ca. 10 M kosten. Im fertigen Gerät kann der Aufwand bereits 100 M betragen. Bei einem Gerät im Einsatz kann der Wert mit Folgeschaden vielleicht auf 1000 M ansteigen. Es kommt also darauf an, Fehler möglichst früh zu entdecken und zu beseitigen oder besser gleich zu vermeiden. Hierbei ist zu beachten, daß Fehler ebenso in der Aufgabenstellung, im Entwurf, in der Konstruktion des Gerätes auftreten können. Wie die Analysen der japanischen Qualität gezeigt haben, ist hierbei die Einstellung zur Qualität der eigenen Arbeit z. T. entscheidend.

Wie groß der menschliche Einfluß ist, möge das folgende Beispiel bezüglich der Montage elektronischer Geräte demonstrieren. Auf 100 000 Operationen fallen im Mittel folgende Fehlerzahlen:

Vergessen, eine Komponente zu montieren	3
Zwei Drähte vertauscht	60
Eine Komponente mit Lötkolben beschädigt	100

Um Fehler zu entdecken, sind Test- und Prüfmethoden einzusetzen. Sie kosten ebenfalls Aufwand und verteuern ein Gerät. Weiter können Maßnahmen eingeleitet werden, um Fehler vorher schon zu verhüten. Alle diese Fakten hängen von der Qualität, dem Vollkommenheitsgrad des Produkts ab.

Insgesamt ergibt sich in etwa ein Verlauf, wie ihn Bild 9.5.1 a) zeigt. In diesem Beispiel liegt das Optimum vor, wenn

50% für Fehlerverhütung (Planung)
40% für Prüfung, Testung
10% für Fehlersuche und -behebung

verwendet wird. Allgemein können diese Werte nur als grober Richtwert betrachtet werden.

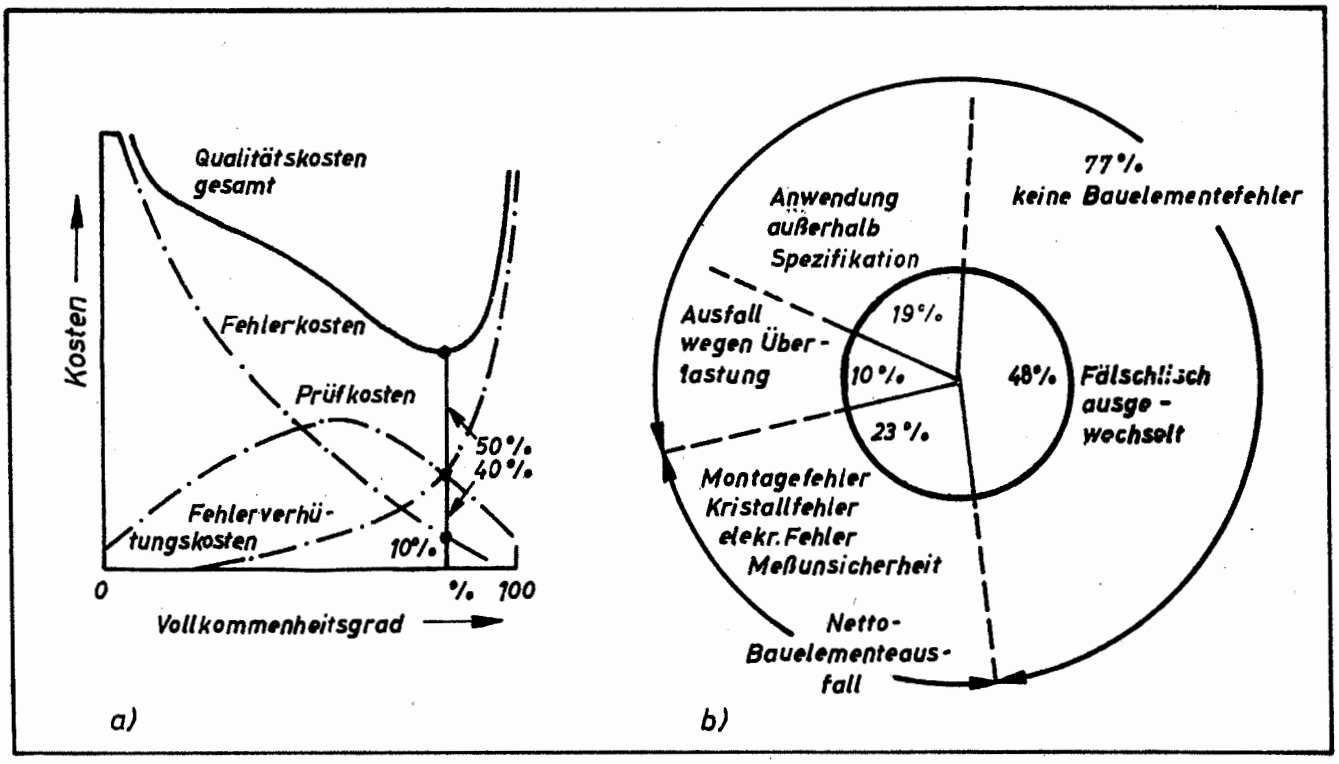


Bild 9.5.1. Zur Qualität von Bauelementen und Geräten

a) Einfluß verschiedener Kostenanteile auf die Qualität

b) Von Valvo 1982 ermittelte Anteile bezüglich „fehlerhafter" Bauelemente

Ein weiteres Beispiel für den Einfluß des Menschen und der Prüfmethoden zeigt Bild 9.5.1 b). Bei als „defekt" bezeichneten bipolaren integrierten Schaltkreisen stellte Valvo fest, daß fast die Hälfte fälschlich ausgewechselt wurde. Dies bestätigt wiederum die Bedeutung einer guten Prüftechnik bzw. Diagnose. Sie kann in zwei Gruppen eingeteilt werden: Einmal, um nur einen Fehler zu erkennen, also um zur Aussage gut — schlecht zu gelangen. Zum anderen, um auch die Ursache des Fehlers zu finden und ihn zu beheben. Für diesen weiteren Zweck wird bei komplexen Schaltungen mit der Entwicklung oft zugleich ein Fehlersuchsystem (z. T. automatisiert) erarbeitet. Hierdurch können die Fehlerkosten erheblich gesenkt werden.

Die Einteilung der Begriffe um die Zuverlässigkeit ist recht schwierig. Hier sei einmal eine Einteilung gegeben, die von der Qualität ausgeht (Bild 9.5.2 a)), und eine zweite, in deren Zentrum der Fehlerbegriff steht (Teilbild b)). Aus Teilbild a) wird insbesondere das Verhältnis von Zuverlässigkeit, Verfügbarkeit und Wartbarkeit deutlich. Es hat nur dann einen Sinn, wenn das Gerät im Zusammenhang mit Ausfällen (Fehlern) reparierbar ist. Verfügbarkeit kann als der relative Zeitanteil betrachtet werden, in der das Gerät genutzt werden kann. Die Wartbarkeit besteht aus der Überwachung der Betriebsfunktion, der Instandhaltung im Sinne von korrigierender und vorbeugender Wartung und der Reparatur (Instandsetzung).

Tab. 9.5.1 gibt abschließend einen kurzen Überblick zur Geschichte der Zuverlässigkeit.

9.5.2. Grundlagen

9.5.2.1. Fehlerverteilung, Ausfallrate

Wird eine sehr große Zahl N_0 von Bauelementen unter gleichen Bedingungen ständig betrieben, so fallen in zufälliger Weise einige aus. Werden *die noch überlebenden*, betriebsfähigen Bauelemente mit $N(t)$ bezeichnet, so entsteht in der Regel eine Kurve, wie sie etwa Bild 9.5.3 zeigt. Der Abfall ist kurz nach der Inbetriebnahme aller Bauelemente relativ steil, geht dann in einen langen linearen Teil über, um schließlich wieder steiler abzufallen. Dieser Verlauf läßt sich wie folgt interpretieren:

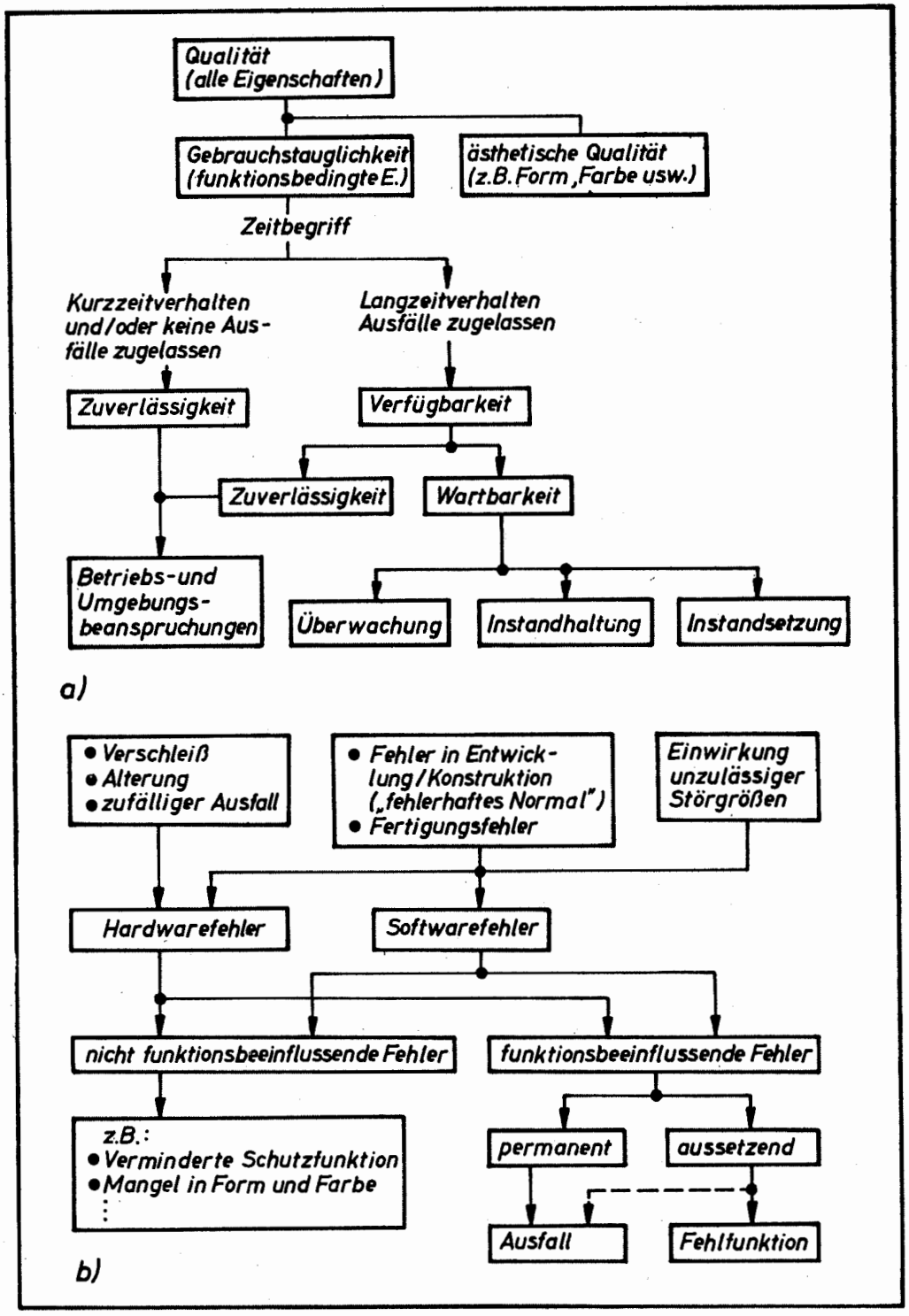


Bild 9.5.2. Zur Systematik einiger Begriffe um die Zuverlässigkeit

a) Aus der Sicht der Qualität

b) Aus der Sicht der Fehler

Zu Anfang überwiegen sogenannte *Frühausfälle*, die im wesentlichen auf Mängel bei der Herstellung zurückzuführen sind. Dies erfolgt in der Regel in den ersten 100 bis 2000 Betriebsstunden.

Der zweite Teil wird durch die *Gesetze des Zufalls* bestimmt. Für ihn gilt meist mit guter Näherung

$$N(t) = N_0 \cdot \exp(-t/m) = N_0 \cdot \exp(-t\lambda) . \tag{1}$$

Der Wert λ wird als *Ausfallrate* oder auch *Fehlerrate* bezeichnet. Er gibt an, wieviel Bauelemente relativ von der vorhandenen Anzahl in einer bestimmten Zeit ausfallen.

Tabelle 9.5.1. Zur Geschichte der Zuverlässigkeit

1930	Erste Arbeit von V. M. MONTSINGER (USA) zur 8-K-Regel
1942	BÜSSING entwickelt wesentliche theoretische Zusammenhänge
1950	Beginn umfangreicher experimenteller Untersuchungen bez. Bauelementen
1951	Zuverlässigkeit von Radar- und Feuerleitgeräten wird für die USA im Korea-Krieg entscheidend
1955	Zuverlässigkeit von Bauelementen bei Rechnern mindert erheblich deren Betriebszeit
1958	Störungen im Funkverkehr durch Atomexplosionen
1960	Bedeutung der Zuverlässigkeit bei der Raumfahrt beginnt
1962	Elektromagnetischer Impuls wird festgestellt
1964	IEC bildet TC 56 „Reliability"
ca. 1970	Intensive Arbeiten zur elektromagnetischen Verträglichkeit
1978	Soft-Error-Rate bei dynamischen RAM wird aufgeklärt
1978	Optimale Arithmetik genormt

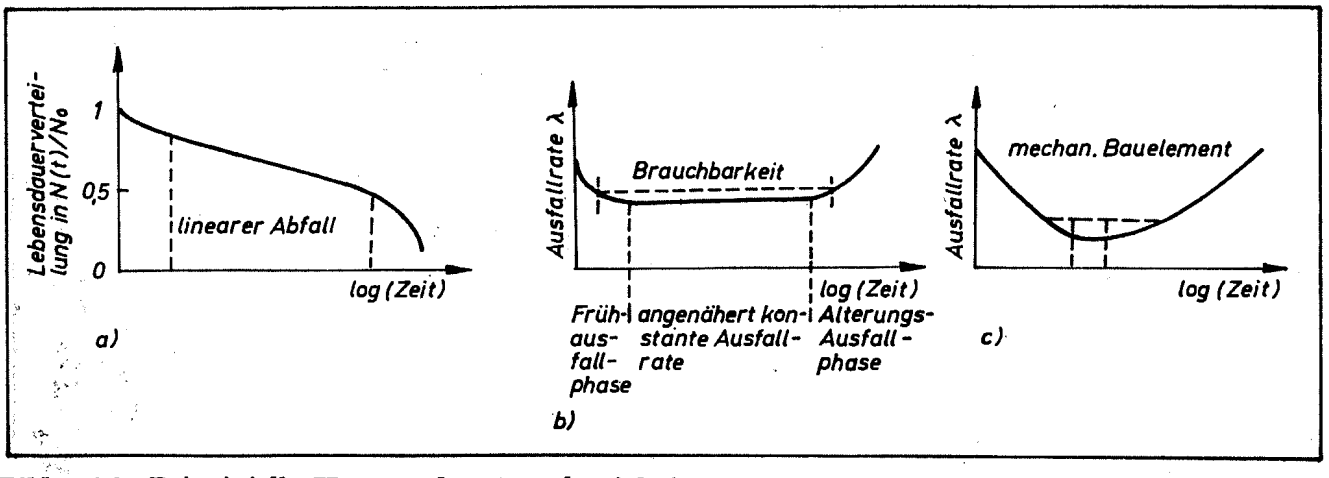


Bild 9.5.3. Prinzipielle Kurven der Zuverlässigkeit

a) Lebensdauererwartung von N Bauelementen

b) Kurve der Ausfallsrate mit ihren drei typischen Bereichen

c) Bei mechanischen Bauelementen gibt es nur einen extrem kurzen Bereich konstanter Ausfallsrate

Im allgemeinen wird dieser Wert auf 10^6 Stunden bezogen. So bedeutet $\lambda = {}^1/_{10^6}$ h, daß von 10^6 Bauelementen im Mittel 1 je Stunde ausfällt oder anders ausgedrückt, daß jedes dieser Bauelemente im Mittel 10^6 h Betriebszeit besitzt.

Der Faktor m ist gemäß Bild 9.5.3a) in zwei Arten zu messen. Erfolgt keine Reparatur — dies gilt also vor allem bei Einzelbauelementen —, so gilt der Wert

MTTF = meantime to failure
= mittlere Zeit bis ein Fehler auftritt.

Bei Systemen aus vielen Bauelementen kann eine Reparatur vorgenommen werden, dann wird verwendet

MTBF = meantime between failures
= mittlere Betriebszeit zwischen zwei Fehlern.

In Anbetracht des exponentiellen Gesetzes von Gl. (1) ist bei beiden Werten die Wahrscheinlichkeit für den Fehler $1 - 1/e \approx 63\%$.

Der *dritte Teil* in Bild 9.5.3 zeigt wieder einen erhöhten Ausfall. Hier sind die Bauelemente am *Ende ihrer Lebensdauer*. Es treten deutliche Verschleißerscheinungen auf.

In Bild 9.5.3b) ist der Verlauf der Ausfallrate als *Ableitung nach der Zeit* in Bezug zu Teilbild a) gezeigt:

$$\lambda(t) = -\frac{1}{N(t)} \cdot \frac{\mathrm{d}N(t)}{\mathrm{d}t} . \tag{2}$$

Hier wird der lineare Teil noch deutlicher sichtbar. Es gibt damit auch Abschnitte, in denen die Ausfallrate zu- bzw. abnimmt. Dies wird noch deutlicher, wenn ein mechanisches System betrachtet wird. Hierfür gilt schematisch Teilbild c). Es gibt also fast keinen Bereich, in dem die Ausfallrate konstant ist. Dies hat dazu geführt, daß neben dem exponentiellen Gesetz auch andere Verteilungen in die Zuverlässigkeitstheorie einbezogen wurden. Besonders wichtig sind die WEIBULL- und die Normalverteilung. Hier sei jedoch weiterhin nur die exponentielle Verteilung betrachtet.

Für die weiteren Betrachtungen ist es nützlich, aus Gl. (1) zwei *weitere Kenngrößen* einzuführen: die Zuverlässigkeit im engeren Sinne als Quotient aus $N(t)$ und N_0,

$$R(t) = \exp(-\lambda t) , \tag{3}$$

und die Verteilungsfunktion der Lebenszeit (Fehlerwahrscheinlichkeit),

$$F(t) = 1 - \exp(-\lambda t) . \tag{4}$$

9.5.2.2. Bauteile und Systeme

Ein System bestehe aus mehreren Bauelementen. Für jedes Bauelement sei die Ausfallrate bekannt. Die Frage ist, wie groß dann die statistische Ausfallrate des Systems ist. Dies hängt u. a. davon ab, wie wichtig die einzelnen Bauelemente für die Funktion des Systems sind. Ein besonders übersichtlicher Fall liegt vor, wenn alle Bauelemente so wesentlich sind, daß auch der Ausfall eines beliebigen sofort die Funktion des Gerätes aufhebt. In diesem Fall spricht man von einem *nicht redundanten Seriensystem*. Seriensystem deshalb, weil bezüglich der Zuverlässigkeit die Bauelemente wie eine Serienschaltung von Widerständen aufgefaßt werden können. Wenn dann einer defekt ist, kann kein Strom mehr fließen. In diesem Sinne existieren beliebige Zusammenschaltungen. Ein anderes Beispiel ist die Reihenschaltung von zwei Kondensatoren in einer Siebkette. Wenn einer von beiden durchschlägt und dann isoliert ausscheidet, bleibt die Siebkette weiterhin funktionstüchtig. Dies kann als Parallelschaltung der Zuverlässigkeit oder in anderem Sinne als *Redundanz* bzw. *Reserve* aufgefaßt werden. Zur Vereinfachung seien im weiteren nur Serienschaltungen der Zuverlässigkeit betrachtet.

Es mögen n Bauelemente in einem System mit den Ausfallraten λ_1 bis λ_n existieren. Für die *Zuverlässigkeit des Systems* gilt dann der multiplikative Ansatz

$$R_S(t) = R_1(t) \cdot R_2(t) \ldots R_n(t) . \tag{5}$$

Bezüglich der λ-Werte bedeutet dies, daß die Ausfallrate des Systems die Summe der Ausfallraten der Bauelemente ist.

$$\lambda_S = \sum_{i=1}^{n} \lambda_i . \tag{6}$$

Wird für alle Bauelemente eine einheitliche Ausfallrate λ_E angenommen, so gilt

$$R_S(t) = \exp(-n\lambda_E t) \tag{7}$$

bzw.

$$\lambda_S = n\lambda_E . \tag{8}$$

Für die *Lebenszeit* (Wahrscheinlichkeit der Betriebssicherheit) ist die Rechnung komplizierter:

$$F_S = 1 - (1 - F_E)^n . \tag{9}$$

Dies sei an einem einfachen Beispiel erklärt: Auf einer Leiterplatte seien 50 Bauelemente. Jedes sei zu einer bestimmten Zeit (kann auch beim Einbau sein) mit einer Wahrscheinlichkeit von 1% defekt, dann beträgt die Wahrscheinlichkeit für ein fehlerhaftes System $1 - (0{,}99)^{50} \approx 39{,}5\%$. Selbst bei einer Fehlerwahrscheinlichkeit von 10^{-4}, aber für ein System mit 10000 Bauelementen folgen 63%.

Hieraus wird deutlich, daß große Systeme extreme Forderungen an die Zuverlässigkeit von Bauelementen stellen. Dies war ein Hauptgrund, warum die *Zuverlässigkeitsbetrachtungen mit den größer werdenden Systemen immer wichtiger wurden.*

Kurven bezüglich der MTBF und der Zuverlässigkeit zeigt Bild 9.5.4a) und b).

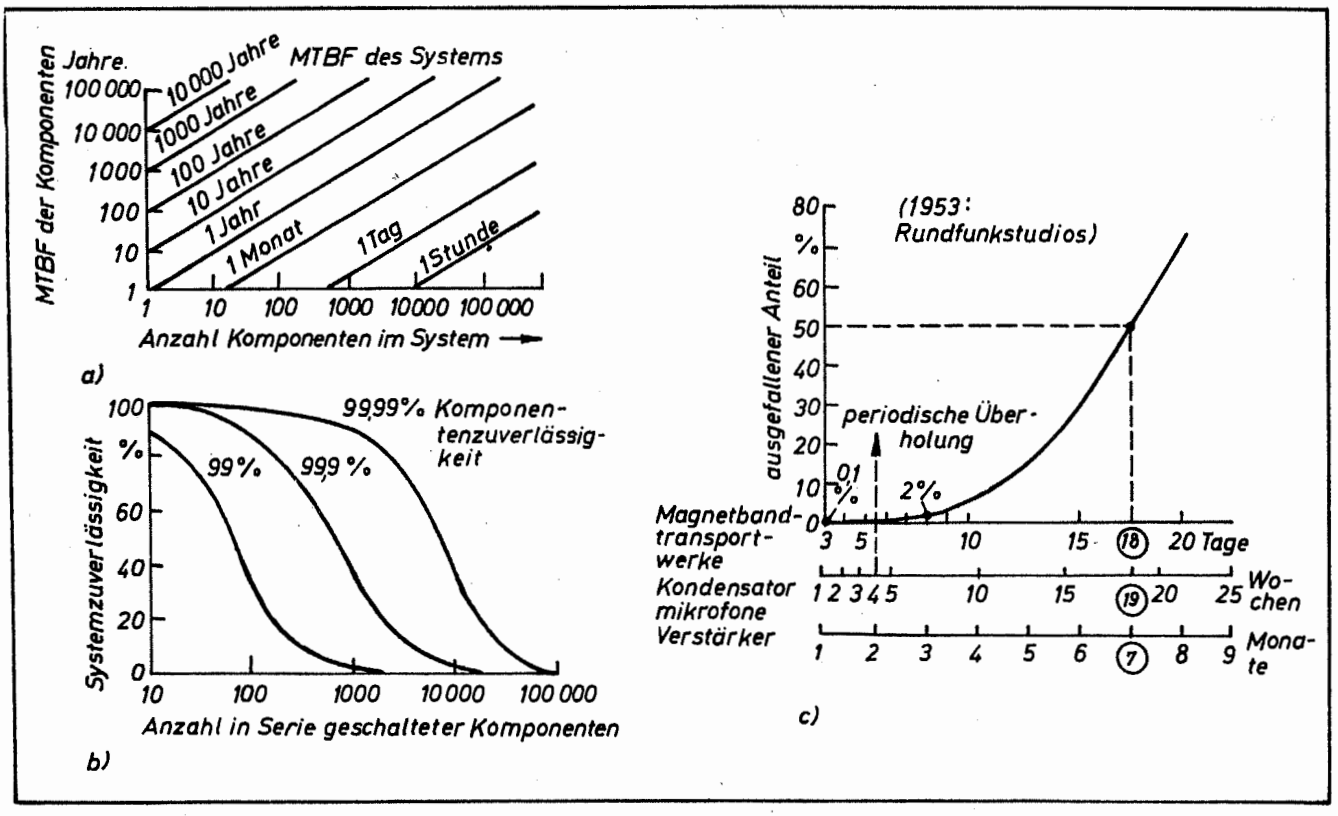


Bild 9.5.4. Zu Systemen aus vielen Komponenten

a) MTBF eines Systems, welches sich aus einer Anzahl von Komponenten mit gleichen MTBF der Komponenten zusammensetzt

b) Zuverlässigkeit eines Systems, das aus einer Anzahl von Komponenten mit gleicher Zuverlässigkeit (Werte 99; 99,9 und 99,99%) zusammengesetzt ist

c) Typisches Verhalten von drei Rundfunkstudiogeräten im Jahre 1953 sowie den daraus abgeleiteten Zeiten der periodischen Überholung

9.5.2.3. Wartung und Reserve

Das zuvor betrachtete Seriensystem der Zuverlässigkeit erfordert bei vielen Komponenten eine extrem hohe Zuverlässigkeit der Komponenten. Das ist nicht immer zu realisieren. Deshalb haben zwei Varianten des Systemverhaltens Bedeutung.

Ein System mit *Reserve* liegt dann vor, wenn für den Ausfall eines Bauelements ein anderes unmittelbar die Funktion übernehmen kann. Diese Übernahme kann unmittelbar oder automatisch über ein zusätzliches Diagnosesystem erfolgen. Es werden je nach der Belastung des Reservesystems in seinem Reservezustand unterschieden:

Bei der *heißen Reserve* wird es ständig unter normalen Betriebsbedingungen betrieben. Es ist also genauso belastet, wie jenes Bauelement, dessen Funktion es im Falle eines Fehlers übernehmen soll.

Bei der *kalten Reserve* steht es nur bereit, und es erfolgt bei ihm während dieser Zeit kein Verschleiß.

Zwischen den beiden genannten Fällen liegt die warme Reserve.

Eine abweichende Methode ist die *Wartung* von Geräten. Aus statistischen Untersuchungen sei der Fehlerverlauf bekannt. Dann kann zu genau festgesetzten Zeiten eine systematische Wartung erfolgen. Dies ist z.B. seit langem in den Studios der Rundfunkanstalten üblich. Eine bewußt alte Kurve hierzu zeigt Bild 9.5.4c). Die periodische Überholung der drei Gerätetypen erfolgt zu dem Zeitpunkt, wo die Ausfallwahrscheinlichkeit bei ca. 0,3% liegt; die Betriebssicherheit der Geräte liegt dann ständig bei 99,7%.

9.5.3. Ursachen und Einflüsse

9.5.3.1. Analyse der Ursachen

Die Ermittlung von Ursachen bei Ausfällen von Schaltungen, Schaltkreisen und Bauelementen ist der wichtigste Hinweis zur Verbesserung der Qualität. Sie erfolgt in der Regel durch Inspektion des fehlerhaften Bauelements mit Lupe, Mikroskop und Rasterelektronenmikroskop. Für die weitere Analyse müssen jene Fakten ausgeschlossen werden, die durch unzulässige Verwendung des Bauelements oder infolge von unzulässigen Werten infolge fehlerhafter Betriebsweise, eventuell auch durch Störungen in der Schaltung entstehen. Hierfür gab Bild 9.5.1b) wichtige Hinweise. Ein ähnliches Ergebnis aus anderer Sicht und für bestückte Leiterplatten zeigt Bild 9.5.5b).

Werden derartige Fakten ausgeklammert, so gelangt man zur Fehleranalyse der Herstellung von Bauelementen. Ein Beispiel für IC zeigt hierzu Bild 9.5.5c) nach zwei verschiedenen Quellen. Aus einer Vielzahl solcher Quellen und unter Berücksichtigung des Zeiteinflusses können dann Kurven entsprechend Bild 9.5.5a) auch in die Zukunft prognostiziert werden. Aus der Fülle des aus solchen Analysen gewonnenen Materials wurde noch Bild 9.5.5d) aufgenommen, welches sowohl Aussagen zu den Hybridtechnologien als auch unterteilt nach analogen und digitalen Schaltungen enthält.

9.5.3.2. Auswirkungen äußerer Einflüsse

Äußere Einflüsse auf die Zuverlässigkeit sind vielfältig vorhanden. Eine grobe Zusammenstellung gibt Tab. 9.5.2. Darin wird auch auf einige noch speziell zu behandelnde Gebiete hingewiesen. Hier folgen nur wenige Ergänzungen. Mehr Details enthalten u. a. [A 2, D 6, D 12, P 24].

Bei den klimatischen Einflüssen müssen vor allem die *Klimazonen* unterschieden werden: Tropen, Subtropen, gemäßigte Zone und Polarzone sowie zusätzlich See- und Kontinentalklima.

Die *Temperaturunterschiede* sind z.B. im Kontinentalklima Asiens von −70 bis +30 °C mit 100 K extrem groß. Da alle Halbleiterbauelemente temperaturabhängig sind, wirken sich solche Einflüsse beträchtlich aus.

Die *Feuchte* ist u. a. bei Seeklima oft hoch. Kritisch wird sie infolge möglicher Betauung bei Temperaturwechsel und dudurch bedingtem Eindringen von Wasser in Poren und Haarrisse. Wenn dann noch tiefe Temperaturen hinzukommen, wirkt die Ausdehnung des Wassers beim Gefrieren zusätzlich mechanisch. Durch Feuchte werden darüber hinaus Kriechwege erzeugt. Der Durchschlag von Papierkondensatoren ist oft eine typische Folge von eingedrungener Feuchte. Schließlich beschleunigt Feuchte die Korrosion.

Luftdruckänderungen haben — von evakuierten oder unter Schutzgas stehenden Bauelementen abgesehen — nur mittelbaren und meist geringen Einfluß. Sie sind jedoch bei Geräten für die Satellitentechnik und Tiefseeforschung zu beachten.

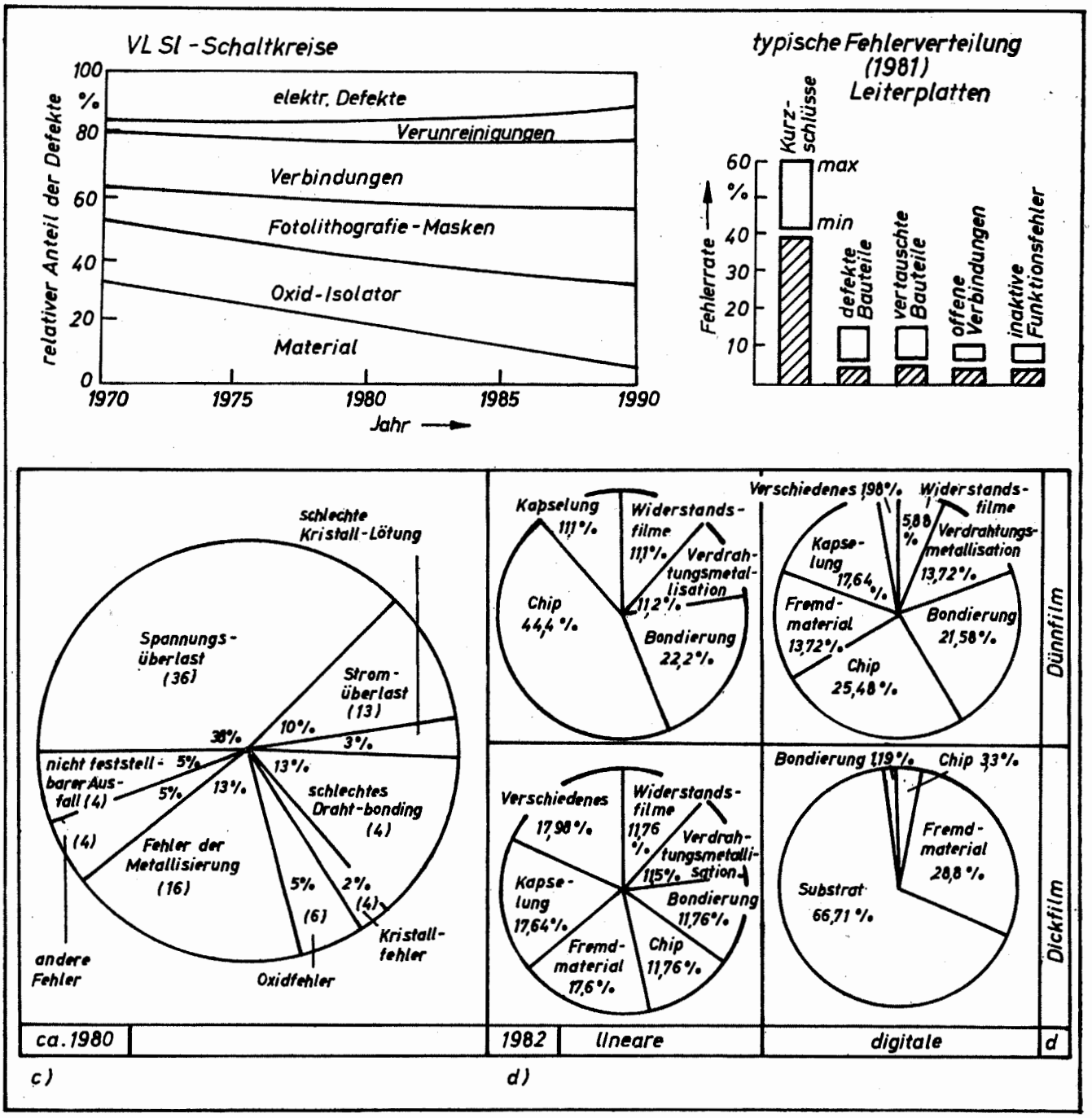


Bild 9.5.5. Zu einigen Ausfallursachen (vgl. auch Bild 9.5.1 b))

a) Entwicklung der Ursachen bei VLSI-Schaltkreisen
b) Ausfallursachen bei bestückten Leiterplatten
c) Ausfallursachen bei integrierten Schaltkreisen (Klammerwerte aus anderer Quelle)
d) Hybridschaltkreise nach Dick- und Dünnfilmtechnik bzw. bei linearen und digitalen Schaltungen

Tabelle 9.5.2. Zusammenstellung wichtiger äußerer Einflüsse auf die Zuverlässigkeit elektronischer Bauelemente

Klima(-zonen)	Temperatur, Feuchte, Luftdruck
Mechanisch	Schlag, Fall, Stoß; Vibration, Beschleunigung, Lärm; Staub
Chemisch	Gase, insbesondere mit Cl-, S-, C-Gehalt sowie Ozon; Säuren; Salze (Ionen)
Elektrisch	Gewitter, EMV, EMP (9.5.6.)
Strahlung	Licht, γ-Strahlen, α-Teilchen, Elektronen, Protonen (9.5.6.)
Biologisch	Schimmel- und Pilzbefall, Insekten und Nagetiere

Mechanische Beanspruchungen hängen stark von den Einsatzbedingungen ab. Sie sind z.B. im Labor und bei stationären Geräten recht gering, nehmen aber bei Fahrzeugen, Flugzeugen und Raketen beträchtliche Werte an. Bei Schwingungen sind besonders Eigenresonanzen unterschiedlichster Art gefährlich. Sie führen z.B. zum Aufreißen von Schweiß- und Lötstellen, Brechen von Zuleitungen und Durchscheuern von Isolierungen. Durch Lärm können neben der mechanischen Beanspruchung auch störende Mikrophonieeffekte auftreten. Staub bewirkt u. a. Störungen bei mechanischen Bauteilen, in Filtern, auf Kontakten usw. Er verschlechtert weiter die Wärmeableitung.

Chemische Einflüsse sind durch hohe Anteile von Schadstoffen in der Luft immer störender geworden. Schwefelgase wirken sich stark durch Silbersulfidbildung auf Kontakte wie Stecker, Schalter usw. aus. Es mußten deshalb neue Kontaktwerkstoffe entwickelt werden. Für hohe Ansprüche ist Gold eine Möglichkeit. Bei Kunststoffen ist zu beachten, daß sie zu Anfang oft Gase, u. a. Chlor, abgeben. Säuren und Salze sind z. T. in Niederschlägen, aber auch in starkem Maße in der Seeluft enthalten. Sie besitzen eine hohe Korrosionswirkung.

Biologische Einflüsse durch Schimmel- und Pilzbefall erhöhen vor allem die Feuchte und damit die Korrosion. Oft sind Staubschichten für sie der Nährboden. Organische Isoliermaterialien konnten weitgehend dagegen beständig gemacht werden. Insekten stören auf mehrfache Weise: mechanisch durch ihre toten Körper, verschmutzend durch ihre Lebensweise, und schließlich zerfressen sie ähnlich wie Nagetiere Materialien.

Eine verdichtete Zusammenfassung dieser Einflüsse ergibt sich, wenn die Einsatzbedingungen aufgestellt werden. Eine entsprechende Übersicht zeigt Tab. 9.5.3. Werden diese Einsatzklassen noch etwas zusammengefaßt, so lassen sich für unterschiedliche Bauelemente Betriebsfaktoren angeben (Tab. 9.5.4). Sie weisen in etwa aus, mit welchem Faktor eine mittlere Ausfallrate zu erhöhen ist.

9.5.3.3. Betriebsbelastungen

Alle Bauelemente haben eine Belastungsgrenze, die u. a. bezüglich Temperatur, Spannung, Strom, elektrischer Leistung usw. gilt. Beispiele hierfür sind u. a. die Derating-Kurven (z.B. Bild 1.1.14), der sichere Arbeitsbereich (z.B. Bild 5.5.1, 5.7.16) und der schon z. T. behandelte Zusammenhang mit der Lebensdauer (z.B. Bild 1.5.11).

Bei den Betriebsbelastungen ist aber auch die notwendige *Einsatzdauer* zu beachten. Bei einem Flugzeug gibt es Elektronik, die nur während der wenige Minuten dauernden Lande- bzw. Startphase eingesetzt ist. Elektronik in Raumflugkörpern muß über Jahre hinweg wartungsfrei arbeiten, Elektronik in Fernmeldeeinrichtungen, z.B. Telefonzentralen bzw. Überseekabeln, muß dies z. T. mit und ohne Wartung für Jahrzehnte leisten. Bei vielen Geräten wird eine bestimmte MTBF gefordert. Sie liegt bei vielen Geräten bei einigen tausend Betriebsstunden.

Bezüglich der *Belastung* können drei Bereiche unterschieden werden: Über-, Nenn- und Unterbelastung. Daten gelten im allgemeinen für die Nennlast. Bei Überlast sind die auftretenden Störungen nicht anders als bei Nennlast, jedoch weitaus häufiger. Bei Unterlast sind die Verhältnisse komplizierter. Einmal vermindert sich hier die typische Ausfallrate bezüglich der Belastung. Es können jedoch zusätzliche typische Fehler auftreten. Sie sind besonders deutlich beim Elko, da hier dann die Nachformierung wesentlich reduziert wird. Bei Kontakten, z. B. in Relais und Schaltern, nimmt bei starker Unterlast sogar die Zuverlässigkeit erheblich ab. Bei der Signalverarbeitung ist zu beachten, daß kleinere Signale erheblich störanfälliger u. a. gegenüber Rauschen und zufälligen Impulsen sind. Wird die Unterbelastung durch zusätzliche Bauelemente

Tabelle 9.5.3. Übersicht über die verschiedenen Umgebungsbeanspruchungen elektronischer Geräte

Geräteklasse	Temperaturbereich		Relative Luftfeuchte	Vibrationen maximal		Schock, g		Lärmbeanspruchung	Druckverhältnis	Sand und Staub
	minimal °C	maximal °C	% r. F.	Frequenz Hz	Beschleunigung g	bei 11 ms Dauer	bei 0,5 ms Dauer	dB	mbar	
Stationär, gepflegte und temperaturbeständige gereinigte Räume	+5°	+30°	60%	10 ... 55	2	15	0	70	800	kein
Stationär sonstige	−10°	+40°	80%	10 ...150	2	15	1500	40 ... 70	800	mäßig
Tragbare Geräte	−25°	+50°	95%	10 ... 150	5	30/18 ms	1500	40 ... 100	550	stark
Fahrbare Geräte	−25°	+60°	95%	10 ... 500	10	30/18 ms	1500	40 ... 100	4700	stark
Luftbordgeräte (in Kabine, Cockpit)	−4°	+70°	95%	10 ... 2000	10	50	1500	40 ... 90	300	gering
Luftbordgeräte (übrige Orte)	−65°	+100°	100%	10 ...2000	10	50	1500	50 ... 150	50	mäßig
Raumfahrzeuge Trägerraketen (Elektronikbehälter)	−15°	+60°	75% (90%)	5 ... 2000	25	45 ... 100	1500	70 ... 165	10^{-6} mm Hg	gering
Satelliten (Abschußphase)	−55°	+85°	75%	10 ... 3000	5 ... 60	15 ... 75	1500	50 ... 130	10^{-6} mm Hg	kein
Satelliten (im Orbit)	−55° (−20°)	(+50°) +85°	0%	0	0	0	0	0	10^{-8} mm Hg	kein
Schiffsgeräte	−25°	+55°	100%	2 ... 150	5	30/18 ms	—	—	950	mäßig

Tabelle 9.5.4. Zusammenstellung der Betriebsfaktoren zur Berücksichtigung der unterschiedlichen Umgebungsbedingungen für Ausfallraten
Die Beanspruchungen für Satelliten können denen der ersten Spalte gleichgesetzt werden, da die Beanspruchungen während der Antriebsphase (vergleiche Spalte 6) nur sehr kurz im Verhältnis zur gesamten Betriebszeit auftreten.

Typ	Boden, stationär	Boden, beweglich	Schiff	Luftbord	Flugkörper
Röhren	1	6	4	6	80
Dioden					
Germanium-	0,75	3,5	0,21	3,5	10
Silicium-	1,5	3,5	4,33	3,5	10
Leistungs-	1,0	12	3,26	12	18
ZENER-	1,0	3	2,75	3	10
Transistoren					
Germanium-	1,5	8	0,09	8	25
Silicium-	1,5	8	1,03	8	25
Leistungs-	1,0	8	0,51	8	75
Widerstände					
Masse-	6	10	21,2	10	50
Schicht-	0,03	0,3	0,67	0,3	1,5
Kondensatoren					
Glas-	1,4	15	11,5	15	25
Papier-	1,3	6	12,5	6	20
Keramik-	1,0	5	16,8	5	15
Tantal-	3	30	0,52	30	50

erkauft, so bringt die höhere Bauelementezahl zusätzliche Störungsmöglichkeiten, wenngleich auch eventuell Redundanz. Auf alle Fälle ist durch Unterbelastung nur z. T. eine Erhöhung der Zuverlässigkeit zu erreichen.

Bezüglich der *Dynamik der Belastung* können statische Dauerlast, ständig oder zeitweilig wechselnde Last und Impulslast unterschieden werden. Die Änderung der Last kann oft vorteilhaft genutzt werden, da Bauelemente im Impulsbetrieb, also kurzfristig, oft ein Vielfaches der Nennlast ohne Änderung der Zuverlässigkeit vertragen. Andere Bauelemente benötigen direkt eine Erholphase. Diesbezüglich sei nur auf die Depolarisation bei Batterien verwiesen.

Die Änderungen der Belastung bezüglich der Lebensdauer sind vielfältig untersucht. Hier seien nur zwei Ergebnisse mitgeteilt. Das ARRHENIUS*sche Gesetz* der Chemie lautet bezüglich der Reaktionsgeschwindigkeit

$$\frac{\mathrm{d}M}{\mathrm{d}t} = A \cdot \exp\left(-E/RT\right). \tag{10}$$

Darin bedeuten A eine Materialkonstante, E die Aktivierungsenergie, R die Gaskonstante 8,317 J/K · mol und T die absolute Temperatur in K. Die Integration und Umformung dieser Gleichung liefert

$$\ln\left(t/t_0\right) = \ln\frac{M}{A} + \frac{E}{RT}. \tag{11}$$

Hieraus kann abgeleitet werden, daß etwa je 7 bis 10 K eine Verdopplung bzw. Halbierung der Reaktionszeit entsteht. Dies ist die oft zitierte 7-K-Regel. Sie wird

z.B. in etwa bei Ge-Transistoren, Widerständen und Kondensatoren eingehalten, während bei Si-Bauelementen z. T. 40 K ermittelt wurden. Relativ gut experimentell ist die folgende Formel abgesichert:

$$\lambda_B = 2\lambda_0(U_B/U_0)^n \cdot (T_B - T_0)/k\,. \tag{12}$$

Die Werte mit dem Index 0 sind Bezugs- und die mit B Betriebswerte, T ist die Temperatur und U die Spannung.

Für Kondensatoren zeigt Tab. 9.5.5 typische Werte für die Konstanten n und k.

Tabelle 9.5.5. Raffungsfaktoren bei Kondensatoren

Kondensatoren	Temperaturbeanspruchung		Spannungsfaktor n in Gleichung (12)
	Bereich °C	Faktor k in Gleichung (12)	
Papier-	ohne Angabe	8,9	5
Sikatrop-	ohne Angabe	10	12
Keramik-	85	20	
(Bauform X)	85 ... 125	10	3
	125 ... 200	5	
Glimmer-	ohne Angabe	20	5
Glimmer-	ohne Angabe	5	85 °C : 6,5 125 °C : 9 150 °C : 11
Kunstfolien-	ohne Angabe	5,5	8

9.5.4. Werte der Ausfallrate

Es sind vielfältige Methoden bekannt, um die Fehlerraten von Bauelementen zu bestimmen. Allen haftet eine beträchtliche Unsicherheit schon deswegen an, da es sich hier bestenfalls um statistisch abgesicherte Mittelwerte handelt. Die zuverlässigste Methode besteht darin, eine sehr große Anzahl von Bauelementen unter den Nennbelastungen zu betreiben und die Ausfälle zu registrieren. Dabei entsteht in etwa eine Einlaufkurve, wie sie Bild 9.5.6a) zeigt. Nach ca. 2000 Stunden (Mittelwert für Halbleiterbauelemente) läuft die Kurve der Charge in einen etwa konstanten Wert, den Betriebsausfallwert, ein. Um diesen Wert zu erhalten, sind also eine Vielzahl von Bauelementen und eine beträchtliche Meßdauer notwendig. Dieser Aufwand kann folglich nur in Ausnahmefällen getrieben werden. Dieser Wert wird sich weiter von Charge zu Charge und noch mehr von Hersteller zu Hersteller ändern. Deshalb gibt es in der Regel nur Richtwerte oder garantierte Werte (die natürlich einen beträchtlichen Sicherheitsfaktor enthalten).

Unter anderem aufgrund der Betrachtungen des vorigen Abschnitts sind zeitraffende Methoden möglich. Sie sind aber nicht ganz einfach statistisch abzusichern. Unter Beachtung all dieser Fakten und der des vorigen Abschnitts gibt Tab. 9.5.6 einen Überblick zu Ausfallraten verschiedener Bauelemente. Bild 9.5.6 ergänzt dazu den Verlauf der Halbleiterwerte im Laufe der zeitlichen Entwicklung.

Sowohl bei speziellen Bauelementen (Teilbild b)) als auch über Bauelementetypen (Teilbild c) und d)) sind solche Verläufe deutlich nachzuweisen. Interessant ist, daß

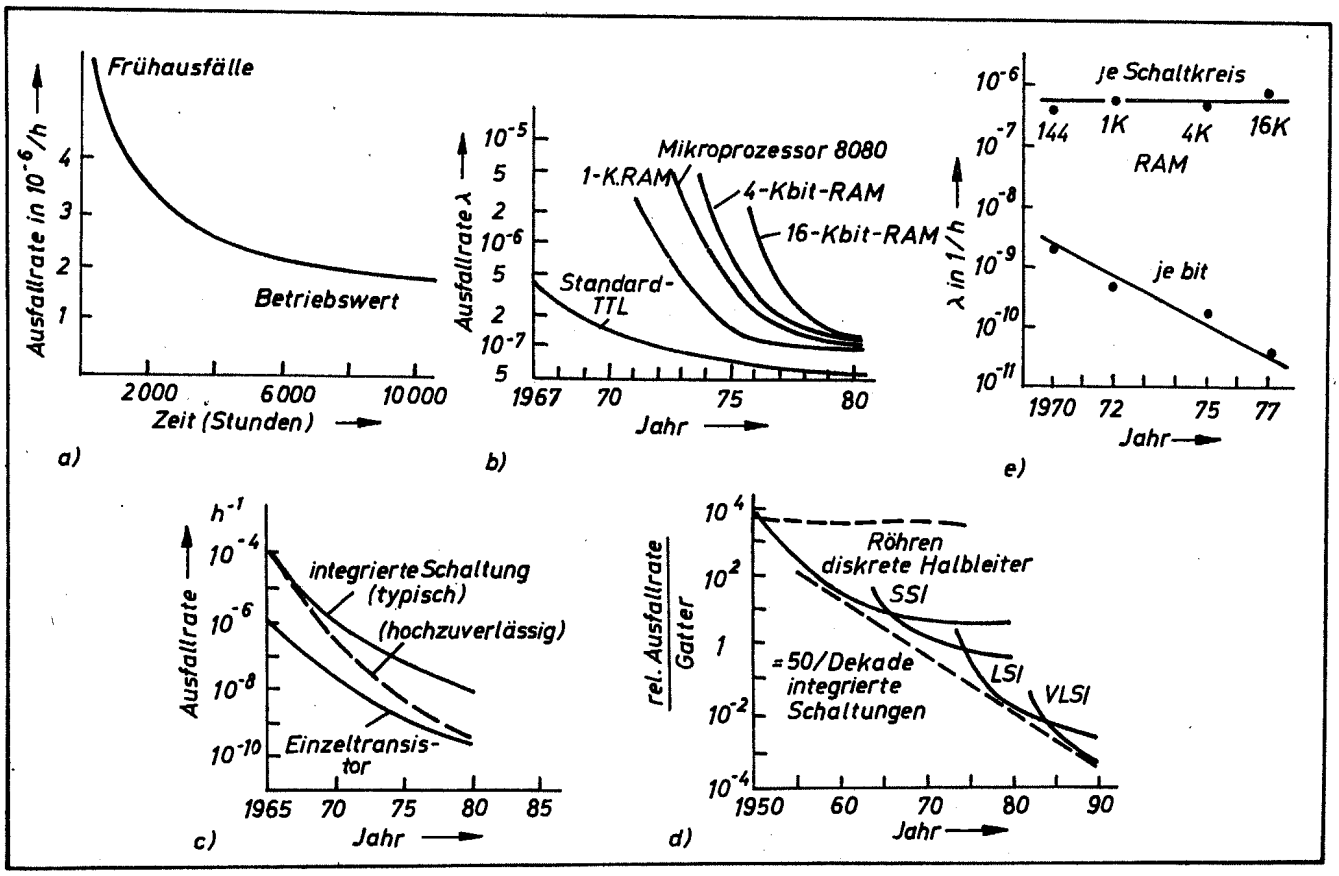


Bild 9.5.6. Zur Ausfallrate

a) Typischer Verlauf von Frühausfällen bei Halbleiterbauelementen
b) Mittelwerte einer Einlaufkurve der Ausfallrate bei einigen ausgewählten Halbleiterbauelementen
c) Mittlerer typischer Verlauf der Ausfallrate für Einzeltransistoren und integrierte Schaltkreise
d) Der ständige Abfall der Ausfallrate mit steigender Komplexität erbrachte die großen Vorteile der integrierten Schaltkreise
e) Vergleich der Ausfallraten je Speicher und je Bit an vier typischen Speichern

Tabelle 9.5.6. Richtwerte für Betriebsausfallraten einiger Bauelemente unter Nennlast und Normalbedingungen je Stunde (ohne Frühausfälle)

Schichtwiderstände	$5 \cdot 10^{-9}$ ... 10^{-7}	Röhren	10^{-6} ... $5 \cdot 10^{-4}$
Drahtwiderstände	10^{-8} ... 10^{-7}	Glühlampen	$5 \cdot 10^{-6}$ ... $5 \cdot 10^{-5}$
Potentiometer	$5 \cdot 10^{-8}$ ... 10^{-6}	Quarze	10^{-7} ... 10^{-6}
Kondensatoren	10^{-10} ... 10^{-7}	Relais	10^{-7} ... 10^{-5}
Trimmer	$2 \cdot 10^{-8}$ ... $2 \cdot 10^{-7}$	Reedkontakte	10^{-9} ... 10^{-7}
Al-Elkos	$2 \cdot 10^{-7}$ ... 10^{-6}	Schalter	ca. 10^{-7}
Ta-Elkos	$2 \cdot 10^{-8}$ ... 10^{-7}	Steckverbinder	10^{-8} ... $5 \cdot 10^{-7}$
Dioden (Signal)	10^{-8} ... 10^{-7}	Lötverbindung	10^{-9} ... 10^{-8}
Dioden (Leistung)	10^{-7} ... 10^{-6}	Wickelverbindung	10^{-10} ... 10^{-9}
Bipol. Transistoren (Signal)	$2 \cdot 10^{-9}$ ... 10^{-7}	Spulen, Drosseln	10^{-9} ... 10^{-7}
Bipol. Transistoren (Leistung)	10^{-8} ... 10^{-6}	Übertrager, Trafos	10^{-8} ... 10^{-6}
FET	10^{-8} ... 10^{-6}	Meßinstrumente	10^{-8} ... 10^{-6}
Thyristoren	$5 \cdot 10^{-8}$ ... $5 \cdot 10^{-6}$	Motoren	10^{-7} ... 10^{-6}
TTL-Schaltung	$3 \cdot 10^{-8}$ ... $3 \cdot 10^{-7}$	Akkumulatoren	ca. 10^{-6}
Analoge IC	10^{-7} ... $3 \cdot 10^{-6}$		
Unipolare IC	10^{-7} ... 10^{-6}		
LSI-Schaltkreise	10^{-7} ... 10^{-5}		
Dickschichtschaltungen	$2 \cdot 10^{-9}$ ... 10^{-7}		
Lumineszenzdioden	ca. 10^{-6}		

offensichtlich im Mittel und nach Einlauf die Fehlerrate der integrierten Schaltkreise trotz wachsender Komplexität bei ca. 10^{-7} bis 10^{-6} je h gehalten wurde (Teilbild e)). Dementsprechend fiel die Fehlerrate je Bit bzw. je Gatter im Laufe der Entwicklung steil ab (Teilbild d) und e)). Es ist zur Zeit nicht sicher, ob dies ständig der Fall sein wird. Auf ein spezielles Problem, wurde ja bereits bei den dynamischen Speichern im Zusammenhang mit Bild 8.3.16c) hingewiesen, nämlich auf die Soft-Errors infolge von α-Teilchen.

9.5.5. Zeitraffende Methoden und Frühausfälle

Je höher die Qualität eines Bauelements ist, desto schwieriger, aufwendiger ist es, diese Qualität nachzuweisen. Deshalb haben sich im Prinzip drei Methoden zur Bestimmung der Zuverlässigkeit herausgebildet:

- Laboruntersuchungen unter Normalbedingungen
- Zeitraffungsmethoden.
- Untersuchungen der Ausfallbauelemente im praktischen Einsatz.

Die letzte Methode führt zu Ergebnissen, wie sie im Abschnitt **9.5.3.1.** beschrieben wurden. Hier bestehen folgende Probleme:

Die Einsatzbedingungen sind kaum reproduzierbar, und die statische Datenerfassung ist schwierig.

Die *Laboruntersuchungen* sind eigentlich ideal, da sie in online mit einem Rechner durchgeführt werden können. Nachteilig ist jedoch die meist sehr kleine Stückzahl, an der diese Untersuchungen möglich sind, sowie die sehr große Zeitdauer der Untersuchung.

Die *Raffungsmethode* geht von hohen Belastungen der Bauelemente aus und ist daher für unterschiedliche Fehler unterschiedlich zu gestalten. In Gl. (12) sind z.B. n und k unterschiedlich für jedes Bauelement. Ebenso wirkt der Einfluß von mechanischer Belastung auf andere Fehler als auch eine Spannungsüberlast verkürzend.

Insgesamt folgt aus dem Gesagten, daß die drei Methoden nur teilweise korrelieren. Sie besitzen andererseits aber alle drei ihre spezifischen Vorteile und Nachteile. Die Zeitraffungsmethode ist u. a. besonders gut zum Überwachen einer laufenden Produktion geeignet, während die Laboruntersuchungen für Garantiewerte große Vorteile besitzen und die Untersuchung der Ausfallbauelemente gute Hinweise zur Verbesserung der Qualität liefert.

Eng im Zusammenhang mit den Zeitraffungsmethoden steht auch die Ausschaltung von Frühausfällen. Sie lassen sich einmal vermeiden, wenn die Bauelemente die entsprechende Zeit normal betrieben werden. Sie läßt sich in gewissem Umfang aber auch durch Überlast, z.B. hohe Betriebsspannung, hohen Strom, hohe Temperatur, mechanische Beanspruchung, thermische Schocks usw., erreichen. Diese Methode heißt im Englischen "burn in" (etwa: Einbrennen). Dabei ist aber zu beachten, daß eine zu hohe Überlast das Bauelement auch bereits so verändern kann, daß die stabile kleine Ausfallrate nicht mehr eintritt. Deshalb werden solche Verfahren oft im Impulsbetrieb durchgeführt. Ob die Frühausfälle bereits behoben sind, wird durch den Screening-Test festgestellt. Weiter sind Methoden bekannt, die mittelbar auf den Zustand des Bauelements schließen lassen. Dies sind z.B. Rauschspannungen, Restströme usw. Sie besitzen oft eine besondere Bedeutung für die Feststellung, ob sich ein Bauelement dem Ende seiner Lebensdauer nähert. Durch sie läßt sich eine vorbeugende Instandhaltung gut realisieren.

9.5.6. Strahlung, EMV, EMP

Strahlung kann in zwei Gruppen eingeteilt werden:

- elektromagnetische Strahlung aller Wellenlängen;
- Korpuskularstrahlung, also u. a. aus Elektronen, Protonen, Neutronen und α-Teilchen.

Infolge dieser Vielfalt treten auch unterschiedliche Auswirkungen in Bauelementen und Schaltungen der Elektronik auf. Hier sollen entsprechend der Überschrift drei Teilgebiete vorrangig behandelt werden.

9.5.6.1. Korpuskel- und γ-Strahlung

Im Abschnitt 4.2.5. und 4.2.6. wurde gezeigt, wie Korpuskelstrahlung (Elektronen, Ionen bzw. Neutronen) zur Dotierung, also zur Herstellung von Halbleitern, eingesetzt werden kann. Hier geht es aber um die Wirkung von Korpuskel- und γ-Strahlung, also energiereicher Strahlung, auf einen fertigen Halbleiter. Dabei können zwei Effekte auftreten:

Soft-Errors und Defekte.

Soft-Errors (weiche Fehler) wurden bezüglich Speicher und α-Teilchen bereits im Abschnitt 8.3.4.8. behandelt. Sie treten nur kurzfristig — in erster Näherung — während der Einwirkung der Strahlung auf und klingen danach sehr schnell ab.

Die Defekte bewirken strukturelle (dotierende) Änderungen im Halbleitermaterial, und dadurch sind dann seine Eigenschaften verändert, eventuell so stark, daß das Bauelement ausfällt. Die Grenzen für diese beiden Fehlerarten hängen sowohl vom Bauelement (-typ) als auch von der Strahlungsart ab. Die Angaben zur Strahlungseinwirkung schwanken noch beträchtlich. Einen neueren Überblick gibt Tab. 9.5.7.

Bei Neutronen stört die Anzahl je cm^2. Hier gibt es keinen Soft-Error. Bei ionisierender Strahlung (α-, β-, γ-Protonen usw.) treten ab einem relativ kleinen Strahlungsfluß (Gy/s) Soft-Errors auf. Hier sind CMOS-Schaltungen am empfindlichsten. Bleibende Fehler können sowohl durch eine hohe Gesamtdosis als auch durch einen vorübergehenden Fluß wirksam werden. Hier sind ECL und TTL bzw. MOS besonders robust.

Aus Quellen von 1970 bis 1975 stammen die Werte von Tab. 9.5.8.

9.5.6.2. EMV

Auf jedes Gerät, jede Anlage wirken elektromagnetische Felder ein. Ein 250-kW-Mittelsender erzeugt immerhin noch 1 V/m in 7 km Entfernung. Feldstärken um 100 mV/m sind eine übliche Umweltbelastung. Weitere Angaben enthält Bild 5.7.6. Bei EDVA werden heute im allgemeinen Feldstärken gefordert, die kleiner als 2,5 V/m bis 30 MHz und darüber hinaus kleiner als 1,5 V/m sind. So gibt es für alle Geräte Grenzwerte, bis zu denen sie einwandfrei funktionieren. Diese Zusammenhänge werden seit Anfang der sechziger Jahre systematisch im Zusammenhang mit den Begriffen EMC = elektromagnetic-compatibility bzw. EMV = elektromagnetische Verträglichkeit untersucht. Sie sind aber im Grunde genommen aus einer Sicht entstanden, die im nächsten Abschnitt behandelt wird. Aus militärischen Interessen gelangten sie aber erst um 1975 an die Öffentlichkeit.

Tabelle 9.5.7. Strahlungsempfindlichkeitsvergleich von verschiedenen Halbleitern und Technologien

Umwelt		Technologie								
		IFET- und diskrete, bipolare Transistoren	Thyristoren	TTL	Low-power-SCHOTTKY-TTL	Lineare IC	CMOS	*n*-MOS	LED	Isoplanar-II-ECL
Neutronen (1/cm³)		$10^{10}-10^{12}$	$10^{10}-10^{12}$	10^{14}	10^{14}	10^{13}	10^{15}	10^{15}	10^{13}	$> 10^{15}$
Ionisierungsstrahlung	Gesamtdosis (Gy (Si))	$> 10^2$	10^2	10^4	10^4	$5 \times 10^2-10^3$	$10-10^2$	10^1	$> 10^3$	10^5
	Vorübergehende Dosis (Gy (Si)/s) (Umkehrung oder Sättigung)	—	10	10^5	5×10^5	10^4	10^5	10^3	—	$> 10^6$
	Vorübergehende Dosis (Gy (Si)/s) (Überleben) (Soft-Error)	10^8	10^8	$> 10^8$	$> 10^8$	$> 10^8$	10^7	10^8	$> 10^8$	10^9

Tabelle 9.5.8. Typische Werte für störende Wirkungen von kumulierter Strahlung. Eine Neutronenstrahlung kann etwa den zehnfachen Wert der Protonenzahl besitzen.

Bauelement	Ausfall		Soft-Error Gy
	Protonen je cm^2	Gy	
Metallschichtwiderstände		10^9	10^8
Kohleschichtwiderstände	10^{14}	10^6	10^4
Papierkondensatoren		$5 \cdot 10^7$	$5 \cdot 10^4$
Mylarkondensatoren		$5 \cdot 10^8$	10^4
Keramikkondensatoren		$5 \cdot 10^{10}$	10^8
Dioden	10^{12}	10^6	10^4
NF-Transistoren	10^{11}	10^4	10^3
HF-Transistoren		10^6	10^4
Quarze	10^{12}	10^9	10^6
Ferromagnetika	10^{16}	10^{10}	10^6
Polyester		10^7	10^5
Teflon	10^{12}	10^5	10^4
Glimmer		10^9	10^6
Aluminium		—	10^{10}

9.5.6.3. Geschichte des EMP

Um 1958 zündeten die USA zwei Kernexplosionen in 40 bzw. 75 km Höhe. Dabei traten beträchtliche Störungen im Funkverkehr und bei der Radarortung auf. Etwa zu dieser Zeit trat ein Abkommen zum Testverbot für Kernwaffen ein, das bis 1961 eingehalten wurde. Bei späteren Atombombenversuchen Anfang der sechziger Jahre fielen sieben USA-Nachrichtensatelliten aus. Besonders störend wirkte sich aber der Versuch mit einer relativ kleinen Wasserstoffbombe von etwa 1,4 MT TNT aus, die am 8. 7. 1962 von den USA in 350 km Höhe gezündet wurde. Sie bewirkte, daß in Honolulu, das etwa 1500 km vom Zentrum entfernt war, die Stromversorgung zusammenbrach, weil Sicherungen durchbrannten und viele elektronische Anlagen bis zu den unterschiedlichsten Alarmanlagen seltsame Störungen zeigten. Dies war die erste bewußte Beobachtung des nuklearen elektromagnetischen Pulses (NEMP bzw. EMP). Er wurde zuvor aus mehreren Gründen nicht beobachtet. Einmal überwogen zuvor — generell gilt das für alle erdnahen Kernexplosionen — die direkten zerstörenden Wirkungen. Zum anderen waren bis dahin elektronische Schaltungen noch weitgehend mit Röhren aufgebaut, die für den EMP etwa 1000fach unempfindlicher sind. Außerdem ist das Spektrum des EMP so hochfrequent, daß es nur mit speziellen Schaltungen unmittelbar meßbar ist. Noch heute werden die Ergebnisse des EMP im wesentlichen geheimgehalten, so daß nur grundsätzliche Aussagen bekannt sind.

9.5.6.4. Wirkungsmechanismen des EMP

Generell gehen von einer Kernexplosion γ-Quanten hoher Energie geradlinig in alle Richtungen aus (Bild 9.5.7). Wird sie oberhalb der Atmosphäre gezündet, so treffen die Teilchen in einer Höhe von 50 bis 20 km auf die Lufthülle der Erde und lösen dort an den Atomen COMPTON-Elektronen aus, die mit fast Lichtgeschwindigkeit das Atom verlassen. Infolge des magnetischen Erdfeldes werden sie abgelenkt und bewegen sich bis zu ihrer Abbremsung auf einer spiralförmigen Bahn. Sowohl die Trennung von Elektron und Atom als auch die spiralförmig beschleunigte Bewegung des Elektrons bewirken eine Strahlung jedes Elektrons. Da sich auch diese mit Lichtgeschwindigkeit,

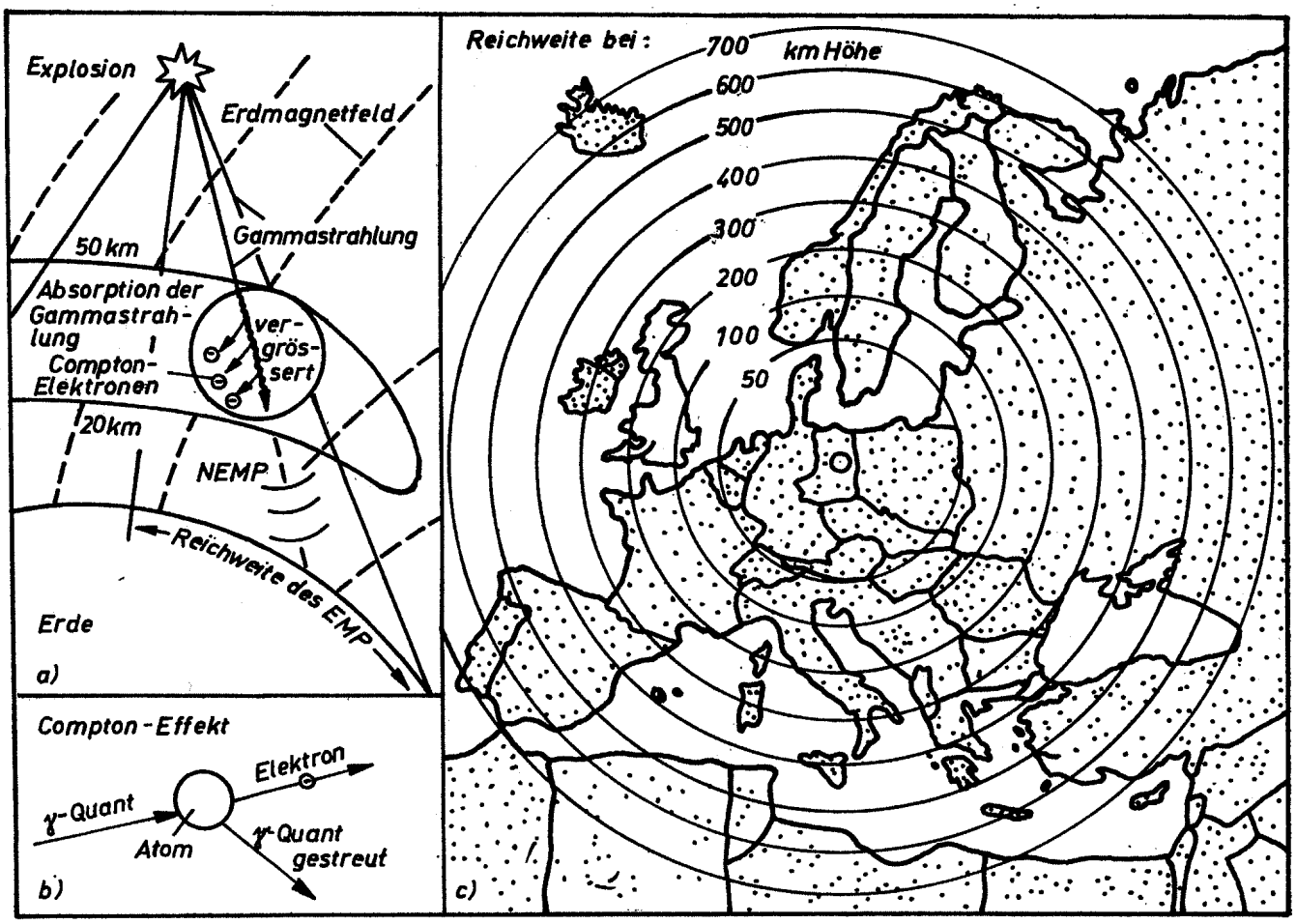


Bild 9.5.7. Zum EMP (electromagnetic pulse)

a) Prinzipielle Wirkungskette: Vom Explosionszentrum gehen γ-Strahlen aus, die an der äußeren Atmosphäre in 50 bis 20 km Höhe (konkav-konvexe Linsenform) COMPTON-Elektronen auslösen. Sie nehmen Spiralbahnen im Erdfeld an und lösen so durch Strahlung und Ladungstrennung den EMP im „Sichtbereich" auf der Erde aus

b) Prinzip des COMPTON-Effektes: Ein γ-Quant wird an einem Atom gestreut und löst dabei ein Elektron aus, das mit fast Lichtgeschwindigkeit sich vom fast ruhenden Atomrumpf entfernt

c) Gebiete, in denen der EMP wirksam wird, je nachdem in welcher Höhe die Kernexplosion gezündet wird

wie die ursprünglichen γ-Quanten, ausbreitet, addieren sich die kleinen Impulse der einzelnen Elektronen auf der Erde zu einem extrem starken und kurzen Impuls auf der Erdoberfläche. Dieser Impuls erscheint soweit auf der Erdoberfläche, wie der „Sichtbereich" der Explosion ist. Dabei treten Feldstärken von ca. 50 kV/cm für etwa 100 ns mit einer Anstiegflanke von ca. 1 ns auf. Dieser Sichtbereich ist sehr groß. Sein Radius beträgt in Abhängigkeit von der Höhe der Explosion etwa (Bild 8.5.7c)):

Höhe in km:	50	100	200	500
EMP-Radius in km:	800	1100	1600	2500

In diesem Bereich gibt es zuweilen kleine Bereiche, wo Erdfeld und Strahlung so zur Erdoberfläche stehen, daß hier der EMP ausbleibt.

Vergleicht man die Zahlenwerte des EMP mit den maximalen Feldstärken von Sendern, so sind sie um viele Zehnerpotenzen größer. Selbst ein Vergleich mit Gewitterblitzen ist bedeutsam. Er erzeugt zwar in seiner näheren Umgebung ähnliche Feldstärken, aber seine Anstiegszeit ist mit rund 1 μs wesentlich langsamer. Auch seine Dauer mit rund 100 μs ist ca. 1000mal so lang. Das Spektrum des EMP enthält also Komponenten mit bis zu 1000mal so hohen Frequenzen wie ein Blitz. Darüber hinaus wirkt der EMP nicht örtlich, sondern gleichzeitig auf einer großen Fläche. Auf allen Leitungen induziert er daher hohe elektrische Spannungen, die nicht nur empfindliche

Geräte und Anlagen zerstören, sondern, wie Honolulu zeigt, das zivilisierte Leben einer ganzen Stadt und bei systematischem Einsatz eines ganzen Landes, ja eines ganzen Erdteils vorübergehend nicht mehr funktionieren lassen. Die Kommunikation, die Energieversorgung, das Transportwesen usw. würden zusammenbrechen. Daher die zunächst geheimgehaltenen Untersuchungen zum EMC und die noch heute kaum bekannten Ergebnisse des EMP.

9.5.6.5. Schutz gegen EMP

Prinzipiell ist alle Elektronik EMP-fest zu machen, aber kaum im Nachhinein, wie dies z.B. gegenüber Blitzstörungen heute der Fall ist. Dafür ist das Spektrum des EMP zu hochfrequent. Andererseits sind gut abgeschirmte Geräte ohne Zuleitungen kaum EMP-gefährdet. Ein Transistorradio ohne Zuleitungen, nur mit eingebauten Batterien, dürfte fast immer EMP-fest sein. Wesentlich ist also einmal, „Antennen" zu vermeiden, die den EMP empfangen, und zum anderen, schnelle Kurzschlußstrecken zu finden, die den EMP in ca. 1 ns überbrücken. Beides ist heute in gewissem Umfang möglich und wird auch schon bei kommerziellen Anwendungen realisiert. Weiter sind Lichtleiter generell EMP-fest. Die moderne Entwicklung besitzt also Möglichkeiten gegen den EMP. Doch es ist bereits viel Technik installiert, die nicht EMP-fest ist. Hier liegt die Gefahr. Wesentlich für die EMP-Empfindlichkeit ist die Energie, mit der ein EMP über Antennen jeglicher Art auf die Bauelemente wirken muß, damit eine Zerstörung eintritt. Es gilt etwa in Joule:

SGFET:	10^{-5};	Leistungselektronik:	0,1 ... 10
Digitale IC:	10^{-4};	Elektronenröhren:	0,1 ... 10
Transistoren: (Kleinsignal)	10^{-4} ... 10^{-2};	Gleichrichter:	1
Signaldioden:	10^{-5} ... 10^{-2};	Motoren, Transformatoren:	10^{3} ... 10^{6}
Relais:	0,01 ... 10		

In den USA gibt es seit 10 Jahren spezielle Einrichtungen, die EMP simulieren können.

Es sei noch erwähnt, daß auch bei bodennahen Explosionen ein EMP auftritt. Seine Reichweite beträgt jedoch nur einige km. Daher ist er gegenüber den anderen hier auftretenden Wirkungen vernachlässigbar. Eine Kernexplosion in ca. 100 km Höhe ist dagegen weder zu hören noch sonst zu spüren und eventuell nicht einmal zu sehen. Hier bleibt also allein die Wirkung des EMP gefährlich.

9.5.7. Software

9.5.7.1. Einteilung möglicher Fehler

In der Rechentechnik können vier Arten von Fehlern unterschieden werden:

Hardwarefehler wurden bereits behandelt. Sie gehen auf Alterung von Bauelementen und äußere Störeinflüsse zurück. Sie können weder theoretisch noch praktisch voll vermieden werden, sondern lassen sich nur auf einen kleinen Wert reduzieren und leicht feststellbar gestalten.

Softwarefehler bestimmen die Qualität eines Programms. Sie sind theoretisch zwar vermeidbar, werden aber in der Praxis immer auftreten. Auf sie soll im folgenden genauer eingegangen werden.

Eingabefehler treten sehr häufig auf und beruhen meist auf menschlichem Versagen. Durch Lernen des Eingebenden lassen sie sich zunächst erheblich reduzieren, bleiben

dann aber etwa bei einem konstanten Wert. Eingabefehler können bei einem Online-Betrieb auch von den Sensoren herrühren, dann handelt es sich um Hardwarefehler.

Interpretationsfehler. Die Rechenergebnisse sind vom Menschen zu interpretieren. Hierbei treten oft Fehler auf, wenn nicht in ausreichendem Maße die genaue Funktion des Programms bekannt ist, und dies ist bei komplexen Programmen wohl meistens der Fall. Ihre Ursache liegt also gemischt bei menschlichem Versagen und Software.

9.5.7.2. Qualitätsmerkmale

Die Qualität einer Software wird durch mehrere Anteile bestimmt. Hier seien die wichtigsten kurz behandelt:

Der *Funktionsumfang* beschreibt die verschiedenen Möglichkeiten eines Programms. Bei einem Numerikprogramm ist es z.B. bedeutsam, ob nur die Standardfunktionen implementiert sind oder ob auch Matrizen- und Vektoroperationen, komplexe Arithmetik, Integration usw. vorhanden sind.

Die *Effizienz* ist z.B. dadurch gekennzeichnet, wieviel Platz im Arbeitsspeicher benötigt wird und wie schnell die Algorithmen eine Aufgabe bzw. Aufgabenkomplexe realisieren. Hierzu gibt es bestimmte standardisierte Methoden der Messung. Bekannt sind z.B. GIBSON-Mix, GAMM-Mix, ESER-Mix und BENCHMARK-Test. Der GIBSON-Mix wurde 1955 von J. GIBSON angegeben. Er enthält 12 Operationstypen, die mit bestimmten Gewichten g_i belegt sind. Sind T_i die Ausführungszeiten der Anlage für eine Operation mit dem Gewicht g_i, so gilt für die mittlere Operationszeit der Anlage

$$T_{\mathrm{GB}} = \sum_{i=1}^{12} T_i g_i \,.$$

Zu den Operationen gehören u. a. Festkomma- und Gleitkommaaddition, Multiplikationen, Divisionen, Transporte, bedingte Sprünge, Vergleiche, Schiebeoperationen und logische Operationen. Mit Ausnahme des BENCHMARK-Tests sind alle Mix genau festgelegt. Der BENCHMARK-Test dient dagegen nur dem mittelbaren Vergleich von Rechnern an speziell zusammengestellten Programmen.

Die *Benutzerfreundlichkeit* bezieht sich im allgemeinen auf den Komfort und die Robustheit. Sie wird besonders deutlich bei interaktiven Systemen.

Die *Flexibilität* hängt eng mit der bereits im Abschnitt 8.8. behandelten strukturierten Programmierung zusammen, denn sie beeinflußt ganz wesentlich die Möglichkeit, Programme zu warten und an neue Probleme anzupassen.

In gewissem Umfang wird auch die *Portabilität* zu den Qualitätsmerkmalen von Software gezählt. Durch sie wird angegeben, wie leicht ein vorhandenes Programm auf andere Rechnertypen übertragbar ist.

Ein hier noch genauer zu behandelndes Merkmal ist die *Zuverlässigkeit* der Software.

Schließlich gibt es noch *Zusatzleistungen*, welche die Qualität von Software nicht unerheblich beeinflussen:

Dokumentation
Schulung
Beratung und Service (Softwarepflege)
Wartung und Weiterentwicklung.

9.5.7.3. Zuverlässigkeit

Die Software zeigt meist ein ähnliches Ausfallverhalten wie Hardware. Das Ausfallverhalten wird dabei durch zwei gegenläufige Tendenzen bestimmt:

- Im Rahmen der Tests und des praktischen Einsatzes werden Fehler erkannt, die durch Programmänderungen ausgemerzt werden. Dadurch fällt ständig die Fehler- und Ausfallrate und strebt theoretisch gegen Null.
- Infolge der Fehlerbehebung sind Eingriffe und Änderungen im Programm notwendig. Durch diese Änderungen können Folgefehler auftreten, die nicht unmittelbar erkennbar sind. Mit zunehmendem Alter des Programms nehmen folglich die Folgefehler zu, und die Ausfallrate steigt an.

Aus diesen beiden Tendenzen sind also — ähnlich wie bei der Hardware drei Fehlerbereiche zu unterscheiden:

Bei den *Frühausfällen* überwiegen die wirksamen Korrekturen, und die Ausfallrate sinkt ab.

Bei dem *Normalfall* werden mit jeder Fehlerkorrektur wieder genauso viele Folgefehler erzeugt. Es liegt also in etwa eine konstante Fehlerrate vor.

Das *Ende der Betriebszeit* tritt dann ein, wenn mit jeder Korrektur mehr Folgefehler erzeugt als eigentliche Fehler behoben werden. Das Programm wird hier also instabil und muß durch ein neues ersetzt werden.

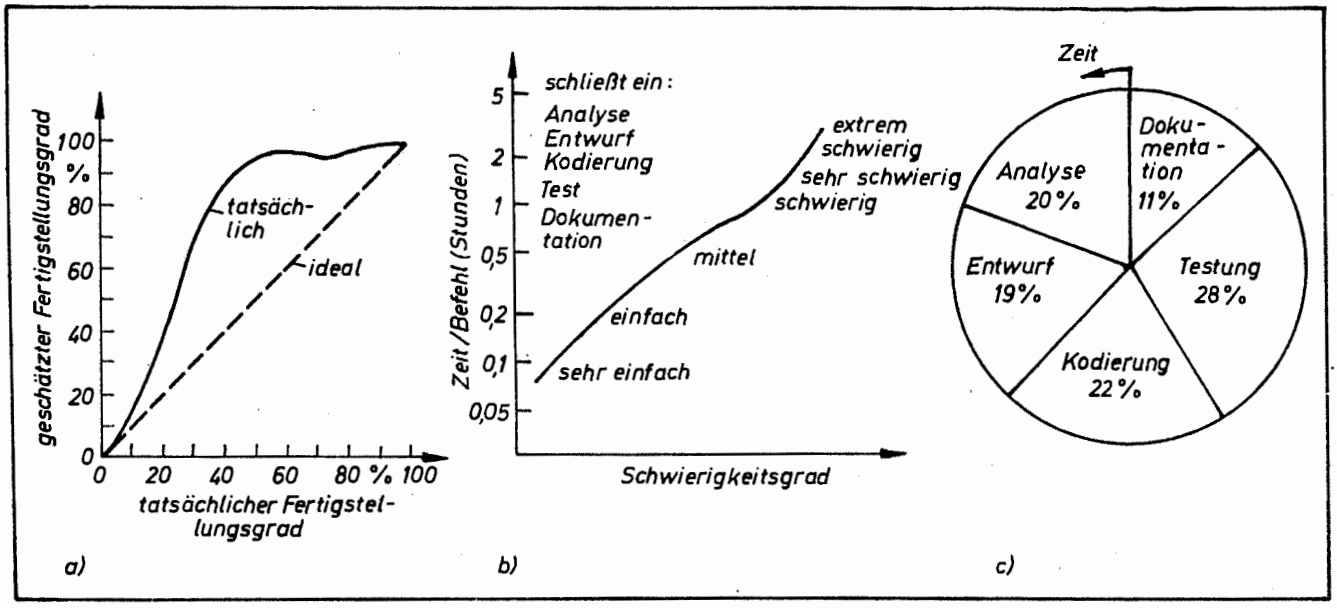


Bild 9.5.8. Zur Softwareentwicklung

a) Verhältnis von subjektiver und objektiver Fertigstellung eines Programms
b) Zeitaufwand für einen Befehl als Funktion der Schwierigkeit eines Programms
c) Gemittelte Anteile der Phasen bei der Entwicklung eines Programms

Gegen alle Fehler gibt es spezielle Testmethoden, Arbeitshilfsmittel. Ein Beispiel ist die schon behandelte strukturierte Programmierung und das Top-down. Dennoch sind ideale Programme, bei denen die Fehlerrate gegen Null strebt, eine sehr seltene Ausnahme.

Bezüglich der Entwicklung der Software seien noch die typischen Kurven von Bild 9.5.8 angegeben. Sie zeigen auf, daß generell der noch zu leistende Aufwand — also die Fehlerbehebung — unterschätzt wird (Teilbild a)) und wie mit dem Schwierigkeitsgrad eines Programms die Zeit zunimmt, die für einen richtigen Befehl aufzuwenden ist (Teilbild b)). Eine Untersetzung zu Teilbild a) gibt Teilbild c) durch den mittleren Aufwandsanteil in den verschiedenen Entwicklungsphasen.

9.6. Einige Grundlagen der Informationstheorie

Die Informationstheorie ist heute ein sehr umfangreiches theoretisches Gebäude der Nachrichtentechnik. Sie hat Anwendungen in sehr vielen Wissenschaften gefunden. Daher ist es hier nicht möglich, auch nur alle Grundlagen gestrafft wiederzugeben. Was hier mitzuteilen ist, bezieht sich vor allem auf zwei Ergebnisse, auf die mehrmals im Text Bezug genommen werden mußte, nämlich die Entropie und Kanalkapazität, d. h. die Grundlagen der Definition des Bit (bit) und des Informationsflusses in bit/s. Weitere Details müssen aus den Standardwerken, z.B. [F 30, M 29, P 28, S 71, V 12, Z 8], entnommen werden. Hier wird der Einfachheit halber eine besonders anschauliche, also weniger mathematische Herleitung der Zusammenhänge versucht. Damit sollen gleichzeitig deutlicher einige inhaltliche Bezüge bewußt gemacht werden.

9.6.1. Kanalmodell

Die grundlegenden Arbeiten von SHANNON setzen ein Kanalmodell gemäß Bild 9.6.1 voraus. Ein Sender ist die Quelle der Information. Sie wird über einen gestörten Kanal dem Empfänger zugeführt. Alle Störungen, die in dieser Strecke entstehen, also auch das Empfängerrauschen, werden hierbei zur Vereinfachung in der Störquelle des Kanals zusammengefaßt. Bei diesen Betrachtungen ist es nicht notwendig auszusagen,

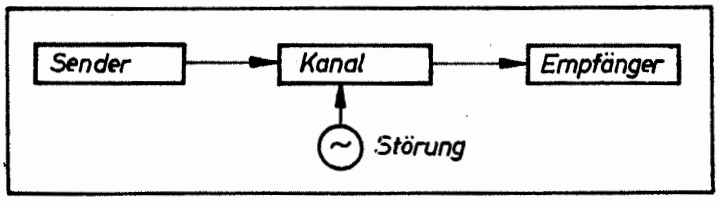


Bild 9.6.1. Kanalmodell der Informationstheorie

was Information sei. Es werden nur die physikalischen Eigenschaften von Informationsträgern und ihre Wahrscheinlichkeit betrachtet, z.B. wie oft in einer großen Anzahl von Übertragungen die Pegel L bzw. H vorkommen. Sie werden allgemein Alphabet genannt. Bei kontinuierlichen Signalen ist das Alphabet unendlich. Es erfüllt z.B. den Spannungsbereich von −10 V bis 10 V beliebig dicht. Dann sind die Wahrscheinlichkeiten für einzelne Symbole des Alphabets durch Wahrscheinlichkeitsdichten zu ersetzen. Weiter ist die große Relativität der Schnittstellen Sender—Kanal bzw. Kanal—Empfänger wichtig. Sie können in einer komplizierten Übertragungsstrecke nahezu beliebig gewählt werden. Das ist für den Informationsfluß über diese Schnittstellen bedeutsam. So wie die Informationstheorie keine Aussagen über die Information trifft, sagt sie auch nichts dazu aus, wie Informationen entsteht, genauer erzeugt wird bzw. wie sie genutzt wird, d. h., wie sie im Empfänger wirkt.

Die Informationstheorie von SHANNON wurde vor allem dazu geschaffen, zu berechnen, wie gut ein Nachrichtenkanal ausgenutzt wird. Sie besitzt damit große Ähnlichkeit zum CARNOTschen Kreisprozeß der Thermodynamik. Mit ihm kann berechnet werden, wie hoch bestenfalls der Wirkungsgrad einer Wärmekraftmaschine werden kann. Dadurch ist es möglich, Aussagen zu machen, wie nahe eine reale Wärmekraftmaschine der theoretisch idealen kommt. Analoges leistet die Informationstheorie. Sie gibt an, wieviel Information im Idealfall über einen Kanal übertragbar ist. Dadurch ist eine Einschätzung (Leistung) eines realen Nachrichtensystems möglich.

Genau wie die CARNOT-Theorie keine konstruktiven Hinweise zur Verbesserung gibt, ist es im wesentlichen auch bei der SHANNON-Theorie. Wichtig für den Wirkungsgrad

eines Nachrichtensystems ist die jeweils verwendete Kodierung (Modulation) der zu übertragenden Symbole. Deshalb ist die Kodierung in der SHANNON-Theorie ein wichtiger Bestandteil.

9.6.2. Entropie

Die Entropiebetrachtungen beziehen sich auf die Schnittstelle Sender—Kanal in Bild 9.6.1. Der Sender verfüge über ein Alphabet X mit den Symbolen x_n, also

$$x_n \in X\,, \qquad 1 \leqq n \leqq m\,. \tag{1}$$

Die Anzahl der Symbole steht also fest. Ihr Wert ist m. Beispiele sind die binären Signale 1 und Ø, also $m = 2$, die Augen eines Würfels, also $m = 6$, die Buchstaben eines Schriftalphabets a, b, c, d, ... , z, also $m = 26$. Die Festlegung dieses Alphabets unterliegt einer gewissen Willkür, die durch das Ziel der Betrachtungen bestimmt ist. So können die 32 Blatt eines Kartenspiels einzeln, jede für sich als Symbol betrachtet werden. Dann wird aus einem gut gemischten Stapel jede einzelne Karte mit einer Wahrscheinlichkeit von $^1/_{32} = 3{,}13\%$ auftreten. Es kann aber auch eine Zusammenfassung erfolgen:

Benennung	Karten	Wahrscheinlichkeit
Z	Zahlen: 7; 8; 9; 10	1/2
M	männlich: Bube, König	1/4
D	Dame	1/8
A	As	1/8

Jedem verwendeten Symbol x_n des Alphabets ist also immer eine Wahrscheinlichkeit p_n zugeordnet. Notwendige Voraussetzung für die Anwendung der Informationstheorie ist folglich die Existenz des Paares

$$x_n/p_n \tag{2}$$

mit

$$\sum_{n=1}^{m} p_n = 1\,. \tag{3}$$

Über den Kanal sollen nun die Symbole x_n übertragen werden, und der Empfänger soll dabei nach jeder Symbolübertragung feststellen, um welches Symbol es sich gehandelt hat. Es sei vorausgesetzt, er kann hierbei bezüglich spezieller Eigenschaften der Symbole nur entscheiden, ob sie vorliegen oder nicht. Bezüglich des oben genannten Kartenspiels bedeutet es, zu erraten, welche von den vier Möglichkeiten: Z, M, D, A vorliegt. Dabei sind an den Sender nur Fragen zulässig, die mit ja oder nein eindeutig beantwortbar sind. Weiter ist eine solche Fragestrategie zu wählen, daß in einem langen Spiel mit möglichst wenig Fragen möglichst viel Zuordnungen bestimmt werden. Dieser Problemstellung entspricht der Wirkungsgrad. Je Symbol sind dabei zwei Fälle zu unterscheiden. Die mittlere Anzahl der Fragen je Symbol, die im konkreten Fall benötigt wurden, und jene, die theoretisch minimal möglich ist. Die letzte Anzahl ist die Entropie je Symbol. Sie sei im folgenden bestimmt. Beide Größen können in der Anzahl der mittels zwei möglicher Antworten (ja/nein) notwendigen Fragen gemessen werden. Die Anzahl der Fragen können auch als Entscheidungen oder Schritte bezeichnet werden. Daraus leitet sich die Maßeinheit des Bit gemäß *b*inary dig*it* (englisch digit = Finger, Ziffer, Schritt) ab.

9.6.2.1. Gleichverteilung

Es liegen m Symbole vor. Sie treten alle mit der gleichen Wahrscheinlichkeit $1/m$ auf. Ein Beispiel ist der Würfel. Hierbei läßt sich folgendermaßen vorgehen. Die Symbole werden zur ersten Frage in zwei möglichst gleich große Gruppen geteilt:

$$\left.\begin{aligned} \mathrm{a} &:= x_1, x_2 \ldots x_{m/2}\,, \\ \mathrm{b} &:= x_{m/2+1} \ldots x_{m-1}, x_m\,. \end{aligned}\right\} \tag{4}$$

Die erste Frage lautet:

Gehört das Symbol zur Gruppe a ? Je nach der Antwort ist nur noch die richtige Gruppe weiter zu betrachten. Sie wird wiederum geteilt in zwei möglichst gleich große Gruppen c und d. Auch hier wird die Frage gestellt: Ist es c ? So läßt sich das unbekannte Symbol am schnellsten erfassen. Bei den m Symbolen sind dann genau

$$q = \{\mathrm{ld}\, m\} \tag{5a}$$

Fragen notwendig. Dabei bedeutet ld den Logarithmus zur Basis 2 und $\{\,\}$ die nächsthöhere ganze Zahl. Mittelt man über viele Spiele und läßt eine weiter optimierte Fragestrategie zu, so gilt exakt der gemittelte Grenzwert der minimal möglichen Fragen

$$q = \mathrm{ld}\, m\,. \tag{5}$$

Umgekehrt lassen sich nämlich mit q Fragen

$$2^q = m^+ \tag{6}$$

Symbole unterscheiden. Also immer, wenn $m = m^+$ ist, ergibt sich sofort ein übersichtliches Verfahren. Dies sei mit Bild 9.6.2 für die acht Symbole 1 bis 8 dargestellt. Liegen nur sechs Symbole vor, so wird bei der ersten Frage der gestrichelte Weg beschritten. Die Frage „< 7 ?“ ist nicht mehr notwendig. Da alle Symbole 1 bis 6 mit gleicher Wahrscheinlichkeit auftreten sollen, sind im Mittel

4mal 3 Fragen und 2mal 2 Fragen

notwendig. Daraus folgt, daß

$$4 \cdot 3 \cdot \tfrac{1}{6} + 2 \cdot 2 \cdot \tfrac{1}{6} \approx 2{,}67$$

Fragen im Mittel je Symbol auftreten werden. Gemäß Gl. (5) folgt dagegen etwa 2,385. Die Fragestrategie ist also nicht optimal. Ein Beispiel hierzu folgt im nächsten Abschnitt. Zuvor sei jedoch noch die obige Zahlenrechnung verallgemeinert. Sie kann als realisierte Entropie gemäß

$$H_q = \sum_{n=1}^{m} q_n p_n \tag{7}$$

betrachtet werden, wobei q_n die für das Symbol p_n im Kodebaum vorhandene Fragenanzahl ist.

9.6.2.2. Unterschiedliche Wahrscheinlichkeiten

Es sei wieder die oben genannte Einteilung des Skatspiels verwendet. Dann wird man zunächst die Fragestrategie gemäß Bild 9.6.2b) versuchen. Sie liefert $H_q = 2$. Wird dagegen die „unübliche“ Fragestrategie gemäß Bild 9.6.2c) verwendet, so fällt H_q auf 1,75. Die Fragestrategie ist folglich besser, und es entsteht die Frage: Ist sie bereits optimal ? Dies muß die Informationstheorie beantworten.

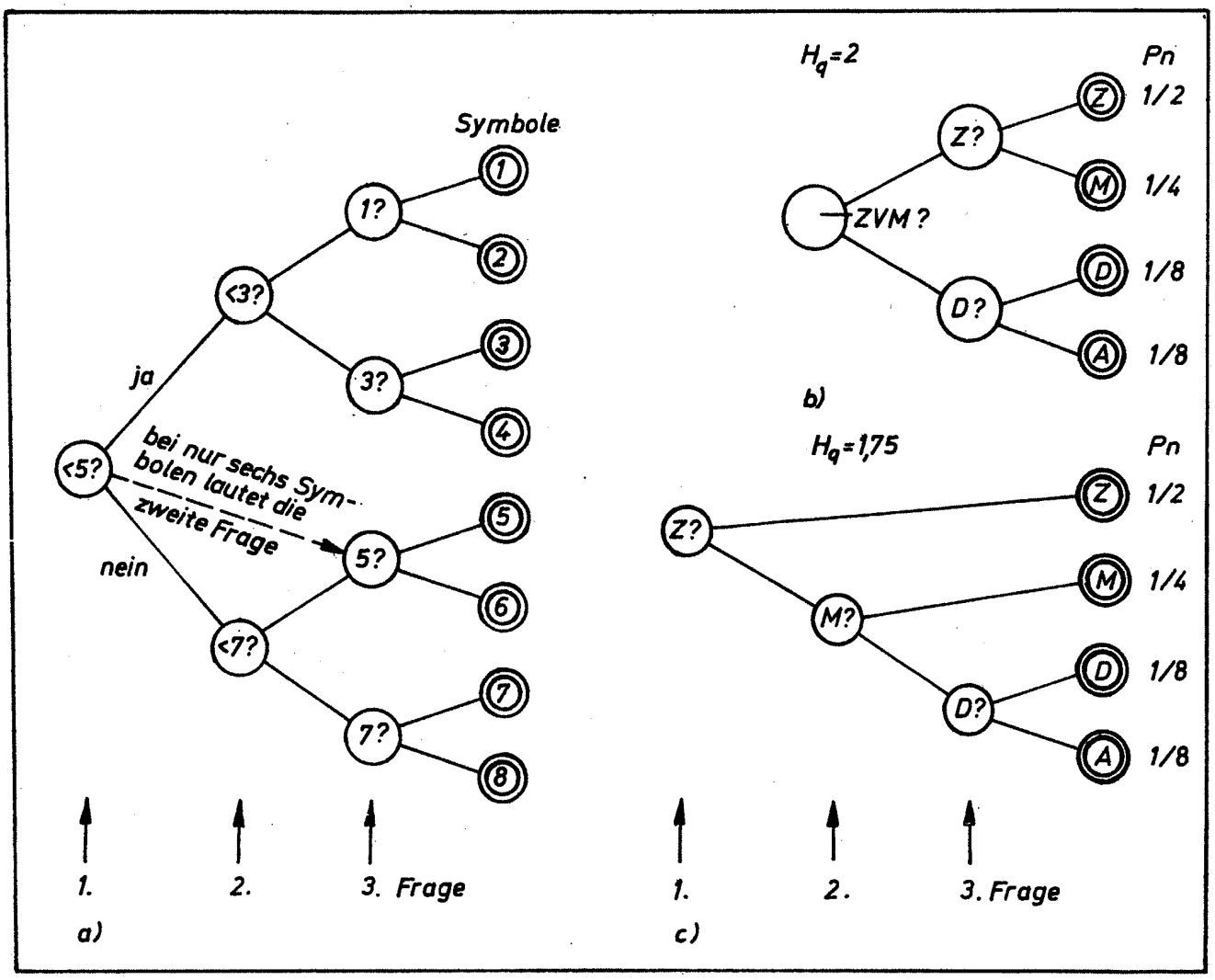


Bild 9.6.2. Zur Kodierung in der Form von Fragebäumen

a) Zur Erklärung der Fragestrategie, um mit möglichst wenig Fragen das Ziel zu erreichen. Nach der Frage bedeutet ein Schritt nach oben „ja", einer nach unten „nein". Existieren nur die Symbole 1 bis 6, so kann die Frage „< 7?" entfallen, und es erfolgt der gestrichelte Übergang bei nein auf „5?"

b) Naheliegende Entscheidungsstrategie für vier Klassen, wie sie für das Skatspiel im Text gewählt wurde

c) Wie b), jedoch optimale Strategie

9.6.2.3. Eine anschauliche Herleitung

Für die konkreten Fälle entstanden die beiden Beziehungen von Gl. (5) und (7). Bei Gleichverteilung könnte in Gl. (5) statt m auch die Wahrscheinlichkeit p eingesetzt werden. Sie lautete dann

$$q = \mathrm{ld}\,\frac{1}{p} = -\,\mathrm{ld}\,p\,. \tag{8}$$

Wird dieses Ergebnis in Gl. (7) jedoch mit den auf das jeweilige Symbol bezogenen Wahrscheinlichkeiten eingesetzt, so folgt die Entropieformel von SHANNON

$$\boxed{H = -\sum_{n=1}^{m} p_n\,\mathrm{ld}\,p_n}\,. \tag{9}$$

Sie gibt theoretisch die minimal mögliche Fragezahl an. Auf den mathematischen Beweis sei hier verzichtet. Für das Beispiel von Bild 9.6.2b), c) folgt $H = 1{,}75$. Die obere Fragestrategie ist also optimal. Wie die optimale aussieht und ob sie überhaupt

existiert, darauf gibt die Entropieformel keine Aussage. Sie sagt nur aus, daß der durch sie berechnete Wert prinzipiell nicht zu unterschreiten ist.

9.6.2.4. Verlauf von $-p \operatorname{ld} p$

In der Entropieformel steht der Ausdruck

$$h = -p \operatorname{ld} p \,. \tag{10}$$

Sein Verlauf ist also für die Informationstheorie von grundlegender Bedeutung. Er ist deshalb in Bild 9.6.3a) gezeigt. Die Funktion besitzt einen unsymmetrischen Verlauf. Ihr Maximum nimmt sie bei der Wahrscheinlichkeit $p = e^{-1} \approx 0{,}37$ mit $h \approx 0{,}53$ an. Dieser Fakt hat durchaus praktische Bedeutung. Die 37% gelten als Überraschungswert. Tritt z.B. in einem Text der Selbstlaut e mit etwa 37% auf, so ist dies e maximal betont. Ähnliches gilt für Farben in einem Bild. Wenn z.B. gelb 37% ausmacht, wird das Bild als gelb bezeichnet.

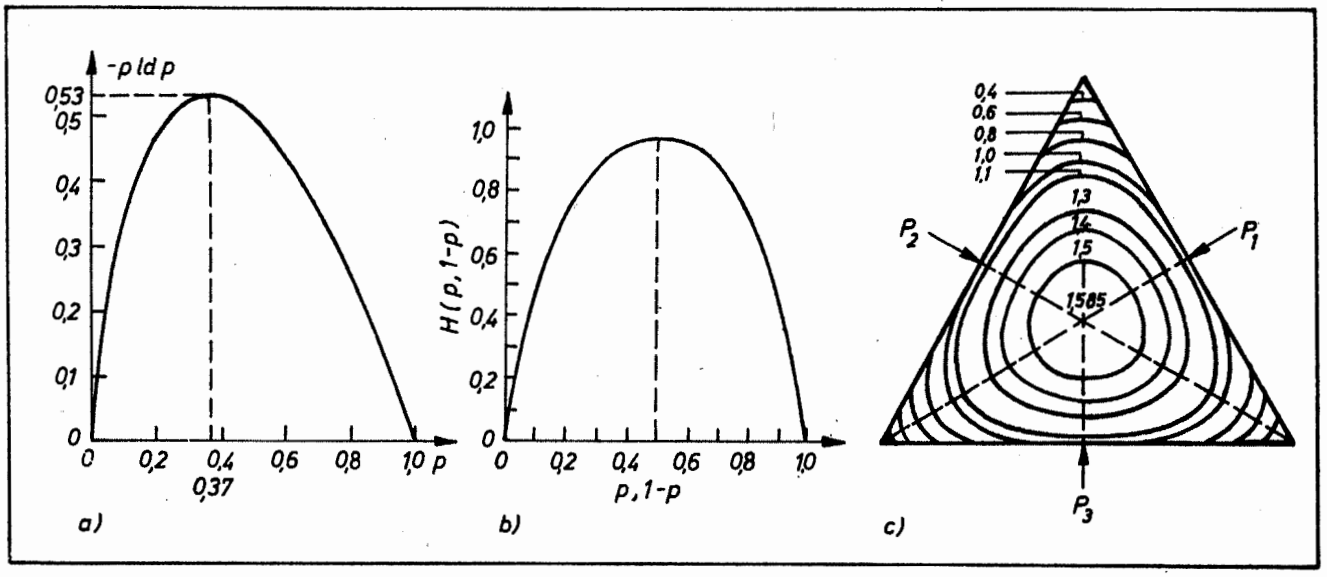


Bild 9.6.3. Kurvendarstellungen zur Entropie

a) Verlauf der Grundfunktion $h = -p \operatorname{ld} p$

b) Entropie einer Binärquelle, bei der die beiden Symbole mit den Wahrscheinlichkeiten p bzw. $1-p$ auftreten

c) Entropie eines ternären Kanals

9.6.2.5. Binäre Quelle

Kann nun eine Quelle nur die beiden Symbole 1 und Ø aussenden, so gilt

$$p_1 = 1 - p_0 \,. \tag{11}$$

Die Entropie ist damit nur von einem p, z.B. p_1 abhängig, und es gilt

$$H_{\text{binär}} = -p \operatorname{ld} p - (1-p) \operatorname{ld} (1-p) \,. \tag{12}$$

Den Verlauf dieser Funktion zeigt Bild 9.6.3b). Sie hat ihr Maximum bei gleicher Wahrscheinlichkeit für beide Symbole, alsp $p = 0{,}5$. Hier ist genau eine Frage je Symbol notwendig. Überwiegt eine Wahrscheinlichkeit, so sind im mittel weitaus weniger Fragen notwendig. So etwas läßt sich z.B. realisieren, wenn nach mehreren gesendeten Symbolen zusammenhängend gefragt wird, z.B.: Ist unter jeweils 10 Symbolen wenigstens ein L ? Diese Frage ist z.B. sinnvoll, wenn L mit einer Wahrscheinlichkeit von 90% auftritt. Für sehr kleine Wahrscheinlichkeiten gilt für Gl. (12) die

Näherung

$$H_{\text{binär},\, p\to 0} = p \operatorname{ld} \frac{p}{e}. \tag{13}$$

9.6.2.6. Ternäre Quelle

Liegt eine Quelle mit drei Symbolen 1, 2, 3 vor, so gilt für die Wahrscheinlichkeiten

$$p_1 + p_2 + p_3 = 1\,. \tag{14}$$

Ein derartiger Zusammenhang führt zu den Dreieckskoordinaten, wie sie unter anderem in der Chemie für ternäre Verbindungen verwendet werden. Die Entropie ist hierin durch Linien konstanter Entropie, also Isoentropien darstellbar. Den entsprechenden Zusammenhang zeigt Bild 9.6.3c). Bei Gleichverteilung gilt

$$p_1 = p_2 = p_3 = \tfrac{1}{3}\,, \tag{15}$$

und die zugehörige Entropie beträgt

$$H_{\text{tern, max}} = \operatorname{ld} 2 \approx 1{,}585\,. \tag{16}$$

9.6.2.7. Kontinuierliche Quelle

Für eine diskrete Quelle gilt Gl. (9) für die Entropie. Für einen kontinuierlichen Kanal sind die Wahrscheinlichkeiten durch Wahrscheinlichkeitsdichten $p(x)$ zu ersetzen. Dabei stellt x den unabhängig variablen Parameter, z.B. eine Spannung, dar. Als entsprechender Grenzübergang zu kleinem Δx folgt

$$H = -\frac{1}{x_0} \int_{-\infty}^{+\infty} p(x) \operatorname{ld} (p(x))\, \mathrm{d}x\,. \tag{17}$$

Die Größe x_0 folgt aus der Divergenz des Grenzüberganges. Sie ist also frei wählbar. Dies bedeutet, daß die Entropie einer kontinuierlichen Quelle bis auf einen Faktor unbestimmt bleibt. Es sind also nur Relationen zwischen zwei Entropien genau bestimmbar. Bei den diskreten Quellen ergab sich, daß eine Gleichverteilung die maximal mögliche Entropie gibt. Es läßt sich zeigen, daß sie bei einer kontinuierlichen Quelle für die GAUSS-Verteilung

$$p(x) = \frac{1}{2\pi} \frac{x_0}{x} \cdot \mathrm{e}^{-x^2/2\bar{x}}\,. \tag{18}$$

eintritt. Dabei bedeutet $\bar{x}$ den zeitlichen, quadratischen Mittelwert

$$\bar{x}^2 = \frac{1}{T} \int_{\lim T\to\infty}^{T} x^2(t)\, \mathrm{d}t\,. \tag{19}$$

Unter diesen Bedingungen geht Gl. (12) über in

$$H_{\max} = \operatorname{ld} \sqrt{2\pi\, \mathrm{e}}\, \frac{\bar{x}}{x_0}. \tag{20}$$

Auch hier ist noch die unbestimmte Größe x_0 enthalten. Der Faktor beträgt

$$\sqrt{2\pi\, \mathrm{e}} \approx 4{,}133\,. \tag{21}$$

Es kann nun aber vorausgesetzt werden, daß x nur auf eine Größe $\bar{x}_{st}$ genau bestimmt werden kann. Sie kann z.B. als thermodynamisches Rauschen aufgefaßt werden. Sie soll ebenfalls der Maximalentropie entsprechend der GAUSS-Verteilung gehorchen. Dann gelte als quadratische Summe (Rauschaddition)

$$\bar{x}_g = \sqrt{\bar{x}^2 + \bar{x}_{st}^2}\,. \tag{22}$$

Hierfür gilt

$$H_g = \mathrm{ld}\,\sqrt{2\pi\,\mathrm{e}}\,\frac{\bar{x}_g}{x_0} \tag{23}$$

und

$$H_{st} = \mathrm{ld}\,\sqrt{2\pi\,\mathrm{e}}\,\frac{\bar{x}_{st}}{x_0}. \tag{24}$$

Die Differenz aus beiden beträgt

$$H_{\mathrm{Nutz}} = H_g - H_{st} = \mathrm{ld}\,\frac{\bar{x}_g}{\bar{x}_{st}}. \tag{25}$$

Wenn also der Mittelwert des Nutzsignals und des Störsignals bekannt sind, kann die Entropie der kontinuierlichen Quelle genau berechnet werden.

Bei der eben durchgeführten Berechnung war vorausgesetzt, daß beide, die Störung und der Signalwert, der GAUSS-Verteilung folgen. Ist dies nicht der Fall, so werden genauso Abweichungen wie bei der Gleichverteilung von diskreten Quellen auftreten. Die Entropie wird kleiner, wenn die Signalquelle einer anderen Verteilung gehorcht.

Die Gl. (25) läßt sich nun besonders gut interpretieren, wenn $\bar{x}_g$ und $\bar{x}_{st}$ als Effektivwerte von Spannungen aufgefaßt werden. Dies ist wegen Gl. (19) und (22) besonders einsichtig. $\bar{x}_g$ bedeutet dann den Meßwert und $\bar{x}_{st}$ die ihm überlagerten Schwankungen. Das Verhältnis

$$n_{AS} = \frac{\bar{x}_g}{\bar{x}_{st}} \tag{26}$$

ist dann der reziproke Fehler. Dieser Quotient wird in der Informationstheorie als Anzahl unterscheidbarer Amplitudenstufen bezeichnet. Er entspricht also in etwa den unterscheidbaren Signalen der diskreten Quelle. Damit ist wieder der Zusammenhang zur Gl. (5) hergestellt. Sie war ja der Spezialfall der Entropieformel von Gl. (9) für den Fall der Gleichverteilung.

9.6.2.8. Verbundentropie

Bisher war stets vorausgesetzt, daß jeweils nur ein Symbol gesendet wird. Das beim zweiten Versuch gesendete Symbol sollte dabei vom ersten unabhängig sein. Dies ist z.B. beim Würfeln erfüllt. Bei der Sprache ist es anders; auf ein Q folgt fast mit Sicherheit ein U, während auf ein S mit unterschiedlicher Wahrscheinlichkeit C, T, P, A usw. folgen können. Es sind also in solchen Fällen drei Wahrscheinlichkeiten zu unterscheiden:

- eine Wahrscheinlichkeit für das Auftreten eines Symbols $p(x)$,
- eine *bedingte Wahrscheinlichkeit* $p(y|x)$.
  Sie gibt an, wie groß die Wahrscheinlichkeit für ein y ist, wenn vorher bereits x aufgetreten war.

— eine *Verbundwahrscheinlichkeit* $p(y, x)$.
Sie gibt an, wie groß die Wahrscheinlichkeit ist, daß x und y bei einem Versuch hintereinander auftreten. Für sie gilt offensichtlich

$$p(x, y) = p(x) \cdot p(y/x) \,. \tag{27}$$

Eine derartige Kopplung zwischen aufeinanderfolgenden Ereignissen wird oft mit MARKOFF-Kette bezeichnet. Die Abhängigkeit kann über h Symbole bestehen, dann liegen MARKOFF-Ketten h-ter Ordnung vor.

Neben der zeitlichen Verkopplung von Symbolen kann sie auch bei mehreren Quellen existieren, die gleichzeitig Symbole aussenden. Für beide Quellen kann nun die Verbundentropie berechnet werden. Allgemein gilt dann

$$H(x, y) = -\sum_i \sum_j p(x_i, y_j) \operatorname{ld} p(x_i, y_j) \,. \tag{28}$$

Mit Gl. (27) folgt

$$H(x, y) = -\sum_i \sum_j p(x_i) \cdot p(y_j/x_i) \left[\operatorname{ld} p(x_i) + \operatorname{ld} p(y_j/x_i)\right] . \tag{29}$$

Allgemein gilt

$$\sum_j p(y_j/x_i) = 1 \,. \tag{30}$$

Wenn x_i eingetreten ist, herrscht Gewißheit, daß irgendein y_j auftritt. Deshalb folgt

$$H(x, y) = -\sum_i p(x_i) \operatorname{ld} p(x_i) - \sum_i p(x_i) \left[\sum_j p(y_j/x_i) \operatorname{ld} p(y_j/x_i)\right] . \tag{31}$$

Die Verbundentropie besteht also aus zwei Teilen:

$$H(x, y) = H(x) + H(y/x) \,, \tag{32}$$

wobei $H(y/x)$ bedingte Entropie heißt. Damit wird eine fundamentale Ungleichung unmittelbar einsichtig:

$$H(x) \leqq H(x, y) \leqq H(x) + H(y) \,. \tag{33}$$

Das linke Gleichheitszeichen gilt, wenn beide Prozesse voll abhängig sind, also in etwa wie bei Q und U. Das rechte Gleichheitszeichen gilt bei fehlender Abhängigkeit, also wie beim Würfel.

Ein Beispiel für das Absinken der Entropie bei MARKOFF-Ketten zeigt Bild 9.6.4. Es gilt für die deutsche Sprache und wurde von KÜPFMÜLLER bestimmt. Es gilt hier eine ungefähre Kettenlänge von $h \approx 7$. Es sei erwähnt, daß der Wert 4,037 bereits auf der Statistik der Buchstaben in der deutschen Sprache beruht. Wären alle Buchstaben gleichwahrscheinlich, so betrüge die Entropie ld 26 ≈ 4,70.

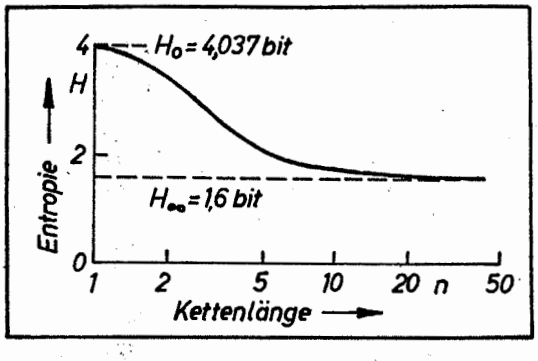


Bild 9.6.4. Abhängigkeit der Entropie H bezüglich der Buchstabenhäufigkeit der deutschen Sprache sowie Verbundentropie für die entsprechenden MARKOFF-Ketten der Länge n

9.6.3. Abläufe in der Zeit

In den meisten Anwendungsfällen dürfte nicht nur der Informationsgehalt, die Entropie eines Symbols interessieren, sondern auch, wieviel Information in der Zeit über einen bestimmten Kanal übertragen werden kann. Auch hier sind wieder der reale Wert unter jeweils gegebenen Bedingungen, der Informationsfluß J, und die theoretische Grenze, die Kanalkapazität C, zu unterscheiden. Beide Größen werden in bit/s gemessen. In der Telegraphie wird statt dessen Baud (Bd) gewählt.

9.6.3.1. Abtasttheorem

Die einzelnen Symbole müssen über einen z.B. elektrischen Kanal in der Regel nacheinander übertragen werden. Dies möge durch sprunghaftes Umschalten auf genau definierte Spannungspegel erfolgen. Am Ausgang des Kanals vollziehen sich dann Einschwingvorgänge, wie sie unter anderem bereits im Abschnitt 1.5.2. und 2.2.3. behandelt sind. Sie begrenzen die Übertragungsgeschwindigkeit, denn es muß jeweils solange gewartet werden, bis das Symbol eindeutig zugeordnet werden kann. KÜPFMÜLLER fand bereits 1924, daß zwischen Einschwingzeit τ und Bandbreite B eines Kanals die Beziehung

$$2B\tau = 1 \tag{34}$$

besteht. Diesen Fakt präzisierte KOTELNIKOW 1933 mathematisch exakt mittels des Abtasttheorems. Es heißt zuweilen auch Probensatz oder Samplingtheorem. Es sagt aus, daß zur Beschreibung eines Signals mit der oberen Grenzfrequenz f_g Meßwerte (Abtastwerte) mit einem zeitlichen Abstand

$$\Delta T = \frac{1}{2f_g} \tag{35}$$

völlig ausreichen. Während die KÜPFMÜLLER-Beziehung also eigentlich für zeitlich diskrete Signale gilt, erfaßt das Abtasttheorem auch zeitlich und in der Amplitude kontinuierliche Signale. Im Samplingtheorem kann bei Signalen, die oben und unten im Frequenzbereich begrenzt sind, auch wieder die Bandbreite B eingeführt werden, so daß

$$\Delta T = \frac{1}{2B} \tag{36}$$

gilt.

9.6.3.2. Kanalkapazität

Gemäß Gl. (36) können pro Zeiteinheit

$$x = 2B \tag{37}$$

Proben übertragen werden. Die Proben sind dabei unabhängig, können also jeweils einem Symbol zugeordnet werden. Hat jedes Symbol nun die Entropie H, so folgt die Kanalkapazität zu

$$C = 2BH\,. \tag{38}$$

Für den diskreten Kanal gilt daher mit Gl. (9) der Zusammenhang

$$C_{\mathrm{dis}} = -2B \sum_{n=1}^{m} p_n \operatorname{ld} p_n\,. \tag{39}$$

Für den kontinuierlichen Kanal gilt mit Gl. (17) entsprechend

$$C_{\text{kon}} = -\frac{2B}{x_0}\int_{-\infty}^{+\infty} p(x)\,\text{ld}\,p(x)\,\text{d}x\,. \tag{40}$$

Bei Gaussschem Rauschen mit dem Effektivwert $\bar{x}_{\text{st}}$ und einem gaußverteilten Signal mit dem Effektivwert $\bar{x}_{\text{g}}$ gemäß Gl. (25) gilt

$$C_{\text{kon}} = 2B\,\text{ld}\frac{\bar{x}_{\text{g}}}{\bar{x}_{\text{st}}}. \tag{41}$$

Hier können nun die Signalwerte durch die Signalleistung P_{N} und die Störleistung P_{st} ersetzt werden. Dabei sei $\bar{x}$ ein Spannungswert. Wegen Gl. (22) folgt dann

$$\bar{x}_{\text{g}} = \frac{\sqrt{P_{\text{N}} + P_{\text{st}}}}{R} \tag{42}$$

bzw.

$$\bar{x}_{\text{st}} = \frac{\sqrt{P_{\text{st}}}}{R}, \tag{43}$$

wobei R den Abschlußwiderstand des Kanals darstellt. Wird beides in Gl. (41) eingesetzt, so folgt die bekannte Shannon-Gleichung

$$\boxed{C_{\text{kon}} = B\,\text{ld}\left(1 + \frac{P_{\text{N}}}{P_{\text{st}}}\right)} \tag{44}$$

Das Verhältnis $P_{\text{N}}/P_{\text{st}}$ entspricht dem bekannten Störabstand. Es ist zunächst erstaunlich, daß auch dann, wenn er klein gegen Eins ist, immer noch fehlerfreie Signalübertragungen möglich sind. Dies beweisen aber andererseits Methoden, wie sie z. B. im Abschnitt 7.7.3. behandelt wurden.

Im allgemeinen wird nun aber nicht die Rauschleistung und Signalleistun einer Gauss-Verteilung gehorchen. Auch brauchen beide nicht gleich intensiv innerhalb der Bandbreite B aufzutreten. Sie müssen also, anders ausgedrückt, nicht weiß sein. Alle solche Effekte können durch einen Faktor k^+ berücksichtigt werden. Dann gilt

$$C = B\,\text{ld}\,k^+\left(1 + \frac{P_{\text{N}}}{P_{\text{st}}}\right). \tag{45}$$

9.6.3.3. Information und Energie

Nur allzuoft wird angenommen, daß Information nichts mit Energie zu tun hat. Dies folgt aus der Formulierung von Wiener „information is information, not energy or matter". Es besteht jedoch ein gewisser Zusammenhang. Er sei aus Gl. (44) abgeleitet. Dazu werde die Störleistung thermodynamisch bedingt gewählt:

$$P_{\text{st}} = B\,kT\,. \tag{46}$$

Hierin bedeuten k die Boltzmann-Konstante mit $k = 1{,}381 \cdot 10^{-23}\,\text{JK}^{-1}$ und T die absolute Temperatur. Die Nutzleistung sei z-mal so groß:

$$P_{\text{N}} = zP_{\text{st}}\,. \tag{47}$$

Dann folgt aus Gl. (44)

$$C = B\,\text{ld}\,(1 + z)\,. \tag{48}$$

Es werde nun das Verhältnis aus Nutzleistung und Kanalkapazität bestimmt;

$$\frac{P_N}{C} = kT \frac{z}{\mathrm{ld}\,(1+z)}. \tag{49}$$

Dieser Wert besitzt die Einheit J bit^{-1}. Er kann also auch als Energie W je Information I betrachtet werden.

$$\frac{W}{I} = \frac{P_N}{C}. \tag{50}$$

Einer genaueren Betrachtung bedarf noch der Ausdruck mit z. Hierin werde zunächst der duale Logarithmus durch den natürlichen ersetzt, dann gilt die Reihenentwicklung

$$\ln(1+z) = z - \frac{z^2}{2} + \frac{z^3}{3} - \frac{z^4}{4} + - \ldots . \tag{51}$$

Damit gilt weiter

$$1 < \frac{z}{\ln(1+z)} \xrightarrow{\text{für } z \to 0} 1 . \tag{52}$$

So folgt aus Gl. (49) die Grenzenergie für 1 bit:

$$W_{gr} = kT \ln 2 . \tag{53}$$

Sie muß mindestens für ein Bit aufgewendet werden. Sie ist natürlich extrem klein. Bei 300 K sind es etwa

$$3 \cdot 10^{-21}\,\mathrm{J} \approx 7 \cdot 10^{-22}\,\mathrm{cal} \approx 19\,\mathrm{meV} .$$

Diese Energie entspricht der thermodynamischen Grenzspannung eines pn-Überganges: 19 mV $\approx$ 27 mV $\cdot$ ln 2.

10. Verzeichnisse

10.1. Bücherverzeichnis

Der Umfang der Literatur zur Elektronik ist so stark gewachsen, daß es schon aus Platzgründen nicht mehr möglich war, die relativ vollständige Erfassung, wie sie in den vergangenen Auflagen erfolgte, fortzusetzen. Deshalb wurde hier eine speziell eingeengte Buchauswahl getroffen. Sie umfaßt im wesentlichen jene Fachbücher, die zu den behandelten Gebieten der Elektronik und zu wichtigen Randgebieten gehören, wobei ausschließlich Bücher in deutscher Sprache berücksichtigt wurden. Um die Aufstellung weiter zu verkürzen, wurden für häufiger auftretende Verlage folgende Kurzformen verwendet:

AV	Akademie-Verlag GmbH, Berlin
BIL	VEB Bibliographisches Institut, Leipzig
BIM	Bibliographisches Institut, Mannheim, Wien, Zürich
Bir	Birkhäuser-Verlag, Basel, Stuttgart
BU	Verlag Berliner Union GmbH Verlag W. Kohlhammer GmbH, Stuttgart
DVG	VEB Deutscher Verlag für Grundstoffindustrie, Leipzig
DVW	VEB Deutscher Verlag der Wissenschaften, Berlin
FV	VEB Fachbuchverlag, Leipzig
Fra	Franzis-Verlag, München, Stuttgart
GP	Akademische Verlagsgesellschaft Geest und Portig, K.-G., Leipzig
Han	Carl Hanser Verlag, München, Wien
HirL	S. Hirzel Verlag, Leipzig
HirS	S. Hirzel Verlag, Stuttgart
Hü	Dr. Alfred Hüthig Verlag, Heidelberg
HüPf	Hüthig und Pflaum, München, Heidelberg
MiV	Militärverlag der Deutschen Demokratischen Republik, Berlin
Old	R. Oldenbourg-Verlag, München und Wien
RFK	Verlag für Radio-Foto-Kinotechnik, Berlin-Borsigwalde
SprH	Springer-Verlag, Berlin, Heidelberg, New York
SprW	Springer-Verlag, Wien, New York
TeuL	BSB B. G. Teubner Verlagsgesellschaft, Leipzig
TeuS	B. G. Teubner, Stuttgart
VDI	VDI-Verlag GmbH, Düsseldorf
Vie	Friedrich Vieweg und Sohn Verlagsgesellschaft, Braunschweig
VT	VEB Verlag Technik, Berlin

[A 1] Achterberg, H.: Operationsverstärker: Grundlagen. Valvo GmbH, Hamburg 1974.
[A 2] Ackmann, W.: Zuverlässigkeit elektronischer Bauelemente. Hü, Heidelberg 1976.
[A 3] Ahlers, H., u. a.: Bauelemente und Schaltungsentwurf. VT, Berlin 1978.
[A 4] Anonym: Hütte IV B — Fernmeldetechnik. Verlag Wilhelm Ernst & Sohn, Berlin, München 1962.
[A 5] Anonym: Rauschen. NTZ-Beiheft 2. Vie, Braunschweig 1955.
[A 6] Halbleiter-Handbuch. Kombinat VEB Keramische Werke Hermsdorf, 1973.
[A 7] Anonym: Ferrite-Handbuch. Kombinat VEB Keramische Werke Hermsdorf, 1976.
[A 8] Anonym: Piezolan-Handbuch. Kombinat VEB Keramische Werke Hermsdorf, 1977.

[A 9] Anonym: Magnetische Eigenschaften von Festkörpern. DVG, Leipzig 1975.
[A 10] Anonym: Telefunken Laborbuch, Bd. 1—15. Fra, München 1969—1971.
[A 11] Anonym: Probleme der Festkörperelektronik, Bd. 1—11. VT, Berlin 1972—1981.
[A 12] D'Ans, J., Lax, E.: Taschenbuch für Chemiker und Physiker. SprH, Berlin u. a. 1949.
[A 13] Anstreich, K.: Rechnergestützter Schaltungsentwurf. NTG-Tagung 1975, Stuttgart. VDE, Berlin(West) 1975.
[A 14] Ardenne, M. v.: Tabellen zur angewandten Physik, Bd. 3. DVW, Berlin 1973.
[A 15] Arnolds, F.: Elektronische Meßtechnik. BU, Stuttgart 1976.
[A 16] Ausborn, W.: Elektronikbauelemente (6. Aufl.). VT, Berlin 1979.
[A 17] Autorenkollektiv: Grundlagen aktiver elektronischer Bauelemente. DVG, Leipzig 1972.
[A 18] Autorenkollektiv: Grundlagen passiver elektronischer Bauelemente. DVG, Leipzig 1973.
[A 19] Autorenkollektiv: Zuverlässigkeit elektronischer Bauelemente. DVW, Berlin 1974.
[A 20] Autorenkollektiv: Magnetische Eigenschaften von Festkörpern. DVG, Leipzig 1975.
[A 21] Autorenkollektiv: Entwurf mikroelektronischer Schaltungen. AV, Berlin 1976.
[A 22] Autorenkollektiv: Messung elektronischer Parameter an Halbleiterbauelementen. FV, Leipzig 1976.
[A 23] Autorenkollektiv: Halbleiterbauelementeelektronik. AV, Berlin 1977.
[A 24] Autorenkollektiv: Vakuumelektronik. AV, Berlin 1978.
[A 25] Autorenkollektiv: Grundlagen der Schaltungsentwicklung. AV, Berlin 1979.
[A 26] Axmann, H. P.: Einführung in die technische Informatik. Funktionsweise digitaler Bausteine und deren Verwendung in Datenerfassungssystemen. SprW, Wien 1979.
[A 27] Anonym: Information — Applikation. Mikroelektronik, H. 1 bis H. 9. VEB Halbleiterwerk Frankfurt/Oder, Frankfurt/Oder 1982/93.
[A 28] Appel, U.: Mikrocomputer und Minicomputer. Old, München, Wien 1981.
[B 1] Baily, F. J.: Halbleiterschaltungen. Old, München 1974.
[B 2] Balcke, E., Krause, H.: Grundlagen der analogen Schaltungstechnik, VT, Berlin 1981.
[B 3] Barkhausen, H.: Elektronenröhren, Bd. 1, Allgemeine Grundlagen (12. Aufl.) HirL, Leipzig 1969.
Dito, Bd. 2. Verstärker (10. Aufl.). 1968.
Dito, Bd. 3. Rückkopplungen (10. Aufl.). 1969.
[B 4] Barthold, H., Bäurich, H.: Mikroprozessoren — Mikroelektronische Schaltkreise und ihre Anwendung, Bd. 1—3. MiV, Berlin 1980/1981.
[B 5] Bauer, F. L., Goos, G.: Informatik. SprH, Berlin 1973.
[B 6] Bauer, A., u. a.: Technologie und Anwendung von Ferroelektrika. GP, Leipzig 1976.
[B 7] Baumann, P.: Halbleiterpraxis. VT, Berlin 1976.
[B 8] Baumann, W., u. a.: Standard-Interfaces der elektronischen Meßtechnik. VT, Berlin 1974.
[B 9] Baeuclair, W. de: Rechnen mit Maschinen. Vie, Braunschweig 1968.
[B 10] Bechstein, E.: Aus der Geschichte der Automatisierungstechnik. VT Berlin 1974.
[B 11] Becker, D., Mäder, H.: Hochintegrierte MOS-Schaltungen. BU, Stuttgart 1973.
[B 12] Becker, E., Beyer, P.: Weichmagnetische Ferrit-Bauelemente und ihre Anwendung. MiV, Berlin 1974.
[B 13] Becker, R.: Theorie der Elektrizität, Bd. 2. TeuL, Leipzig 1949.
[B 14] Bender, D.: Werkstoffkunde. Elektroberufe (7. Aufl.). DVG, Leipzig 1977.
[B 15] Bender, K., u. a.: Mikrorechner — Struktur und Programmierung. VDI, Düsseldorf 1977.
[B 16] Beneking, H.: Feldeffekttransistoren. SprH, Berlin 1973.
[B 17] Bening, F.: Negative Widerstände in elektronischen Schaltungen. VT, Berlin 1971.
[B 18] Benndorf, H., u. a.: Fernwirktechnik. VDI, Düsseldorf 1975.
[B 19] Berg, C.: Einführung in die Operatorenrechnung. DVW, Berlin 1965.
[B 20] Bergtold, F.: Umgang mit Operationsverstärkern. Old, München 1973.
[B 21] Bergtold, F.: Schaltungen mit Operationsverstärkern, Bd. 1 u. 2 (2. Aufl.). Old, München 1975.
[B 22] Bernhard, J. M.: Klein-Computer, Bd. 2. Vogel, Würzburg 1972.
[B 23] Bernhard, J. M., Knuppertz, B.: Analoge Informationsverarbeitung — kurz und bündig. VT, Berlin 1974.
[B 24] Bernstein, H.: Hochintegrierte Digitalschaltungen und Mikroprozessoren. Pflaum, München 1978.

[B 25] Bernstein, H.: Mikroprozessor (2. Aufl.). Hofacker, München 1978.
[B 26] Bishop, G. D.: Einführung in lineare elektronische Schaltungen. Vie, Braunschweig 1977.
[B 27] Bitterlich, W.: Elektronik. SprW, Wien, New York 1967.
[B 28] Bladowski, R.: Integrierte Analogschaltungen. Telekosmos-Verl., Stuttgart 1970.
[B 29] Blatt, E., Fleissner, H.: Prozeßdatenverarbeitung. VDI, Düsseldorf 1976.
[B 30] Bleicher, M.: Halbleiter — Optoelektronik. Hü, Heidelberg 1978.
[B 31] Blomeyer-Bartenstein, H. P.: Mikrocomputertechnik (4. Aufl.). Hofacker, München 1979.
[B 32] Bocker, P.: Datenübertragung. SprH, Berlin 1976.
[B 33] Bochmann, D.: Einführung in die strukturelle Automatentheorie. VT, Berlin 1975.
[B 34] Boerner, H.: Phasenkopplungssysteme in der Nachrichten-, Meß- und Regelungstechnik. VT, Berlin 1976.
[B 35] Bohm, J.: Kristalle. DVW, Berlin 1975.
[B 36] Boll, R.: Weichmagnetische Werkstoffe (3. Aufl.). Vakuumschmelze, Hanau 1977.
[B 37] Bopp, A.: Grundschaltungen der Analog-Elektronik (Lehrbuchreihe Elektronik). BU, Stuttgart 1979.
[B 38] Borodin, L. F.: Einführung in die Theorie der störsicheren Kodierung. BP, Leipzig 1972
[B 39] Borucki, L.: Grundlagen der Digitaltechnik. TeuS, Stuttgart 1977.
[B 40] Bystron, K.: Technische Elektronik. Han, München 1974.
[B 41] Breuer, R., Lechleitner, H.: Der lautlose Schlag. Meyster-Verlag, München 1982.
[B 42] Bläsing, K.-H., Schlenzig, K.: Experimente mit Speicherschaltkreisen. MiV, Berlin 1982.
[B 43] Boll, R., u. a.: Magnettechnik. Expert-Verlag, Grafenau/Württ. 1980.
[B 44] Beuth, K., Schmusch, W.: Grundschaltungen der Elektronik. Vogel, Würzburg 1976.
[C 1] Cauer, W.: Theorie linearer Wechselstromschaltungen (2 Bände). AV, Berlin 1954 u. 1960.
[C 2] Cedighian, S.: Die magnetischen Werkstoffe. VDI, Düsseldorf 1973.
[C 3] Celio, T.: Die photoelektronischen Abtastmethoden in der Technik der Bildwiedergabe. Bir, Basel, Stuttgart 1975.
[C 4] Classen, L.: Programmierung des Mikroprozessorsystems U 880 — K 1520. VT, Berlin 1981.
[C 5] Conrad, W.: Meyers Taschenlexikon Elektronik. Funktechnik. BIL, Leipzig 1978.
[C 6] Czech, J.: Oszillographen-Meßtechnik. RFK, Berlin 1959.
[C 7] Czirr, E.: NF-Meßpraxis. VT, Berlin 1966.
[D 1] Danner, G. M., Gatermann, H. G.: Methodischer Entwurf digitaler Funktionsgruppen, Geräte und Anlagen. Old, München, Wien 1978.
[D 2] Delfs, H.: Hybridschaltungen. BU Stuttgart 1973.
[D 3] Dirks, C., Krinn, H.: Microcomputer II (Lehrbuchreihe Elektronik). BU, Stuttgart 1978.
[D 4] Dobesch, H., Sulanke, H.: Zeitfunktionen — Theorie und Anwendung. VT, Berlin 1962.
[D 5] Dobesch, H.: Laplace-Transformation (3. Aufl.). VT, Berlin 1966.
[D 6] Dombrowski, E.: Einführung in die Zuverlässigkeit elektronischer Geräte und Systeme. AEG-Telefunken.
[D 7] Domsch, G.: Der Übertrager der Nachrichtentechnik. GP, Leipzig 1953.
[D 8] Domsch, G.: Die Abschirmung magnetischer Felder in der Nachrichtentechnik. VT, Berlin 1955.
[D 9] Dosse, J., Mierdel, G.: Der elektrische Strom im Hochvakuum und in Gasen. HirL, Leipzig 1945.
[D 10] Drachsel, R.: Grundlagen der elektrischen Meßtechnik. VT, Berlin 1977.
[D 11] Dreszer, J.: Mathematik-Handbuch. FV, Leipzig 1975.
[D 12] Dummer, W. A., Griffin, N. B.: Zuverlässigkeit in der Elektronik. VT, Berlin 1968.
[D 13] Duske, J., Juergensen, H.: Codierungstheorie. BIM, Mannheim u. a. 1977.
[D 14] Dworatschek, S.: Einführung in die Datenverarbeitung (3. Aufl.). De Gruyter, Berlin 1970.
[E 1] Ebert, J., Jürris, E.: Digitale Meßtechnik, VT, Berlin 1971.
[E 2] Eckhardt, D., u. a.: Digitale Grundschaltungen und ihre Anwendung. VT, Berlin 1974.
[E 3] Eckhardt, D., Gross, W.: Grundlagen der digitalen Schaltungstechnik (2. Aufl.). MiV, Berlin 1975.

[E 4] Eckhardt, D., u. a.: Entwurf komplexer digitaler Schaltungen. VT, Berlin 1976.
[E 5] Eckhardt, D., u. a.: Hochintegrierte digitale Schaltungen und ihre Anwendung (3. Aufl.). VT, Berlin 1980.
[E 6] Eckhardt, D., u. a.: Mikroelektronik — Stand und Entwicklung. AV, Berlin 1981.
[E 7] Ehrenstrasser, G.: Stochastische Signale und ihre Anwendung. Hü, Heidelberg 1974.
[E 8] Elschner, H., Möschwitzer, A.: Neue Bauelemente der Informationselektronik. GP, Leipzig 1974.
[E 9] Erlekamp, R., u. a.: Mikroelektronik in der Amateurpraxis. MiV, Berlin 1980.
[E 10] Ergelage, D., Dallwitz, L.: Grundlagen supraleitender elektronischer Bauelemente und Schaltungen. GP, Leipzig 1978.
[E 11] Entenmann, W.: CCD-Filter. Old, München, Wien 1980.
[F 1] Fasol, K. H., Vingron, R.: Synthese industrieller Schaltungen. AV, Berlin 1975.
[F 2] Fedetow, J. A., Pohl, H. J.: Fotolithografie — Grundlagen und Anwendung in der Halbleitertechnologie. DVW, Berlin, Moskau 1974.
[F 3] Feldkeller, R.: Vierpoltheorie (5. Aufl.). HirS, Stuttgart 1948.
[F 4] Feldkeller, R.: Einführung in die Siebschaltungstheorie. HirS, Stuttgart 1956.
[F 5] Feldkeller, R.: Theorie der Spulen und Übertrager. HirS, Stuttgart 1963.
[F 6] Fender, M.: Fernwirken. TeuS, Stuttgart 1981.
[F 7] Fetzer, V.: Einschwingvorgänge in der Nachrichtentechnik. VT, Berlin 1958.
[F 8] Fetzer, V.: Integraltransformationen. Hü, Heidelberg 1977.
[F 9] Fliege, N.: Lineare Schaltungen mit Operationsverstärkern. SprH, Berlin 1979.
[F 10] Franz, D.: Chemische Stromquellen. MiV, Berlin 1968.
[F 11] Frey, H.: Einführung in die Ortskurventheorie. MiV, Berlin 1974.
[F 12] Fricke., H., u. a.: Grundlagen der elektrischen Nachrichtenübertragung. TeuS, Stuttgart 1979.
[F 13] Friebel, W., u. a.: Programmierung von Mikroprozessoren. VT, Berlin 1978.
[F 14] Friedrich: Tabellenbücher Elektrotechnik (19. Aufl.). FV, Leipzig 1981.
[F 15] Frieser, H.: Photographische Informationsaufzeichnung. Old, München 1975.
[F 16] Fritzsch, W.: Prozeßrechentechnik. VT, Berlin 1981.
[F 17] Fritzsche, G.: Entwurf linearer Schaltungen. VT, Berlin 1962.
[F 18] Fritzsche, G.: Systeme — Felder — Wellen. VT, Berlin 1975.
[F 19] Fritzsche, G.: Theoretiche Grundlagen der Nachrichtentechnik. VT, Berlin 1976.
[F 20] Fritzsche, G.: Informationsübertragung (Wissensspeicher). VT, Berlin 1977.
[F 21] Fritzsche, G.: Grundlagen und Entwurf passiver Analog-Zweipole. AV, Berlin 1979.
[F 22] Fritzsche, G.: Entwurf passiver Analog-Vierpole. AV, Berlin 1979.
[F 23] Fritzsche, G.: Entwurf aktiver Analogsysteme. AV, Berlin 1980.
[F 24] Fritzsche, G.: Zeitdiskrete und digitale Systeme. AV, Berlin 1981.
[F 25] Frommhold, E. A.: FET-Analogschalter. MiV, Berlin 1980.
[F 26] Fruhauf, W.: Grundlagen der elektronischen Meßtechnik. GP, Leipzig 1977.
[F 27] Fuchs, H., Könitzer, L.: Digitale Meßwerterfassung (2. Aufl.). VT, Berlin 1975.
[F 28] Funke, R., Liebscher, S.: Grundschaltungen der Elektronik (8. Aufl.) VT, Berlin 1979.
[F 29] Fritzsche, G., Seidel, V.: Aktive RC-Schaltungen in der Elektronik. VT, Berlin 1981.
[F 30] Fey, P.: Informationstheorie. AV, Berlin 1977.
[G 1] Gad, H.: Feldeffektelektronik. TeuS, Stuttgart 1976.
[G 2] Gaedler, J.: Konstruktion elektronischer Geräte, Bd. 1 u. 2. Inst. f. Nachrichtentechnik, Berlin 1981.
[G 3] Geisselhardt, W.: Fehlerdiagnose in Geräten der Digitaltechnik. Han, München 1978.
[G 4] Geist, D.: Halbleiterphysik, Bd. 1 u. 2. Vie, Braunschweig 1969/1970.
[G 5] Gerlach, W.: Thyristoren. SprH, Berlin u. a. 1979.
[G 6] Giloi, W. K., Liebig, H.: Logischer Entwurf digitaler Systeme (2. Aufl.). SprH, Berlin u. a. 1973.
[G 7] Giloi, W. K.: Rechnerarchitektur. SprH, Berlin 1980.
[G 8] Glade, H., Manteuffel, K.: Am Anfang stand der Abacus. Urania, Leipzig u. a. 1973.
[G 9] Glaser, W., Kohl, G.: Mikroelektronik (3. Aufl.). FV. Leipzig 1973.
[G 10] Glaser, W.: Lichtleitertechnik. VT, Berlin 1981.

[G 11] Goercke, P., Mischel, P.: Optoelektronische Bauelemente für die Automatisierung. Hü, Heidelberg 1976.
[G 12] Göldner, K.: Mathematische Grundlagen der Systemanalyse. FV, Leipzig 1981.
[G 13] Gorjunowa, N. A.: Halbleiter mit diamantähnlicher Struktur. TeuL, Leipzig 1971.
[G 14] Görke, W.: Mikrorechner — Eine Einführung in ihre Technik und Funktion. BIM, Mannheim 1978.
[G 15] Görlich, P.: Die Fotozellen. GP, Leipzig 1964.
[G 16] Gössel, M.: Angewandte Automatentheorie, Bd. 1 u. 2. AV, Berlin 1972.
[G 17] Gottschalk, H.: Elektronische Bausteine der Digitaltechnik. VT, Berlin 1970.
[G 18] Gould, I. H.: IFIP — Sachwörterbuch der Datenverarbeitung. TeuL, Leipzig 1977.
[G 19] Grass, W.: Steuerwerke — Entwurf von Schaltwerken mit Festwertspeichern. SprH, Berlin 1978.
[G 20] Grave, H.: Elektrische Messung nichtelektrischer Größen. GP, Leipzig 1962.
[G 21] Greig, D.: Elektronen in Metall und Halbleitern. AV, Berlin 1971.
[G 22] Greuel, O.: Mathematische Ergänzungen und Aufgaben für Elektrotechniker (9. Aufl.). FV, Leipzig 1978.
[G 23] Grosse, E., Weissmantel, C.: Der gestörte Kristall. Urania, Leipzig 1979.
[G 24] Gruhle, W.: Elektronische Hilfsmittel des Physikers. SprH, Berlin 1960.
[G 25] Günther, H.: Eigenschaften des RC-Doppel-T-Vierpols. VT, Berlin 1953.
[G 26] Göllnitz, H., u. a.: Vakuumelektronik. AV, Berlin 1978.
[H 1] Haack, O.: Einführung in die Digitaltechnik. TeuS, Stuttgart 1972.
[H 2] Hadamovsky, H. F.: Halbleiterwerkstoffe. DVG, Leipzig 1968.
[H 3] Hahn, L., u. a.: Werkstoffkunde für die Elektrotechnik und Elektronik (2. Aufl.) VT, Berlin 1976.
[H 4] Hahn, W.: Elektronik — Praktikum für Informatiker. SprH, Berlin u. a. 1970.
[H 5] Hahn, W., Bauer, F. L.: Physikalische und elektrotechnische Grundlagen. SprH, Berlin u. a. 1975.
[H 6] Hailpert, H. W.: Halbleiterbauelemente (2. Aufl.) TeuS, Stuttgart 1976.
[H 7] Hamann, C., u. a.: Organische Leiter, Halbleiter und Photoleiter. AV, Berlin 1980.
[H 8] Hanke, H. J., Fabian, H.: Technologie elektronischer Baugruppen. VT, Berlin 1975.
[H 9] Hantzsch, H.: Wärmeableitung bei Halbleitern. MiV, Berlin 1978.
[H 10] Harms, G.: Digitale Schaltkreise. Vogel, Würzburg o. J.
[H 11] Hart, H.: Einführung in die Meßtechnik. VT, Berlin 1977.
[H 12] Harth, W.: Halbleitertechnologie. TeuS, Stuttgart 1972.
[H 13] Harth, W., Classen, M.: Aktive Mikrowellendioden. SprH, Berlin 1981.
[H 14] Hartl, P.: Fernwirktechnik der Raumfahrt. SprH, Berlin 1977.
[H 15] Haug, A.: Elektronisches Messen mechanischer Größen. Han, München 1969.
[H 16] Heck, C.: Magnetische Werkstoffe und ihre technische Anwendung (2. Aufl.). Hü, Heidelberg 1975.
[H 17] Heine, K.: Laufzeit-Dioden. Old, München 1976.
[H 18] Hellwege, K. H.: Einführung in die Festkörperphysik. SprH, Berlin 1976.
[H 19] Henkler, O.: Anwendung der Modulation beim Trägerfrequenzfernsprechen auf Leitungen. HirL, Leipzig 1948.
[H 20] Henne, W.: Rauschkenngrößen der Antennen, HF- und NF-Verstärker. Old, München 1972.
[H 21] Henne, W.: Empfänger-Elektronik. Hü, Heidelberg 1974.
[H 22] Hermann, H.: Zuverlässigkeitsverfahren für die Prozeßmeßtechnik. Old, München 1972.
[H 23] Herold, H.: Elektronische Verstärker in der Automatisierungstechnik. VT, Berlin 1972.
[H 24] Herpy, M.: Analoge Integrierte Schaltungen. Kiadó, Budapest 1976.
[H 25] Herrmann, G., Wagener, S.: Die Oxydkatode, Bd. 1 u. 2. Barth, Leipzig 1948/1950.
[H 26] Herter, E., u. a.: Nachrichtentechnik (2. Aufl.). Han, München 1981.
[H 27] Hertwig, H.: Induktivitäten. RFK, Berlin 1954.
[H 28] Hertzsch, A.: TTL-Vergleichsliste (2. Aufl.). MiV, Berlin 1980.
[H 29] Hertzsch, A.: Elektronische Zähler, Bd. 1 u. 2. MiV, Berlin 1981.
[H 30] Heywang, W., Pötzl, H. W.: Bänderstruktur und Stromtransport. SprH, Berlin u. a. 1976.
[H 31] Hilberg, W.: Elektronische digitale Speicher. Old, München 1975.

[H 32] Hilberg, W., Piloty, R.: Mikroprozessoren und ihre Anwendungen. Old, München 1977.
[H 33] Hilberg, W., Piloty, R.: Grundlagen digitaler Schaltungen. Old, München 1978.
[H 34] Hirschmann, W.: Schaltungen mit Halbleiterbauelementen, Bd. 5. NF/Hi-Fi, HF- u. Fernsehempfängerschaltungen. Siemens AG, München 1977.
[H 35] Hoffmann, R.: Rechenwerke und Mikroprogrammierung. Old, München 1977.
[H 36] Hoffmann, D.: Handbuch der Meßtechnik und Qualitätssicherung. VT, Berlin 1979.
[H 37] Höft, H.: Passive elektronische Bauelemente. VT, Berlin 1977.
[H 38] Höft, H.: Elektrische Kontakte: Werkstoffe, Einsatzbedingungen, Zuverlässigkeit. VT Berlin 1980.
[H 39] Hoftlinger, B.: Großintegration — Technologie, Entwurf, Systeme. Old, München 1978.
[H 40] Holzmüller, W.: Technische Physik, Bd. II/2. TeuL, Leipzig 1966.
[H 41] Hölzler, E., Holzwarth, H.: Pulstechnik, Bd. 1 u. 2. SprH, Berlin 1980, 2. Aufl.
[H 42] Hummitzsch, P.: Zuverlässigkeit von Systemen. VT, Berlin 1968.
[H 43] Hurst, S. L.: Schwellwertlogik. Hü, Heidelberg 1974.
[H 44] Hübner, D., Schönherr, E.: Diagnostik in der Digitaltechnik. VT, Berlin 1982.
[H 45] Hiller, H.: Operationsverstärker — Schaltungen und Anwendungen. VT, Berlin 1982.
[H 46] Hälsig, C.: Moderne mechanische Frequenzselektion. VT, Berlin 1970.
[I 1] Ilmer, H.-U.: Signalflußgraphen in der Elektrotechnik. VT, Berlin 1977.
[J 1] Jakubaschk, H.: MOSFET — Schaltungspraxis. MiV, Berlin 1976.
[J 2] Jakubaschk, H.: MOSFET — Schaltungstechnik — Schaltbeispiele mit Feldeffekttransistoren. MiV, Berlin 1976.
[J 3] Jakubaschk, H.: Komplimentär-Schaltungstechnik. MiV, Berlin 1977.
[J 4] Jakubaschk, H.: Das große Schaltkreis-Bastelbuch. MiV, Berlin 1978.
[J 5] Jakubaschk, H.: Elektronikschaltungen mit dem Operationsverstärker A 109. MiV, Berlin 1980.
[J 6] Jaworski, B. M., Detlaf, A. A.: Physik griffbereit. AV, Berlin 1972.
[J 7] Jemeljanow, G. A., Schwarzman, W. O.: Übertragung diskreter Signale. VT, Berlin 1978.
[J 8] Jentsch, H. u. a.: Elektrische Meßgeräte und Meßverfahren. SprH, Berlin u. a. 1978.
[J 9] Jessen, E.: Assoziative Speicherung, elektronische Datenverarbeitung, Beiheft 5. Vie, Braunschweig 1965.
[J 10] Jessen, E.: Architektur digitaler Rechenanlagen. SprH, Berlin u. a. 1975.
[J 11] Judmann, K. P., u. a.: Mikroprozessoren — Grundlagen und Anwendungen. Verl Erb, Wien 1978.
[J 12] Jugel, A.: Mikroprozessorsysteme. VT, Berlin 1978.
[J 13] Junge, H. D.: Brockhaus abc — Automatisierung. Brockhaus, Leipzig 1975.
[J 14] Jungnickel, H.: Stromversorgung elektronischer Geräte. VT, Berlin 1974.
[J 15] Jordan, W., Urban, H.: Strukturierte Programmierung. SprH, Berlin u. a. 1978.
[K 1] Kalantarow, P. L., Neumann, L. R.: Theoretische Grundlagen der Elektrotechnik. VT, Berlin
[K 2] Kammerloher, J.: Hochfrequenztechnik, Bd. 1. Elektromagnetische Schwingkreise. C. F. Wintersche Verlagsbuchhandlung, 1941.
[K 3] Kämmerer, W.: Einführung in die mathematischen Methoden der Kybernetik. AV, Berlin 1971.
[K 4] Kämmerer, W.: Digitale Automaten. AV, Berlin 1973.
[K 5] Kanton, D.: Mikroprozessorsysteme in der Automatisierungstechnik. VT, Berlin 1978.
[K 6] Kaufmann, H.: Datenspeicher. Old, München 1973.
[K 7] Khambata, H. J.: Einführung in die Gruppenintegration. VT, Berlin 1971.
[K 8] Kirenski, I. W.: Magnetismus. TeuL, Leipzig 1969.
[K 9] Klar, R.: Digitale Rechenautomaten. De Gruyter, Berlin 1976.
[K 10] Klein, G., Zaalberh van Zelst, J. J.: Präzisionselektronik. Philips Fachbücher, Hamburg 1972.
[K 11] Klein, R. D.: Mikrocomputersysteme (2. Aufl.). Fra, München 1979.
[K 12] Klein, R. D.: Mikrocomputer — Hard- und Software. Fra, München 1981.
[K 13] Klein, R. D.: BASIC-Interpreter. Fra, München 1982.
[K 14] Klein, W.: Grundlagen der Theorie der elektrischen Schaltungen. AV, Berlin 1961.
[K 15] Klein, W.: Finite Systemtheorie. TeuS, Stuttgart 1976.
[K 16] Kneller, E.: Ferromagnetismus. SprH, Berlin u. a. 1962.

[K 17] Kobitzsch, W.: Mikroprozessoren. Aufbau und Wirkungsweise, T. 1 u. 2. Old, München 1977/1980.
[K 18] Koepp, S., Fröhlich, H. J.: Grundlagen und Anwendungen der Oberflächen-Akustoelektronik, Bd. 8. Probleme der Festkörperelektronik. VT, Berlin 1976.
[K 19] Köstner, R., Möschwitzer, A.: Elektronische Schaltungstechnik; Analogschaltkreise. GP, Leipzig 1978.
[K 20] Kohl, G.: Schicht- und Schicht-Hybrid-Technik. Inst. f. Nachrichtentechnik, Berlin 1979.
[K 21] Kowalski, J. H.: Aktive RC-Filter. MiV, Berlin 1981.
[K 22] Kovács, F.: Hochfrequenzanwendungen von Halbleiterbauelementen. Akadémiai Kiadó, Budapest 1978.
[K 23] Kramer, K. D.: Elektronik-Praktikum. Bertelsmann, Düsseldorf 1973.
[K 24] Kramer, M.: Impulstechnik mit TTL-Schaltkreisen. MiV, Berlin 1980.
[K 25] Kremer, H.: Numerische Berechnung linearer Netzwerke und Systeme. SprH, Berlin u. a. 1978.
[K 26] Kress, D.: Theoretische Grundlagen der Signal- und Informationsübertragung. AV, Berlin 1977.
[K 27] Kress, D.: Theoretische Grundlagen der Übertragung digitaler Signale. AV, Berlin 1979.
[K 28] Kroschel, K.: Statistische Nachrichtentheorie. SprH, Berlin 1973.
[K 29] Krupicka, S.: Physik der Ferrite und der verwandten Oxide. Vie, Braunschweig 1973.
[K 30] Kühn, E., Schmied, H.: Handbuch Integrierte Schaltkreise. VT, Berlin 1978.
[K 31] Kühn, R.: Der Kleintransformator. C. F. Wintersche Verlagsbuchhandlung, Prien 1965.
[K 32] Kühne, H.: Schaltbeispiele mit digitalen p-Kanal-MOS-Schaltkreisen. MiV, Berlin 1978.
[K 33] Kühne, H.: Schaltbeispiele mit dem Operationsverstärker A 109. MiV, Berlin 1979.
[K 34] Kuhrt, F., Lippmann, H. J.: Hallgeneratoren — Eigenschaften und Anwendungen. SprH, Berlin 1968.
[K 35] Kulp, M.: Schaltungsprinzipien von Digitalanlagen. Universität Stuttgart, 1976.
[K 36] Küpfmüller, K.: Einführung in die theoretische Elektrotechnik (9. Aufl.), SprH, Berlin u. a. 1968.
[K 37] Kurz, G.: Analoge Schaltungen (2. Aufl.). MiV, Berlin 1980.
[K 38] Kussl, V.: Digitaltechnik III: Steuerungstechnik mit elektronischen Funktionselementen (2. Aufl.). VDI, Düsseldorf 1973.
[K 39] Kieser, H., Meder, M.: Mikroprozessortechnik — Aufbau und Anwendung des Mikroprozessorsystems U 880. VT, Berlin 1982.
[K 40] Kurz, G.: Grundlagen und Schaltungsbeispiele der Stromversorgung. MiV, Berlin 1982.
[K 41] Koslow, B. A., Uschakow, I. A.: Handbuch zur Berechnung der Zuverlässigkeit in Elektronik und Automatentechnik. AV, Berlin 1978.
[K 42] Kühnel, C.: Lineare und nichtlineare Analogschaltungen mit Operationsverstärkern. MiV, Berlin 1982.
[L 1] Lacroix, A., Witte, K. H.: Zeitdiskrete normierte Tiefpässe. Hü, Heidelberg 1980.
[L 2] Lange, F. H.: Korrelationselektronik. VT Berlin 1959.
[L 3] Lange, F. H.: Signale und Systeme, Bd. 1—3. VT, Berlin 1969/1970.
[L 4] Lange, F. H.: Methoden der Meßstochastik. AV, Berlin 1978.
[L 5] Lange, W. R.: Digital-Analog- und Analog-Digital-Wandlung. Old, München 1974.
[L 6] Lehmann, J.: Dioden und Transistoren (4. Aufl.). Vogel, Würzburg 1975.
[L 7] Lennartz, H., Träger, W.: Transistorschaltungstechnik. RFK, Berlin 1963.
[L 8] Lessen, M.: Feldeffekttransistoren — Eigenschaften und Anwendungen. MiV, Berlin 1974.
[L 9] Liebig, H.: Rechnerorganisation. SprH, Berlin 1976.
[L 10] Liebscher, S.: Fachzeichnen Informationstechnik (6. Aufl.). VT, Berlin 1978.
[L 11] Lindner, H.: Grundriß der Festkörperphysik. FV, Leipzig 1978.
[L 12] Lindner, H., Balcke, E.: Elektroaufgaben, Bd. 1—3. FV, Leipzig 1979.
[L 13] Luder, E.: Bau hybrider Mikroschaltungen. SprH, Berlin 1977.
[L 14] Lund, P.: Präzisionsvorlagen für Leiterplatten. Han, München 1979.
[L 15] Lunze, K., Wagner, E.: Einführung in die Elektrotechnik. VT, Berlin 1974.
[L 16] Lunze, K.: Theorie der Wechselstromschaltungen (3. Aufl.). VT, Berlin 1979.
[L 17] Lehmann, J. G.: Feldeffekttransistoren. Vogel, Würzburg o. J.
[L 18] Lesea, A.: Mikroprozessor-interface-techniken. Sybex, Düsseldorf 1981.

[L 19] Lancaster, D.: Das C-MOS-Kochbuch. IWT-Verlag, Vaterstetten 1980.
[L 20] Lampe, B., u. a.: Algorithmen der Mikrorechentechnik. VT, Berlin 1983.
[M 1] Mache, W.: Lexikon der Text- und Datenkommunikation. Old, München 1980.
[M 2] Madelung, O.: Festkörperprobleme VIII. AV, Berlin 1968.
[M 3] Madelung, O.: Grundlagen der Halbleiterphysik. SprH, Berlin u. a. 1970.
[M 4] Marko, H.: Methoden der Systemtheorie. SprH, Berlin 1977.
[M 5] Marsal, D.: Kleincomputer. Han, München 1972.
[M 6] Martin, H. J.: Die Ferroelektrika. GP, Leipzig 1964.
[M 7] Martin, W.: Mikrocomputer in der Prozeßdatenverarbeitung: Aufbau und Einsatz der Mikrocomputer zur Überwachung, Steuerung und Regelung. Han, München 1977.
[M 8] Meiling, W.: Mikroprozessor — Mikrorechner. AV, Berlin 1978.
[M 9] Meiling, W.: Digitalrechner in der elektronischen Meßtechnik, Bd. 1 u. 2. AV, Berlin 1978.
[M 10] Meinke, H., Grundlach, F. W.: Taschenbuch der Hochfrequenztechnik. SprH, Berlin u. a. 1956 u. 1962.
[M 11] Mennenga, H.: Schaltungstechnik mit Operationsverstärkern. VT, Berlin 1979.
[M 12] Metz, J., Merbeth, G.: Schaltalgebra. Grundlagen digitaler Schaltungen. FV, Leipzig 1970.
[M 13] Miel, G.: Elektronik des Modellantriebes, T. 2. Spannungsquellen. MiV, Berlin 1979.
[M 14] Mildenberger, D.: Analyse elektronischer Schaltkreise. HüPf, München 1973.
[M 15] Millner, R.: Katodenstrahloszillogrpahen. VT, Berlin 1967.
[M 16] Möschwitzer, A.: Wissensspeicher Halbleiterelektronik. VT, Berlin 1971.
[M 17] Möschwitzer, A.: Elektronische Halbleiterbauelemente (2. Aufl.). VT, Berlin 1975.
[M 18] Möschwitzer, A., Rumpf, K. H.: Einführung in die Elektronik. VT, Berlin 1978.
[M 19] Möschwitzer, A., Jorke, G.: Mikroelektronische Schaltkreise: Entwurf und Anwendung hochintegrierter Digitalschaltkreise und Mikroprozessoren. VT, Berlin 1979.
[M 20] Möschwitzer, A., Lunze, K.: Halbleiterelektronik (4. Aufl.). VT, Berlin 1979.
[M 21] Möschwitzer, A.: Integration elektronischer Schaltungen und Mikrorechner (3. Aufl.). VT, Berlin 1981.
[M 22] Mörder, C.: Transistorrechenpraxis, Hü, Heidelberg 1970.
[M 23] Müller, R.: Rauschen. SprH, Berlin u. a. 1979.
[M 24] Müller, R.: Grundlagen der Halbleiterelektronik (3. Aufl.). SprH, Berlin u. a. 1979.
[M 25] Müller, R.: Bauelemente der Halbleiterelektronik (2. Aufl.). SprH, Berlin u. a. 1979.
[M 26] Müller, W.: Elektronische Anzeigebauelemente. MiV, Berlin 1979.
[M 27] Münch, W. V.: Technologie der Galliumarsenid-Bauelemente. SprH, Berlin u. a. 1969.
[M 28] Müseler, H., Schneider, Th.: Elektronik — Bauelemente und Schaltungen. (1. Aufl.). Han, München, Wien 1975.
[M 29] Meyer-Eppler, W.: Grundlagen und Anwendungen der Informationstheorie (2. Aufl.). SprH, Berlin u. a. 1969.
[N 1] Naumann, G., u. a.: Standard-Interfaces der Meßtechnik. VT, Berlin 1980.
[N 2] Neuburger, E.: Einführung in die Theorie des linearen Optimalfilters (Kalman-Filter). Old, München 1972.
[N 3] Neufang, O.: Digitale Systeme, T. 2, Schaltwerke. Hü, Heidelberg 1979.
[N 4] Nüchtel, W.: Entwurf von Schaltwerken mit Mikroprozessoren. TeuS, Stuttgart 1979.
[N 5] Nüsslein, G.: Gleichspannungsverstärker. VT, Berlin 1952.
[N 6] Neumann, H., Schäfer, K.: Elektrische und elektronische Meßtechnik, AV, Berlin 1982
[O 1] Oberst, E.: Entwurf von Kombinationsschaltungen. VT, Berlin 1972.
[O 2] Ormont, B. F.: Einführung in die physikalische Chemie und Kristallchemie der Halbleiter. AV, Berlin 1979.
[O 3] Osborne, A.: Einführung in die Mikrocomputer-Technik. Grundlagenbuch der Mikrocomputer-Technik. te-wi Verl. GmbH, München 1977.
[O 4] Oefler, U., Classen, L.: Mikroprozessor-Betriebssysteme. VT, Berlin 1982.
[P 1] Pabst, D.: Operationsverstärker (3. Aufl.). VT, Berlin 1976.
[P 2] Pascoe, R. D.: Halbleiter-Schaltkreise in diskreter und integrierter Technik. Old, München 1978.
[P 3] Passynkow, W. W.: Nichtlineare Halbleiterwiderstände. GP, Leipzig 1965.
[P 4] Paul, M.: Schaltungsanalyse mit s-Parameter. Hü, Heidelberg 1977.

[P 5] Paul, R.: Transistoren. VT, Berlin 1964.
[P 6] Paul, R.: Transistoren — Meßtechnik. VT Berlin 1968.
[P 7] Paul, R.: Feldeffekttransistoren. Physikalische Grundlagen und Eigenschaften. VT, Berlin 1972.
[P 8] Paul, R.: Halbleiterphysik, VT, Berlin 1974.
[P 9] Paul, R.: Halbleiterdioden. VT, Berlin 1976.
[P 10] Paul, R.: Transistoren und Thyristoren. VT, Berlin 1977.
[P 11] Paul, R.: Halbleitersonderbauelemente. VT, Berlin 1981.
[P 12] Paul, R.: Mikroelektronik — Eine Übersicht. VT, Berlin 1981.
[P 13] Pauli, W.: Vierpoltheorie. AV, Berlin 1974.
[P 14] Peibst, H., u. a.: Spezialwerkstoff der Elektrotechnik. AV, Berlin 1964.
[P 15] Pelka, H.: Von der Schaltalgebra zum Mikroprozessor. Fra, München 1977.
[P 16] Peters, J.: Einschwingvorgänge, Gegenkopplung, Stabilität. SprH, Berlin u. a. 1954.
[P 17] Pfeifer, H.: Elektronik für Physiker, Bd. 1. Theorie linearer Bauelemente; Bd. 2. Die Elektronenröhre; Bd. 3. Schaltungen mit Elektronenröhren. AV, Berlin 1970.
[P 18] Pfüller, S.: Halbleitermeßtechnik. VT, Berlin 1977.
[P 19] Philippow, E.: Taschenbuch der Elektrotechnik, Bd. 1—6. VT, Berlin 1976—1982.
[P 20] Philippow, E.: Nichtlineare Elektronik (2. Aufl.). GP, Leipzig 1970.
[P 21] Pilz, G.: Basteln mit Thyristoren. MiV, Berlin 1981.
[P 22] Pitsch, H.: Hilfsbuch für die Funktechnik. GP, Leipzig 1953.
[P 23] Porst, A.: Bipolare Halbleiter. HüPf, München 1979.
[P 24] Preuss, H.: Zuverlässigkeit elektronischer Einrichtungen. VT, Berlin 1976.
[P 25] Profos, P.: Handbuch der industriellen Meßtechnik. Vulkan-Verl., Essen 1974.
[P 26] Palm, A.: Elektrische Meßgeräte und Meßeinrichtungen (3. Aufl.). SprH, Berlin u. a. 1948.
[P 27] Pflier, P. M.: u. a.: Elektrische Meßgeräte und Meßverfahren (4. Aufl.). SprH, Berlin u. a. 1978.
[P 28] Peters, J.: Einführung in die allgemeine Informationstheorie. SprH, Berlin u. a. 1967.
[R 1] Racho, R. u. a.: Werkstoffe der Elektrotechnik. VT, Berlin 1968.
[R 2] Rauchfuss, S., Sack, R.: Elektrostatische Aufladungen — erkennen, vermeiden, beseitigen. Tribüne, Berlin 1976.
[R 3] Recknagel, A.: Physik — Elektrizität und Magnetismus (6. Aufl.). VT, Berlin 1970.
[R 4] Rein, H. M., Ranft, R.: Integrierte Bipolarschaltungen. SprH, Berlin u. a. 1980.
[R 5] Reinboth, H.: Technologie und Anwendung magnetischer Werkstoffe (3. Aufl.). VT Berlin 1975.
[R 6] Reiner, R.: Stabilisierung von Gleichspannungen und Gleichströmen. Old, München 1972.
[R 7] Reiss, K., u. a.: Integrierte Digitalbausteine. Siemens AG, Berlin, München 1970.
[R 8] Rhein, D.: Grundlagen der digitalen Speichertechnik. GP, Leipzig 1974.
[R 9] Richard, B.: Datenverarbeitung mit Mikroprozessoren. T. 1: Hardware. Mikroprozessor, Mikrorechner, Mikrorechnersystem. Han, München 1980.
[R 10] Richards, R. K.: Elektronische Bauelemente und Schaltungen. AV, Berlin 1972.
[R 11] Richter, H.: Praxis der integrierten Schaltungen. Telekosmos-Verl., Stuttgart 1970.
[R 12] Rint, C.: Handbuch für Hochfrequenz- und Elektrotechniker, Bd. 1—5 (neue Aufl.). HüPf München 1978—1981.
[R 13] Rohrbach, C.: Handbuch für elektrisches Messen mechanischer Größen. VDI, Düsseldorf 1967.
[R 14] Rost, R.: Silicium als Halbleiter. BU, Stuttgart 1966.
[R 15] Roth, M.: Mikroprozessoren (2. Aufl.). Technische Hochschule Ilmenau, 1978.
[R 16] Rothe, H., Kleen, W.: Grundlagen und Kennlinien der Elektronenröhren. GP, Leipzig 1948.
[R 17] Rothe, H., Kleen, W.: Elektronenröhren als Anfangsstufenverstärker. GP, Leipzig 1948.
[R 18] Rothe, H., Kleen, W.: Elektronenröhren als End- und Sendeverstärker. GP, Leipzig 1948.
[R 19] Rothe, H., Kleen, W.: Elektronenröhren als Schwingungserzeuger und Gleichrichter. GP, Leipzig 1948.

[R 20] Ruge, I.: Halbleitertechnologie. SprH, Berlin 1975.
[R 21] Rühl, H.: Matrizen und Determinanten in elektronischen Schaltungen. Hü, Heidelberg 1977.
[R 22] Rühl, H., Nguyen, H. H.: Praktischer Entwurf von Wellenparameterfiltern. Hü, Heidelberg 1977.
[R 23] Rumpf, K. H., Pulver, H.: Anwendung von Halbleiterbauelementen im Schaltbetrieb. VT, Berlin 1970.
[R 24] Rumpf, K. H.: Bauelemente der Elektronik (9. Aufl.) VT, Berlin 1978.
[R 25] Rupprecht, W.: Netzwerksynthese. SprH, Berlin u. a. 1972.
[S 1] Saal, R.: Handbuch zum Filterentwurf. Hü, Heidelberg 1979.
[S 2] Sahner, G.: Digitale Meßverfahren (2. Aufl.). VT, Berlin 1979.
[S 3] Salow, H., u. a.: Der Transistor, physikalische und technische Grundlagen. SprH, Berlin u. a. 1963.
[S 4] Schade, K.: Halbleitertechnologie. VT, Berlin 1981.
[S 5] Schaefer, E.: Magnettechnik — kurz und bündig. Vogel, Würzburg 1969.
[S 6] Schapaschnikow, A. A.: Elektronen- und Ionenröhren. VT, Berlin 1961.
[S 7] Schatt, W.: Einführung in die Werkstoffwissenschaft. DVG, Leipzig 1981.
[S 8] Schauer, H., Barta, G.: Methoden der Programmerstellung für Tisch- und Taschenrechner. SprW, Wien 1979.
[S 9] Schecher, H.: Funktioneller Aufbau digitaler Rechenanlagen. SprH, Berlin u. a. 1973.
[S 10] Schiller, D.: NF-Verstärker-Praxis. MiV, Berlin 1981.
[S 11] Schiller, S., Heisig, U.: Bedampfungstechnik. VT, Berlin 1975.
[S 12] Schlenzig, K., Blasing, K. H.: Schaltbeispiele mit dem Rechnerschaltkreis U 820 D/U 821 D. MiV, Berlin 1980.
[S 13] Schleuder, G.: Periphere Geräte in der Datenverarbeitung. Han, München 1972.
[S 14] Schmellenmeier, H.: Technologie elektronischer Bauelemente. VT, Berlin 1970.
[S 15] Schmid, D., u. a.: Technische Informatik, T. 1. Old, München 1973.
[S 16] Schmid, H.: Elektronische Dezimalrechner. Old, München 1978.
[S 17] Schmidt, K. H.: Entwurf digitaler Systeme. Eine praktische Einführung (Lehrbuchreihe Elektronik). BU, Stuttgart 1979.
[S 18] Schmidt, V.: Digitalelektronisches Praktikum. TeuS, Stuttgart 1973.
[S 19] Schmidt, V., u. a.: Digitalschaltungen mit Mikroprozessoren. TeuS, Stuttgart 1978.
[S 20] Schneider, H. G., Hässner, A.: Werkstoffe und Technologie elektronischer Bauelemente. DVG, Leipzig 1974.
[S 21] Schneider, H. G., u. a.: Hybridtechnik. DVG, Leipzig 1977.
[S 22] Schneider, H. G., u. a.: Halbleiterbauelementeelektronik. AV, Berlin 1977.
[S 23] Schnell, G.: Magnete. K. Thiemig-Verl., München 1973.
[S 24] Schnell, G., Hoyer, K.: Mikrocomputerfibel. Vie, Braunschweig 1982.
[S 25] Schoenfeld, R.: Hybrid-Simulation. AV, Berlin 1979.
[S 26] Schröder, H., Rommel, G.: Elektrische Nachrichtentechnik: Eigenschaften und Darstellung von Signalen, Bd. 1a. HüPf, München 1978.
[S 27] Schrenk, H.: Bipolare Transistoren. SprH, Berlin 1978.
[S 28] Schubert J.: Gleichspannungsverstärker. VT, Berlin 1971.
[S 29] Schubert, K. H.: Das große Radiobastelbuch (5. Aufl.) MiV, Berlin 1980.
[S 30] Schüssler, H. W.: Digitale Systeme zur Signalverarbeitung. SprH, Berlin u. a. 1973.
[S 31] Schultz, J.: Elektrische Meßtechnik (2. Aufl.). VT, Berlin 1979.
[S 32] Schummy, H.: Digitale Datenverarbeitung für das technische Studium. Vie, Braunschweig 1975.
[S 33] Schukky, H.: Signalübertragung. Lehrbuch der Nachrichtentechnik und Datenverarbeitung. Vie, Braunschweig 1978.
[S 34] Schunk, H., Engel, W.: Grundlagen der Impulstechnik. HüPf, München 1979.
[S 35] Schuon, E., Wolf, H.: Nachrichten-Meßtechnik. SprH, Berlin 1981.
[S 36] Schuppe, W., Leistner, D.: Microrechnerprogrammierung. Inst. f. Nachrichtentechnik, Berlin 1978.
[S 37] Schwarz, H.: Einführung in die moderne Systemtheorie. Vie, Braunschweig 1969.
[S 38] Schwarz, W., u. a.: Mikrorechner — Wirkungsweise, Programmierung und Applikation. VT, Berlin 1980.

[S 39] Schwertfeger, H. J.: Grundlagen der elektrischen digitalen Schaltungstechnik. Transpress, Berlin 1971.
[S 40] Seeger, A.: Theorie der Gitterfehler. Handbuch der Physik, Bd. 7, Teil 1. SprH, Berlin u. a. 1955.
[S 41] Seifart, M.: Analoge Schaltungen und Schaltkreise. VT, Berlin 1980.
[S 42] Seifert, W.: Digitaltechnik II: Schaltelemente (2. Aufl.). VDI, Düsseldorf 1973.
[S 43] Seitzer, D.: Arbeitsspeicher für Digitalrechner. SprH, Berlin u. a. 1975.
[S 44] Seitzer, D.: Elektronische Analog-Digital-Umsetzer. SprH, Berlin u. a. 1977.
[S 45] Seliger, N. B.: Kodierung und Datenübertragung. VT, Berlin 1973.
[S 46] Shah, A., u. a.: Integrierte Schaltungen in digitalen Systemen. Bir, Basel 1977.
[S 47] Shea, R. F.: Transistortechnik. VT, Berlin 1971.
[S 48] Simonyi, K.: Theoretische Elektrotechnik. DVW, Berlin 1970.
[S 49] Sonin, A. S., Strukow, B. A.: Einführung in die Ferroelektrizität. AV, Berlin 1974.
[S 50] Speiser, A. P.: Digitale Rechenanlagen. SprH, Berlin 1961.
[S 51] Spenke, E.: pn-Übergänge. SprH, Berlin u. a. 1979.
[S 52] Stamm, H.: Elektronische Meßverfahren. VT, Berlin 1969.
[S 53] Stanek, J.: Technik elektrischer Meßgeräte. VT, Berlin 1961.
[S 54] Steinbuch, K., Rupprecht, W.: Nachrichtentechnik. SprH, Berlin 1967.
[S 55] Steinbuch, K.: Taschenbuch der Nachrichtenverarbeitung (2. Aufl.). SprH, Berlin u. a. 1967.
[S 56] Stiefken, H.: Analoge Schaltkreise: Aufbau — Methoden — Anwendungen. Hü, Heidelberg 1972.
[S 57] Streng, K.: ABC von Elektronenröhre und Halbleiterbauelement. MiV, Berlin 1972.
[S 58] Streng, K.: abc der Stromversorgungstechnik. MiV Berlin 1972.
[S 59] Streng, K.: Transistordaten. MiV, Berlin 1975.
[S 60] Streng, K.: Analoge integrierte Schaltungen von TESLA. MiV, Berlin 1976.
[S 61] Streng, K.: Diodendaten. MiV, Berlin 1977.
[S 62] Stürz, H., Cimander, W.: Automaten. VT, Berlin 1972.
[S 63] Swoboda, R.: Thyristoren. Francksche Verlagsbuchhandlung, Stuttgart 1968.
[S 64] Swoboda, J.: Codierung zur Fehlererkennung. Old, München, Wien 1973.
[S 65] Schmidt, W., Feustel, O.: Optoelektronik kurz und bündig. Vogel, Würzburg 1975.
[S 66] Schüssler, H. W.: Netzwerke, Signale und Systeme, Bd. 1 u. 2. SprH, Berlin 1981/82.
[S 67] Seifart, M.: Digitale Schaltungen und Schaltkreise. VT, Berlin 1982.
[S 68] Streng, K. U.: Daten linearer Schaltkreise. MiV, Berlin 1982.
[S 69] Sarkowski, H.: Dimensionierung von Halbleiterschaltungen. Lexica-Verlag, Grafenau-Döffingen 1973.
[S 70] Stuhlmüller, P.: 16-bit-Generation Z 8000, Aufbau und Anwendung. te-wi Verl. GmbH, München 1980.
[S 71] Shannon, C. E., Weaver, W.: Mathematische Grundlagen der Informationstheorie. Old, München, Wien 1976.
[T 1] Teichmann, H.: Angewandte Elektronik, Bd. 1 u. 2. Bir, Basel 1975.
[T 2] Thiel, R.: Elektrisches Messen nichtelektrischer Größen. TeuS, Stuttgart 1977.
[T 3] Tholl, H.: Bauelemente der Halbleiterelektronik, Bd. 1 u. 2. TeuS, Stuttgart 1976/1978.
[T 4] Tietze, U., Schenk, C.: Halbleiterschaltungstechnik (3. Aufl.). SprH, Berlin 1976.
[T 5] Tränkler, H. K.: Die Technik des digitalen Messens. Old, München 1976.
[T 6] Tröndle, K. H., Weiss, R.: Einführung in die Puls-Code-Modulation. Old, München 1974.
[T 7] Trzeba, E.: Grundlagen der Schaltungstechnik. HirL, Leipzig 1974.
[U 1] Ulrich, D.: Grundlagen der Digitalelektronik und digitalen Rechentechnik. Fra, München 1975.
[U 2] Unbehauen, R.: Synthese elektrischer Netzwerke. Old, München 1972.
[U 3] Unbehauen, R.: Systemtheorie (3. Aufl.). AV, Berlin 1980.
[U 4] Unger, H. G., Harth, W.: Hochfrequenz-Halbleiterelektronik. HirS, Stuttgart 1972.
[U 5] Unger, H. G.: Optische Nachrichtentechnik. Elitera-Verl. Berlin(West) 1976.
[U 6] Unger, H. G., Schultz, W.: Elektronische Bauelemente und Netzwerke II. Vie, Braunschweig 1969.
[V 1] Vahldiek, H. J.: Aktive RC-Filter. Old, München 1972.

[V 2] VAHLDIEK, H. J.: Übertragungsfunktionen. Old, München 1973.
[V 3] VAHLDIEK, H. J.: Elektronische Signalverarbeitung. Old, München 1977.
[V 4] VASKE, P., u. a.: Programmierbare Taschenrechner in der Elektrotechnik. TeuS, Stuttgart 1981.
[V 5] VIEH, R.: Z-Transformationen. VT, Berlin 1964.
[V 6] VILBIG, F.: Hochfrequenzmeßtechnik. Han, München 1953.
[V 7] VILBIG, F.: Lehrbuch der Hochfrequenztechnik. GP, Leipzig 1960.
[V 8] VOLLMEYER, A.: Fertigung in der Elektronik. VDE-Verlag, Berlin(West) 1973.
[V 9] VÖLZ, H.: Elektronische Spannungsstabilisation. VT, Berlin 1965.
[V 10] VÖLZ, H.: Grundlagen der magnetischen Signalspeicherung, Bd. 1—3. AV, 1968, 1970, 1972.
[V 11] VOORHOEVE, N. A. J.: Niederfrequenzverstärkertechnik. Philips Techn. Bibliothek, Eindhoven 1952.
[V 12] VÖLZ, H.: Information I und II. AV, Berlin 1982.
[W 1] WAGNER, S. W.: Stromversorgung elektronischer Schaltungen und Geräte. Decker-Verl., Hamburg 1964.
[W 2] WALDSCHMIDT, K.: Schaltungen der Datenverarbeitung. TeuS, Stuttgart 1980.
[W 3] WALLOT, J.: Größengleichungen, Einheiten und Dimensionen. Barth, Leipzig 1947.
[W 4] WEBER, W.: Einführung in die Methoden der Digitaltechnik (5. Aufl.). Elitera-Verl., Berlin(West) 1977.
[W 5] WEHRMANN, W., u. a.: Korrelationstechnik — ein neuer Zweig der Betriebsmeßtechnik Lexika-Verlag, Grafenau 1977.
[W 6] WEIDEL, J.: Halbleiter erobern die Technik. Hü, Heidelberg 1966.
[W 7] WEIL, G.: Digitale integrierte Schaltkreise. VDE-Verlag, Berlin(West) 1977.
[W 8] WELLER, W.: Anwendung der Mikroelektronik in der Prozeßautomatisierung. VT, Berlin 1979.
[W 9] WELLER, W., WILKE, H.: Programmierbare Steuereinrichtungen. VT, Berlin 1981.
[W 10] WENDT, S.: Entwurf komplexer Schaltwerke. SprH, Berlin 1974.
[W 11] WIESENER, K., u. a.: Elektrochemische Stromquellen. AV, Berlin 1981.
[W 12] WINSTEL, G., WEYRICH, C.: Optoelektronik, Bd. 1. SprH, Berlin 1980.
[W 13] WINKLER, H.: Elektronische Analogieanlagen. AV, Berlin 1963.
[W 14] WIRK, H., THILO, H. G.: Niederfrequenz- und Mittelfrequenz-Meßtechnik. HirS, Stuttgart 1956.
[W 15] WOLF, G.: Digitale Elektronik. Fra, München 1969.
[W 16] WOLF, H.: Thyristoren in der Automatisierungstechnik (2. Aufl.). VT, Berlin 1973.
[W 17] WOLF, H.: Lineare Systeme und Netzwerke. SprH, Berlin u. a. 1978.
[W 18] WOSCHNI, E. G.: Meßdynamik. HirL, Leipzig 1954.
[W 19] WOSCHNI, E. G.: Meßgrößenverarbeitung. HirL, Leipzig 1969.
[W 20] WOSCHNI, E. G.: Information und Automatisierung. VT, Berlin 1970.
[W 21] WOSCHNI, E. G.: Informationstechnik, Signal, System, Information. VT, Berlin 1973.
[W 22] WOSCHNI, E. G., KRAUS, M.: Informationstechnik. Arbeitsbuch. VT, Berlin 1976.
[W 23] WUNSCH, G.: Moderne Systemtheorie. GP, Leipzig 1962.
[W 24] WUNSCH, G.: Systemtheorie der Informationstechnik. GP, Leipzig 1971.
[W 25] WUNSCH, G.: Systemtheorie. GP, Leipzig 1975.
[W 25] WEISS, H., HORNINGER, K.: Integrierte MOS-Schaltungen. SprH, Berlin u. a. 1977.
[W 26] WÜSTENHUBE, J., u. a.: Schaltnetzteile: Grundlagen, Entwurf, Schaltungsbeispiele (2. Aufl.). Expert-Verlag, Grafenau/Württ., VDE-Verlag, Berlin(West) 1982.
[Z 1] ZACH, F.: Leistungselektronik. SprW, Wien 1979.
[Z 2] ZAKS, R.: CP/M Handbuch mit MP/M. Sybex Verl., Düsseldorf 1981.
[Z 3] ZANDER, H. J.: Entwurf von Folgeschaltungen. VT, Berlin 1974.
[Z 4] ZIRPEL, M.: Operationsverstärker. Fra, München 1977.
[Z 5] ZUIDERVEEN, E. A.: Handbuch der digitalen Schaltungen. Fra, Stuttgart 1981.
[Z 6] ZUSE, K.: Rechnender Raum. Vie, Braunschweig 1969.
[Z 7] ZUSE, K.: Der Komputer — mein Lebenswerk. Verl. moderne Industrie, München 1970.
[Z 8] ZEMANEK, H.: Elementare Informationstheorie. Old, München, Wien 1959.

10.2. Abkürzungsverzeichnis

Enthält nur Begriffsbezüge
Vollständiger Verweis im Sachwortverzeichnis
≠ in Bezug auf Programmierung
* verweisen auf Signalbezeichnungen

A RDY* 870
A STB* 870
A-Betrieb 534
A-Typ 168
A0 ... A15* 843
AB-Betrieb 539
AC 805
ADC 941
ADC≠ 851
ADD≠ 833, 851
ADU 941
ADV 707
AEP 627
AI 753
AIM 751
AKF 627
AL-WERT 81
ALGOL 923
ALS 697
ALU 719
AM 483
AND 670, 673, 675
AND≠ 834, 851
ANF 679
AS 697
ASBC 277
ASCII 667
AT-Schnitt 421
ATMOS 753
ATN* 895, 897

B RDY* 870
B STB* 870
b≠ 850
B-Betrieb 535
B/A* 870
B = binär 660
BAS 882
BASIC 923
BBD 640
BCA 969
BCD 661
BDI 277f.
BDOS 921
BFZ 830, 842
BH 901
BI 713
BIFET 569
BIOS 921
BIT≠ 834, 849
BORAM 758
BP 124
BS 124
BUSAK* 826, 845, 865
BUSREQ* 826, 843, 863

C≠ 846, 850, 855
c≠ 850
C* 779, 782
C-Endstufe 551
C/A* 870
C4D 640
CAD 206
CALL≠ 835f., 851, 913
CAM 761
CAMAC 888, 900ff.
CAS* 744
CC 900
cc≠ 848
CCD 347, 640
CCD-Filter 651
CCD-Speicher 758
CCF≠ 851
CCITT 667, 902, 905
CCP 920
CDI 277
CE* 870
CID 353
CLK* 877
CMC7 323
CML 698
CMOS 706
CMP≠ 851
CMRR 505
COBOL 923
CP≠ 833, 851
CP/M 920
CPD≠ 851
CPDR≠ 851
CPI≠ 851
CPIR≠ 851
CPL≠ 834, 851
CPU 808
CRC 873, 885
CRT 883
CS* 740
CSFL≠ 834
CT-Schnitt 421
CTC 867
CTD 347, 640
CTS* 875, 902
CY≠ 846

D 660
d≠ 850
D* 779
D-FF 779, 781
d-RAM 729ff.
D-Verstärker 551
D0 ... D7* 843
DAA≠ 849
DAC 941
DAC* 807, 896
DAF 680
DAP 356
DAU 941
DAV* 807, 894, 896

dB 614
DCD* 875, 902
dd# 850
DEAP 752
DEC 893
DEC# 833, 851
DEE 903
DEF# 918
DEMUX 714
DH 344
DI# 851, 865
DIAC 443
DIFMOS 753
DIL-Gehäuse 213
DIN 902
DIO* 807
DJNZ# 849
DLD 553
DMA 826, 835, 867
DMOS 260, 704
DMOS-FET 259
DMS 453
DNF 679, 720
DNL 626
DO 740
DREKO 82
DRO 743
DSM 355
DSR* 902
DT-Schnitt 421
DTR* 875, 902
DÜE 903
DV-FF 784

e# 850
e* 780
E-Verstärker 553
E13B 323
E²CL 699
E²PROM 749
EAROM 749
ECB 356
ECL 698
EDV 808
EE-Kerne 85
EEROM 749
EFG 709
EFL 699
EFSI 276, 709
EI# 853, 865
EI-Schnitt 84
EIA 902
EL 362
ELKO 76
ELSE# 911
EMP 992
EMV 990
EOI* 895, 897
EPROM 749
EQR 905
EQU# 918
ESL 73
ESONE 900
ESR 73
ETC-Filter 645
Ex 753
EX 937
EX# 832, 853
EXX# 842, 853

f# 850
FAMOS 753
FAST 697
FE 890
FEFET 162
FET 162ff.
FF 779ff.
FET 58, 936
FG 753
FL 751
FLAD 360
FM 483, 883
FOR-DO# 911
FOR-TO-NEXT# 911
FORTRAN 922
FPL 719
FPLA 719
FPLS 719
FPRP 719

G-H 356
G-Verstärker 553
GATO 438
GATT 438
GLOBAL# 918
GND* 902
GOSUB# 913
GTO 441

H-TTL 697
H = HEX 660
H = High 685
HALT* 843, 851
HDK 52, 74
HGÜ 433
HMOS 704
HP 124, 893
HP-IL-Bus 906

I-T-Bar 760
I-Y-Bar 760
I²L 700
I³L 704
IC 509
IDO 937
IDW 428
IEC-Bus 686, 893
IEEE 893
IEEE-488 893
IEEE-696 892
IEI* 870
IEO* 870
IF-THEN# 911
IFC* 895
IFF# 865
IFL 719
IGFET 163
ii# 850
IL 906
IM# 853, 865
IMS-1 891
IMS-2 893
IN# 834, 853
INAK* 826
INC# 833, 853
IND# 853
INDR# 853
INI# 853
INIR# 853
INT* 826, 845, 865
IOREQ* 825, 845
IPOS 276
IRQ* 826
ISO 667, 902

J* 779
JEDEC 384, 708
JK-FF 781, 783
JK-MS-FF 783
JP# 836, 853
JPZ# 836
JR# 836, 853
JRZ# 837
JUMP# 835

K* 779
k* 780
KAF 679
KDF 679
KI 753
KKF 628, 680
KNF 679, 720

L-TTL 697
L = Low 685
LARAM 758
LASER 341
LCD 355
LD# 831, 853
LDD# 853
LDDR# 853

LDI# 853
LDIR# 853
LED 335
LEF 362
LID 967
LIFO 764
LOCMOS 709
LOCOS 278
LS-TTL 697
LSB 661
LSI 723
LSS-Theorie 199
LT 713

M# 846, 850, 855
m# 850
m-Halbglied 126f.
M-Schnitt 84
M1* 826, 845
MC 797
MECL 699
MESFET 163
MFM 883
MISFET 163
MKL 75
MKU 75
MNOST 162
MOSFET 163
MPR 805
MR 805
MREa* 825, 843
MS-FF 779, 781
MSB 661, 715
MSI 723
MSSW 429
MTBF 979
MTL 700
MTTF 979
MUX 883

n# 850
n-Kanal 168
N-Pfad-Filter 645
NAND 670, 673, 675
NC# 846, 850, 856
NDAC* 895, 897
NDK 52, 74
NDRO 743
NEG# 833, 853
NEMP 992
NEP 326
NEXT# 911
NMI* 826, 845, 865
NMOS 756
nn# 850
NOP# 853
NOR 670, 673, 675
NOT 670
Np 614
npn-Transistor 174
NRFD* 895
NRZ 883
NTC 223
NTSC 400
NVRAM 749, 757
NZ# 846, 850, 855

O = oktal 660
OCR 323
OE 740
OF 937
OFW 428, 646
ON-GOTO# 911
OR 670, 673, 675
OR# 834, 853
ORG# 918
OTDR# 853
OTIR# 853
OUT# 834, 853
OUTD# 853
OUTI# 853
OV 554
OXIM 276f.
OXIS 277

P# 846, 850, 855
p# 850
P-C 356
p-Kanal 168
P/V# 846
P2CCD 349
PAL 400
PAM 487
PAP 910
PASCAL 924
PC 840
PCCD 349, 640
PCM 489
PDM 487
PDV-Bus 904
PE# 846, 850, 855
PFM 489
pin-Diode 229
PIO 867
PIPO 796
PISO 796
PL/1 923
PLA 719, 934
PLANOX 276
PLL 629
PLM 487
PLZT 364
PM 483
PN-Plan 118ff.
pn-Übergang 149ff.
pnp-Transistor 174
PO# 846, 850, 855
POP# 832, 853
pp# 850
PPM 487
PROM 749
psn-Struktur 159, 229
PTC 223
PUSH# 832, 853
PVDF 419
PVF 419
PVF2 419
PVM 487
PZM 487
PZT 405, 416

Q = oktal 660
qq# 850

r# 850
R* 779
r* 780
R.C* 875
R.D* 875
RAM 749
RAS* 744
RBI* 713
RBO* 713
RC-DP 586
RC-Glieder 109ff.
RC-Verstärker 251
RCVR* 902
RD* 825, 845
RECL 699
REM 814
REN* 895
REPEAT-UNTIL# 911
RES# 834, 854
RESET* 826, 843
RET# 835, 854
RETI# 854
RETN# 854
RETURN# 835, 913
RFSH* 826, 845
RGB 400
RIO 921
RLA# 833
RLT 440
RM-Kerne 85
ROM 749
rr# 850
RRA# 833
rRES# 863
rRL# 863
rRR# 863
RS-232 902, 904

RS-422 904
RS-449 904
RS-FF 780
rSET# 863
rSL# 863
rSR# 863
RST# 836, 854, 8666
RT 809
RTS* 902
RWM 749

S# 844
s# 848
s* 780
S-100-Bus 892
s-RAM 729ff.
S-TTL 697
SAM 757
SAMOS 753, 755
SBC# 854
SBS 440
SC 900
SC-Differenzintegrator 651
SC-Filter 645, 651
SC-Integrator 651
SC-Schnitt 423
SCCD 640
SCF# 852
SCF* 902
SCH 344
SCT 640
SDLC 875, 885
SDS 906
SECAM 400
SET# 834, 854
SEV 386
SFET 163
SH 344
SI 1.2 891
SI-K 890
SI-L 890
SI-S 890
SI2.2 899
SIAL 899
SIF-1000 667, 892
SIMOS 753, 755
SIO 867
SIPMOS 260
SIPO 796
SISO 796
SL# 854
SLA# 833
SOS 704
SPS 758
SQR* 895
SR# 854
SRA# 833
SRV 557
ss# 850
SSI 723
SSM 483
ST 906
St-TTL 697
STEP# 911
STL 704
SUB# 832, 854
SYNC* 875

t# 850
T* 779, 782
T-FF 779, 781
T.C* 875
T.D* 875
T³L 695
T = tenär 660
TC-66 893
TCC-Schnitt 423
TEST# 834
TFT 162, 967
TK 73
TN-FEN 355
TN-LCD 357
TP 124
TRIAC 440
TRIMASK 277f.
TSN 339
tt# 850
TTL 688
TTY 904

UART 843
UI-Schnitt 84
UMOS-FET 259
UNIX 921
USART 865
USV 435
UV 753

V* 779
V-24 902
V-Typ 168
VATE 276, 278
VCO 520, 629
VDR 224
VIL 704
VIP 276ff.
VLSI 723
VMOS 704, 740
VMOS-FET 259f.

WAIT* 826, 843
WE 740
WHILE-DO# 911
WITH-DO# 911
WR* 824, 825, 845

XMIT* 902
XOR 670, 673, 675
XOR# 834, 855

Z# 844, 850, 855
Z* 779
Z-Diode 157
z-Transformation 649
ZIF 937

10.3 Verzeichnis der Bauelemente

2102 739
2816 755
2817 757
2920 955
2 N 3585 300
358 CM 565
4004 822, 828, 886
4417 790
4116 744
5101 738
6502 822, 839f., 886
7400 692
74138 710
74181 718
74251 710
7440 692
7442 710
7446 712
7447 712
7448 712
7449 712
7472 785
7474 784
7481 733f.
7483 716
7492 792f.
7493 786
7495 798
8008 828, 840, 841, 886
8080 822, 826, 828, 839f. 886
8085 840
8086 822, 840f.
8087 880
8212 795
8231 879f.
A 109 555f.
A 210 D 547f.
A 210 K 547f.
A 211 D 545f.
A 244 D 513
A 301 D 513
AHY 10 452
B 080 D 569
B 084 D 569
B 176 D 568f.
B 177 D 568f.
B 222 D 579
B 260 D 609
B 2761 D 568f.
B 340 D 501, 508
B 611 D 568f.
B 621 D 568f.
B 631 D 568f.
B 761 D 568f.
B 861 D 568f.
BFX 82 503
CMOS 708
CTC 875f.
D 120 690
D 172 785
D 174 784
D 181 733f.
D 192 792f.
D 193 786
D 195 798
DMA 877f.
HCS-4 886
iAPX 32 822, 828
IL 72723 598
K 140 MA 1 579
K 140 УД 12 568
K 140 УД 7 568
K 155 Iπ 4 718
K 580 IK 80 826
L 123 598
LF 356 565
LM 3900 508
M 6800 822, 840
MA 78 XX 600
MAA 723 598
MAA 741 568
MAA 748 568
MDA 2010 549
MDA 2020 549f.
MH 3212 795
PIO 868f.
SC 206 263, 266, 267, 268, 270
SC 207 503
SF 126 309
SF 137 274, 289
SF 215 502
SIO 872f.
SM 104 502, 622
SMY 51 248
TA 72741 568
TA 72748 568
TAA 2761 568f.
TAA 761 568f.
TAA 861 568f.
TCA 311 568f.
TCA 321 568f.
TCA 331 568f.
TDA 1060 608
TL 81 ... 84 569
TMS 9900 842
TRA 72400 598
TRV 1723 598
U 256 D 744
U 855 868f.
U 856 872f.
U 857 875f.
U 8810 843
U 8820 843
μA 702 555
μA 709 555
μA 714 565
μA 723 598
μA 741 565f.
μA 748 565f.
μA 776 567f.
μA 777 565f.
Z 8 841
Z 80 828, 839ff.
Z 8000 822, 828, 840, 886

10.4. Sachwortverzeichnis

Es bedeuten:

* Tabelle
≠ Bild
f. + die folgende Seite
ff. + mehrere folgende Seiten

7-bit-Kode 667
7-K-Regel 986
7-Segmentdekoder 711f.
A-Betrieb 534ff.
AB-Betrieb 539
Abakus 821
Abbruchbedingung 912
Abgleichen 615
Ablaufsteuerung 806
Ablenkung von Elektronen 381ff.
Abschlußglied 127
Absorptionskoeffizienten 301*
Absorptionskonstante 325
Abstandsbelichtung 205
Abtast-Halte-Schaltung 651
Abtastfilter 645, 647ff., 953≠
Abtastglied 516
Abtastsysteme 647f., 648≠
Abtasttheorem 1005
Abtastverfahren 627
Adder 715, 716≠
Addierstufe 573, 803≠
Additionstafel 716≠
Adreßraum 824
adress-activation 742
Adressen 726
Adressenbus 808, 824f.
Adressierung, Systematik 838≠
Adreßteil 812, 830
Advanced TTL 697≠
AIKEN-Kode 663≠
Akkumulation 627
Akkumulator 32ff., 804
aktive Filter 586ff.
Akustoelektronik 404ff.
AL-Wert 81
Algebra, abstrakte 671
—, BOOLEsche 672ff.
Algorithmus 818
alkalische Zelle 28f., 31≠
Allpaß, aktive Filter 590≠
Allpaß-System 119≠
ALU 719ff.
ALU-Befehle 832ff., 856*, 857*
Aluminium-Elko 77f.
amorphes Material 46
Amplitudenmodulation 483
Amplitudenstatistik 528
Amplitudenstufe 1003
analog 654
analog—digital, Vergleich 654ff., 659*
Analog-Digital-Wandler 944ff., 952≠
Analogie, Wärme—elektr. Strom 303f.
Analogien 49
Analogrechentechnik 554
Analogschaltungen 489ff.
Analyseverfahren 623≠
Anfangspermeabilität 42
Anisotropie 47
Anlaufstrom 380
Anordnungskode 663
Anpassung, Leistung 15, 16≠
—, optimales Filter 127
Anreicherungstyp, FET 167f.
Antinomie 819
Antivalenz 670≠, 673≠, 675≠
Anwenderprogramm 917
Anwenderprogrammierung 748
Anzeige, elektrochemisch 363ff.
—, magnetisch 365
—, Messung 610
—, Rechner 881
—, Vergleich 365ff., 366≠, 367*
Anzeigen 322ff.
Aperturfehler 120
Äquipotentiallinien 7≠, 49, 50≠

Äquivalenz 673#
Arbeitsbereich, sicherer 299, 300#, 526f., 527#, 591#, 592
Arbeitskennlinie, FET 251#
—, komplex 255#, 298#
Arbeitspunkt, FET 250f., 251#
—, gleitender 533#
—, Transistor 269f., 270#
Arbeitsspeicher 727
Arbeitssteilheit 251
Arbeitswiderstand, induktiv überbrückter 532
—, optimaler 528ff.
Archivspeicher 727
Arithmetikprozessoren 879f.
Armbanduhr, Quarz 425f.
Array-Prozessor 938
ARRHENIUSsches Gesetz 986
ASCII-Kode 667, 668*
Assembler 917
Assoziativspeicher 761ff.
Ätzen 191f.
Aufdampfanlage 197#
Aufdampfen 971
Auffangregistor 794
Auffrischen 743
Aufgaben, kombinatorische 818
Augenempfindlichkeit 315f.#*
Ausfallrate 977f., 988#*
—, Elko 78
Ausfallurschen 983#
Ausgangsalphabet 771
Ausgangslastfaktor 689
Ausgangsstufen 526ff.
Ausgangstransformator 93
Ausprägungsgrad 610
Ausschlagmethode 614
Aussagenlogik 669
Aussteuerung, Endstufe 530f.
Austauschbefehle 832
Austauschintegral 36, 37#
Austrittsarbeit 374*
Autokorrelation 59, 627, 628#
Automat, Entwurfsprinzipien 775f.
—, Typen 772#
Automatentafel 766
Automatentheorie 768, 771ff.
automatic-power-down 742
Automatisierung 810
Avalanche-Effekt 156f., 436, 751
Avalanche-Transistor 296
Averaging 627

B-Betrieb 535ff.
back-up 818
Backward-Diode 457#
Band, verbotenes 138
band-gap-reference 602
Bandabstandsreferenz 602
Bandbreite, FM 485f.
—, relative 103
Bändermodell 138, 139#
Bandkante 325#
Bandkern 84
Bandpaß 124
Bandsperre 124
bank-select 839
BARKHAUSEN-Beziehung 249, 463
Barrierenhöhen 161#
Basis-Schaltung 262#
Basisweite 179f.#
Batterieverbrauch 27
Bauelement, Definition 66
—, diskretes 975
—, elektronisch erzeugt 475ff.
—, Hybridtechnik 965ff.
—, Kosten 215#
—, Vollständigkeit 480
Baumrechner 939
BCD-Addition 915
BCD-Zahl 661
BCD-Zähler 792f., 793#
beam-lead 967
Bedingungsschleifen 912ff.
Befehl, Kurzzeichen 850*
—, Struktur 829
—, Tabellen 857ff.
—, Typen 829f.#
—, Übersicht 852#
Befehlsregister 812
Befehlssatz Z 80 849ff.
Befehlszähler 812, 826f., 830, 840
Begrenzer 524f., 582f.
Belastung von Quellen 14
Beleuchtungsstärke 317f., 318#
BENCHMARK-Test 995
Benutzerfreundlichkeit 995
Berechenbarkeit 819
BESSEL-Filter 130
Betriebs-Matrix 20
Betriebsfaktoren 984, 986*
Betriebsparametertheorie 129ff.
Betriebswerte, Transistor 263f.
Beweglichkeit 147, 149#
Bezeichnungsaufwand 660f.
Bezugspegel 614
Bildaufnahmeröhre 388
Bildröhre 396ff., 397#
Bildschirm, flacher 359
Bildschirmcontroller 868, 883
Bildverarbeitung 810
Bildverstärker 387
Bildwandler 387#
binär 656f.
Binär-Zahl 657

binäre Logik 669
binary digit 657
Binder 919
biphase 883f.
Bipolartransistor 174ff., 262ff.
biquadratischer Ausdruck 118
Bitbefehl 834
Bitbreite 825
Bitkosten 729#
Bitleitung 737
Bitorganisation 731
Blindkopplung 463
Blindlast, Schalten 299#
Blindzelle 744
Block 210
BODE-Diagramm 99
BOHRsches Magneton 36
BOLTZMANN-Verteilung 142, 152, 377#
Bondtechnik 214#
BOOLEsche Algebra 672ff.
— Darstellungsformen 674ff.
— Funktionen 672ff., 673#, 676#
— Funktionen, Minimierung 681ff.
— Grundfunktionen 675f.
— Rechenregeln 677
Bootstrap-Lader 920
Bootstrap-Prinzip 477, 493, 503, 542, 571
Branch-highway 901
Breakpoint 919
Breitbandphasenschieber 115#, 589f.
Breitbandverstärker 256f.#
Brückengegentaktschaltung 535
Brückenmethode 615
Brückenschaltung 493
Brummspannung 236
Bubbles 759f., 760#
bucket brigade device 640
Buffer 685, 794
Bulkspannung 169
bump 967
burn in 989
burst 665
Buskoppler 906
Bussysteme 807f.
Bussysteme, Vergleich 907#
BUTTERWORTH-Filter 129
Byte 665

C-Betrieb 551
CMOS-Schaltungen 707f.#
Cache 840
CAMAC 900ff.
CAUER-Filter 129, 133
CCD-Bauelement 173, 347ff.
CCD-Filter 651, 652#
CCD-Matrix 350f., 351#
CCD-Zeile 350f., 351#
Cermet 969
Cermet-Widerstand 23
Channeling 200#
charge coupled device 640
charge transfer devices 640
chemische Reaktion bei Batterien 27
Chip-Prozesse 212ff.
Chirp-Signal 431
cholesterisch 355#
Chopperbetrieb, Oszillograph 637
Chopperverstärker 584
Chromatographie 623#
CID-Bild-Matrix 352#
CLAPP-Oszillator 518#
Clipper 582
COLPITZ-Oszillator 518#
Compiler 921
Control-Bus 825
Controller 895f.
COOPER-Paar 458f., 459#
CORBINO-Scheibe 451#
Cordic-Algorithmen 881
Corona-Entladung 403
crate 900
Cross-Assembler 924
Cross-Compiler 924
Cross-reference 919
CRT-Controller 883
CTC 875f.
CuO_2-Diode 232
current mode logic 698
current switshing mode 297
current-loop-interface 904
Cursor 920
CZOCHRALSKI-Verfahren 189#
D-Verstärker 551ff.
Dachschräge 71
Daisy-chain 865
Dämpfung, überaperiodische 99f.
Dämpfungsfaktor 116
DARLINGTON-Transistor 498ff., 500#, 543
Daten 810
Datenbus 808, 824f.
Datenleitung 737
Datenselektor 714
Datenstruktur 912
Datenübertragung 806ff.
Datenwähler 714
Dauermagnetkreis 43#
debugging 919
Default-Werte 920
Dehnungsmeßstreifen 453ff., 454#
Dekoder 709f., 719#, 731
Dekrement 719
DELON-Schaltung 238
Deltamodulation 489
Demodulation 241

DE-MORGANsches Gesetz 678
Demultiplexer 714, 731
Depletion-FET 168
derating-curve 23#
Detektivität 327
Dezibel 614
DIAC 614
Diagnoseprogramm 920
Diamanttyp 139
Dickschichttechnik 962*, 968ff.
Dickschichtwiderstand 21ff.
Dielektrizitätskonstante 49
Differentiator 573#
Differenz-Verstärker 504ff.
Differenzierglied 107#, 478
Differenzmethode 615
Diffusion 194f.
Diffusionskonstanten 186*
Diffusionsspannung 151
Diffusionsstrom 151
digital 656
Digital-Analog-Wandler 940ff.
Digital-Oszillograph 637
Digitalfilter 952ff., 953#, 954#
Digitron 390ff., 391#
DIL-Gehäuse 213#
Dioden 228ff.
—, Anwendung 228*
—, Aufbau 230#
—, Rauschen 284f.
—, Schalter 242f.
—, Sperrbetrieb 154f.
Diodenbrücke 516, 586
Diodenlogik 683
Diodenmatrix 684#
Dipol, magnetischer 36
DIRAC-Impuls 59
Disassembler 917
Disjunktion 670#, 673#, 675#
diskret 655
Display 322ff.
Division 804
Divisionsschaltung 578
DMA 877f.
DMOS-FET 259
Dolby 626
don't care 761
Doppel-Gate-MOSFET 261#
Doppelbasisdiode 443
Doppelbrechung, LCD 356
Doppelrechner 818
Dosenelektrometer 412
Dotierung 144
Dotierung, Neutronen 436
Dotierung, Vergleich 201f.
Dotierungsniveaus 146#
Drahtbonden 214#
Dreheiseninstrument 412f., #
Drehfeld 760
Drehkondensator 82#
Drehprozesse 40
Drehspulinstrument 410f.#
Drehstromnetz 55
Dreieck-Stern-Umwandlung 11*
Dreieckschwingung 53
Dreiecksgenerator 521#
Dreiexzeßkode 663#
Dreileiter-handschake 895#f.
Dreiphasen-CCD 348#
Dreipolschaltungen 491#
Dreipunkt-Oszillator 518#
Dreko 82#
drop-out 665
Drossel, vormagnetisiert 84
Drosselanfangsschaltung 235#, 236
DRUDE-Theorie 4f., 137
dual 656f.
dual-slope 948#
Dual-Zähler 790
Dualitätsprinzip 10, 15, 69
Dummy-Zelle 744, 746
Dunkelraum 401
Dünnfilmtransistor 967ff.
Dünnschichttechnik 962f., 971ff.
Durchbruch, zweiter 299f.#
Durchbruchverhalten, Transistor 296#
Durchflußwandler 605f.
Durchflutung 36, 42
Durchführungskondensator 80, 82#
Durchgriff, FET 171
Durchgriff, Rückkopplung 465
Durchlaßbereich 107, 125
Durchlaßerholzeit 158
Durchlaßrichtung, Diode 153f.
Durchmesser, Si-Kristalle 189#
dynamische Logik 669
Dynamoblech 84
Dynamometer 411

EARLY-Spannung 268
EBERS-MOLL-Modell 176, 288
Echtzeituhr 875
ECL-Technik 698ff.
Editor 917
EE-Kern 85
Effektivwert 65
Effektivwertmessung 581
Effizienz 908
EI-Schnitt 84
Eichen 615
Eigenkonzentration 144
Eigenleistung 140, 144
Eimerkettenspeicher 641ff.
Einbrennen, Leuchtschirm 386

Einchip-Prozessor 842
Eindringtiefe, Ionen 199*
Einerkompliment 666
Eingangsalphabet 771
Einheiten, lichttechnische 316*
Einheiten, magnetische 44
Einheitssprung 59
Einkristalle 185
Einkristallherstellung 187ff.
Einsatzfrequenz 272
Einschaltvorgang 67f., 100#
einschrittige Kodes 663
Einseitenbandmodulation 483
EINSTEIN-Gerade 376#
Einstellregler 25
Einsverstärker 587
Einzelmessung 612
Eisenquerschnitt 93
Elektret 52
Elektrik 1f.
Elektrolyt, Batterien 27
Elektrometerverstärker 570#f.
Elektromigration 199, 290#
Elektronen, freie 4, 373f.
Elektronengas 4f.
Elektronenoptik 383#
Elektronenstrahllithografie 206ff., 207#
Elektronik 1f.
—, industrielle 434
Elektrostatik 49ff.
Elementarsignale 59
Elementarterm 679
Elko 76ff., 79#
Ellipsenauswertung 631#
elliptische Filter 129
Emission, Elektronen 373
emitter coupled logic 698
Emitterschaltung 262#
Emulation 924
enable 782
Endstufe 526ff.
—, integriert 545ff.
—, kurzschußstromsicher 545
—, quasikomplimentär 543f.
Energie 2
Energiedichte bei Batterien 27
Energiedichte magnetischer Materialien 43#
Energieprodukt, magnetisches 42
Energiequellen 26ff.#
Energieterme 137
Enhencement-FET 168
Entblößungs-FET 168
Entmagnetisierung 44
Entprellen 721
Entropie 998ff.
Entscheidung 909
Entschlüßler 940
Entwicklungssystem 917
Entwurf, optimierter 932
Entwurfsmethoden, Vergleich 933#, 934*
Epitaxie 193ff.
Epitaxie-Planar-Diode 230f.#
Epitaxie-Planar-Technik 208ff.
Epitaxie-Planar-Transistor 210#
EPROM 753f., 754#
Erdnung 291f.
Ergebnisfunktion 771
Ermüdung 324#
Ersatzquellen 14, 16#
Ersatzschaltbild, Elko 77#
—, FET 172f., 173#, 249f.#
—, Kondensator 73
—, Spannungsquelle 13#
—, Stromquelle 13#
—, Transformator 89ff.#
—, Transistor 269ff., 273#
—, Wärmeleitung 303#
ETC-Filter 645
Eventoskope 637
Exponentialfunktion 68
exponentielle Verstärker 575f.#

face down 966#, 967
Fallbügelschreiber 413, 414#
fan-out 686f., 689
Fangbereich 668#
Farbart 319
Farbbandbreite 400
Farbbildröhre 397ff.
Farbcode, Widerstände 25*
Farbdreieck 321#
Farbfilter-CCD 352#
Farbmetrik 319
Farbton 319
Farbwahrnehmung 319
FAST 697#
feed-back 461
feed-forward-error-correction 554#
Fehlerbetrachtungen 611
fehlererkennender Kode 664
Fehlerkontrolle 192#
fehlerkorrigierende Kodes 664f.
Fehlerrate 978
Fehlersicherung 885
Fehlerverteilung 977f., 987ff.
Fehlstellen 187
Feld, elektrostatisches 49
—, entmagnetisierendes 44
Feldenergie, magnetische 45
Feldplatte 450
Feldrechner 938
Feldstärke, elektrische 8#
—, magnetische 36, 39
Feldstrom 151

FERMI-Energie 142
FERMI-Verteilung 141 ff.
Fernschreibkode 905*
Fernsehbildröhre 397#
Fernsehen, industrielles 637
Fernsehinterface 882
Fernwirken 616
Ferrite 46
Ferritkern, Materialien 85*
Ferritkernspulen 85 ff.
Ferroelektrizität 416
Ferromagnetismus 36, 37#
Festspannungsregler 601
Festwertspeicher 749 f.
FET 161 ff.
—, Daten 247
—, Überblick 162#
—, Widerstand 220#
Filmtechniken 963
Filter, aktive 586 ff.
—, Beispiele 130 f.#
—, mechanische 426 f., 427#
—, nichtklassische 644 ff.
—, Oberflächenwellen 428 ff.
—, piezoelektrisch 427
—, rekursive 646
—, Theorie 123 ff.
—, Theorie, klassische 630
—, Vergleich 653 f.
Filterung von Nutzsignalen 625
Fingerstruktur 428
Firmware 816
Flächenwiderstand 218
Flachrelais 408
Flachzelle 28 f., 29#
Flag, Z 80 846
flat-pack 213#, 967
Flimmergrenze 324#
Flip-chip 966#, 967
Flipflop 522#, 799 ff.
—, taktflankengesteuert 783
—, Typen 781#
floating gate, Speicher 753
Floppy-disc 638#, 639, 822, 884
Fluoreszenz 335
fluoreszenzaktiviertes Display 360 f., 361#
Fluoreszenzanzeigeröhre 390 ff.
Fluß, magnetischer 39
Flußänderung 67
Flußbilddarstellung 910
Flußdichte, magnetische 39
Flußlinien, magnetische 42
Flußquant 460
Flüssigkeit, magnetische 47
Flüssigkristall 353 ff., 356*
Foldback-Kennlinie 597
Folgeschaltungen 669
Folienkondensatoren 75 f., #
Folienspeicher 638#, 639, 822, 884
Folientasten 455 f.#
Formfaktor 66
Formierung, Elko 76
Formstrahl 207#
Fotodiode 331 ff.
Fotoduodiode 334#
Fotoeffekte, Einteilung 327
Fotoelemente 329 ff.
Fotoemission 375 ff.
Fotoempfänger 325 f.
Fotokatoden 376#, 377*
Fotolack 202
Fotolithografie 202 ff.
Fotorepeater 204#
Fotoresist 202
Fotothyristor 441 f., 442#
Fototransistor 333#
Fotowiderstände 327 f., 328*, 329#
Fotozelle 386 f.
FOURIER-Analyse 56 ff.
frame 907
Freilaufdiode 606
Freiwerdezeit 439
Fremdatome 140
Frequenz, komplexe 54, 101, 116#
—, negative 116
—, normierte 98, 125
Frequenzbereichsfilterung 625 f.
Frequenzgang, RC-Verstärker 253#
—, Transformator 92
Frequenzkorrektur, OV 559 ff., 561#
Frequenzmodulation 483 ff., 485#
Frequenzteiler 522, 785 ff.
Frequenzverdopplung 415
Frequenzvervielfachung 63
Frühausfall 978, 989 ff.
Fundamentalterm 679
Fünfschichtdiode 440
Funkel-Rauschen 282
Funkenlöschung 241
Funktion und Struktur 928
Funktionsgenerator 576 f., 584 f.
Funktoren 669, 670#
fusable link 751

GaAs-Transistor 276
Gabellichtschranke 347#
Gap 460
Gasentladungsröhre 401 ff.#
Gatestrom 437
Gatter 687
—/Chip 729#
Gatterlaufzeit 724#
Gauß 39
GAUSS-Ebene 54, 116, 650#

GAUSS-Filter 129
Ge-Transistor 275, 527
Gebereffekte 448#
Gegeninduktivität 67, 88f.
Gegenkopplung 463
—, Grenze 472ff., 473#
Gegentaktendstufe 534ff.
Gegentaktmodulator 244#
Gegentaktschaltungen, Verzerrungen 539
Gegentaktwandler 607
Gehäuseeinbau 213#
Generation 144
Generatorpolynom 885
Germanium als Material 183
Geschichte, Batterien 35*
—, Bauelementeanzahl u. -dichte 959#
—, Dioden 229*
—, EMP 992
—, Hybridtechnologien 958*
—, Laser 344#
—, Leistungselektronik 433
—, Leistungsgrenzen Transistor 311#
—, Mechano-, Akusto-Elektronik 405f.*
—, Mikroprozessoren 886f.
—, Opto-, Vakuumelektronik
—, Rechentechnik 820ff.
—, Relais 406#
—, Röhren 371#
—, Sensoren 447
—, Speicher 724#, 728#, 729#
—, Tieftemperaturelektronik 447*
—, Transistor 276*
—, Verbindungstechniken 958*
—, Zuverlässigkeit 979*
Geschwindigkeit von Elektronen 381
Geschwindigkeit-Leistungs-Produkt 697ff., 703#
Gewichtsfunktion 648
GIACOLETTO-Schaltung 271f., 273#
Glasbildner 46
Gläser, metallische 46
gleichgewichtiger Kode 663
Gleichrichter, Anwendungsgebiete 237#
—, phasenempfindlicher 628
Gleichrichterschaltungen 235ff.#
Gleichrichtertechnik 434
Gleichrichtung 63, 234
—, ideale 580
Gleichspannungswandler 608
Gleichstromverstärker, Vergleich 585*
Gleichtaktverstärkung 506
Gleitkommastandard 879
Gleitkommazahlen 667
Glimmlicht 401
GLIXAN-Kode 663#
globale Daten 913
GOEDEL-Theorem 658
Golddrähte 214
GRAETZ-Schaltung 238, 433
Graph, Automat 766
—, markierter 770
GRAY-Kode 663#
GREINACHER-Schaltung 238#
Grenzdaten, Transistoren 279f.
Grenzenergie 1007
Grenzfall, aperiodischer 99f.
Grenzfrequenz 104
—, Ferrit 87
Grenzgebiet 152
Grenzhysterese 40
Grenzverstärkung 250
Grenzzyklen 955
GROSCHsches Gesetz 817
Größen, lichttechnische 314f., 316*
—, magnetische 44
Grundeinheit 613
Grundschaltungen, IC
Grundsoftware 920ff.
Gruppenlaufzeit 120
guard board 709
Gummi, magnetischer 47
GUNN-Element 456
Güte, aktiver Verstärker 589
—, FET 254f.
—, Quarz 422
—, *RC*-Glied 109, 111#
—, Schwingkreis 98, 109
—, Spule 82
Gyrator 478ff.

Halbadder 717
Halbglied 124
Halbleiter, Bauelement 138ff.
—, Laser 342
—, Material, Daten 138ff., 147ff.##
—, vergiftete 146
—, Widerstand 217ff., 218#, 219*
HALL-Effekt 449ff.#
HALL-Konstante 450#
Haltebereich 628#
Haltepunkt 919
Haltestrom 436
HAMMING-Abstand 665#
handshake 806ff., 895f.
Hardware 815ff.
—, minimale 816
Härteskala 610
HARTLEY-Oszillator 518f.
hartmagnetische Materialien 46
Hazard 722
—, essentieller 776
heat-pipe 307f.
Heißleiter 220ff.
Hex-Zahl 658, 660

HF-Bandfilter 136#
HF-Transistor 275
High-Com 626
HILBERT-Transformation 120
Hochpaß 106#, 124
Hochspannungstransistor 279#
Hochspannungsübertragung 433
Hochvolttechnik bei MOS 704
Hohlraumstrahler 318
Hostcomputer 924
HP-IL-Bus 906f.
Hub, relativer 485
HURWITZ-Polynom 118
Hybrid-Matrix 18*, 181
Hysteresekurve 40#, 632#
Hystereseverluste 82

I^2L-Technik 700ff.
IBM-Format 884
IC, Grundschaltungen
Identität 673#
IEC-Bus 893ff.
Ikonoskop 388f.
Imetanzfunktion 117
Impedanz, Kondensator 73#
Impedanzconverter 478
Impedanzwandlung 493
Impfkeim 189
Implikation 670#, 673#
Implosionsgefahr 397
Impulsmodulation 486ff.
Impulsschaltung 521ff., 77ff.
Impulsschlag 992ff.
Impulsverdichtung 431, 630#
Impulsverhalten, Thyristor 438
Impulswärmewiderstand 307f., 309#
Impulszeit 523f.
In-out-Befehle 834, 857*
Index-Register 839f.
Indexröhre 399#
Induktion, magnetische 39
Induktionsgesetz 67
Induktivität 67
—, integrierte 227#
—, Meßbrücke 112#
—, negative 479
—, reale 105#
—, technische 80ff.
Influenz 49
Information 2, 885f.
Informationselektronik 2#
Informationstheorie 997ff.
Infrarotdioden 339
Inhibition 673#
Injektor 702
Injektorstromstärke 703
Inkrement 719
Innenboardschneiden 190#
Innenwiderstand, FET 171
—, negativer 466f.#
—, Null 465f.
—, Null, Verstärker 504#
—, Quelle 13
Instrumentationsrecorder 639
Integrationsgrad 723*, 724#
Integrationstechniken, Vergleich 725*
Integrator 573#
Integrierglied 107#, 478
integrierte Injektionslogik 700ff.
Interdigitalwandler 428
Interface-Schaltkreise 867ff.
Interferometer 343
Intermodulationsfaktor 62
Interpreter 921
Interrupt 835, 863ff.
Interruptadresse 866#
Interruptflipflop 865#
Interruptkette 847
Interruptlogik 869
Intrinsiczahl 144
Inversionsdichte 144
Inverter 685
Ionenimplantation 199f., 200#, 211#
Ionisationsenergien 146#
Ionisierungsspannung 442*
Isolationswiderstand 72
Isolator 141#
Isolierschichten 196f.
Isolierstoffe 51, 51*
Isoliertechnologien 277ff.

Ja-nein-Entscheidung 819
JEDEC-B-Standard 708
JOHNSON-Rauschen 282
JOHNSON-Zähler 789#
JOSEPHSON-Element 460#
JOULEscher Effekt 414
Justieren 615

Kalibrieren 615
KALMAN-BUCY-Filter 644
Kaltleiter 223f.#
Kaltwalzen 47
Kanalkapazität 1005f.
Kanalmodell 997#
Kanalstopper 709
kanonische Realisierung 122f.
Kapazität 49, 66
—, Diode 156#, 226#
—, integrierte 225#, 227*
—, Meßbrücke 112#
—, negative 479
—, spezifische 457
—, technische 72

KARNAUGH-Diagramm 674
Kaskode-Verstärker 501 ff.#
Kbyte 666
Kellerspeicher 764#
Kennlinie, Diode 232*, 234#
—, FET 165#, 247#
—, Heißleiter 222#
—, LED 340#
—, nichtlineare 63
—, Transistor 180 f., 181#, 266#, 268#, 274#
—, Z-Dioden 157#
Kennlinienfeld, Rückkopplung 466 f.#
Kennlinienschar 632
Keramik, ferroelektrische 364 f., 365#
keramische Kondensatoren 73 f.
— Materialien 52
Kernblech, normiertes 83#
Kernmagnetinstrument 411
Kernspeicher 821
Kernwiderstand, Vierpol 20
Ketten-Matrix 18*
Kettenbruch 123
Ketteninterface 890
Kettenschaltung 123
Kettenverstärker 257#
Kippgerät 634 f.
KIRCHHOFFsche Sätze 9, 17 f.
Klärpunkt 355#
Kleinsignalparameter, Transistor 181 ff.
Kleinsignalverhalten, FET 171 ff.
—, Transistor 264 f.
Klemmenspannung 13
Klemmschaltung 584
Klimazonen 982 f., 983*
Klirrfaktor 62, 71
Klirrfaktormessung 113
Klirrkoeffizient 98
Knickkennlinie 98
Knopfzellen 31 f., 32#
Knotenregel 18
Kodeeigenschaften 662
Koder 709 f., 710#
Kodierungen 661 ff., 883
Koerzitivfeldstärke 40, 46
Kohärenzgrad 335
Kohlemassewiderstand 23
Kohleschichtwiderstand 21 f.
Kollektor-Schaltung 266#
Kombinationsfrequenzen 242 f.
kombinatorische Schaltungen 668 ff.
Kommentarfeld 918
Kommunikation 907
Komparator 521 ff., 522#, 524 f., 582 f., 715
Kompensation 477
Kompensationschreiber 413, 414#
Kompensationsmethode 615
Kondensatoren, dämpfungsarm 75
—, Daten 81#
—, technische 72
Konflikt 771
Konjunktion 670#, 673#, 675#
Kontaktprellen 721 f.
Kontaktschaltung 674
kontinuierlich 655
Konvektion 301
Koordinatenschreiber 413
Kopplungsfaktor 89, 417
Korngrenze 188#
Korrektur-Verstärker 554
Kraftlinien 50#
Kraftwirkung, magnetische 45
Kreis, magnetischer 42, 80
Kreisdiagramm 102
Kreisfrequenz 54
Kreisstrom 35
Kreuzkorrelation 627, 628#
Kristallanisotropie 38, 39*
Kristallaufbau 139
Kristallfehler 188#
Kühlblech 305 f.
Kundenwunschschaltkreise 930 ff.
Kunststoffe 52
Kunststoffverkapselung 216
Kupferoxidul-Gleichrichter 433
KÜPFMÜLLER-Beziehung 59, 1005
Kurveneinfluß bei Messungen 65#
Kurzschlußwiderstand 124
Kybernetik 821

Lackfolienkondensatoren 75
Ladebefehle 831 f., 855
Ladeendspannung 32, 34*
Ladekondensator 234
Lader 919
Ladung, freie 49
Ladungsbilder 50#
Ladungskopplung 347 ff.
Ladungssteuerung 170
Ladungsträger 4*
Ladungsträgerverteilung 143#
Ladungstransport 349 f.
Ladungstrennung 49
ladungsverschiebende Schaltungen 177, 639 ff
Längsregler 592, 594
Langzeitkonstanz 73
Läppen 190 ff.
Laser 341 ff., 342#, 344#, 345#
latch 732
latch-up 709
Laufparameter 814
Laufzeitfilter 130, 133
Lawinenbetrieb, Transistor 296
Lawinenfotodiode 333#

LCD 355ff., 357#, 359#
Leak-proof-Zellen 28
Lebensdauer von Trägern 158
Lebenszyklus, Kondensatoren 80f.
LECLANCHÉ-Element 31#
LED-Anzeigen 344#
Leerlaufwiderstand 124
Legierungsdiode 230f.#
Legierungstransistor 279#
Leistung, umsetzbare 528
Leistungsanpassung 15, 16#
Leistungsbauelemente, Vergleich 445#
Leistungsbilanz, Endstufe 530f.
Leistungselektronik 2#, 432ff., 434ff.
—, Schaltungen 444ff.
Leistungshyperbel 295, 527
Leistungstransistor 275
Leistungsverstärker 526ff.
Leistungsverstärkung 265f.
Leiter 141#
Leiterplatte 958, 960#, 961ff.
Leitfähigkeit 147
Leitungsband 138
Leitungsmechanismen 147f.
Leitwert 9
Leitwert-Matrix 18*
LENZsche Regel 67
Lesestelle 770
Leuchtdichte 317f., 318#
Leuchtdiode 334ff.
Leuchtkondensator 362#
Leuchtstoff 383ff., 385#
Lichtbogen 403
Lichtleitfasern 366ff., 368#, 369*, 370*
Lichtschranke 346f., 347#
Linearfaktoren 118
Linearität, System 117
Linearitätsfehler 62
Linieninterface 890, 893f.
Linker 919
LISSAJOUS-Figuren 633
Listener 894
Lithium-Zelle 31
Loch, Halbleiter 140
Lochkarte 821
Lochmaskenröhre 397f., 398#
Lock-in-Schaltung 245
logarithmische Verstärker 575f.#
Logik, innere und äußere 687f.
—, positive und negative 685f., 686#
Logikbefehle 834
logische Schaltungen 669
Logons 59
lokale Daten 913
look-ahead-carry 717
Low-power-SCHOTTKY-TTL 991*
LSS-Theorie 199
Luftsauerstoffzellen 28, 31
Luftspalt 44
—, Spule 84
Luftspulen 87, 88#
Lumineszensdiode 334ff., 337#, 338#, 341*

m-Halbglieder 127
M-Schnitt 84
Macro 918
Magnet-Dioden 452#
Magnetband 47
Magnetbandgerät, Steuerung 773ff.
Magnetbandspeicherung 637
Magnetfeldglühen 47
Magnetisierung 38f.
—, spontane 38
Magnetisierungsstrukturen 38#
Magnetmaterialien 48*
magnetoresistiver Effekt 451f.
Magnetostatik 35ff.
Magnetostriktion 414f., 415#
Majoritätsträger 141
Mark 873
Markenfeld 918
MARKOFF-Kette 1004
Maschenregel 17
Maschinenzyklen 847
Maßeinheit 610
Maskenprogrammierung 748, 750#, 930f.
Maskentechnik 204ff.
maskierbare Kodes 663
maskieren 761
Maßsystem 613
Master-Slave-Flipflop 783f.
matched Filter 644
Materiewelle 137
Matrixorganisation 731
Maxterm 679, 720
Maxwell 39
MAXWELL-WIEN-Brücke 112#
MEALY-Automat 767
Mechanoelektronik 404ff.
MEDWEDEV-Automat 773
Mehrprozessorsystem 938
Mehrschichtkeramikmodule 962
Mehrwegefilter 645
MEISSNER-Oszillator 518#
Memristor 480#
Mengentheorie 671
Mesa-Diode 230#
Mesa-Transistor 279f.
Meßbrücke, Abgleich, komplex 114f.
Meßgenauigkeit 658
Meßinstrumente 410ff.
Meßkette 611f., 612#
Meßmethoden, allgemein 625#
Mesophasen 354f.

Meßprinzipien, der Elektronik 624#
Meßreihe 612
Meßschaltungen 616
—, komplexe 664ff.
Messung, Wechselgrößen 63f.
Messungen 610ff.
Meßtechnik, automatisierte 810
Meßverstärker 622
Meßwertaufnehmer 446
Meßwertspeicherung 613, 639
Meßzeitverlängerung 626
Metall-Halbleiter-Kontakt 160f.
Metall-Papier-Kondensatoren 75
Metalloide 46
Metalloxidwiderstand 21ff.
Metallschichten 198f.
Metallschichtwiderstand 21ff.
mikro-OHMsches Gesetz 6
Mikrocomputer 808
Mikroelektronik 1, 808
Mikromodultechnik 962
Mikroprogrammierung 820, 926
Mikroprozessor 808, 824ff.
Mikrorechentechnik 808ff.
Mikrosonde 623#
MILLER-Effekt 254, 574
MILLER-Integrator 477
MILLER-Kapazität 477
MILLER-Kode 823f.
MILLERsche Indizes 139
Miniaturtechnik 961
Minimal-Phasen-System 119#, 120
Minimalsoftware 917
Minoritätsträger 141
Minterm 679, 720
Mitkopplung 463
Mittelpunktsschaltung, Gleichrichter 239#
Mittelungsfilter 133
Mittelwerte bei Messungen 63ff.
Mittenspannung 542f.
Mnemonics 916ff.
MÖBIUS-Zähler 790
Modem 903
Modulation 481ff., 482#
Modulationsindex 485
Modulationsverstärker 583
Module 900
Modulo-n-Zähler 787#
Modultechnik 962
Momentanwert 54
Momente, magnetische 37#
Monitor 917
Monoflop 522ff.
MOORE-Automat 767
MOS-Inverter 705#
MOS-Technologien 258ff., 704ff.
MOSFET 165ff.
Multavi II 621
Multichiptechnik 963, 975
Multiplexer 515#, 714
Multiplikation 450, 804, 944
Multiplikatoren 576f.
Multivibratoren 519ff.#
Multiwire 961
Multizellularinstrument 412
Musikleistung 537
Mustergenerator 883

n-aus-m-Kode 664
N-Pfad-Filter 645
Nachbeschleunigung 395
Nachleuchtdauer 384
Nachrichtentechnik 2#
Nachtsichtgerät 388
Nachwirkungsverluste 83
Nacktchip 965
Nacktwickel 75
Negation 673#, 6753
Negativlack 202
Negator 684
nematisch 684
nematisch 355#
Nennspannung, Elko 78
Neper 614
Netzstörungen 294
Netztransformator 93
Neukurve, magnetische 40
NEUMANN-Rechner 811
Neutronendotierung 201f.
Nibble 732
NiCd-Akkumulator 32
nichtlineare Schaltungen mit OV 574ff.
Niederfrequenztransistor 275
NIXIE-Röhre 404
Nonius 610
Normalformen 679ff.
Normierung, Filter 124
Normreihe von Widerständen 22*
NORTHON-Verstärker 508f.#
Notation, polnische 671
Nullstellen 101
Numerik 809
NYQUIST-Rauschen 282

Oberflächenschichten 193ff.
Oberflächenstrukturen, Transistoren 280#
Oberflächenwellen 428f.
Oberflächenwellen, magnetische 429
Oberflächenzustände 168
Oberwellen, Entstehung 62
Oberwellensonden 449
O'BRIEN-Kode 663#
Oersted 366
Offsetdrift 506ff.

Ohmmeter, direktanzeigendes 619
—, lineares 572#
Open-collector-Stufen 896
Operandenfeld 918
Operandenteil 830
operating power 740
Operationsfeld 918
Operationskode 830
Operationsteil 812
Operationsverstärker 554ff., 571f., 572#
—, Grundschaltungen 569f., 560*
Optimalfilter 644
optische Darstellungen 322f.
Optoelektronik 3, 314ff.
Optokoppler 346f.*
Originalschablone 205
Orthikon 388f.
Ortho-Null 115
Ortskurve 95ff.
Oszillatoren 516ff.
—, Spannungsbegrenzung 517f.
Oszillograph 630ff.
—, dreidimensional 635#
—, Prinzipaufbau 633ff., 634#
Oszillographenröhre 392#, 394ff.
Oxydation, anodische 76
Oxydationsanlage 197#

Packungsdichte 724#, 960
paging 840
paperlined Zellen 28
Parallelschaltung, Transistoren 278
parity-check 664, 873
Partialbruch 118
Pastensysteme 968f., 969*
patching 919
PAULI-Verbot 137
PAULINGsche tetraedrische Radien 190#
PDV-Bus 904ff.
PEIRCE-Funktion 673#, 675#
Pendelrückkopplung 474ff.
Penetrationsröhre 399#
Periodendauer 70
Permeabilität 39f., 41#
—, komplexe 105
Perowskitgitter 417
PETRI-Netz 769ff.
Phasenaddition 507f.#
phasenempfindlicher Gleichrichter 245
Phasenkompensation, OP 560f.
Phasenlaufzeit 120
Phasenmessung 246
Phasenminimumsystem 120
Phasenmodulation 483ff.
Phasenrand 562
Phasenregelkreis 629#
Phasenschieber 115#
Phasenschieber-Generator 518#
Phasenschieber-Schaltung 493
Phasenspielraum 566
Phasensteilheit 98
Phasenverschiebung 72
Phasenwinkel, Frequenzgang 98
Phosphoreszenz 335
Piezo-Effekt 415f., 416#
Piezo-Elektrizität 52
Piezo-Folien 419
Piezo-Keramik 418#, 419f., #, 420*
Piezo-Relais 409
piezoresistiver Effekt 453
Pietoresistivität 416
Pillenkondensator 73
pin-hole 204
Pinch-off-Bereich 250
pinkompatibel 697
PIO 868f.
Pipelining 937
pixel 882
PLA 719f., 720#, 934*
PLANCKsche Formel 318
Planschirm 394
Plasmabauelement 400ff.
Plasmapanel 403#
Plastverkapselung 215
Plotter 413
PN-Plan 101#, 118ff.
—, anschaulicher 121
pn-Übergang 149ff.
Pol-Nullstellen-Plan 118ff.
Polaritätstrenner 581
Pole 101
Polieren 191f.
polling 835, 865
polnische Notation 671
Portabilität 995
Positivlack 202
Potential, elektrisches 8
—, magnetisches 39
Potentialanalogien, PN-Plan 120
Potentialtopf 137f., 173
Potentialverschiebung 511f.
Potentiometer 25
—, kontaktfreies 451
Potenzfilter 129, 132f.
Potenzfunktion mit OV 576
Prädikatenlogik 669
Präzisionswiderstände 24
Prellen 721
Primärelemente 28ff., 30*, 33*
Primärvalenzen 319
Priorität 863, 869
Probensatz 1005
Programmiersprache, höhere 917, 922ff.
Programmablaufplan 814, 910

Programmdokumentation 926
Programme 810
Programmiersprachen 817
Programmierung 908ff.
—, Phasen 925
—, rekursive 814
—, strukturierte 909
Projektionsoptiken 206#
Projektionsoszillograph 636
PROM, Überblick 748#, 751ff., 752#
Prompt-Zeichen 920
Proximity-Belichtung 205
Prozedur 910
Prozeßfolge, Bipolartransistor 209#
—, FET 211#
Prüfbit 665
Prüfen 615
Pseudooperator 918
Pseudotetrade 662
Pseudozufall 801
psn-Struktur 158f.
Puffer-register 794
Pufferung von Zellen 35
Pulscodemodulation 489
Pulslängenmodulator 525ff., 526#
Pulsmodulation 486ff.
Pulverkerne 47
punch-through 169f., 179F
Punktdefekt 188#
Punktraster 323#
Purpurpest 214
push-pull 535
Pyroelektrizität 416
PZT-Keramik 416f., 417#

Quadrantenelektrometer 412#
Qualtität 977#, 978#
Quantentheorie 138
Quantisierung 488ff., 656
Quantisierungsgeräusch 488
Quarz 416, 422f.
—, künstlicher 426
—, Ziehen 424
Quarzschnitt 421#, 423#, 424#
Quarzuhr 425f.
quasikomplementäre Endstufen 543f.
Quecksilber-Zink-Zelle 31
Quellprogramm 918
Querregler 592, 595
Querschnitt, Eisen 93
QUINE-MCCLUSKEY-Methode 682
Quittierungsbus 897

Race 776
Raffungsfaktoren 987*
Rahmen 884
RAM, dynamischer 742ff.
—, statischer 773ff.
Randomfolgen 801
Randomlogik 929
Raumladekapazität 156f.
Raumladestrom 380f.
Raumladungszone 152
Rauschen 281ff.
—, aktive Bauelemente 285#
—, Fotoempfänger 326
—, Kennzahlen 283
—, Widerstände 23f.
Rauschmaß 283
RC-Dreipole 587
RC-Kopplung 252
RC-Siebglieder 109ff., 111#
RC-Verstärker 251ff.
Reaktanzstufe 478, 479#
Realisierung, kanonische 122f.
Rechenschieber 821
Rechentechnik 809ff.
Rechenwerk 811, 826f.
Rechner, Architektur 824#
—, Generationen 823*
—, Grenzen 818f.
—, Größeneinteilung 817f.
—, klassisch 810
—, Schaltungen 802ff.
Rechnungsablauf 813f.
Rechteckgenerator 519ff.#
Rechteckschwingung 53
Rechteckströme 69ff.
Reduktionskamera 204#
REED-Kontakt 408
Reflexlichtschranke 347#
refresh 743, 746
—, automatisch 747
refresh-amplifier 758
refresh-Register 842
Regenerieren von Primärzellen 28
Register 394ff.
—, bei Mikroprozessoren 840ff., 841#
—, dynamisches 799
—, quasistatisches 798f.
Register-ALU 803
Registrierinstrumente 413
Reihenschwingkreis 97f.
Reinheitsstufen 148#
Reinigung, physikalische 186
Reinstsilicium 185
Rekombination 144
rekursive Filter 955
Relais 406ff., 409#
Remanenz 40
Reparatur 977
Reserve 980
Resist 202
Resonanzfrequenz 97f.

Resonanzverluste 83
Resonanzverstärker 256f.
restart 836
restore 743
Reststrom 287ff.
—, Elko 76*
reticle 205
Reziprozität 88
RGB-Signal 400
RICHARDSON-Gleichung 378
RICHARDSON-Konstante 374
RIEMANN-Fläche 650#
Ringmodulator 243f., 244#
Ringmodulatorschaltungen 581
Ringzähler 789
Röhrchenkondensator 73
Röhrenkatoden 379#
Rohsilicium 185
ROM 748#, 750f.
Röntgenstrahllithografie 208
Rotatebefehle 833
Rückführung 461
Rückkopplung 461ff.
—, Einteilung 474#, 481#
—, Kennwerte 464ff.
—, Nichtlinearität 468f., 469#
—, Stabilität 468
Rückkopplungshazard 776, 783
Rückspeichern 743
Rückstellung 788
Rückwirkungsfehler 611, 618
Ruhestromeinstellung 540f.
Rundgehäuse 213#
Rundungsrauschen 955
Rundzellen 28f., 29#

Sägezahngenerator 443, 444#, 521#, 585#
Sägezahnverschlüßler 947, 948#
Sammelleitung 807
sample-hold 515f.#
Samplimg-Oszillograph 636
Samplingtheorem 488, 1005
Sättigungsbereich, FET 171
Sättigungsbetrieb 297
Sättigungsmagnetisierung 46
Sättigungssperrstrom 152
Sättigungsstrom 380
saturated mode 297
SC-Filter 645, 651f., 652#
scale down 747
scale of two 522
Schalenkerne 85f., 86#
Schallgeschwindigkeit 421*
Schaltalgebra 672
Schalterfilter 645, 651, 652#
Schalterprinzipien 297#
Schaltkreisentwurf 183
Schaltkreisfamilien 724
Schaltnetze 669
Schaltnetzteil 604ff.
Schaltnetzwerke 669
Schaltregeln 770
Schaltregler 604ff.
Schalttransistor 275
Schaltungen, arithmetische 715
—, digitale, Systematisierung 927#
—, sequentielle 765ff.
Schaltungsäquivalente 123
Schaltverhalten, Diode 158#
—, FET 297#f.
—, Transistoren 295
SCHEFFERscher Strich 670#, 673#, 675#
Scheibenkondensator 73
Scheibenprozesse 189ff.
Scheibenverschlüßler 947
Scherung 44
Schieberegistor 794
Schirmung 293
Schleifen 911
Schleifenverstärkung 474
Schlitzmaskenröhre 397f., 398#
SCHMITT-Trigger 444, 521ff., 522#, 524, 582f. 585#, 778#
Schnelladung 34
schnelle FOURIER-Transformation 58
Schnittbandkerne 84
SCHOTTKY-Diode 230f.#
SCHOTTKY-Transistor 277
SCHOTTKY-TTL 697
Schreibleitung 737
Schriftarten 323#
Schrot-Rauschen 283
Schutzgaskontakt 408
Schutzringdiode 230#
Schutzschaltung 544f., 596
schwarzer Strahler 317f.
Schwellenspannung, FET 168
Schwellstrom, Laser 343f.
Schwellwertlogik 676
Schwellwertschalter 778#
Schwingkreis 97f.
Schwingkreisspannungsteiler 108#
Schwingung, Darstellungsformen 58
—, freie 100
—, gedämpfte, Darstellung 635#
Schwingungsarten 53#
Schwingungserzeugung 471f.
screen and fire 968
sedezimal 658
Segmentschrift 323#
Seignettesalz 419
Seitenadressierung 840
Sektorsteuerung 433
Sekundärelement 32ff., 34*

Sekundäremission 373#
Selbstentladung 34
Selbstheilung, Kondensatoren 75
Selbstinduktion 67
selbsttaktender Kode 663
select 714
Selen-Gleichrichter 433
Sensoren 446ff.
serielle Busse 910ff.
Seriensystem 980
Serpentinen 758#
SEV 386f.
SHANNON-Formel 1000
SHANNON-Gleichung 1006
SHANNONscher Satz 678
SHELGAKIN-Algebra 675
Shiftbefehle 833
shunt 617
Sicherungen, elektronische 596
Sichtspeicherröhre 395#
Siebdruck 969
Siebfaktor 236
Siebkette 236
Siemens 9
Signale, Z 80 845
—, quantisierte 488ff.
Signaleinteilung 656#
Signalprozessor 955
Signalsynthese 630
Silber-Zelle 31
Silicium als Material 183
Simulator 925
single step 919
single slope 947
Sinnbilddarstellung 910
Sinusschwingung 53
SIO 872f.
SiO_2-Schichten 198*
Skineffekt 82
Slew-Rate 563
smektisch 355#
SMITH-Diagramm 103#
Soft-Error 748f., 990
Software 815ff., 908ff., 994ff.
Softwareentwicklung 996#
Softwarefehler 994
Softwarekrise 816, 820
Softwarepflege 926
Softwaretechnologie 816f.
Softwarezuverlässigkeit 995ff.
Solarzellen 331*
Sondertechnologien 958ff.
Sortieren 615
SOS-Technik 707#, 709
Sourcefolger 493
Sourcekode 918
Spalt 44
Spannband 411
Spannung 8#
—, Anstiegsgeschwindigkeit 439
—, Begrenzung 240
—, erdsymmetrische 595
—, Messung 616
—, Quelle 12
—, —, elektronische 496ff., 497#
—, —, integrierte 513f.#
—, —, programmierbare 594
Spannungs-Frequenz-Wandler 948
Spannungs-Strom-Konverter 572#
Spannungsrückkopplung 464
Spannungsstabilisation 223, 591ff.,
Spannungsteiler, komplexe 106ff.
Spannungsübertragung 20
Spannungsverstärkung, Transistor 264
Spartransformator 93
Speicher 616, 723ff.
—, 1-Transistorzelle 743#, 744#
—, 3-Transistorzelle 742f.
—, allgemeiner 725
—, assoziativer 761ff.
—, bipolarer 732ff.
—, CCD 758
—, direkter Zugriff 835
—, dynamische 742
—, Eigenschaften 726
—, Funktionen 725
—, Hauptdaten 727*
—, Hierarchie 727
—, I^2L 735f., 736#
—, idealer 726
—, magnetomotorischer 884
—, minimaler 725
—, Minimalstrom 737
—, MOS 738f.
—, Organisation 730f., 731#
—, programmierbarer 749f.
—, Redundanz 747
—, Reparatur 747
—, serieller 757f., 758#
—, Soft-Error 748f.
—, Theorie 725
—, TTL 774#
—, Tunneldiode 457
—, VMOS 740f.
—, Zelle 726, 737#
Speicherdrossel 606
Speicherfilter 645
Speicheroszillograph 637
Speicherwerk 811f., 828f.
Speicherzeit, Diode 158#
Spektraldarstellung 58
Spektraldichte 59
Spektralfarben 321
Spektroskopie 623#

Sperrbereich 107
Sperrbetrieb, Diode 154f.
Sperrerholzeit 158#
Sperrschicht-FET 163
Sperrschichtkondensator 73
Sperrspannung, Transistor 296#
Sperrwandler 606f.
Sperrzone 152, 154
Spezialrechner 816, 936ff.
Spiegelbild, Ladung 49
Spitzendiode 230f.#
Spitzenwerte 65
Spitzenwertmessung 581
Splitelektrode 651, 652#
Sprachvergleich 923#
Sprungantwort, OV 562f.
Sprungbefehle 835, 849, 852
Sprünge, irreversible 40
Spulenverluste 82, 124
Sputtern 971f.
Stabilisation 591ff.
Stabilisation von Spannungen 239ff.
Stabilisationsfaktor 240, 593
Stabilität bei Systemen 117
—, thermische 309f.#
Stack 764#
Stackbefehle 832
Stackpointer 765
Standard-Busse 888f.
Standard-Interface 888ff.
Standardfilter 128#
Standardzellen 931
standby-power 740
Stapelspeicher 764#
Stapelzeiger 765
Starkstromtechnik 2#
Startbit 873
statische Logik 669
Steilheit, FET 171
Stelltransformator 93
step and repeat 205
STEPHAN-BOLTZMANN-Gesetz 301, 319
Stern-Dreieck-Umwandlung 11*
Sterninterface 890f.
stetig 655
Steuerbefehle 835
Steuerbus 808, 825f.
Steuerer 895f.
Steuergleichung 249
Steuerung 594
Steuerwerk 811f., 826
STIEBITZ-Kode 663#
STOKESsches Gesetz 5
Stoppbit 873
Störbefreiung 488
Störspektren 292#
Störstellen, isoelektrische 336
Störstellenleitung 140, 144
Störstellenzahl 148
Störungen 291ff.
Störwertaufschaltung 594
Stoßzeit 5
Strahlung 300, 315ff., 316*
Strahlungsausbeute 338
Strahlungseffekte 990
Strahlungsempfindlichkeit 991*
Streifentransistor 966#, 967
Streufaktor 89
Streumatrix 20
Streuung, dynamische 355
stripping 205
strobe 714
Strom, stationärer 4ff.
Strom-Spannungs-Konverter 572#
Stromanstiegsgeschwindigkeit 439
Strombank 511
Strombegrenzung 596
Stromdämmung 616
Stromdichte 6, 8#
Stromflußwinkel 234
Stromkennlinie, Gasentladung 402#
Strommesser 572#
Strommessung 617
Strompfad 618
Stromquelle 12
—, elektronische 494ff.#
—, ideale 465
—, integrierte 514f.#
Stromregler 603ff.
Stromrichtertechnik 434
Stromröhre 7
Stromrückkopplung 464
Stromschalterprinzip 297, 698
Stromschleife 904
Stromspiegel 510f.#
Stromstärke 6, 8#, 9
Stromtransport 5
Stromübertragung 20
Strömungsfeld 6ff., 7#
Stromversorgung 591ff.
Stromversorgung, geschaltete 604ff.
Stromverstärkung, Transistor 264
Struktogramm 910
Struktur von Schaltungen 928
Stufenverschlüßler 949ff.
Subminiaturtechnik 961
Substrat 964f.*
Substratspannung 169f.
Suchtonverfahren 246
Summierwandler 607
Superstrahlung 343
Supraleitung 460
Symboltabelle 918
Symmetrien bei Schwingungen 57#

Symmetrierung 293f.
Synchrondetektor 246
Synchronisation 634
System International 610
Sytemtheorie 116ff.

T-Zyklen 847
Takteingang 779, 782
Taktgeber 812
Talker 894
Tantal-Elko 77f., 78#
Tantalsystem 973
Taschenrechner 821, 934ff.
Tastatur 881
Tasten 454ff.
Technologie, Bipolartranstor 272f.
—, FET 258ff.
—, Grobschema 184#
—, Halbleiter 183ff.
—, *pn*-Übergänge 277
—, Vergleich 975ff.
Teiler 785ff.
—, frequenzunabhängig 109
Temperatur, kritische 459*
Temperaturabhängigkeit, *pn*-Übergang 155f.
Temperaturbeiwert 221
Temperatureinfluß bei Transistoren 289
Temperaturgeschwindigkeit 4
Temperaturkoeffizient 73
Temperaturspannung 378
Terme 137
Terminal, intelligentes 360#
Tesla 39
Testphase 919
Tetrade 662
Textur 47
Textverarbeitung 809
Thermo Point 961
Thermoemission 377ff.
Thermokompression 214
Thermorelais 449
Thermostat 458
Thyratron 433
Thyristor 437ff., #
Thyristor, abschaltbarer 441#
Thyristor-Familie 439f., #
Tiefpaß 106#, 124
Tiefpaß, diskret 649
Tieftemperaturelektronik 458ff.
Time-devision-Prinzip 579
TN-Zellen 357
Toleranzfelder digitaler Pegel 687
Top-down 908f.
Trace 919
Trafoblech 83#, 85
Trägerfrequenzverfahren 626
Trägerkonzentration, Diode 154#, 155#
Trägerpaar 140
Transformator 88ff.
Transformatorblech 84
Transientenrecorder 637
Transistor, als Diode 231#
—, Minimalabmessungen 289ff.
—, physikalisches Modell 178
—, Sperrspannung 296#
—, Verstärkerwerte 265f.*
Transistorsound 553f.
Transistorverzerrungen 553ff.
Transistorzahl 828
Transitfrequenz 271
Transition 769
Transmissionsgatter 515#, 708, 799
Transportbefehle 830ff., 855ff.
Transversalfilter 645
Transverter 605f.
Trapezfehler 393f.
Traveling-wave-Röhre 395
Treppengenerator 585#
TRIAC 440
Trichlorsilan 185, 186#
Triggerdiode 443
Triggerung 634
Trimmer 25, 80, 82#
Trinitron 399#
Triodenbereich, FET 171
Tristate-TTL 695#
Trommelspeicher 821
TSCHEBYSCHEFF-Filter 129, 132f.
TTL-Ausgangsstufen 694, 695#, 689
TTL-Familien 697
TTL-Generationen 697#
TTL-Kenndaten 691#
TTL-Kennlinien 692#
TTL-Leistungsgatter 692#
TTL-Prinzip 688f.
TTL-Schaltpunkt 693
TTL-Technik 688ff.
TTL-Transistoraufbau 281
TTL-Versorgungsspannungen 694
TTL-Zustände 690
Tunneldiode 456#
Tunneleffekt 456
TURING-Automat 816, 819
Turmalin 419

Überanpassung 16
Übergabefunktion 771
Übergangsautomat 768
Überlagerungsgesetz 17
Überlaufverzerrungen 955
Übernahmeverzerrungen 539#
Überraschungswert 1001
Übersetzungsverhältnis 90
Überspannungsschutz 597

Übertrager 88ff.
Übertragungsfunktion 117, 588
UI-Schnitt 84
Umgebungsbeanspruchung 985*
Umkehrverstärker 570#f.
Umlaufregister 797
Umlaufspeicher 757f.
UMOS-FET 259
Umrichter 434
Unentscheidbarkeit 819
Unijunction-Transistor 443
Universalität, praktische 817
Universalrechner 811
Univibrator 522#, 523#
Unteranpassung 16
Unterlastkurve 23#
Unterprogramm 836#, 913
Urlader 920

Vakuumelektronik 370ff.
Valenzband 138
Variometer 88#, 94
Varistor 224#
VDR-Widerstand 224#
Vektorscan 207#
VENN-Diagramm 674
Verarbeitungsschritt 909
Verarmungstyp, FET 167f.
Verbindungstechniken 960#
Verbundentropie 1003
verdrahtetes Oder 694
Verfügbarkeit 818, 977
Vergleich elektromagn. Felder 50*
Vergleichen 610
Vergleicher 715
Vergleichsbefehle 833
Verhältnisskalen 610
Verluste, Hysterese 82
—, Kondensator 73
—, Nachwirkung 83
—, Spulen 82f., 124
—, Ummagnetisierung 41#, 93#
—, Wirbelstrom 82f.
Verlustwinkel 73
VERNEUIL-Verfahren 189#
Verschiebelinien 49
Verschiebung, dielektrische 49
Verschlüßler 944
Versetzung 188#
Verstärker, einstufiger 490ff., 491#, 492#
—, frequenzselektiver 470f.#
—, parametrischer 226
Verstärkerröhren 371ff.
Verstärkungswerte, Transistor 265f.*
Verteilungskonstante 186*
Verzögerungsglieder 649
Verzögerungszeit 724#
Verzweigungen 911
Verzweigungsnetzwerke 646ff.#
Video-Interface 868, 883
Vidikon 388f.
Vielfachinstrument 619f., 620#
Vielschicht-Keramik-Kondensator 73
Vierpol 18ff.
—, Systematik 20f.
—, Umrechnungen 19*
Vierschichtdiode 435f., #
VILLARI-Effekt 414
VMOS-FET 259
Volladder 717, 749
Vollterm 679
Vor-Rückwärtszähler 789, 791#, 792
Vorbereitungseingang 779, 782
Voreinstellung 788
Vorratskatode 379
Vorspeicher 783
Vorwärtskopplung 554#

Waferprozesse 189ff.
—, direkte 250f.
wägbarer Kode 663
Wägeprinzip 948ff., 949#
Wahrheitstafel 674
Wahrheitswert 670#
Wahrscheinlichkeitsdichte 1002
WALSH-Funktion 53, 60
Wandler 446
Wandverschiebung 40
Wärme, spezifische 302#
Wärmeabfuhr 5
Wärmeableitung 300ff.
Wärmeleitfähigkeit 302*
Wärmeleitrohr 307f.
Wärmeübergangszahl 302*
Wärmewiderstand 303ff., 305*, 306#*
Wartbarkeit 977
Wartung 981
Weber 39
Wechselgrößen, Messung 63f.
Wechselplattenspeicher 821
Wechselrichter 434
Wechselstrom 52ff.
Wechselstromschaltungen, lineare 95ff.
Wechseltaktschrift 883f.
Weglänge, freie 5
Wegumsetzer 947
Weicheiseninstrument 412#
weichmagnetische Materialien 46
WEISSsche Bezirke 38#
Wellenpaket 335
Wellenparametertheorie 124ff.
Wellenwiderstand 125, 127
Wertetabelle 674, 768
Wettläufe 776

WHEATSTONE-Brücke 11, 112#
WHITE-Folger 502#
Wickeltechnik 961
Wicklung 80
—, bipolare 84
Wicklungskapazität 83#
Widerstand, differentieller 61
—, Eigenschaften 24*
—, elektrostatischer 49
—, komplexer 95
—, magnetischer 42, 66, 81, 90
—, Matrix 18*
—, Netzwerk 10, 11#, 23
—, nichtlinearer 61
—, ohmscher 9ff.
—, Parallelschaltung 10
—, Reihenschaltung 10
—, spezifischer 7*, 8#
—, —, Halbleiter 149#
—, technischer 21ff.
Wiederholstruktur 829
WIEGAND-Sensor 452f., 453#
WIEN-Brücke 112#
WIEN-Glied 109ff.
WIEN-ROBINSON-Generator 518#
WIEN-Spannungsteiler 110
WIENsches Verschiebungsgesetz 318
WINCHESTER-Speicher 822
Wirbelstromverluste 82f.
wired or 694
Wirkungsgrad, Endstufe 528
Wirtscomputer 924
WOLMAN-Frequenz 105
Worst-case-Störabstand 693
Wortleitung 737
Wortorganisation 731

XY-Schreiber 413, 414#

Z-Diode 231
Z-Diode, einstellbare 497f., 498#
z-Transformation 649f., 650#
Zähl-*FF* 785
Zählen 615
Zahlendarstellungen 666
Zahlensysteme 659f.
Zähler 788ff.
Zählermode 876
Zählregistor 794
Zählschleifen 911
Zahlwörter 657
Zebragummi 359#
Zeichenerkennung 810
Zeigerdarstellung 58
Zeitbereichsfilterung 627
Zeitfunktion 58
Zeitgebermode 877
Zeitinvarianz 117
Zeitkonstante 68
Zeitmultiplex 486
zeitraffende Methoden 989f.
Zeitverhaltensfilter 133
Zellenentwurf 931f.
Zellrechner 938
ZENER-Effekt 156f.
Zentraleinheit 812
Zerhackerverstärker 584
zero-page 839
Ziffernanzeigeröhre 404
ZOBEL-Halbglieder 126#
Zonenfloaten 189#
Zonenschmelzverfahren 187
Zufallsfolgen 801
Zugriffszeit 726
Zündimpulse 438
Zündstrom 436
Zungenfrequenzmesser 412#
Zuordner 669
Zustand, innerer 766, 771
Zustandsautomat 768
Zustandsdichte 143
Zustandsknoten 769
Zustandskodierung 771
Zuverlässigkeit 976ff.
Zweierkomplement 666
Zweiflankenwandler 948#
Zweipol 13ff., 95ff.
—, Eigenschaften 105#
Zweipolfunktion 117
—, zulässige 121
Zweipolverstärker 467#
Zweiweggleichrichter 238f.#
Zwischengitterplatz 188#
Zwischenschablone 205
Zyklen 847
—, Halbleitertechnologie 183
zyklischer Kode 663
Zyklus stehlen 878

Berichtigungen zur 5. Auflage

S. 11, Tab. 1.1.3, 2. Zeile in 1. Spalte muß lauten:

$$R_2 = \frac{r_1 \cdot r_3}{r_1 + r_2 + r_3}$$

S. 18, Tab. 1.1.4, 3. Zeile in 3. Spalte muß lauten:

$$\begin{pmatrix} U_1 \\ I_1 \end{pmatrix} = \mathbf{A} \begin{pmatrix} U_2 \\ I_2 \end{pmatrix}$$

S. 68, Gl. (13) muß lauten:

$$U_C = U_0 - RC \frac{dU_C}{dt}$$

S. 203, Im ersten Satz des Textes sind die Lichtwellenlängen in nm statt in µm zu lesen.

S. 318, In Gl. (7) muß im Nenner λm durch λ_m ersetzt werden.

S. 333, Bilder 6.1.14 und 6.1.15
Die Dotierungswerte N und P müssen n und p lauten.

S. 334, Bild 6.1.16a) Änderung wie S. 333

S. 422, Bild 6.3.11b)
Das linke f_p ist durch f_s zu ersetzen.

S. 442, Bild 6.4.9h) Änderung wie S. 333

S. 473, Bild 7.1.8b)
V_k und V_{kt} sind durch v_k und v_{kt} zu ersetzen.

S. 475, Bild 7.1.10
statt U_{MF} lies U_{NF},
statt U_p lies f_p,
statt v_p lies f_p.

S. 477, Gl. (39) muß lauten:

$$i_a = \pm S u_i$$

S. 497, 6. Zeile v. u.
lies: Sie müssen mit . .
statt: Sie muß mit . . .

S. 543, Bild 7.4.10
Bei der Kreuzung der Verbindung T_1 und T_2 und der Ausgangsleitung fehlt der Punkt.

S. 543, Bild 7.4.11b)
T_5 ist ein *pnp*-Transistor

S. 546, Bild 7.4.13b)
T_{16} ist ein *npn*-Transistor

S. 548, Bild 7.4.14c)
Der Transistor ganz rechts, unten heißt T_{14}.

S. 564, Tab. 7.5.3
Der unbezeichnete Widerstand im Bild ist R_k.

S. 566, Bild 7.5.7
An der Verbindung der Transistoren T_{13}, T_{14}, T_{15} und T_{18} fehlt je ein Kreuzungspunkt

S. 583, Bild 7.5.18b)
Die Spannung an den ZENER-Dioden lautet U_Z statt U_2

S. 589, Tab. 7.5.9
In der 1. Spalte unter TP lautet die zweite Formel $1/R_2 = 2\omega_0 CVG$

S. 711, In der Tabelle muß es unter Spalte $n = 16$ lauten: 1048592
262152

S. 741, Bild 8.3.11a)
Der FET bei der Stromquelle hat am unteren Ende einen abwärts weisenden Pfeil.

S. 756, Bild 8.3.22e)
Unter Select muß stehen Si_3N_4.

S. 805, Bild 8.5.23c)
Vor dem Volladder muß stehen $\ddot{u}_{n-1}$ statt $\ddot{u}_{-1}$.

S. 898, Bild 8.7.4b)
Statt NDFR lies NRFD.

S. 956, Bild 9.3.3a)
Statt: Zurückgesetzen lies: Zurücksetzen,
statt: grithm.-Log.
lies: Algorithmisch-logische

Im Sachwortverzeichnis fehlen:
Low-power-TTL 697 ≠, 991 *
Schneiderman, Nassi 910