S-Information

Das S steht für Shannon: als seine wichtige Arbeit [Sha48] mehr als Wieners [Wie48].
Sie ist ***die*** theoretische Grundlage der gesamten Nachrichtentechnik als Teil der Informationstheorie.
So wird der Raum für eine Kommunikation fast bis ins Unendliche erweitert.
Dabei werden die Zeichen bevorzugt elektromagnetisch, aber immer stofflich-energetisch übertragen.
Erst am Empfangsort wirken sie als Informer auf das „richtige“ System ein, wobei das Informat entsteht.
Paradox scheint es dabei zu sein, dass für die Übertragung nicht ihr Inhalt (ihre Bedeutung) wichtig ist.
Denn hier ist nicht das Informat zu erzeugen, sondern nur das Zeichen, also der Informer zu übertragen.

Dafür reichte die Wahrscheinlichkeiten aller Zeichen als wesentliche Kenngröße.

Es ist nicht geklärt, wie Shannon zu dieser Entscheiddung kam.
Wiener schreibt in seinem Vorwort zu [Wie48]: bezüglich der Theorie des Informationsgehaltes seien R. Fischer, C. Shannon und er ungefähr gleichzeitig zur statistischen Beschreibung übergegangen
An anderer Stelle verweist er auch noch auf J. v. Neumann.
Genaueres dürfte kaum noch zu erfahren sein, denn alle vier waren im Krieg mit der Kryptographie befasst und daher zur strengsten Geheimhaltung auch nach dem Krieg verpflichtet.

Ohne den obigen Hintergrund ergibt sich folglich der ***scheinbare*** **Widerspruch**:

- Die ***Nachrichtentechnik dient*** primär der ***Übermittlung von Aussagen, Fakten, Inhalten, Wissen usw***.
- Ihre (informations-) ***theoretische Beschreibung*** sieht jedoch hiervon völlig ab, sondern benutzt ausschließlich ***die Statistik der Zeichen.***

Fortsetzung S-Information

Es ist unerklärbar warum Weaver im Vorwort zu [Sha48] genau das Gegenteil zu zeigen versucht,
und damit sogar im deutlichen Widerspruch zu allen entsprechenden Aussagen Shannons steht.

Leider ist das Vorwort von Weaver auch sonst wissenschaftlich weitgehend unsolide.
Z. B. fehlt jede Literaturangabe, es kann daher bestenfalls als mittelmäßiger Essay betrachtet werden.
Weiter sei erwähnt, dass Shannon sogar einen Streit mit seiner Frau hatte, weil sie als Programmiererin seine Arbeit für die Generierung von Musik benutzen wollte

Wegen der Statistik ist es notwendig, von endlich vielen (n) diskreten Zeichen z_i auszugehen
dabei muss deren Wahrscheinlichkeiten als p_i der zu übertragenden Datei gegeben sein.
Beide Kennwerte der n zu übertragenden Zeichen (z_i und p_i) werden (Zeichen-) Alphabet genannt.

Das verlangt automatisch die damals noch nicht existierende ***Digitaltechnik*** als wesentliche Grundlage.
Die Summe der für die damalige Zeit ungewöhnlichen Annahmen, ist ein Beweis für Shannons Weitblick.

Die bestmögliche Übertragung

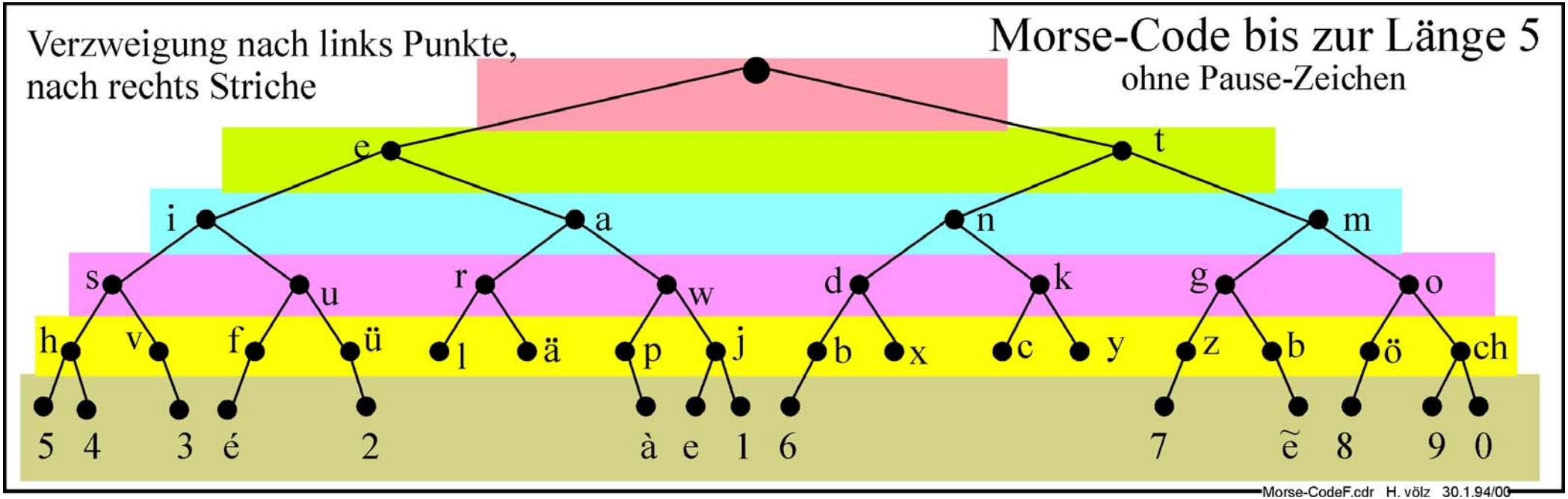
Bei der Übertragung der Zeichen über den Kanal müssen sie empfangsseitig wiedererkannt werden.
Dazu müssen sie getrennt einzeln auftreten und daher durch Pausen oder Sonderzeichen getrennt werden.
In der Schrift gibt es hierzu die Leer- und Interpunktszeichen.
Bei einer technischen Übertragung verlängern sie aber beträchtlich die Übertragungszeit.
Um das zu vermeiden erfand Shannon eine neuartige Zeichendarstellung, den ***Präfix-Code.***
Er heißt auch ***irreduzibel, kommafrei*** und ***natürlich*** und kommt ohne diese „Hilfs-“ Zeichen aus.
Dadurch folgen die Zeichen unmittelbar dicht aufeinander.
Dennoch muss der Empfänger sie zuverlässig erkennen?
Dafür müssen aber nach Shannon die Zeichen einer zusätzlichen Regel genügen:

Kein Zeichen darf der Anfang eines anderen sein.

Z. B. dürfen neben dem Zeichen „Haus“ nicht mehr „Haushalt“, „Hausordnung“, „Hause“ usw. auftreten.
Das widerspricht zwar völlig unserem Sprachgebrauch, ist aber technisch relativ gut zu verwirklichen.
Zur Übertragung müssen die ursprünglichen Zeichen auf spezielle Zeichen (Codewörter) umgesetzt werden.
Das verlangt eine Methode zur systematisch Generierung der Codewörter
Wesentlich sind dafür digitale Codebäume, bei denen aber an den Endknoten gültige Codewörter stehen.

Der Morse-Code

Der Morse-Code ist zwar kein Präfixcode, lässt aber gut dessen Forderung erkennen.
Er ist auch nicht – wie zuweilen fälschlich angenommen – ein binärer Code aus Punkten und Strichen.
Er benötigt sogar drei Trennzeichen für Pausen:
1. zwischen Punkten/Strichen, 2. den Code-Zeichen und 3. Wörtern.
Den größten Teil (bis zur Tiefe 5) des Codebaumes zeigt das Bild.



Ein Beispiel Morsecode

Gegeben sei die Punkt-Strich-Folge

••••••-••-••--••••-••

Gemäß den ausgewählten „Text“ gilt für deren Zeichen

a	d	e	h	i	l	n	r	s	w
.-	-..	.	....	..	.-..	-.	.-.	...	.--

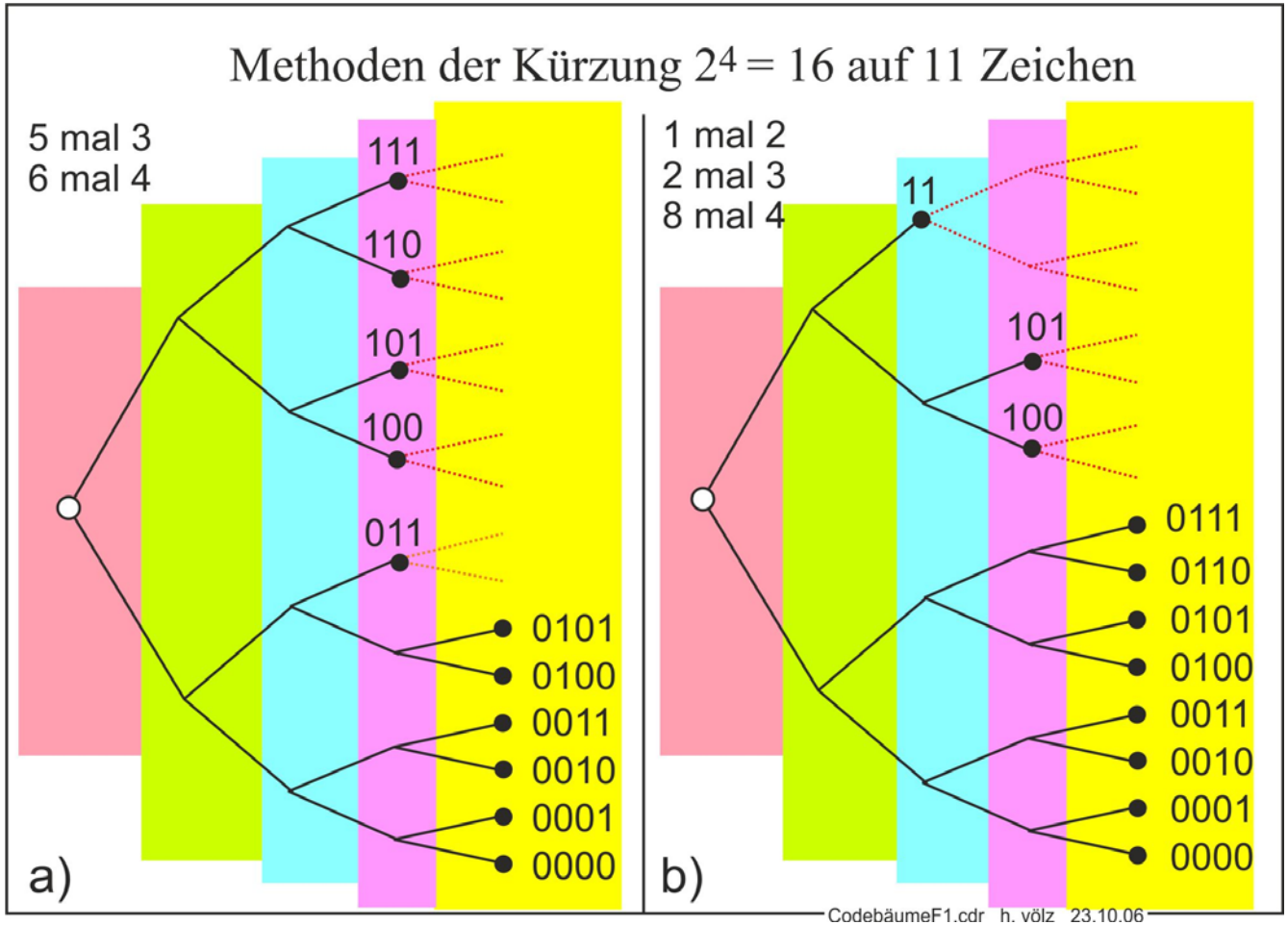
Daher sind u. a. die drei folgenden Decodierungen möglich:

„seinen adel“ oder „herr weil“ oder „ies nah d“

Die Ursache dieser Unbestimmtheit entsteht dadurch, dass beim Morse-Code nicht nur die Endknoten, sondern auch andere Knoten des Baumes für die Code-Zeichen benutzt werden.

Der Shannon-Code 1

Shannon geht von einem vollständigen binären Codebaum aus.
Bei einer Tiefen m entstehen dann 2^m Endknoten.
Für $n < 2^m$ verkürzt er vollständigen Baum durch Zusammenlegen von Endknoten, im Bild 16 $\rightarrow$ 11.
Jetzt besitzt kein Codewort den Anfang eines anderen. Daher sind Trennzeichen überflüssig



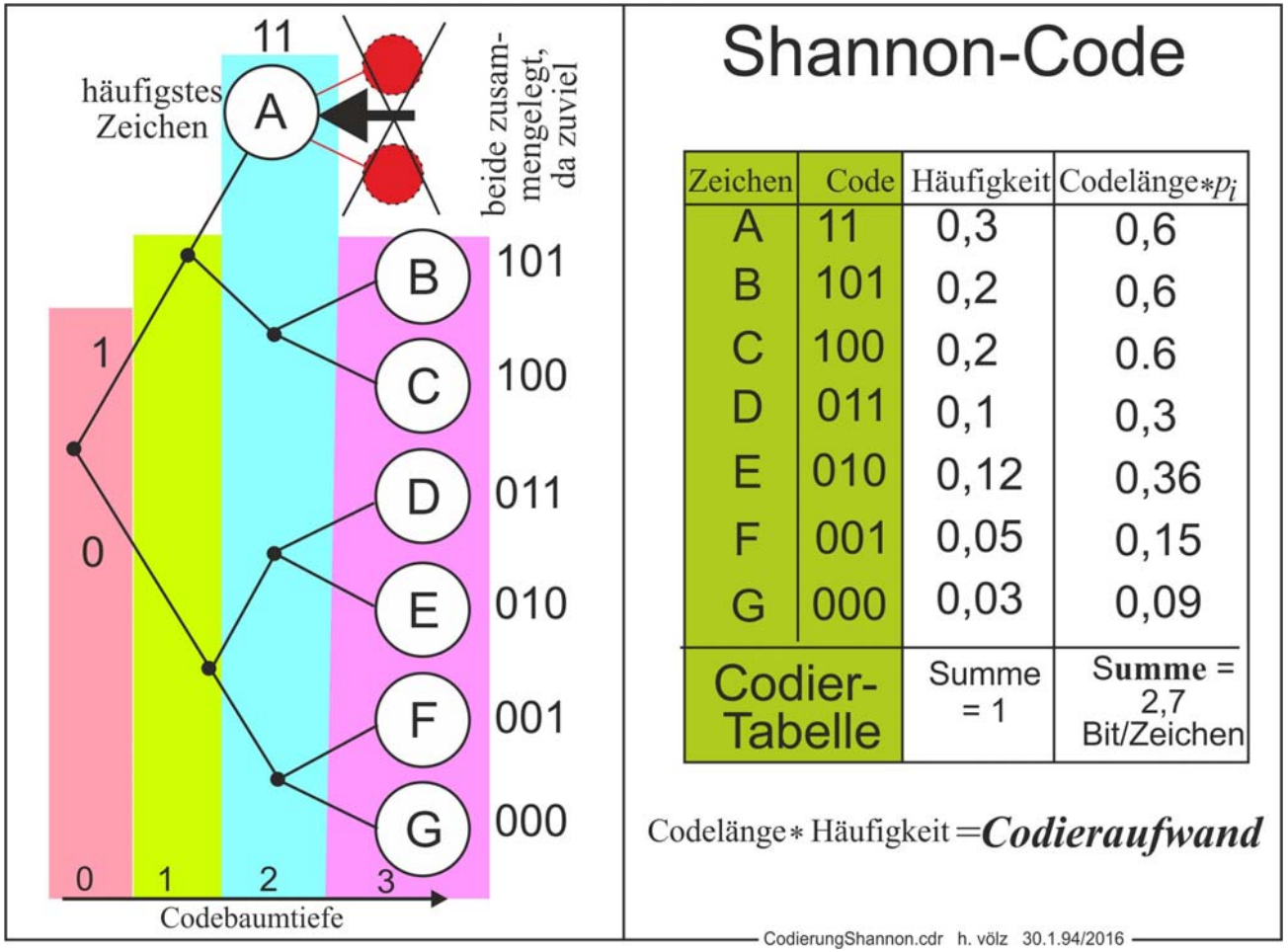
Der Shannon-Code 2

Eine Untersuchung kann z. b. mittels der Tabelle aus 7 Zeichen z_i = A bis G und den dortigen p_i erfolgen.
Der vollständige Codebaum hat die Tiefe 3 mit $2^3 = 8$.
Für die 7 Zeichen werden 2 Endknoten zusammengelegt.
Der Baum erzeugt dann die 7 Codewörter von 11 über 101 bis 000.
Die kürzeste Bitfolge 11 erhält häufigste Zeichen A.

Das Produkt aus Häufigkeit p_i und Codelänge c_{Li} bestimmt die Übertragungszeit des Zeichens.
Die Summe bestimmt einen gemittelten ***Codeaufwand*** der Übertragung.

$$C_A = \sum_{i=1}^{n} p_i \cdot c_{Li} \ .$$

Er beträgt 2,7 **Bit/Zeichen**.
Für die gesamte Übertragungszeit ist er nur noch mit der Anzahl der zu übertragenden Zeichen zu multiplizieren.
Es ist schnell zu prüfen, dass jede andere Codezuordnung mehr Codeaufwand verlangt.



Shannon-Code

Zeichen	Code	Häufigkeit	Codelänge$*p_i$
A	11	0,3	0,6
B	101	0,2	0,6
C	100	0,2	0.6
D	011	0,1	0,3
E	010	0,12	0,36
F	001	0,05	0,15
G	000	0,03	0,09
Codier-Tabelle		Summe = 1	**Summe** = 2,7 Bit/Zeichen

Codelänge$*$ Häufigkeit =***Codieraufwand***

CodierungShannon.cdr h. völz 30.1.94/2016

Shannonsche Entropieformel

Shannon stellte auch die Frage nach dem kleinstmöglichen Codeaufwand = minimale Übertragungszeit.
Wie er sein Ergebnis fand, ist leider nicht bekannt. Es gibt auch keinen einsichtigen Beweis dafür.
Deshalb wird das Ergebnis zunächst ohne Beweis übernommen, aber anschaulich erklärt.
Dazu sei eine nichtganzzahlige Codebaumtiefe c_{tif} erlaubt, die genau n Endknoten erzeugt:
Dafür müsste gelten c_{tif} = ld (n).
Für die Einzelzeichen mit den p_i, folgt eine ebenfalls eine nichtganzzahlige Codelänge.
Sie erzeugt den idealen Codeaufwand je Zeichen

$$h_i = C_{Ai} = p_i \cdot \mathrm{ld}(1/p_i) = -p_i \cdot \mathrm{ld}(p_i).$$

Durch Summierung ergibt sich dann die Shannonsche Entropieformel:

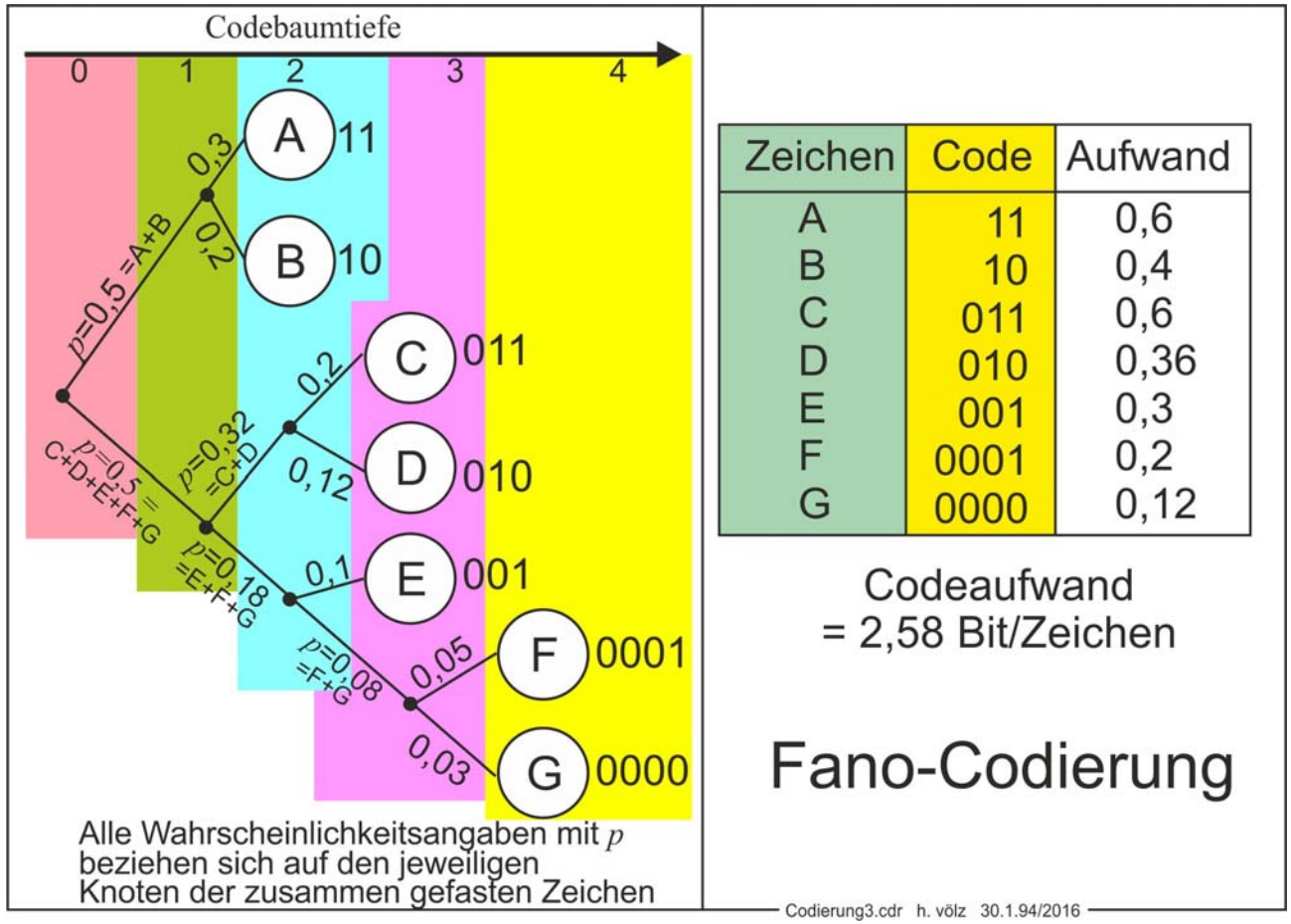
$$H = -\sum_{1=1}^{n} p_i \cdot ld(p_i).$$

Das vielfach hervorgehobene negative Vorzeichen folgt zwangsläufig wegen $p_i < 1$ mit $\mathrm{ld}(p_i) < 0$.
Es ist also notwendig, damit die Entropie positiv ist.
Der Name Entropie wurde Shannon von Wiener vorgeschlagen und wird später noch genau erklärt.
Für den vorigen Zeichenvorrat beträgt die Entropie H = 2,52 Bit/Zeichen.
Wegen des obigen Codeaufwands mit C_A= 2.7 Bit/Zeichen besteht Verbesserungsbedarf.
Hierfür mussten aber erst andere Verfahren zur Erzeugung anderer Codebäume entstehen.

Der Fano-Code

Zunächst fand 1849 Fano dafür den folgenden Algorithmus [Fan61]:

1. Die Zeichen sind nach fallenden Wahrscheinlichkeiten zu sortieren.
2. Es werden zwei Gruppen (als gedankliches Superzeichen) mit möglichst gleicher Wahrscheinlichkeit gebildet.
3. Alle Zeichen der „oberen“ Teilgruppe werden mit 1 codiert, alle der unteren mit 0.
4. Für jede Teilgruppe ist nach 2. fortzufahren, dabei werden die 1 bzw. 0 jeweils hinten ergänzend angefügt.
5. Es ist solange mit jeder entstandenen Teilgruppe fortzufahren, bis jeweils ein einzelnes Zeichen erreicht ist.



Zeichen	Code	Aufwand
A	11	0,6
B	10	0,4
C	011	0,6
D	010	0,36
E	001	0,3
F	0001	0,2
G	0000	0,12

Codeaufwand
= 2,58 Bit/Zeichen

Fano-Codierung

Auffällig ist: Die Zeichen mit kleiner Wahrscheinlichkeit besitzen eine größere Codebaum-Tife

Der Codeaufwand beträgt nur noch 2,58 Bit/Zeichen, erreicht aber noch nicht die theoretische Grenze.

Der Huffman-Code

Den nächsten Algorithmus schuf Huffman 1952 [Huf52]:

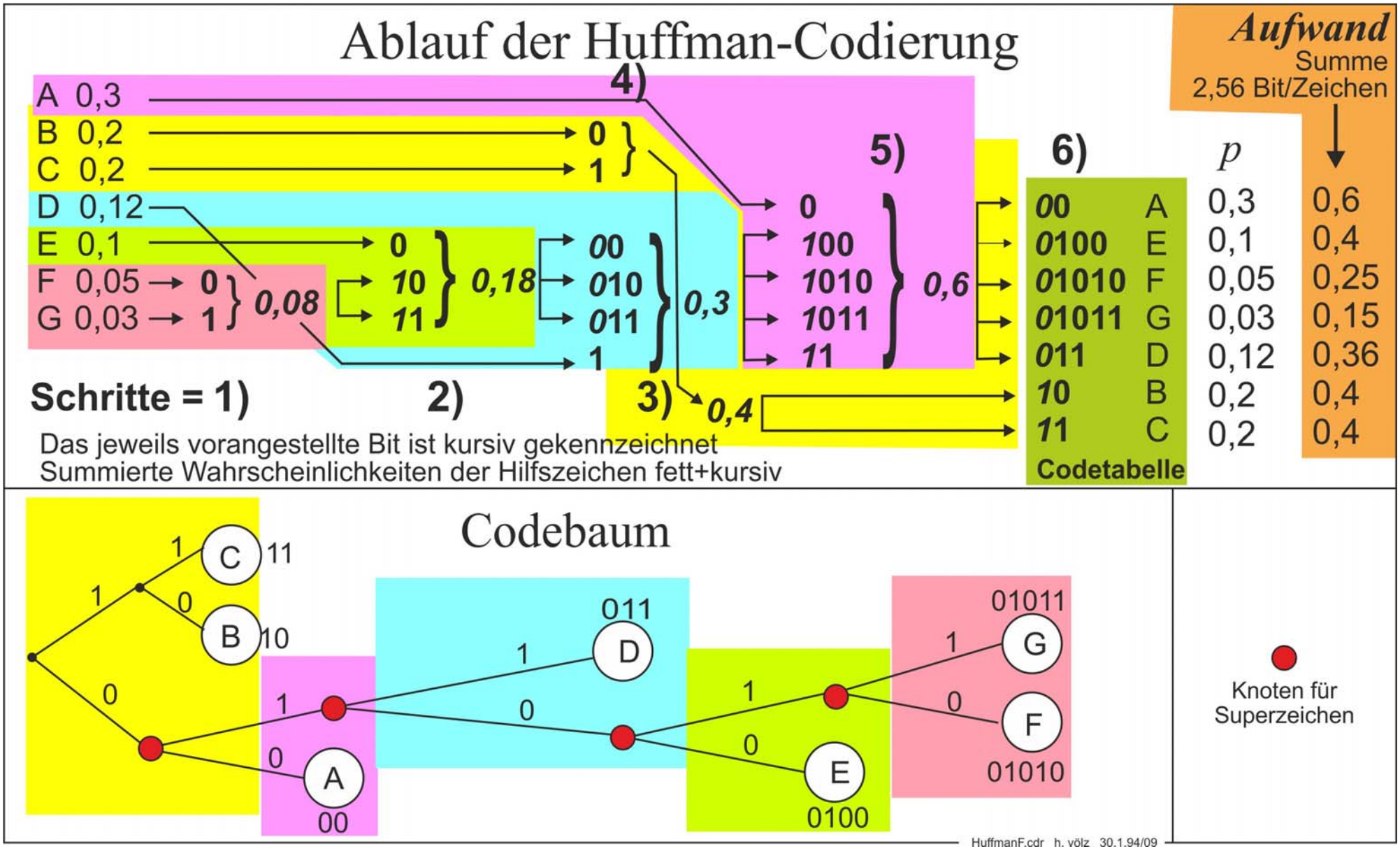
1. Die Zeichen werden nach fallender Wahrscheinlichkeit sortiert.
2. Die beiden Zeichen mit kleinster Wahrscheinlichkeit werden mit 0 bzw. 1 codiert. Bei Wiederkehr hierher wird der zu ergänzende Code immer vor dem schon vorhandenen eingefügt.
3. Beide Zeichen sind aus dem Symbolvorrat zu entfernen und als ein neues Hilfszeichen (Superzeichen) mit den addierten Wahrscheinlichkeiten eingefügt.
4. Bei 1. ist solange fortzufahren, bis nur zwei Hilfssymbole existieren.

Das führt zu einem kompliziert zusammenhängenden und damit weniger übersichtlichen Ablauf. Für das obige Beispiel ergibt sich das folgende Bild

Der Codeaufwand sinkt damit auf C_A = 2,56 Bit/Zeichen, ist aber immer noch deutlich kleiner H = 2,52

Seit Huffman ist aber kein besserer Algorithmus mehr gefunden worden.

Vielleicht existiert er auch nicht. Es nur wenig überzeugende Versuche, das zu beweisen [Fan66].


Ablauf der Huffman-Codierung
Aufwand
Summe
2,56 Bit/Zeichen
A 0,3
B 0,2
C 0,2
D 0,12
E 0,1
F 0,05
G 0,03
0,08
0,18
0,3
0,6
0,4
Schritte = 1)
2)
3)
4)
5)
6)
Das jeweils vorangestellte Bit ist kursiv gekennzeichnet
Summierte Wahrscheinlichkeiten der Hilfszeichen fett+kursiv
Codetabelle
p
Codebaum
Knoten für
Superzeichen
HuffmanF.cdr h. völz 30.1.94/09

Senkung der Redundanz

Wegen $H \le C_A$ wird die Redundanz einer Codierung $R = C_A - H$ bzw. die relative Redundanz eingeführt

$$r = \frac{C_a - H}{H}.$$

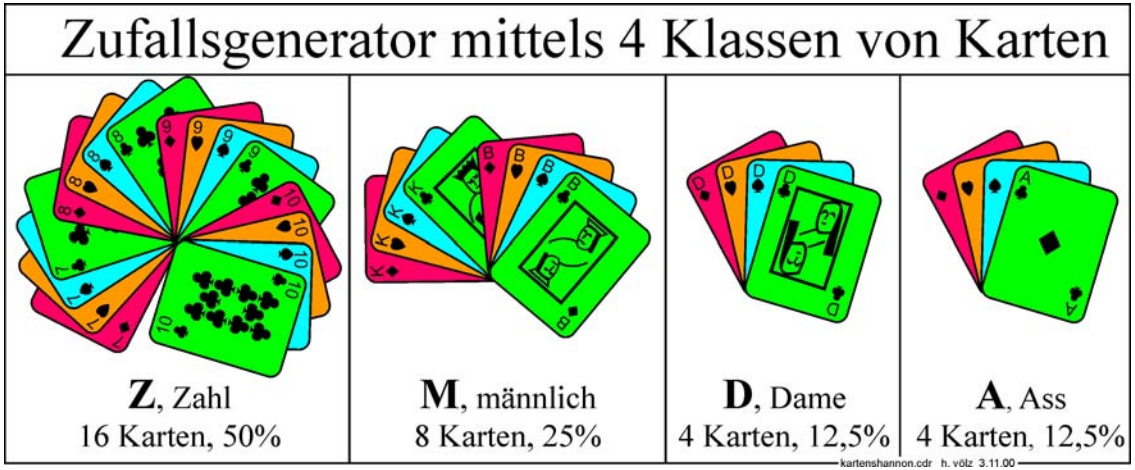
Bereits Shannon bewies: theoretisch ist $r \to 0$ möglich.
Sein Beweis benutzt Codekombination mit Gedächtnis.
Das sei an einer Datei mit 2 Zeichen (A, B) und den Wahrscheinlichkeiten $p_A = 0{,}75$ und $p_B = 0{,}25$ gezeigt:
Ihr Coderaufwand ist $C_A = 1$ und die Redundanz $r \approx 23$ %.
Die 1. abgewandelte Datei benutzt AA, AB, BA und BB.
Für sie sinkt auf ≈1,4 %.

Das Prinzip der Kombinationszechen ist fortsetzbar.
Dabei werden die Codewörter immer länger (theoretisch ∞)
Für die Codierung und Decodierung benötigen sie zusätzlichen Speicherplatz.
So verzögert sich auch die Übertragung, denn die Decodierung kann erst dann erfolgen, wenn ein vollständiges Kombinationszeichen vorliegt.
Daher wird dieses Prinzip nur selten und immer begrenzt auf wenige Kombinationen benutzt.

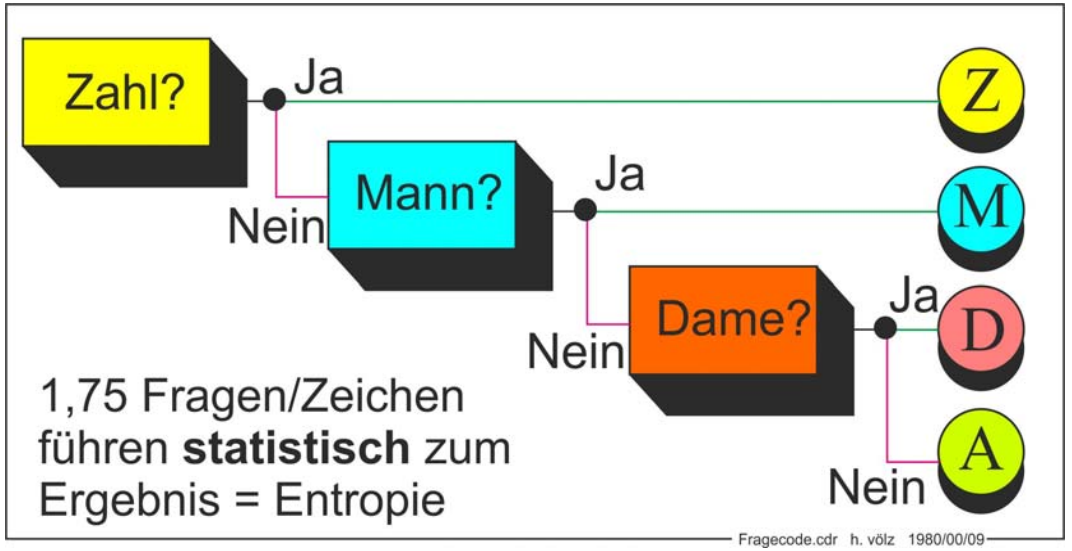
Zeichen	Wahrschein-lichkeiten	Codes	Aufwand, Redundanz
Einzelzeichen			
A	0,75	0	1
B	0,25	1	≈23 %
Zweierzeichen			
AA	0,5625	0	
AB	0,1875	11	0,84375×2
BA	0,1875	100	≈4 %
BB	0,0625	101	
Dreierzeichen			
AAA	0,421875	1	
AAB	0,140625	001	
ABA	0,140625	010	
BAA	0,140625	011	0,82292×3
BBA	0,046875	00000	≈1,4 %
BAB	0,046875	00001	
ABB	0,046875	00010	
BBB	0,015625	00011	

Das kann auch als zusätzliches Argument für die Entropieformel gelten.

Beispiel redundanzfreie Codierung



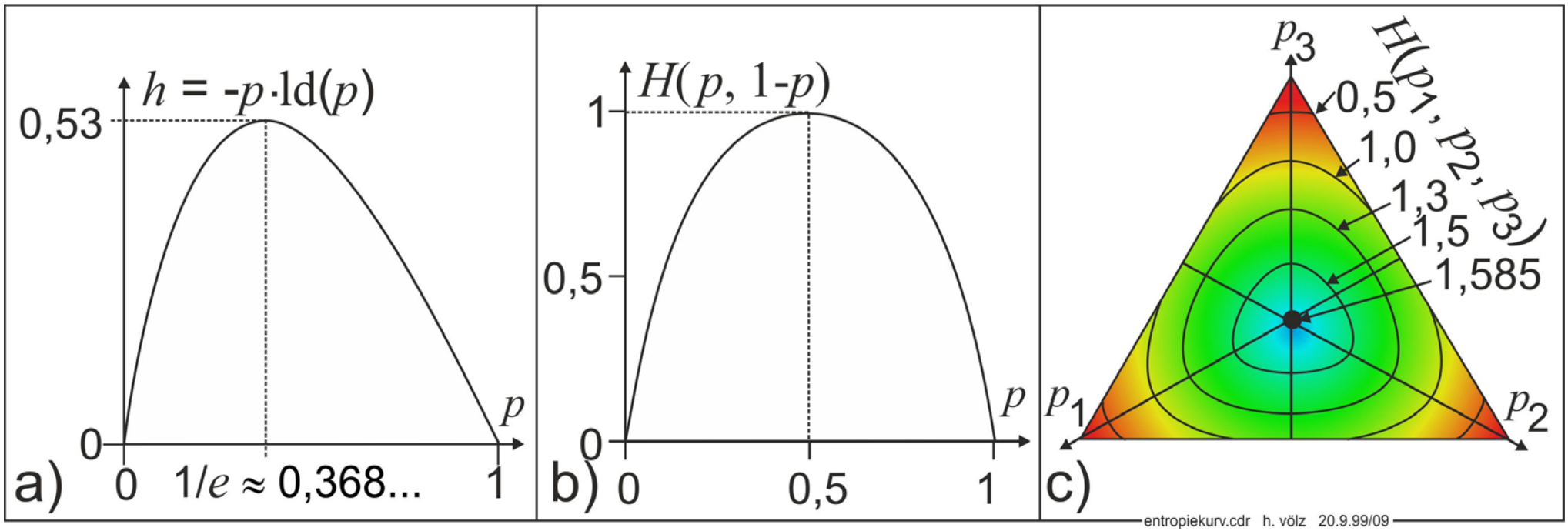
Zuweilen sind auch redundanzfreie Codierungen möglich. Ein Ratespiel ist ein Beispiel.
Dazu werden die Skatkarten in 4 Klassen geteilt.
Nach dem Mischen tritt für jede Klasse das p_i auf.
Es bestimmt den Aufwand für das Erraten der Klasse einer zufällig gezogenen Karte.
Er wird mittels Ja/Nein-Fragen bestimmt.
Bei einer optimalen Fragestrategie entsteht der Codebaum.
Der Codeaufwand berechnet sich daraus zu 1,75 Bit/Zeichen.
Die Entropie besitzt denselben Wert.
Es liegt also ein bestmögliche Fragstrategie (Codierung) vor.
Zwar kann es noch andere gleichwertige, aber keine bessere Codierungen geben.



1 Bit entspricht so genau der Antwort auf eine Ja/Nein-Frage

Entropie-Verläufe

Sie sind bei kontinuierlich veränderlichen Wahrscheinlichkeiten für 1, 2 und 3 Zeichen drstellbar.
Auffällig ist der Verlauf für den Einzelwert $h = -p \cdot \mathrm{ld}(p)$ mit ausgeprägtes Maximum bei $p = 1/e \approx 0{,}368$.
Sie fallen subjektiv auf. Das entdeckte Helmar Frank und benannte den Wert „Auffälligkeit“ [Fra69]
Ansonsten ergibt sich immer ein Maximum für H bei Gleichverteilung: $p_i = 1/n$ *mit* $H_{max} = \mathrm{ld}(n) \geq C_A$.
Das entspricht zugleich der Obergrenze für H. Hierfür gilt die Tabelle

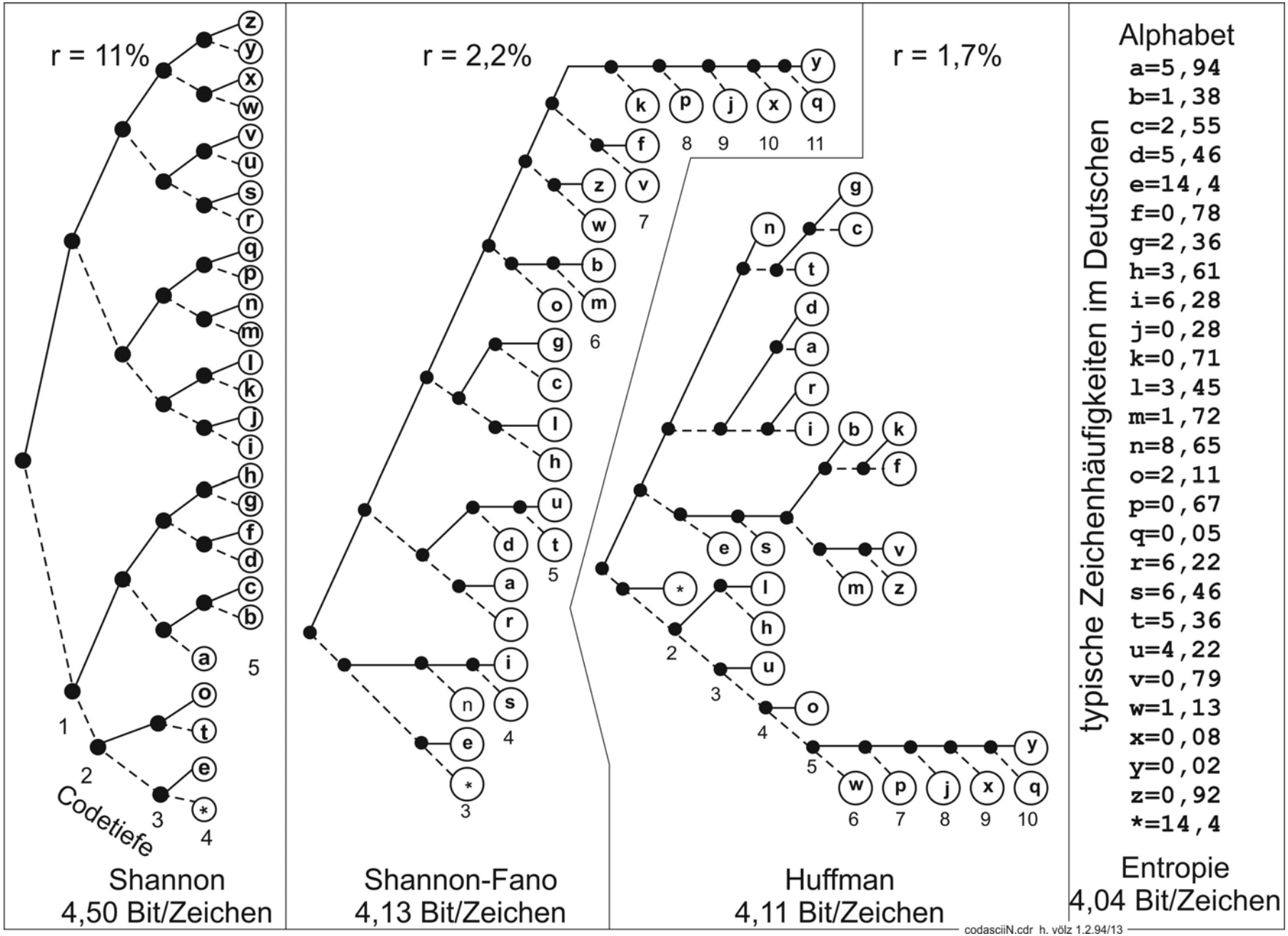


n	2	3	4	8	1024	1048576
$\boldsymbol{H_{max}}$	1	1,585	2	3	10	20

Fano ↔ Huffman

Bei kleiner Zeichenanzahl führen die drei Codierverfahren vielfach zum gleichen Codeaufwand.
Für Unterschiede ist der Codebaum dann nicht tief genug
In der Literatur ist daher üblich, den Vorteil der Huffman-Codierung bei 7 Zeichen zu demonstrieren.
Jedoch für die Grenze $H_{Huffman} < H_{Fano}$ war keine Untersuchung bekannt.
Darauf wies ich bei einem Vortag in der Heinrich-Hertz-Oberschule Berlin 1985 hin.
Kurz darauf lieferten zwei zwei Oberschüler das exakte Ergebnis ab 6 Zeichen [Bal89].

Den Fortschritt der Codierung von Shannon über Fano bis Huffman zeigen die obigen Bilder.
Dabei kann der Eindruck entstehen, dass mit zunehmender **Codetiefe die Codierung** auch effektiver wird.
Das gilt nicht allgemein und wird durch die Codierung der Buchstabenhäufigkeit der deutschen Sprache (nächstes Bild) belegt.


r = 11%
r = 2,2%
r = 1,7%
Codetiefe
Shannon
4,50 Bit/Zeichen
Shannon-Fano
4,13 Bit/Zeichen
Huffman
4,11 Bit/Zeichen
Alphabet
a=5,94
b=1,38
c=2,55
d=5,46
e=14,4
f=0,78
g=2,36
h=3,61
i=6,28
j=0,28
k=0,71
l=3,45
m=1,72
n=8,65
o=2,11
p=0,67
q=0,05
r=6,22
s=6,46
t=5,36
u=4,22
v=0,79
w=1,13
x=0,08
y=0,02
z=0,92
*=14,4
typische Zeichenhäufigkeiten im Deutschen
Entropie
4,04 Bit/Zeichen
codasciiN.cdr h. völz 1.2.94/13

Superzeichen

Die Kodierung mit Kombinationszeichen (s. o.) besitzt auch für unsere Wahrnehmung Vorteile.
Leider gilt aber für die Ziffernhäufigkeiten das Newcombschen Gesetz (gemäß Intervallänge!):

Ziffern	1	2	3	4	5	6	7	8	9
Häufigkeit	30,1	17,6	12,5	9,7	7,9	6,7	5,8	5,1	4,6

Außerdem existiert bei Preisen noch eine die Überhöhung für $p(9)$.

Hiervon abgesehen, hat Cube die Wahrnehmung von Superzeichen (als Kombinationen) untersucht [Cub65].
Ein Text bestehe dazu aus k Wörtern mit je n Zeichen, also insgesamt $m = k \cdot n$ Zeichen.
Für die Einzelentropie eines Wortes gilt dann $H_1 \approx n \cdot \mathrm{ld}(n)$.
Folglich liefern alle Wörter $H_2 \approx m \cdot \mathrm{ld}(n)$.
Zusammengefasst gilt $H_3 \approx k \cdot \mathrm{ld}(\mathrm{k})$. Wegen $k = m/n$ gilt folgt schließlich

$$H_{total} \approx m \cdot \mathrm{ld}(n) + \frac{m}{n} \cdot ld \left(\frac{m}{n} \right).$$

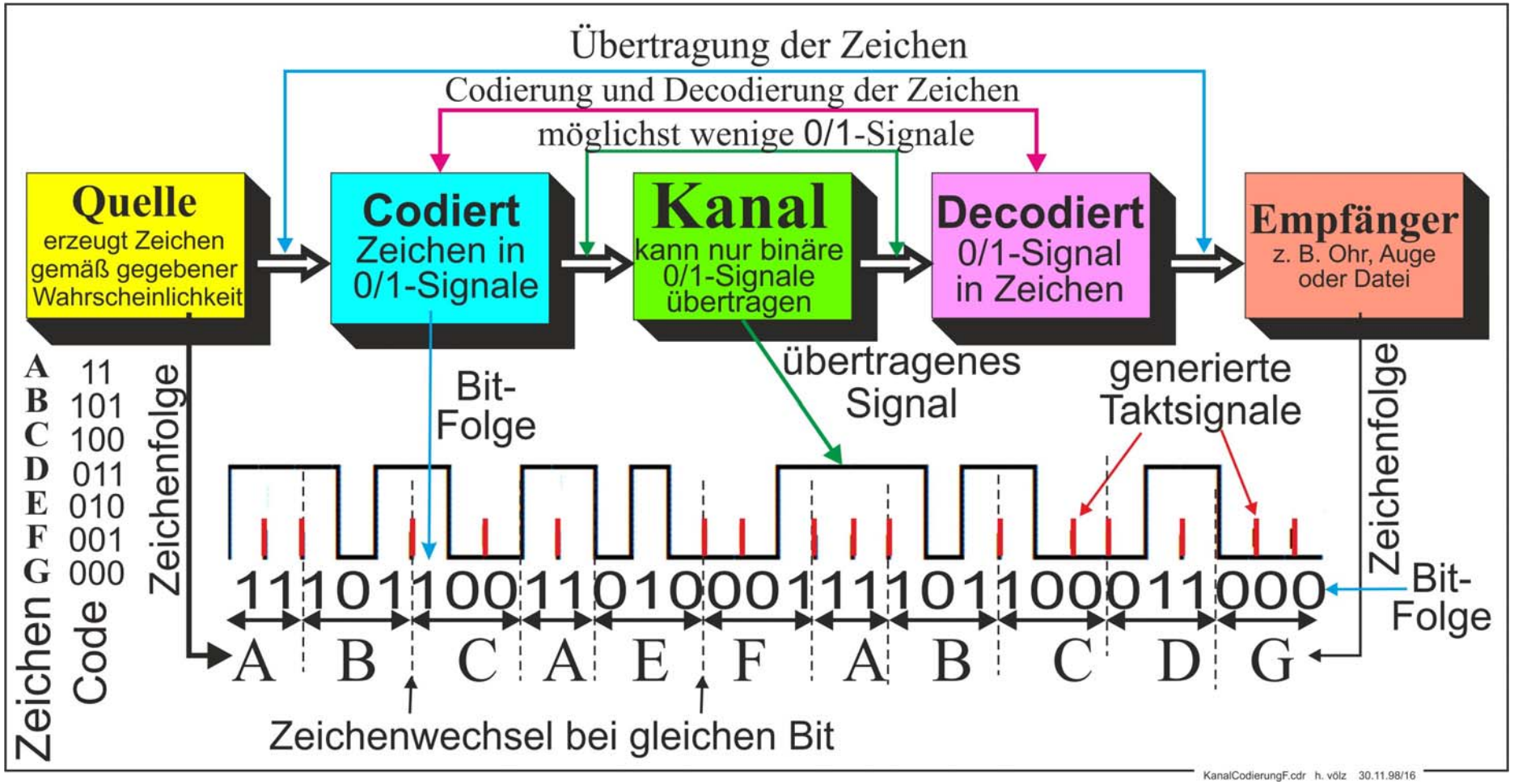
Ihr Minimum liegt bei $m = n \cdot e^{n-1}$. Daraus leiten sich die folgenden Werte ab:

bei maximaler Wortlänge m	5,44≈6	22,2≈23	80,3≈81	273	891	2824	8773
optimale Gruppengröße k	2	3	4	5	6	7	8

So sind die typischen Trennungen bei Telefonnummern z. B. 289 517 091, bei Bankkonten als IBAN z. B. DE77 1018 0000 0916 6311 03 usw. gut erklärbar. Ähnliches gilt auch für Wörter
Auch deshalb sind Silbenwiederholungen (und Ohrwürmer) oft so einprägsam.

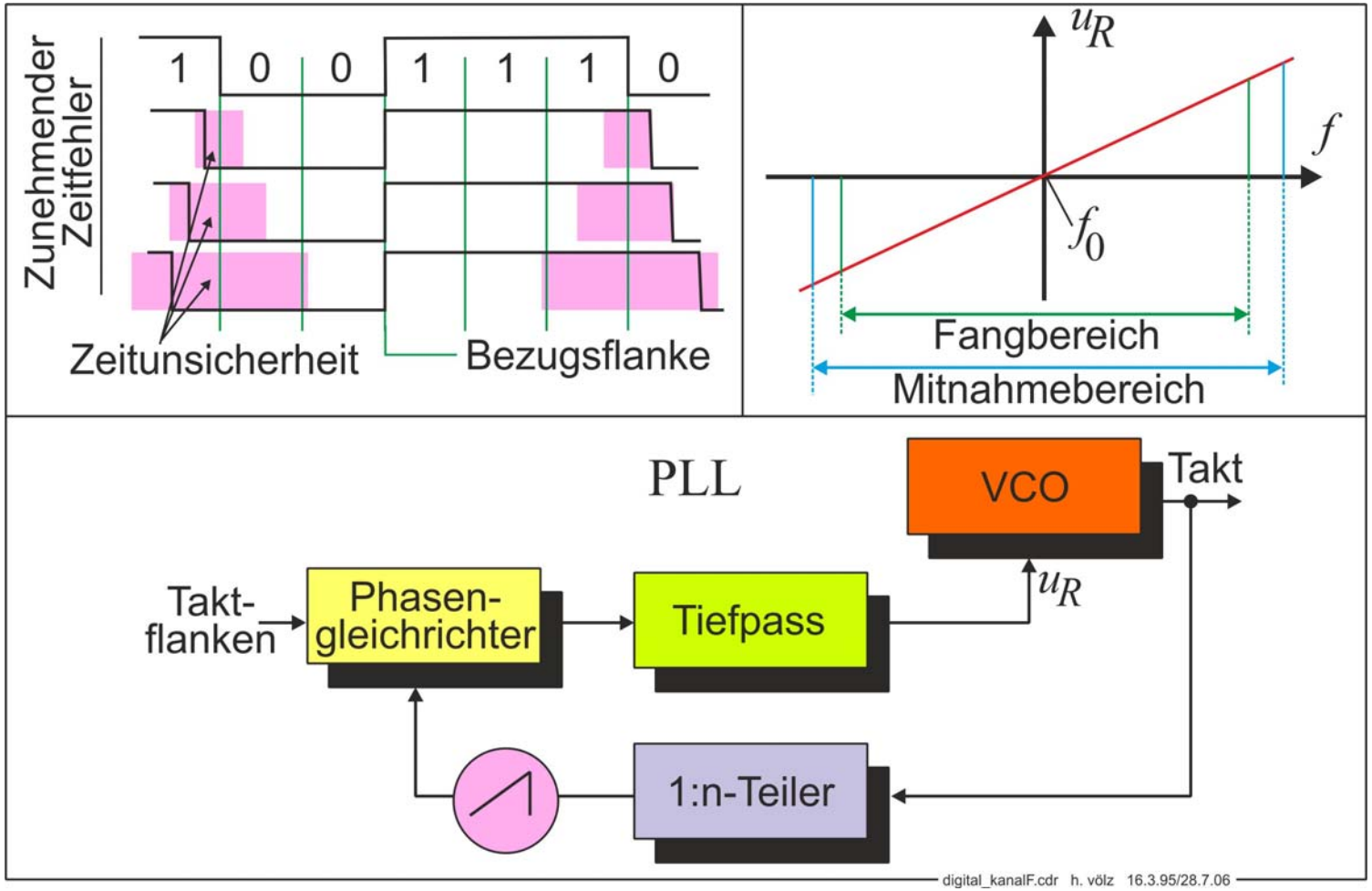
Lauflänge und Takt

Bei der Präfixcodierung können die einzelnen Bitkombinationen ohne Trennzeichen unmittelbar folgen.
Daher können zuweilen mehrere bis viele 0- bzw. 1-Werte unmittelbar aufeinander folgen (Bild).
Im Rechtecksignal sind dann die einzelnen Bit nicht mehr zu erkennen. Es ist erst der Takt zu erzeugen
nur mit ihm lassen sich die einzelnen 0 und 1 eindeutig zurückgewinnen.



Taktgewinnung

ie Taktzurückgewinnung geschieht meist mit einer in der Phase gesteuerten Schaltung
aus PLL (phase locked loop), VCO (.voltage controlled oscillator) und Frequenzteiler
Wegen Zeitfehler muss dafür aber die Lauflänge begrenzt werden
Das ist die maximale Anzahl der unmittelbar aufeinander folgenden 0- und auch 1-Werte



Verschiedenen Wahrscheinlichkeiten

Die Wahrscheinlichkeiten (der Entropieformel) können auf recht unterschiedliche Weise entstehen

1. Sie sind ***objektiv vorhanden:*** a. ***thermisch*** infolge der Wärmebewegung von Teilchen.
b) ***komplexen Ursachen*** z. B.: Katastrophen, Wirbelbildung, Reibung, Wetter z.T. Soziometrie usw.
Dabei sind die Ursachen, ihre Gesetze kaum ausreichend geklärt, sie sind zufällig

2. Mit ***Techniken erzeugt***, z. B. Würfel, Urne, Münzwurf, Kartenspiel, Los, Knobeln.
Anwendungen u. a. bei Lotterie; für Risiken, Spiele und Versicherungen.
Dann sind die Zufallsgesetze und damit Wahrscheinlichkeiten gut und ***im Voraus bekannt***.

3. Simulationen erfolgen mittels ***Rechner (***Zugriff auf die 1/100-tel-Sekunde der Systemuhr).
Es gibt aber auch langperiodische Algorithmen (s. u.) mit Pseudozufall und Zufallstabellen.

4. Zuweilen werden bereits ***vorhandene Werte*** einfach gemessen oder ***abgezählt***.
Dann existieren nur ***Häufigkeiten*** (aposterio Wahrscheinlichkeiten statt sonst apriorie)
Rein formal sollen sei bei großer Anzahl gegen apriori konvergieren (Restfehler immer vorhanden)
Sie ermöglichen aber Abschätzungen.

5. Jeder Mensch hat persönliche Schätzungen für mögliches, künftiges Geschehen.
Das führt zu ***subjektiven*** Wahrscheinlichkeiten (s. Bongard-Weiß-Entropie

6. Die ***bedingten*** Wahrscheinlichkeiten treten erst auf, wenn Voraussetzung mit p_v erfüllt wurden
ihre gesamte (bedingte) Wahrscheinlichkeit beträgt $p_B = p_V \cdot p_{E.}$

7. Ungewöhnlich ist der ***quantenphysikalische*** Zufall (oft ***absolut*** genannt)
Für ihn gibt es grundsätzlich keine ergründbare Ursache. ($\rightarrow \psi$ der Wellenfunktion)

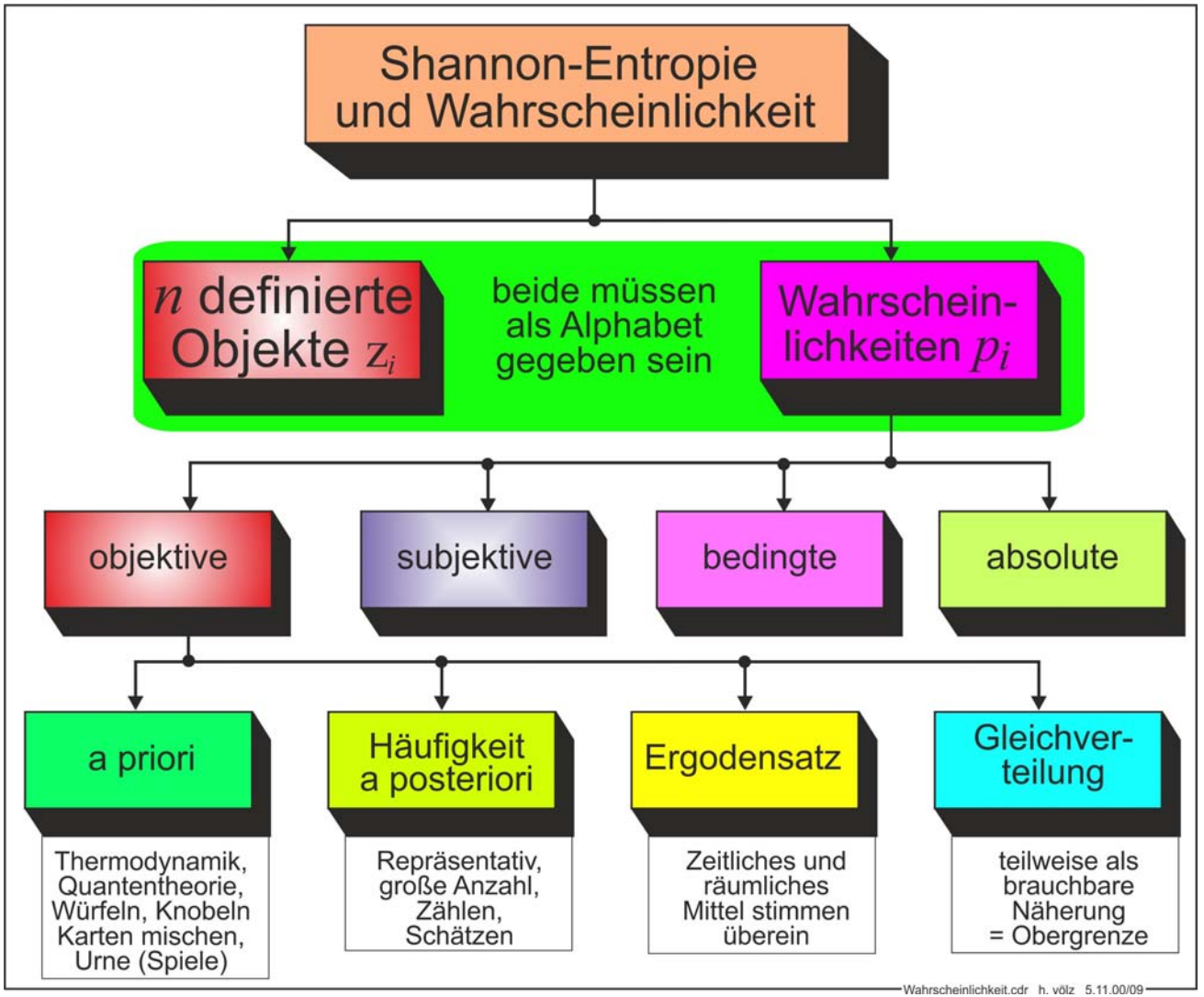
Ergänzungen und Überblick

Ergodensatz (lat. erg Tat, Werk Vorfall) betrifft Bedingungen beim Gewinn von Wahrscheinlichkeiten.
Bei einer ***Urne*** können z. B. die Kugeln nacheinander oder mit einem Mal herausgenommen werden.
Das macht für ihre Wahrscheinlichkeiten keinen Unterschied.
Diese Quelle ist dann ergodisch.

Jedoch bei der ***Sprache*** können sich solche Werte unterscheiden:

Jeder Sprecher ändert immer etwas seinen Wortschatz und seine Wortwahl.

In diesem Fall ist die Quelle nicht ergodisch, Gestrige und heutige Texte sind unterschiedlich.



Von kontinuierlich bis digital

Bereits Shannon beginnt mit diskreten Werten [Sha48], obwohl es damals noch keine Digitaltechnik gab.
Durch sie ist nur die effektive Übertragung mit dem Präfixcode und dazugehörenden Folgerungen möglich.
Heute erfolgen zwar fast alle Signalübertragungen, -speicherungen, -bearbeitungen usw. digital.

Ganz im Gegensatz dazu funktionieren (immer noch) fast alle Sensoren und Aktoren kontinuierlich.
Das gilt auch für die menschlichen Sinne und Muskeln.

Daher müssen Signale oft zwischen kontinuierlich und digital gewandelt werden.
Aus dieser Sicht ist die Digitaltechnik eigentlich ein „Umweg“ mit zwei Wandlungen (ADU und DAU).

Doch es gibt viele Gründe für die Digitaltechnik: verlustloses Kopieren, Fehlerkorrektur und Rechentechnik
Deshalb sollen hier alle diesbezüglichen Begriffe, einschließlich analog und digital erklärt werden [Völ79].

Analog und Analogie

Grie. logos Vernunft, lateinisch ana auf, wieder, aufwärts, nach oben.
analogia mit der Vernunft übereinstimmend, Gleichmäßigkeit auch Entsprechung, Ähnlichkeit, Gleichwertigkeit, Übereinstimmung.
Analog wird in einzelnen Anwendungen unterschiedlich benutzt (betrifft aber immer einen **Vergleich**):

- ***Platon***: menschliche Seele dreigeteilt: Vernunft, Wille, Begierden.
  Gerechter Mensch kontrolliert seine Begierden durch Vernunft und mit Unterstützung des Willens.
  Analogie: Dreiständeaufbau des Staates: Erleuchteter Philosoph/König regiert Gesellschaft mit Kriegern.
- ***Biologie/Medizin***: Analoge Organe (Morphologie/Struktur), Augen: Wirbeltier, Tintenfisch und Insekt.
- ***Psychologie***: Analogien grundlegend für Denken, z. B. bei Maxwell: Wasserströmung ⇔ elektromagnetische Felder + Kekule: Affen ⇔ Benzol-Ring.
- ***Literatur***: Nutzung für Fabeln, Parabeln, Märchen, Gleichnisse usw.
- ***Kybernetik***: Technisch Systeme analog lebenden Organismen. Achtung! Eine angenommene konsequente Analogie zwischen Technik und Mensch (harte KI) führt fast immer zu Fehlschlüssen.
- ***Bionik,*** Analogie der Haut von Delphinen für die Optimierung von Schiffsrümpfen, Lotus-Effekt bzgl. nicht auftretender Verschmutzungen, usw.
- ***Technik***: Analogrechner (Differential-Gl.), elektromechanische Analogien, Analoguhr: drehende Zeiger.
- ***Logik***: Induktive Beweisführung: Größen in einigen Fällen ähnlich sind, dann eventuell auch in anderen.
- ***Rechtsprechung***: Einen juristischen Tatbestand auf einen etwa wesensgleichen übertragen ist in Deutschland unzulässig.

Kontinuierlich

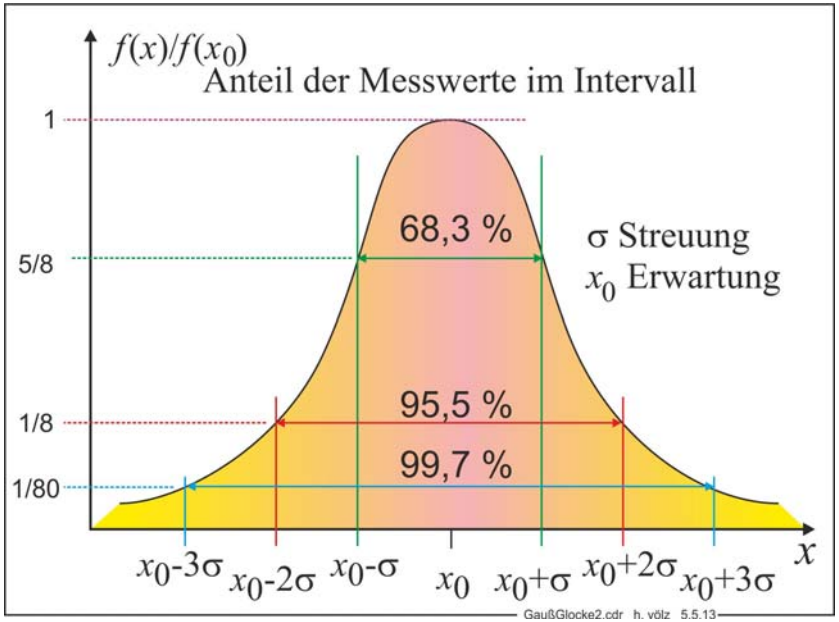
Lat. continens, continuus: zusammenhängend, angrenzend an, unmittelbar folgend, ununterbrochen, jemand zunächst stehend.

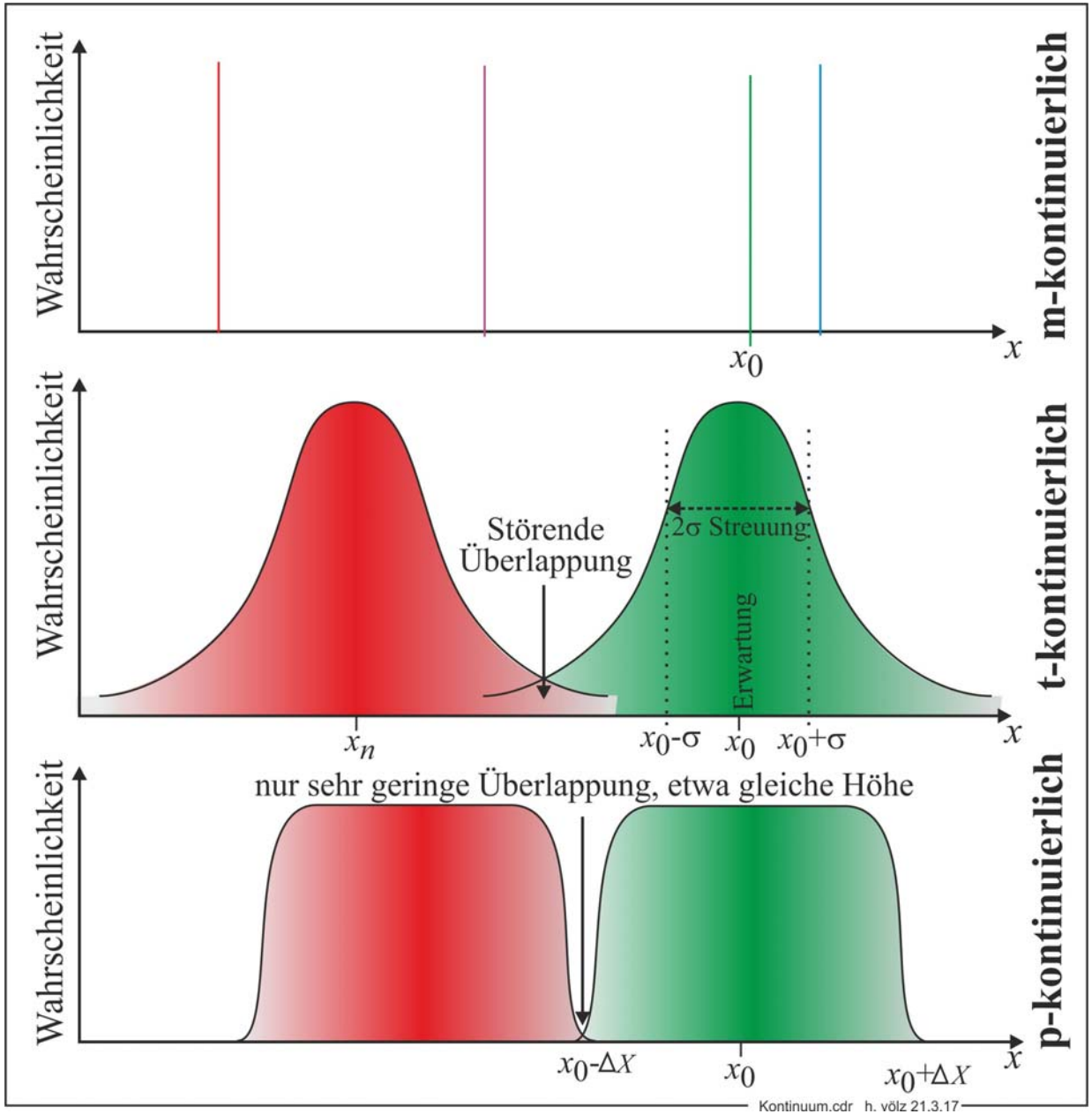
Continuare ≈ aneinanderfügen, verbinden, fortsetzen verlängern, gleich darauf, ohne weiteres.

Contingere ≈ berühren, kosten, streuen, jemandem nahe sein, beeinflussen.

- ***Umgangssprache*** etwa beharrlich, ununterbrochen, ständig, verwandt mit stet, stets von stehen; Gegenteil ist unstet (s. u.).
- ***Mathematik → m-kontinuierlich:*** Kontinuum der reellen Zahlen zwischen zwei Zahlen gibt es immer eine weitere. Reelle Zahlen mit unendlich vielen Ziffern.
  Es sind beliebig kleine Grenzwerte möglich: lim $f(x)$ für $x \to 0$;
  Es besteht Verwandtschaft zu mathematisch **stetig** und betrifft auch die Umgebung von Grenzwerten.
  u. a. den Grenzwert lim $f(x)$ für $x \to 0$. Bietet so die Möglichkeit zur Differentialrechnung.
- ***Technik-Signale:*** *t-kontinierlich* (ermöglicht zwar beliebige Werte bei Zeit und Amplitude (Energie). Sie sind aber immer mit Fehlern behaftet.
  p-kontinuierlich (Physiologie, Prozent) nahezu **rechteckiger Fehlerbereich**, subjektiv ununterscheidbar
- ***Physik***: Als Kontinuumsmechanik; berücksichtigt vereinfachend nicht die Mikrostruktur der Materie, insbesondere die Teilchen.

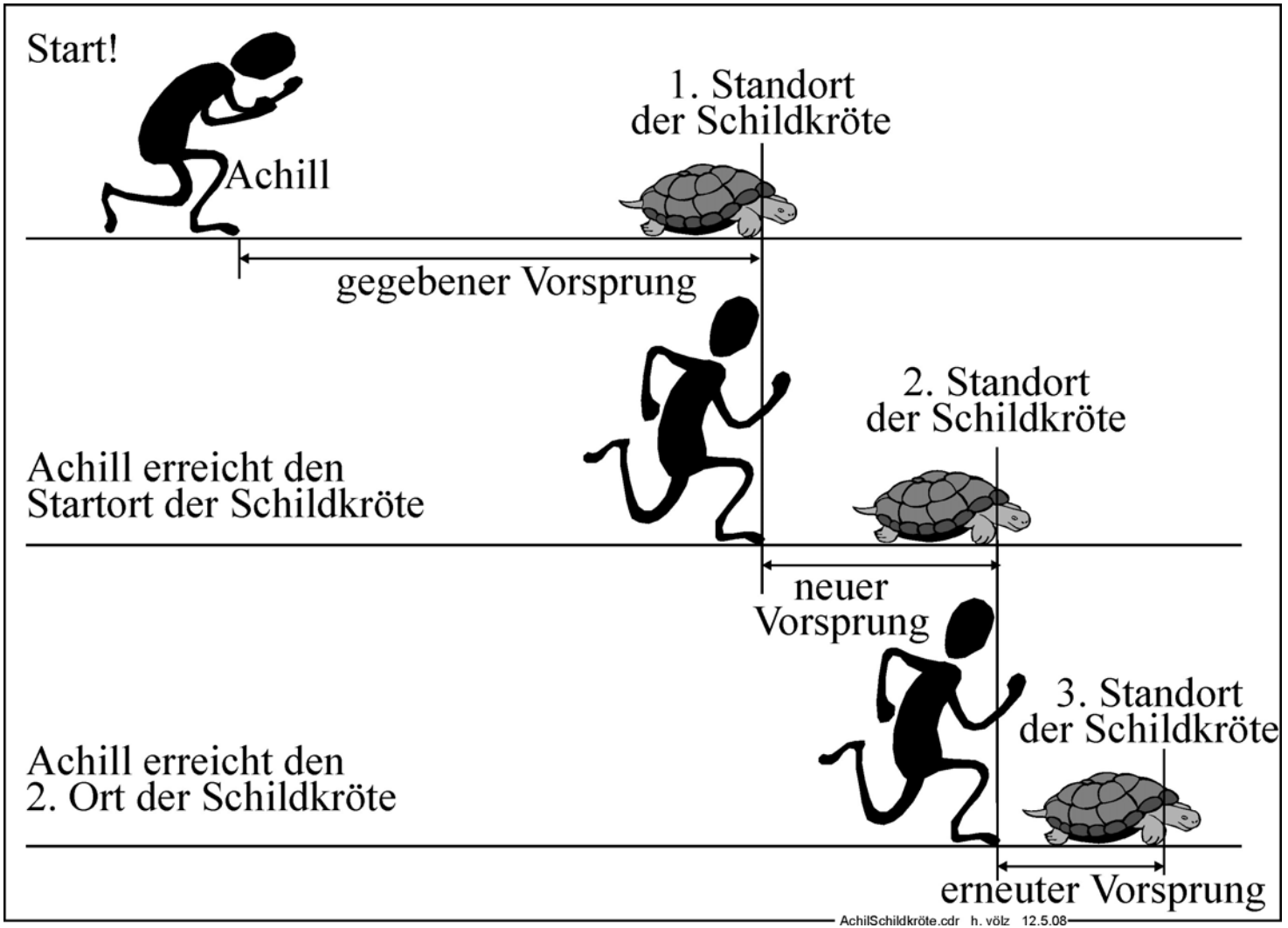
stetig, deutsch von stehen, verwandt mit stet, stets; das Gegenteil von unstet; umgangssprachlich: beharrlich, ununterbrochen, ständig.


f(x)/f(x0)
Anteil der Messwerte im Intervall
1
5/8
1/8
1/80
68,3 %
95,5 %
99,7 %
σ Streuung
x0 Erwartung
x0-3σ
x0-2σ
x0-σ
x0
x0+σ
x0+2σ
x0+3σ
x
GaußGlocke2.cdr h. völz 5.5.13


Wahrscheinlichkeit
x0
x
m-kontinuierlich
Wahrscheinlichkeit
Störende
Überlappung
2σ Streuung
Erwartung
xn
x0-σ
x0
x0+σ
x
t-kontinuierlich
Wahrscheinlichkeit
nur sehr geringe Überlappung, etwa gleiche Höhe
x0-ΔX
x0
x0+ΔX
x
p-kontinuierlich
Kontinuum.cdr h. völz 21.3.17

Einführung des Kontinuums

Sie erfolgte durch Aristoteles in Bezug auf die Paradoxien von Zenon (=Xenon). Danach könnte z. B. der sehr schnelle Achilles die äußerst langsame Schildkröte nicht einholen, geschweige denn überholen.



Diskret ↔ digital

Diskret: Lat. discretion Unterscheidungsvermögen, Urteil und Entscheid.
Discretus abgesondert, getrennt; discernere: scheiden, trennen, unterscheiden, beurteilen, entscheiden:

- ***Umgangssprachlich***: taktvoll, rücksichtsvoll, zurückhaltend, unauffällig, unaufdringlich, vertrauensvoll, geheim, verschwiegen.
- ***Mathematisch***: etwas zerfällt oder gliedert sich in einzelne (*abzählbare* viele) Punkte oder Elemente.
- ***Physikalisch*** betrifft es Größen, die sich nur in endlichen Schrittweiten ändern (vgl. unten spezieller Quant).
- ***Signal***: besitzt nur endlich viele t-kontinuierliche Werte mit Toleranzbereichen, die eigentlich streng gegeneinander abgegrenzt sein sollten.
  Wegen weitreichende „Ausläufe“ (s. o. besteht immer die Gefahr einer Überschneidung, die dann zu Fehlern führt. Einzelne diskreten Werten können bei Z-Information benut werden

Digital: lat. digitus Finger; zählen, ziffernmäßig, in Zahlen angeben. Es betrifft:

- ***England***: Eine alte Maßeinheit etwa Fingerbreite mit 18,5 mm
- ***Biologie***: Eine Blütenpflanze heißt digitalis, deutsch Fingerhut
- ***Signalwerte*** mit einer Zahlendarstellung (-abbildung), z. T. unterschiedliche Zahlenbasen:
  binär = 2, oktal = 8, dezimal = 10, hexadezimal = 16.
  Anzahl der Abstufungen (Speicherzustände) kann infolge Codierungen abweichen:
  Z. B.: *dual* bei zwei ***physikalische Zuständen*** oder BCD für binär codiertes Dezimal.
  Echt digitale Signale sind bezüglich ***Amplitude und Zeit*** diskret.

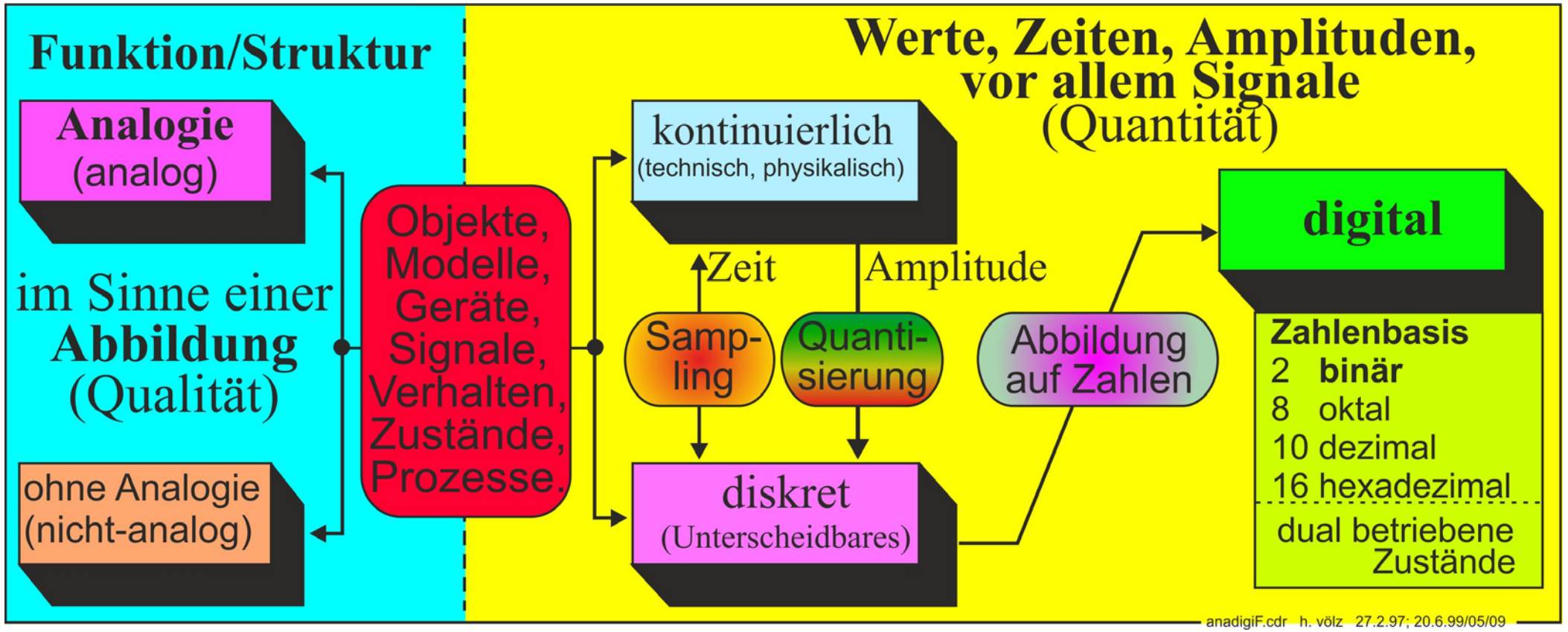
Quantisierung und Bit

Quantisiert: Lat. quantitas Größe, Anzahl; quantum wie viel, so viel wie, inwieweit, irgendwie. usw.:

- ***Philosophie***: Zusammenhang und Unterschied von *Quantität* (≈ Menge) ↔ *Qualität* (≈ Güte).
- ***Physik***: Das Quant wurde 1900 durch Planck eingeführt; Beginn der Quantentheorie.
- ***Messwerte*** = t-kontinuierlich bzgl. der Ausprägung, qualitativ für die Maßeinheit (m, kg, s usw.).
- ***Technisch*** bedeutet quantisieren diskrete Werte erzeugen.
  Quantisierte Signale besitzen diskrete Amplituden (-stufen) und/oder diskrete Zeitpunkte (Takte).

Bit von binary digit, englisch digit Zahl, Ziffer, Finger.
wurde 1948 von J. W. Tukey eingeführt, besitzt nur einen der beiden Werte 0 oder 1, wahr oder falsch
entspricht der Beantwortung einer Ja/Nein-Frage
Zusammenhang oft (eigentlich nicht ganz richtig) von „beseitigter Unsicherheit“.

Zusammenhänge



Varianten der Diskretisierung und Digitalisierung

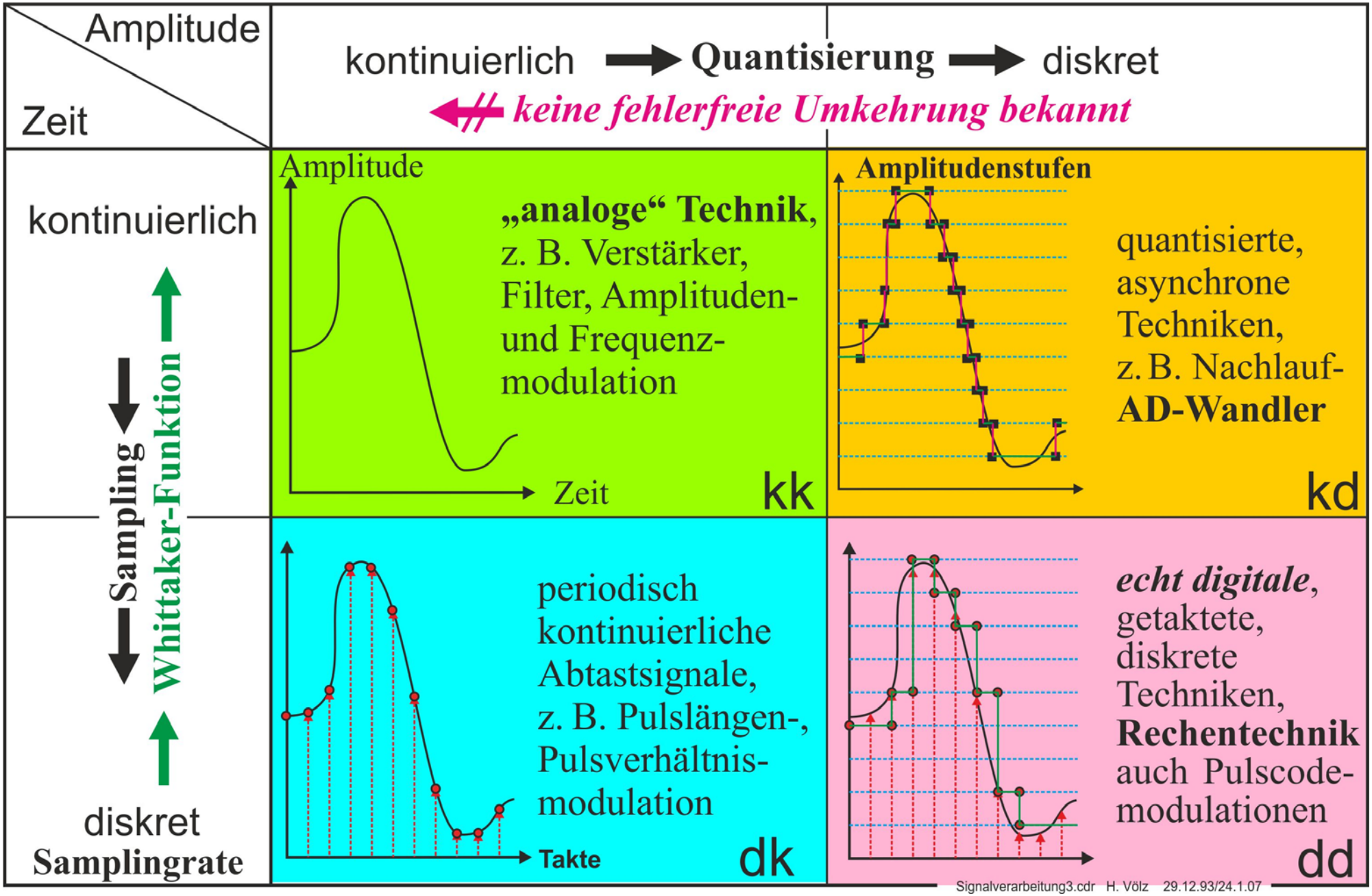
Sowohl die Zeit als auch die Amplitude (Energie) können sowohl einzeln als auch gemeinsam.
t-kontinuierlich, diskret und/oder digital sein. Daher existieren vier Signalvarianten (Bild).
Vereinfacht mit k für kontinuierlich und d für diskret in der Abfolge Zeit - Amplitude benannt:

kk Zeit- und Amplitude kontinuierlich = kontinuierliche (*analoge*) Technik, z. B. klassische Verstärker, Filter, Amplituden- und Frequenzmodulation.
Zu jedem t-kontinuierlichen Zeitpunkt existiert ein t-kontinuierlicher Amplitudenwert.

dk Neue t-kontinuierliche Amplitudenwerte treten nur zu diskret festgelegten Zeiten auf.
Meist zu festgelegten Taktzeiten mittels sample hold (s. u.) ausgewählt und gespeichert.
Bei richtig gewählter Taktfrequenz ist der Schritt im Prinzip rückgängig zu machen
Typisch hierfür sind Abtastsignale sowie Pulslängen- und Pulsverhältnismodulation.

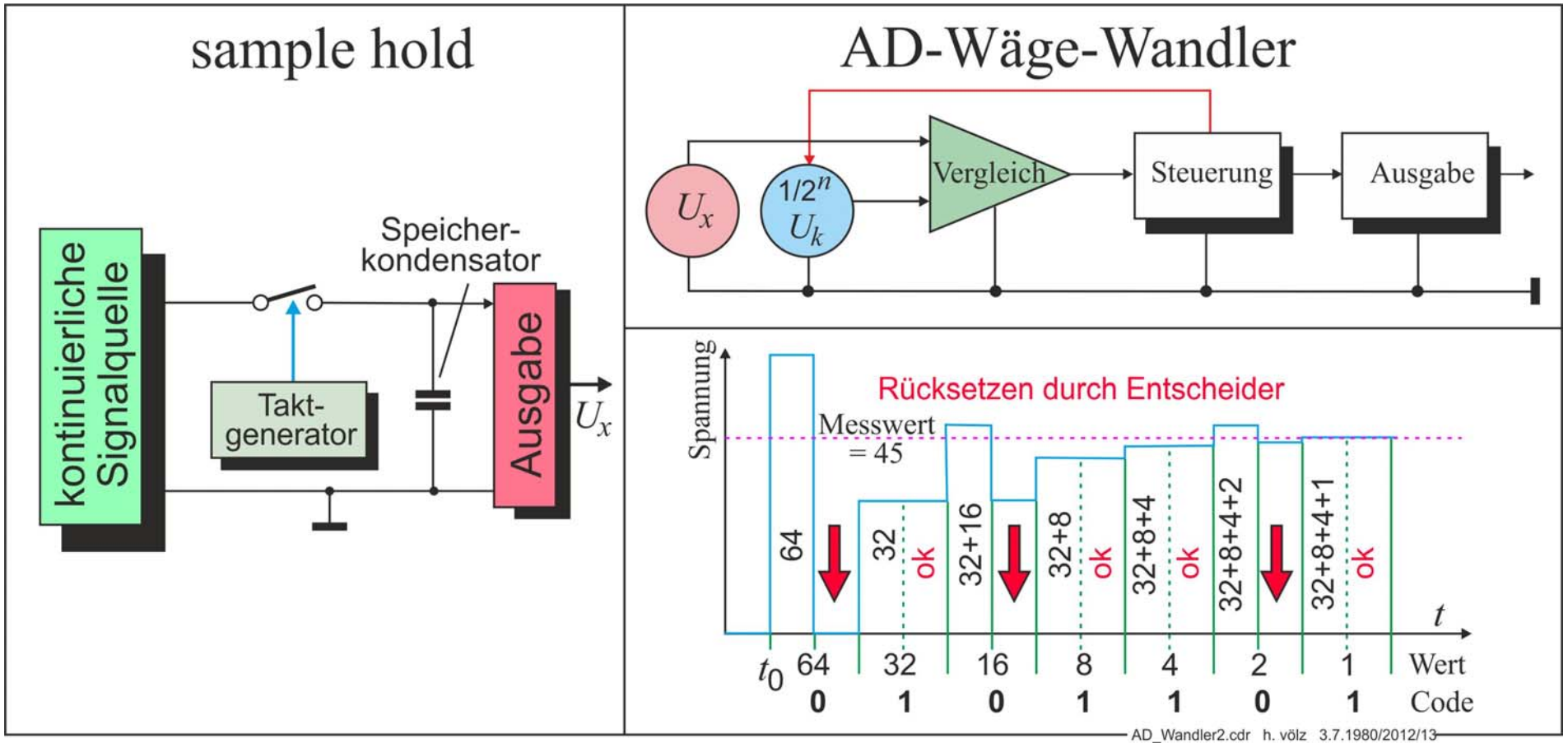
kd Es gibt nur diskrete, t-kontinierliche Amplitudenwerte.
Wechsel erfolgt beim Erreichen ener neuen t-kontinuierliche Amplitude
nimmt dann das Ausgangssignal sprunghaft an. Ist eine reine Amplituden-Diskretisierung
Es gibt kaum Anwendungen.

dd Sprung zum neuen diskreten Amplitudenwert kann nur zu einer diskreten Taktzeit erfolgen.
Nur hier liegt „echte“ Digitalisierung und damit ein echtes digitales Signal vor.
Beschreibung erfordert zwei diskrete Zahlen
Wichtig für Rechentechnik; Außerdem bei Pulscodemodulation und ähnlichen Verfahren.
Ungenauigkeiten (Fehler) der Amplitude Δx (Energie ΔE) und der Zeit Δt bedingen sich gegenseitig
Produkt muss zumindest die Heisenberg-Unschärfe $\Delta x \cdot \Delta t \geq h$ erfüllen (h = Planck-Konstante).


Amplitude
Zeit
kontinuierlich
Quantisierung
diskret
keine fehlerfreie Umkehrung bekannt
kontinuierlich
Sampling
Whittaker-Funktion
diskret
Samplingrate
Amplitude
Zeit
„analoge“ Technik,
z. B. Verstärker,
Filter, Amplituden-
und Frequenz-
modulation
kk
Amplitudenstufen
quantisierte,
asynchrone
Techniken,
z. B. Nachlauf-
AD-Wandler
kd
Takte
periodisch
kontinuierliche
Abtastsignale,
z. B. Pulslängen-,
Pulsverhältnis-
modulation
dk
echt digitale,
getaktete,
diskrete
Techniken,
Rechentechnik
auch Pulscode-
modulationen
dd
Signalverarbeitung3.cdr H. Völz 29.12.93/24.1.07

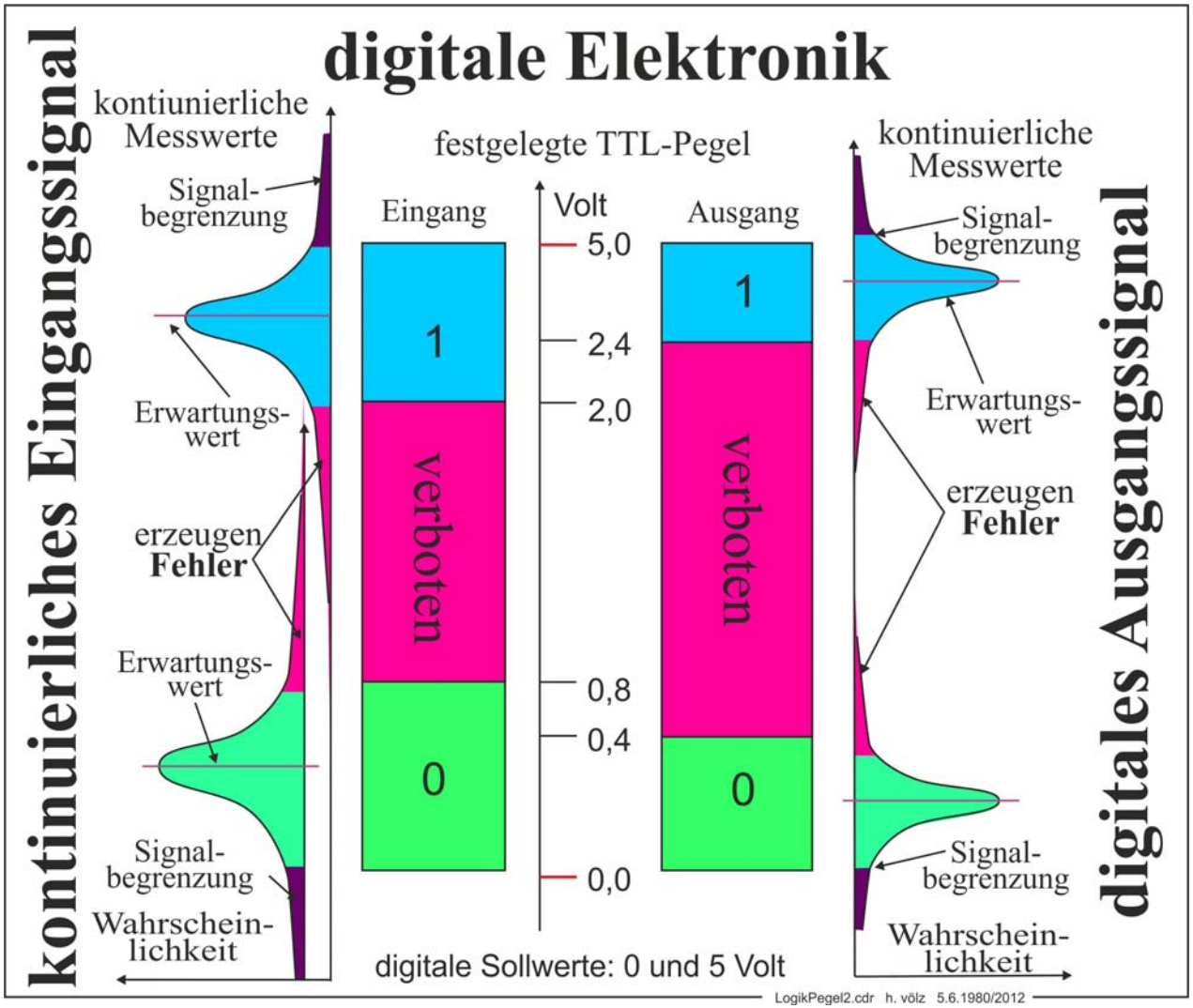
Von kontinuierlich nach diskret (digital)

Für die Zeit-Diskretisierung genügt ein einfacher streng periodischer Zeittakt.
Für die Amplituden-Diskretisierung sind dagegen recht komplexe Schaltungen erforderlich.
Sie heißen Analog-Digital-Wandler (ADU Analog-Digital-Umsetzer). Es gibt etwa 3 Variuanten. [Völ89]
Der Begriff ist eigentlich falsch: es werden fast immer t-kontinuierliche Werte in diskrete umgesetzt.
Hier wird nur der Wägewandler kurz erklärt. Sein Prinzip ähnelt Balkenwaage (immer halbe Größe ±)



Vorbereitung der AD-Wandlung

Sie ist deutlich einfacher. Mit der sample-hold wird in dem Kondensator ein Wert für Taktlänge gespeichert.
Infolge verschiedener Einflüsse treten in beide Richtungen immer Fehler auf (t-kontinuierlich).
Das kann technisch durch die Toleranzbereiche digitaler Werte berücksichtigt werden (Beispiel TTL-Pegel).



Grenzen der Digitalisierung

Im Gegensatz zur AD-Wandlung betrifft Digitalisierung die zeitliche Abfolge von Daten.
Die Rückwandlung in die ursprünglichen kontinuierlichen Signalfolgen soll möglichst fehlerarm sein.
Das verlangt spezielle mathematische Zusammenhänge.
Besonders einfach ist das für alle Geraden mit der Gleichung $y = a + b \cdot x$.
Sie gehen durch den Punkt $y = a$ und $x = 0$ und besitzen die Steigung b.
Die Gleichung $(x - x_0)^2 + (y - y_0)^2 = r^2$ betrifft ähnlich alle Kreise zum Mittelpunkt x_0, y_0 mit dem Radius r.
Für beliebige Signalverläufe ist das viel schwieriger: Es müssen Zeit und Amplitude berücksichtigt werden.
Dann erfordert z. B. Potenzreihen, orthogonale Funktionen cos (n φ) + sin (n φ), Tschebyscheff-Polynome).
Experimentelle Untersuchungen hierzu hatte bereits Küpfmüller um 1924 durchgeführt.
Er fand: Die Einschwingzeit ΔT eines Rechteckimpuls ist grob umgekehrt zur Bandbreite B: $\Delta T \approx 1/2B$.
Um 1930 folgten dann ähnliche Untersuchungen bei Pulsmodulationen von Nyquist.
1933 fand Kotelnikow ähnliche Ergebnisse für eine Signal-Abtastung, waren aber nur in Russland bekannt
Shannon zeigte um 1940, dass für den Probenabstand (sample) bei einer Digitalisierung gelten muss

$$\Delta T \geq \frac{1}{2 \cdot B}.$$

Im Vergleich mit der Quantentheorie (Kapitel 8) ergibt sich dabei eine Analogie zur Heisenberg-Unschärfe.
Dort gilt für das Produkt der Fehler von zwei physikalisch konjugierten Größen, z. B. Zeit Δt und Energie ΔE mit der Planck-Konstanten $\Delta t \cdot \Delta E \geq h/2$. Mit der Photonen-Energie $\Delta E = h \cdot \nu$ folgt daraus

$$\Delta t \geq \frac{1}{2 \cdot \nu}.$$

Whittacker-Funktion

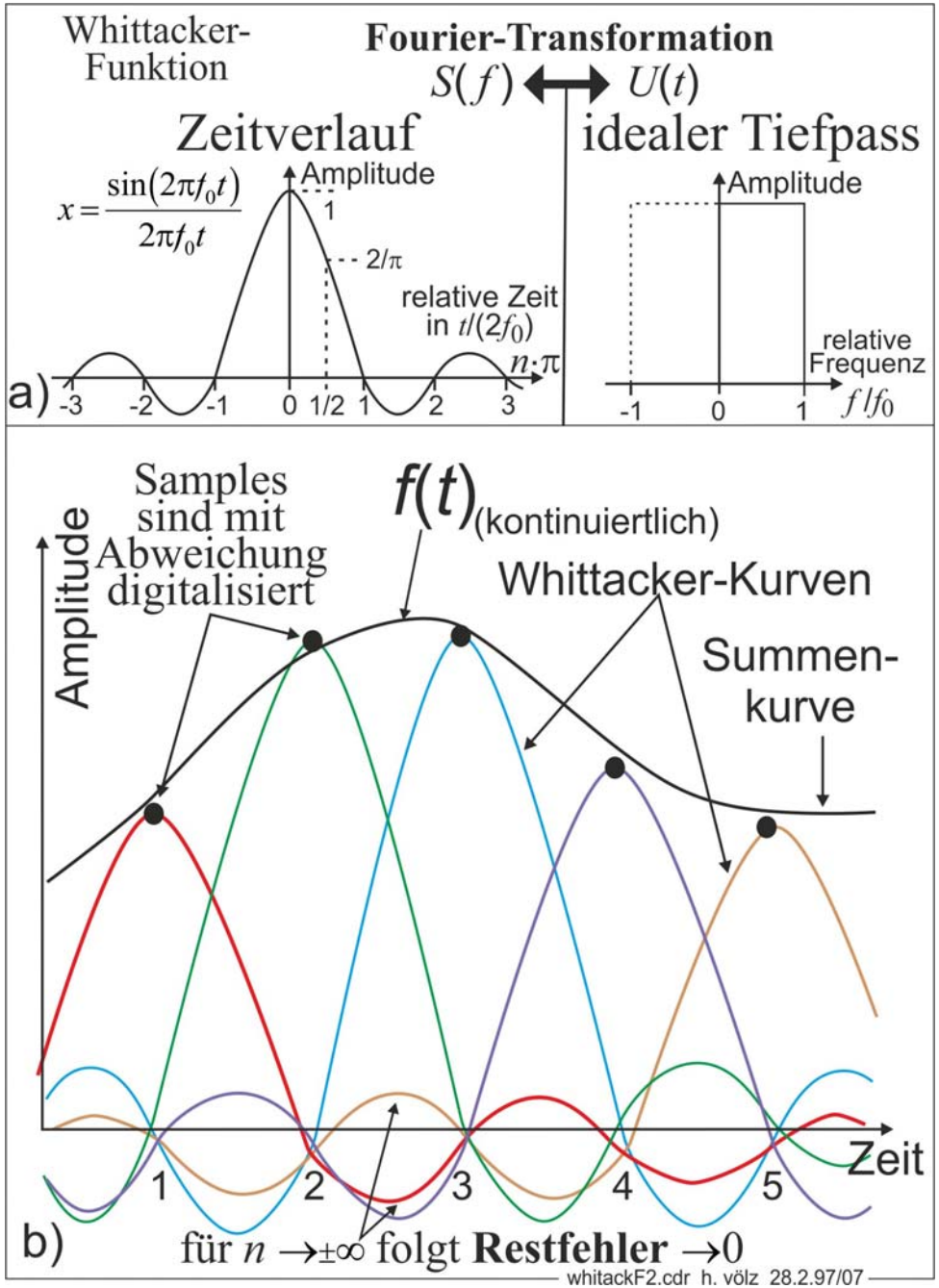
= Spaltfunktion bei Tonband und Tonfilm

Nach Festlegung des minimalen Probenabstandes ist noch die mathematische Beziehung zu finden:
Hierfür wählte Shannon die Whittacker-Funktion:

$$x = \frac{\sin(\alpha)}{\alpha} = \mathrm{Si}(\alpha)\,.$$

Ihr Verlauf besitzt einen Zusammenhang mit dem idealen Tiefpass über die Fourier-Transformation.

Die Überlagerung von Whittacker-Funktion zu den Zeitpunkten lässt exakt das originale kontinuierliche Signal wieder erzeugen.



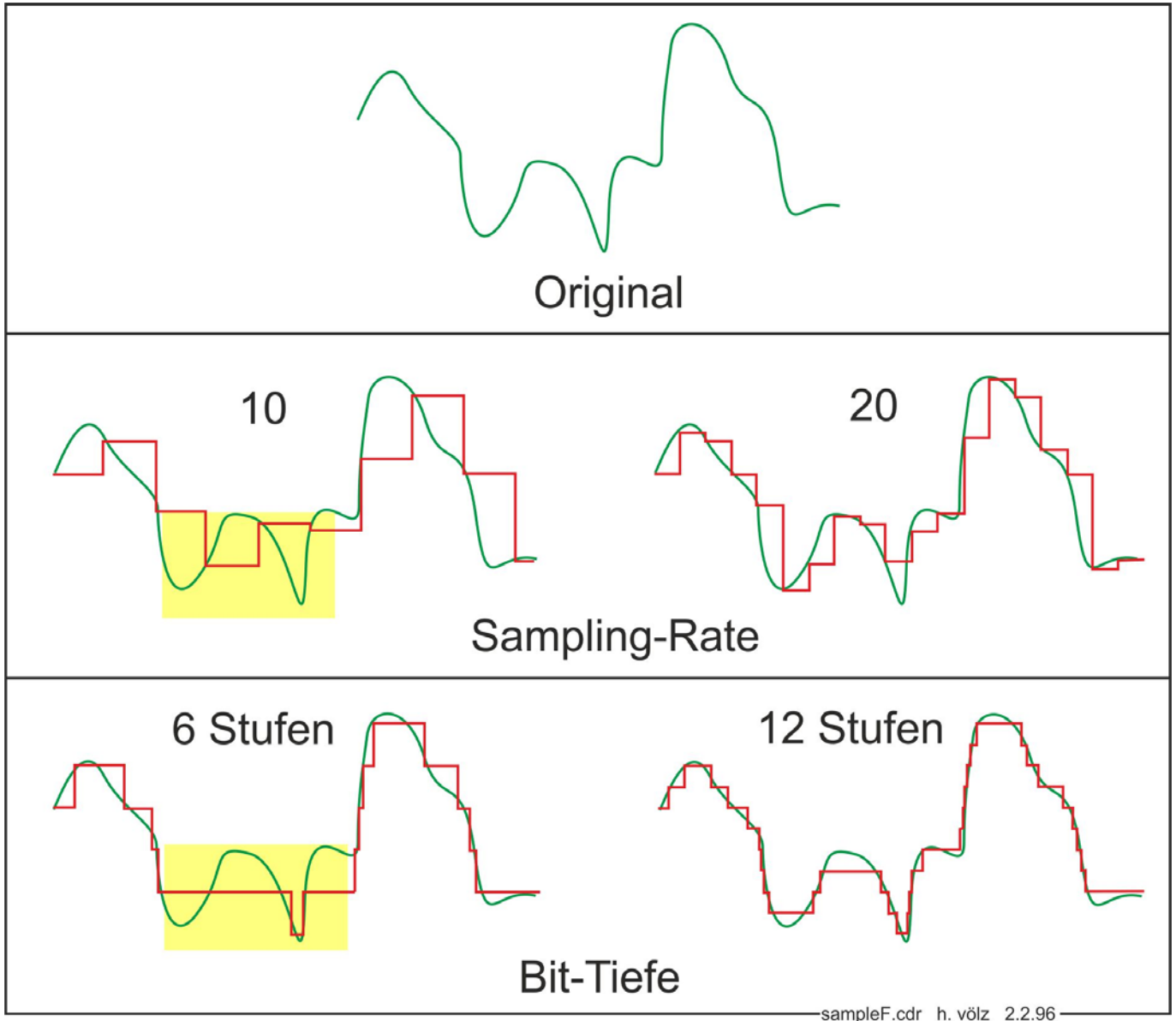
Die Fehler der Rückwandlung

Leider sind die Forderungen für eine exakte Rückgewinnung technisch nur teilweise erfüllbar.
Insbesondere treten folgende **Fehler** auf:

- Um ein exaktes Si(x) zu erzeugen, ist (gemäß der Fourier-Transformation) ein ideal rechteckförmiger **Tiefpass** mit der Grenzfrequenz B und ohne Phasenfehler erforderlich.
  Er ist aber selbst bei extrem großem Aufwand bestenfalls annähernd herzustellen.
  Alle technisch realen Tiefpässe haben keine ideal steile Flanke und besitzen beachtliche Phasenfehler.
- Die Digitalisierung der Amplituden der Samples muss auf eine **endliche Bit-Tiefe** quantisiert werden.
  Dadurch gibt es unerwünschte Änderungen für die Höhe der Nebenmaxima mit Restfehler.
- Es müssten **alle Samples** $-\infty \le t \le +\infty$ überlagert werden, nur dann verschwinden die Restfehler
  Praktisch stehen se aber nur eine endliche Zeit rückwärts und nicht für die Zukunft zur Verfügung.
- Bei der Wiedergabe müssten alle Samples δ-**Impulse** mit $\mathrm{d}\tau \to 0$ auftreten.
  Infolge der Bandbreite des Kanals werden sie aber immer etwas „verschmiert“.
- Die Samples müssen zum exakt korrekten Zeitpunkt wiedergegeben werden,
  infolge von Störungen sind aber gewisse **Zeitfehler** unvermeidbar.

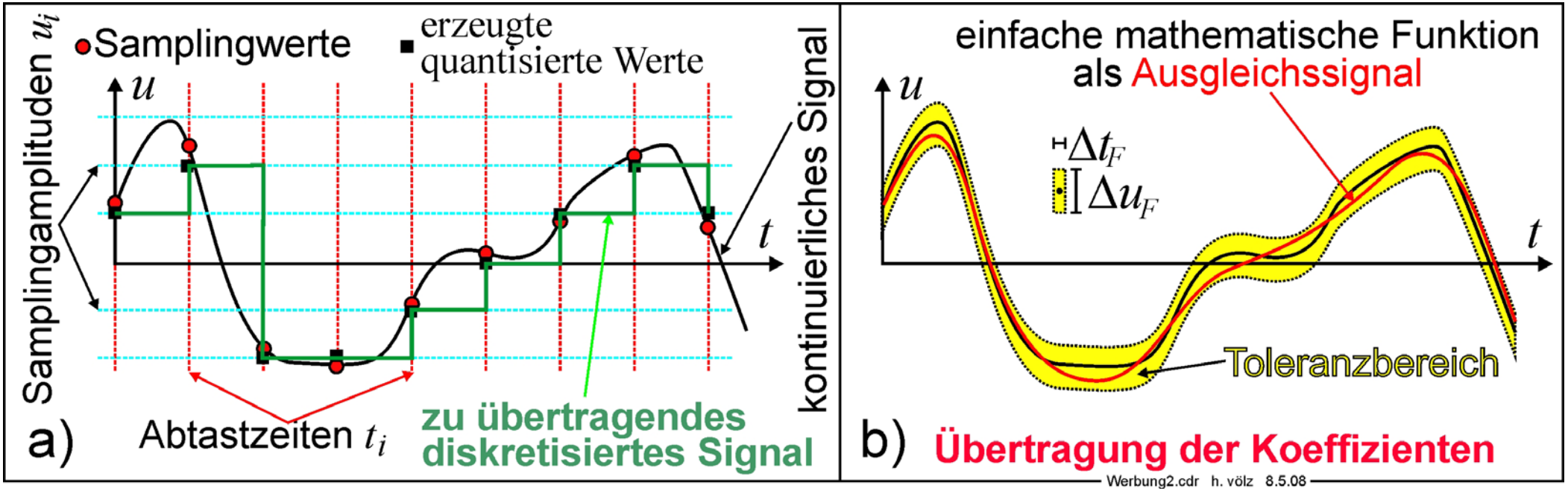
Trotz dieser immer vorhandenen Mängel sind meist die zurück gewonnenen Signale noch gut bis sehr gut.
Für Schallaufzeichnungen bewirken Restfehler ein sehr unangenehmes **Sampling-Geräusch**
Bei den meisten Audio-CDs wird es durch ein **6 dB** stärkeres **thermisches Rauschen** verdeckt.

Qualität und Samplingrate



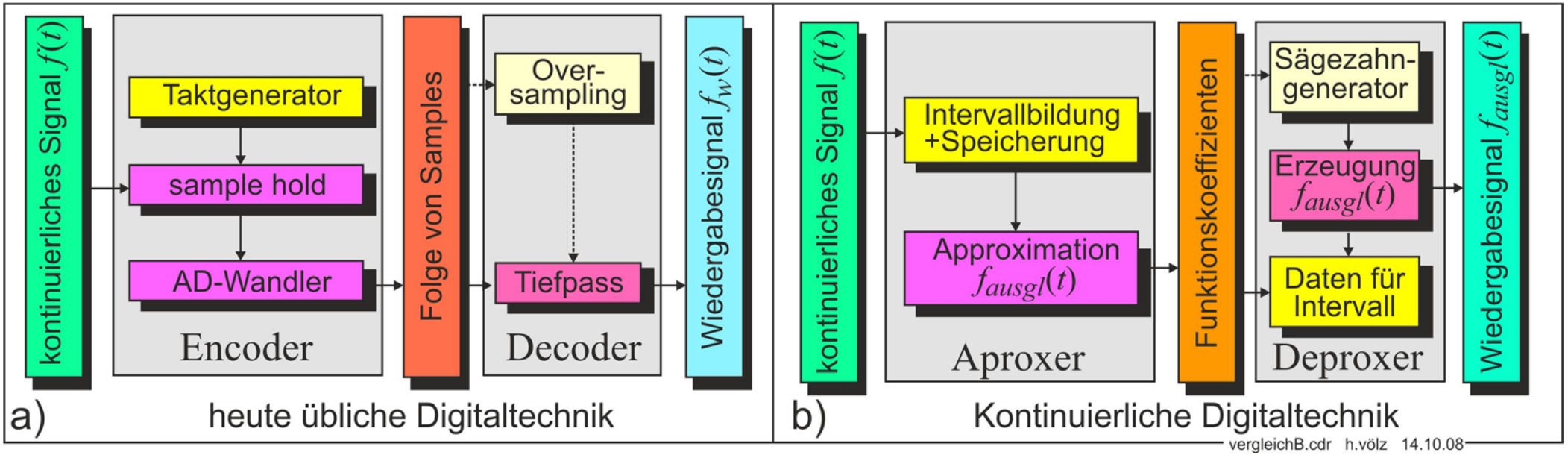
Kontinuierliche Digitaltechnik 1

Warum gibt es keine fehlerfreie Rücknahme der Amplituden-Quantisierung ähnlich wie bei der Zeit?
Zeitverläufe $u = f(t)$ besitzen zu jeder Zeit t nur einen einzelnen Amplitudenwert u (Rückrechnung möglich).
Zur einer Amplitude u kann es aber gemäß $x = f_2(u)$ mehrere t geben.
2007 fand ich nach jahrelangem Suchen einen Umweg, der zugleich das Sampling-Geräusch ausschaltet
Er benötigt sogar weniger zu übertragenden digitalen Werten
Um das t-kontinuierliche Signal wird ein nicht wahrnehmbarer Toleranzbereich gelegt (p-kontinuieröich)
Innerhalb dieses Bereiches wird dann eine möglichst einfache Reihe als Ausgleichssignal approximiert
Für die digitale Übertragung genügen dann die Koeffizienten[Völ08].

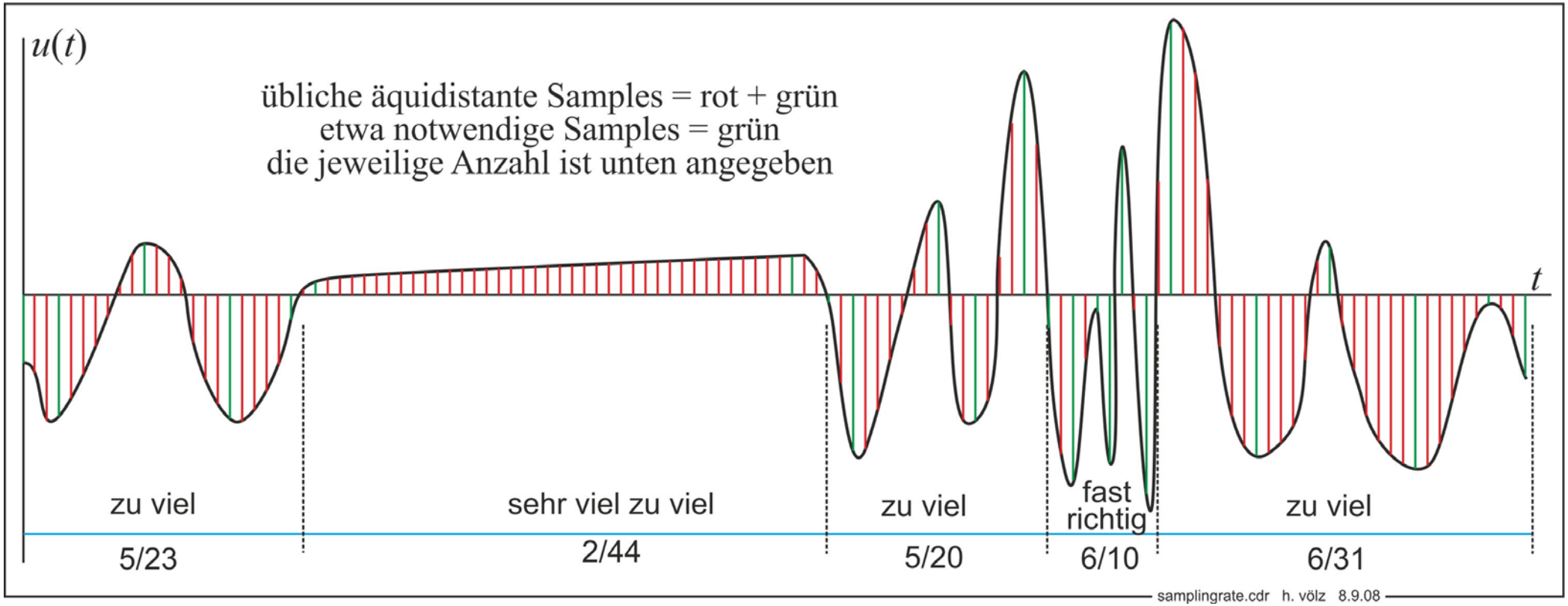


Kontinuierliche Digitaltechnik 2

Es gibt allerdings teilweise Probleme für die Wahl der Teilintervalle und deren Anfügung. Doch auch hierfür sind inzwischen Lösungen mit Gauß-Lobatto-Knoten gefunden.



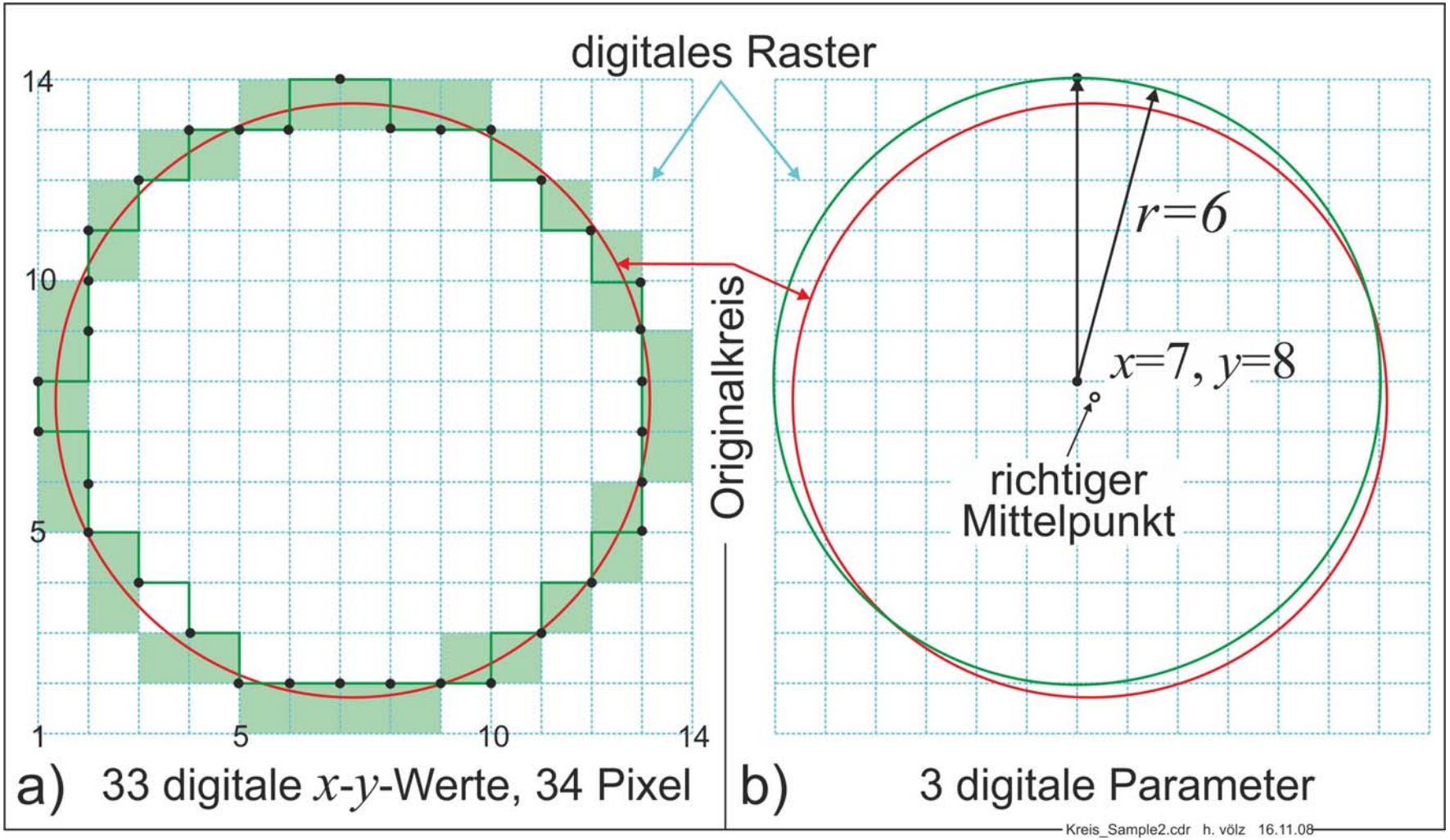
Ein einfacher Vergleich



In der üblichen Digitaltechnik sind eigentlich die roten Samples überflüssig

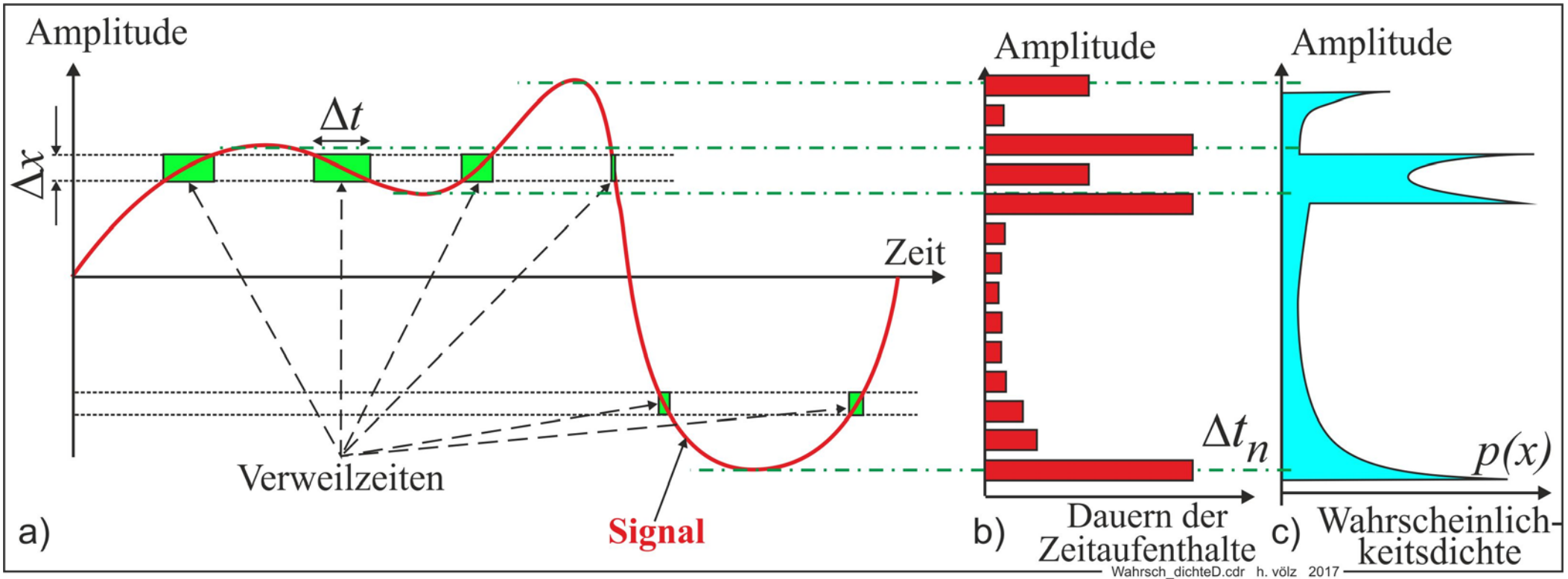
Beispiel Kreis

Bei der Kontinuierlichen Digitaltechnik genügen Mittelpunkt und Radius statt der 33 *x*-*y*-Werte.
Zusätzlich ist der Kreis trotz der etwas abweichenden Werte sogar viel exakter sichtbar.



Kontinuierliche Signale und Wahrscheinlichkeit

Für kontinuierliche Signale gibt es nur Wahrscheinlichkeitsdichten $p(x)$
Sie können über Amplitudenintervalle Δx und mit Grenzwerten gebildet werden



Damit kann aber nur sehr mittelbar eine Entropie bestimmt werden.

Kontinuierliche Entropie

Hierfür hat Zemanek eineen nicht ganz gültige Berechnungsweg aufgezeigt[Zem75].
Als Δx-Pendant mit den einzelnen $p_n = \Delta t_n$ gilt gemäß der diskreten (digitalen) Entropieformel zunächst

$$H(n,\Delta x) = -\underbrace{\lim}_{\substack{\Delta x\to 0\\ n\to\infty}} \sum_{i=0}^{n} p_i(\Delta x)\cdot\Delta x\cdot \mathrm{ld}(p_i(\Delta x))\cdot\Delta x\,.$$

Die Grenzübergänge $\Delta x \to 0$, $n \to \infty$ führen zum ***Integral*** und benötigen die Wahrscheinlichkeitsdichte $p(x)$.
Infolge des Logarithmus kann das Produkt unter dem Integral in eine Summe zerlegt werden:

$$H(x) = -\int_{-\infty}^{+\infty} p(x)\cdot \mathrm{ld}(p(x))\cdot dx - \underbrace{\lim}_{\substack{\Delta x\to 0\\ n\to\infty}} \sum_{i=0}^{n} p_i(\Delta x)\cdot \mathrm{ld}(p_i(\Delta x))\cdot\Delta x\,.$$

Der zweite Teil divergiert wegen $\underbrace{\lim}_{\Delta x\to 0} \log(\Delta x) \to -\infty$.

Deshalb wird nur der erste Term als relative (Nutz-) Entropie weiter benutzt:

$$h_{Nutz}(x) = -\int_{-\infty}^{+\infty} p(x)\cdot \mathrm{ld}(p(x))\cdot dx \quad \text{.mit} \int_{-\infty}^{+\infty} p(x)dx = 1\,.$$

t-kontinuierliche Signale sind von Störungen überlagert → führt dann zur relativen Stör-Entropie $h_{Stör}(x)$.
Sie ist von der oben bestimmten relativen Nutz-Entropie $h_{Nutz}(x)$ abzuziehen. Dann folgt:

$$H_k(x) = h_{Nutz}(x) - h_{Stör}(x)\,.$$

Gehorchen das Signal und die Störung einer Normalverteilung, so folgt nach längerer Rechnung

$$\boxed{H_k = \mathrm{ld}\left(\frac{P_N + P_S}{P_S}\right).}$$

Andere Verteilungen

Die relativen Entropien h_{Nutz} und $h_{Stör}$ sinb recht schwierig zu bestimmen (RandwertProbleme)

Obige Ableitung gilt nur für die Gauß-Verteilung beim Signal und Rauschen.

Für amplitudenbegrenzte Signale $x_1 \le x \le x_2 \rightarrow p(x) = 1/(x_2\text{-}x_1)$ folgt noch relativ einfach $h(x) = \mathrm{ld}(x_2 - x_1)$. Jedoch bei leistungsbegrenzten Signalen mit

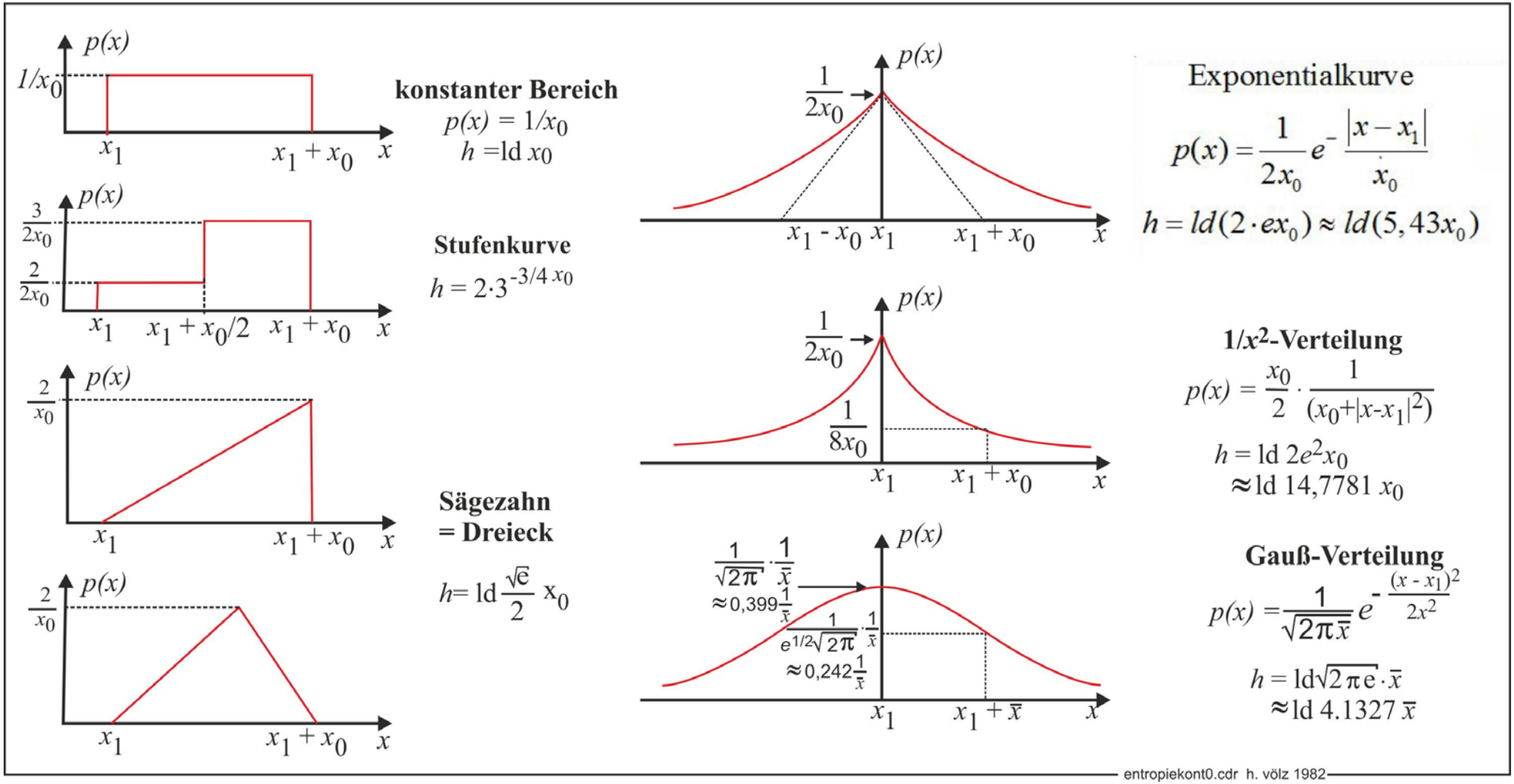
$$x_{m2} = \int_{-\infty}^{+\infty} x^2 \cdot p(x) \cdot \mathrm{d}x \quad \text{folgt bereits die Verteilung} \quad p(x) = \frac{1}{x_m \cdot \sqrt{2 \cdot \pi \cdot e}} \cdot e^{-(x/x_m)^2/2} \; .$$

ergibt sich

$$h_{Nutz}(x) = \mathrm{ld}(x_m \cdot \sqrt{2 \cdot \pi \cdot e} \approx \mathrm{ld}(4{,}133 \cdot x_m) \; .$$

Bekannte Beispiele

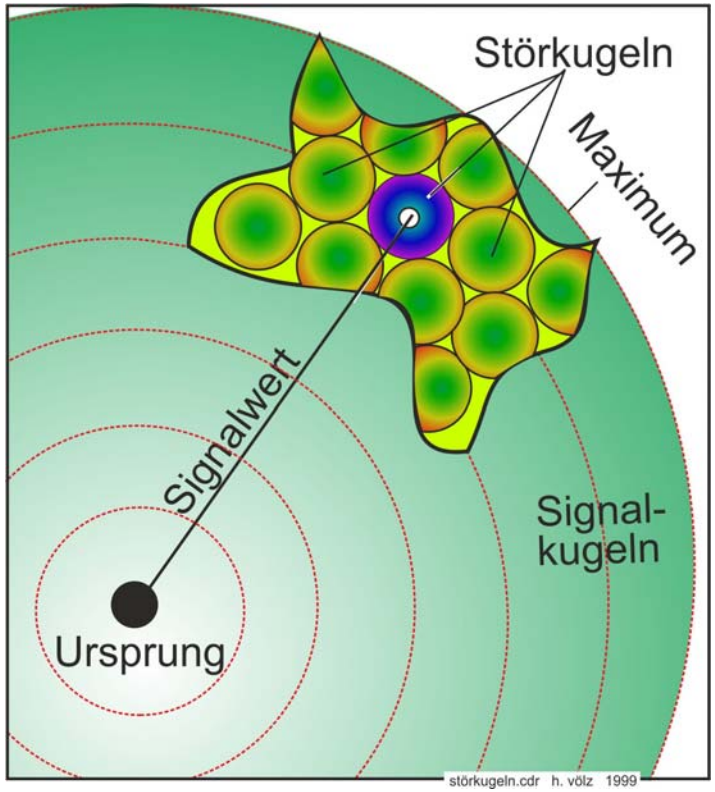
Diese Zusammenstellung zeigt die bisher wohl nur berechneten Verteilungen. [Pet67], [Völ82], S. 28ff.



Berechnung nach Shannon

Shannon leitet die kontinuierliche Entropie mittelbar über die Kanalkapazität C_K in Bit/s ab.
Das Sampling-Theorem erzeugt aus einen Signal der Zeitdauer T aus $n = 2 \cdot B \cdot T$ Samples.
Sie werden in einem Raum mit $n = 2 \cdot B \cdot T$ Dimensionen angeordnet.
Für die Kugel mit dem Radius r beträgt dann das Volumen

$$V = r^n \frac{\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)}.$$



Mit a folgt aus der Leistung P ein Radius $r = a \cdot \sqrt{P}$.
So ist Volumen für Nutzenergie P_N und Störenergie P_S berechenbar.
Daher haben endlich viele Störkugeln in der Nutzkugel Platz.

$$n_{n_{AS}}^{2 \cdot B \cdot T} \le \frac{V_N}{V_S} = \left(1 + \frac{P_N}{P_S}\right)^{2 \cdot B \cdot T}.$$

n_{AS} entspricht der Anzahl unterscheidbarer Zeichen/Sample.
Ihr Logarithmus ist der maximale Anzahl unterscheidbarer Bit.
Für die Kanalkapazität ist noch Normierung auf die Zeit notwendig:

$$C_K = \frac{\mathrm{ld}(n_{AS})}{T} = 2 \cdot B \cdot \mathrm{ld}\left(1 + \frac{P_N}{P_S}\right).$$

Und wegen $n_{AS} = \frac{P_N + P_S}{P_S}$

ergibt sich exakt die kontinuierliche Entropie.

$$H_k = \mathrm{ld}(n_{AS}).$$

Logarithmische Amplitudenstufen

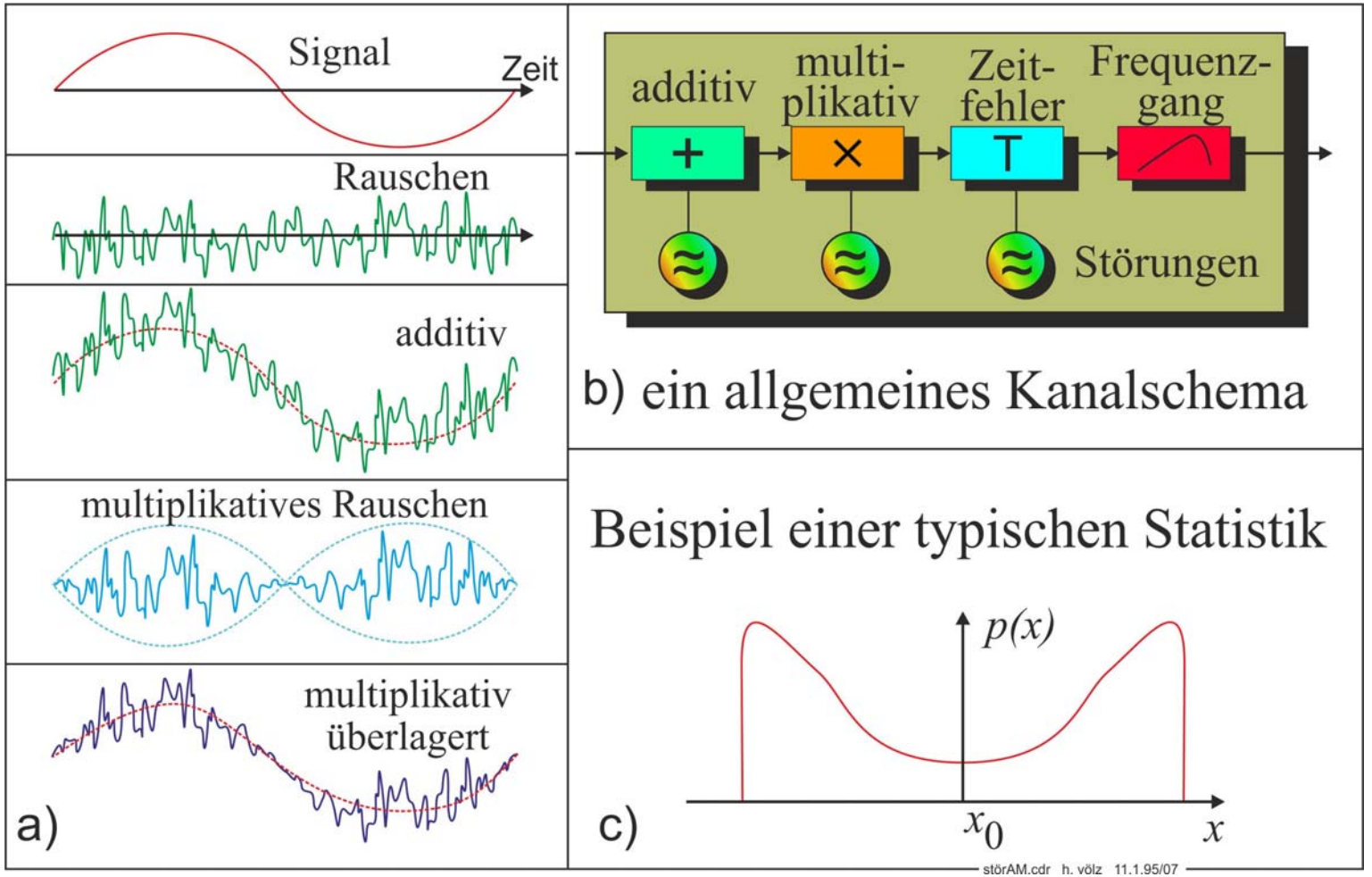
Bei einigen Signalen sind Intensität, Verteilung der Störung abhängig vom Signal (additive Überlagerung). z.B. Magnetbandkanal zwei nahezu unabhängige und deutlich unterschiedliche Störquellen: konstante Rauschquelle additiv zum Signal u_s. + multiplikatives Rauschen mit „Modulationsgrad“ *m:*

$$u_{St} = u_s \pm m \cdot u_n.$$

Es kann auch Zeitfehler auftreten.

Alle meine Versuche hiermit die Entropie über die dazugehörende Verteilung $p(x)$ zu berechnen scheiterten kläglich.

1962 erfolgte in Stuttgart eine Dissertation von Erich Pfeiffer mit fehlerhaften Beweis! Trotzdem wies er mein Ergebnis aus.



Meine Berechnung

Grenze zwischen zwei unterscheidbare Amplitudenstufen [Völ59a]:

$$u_n \cdot (1+2 \cdot m) + u_s = u_{n+1} \cdot (1-2 \cdot m) - u_s.$$

Hierzu Substitution $u_n^+ = u_n + \frac{u_s}{2 \cdot m}$, aus ihr folgt dann $\frac{u_{n+1}^+}{u_n^+} = \frac{1+2m}{1-2m}$.

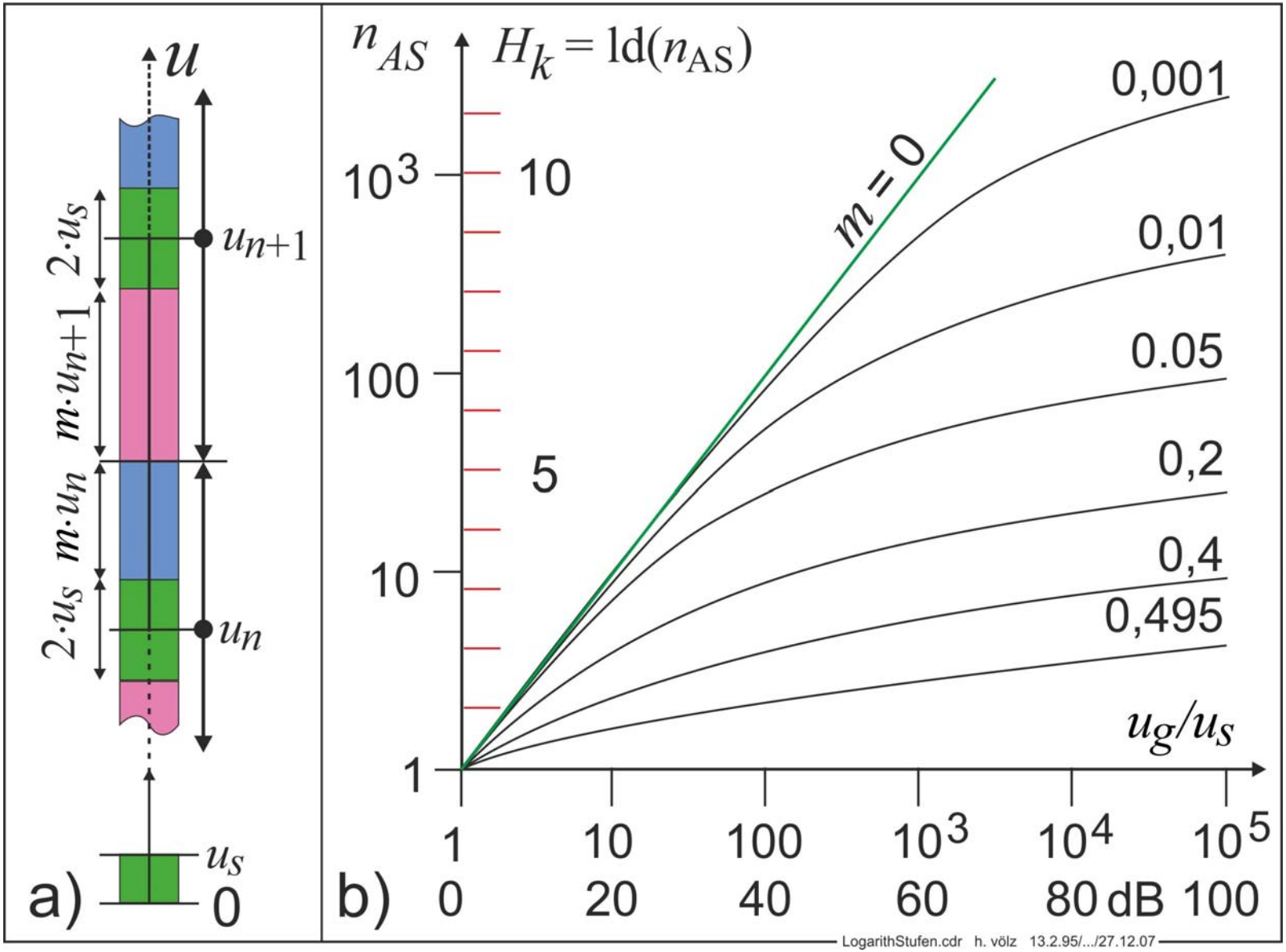
Für n Stufen gilt daher $\frac{u_{n+1}^+}{u_1^+} = \left(\frac{1+2m}{1-2m} \right)^n$.

Substitution zurücknehmen und größtmöglicher Wert $u_{n+1} = u_g$ dann folgt

$$n_{AS} = \frac{2 \cdot \log \frac{2 \cdot m \cdot \frac{u_g}{u_s} + 1}{2 \cdot m + 1}}{\log \frac{1+2 \cdot m}{1-2 \cdot m}} + 1 \approx \frac{\ln\left(2 \cdot m \cdot \frac{u_g}{u_k} + 1 \right) - \ln(1+2 \cdot m)}{2 \cdot m} + 1.$$

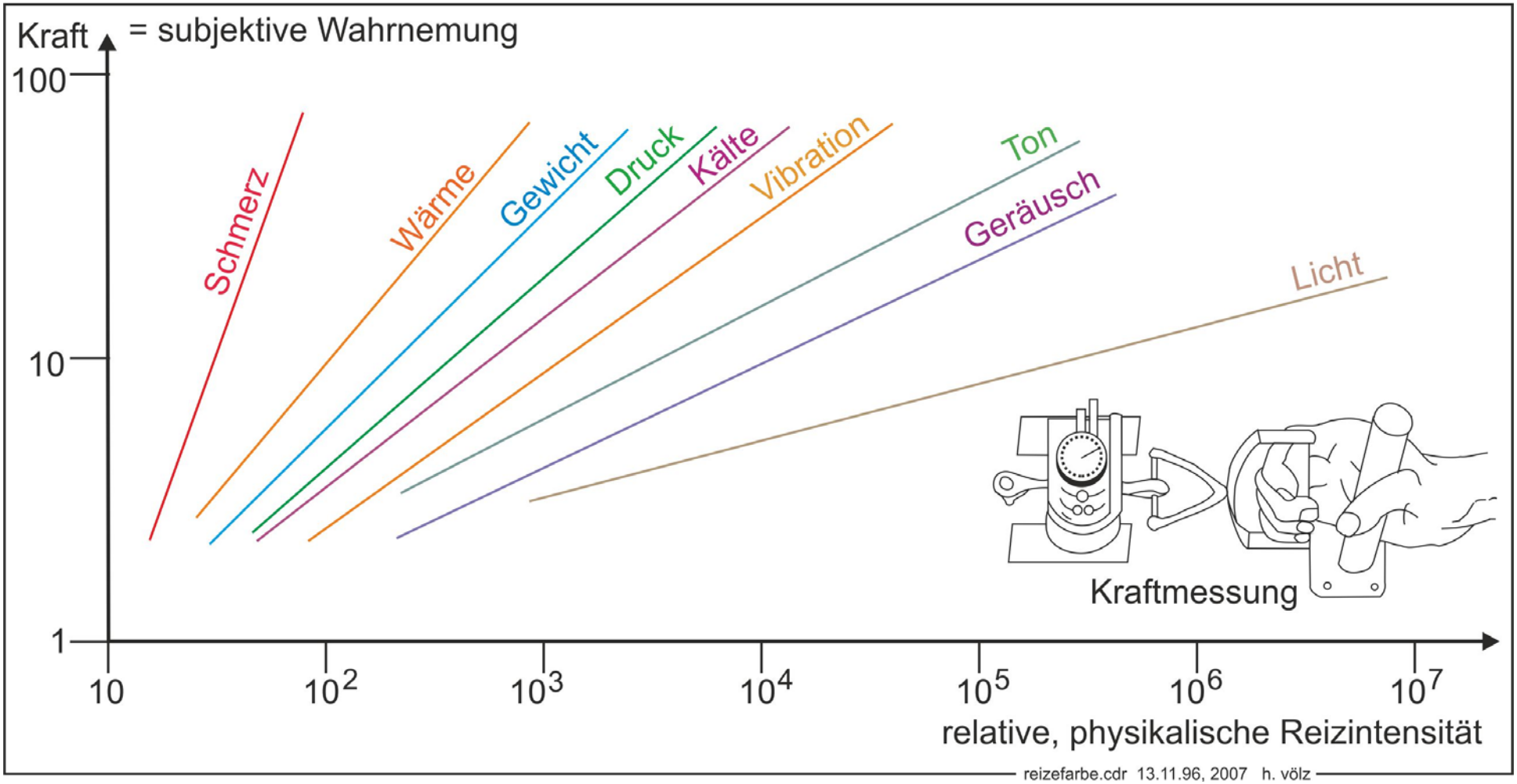
Die Auswertung der Formel für den maximalen Störabstand u_g/u_s zeigt das folgende Bild

Einfluss störender Amplitudenmodulation



Weber-Fechner-Gesetz 1

Die logarithmische Verteilung von Amplitudenstufen gilt auch bei unseren Sinnen
Das Gesetz geht auf Untersuchungen von Weber 1834 und Fechner von 1860 zurück [Sch93].
Die subjektive Wahrnehmung *Y* wird verglichen mit einer Kraft an der Federwaage



Weber-Fechner-Gesetz 2

Hiernach folgt die Kraft Y mit guter Näherung dem Logarithmus der Energie des physikalischen Reizes R:

$$Y = a + b \cdot \log(R).$$

Mit einer Konstanten k und einem Anfangsreiz R_0 kann das auch so geschrieben werden

$$Y \approx k \cdot (R - R_{0S})^n.$$

Für die einzelnen Sinne gilt die Tabelle [Völ99], S. 20.

Reizart	**Exponent *n***	**Dynamikbereich**	**Unterscheidbare Stufen**
Schmerz	2,1	15 dB; 1:8	≈5
Wärme	1,6	33 dB; 1:40	≈20
Schwere	1,5	24 dB, 1:17	≈50
Druck	1,1	20 dB; 1:10	≈50
Kälte	1,0	20 dB; 1:10	≈10
Vibration	0,95	50 dB; 1:300	≈100
Schall	0,6	100 dB; $1:10^5$	≈400 Lautstärken
Licht	0,33	130 dB; $1:3 \cdot 10^6$	≈600 Helligkeiten

Anwendung der kontinuierlichen Entropie

Die diskrete Entropie berechnet viele Bit/Zeichen mindestens für die digitale Codierung notwendig sind.
Mit bekannter Taktrate folgt daraus die kürzest mögliche Übertragungszeit je Zeichen.
Die kontinuierliche Entropie muss aber eine andere Bedeutung haben.
Mit der Kanalkapazität C_K kann mit der Übertragungsdauer $T_{Ü}$ die Informationsmenge I übertragen werden:

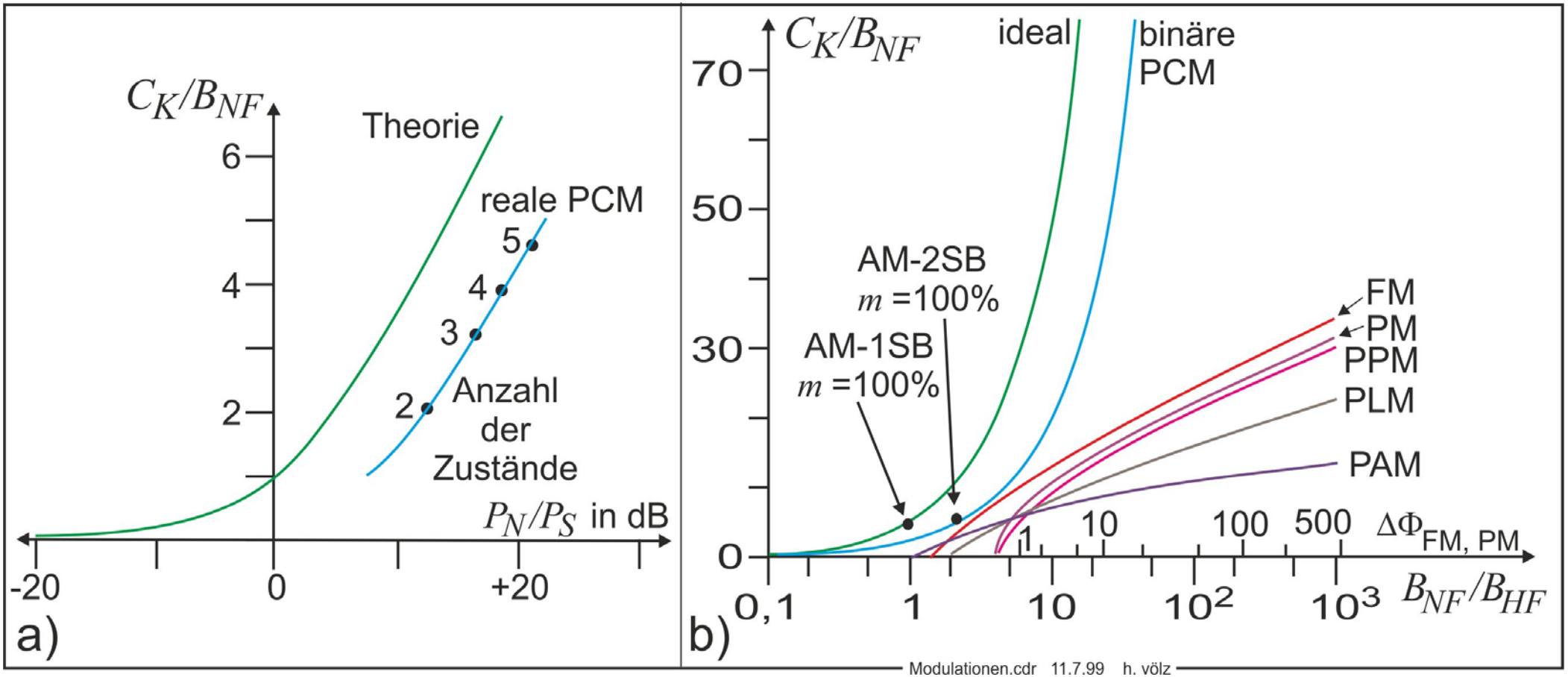
$$I = T_{Ü} \cdot C_K = T_{Ü} \cdot 2 \cdot B \cdot \mathrm{ld}\left(1 + \frac{P_N}{P_S}\right) = T_{Ü} \cdot \frac{\mathrm{ld}(n_{AS})}{T_0}.$$

Die drei darin enthaltenen Größen:

Übertragungszeit $T_{Ü}$, **Bandbreite** B und **Störabstand** P_N/P_S

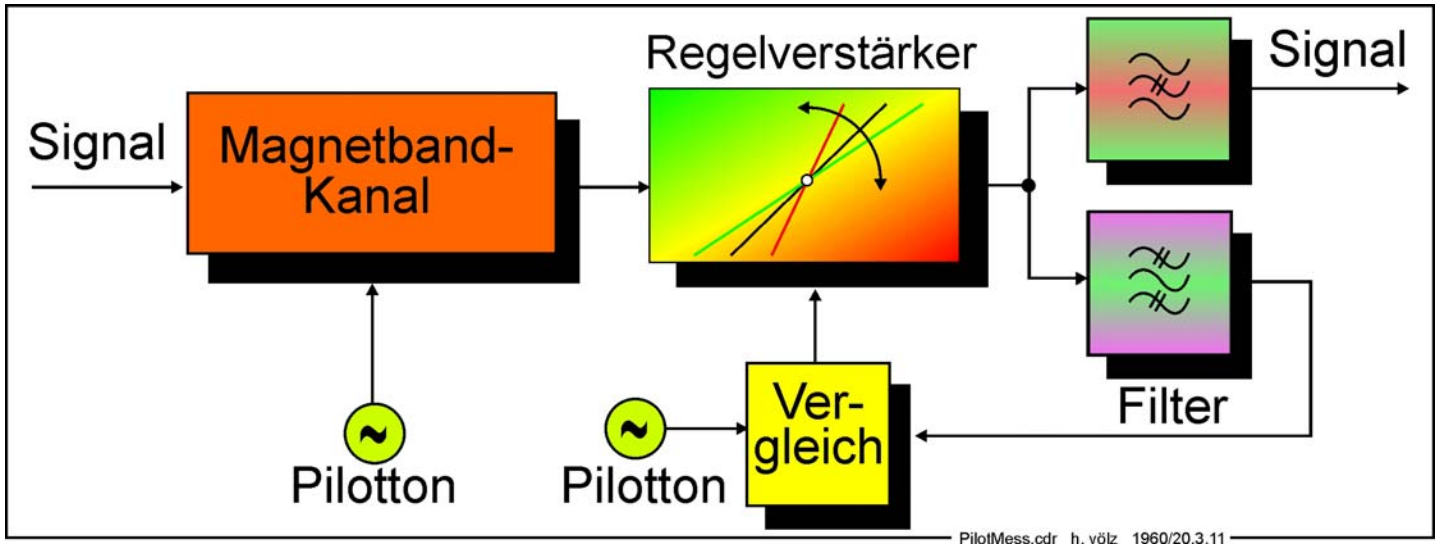
sind durch Modulationen usw. (maximal) gegeneinander austauschbar
Ähnliches gilt auch für die nutzbaren **Amplitudenstufen** n_{AS} und die **Einschwingzeit** T_0 des Systems
Bezüglich Störabstand und Bandbreite zeigt einige Möglichkeiten das folgende Bild.
Auch mit Pulscode-Modulationen PCM sind die erreichbaren Werte immer kleiner als die Theorie.
Lediglich bei der **Einseitenband-Amplitudenmodulation** (AM-1SB) gilt der theoretische Wert
Solche Austauschmöglichkeiten ergänzt die Digitaltechnik durch Stückelungen von Informationsmengen sowie Verschachtelungen unterschiedlicher Informationsarten.

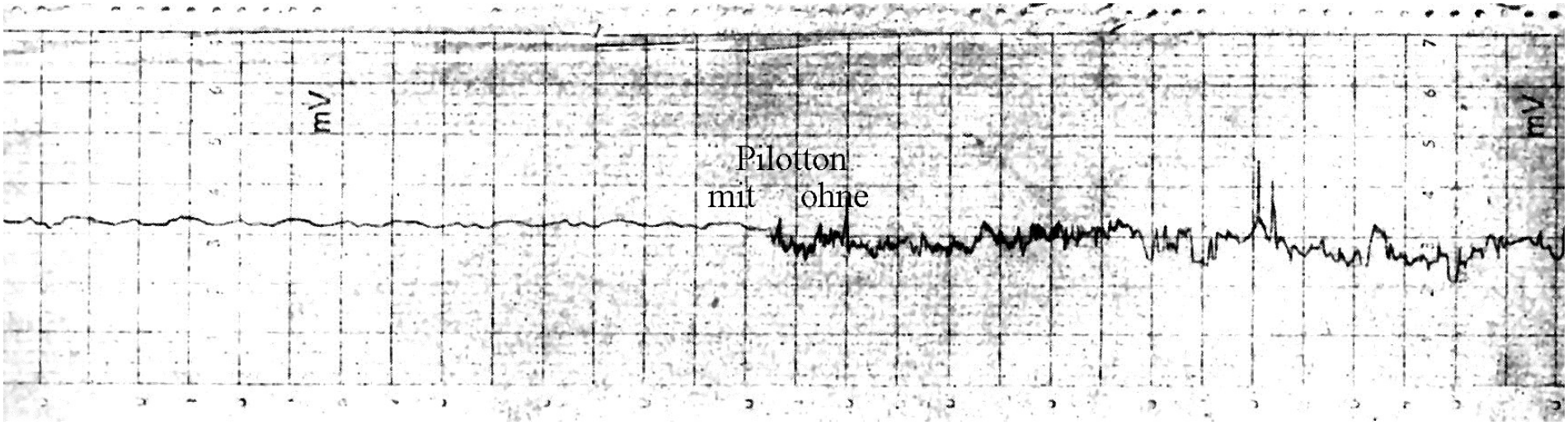
Austauschgrenzen kontinuierlich



Austausch Dynamikregelungen

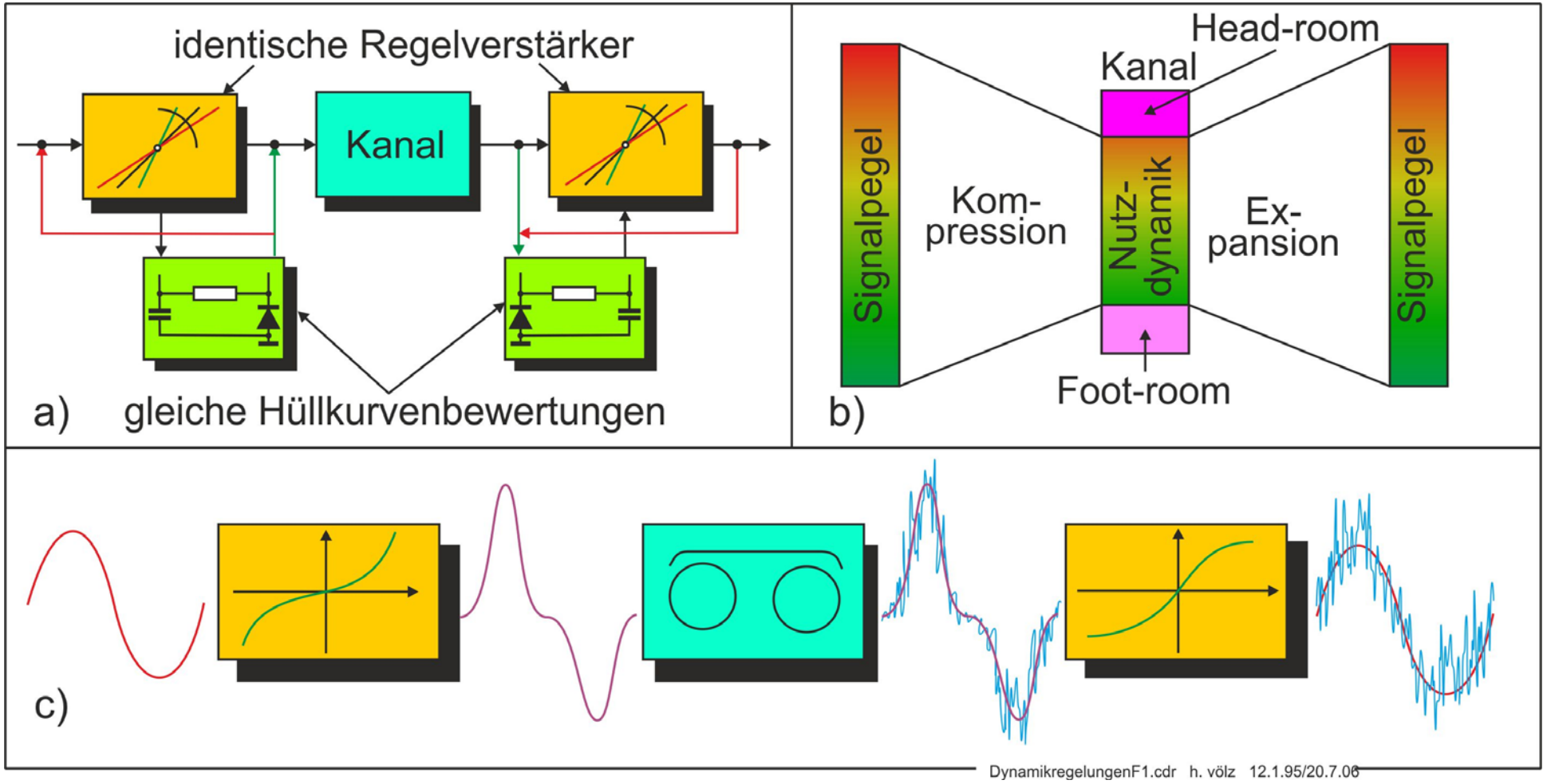
Bei der Magnetband-Speicherung begrenzt die störenden Amplitudenmodulation die Messung auf ±10 %
Mittels Pilotton konnte ich sie bereits in den 1950er Jahren auf 1 % erhöhen [Völ58b].



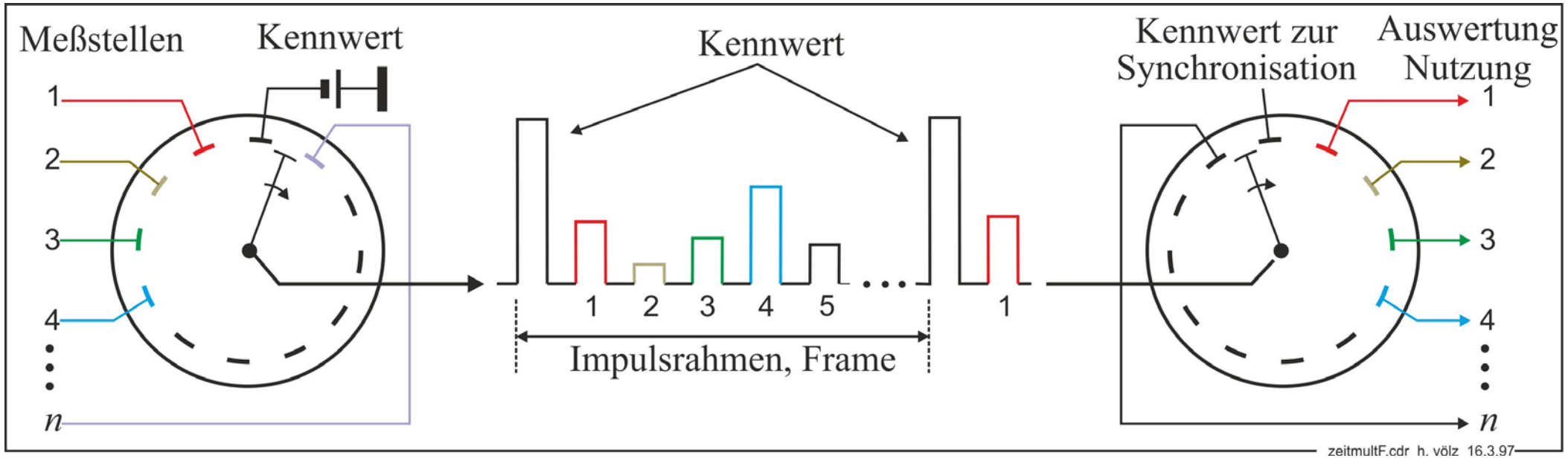


Reziproke Dynamikregelung (Dolby)

Viel universeller ist damals auch von mir erprobte reziproke Dynamikregelung [Völ58a].
Mit ihr kann die Messgenauigkeit oder der Dynamikbereich vergrößert werden.
Erst sehr viel später wurde sie von Dolby benutzt.

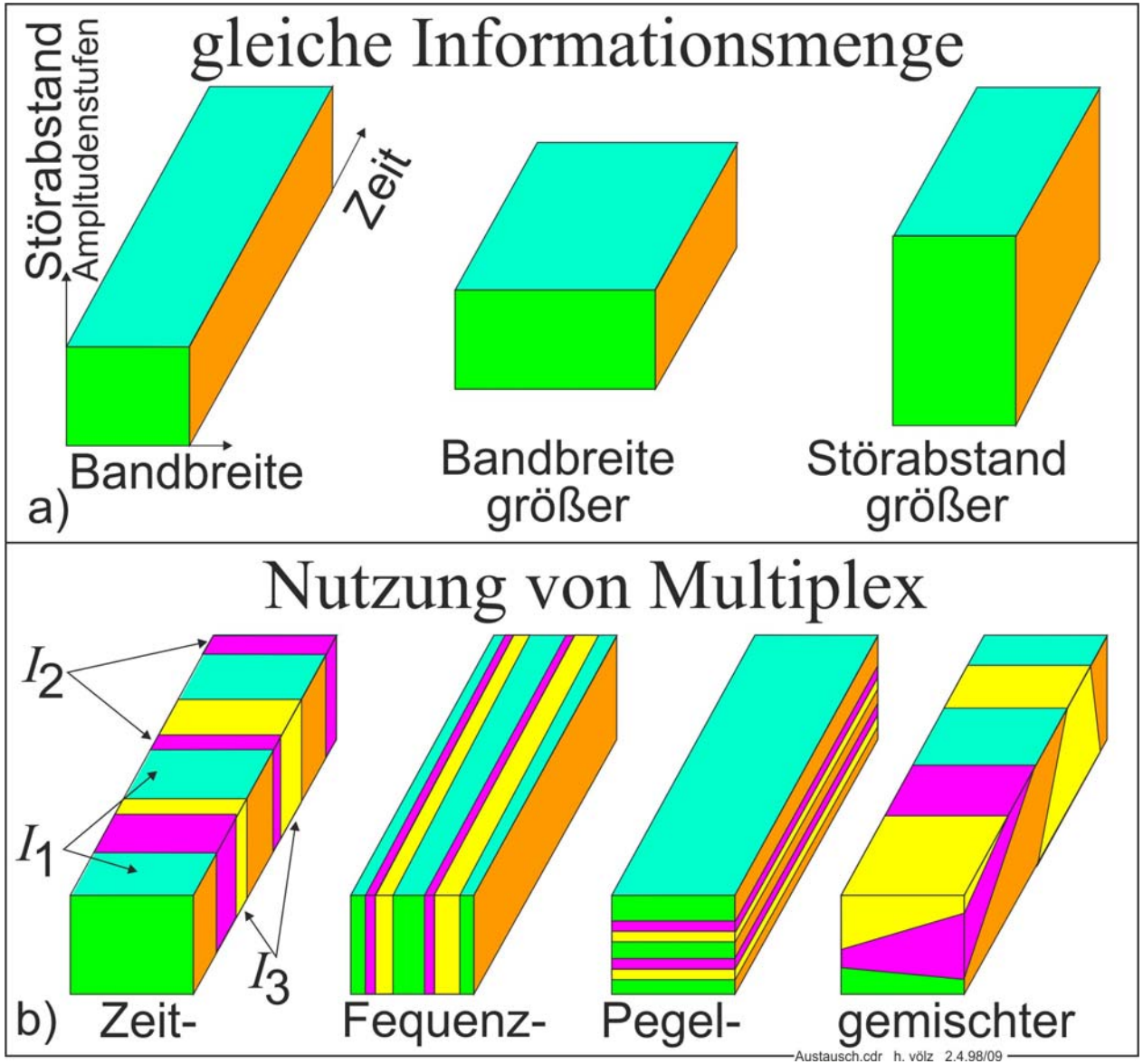


Ein Vorteil der Digitaltechnik



Die Signalverschachtelung

Austauschmöglichkeiten digital



Extrem geringe Energie für Information

Die äußerst geringe Energie bei der Information stellte Buchwald bereits in einem Vortrag 1950 fest.
und zwar in Bezug zu den entsetzlichen Folgen der Kernenergie [Buc50], S. 145:

> *„Man stellt fest, dass eine Stunde Redens energetisch einen Liter Wasser um 2/10 000 ºC erwärmen könnte – um wie viel weniger die Masse der Zuhörer. Dieser geringe Nutzeffekt ist für den Redner wahrhaft erschütternd - als Trost bleibt, dass menschliche Wechselbeziehungen denn doch nicht so einfach in einer physikalischen Bilanz niederzulegen sind.“*

Auch Wiener sagt ähnliches am Anfang vom 1.Kapitel und zwar in Bezug [Wie48], S.192.
zum Leistungsverbrauch des Gehirns für die geistigen Operation:

> *„Trotzdem ist die Energie, die für eine einzelne Operation verbraucht wird, beinahe verschwindend gering und bildet sogar nicht einmal ein angemessenes Maß der Funktion selbst “*

Eine Berechnung ist nun aber mit der hergeleiteten Informationstheorie möglich

Energie Je Bit

Es wird angenommen, dass die Störleistung allein durch thermisches Rauschen bestimmt $P_S = k \cdot B \cdot T$ ist.
benutzt werden B die Bandbreite, T die absolute Temperatur, $k = 1{,}381 \cdot 10^{-23}$ J/K die Boltzmann-Konstante.
Die Nutzleistung P_N sei das z-Fache der Störleistung $P_N = z \cdot P_S$. Für das Verhältnis von Nutzleitung zur Kanalkapazität C_K ergibt sich dann

$$\frac{P_N}{C_K} = k \cdot T \cdot \frac{z}{ld(1+z)} \quad \text{gemessen in} \frac{\text{J}}{\text{Bit}} \text{bzw.} \frac{\text{W}}{\text{Bit/s}}.$$

Mit einer Reihenentwicklung $\frac{z}{\ln(1+z)} = \frac{1}{1 - \frac{z}{2} + \frac{z^2}{3} - \frac{z^3}{4} \pm \ldots}$

ergeben sich die Grenzen des Quotienten zu $1 < \frac{z}{\ln(1+z)} \to 1 \text{ für } z \to 0$.

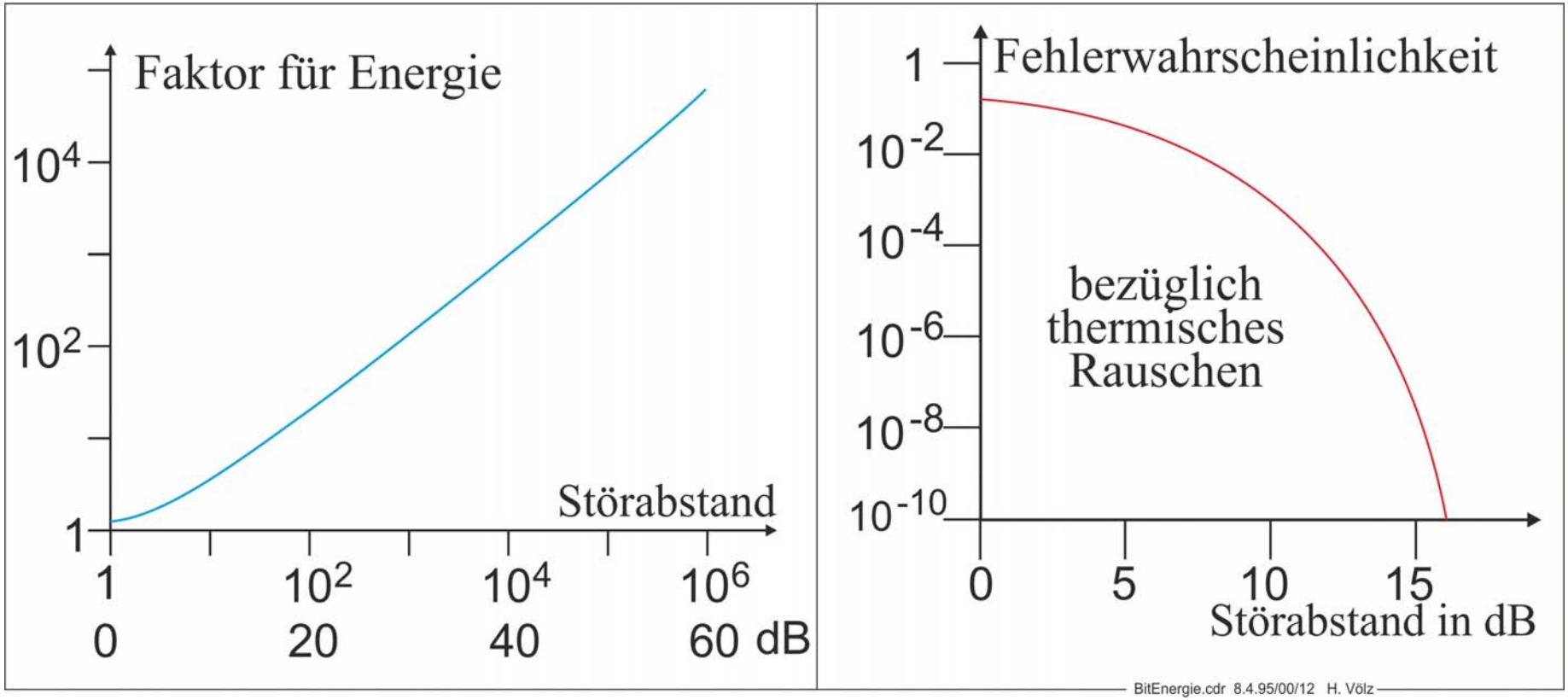
Daher gilt für die Energie je Bit

$$\boxed{\frac{E}{Bit} \geq k \cdot T \cdot \ln(2).}$$

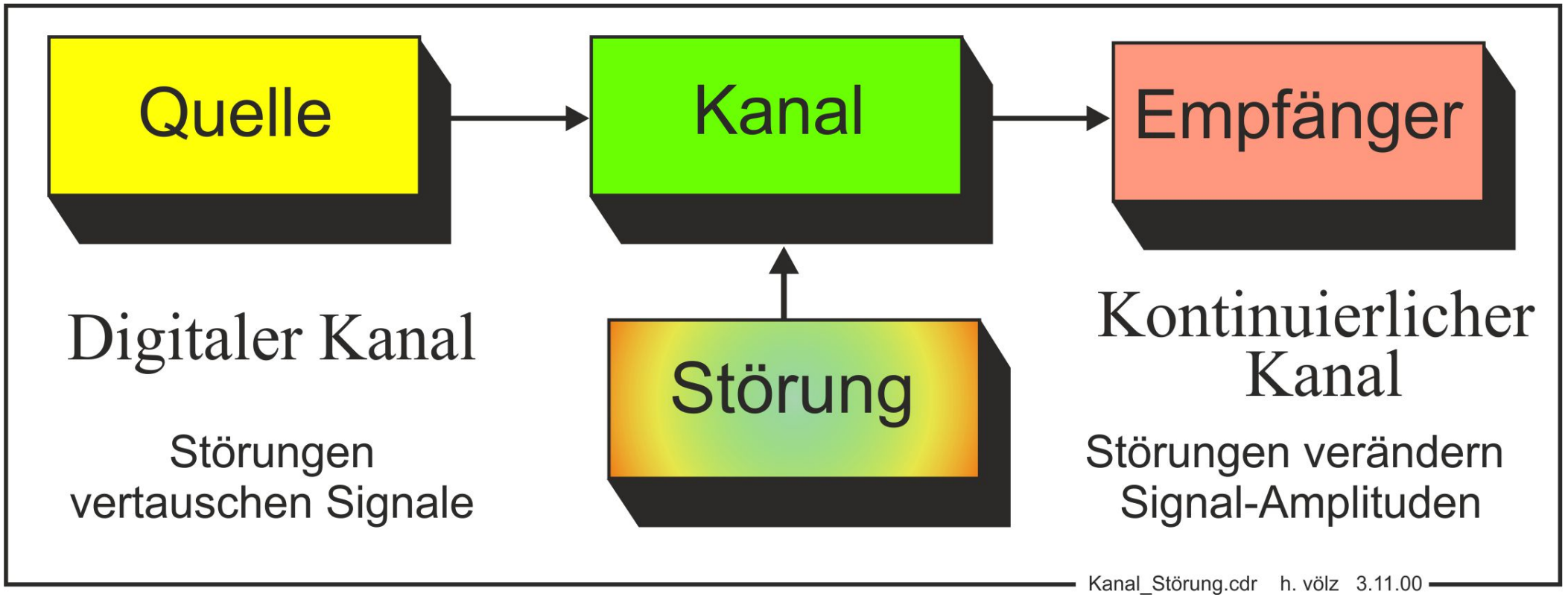
Bei 300 K (≈ Zimmertemperatur) folgt E/Bit ≥ 3.10^{-21} J ≅ $5 \cdot 10^{11}$ Hz ≅ $5 \cdot 10^{-22}$ cal ≅ 26 mV.
Genau dieser Wert lässt sich auch quantentheoretisch herleiten.
Bei dieser minimalen Energie je Bit strebt jedoch der Störabstand des Signals gegen Null.
Das würde extrem große Fehlerraten hervorrufen. Für Anwendungen ist daher eine deutlich größere Energie erforderlich.

Berücksichtigung des Störabstandes

Die minimale Energie muss für einen hinreichenden Störabstand erhöht werden
Für eine brauchbare Fehlerkorrektur (später) sollten die Fehler maximal bei 10^{-5} liegen.
Das entspricht etwa einen Störabstand von 15 dB und verlangt die zehnfache Energie.



Unterschied t-kontinuierlich und digital



Vor- und Nachteile

t-**kontinuierlich**	**digital (diskret)**
Vorteile (Stärken, Nutzen)	
Sie Treten fast ausschließlich bei technischen Sensoren, Aktoren, den menschlichen Sinnen und Handlungen auf. Mit einer Verstärkung sind beliebig kleine und für Aktoren, wie Lautsprecher und Motoren beliebig große Signale anwendbar. Bei akustischen Signalen setzen bei einer Übersteuerung die Verzerrungen relativ weich ein.	Die Signale sind weitgehend fehlerfrei zu speichern, vervielfältigen, übertragen und regenerieren. Sehr unterschiedliche Signale lassen sich problemlos verschachteln (Bild 33). Fehlererkennung, -korrektur und verlustfreie Komprimierung sind möglich. Es gibt wirksamen Datenschutz (Verschlüsselung, Kryptografie).
Nachteile (Schwächen)	
Bei jeder Übertragung, Speicherung und Vervielfältigung kommen Störungen, zumindest Rauschen hinzu. Jeder Signalwert besitzt einen Erwartungswert und eine Streuung.	Die Signale müssen hinreichend groß sein und auf einen eng begrenzten Pegel festgelegt sein. Der Takt darf nicht verloren gehen. Bei der üblichen Digitalisierung tritt störendes Sampling-Rauschen auf.

Ganz allgemein

Zumindest bei Übertragung und Speicherung überwiegen also deutlich die Vorteile der digitalen Technik.
Dazu muss aber alles (u. a. Sprache, Begriffe, Zeit usw.) in Zahlen (-Zeichen) codiert werden
Dennoch sind meist und sogar primär zusätzlich kontinuierliche Signale notwendig, ja unvermeidlich.
Daher ist es nicht verwunderlich, dass bereits seit dem Altertum sich auf fast allen Gebieten die kontinuierliche und diskrete (Welt-) Beschreibung gegenüberstehen.
Planck sprach von einem Akt der Verzweifelung von Differentialgleichungen zu diskreten Quanten überzugehen.

Typ	**kontinuierlich**	**diskret**
Griechen	Aristoteles	Demokrit
Mathematik	unendlich	endlich
Antinomien Paradoxien	Hilbert-Hotel, grot = grün-rot Alle Raben sind schwarz Xenon Wettlauf	Xenon Pfeil Kretaer Lügner Gödel-Unentscheidbarkeit
Methoden	Limes, Differential	Zählen (Messen)
Beschreibung	Differentialgleichungen	Algorithmen
Geräte	Analogrechner	Digital-Computer
Entsprechung	Welle	Korpuskel, Quant
Übergang	Bohr’sches Korrespondenz-Prinzip; Heisenberg-Unbestimmheit	

Weitere Entropien

Der Begriff Entropie ist heute für über zehn mathematische Formeln und deren Kontext gebräuchlich. Dabei gibt es nur selten ähnliche oder gar übereinstimmende Inhalte.

Weil aber der Gehalt der Formeln oft nicht richtig verstanden wird, werden oft – sogar bei Experten mehrerer Gebiete – erheblich falsche Kombinationen und Folgerungen abgeleitet.

Mehrfach betreffen sie einen vermuteten (behaupteten) aber falschen Zusammenhang zwischen thermodynamischer und Shannon-Entropie(s. u.), oder sogar behaupten sogar ihre Gleichheit.

Den ursprünglichen Entropie-Begriff beschrieb Clausius 1865 (s. [Eig83] S. 164):

> *„Sucht man für S (die Entropie) einen bezeichnenden Namen, so könnte man, ähnlich wie von der Größe U (der inneren Energie) gesagt ist, sie sey der Wärme- und Werkinhalt des Körpers, von der Größe S sagen, sie sey der Verwandlungsinhalt des Körpers. Da ich es aber für besser halte, die Namen derartiger für die Wissenschaft wichtiger Größen aus den alten Sprachen zu entnehmen, damit sie unverändert in allen neuen Sprachen angewandt werden können, so schlage ich vor, die Größe S nach dem griechischen Worte »tropae«, die Verwandlung, die Entropie des Körpers zu nennen. Das Wort Entropie habe ich absichtlich dem Wort Energie möglichst ähnlich gebildet, denn die beiden Größen, welche durch diese Worte benannt werden sollen, sind ihren physikalischen Bedeutungen nach einander so nahe verwandt, daß eine gewisse Gleichartigkeit in der Benennung mir zweckmäßig zu seyn scheint.“*

Die(se) Entropie betrifft damit den Anteil der Wärme-Energie der maximal in mechanisch nutzbare (bei der Dampfmaschine gewandelt werden kann.

Beginn der Thermodynamik

Ab dem 16. Jh. war die Dampfmaschine gut bekannt.
Jeder gute Ingenieur konnte damals eine bauen und einsetzen.
Mit ihr begann die industrielle Revolution begann.
Den großen technischen Fortschritt weist schon aus: Die Leistung wird in Pferdestärken PS angegeben.

Für das inhaltliche Verständnis des obigen Zitats sind jedoch thermodynamische Grundlagen notwendig.
Ursprünglich wurde Wärme noch als Stoff betrachtet.
Doch bereits 1738 hatte Bernoulli sie als ungeordnete Bewegung der Moleküle erkannt.
1843 formulierte Mayer den Energie-Erhaltungssatz.
Wenig später präzisierten ihn Joule und Helmholtz.
Das führte schließlich zu den drei Hauptsätzen der Thermodynamik
Zuweilen wird noch ein nullter hinzugefügt
Ganz wichtig und wesentlich (zentral) ist der 2. Hauptsatz, den 1851 Kelvin formulierte

Entscheidend ist jedoch der Carnot’sche Kreisprozess von 1824 [Car09].

Die Sätze der Thermodynamik

0. **Streben nach Gleichgewicht:** Nach hinreichend langer Wartezeit stellt sich in geschlossenen Systemen immer ein Gleichgewicht ein, bei dem alle ***Temperaturdifferenzen ausgeglichen*** sind und sich die makroskopischen Größen, wie Druck, Temperatur, Volumen usw. nicht mehr ändern.

1. ***Energieerhaltung:*** Die Gesamtenergie eines Systems ist $U = {}^3/_2 \cdot N \cdot k \cdot T$; N= Teilchenzahl, k = Boltzmann-Konstante und T = absolute Temperatur. Die Änderung der **inneren** Energie $\Delta U = A + Q$ besteht aus **mechanischer** Energie (Arbeit) A und **Wärme** Q.
Mechanische Energie ist immer vollständig in Wärme umwandelbar, aber nicht umgekehrt (Zeitpfeil)
Es gibt kein **Perpetuum Mobile 1. Art**, das dauernd Energie erzeugt, ohne sie der Umgebung zu entziehen.

2. Thermodynamische ***Entropie*** in zwei Varianten: 1854 Clausius: $\Delta S = \Delta Q/T$ gemäß dem Kreisprozess
1857 Boltzmann: $S = k_B \ln(W)$, auch als H-Theorem bezeichnet.
Nicht die Temperatur T sondern die **Wahrscheinlichkeit *W*** fes Zustandes. k_B = Boltzmann-Konstante.

3. ***Absoluten Nullpunkt*** $T = 0$, 1848 von Kelvin vorgeschlagen wurde.
Zusammenstöße der Moleküle führen letztlich zur Gleichverteilung aller Teilchen (Zeitrichtung)
Am Nullpunkt besitzt das System keine Anregungsenergie mehr. Hier herrscht der Wärmetod.
Experimentell ist er nie erreichbar. Entwicklung dorthin entspricht der Irreversibilität vieler Prozesse.
Unmöglichkeit des **Perpetuum Mobile 2. Art**, das ständig periodisch Wärme in Arbeit wandeln könnte.
Satz zwar nicht beweisbar, wird jedoch immer wieder beobachtet und bestätigt!

Der Carnot'sche Kreisprozess

Annahme: zwei idealen Wärmequellen mit konstanten Temperaturen $T_v > T_u$
= Verbrennungs- und Umwelttemperatur.
Das sehr gut isolierte Zylindergefäß besitzt verschiebbaren Kolben, mechanischen Energie-Ankopplung.
Volumen wird zwischen V_1 und V_2 geändert.
Gefäß wird wahlweise mit den Temperaturen T_v oder T_u gekoppelt.
Bei der jeweils konstanten Temperatur werden durch die Volumenänderung die hyperbelförmigen
Isothermen durchlaufen (Gas-Gesetze).
Die geradliniegen Isochoren entstehen beim Abkühlen bzw. Erhitzen (Volumen konstant gehalten).
Entlang den großen Pfeilen entsteht so der Kreisprozess. Er begrenzt die **Arbeitsfläche**.
Sie entspricht der aus der Wärmeenergie ΔQ gewonnenen mechanischen Energie ΔW.
Eine größere Menge nicht möglich. Daher gilt für den maximalen Wirkungsgrad

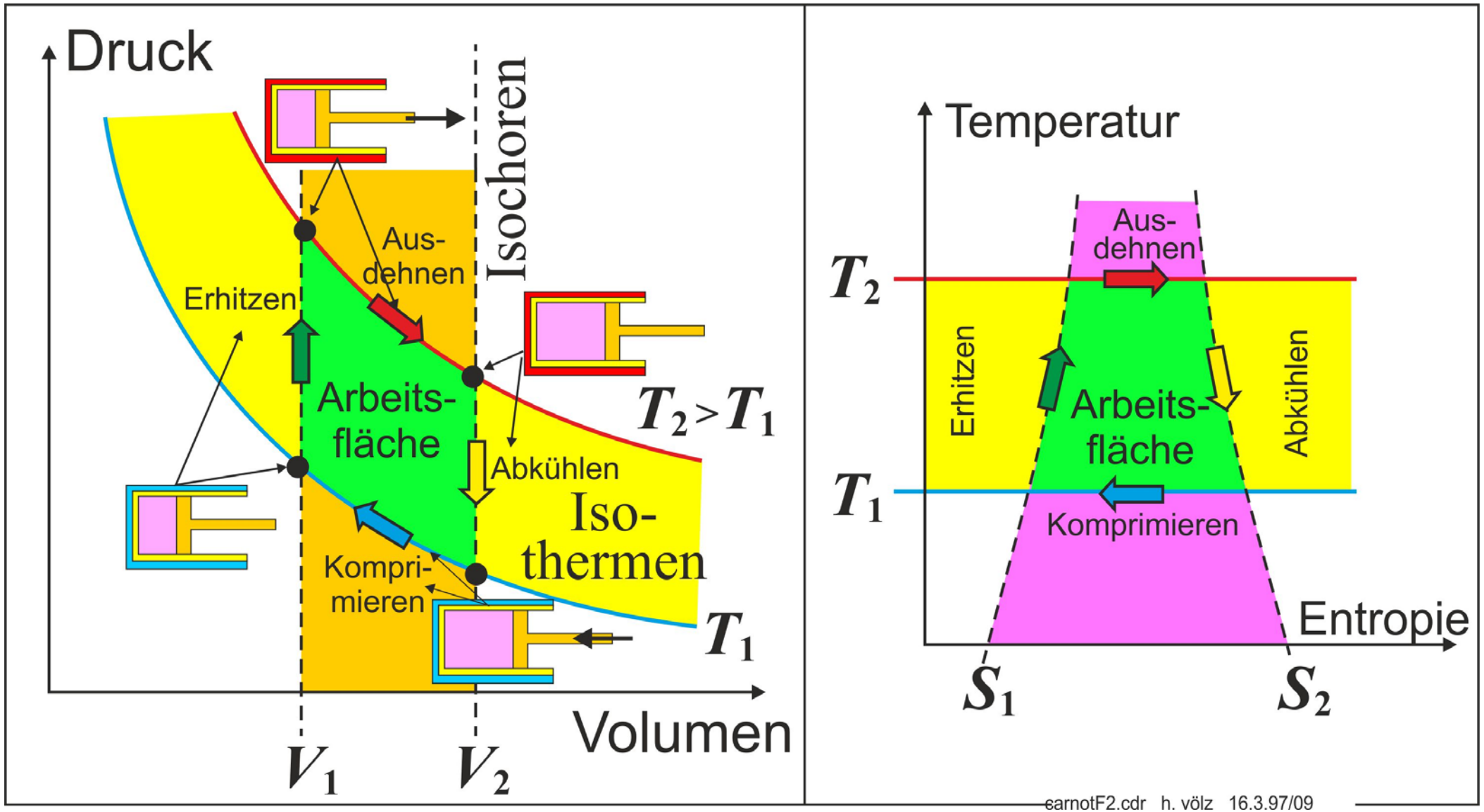
$$\eta = \frac{\Delta Q}{\Delta W} \leq \frac{T_v - T_u}{T_v} .$$

Clausius leitete hieraus seine Entropie ab:

$$\boxed{\Delta S = \Delta Q/T .}$$

Mit dieser Theorie entwickelte **Diesel** den deutlich besseren Verbrennungsmotor (1892 Patent).

Formal kann dieser Zusammenhang auf *alle Entropien* umgedeutet werden.
Dann bleibt nur noch übrig, Entropien betreffen den maximal möglichen *Austausch von Parameter-Werten*.
Wichtig ist aber, dass Boltzmann noch eine andere Berechnung, Interpretation fand.


Druck
Isochoren
Aus-
dehnen
Erhitzen
Arbeits-
fläche
$T_2 > T_1$
Abkühlen
Iso-
thermen
Kompri-
mieren
T_1
Volumen
V_1
V_2
Temperatur
Aus-
dehnen
T_2
Erhitzen
Arbeits-
fläche
Abkühlen
T_1
Komprimieren
Entropie
S_1
S_2
carnotF2.cdr h. völz 16.3.97/09

Boltzmann-Entropie

1859 leitete **Maxwell die Geschwindigkeitsverteilung** der statistischen Bewegungen der Moleküle eines Gases ab. Bei der absoluten Temperatur *T* und der Masse *m* der Moleküle beträgt deren mittlere Geschwindigkeit

$$\overline{v_{th}} = \sqrt{\frac{2 \cdot k \cdot T}{m}}.$$

Diese Betrachtung erweiterte **Boltzmann** ab 1868 und 1872.
Schließlich gelang es ihm, ein statistisches Äquivalent zur Clausius-Entropie abzuleiten.
Mit der Wahrscheinlichkeit *W* des jeweiligen Zustandes gewann er die Formel der Boltzmann-Entropie

$$\boxed{S = k \cdot \ln(W)\ .}$$

Die exakte Übereinstimmung erreichte er mit dem *natürlichen Logarithmus* und der *Boltzmann-Konstanten*

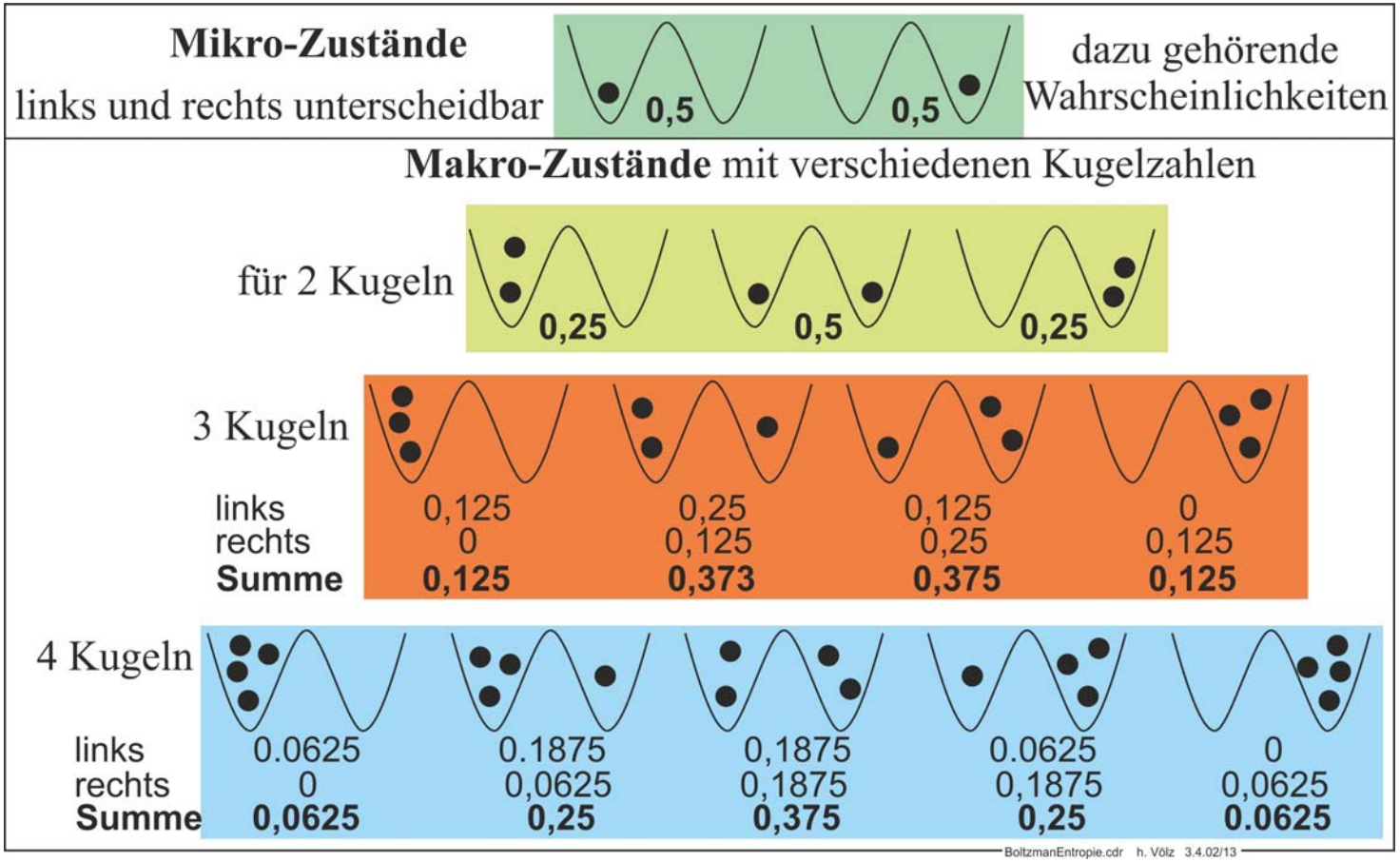
$k \approx 1{,}38065... \cdot 10^{-23}$ Joule/Kelvin

Erstaunlich ist, dass die Energie und Temperatur nicht mehr vorkommen.
Außerdem gilt infolge des Logarithmus, dass sich *Wahrscheinlichkeiten* ***multiplizieren*** $W = W_1 \cdot W_2$,
aber die *Entropien* ***addieren*** $S = S_1 + S_2$.

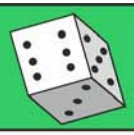
Ähnlich wie es bei Shannon die diskrete und kontinuierliche Entropie gibt, so bestehen auch hier zwei thermodynamische Entropien.

Makro- und Mikrozustände

Beispiel: Urne mit zwei Mulden, *n* Kugeln, die aus beachtlicher Entfernung hineingeworfen werden.
Mikrosystem nur eine Kugel, die gleichwahrscheinlich in die linke oder rechte Vertiefung fällt.
Jedes Mikrosystem kann z_i Zustände annehmen.
Makrozustand gilt für ein System aus *n* Mikrosystemen, die zu einer Einheit zusammengefasst sind.
Der Makrozustand besitzt die auftretende Wahrscheinlichkeit *W*

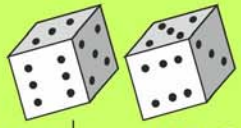


zweites Bespiel



Alle 6 **Mikro-Zusände** Wahrscheinlichkeit = 1/6 ≈ 0,167

Für die **Makro-Zustände** mit 2 oder 3 Würfeln gilt



Würfel werden nicht unterschieden

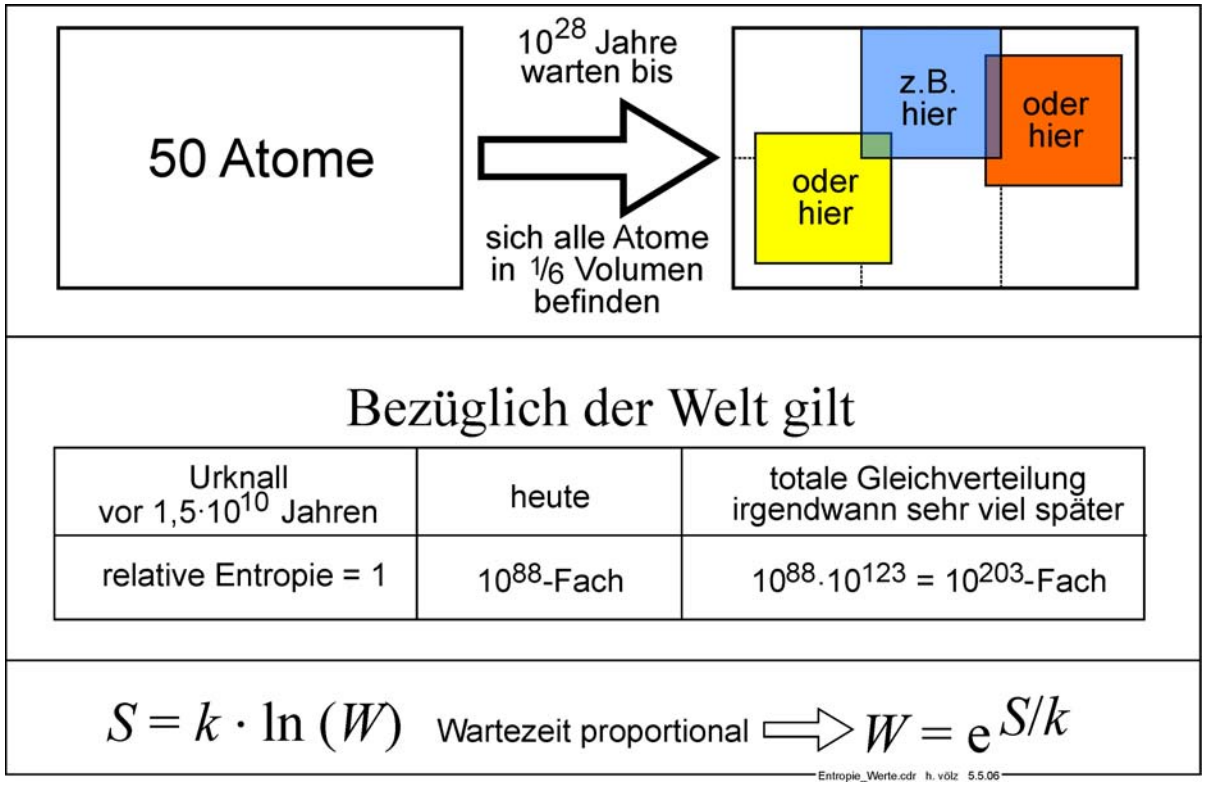
Augen	Wurf (Fälle)	Wahrscheinlichkeit
2	**11** (1)	1/36 ≈ 0,028
3	12 (2)	2/36 ≈ 0,056
4	13 (2); **22** (1)	3/36 ≈ 0,083
5	14 (2); 23 (2)	4/36 ≈ 0,111
6	15 (2); 24 (2); **33** (1)	5/36 ≈ 0,139
7	16 (2); 25 (2); 34 (2)	6/36 ≈ 0,167
8	26 (2); 35 (2); **44** (1)	5/36 ≈ 0,139
9	36 (2); 45 (2)	4/36 ≈ 0,111
10	46 (2); **55**(1)	3/36 ≈ 0,083
11	56 (2)	2/36 ≈ 0,056
12	**66**(1)	1/36 ≈ 0,028

wuerfelwahr2.cdr h. völz 26.12.98/2013

Augen	Wurf (Fälle)	Wahrscheinlichkeit
3	**111** (1)	1/216 ≈ 0,005
4	**11**2 (3)	3/216 ≈ 0,014
5	122 (3); **11**3 (3)	6/216 ≈ 0,028
6	**222** (1); **11**4 (3); 123 (6)	10/216 ≈ 0,046
7	**11**5 (3); 124 (6); 133 (3); 223 (3)	15/216 ≈ 0,069
8	**11**6 (3); 125 (6); 134 (6); 224 (3); 233 (3)	21/216 ≈ 0,097
9	126 (6); 135 (6); 144 (3); 225 (3); 234 (6); **333** (1)	25/216 ≈ 0,116
10	136 (6); 145 (6); 226 (3); 235 (6); 244 (3); 334 (3)	27/216 ≈ 0,125
11	146 (6); 155 (3); 236 (6); 245 (6); 335 (3); 344 (3)	27/216 ≈ 0,125
12	156 (6); 246 (6); 255 (3); 336 (3); 345 (6); **444** (1)	25/216 ≈ 0,116
13	166 (3); 256 (6); 346 (6); 355 (3); 445 (3)	21/216 ≈ 0,097
14	266 (3); 356 (6); 446 (3); 455 (3)	15/216 ≈ 0,069
15	366 (3); 456 (6); **555** (1)	10/216 ≈ 0,046
16	466 (3); 556 (3)	6/216 ≈ 0,028
17	566 (3)	3/216 ≈ 0,014
18	**666**(1)	1/216 ≈ 0,005

Ergänzungen zur Boltzmann-Formel

Vereinfacht ergibt sich so für jeden möglichen Makrozustand die Boltzmann-Entropie (obige Formel).
Unterschied zur Realität: In der Thermodynamik besteht Makrosystem aus sehr vielen Molekülen
Außerdem sind alle Bewegungen möglich
Daher können einzelne Kugeln mit hinreichender Bewegungsenergie über die Trennung hinaus springen
Je länger dabei gewartet wird, desto wahrscheinlicher wird der Zustand mit der größten Wahrscheinlichkeit.
Das entspricht dem Streben zum o. g. Wärmetod. Auch das Hund-Flöhe-Modell ist so erklärbar.
Prinzipiell ist so auch die sehr kurzzeitige Wiederkehr eines sehr unwahrscheinlichen Zustands möglich
Abschätzung zur Entwicklung unserer Welt versucht.



Boltzmanns Freitod

Heute gelten die Ergebnisse von Boltzmann als eine besonders große wissenschaftliche Leistung.
Doch zu seiner Zeit war das deutlich anders.
Von den meisten Physikern wurde sie aus zwei Gründen stark kritisiert.
Zunächst galt die ***Atom-Annahme*** als unbegründet, teilweise sogar als falsch:

„Hat denn schon jemand ein Atom gesehen?“

Das galt sogar, obwohl sie in der Chemie durch Dalton für die Stöchometrie eingeführt und generell sehr erfolgreich benutzt wurden.
Der zweite Widerspruch kam vor allem von Mach, Ostwald, Poincaré und Zermelo.
Er betraf die ***statistische Irreversibilität*** sowie die Ableitung des 2. Hauptsatzes aus reversiblen Mikrozuständen. Das trieb Boltzmann schließlich am 5.9.1906 in den Freitod.

Hierzu stellte Planck später fest, *dass sich eine neue wissenschaftliche Wahrheit normalerweise*

„nicht in der Weise durchzusetzen pflegt, daß ihre Gegner überzeugt werden und sich, als belehrt erklären, sondern vielmehr dadurch, daß die Gegner allmählich aussterben und daß die heranwachsende Generation von vornherein mit der Wahrheit vertraut gemacht wird“.

Thermodynamische kontra Shannon-Entropie

In der Literatur werden häufig – selbst von hervorragenden Wissenschaftlern – die thermodynamische und Shannon-Entropie fälschlich in (sehr) enge Beziehung zueinander gebracht.
Dazu vier Varianten von Formeln.
Die Zeilen – kontinuierlich und diskret/digital – sind dabei allerdings nur z. T. korrekt.
Bei endlich vielen Amplitudenstufen n_{As}, ist das Signal fast immer t-kontinuierlich.
Bei der Boltzmann-Entropie ist die Wahrscheinlichkeit *W* kontinuierlich.
Sie gilt immer nur für einen ***einzelnen Zustand*** bestimmt, bei Shannon für eine **Gesamtheit**.
Die Clausius-Entropie ist bei allen bekannten Vergleichen nicht einbezogen
Sie ermöglicht offensichtlich auf keine Weise einen sinnvollen Vergleich mit der Shannon-Entropie.

Daten-Art	**Shannon-Entropien**	**Thermodynamische Entropien**
kontinuierlich	$H_K = \mathrm{ld}(n_{AS})$ bzw. $h_{Nutz}(x) = -\int_{-\infty}^{+\infty} p(x) \cdot \mathrm{ld}(p(x)) \cdot dx$	$\Delta S = \Delta Q/T$
diskret/digital	$H = -\sum_{1=1}^{n} p_i \cdot ld(p_i)$	$S = k \cdot \ln(W)$

Gemeinsamkeiten

Trotz aller Unterschiede (s. später) gibt es fünf Gemeinsamkeiten:

1. Beide sind Grundlage und Ursache von wissenschaftlich-technischen Revolutionen.
   Die Thermodynamik (Dampfmaschine) für die erste industrielle Revolution.
   Shannon-Entropie für die Informations-, Nachrichten- und Computer-Technik.
2. Beide nutzen den Logarithmus von Wahrscheinlichkeiten.
3. Beide verweisen auf Austauschmöglichkeiten von einigen Parametern.
4. Es existieren teilweise mehrere Beschreibungen (Formeln) und Inhalte gemäß den beiden obigen Zeilen.
5. Beide Entropien betreffen Theoriegrenzen, die verallgemeinert als Wirkungsgrad interpretiert werden können. Das zeigt die folgende Gegenüberstellung. Dabei ist zu beachten, dass Wirkungsgrade oft recht schwierig zu bestimmen sind, denn sie erfordern:
   a) die entscheidenden Parameter zu finden (z. B. Temperaturen. Wahrscheinlichkeiten),
   b) die genaue Kenntnis der typischen Grundlagen (Kreisprozess).
   c) eine Beschreibung (Formel) der Zusammenhänge so, dass eine Berechnung möglich ist.

Für das Beispiel Programmieren und damit für die Informatik fehlt hierzu noch vieles, und vielleicht ist das sogar – wie beim Halteproblem – prinzipiell nicht erreichbar.

Problemvergleich für Wirkungsgrade

Wieviel Wärme kann in mechanische Energie gewandelt werden?

Carnot-Prozess 1824
Parameter: Temperaturen
Wirkungsgrad:

$$\eta = \frac{T_v - T_u}{T_v}$$

Entropie: $\Delta S = \Delta Q / T$
bzw. $S = k_B \ln(W)$

Wieviel Information kann über den Kanal übertragen werden?

Shannon-Theorie 1940/48
Parameter: Statistik der Signale

Entropie: $H = -\sum_{v=1}^{n} p_v \cdot \mathrm{ld}(p_v)$

Kanalkapazität: $C = B \cdot ld\left(\frac{P_n + P_s}{P_n}\right)$

Was ist der beste Algorithmus zur Lösung einer Aufgabe?

Z. Z. keine Ansätze zur Lösung;
Parameter könnten sein:

- Programmgröße,
- Laufzeit,
- Speicherbedarf

wirkgradH.cdr h. völz 27.1.94/17

Die Unterschiede

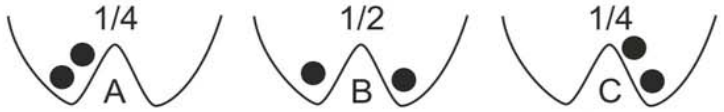
Die **Shannonentropie** fasst immer **alle möglichen** Informer (Zeichen; Z-Information) zusammen.
Sie können dabei diskret (digital) oder z-kontinuierlich auftreten.
Ihre Maßeinheit ist Bit/Zeichen, Anzahl der Amplitudenstufen n_{AS}, oder Wahrscheinlichkeitsverteilung $f(x)$.
Die **Boltzmann-Entropie** betrifft nur **eine** Wahrscheinlichkeit eines Systemzustands
Sie zeigt quasi an, wie weit das System vom Wärmetod entfernt ist.

Ergänzend demonstriert *das folgende Bild* die wesentlichen Unterschiede beider Entropien

Die äquivalenten **Clausius-Entropie** besitzt die Maßeinheit der Energie je Temperatur
Sie entspricht daher gemäß $\Delta S = \Delta Q/T_K$ etwa den reziproken Stromkosten eines Kühlschranks bezüglich seiner Innen-Temperatur T_K: Je tiefer sie ist, desto höher die Stromkosten in kWh.

In eine Urne mit zwei Vertiefungen werden zwei Kugeln geworfen. Es sei gleichwahrscheinlich in welche Vertiefung eine Kugel beim Wurf fällt.
Dann sind 3 Zustände mit den Wahrscheinlichkeiten $W_v = p_v$ möglich.

Kugeln ununterscheidbar



rechts und links unterscheidbar

Boltzmann-Entropie

gilt für nur einen Zustand, besagt, dass B am wahrscheinlichsten ist:

$$S_v = k \cdot \ln(W_v) \quad \text{in J/K}$$

Der Zustand B wird bei freier Entwicklung schließlich angenommen.
Die Geschwindigkeit dorthin hängt von der absoluten Temperatur ab.

diskrete (digitale) Shannon-Entropie

Bestimmt die gemittelt minimal möglichen Bit/Zeichen
für eine ideale Codierung im Präfixcode an

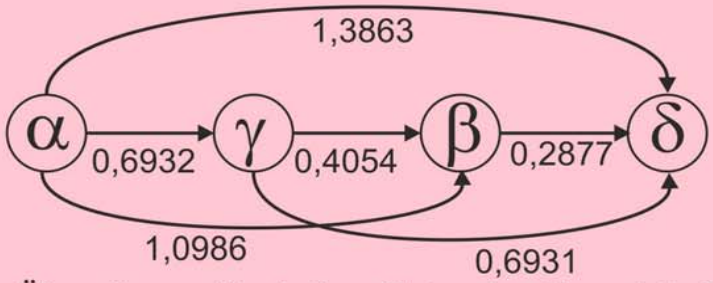
$$H = -\sum_{v=1}^{n} p_v \cdot \mathrm{ld}(p_v) \quad \text{in Bit/Zeichen}$$

Ergibt nur einen Wert für das Ensemble aller Zeichen von $H = 1{,}5$,
anstatt der verschiedenen, einzelnen S_v.

Ein anderes statistisches System besitzt 4 Zustände mit unterschiedlichen Wahrscheinlichkeiten:

α: 0,4 β: 0,3 γ: 0,2 δ: 0,1

$-S_A/k = \ln(0{,}4) \approx 0{,}9163$ $-S_B/k = \ln(0{,}3) \approx 1{,}2040$
$-S_C/k = \ln(0{,}2) \approx 1{,}6094$ $-S_D/k = \ln(0{,}1) \approx 2{,}3026$



Mögliche Übergänge mit relativer Entropie-, Energieänderung.

H = 1,8644 Bit /Zeichen

Entspricht dem möglichen, idealen Wirkungsgrad
für eine Übertragung

Entropie_thermodynamisch.cdr h. völz 19.11.01/21.7.06

Ursachen für die falschen Aussagen

Leider gab bereits Wiener Anlass für Falschaussagen mit dem etwas unklarem Zitat [Wie48], S. 38:

> *„Der Begriff des Informationsgehaltes berührt in natürlicher Weise einen klassischen Begriff in der statistischen Mechanik: den der Entropie. Gerade wie der Informationsgehalt eines Systems ein Maß des Grades der Ordnung ist, ist die Entropie eines Systems ein Maß des Grades der Unordnung; und das eine ist einfach das Negative des anderen.“*

Wahrscheinlich hat Wiener aus dieser Sicht Shannon den Begriff Entropie für die Formel vorgeschlagen.
Aber Ordnung und Unordnung sind immer subjektive Begriffe sind, verlangen wonach etwas geordnet ist.
Bei der Boltzmann-Entropie dürfte Wiener darunter den Endzustand des Wärmetodes (s. o.) gemeint haben.
Bei der Shannon-Entropie das Maximum bei gleichwahrscheinlichen Zeichen. (jede Ordnung belanglos).
So wird die Aussage, das Eine sei das Negative vom Anderen, unverständlich bis falsch.
Schließlich glauben mehrere folgern zu müssen, dass das Minuszeichen inhaltlich nötig sei,
Doch es nur deshalb erforderlicht ist, weil der Logarithmus einer Zahl kleiner als 1 ist
Ein Beispiel von Brillouin und anderen ist der missbrauchten Begriff der Neg-Entropie. Siehe auch [Zeh05].
Es entfällt auch die falsche Folgerung, dass Boltzmann- und Shannon-Entropie zusammen Null ergeben,
also die eine auf Kosten der anderen wächst.
Weiter sei hierzu lediglich noch die falsche Aussage von Mainzer [Mai16], S. 55 zitiert:

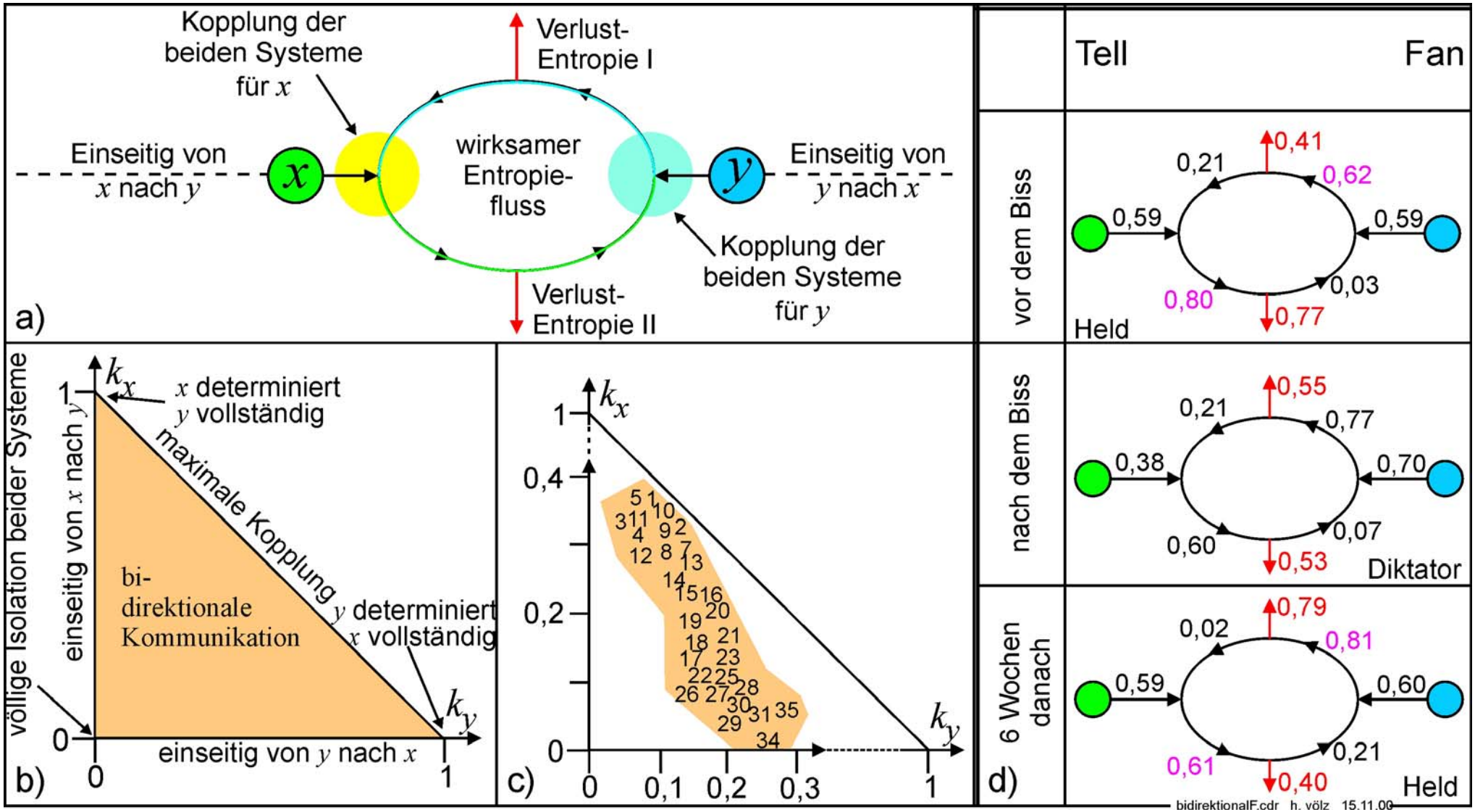
> *Die formale Übereinstimmung der Entropie S eines thermodynamischen ist offenbar. In Shannons Informationsmaß H (...) wird die informationstheoretische Konstante k durch die physikalische Boltzmann-Konstante ersetzt. Man muss dabei allerdings beachten, dass Boltzmanns Entropiebegriff zunächst nur für isolierte Systeme im thermischen Gleichgewicht definiert ist.*

Schließlich Landauers falsche Behauptung: Löschen eines Speichers wird Energie gewonnen [Bet88].

Bidirektionale Entropie

Die übliche Shannon-Theorie gilt nur beim einseitig gerichteten Informationsfluss Sender → Empfänger.
Marko gelang es 1965, die Theorie auf den bidirektionalen Fall zu erweitern [Mar66].
Neuberger erweiterte das Modell auf die Kommunikation beliebig vieler Partner [Neu69], [Neu70].
Dan treten für jeden Teilnehmer 8 Informationsflüsse (Entropien) auf:
4 bilden eine „Ellipse“, 2 betreffen Verluste und 2 weitere gehen zu bzw. von den Partnern *x* bzw. *y* aus.
Zu jedem Partner gehört ein Gesamtfluss, aber nur ein Teil (Trans-Information) gelangt zum Partner.
Es gibt immer Verluste, u. a. durch Übertragungsverluste und Störungen
Deshalb wurden Kopplungen eingeführt. Die Entropie-Flüsse bestimmen die Kopplungsfaktoren k_x und k_y
Aus beiden Kopplungsfaktoren kann das Kopplungsdiagramm aufgestellt werden.
Ist ein Kopplungsfaktoren Null, dann liegen die Werte auf der k_x- bzw. k_y-Achse: einseitige Kommunikation.
Im Ursprung $k_x = k_y = 0$ sind beide Systeme völlig entkoppelt
Andererseits muss immer $k_x + k_y \leq 1$ gelten. bei 1 existiert die maximal mögliche Kopplung beider Systeme.

Das Diagramm



Anwendung Mayer

Er untersuchte hierzu das Verhalten von zwei Totenkopfaffen im Zoo untersucht [May78].
Mit dem üblichen 5-Zeichen-Repertoire aus:

Nebeneinandersitzen, Weggehen, Imponieren, Bedrängen und Sonstiges.

Es gab 217 Beobachtungen von jeweils 15 Minuten Dauer, Nicht alle eingezeichnet
Im Mittel ergab sich eine Gesamt-Entropie von etwa 1,1 Bit/Zeichen.
Besonders aufschlussreich ist ein Dominanzwechsel zwischen den beiden Affen *Tell* und *Fan*
durch einen Biss: Beobachtungszahlen 1 bis 35 und Informationsflüsse in (d).
Zunächst ist *Tell* das dominante Tier. Er ist der „Held“ und hat im Wesentlichen das „Sagen“ (0,80 : 0,62).
Er benutzt die größere Gesamt-Entropie und nimmt mehr Trans-Information (0,21 : 0,03) als Fan auf.
Dann biss *Fan* den *Tell* und wurde dadurch zum „Diktator“.
Kurz danach bestimmt er, wie *Tell* sich zu verhalten hat (0,70 : 0,38).
Etwa sechs Wochen später hat *Fan* die dominante Position mit den dazugehörenden Daten eingenommen.
Auffällig ist an diesem Beispiel, dass ein Diktator mehr Information aufnimmt als er abgibt!
Kontrolle in der Politik?!

Bongard-Weiß-Entropie 1

Sie benutzt sowohl die (üblichen) **objektiven** und die **subjektiven** Wahrscheinlichkeiten.
Ihre Einführung kann auch gut zur Erklärung der Entropieformel dienen.
Für ld(p_v) sind entsprechend der Tabelle m Speicherzellen und die Belegungen als n Zeichen nötig.
Es ist das Zeichen 0110 (= 6) rot hervorgehoben.

Speicher-plätze	zulässige, herstellbare Zeichen Realisiete Zeichen
1	**0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1**
2	**0 0 0 0 1 1 1 1 0 0 0 0 1 1 1 1**
3	**0 0 1 1 0 0 1 1 0 0 1 1 0 0 1 1**
4	**0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1**
m	$n= 2^m$

m Speicherzellen ermöglichen n darstellbaren Zeichen
Es besteht der Zusammenhang m = ld(n).
Er wurde schon 1928 von Hartley als typischer Zusammenhang für die Information betrachtet.

In der Entropie-Formel

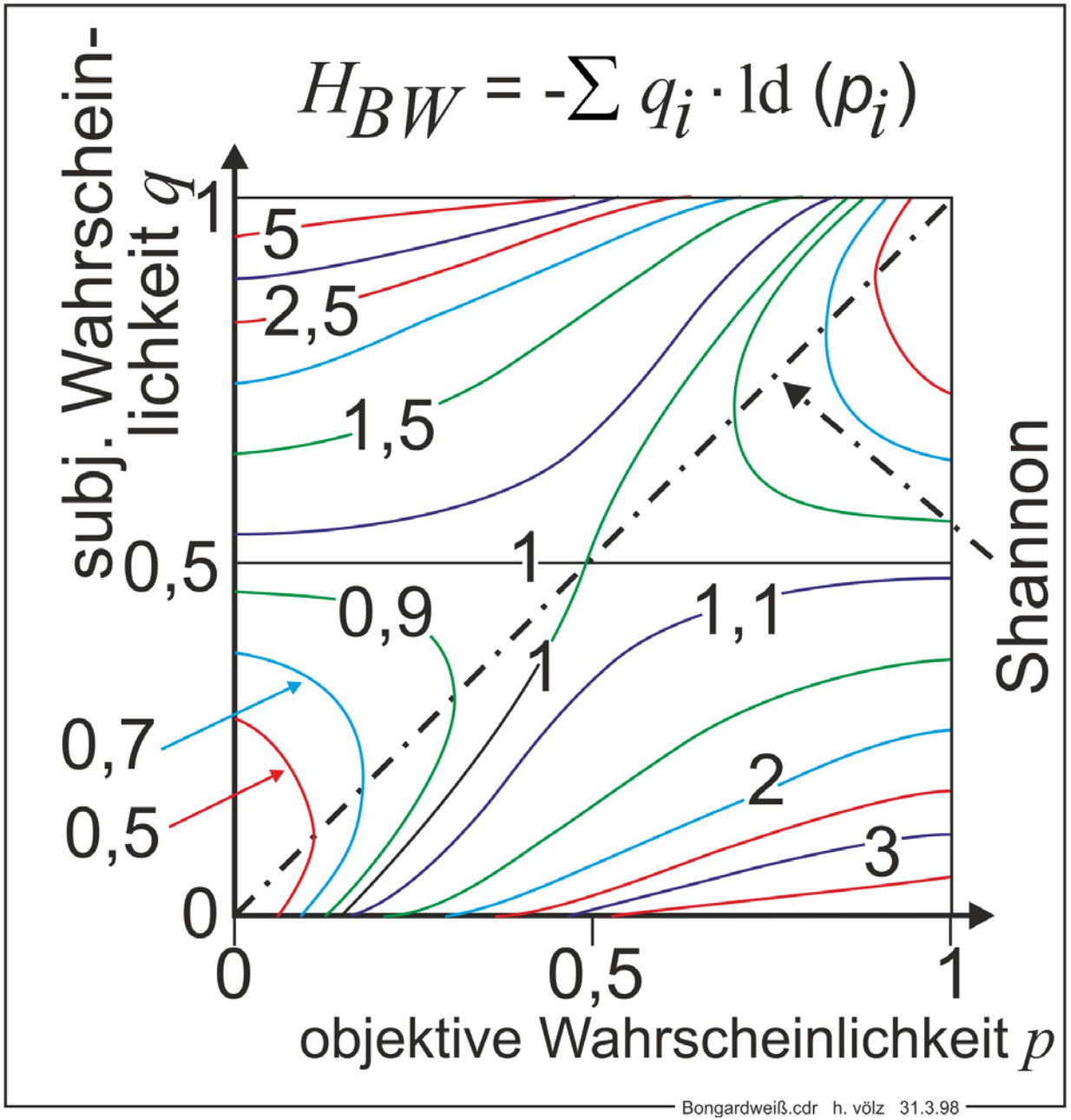
$$H = -\sum p_v \cdot ld(p_v)$$

kann das vorn stehende p_v als ein in der Statistik üblicher Bewertungsfaktor angesehen werden.
Dann ist es nahe liegend genau hier die subjektive Wahrscheinlichkeit q_v zu nutzen.
Das führt zur Bongard-Weiß-Entropie [Bon63]

$$H_{BW} = -\sum q_v \cdot ld(p_v)\,.$$

Ihre grafische Darstellung zeigt das folgende Bild.

Bongard-Weiß- Entropie 2

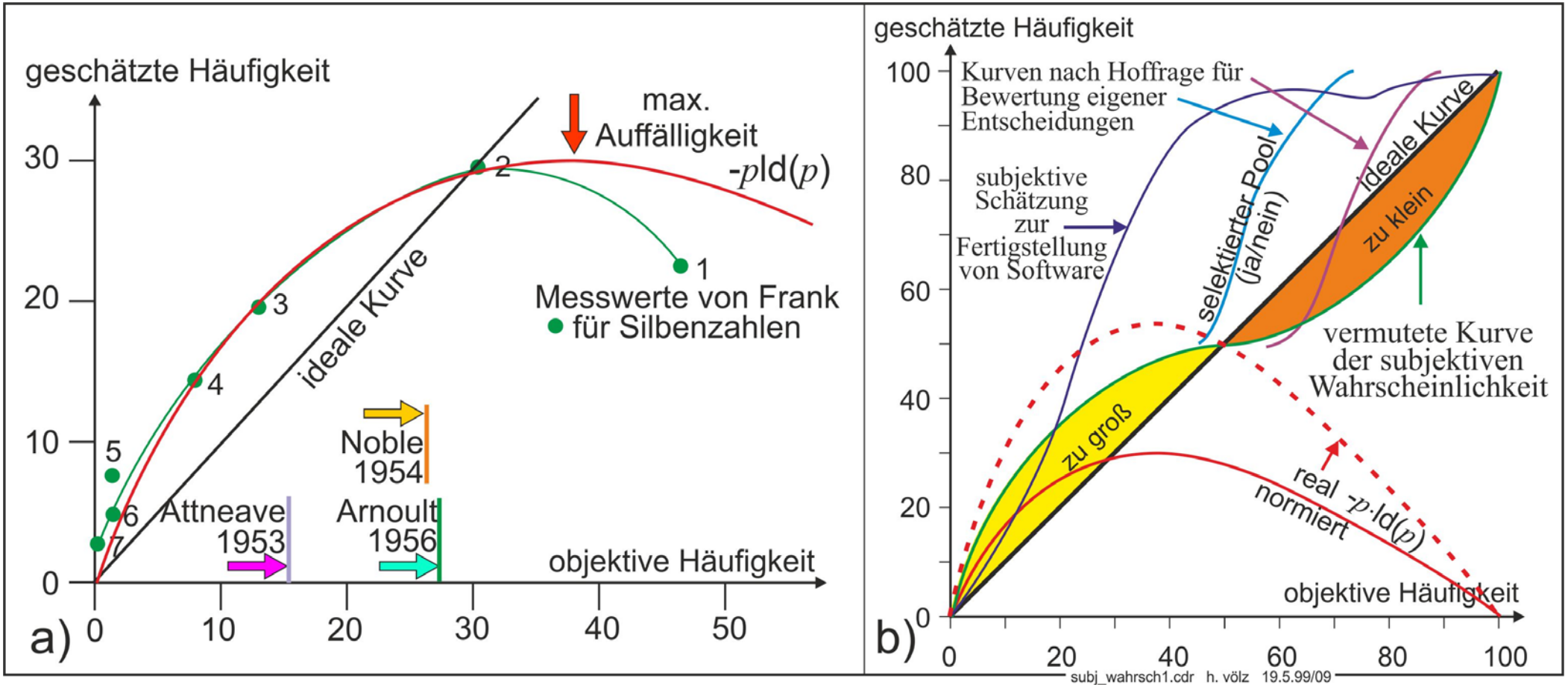


Die subjektive Wahrscheinlichkeit

Der Zusammenhang zwischen objektiver und subjektiver Wahrscheinlichkeit ist nur grob bekannt:
Wenn ein Gewinn lockt, sind wir spontan konservativ, werden aber zu Abenteurern, sobald Verlust droht.
Meist schätzen wir kleine Wahrscheinlichkeiten immer zu groß ein. Wer würde sonst schon Lotterie spielen?
Wohl wenige nur aus Freude! Dagegen werden große Wahrscheinlichkeiten meist zu klein eingeschätzt.
Groß ist der Unterschied, ob vor einer Operation eine Sterberate 7% oder eine Überlebenschance von 93%.
Solche Zusammenhänge in Spielhallen durch dichtes Aufstellen vieler einarmiger Banditen ausgenutzt.
Obwohl auf jedem einzelnen nur selten gewonnen wird, erzeugen die Automaten einen beachtlichen Lärm.
Der dann vermuten lässt, dass oft gewonnen wird.

Ein anderes Beispiel betrifft ein Treffen von 24 Personen.
Mit 50 % Wahrscheinlichkeit haben dann zwei Teilnehmer an einem gleichen Tag Geburtstag
Experimentell gibt es deshalb nur sehr wenige Messwerte für die subjektive Wahrscheinlichkeit.
Sie werden meist recht unsicher geschätzt wird und sind daher schwer zu bestimmen
Einige Beispiele zeigt das Bild 48a auf der nächsten Seite.
Die Werte 1 bis 7 gehen auf Untersuchungen von Frank zurück.
Er ließ die erste Textseite aus Musils „Mann ohne Eigenschaften“ lesen
und dabei für jede Silbe mit dem Bleistift auf den Tisch klopfen.
Danach fragte er völlig unerwartet, wie häufig bei den Wörtern die verschiedenen Silbenzahlen auftraten.
[Fra69] Bd 2. S. 128. Es könnte sabei ein Zusammenhang mit der Auffälligkeit $-p \cdot \mathrm{ld}(p)$ bestehen.
Zusätzlich sind einige Sonderfälle von subjektiven Einschätzungen ergänzt.
Z. B. wird der Zustand der Fertigstellung von Software immer viel zu hoch angegeben

Beispiele subjektiver Wahrscheinlichkeit

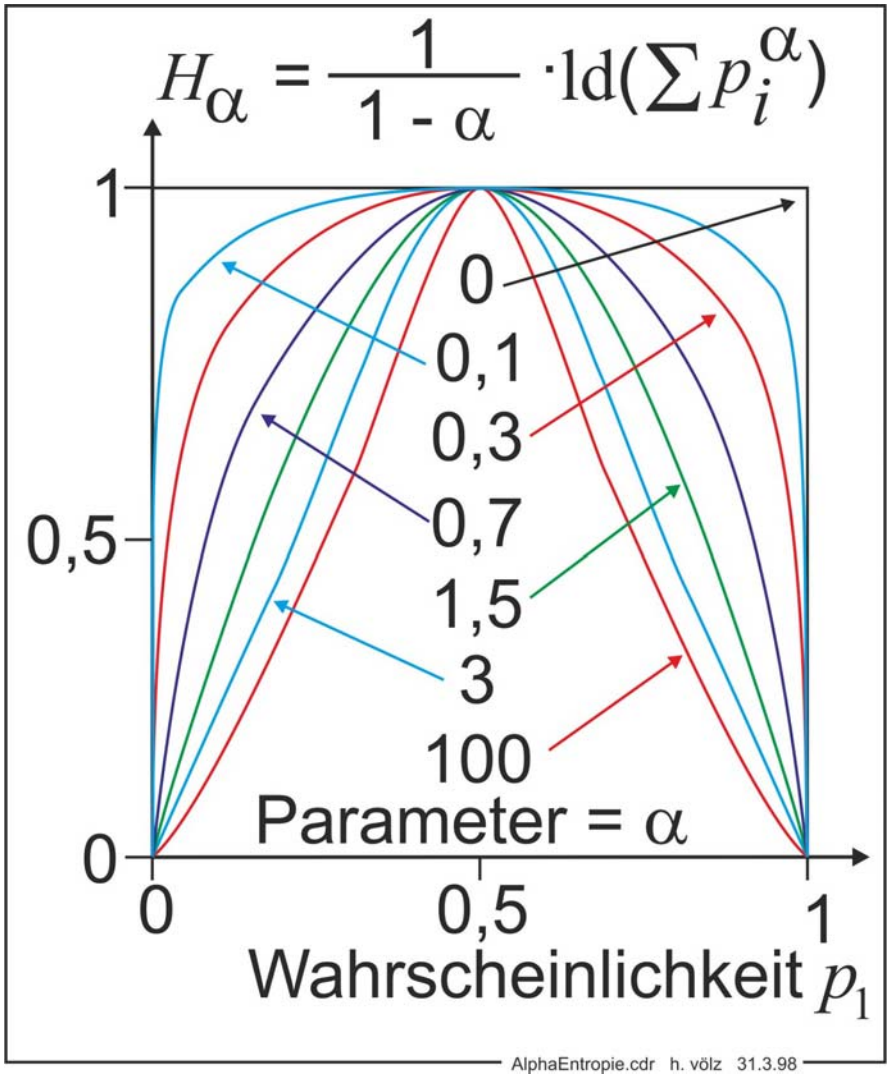


Renyi-Entropie

Ab den 1960er Jahren wurde umfangreich versucht, die Entropie-Formel überall anzuwenden.
Dabei blieben oft die typischen Voraussetzungen unbeachtet. Zuweilen wurde nur Gleichverteilung benutzt.
So kam es zuweilen zu seltsamen Widersprüchen.
Daher wurde nach einer **allgemeinen Formel** gesucht.
Es entstanden mehrere neue Formeln.
Davon hat nur die α-Entropie 1962 mit einer frei wählbaren Konstanten α von Renyi gewisse Bedeutung [Ren82]:

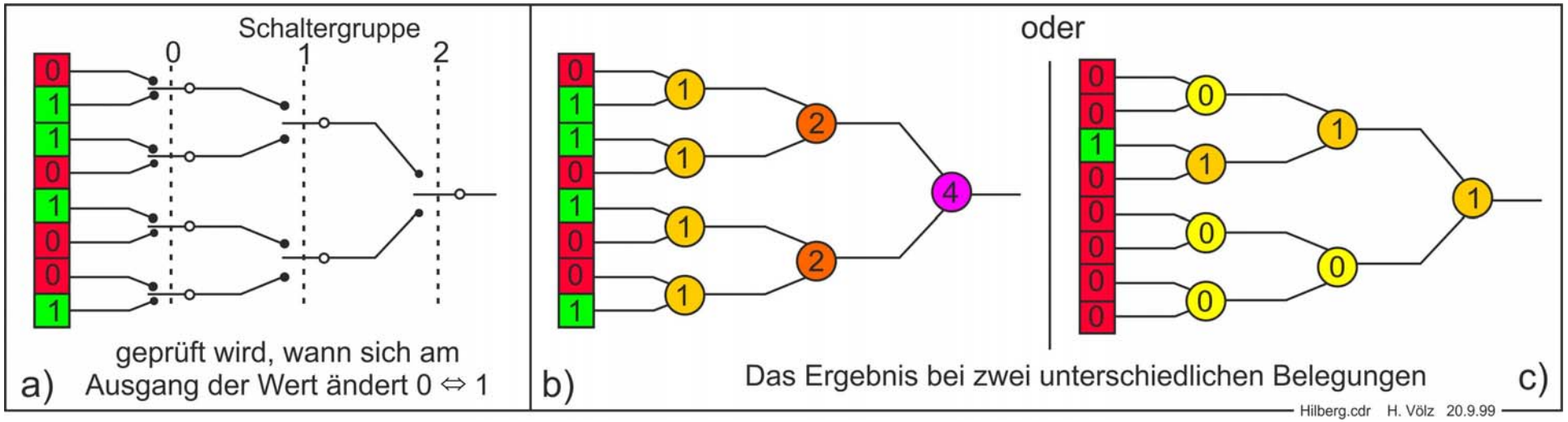
$$H_\alpha = \frac{1}{1-\alpha} \cdot \mathrm{ld}\left(\sum_{i=1}^{n} p_i^\alpha \right).$$

Bei nur zwei Zeichen ergeben sich die Verläufe des Bildes
Für α = 1 entsteht der Verlauf der Shannon-Entropie.
Eine Anwendung der Formel oder inhaltliche Begründung sind nicht bekannt.



Deterministische Entropie 1

Eine total andere, nämlich eine streng deterministische Entropie hat Hilberg eingeführt [Hil84, 84, 00]
Zuweilen nennt er sie auch funktionale Komplexität. Sie sieht von Wahrscheinlichkeiten und Zufall ab.
Er geht davon aus, dass auch feste Strukturen, Gebilde, Zeichen usw. einen Informationswert besitzen
Das widerspricht dem p_i in der Shannon-Formel.
Deshalb verlagert den Zufall auf umschaltbare Schaltergruppen.
So geht er z. B. von der Bit-Folge 01101001 aus.
Gemäß dem folgenden Bild wird sie Bit-weise in übereinander liegenden Kästchen eingetragen.
Dann werden die Kästchen mit einer binären Schalterhierarchie verbunden und statistisch umgeschaltet
Bleibt bei einer Schalterbetätigung der Ausgangswert erhalten, bekommt der jeweilige Schalter 1 sonst 0.
Mit ihrer Addition wird die Statistik einer Signalquelle auf die Schalterstellungen übertragen.
Für zwei Belegungen der Kästchen gelten b) und c).
So lassen sich indirekt „Entropie“-Werte auch für binäre Bauelemente bestimmen



Deterministische Entropie 2

Es ergeben sich folgende Werte:

Bauelement	**„Entropie“**	**Bauelement**	**„Entropie“**
AND	1	Master-Slave-Flipflop	3.31
n-faches XOR	2^n- 1	Speicher mit *m*-Wort, *n* Adressen	$m \cdot 2^n$
n-stufiger Decoder	2^n	*n*-Bit/Addition	$n \cdot 2^n$
RS-Flipflop	1.97	*n*-Bit-Multiplikation	$n \cdot (2^n-1)$

Für *n*-Bit-Folgen kann diese Entropie einen Wert zwischen 0 und *n*/2 Bit annehmen.
In das Ergebnis geht dabei das Gewicht der Folge (Anzahl der **1**) ein.
Für eine Folge aus *n* Zeichen kann das Maximum nur dann erreicht werden, wenn die Anzahl der **1**-Werte gleich *n*/2 beträgt.
Für *m* **1**-Werte oder *m* **0-**Werte mit $m \leq n/2$ beträgt das Maximum *m*.
Weiter geht die Anordnung der **1** in die Folge ein.
Wenn sie alle – vom Beginn oder vom Ende der Folge her – unmittelbar aufeinander folgen, dann kann die deterministische Entropie unabhängig von der Länge der Folge nur den Wert 1 erreichen.
Etwa 1987 konnte Hilberg zusätzlich zeigen:

- Die deterministische und statistische Entropie konvergieren für Bit-Ketten mit $n \rightarrow \infty$ gegen einen gemeinsamen Wert.
- Seine Betrachtungen stehen im Einklang mit übergeordneten mathematischen Arbeiten von Chaitin [Cha75].
- Es besteht ein gewisser Zusammenhang mit der ***Kolmogoroff***-Komplexität des folgenden Abschnitts.

Kolmogoroff-Entropie

Sie wurde 1959 eingeführt und ist ein Maß für den Grad von Chaos,
hat aber kaum inhaltlichen Bezug zu den anderen Entropien.
Lediglich die zu ihr gehörende Formel hat Ähnlichkeit.
Ihr Kennwert K gibt die Wahrscheinlichkeit dafür an, welche benachbarten Orte im nächsten Schritt erreicht werden. Dabei können unterschieden werden:

- K = 0 deterministische Bewegung,
- K = ∞ vollkommen stochastische Bewegung,
- K > 0 und endlich; deterministisches Chaos; hier existiert ein seltsamer Attraktor.

Leider ist keine Normierung bekannt. Mit K ist aber eine Aussage zur möglichen Vorhersagezeit T gegeben:

$$T \sim \frac{1}{K} \cdot \log\left(\frac{1}{l}\right)$$

Darin ist *l* die geforderte Genauigkeit.
Es ist auch eine Anwendung für Objekte möglich.
Sie sind dann durch die kleinstmögliche Beschreibung bestimmt.

Carnap-Entropie 1

Auch Sie hat kaum Bezug zur Shannon-Entropie.
Sie war der Versuch die Semantik quantitativ in die Informationstheorie einzubinden [Car52] [Car54]
Schließlich hat sich auch Carnap wieder von ihr getrennt. [Völ91].
Carnap hat die Bedeutung von Information untersucht und dabei die Extension und die Intension eingeführt:

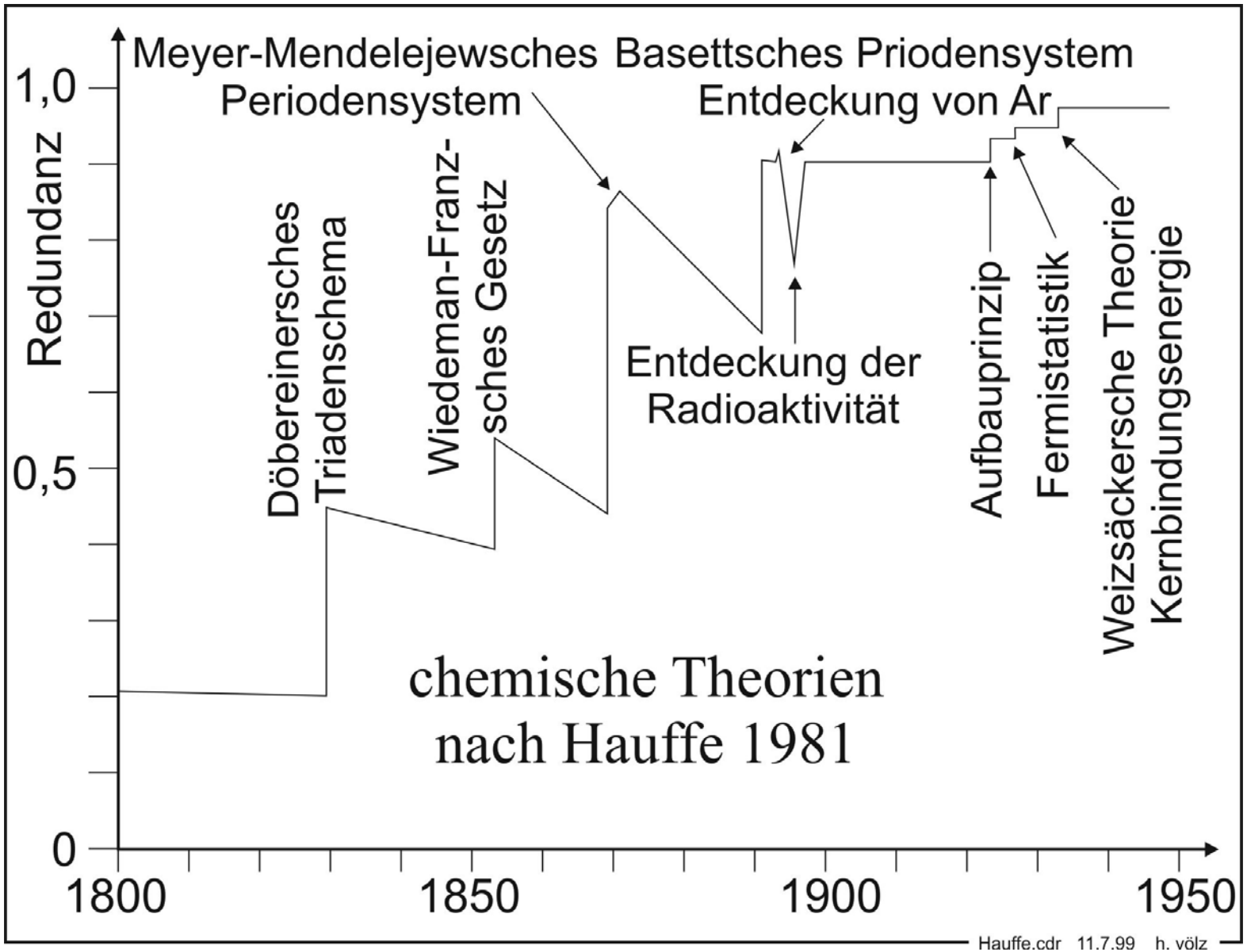
Extension: Objektbezug, Sachbezug etwa Begriffsumfang; gilt für Individuen und Klassen, auf die das Prädikat zutrifft; extensionale Beziehungen untersucht die Prädikatenlogik.

Intension: Abbildbezug, etwa Begriffsinhalt; erfasst die Eigenschaft der Objekte mit einem Prädikat; ist nur im Gedächtnis vorhanden.

Am Periodensystem der Chemie hat Hauffe die Brauchbarkeit/Gültigkeit von Theorien untersucht [Hau81].
Eine gute Theorie soll möglichst viele experimentelle Fakten (hier nur ausgewählte Eigenschaften der Elemente) exakt beschreiben und weiteres dazu voraussagen.
Sie wird dabei immer kompakter und macht so auch alte Aussagen redundant.
Hauffe wählte für seine Analyse sechs Eigenschaftspaare aus:

schwer	↔	leicht
tiefschmelzend	↔	hochschmelzend
wärmeleitend	↔	wärmeisolierend
elektrisch leitend	↔	elektrisch isolierend
leicht ionisierbar	↔	schwer ionisierbar
stabil	↔	instabil.

Carnap-Entropie 2



Entropie-Axiomatik nach Feinstein

Für die Shannon-Theorie ist eine axiomatische Herleitung der Entropieformel wünschenswert:

$$H = -\sum_{i=1}^{n} p_i \cdot {}_x\log(p_i)$$

Es existiert auch eine von Feinstein [Fan66]. Dabei werden vier Axiome gewählt.

1. $H(p, 1\text{-}p)$ ist ***stetig*** in p mit $0 \le p \le 1$ .....2 Zeichen
2. $H(0{,}5, 0{,}5) = 1$ ..... ***Normierung***
3. $H(p_1, p_2, ..., p_n)$ ist unabhängig von der ***Reihenfolge****n* Zeichen
4. ***Verfeinerungstheorem***: Zerlegung $p_n = q_1 + q_2 n \rightarrow n+1$ Zeichen
dann: $H(p_1, p_2, ..., q_1+q_2) = H(p_1, p_2, ..., p_n) + p_n \cdot H(q_1 / p_n + q_2 / p_n)$

Leider gibt es aber für das 4. Axiom keine einsichtige Begründung.
Daher ist die folgende Ableitung auch nur bedingt zulässig:

- $H(1) = 0$: **Gewissheit**
- n **gleichwahrscheinliche** Symbole: $p_i = 1/n$: $H(n_1) > H(n_2)$ wenn $n_1 > n_2$
- Zwei **unabhängige** Urnenversuche: $p(A; B) = p(A) \cdot p(B) \rightarrow H(A; B) = H(A) + H(B)$
- Das erfordert den **Logarithmus**, dessen **Basis** noch unbestimmt ist.
- **Ein Versuch** mit $p = 1/n \rightarrow h = c \cdot {}_x\log(1/n)$ folglich bei Summierung Faktor p notwendig.
- $h = p \cdot {}_x\log(1/n) = -p \cdot {}_x\log(p)$
- Durch Summierung folgt dann die bekannte Entropieformel (oben).
- Zur Normierung (2. Axiom) kommt nur der **binäre Logarithmus** $x = 2$ in Betracht.