INFORMATIONSTHEORIE

Skript zur Vorlesung

Institut für Theoretische Nachrichtentechnik
und Informationsverarbeitung
Universität Hannover

Prof. Dr.–Ing. H.G. Musmann

2. Auflage Frühjahr 1998

Inhaltsverzeichnis

1. Einführung 5
1.1 Quellenmodelle und Quellencodierung 6
1.2 Kanalmodelle und Kanalcodierung 9

2. Quellenmodelle 13
2.1 Diskrete gedächtnislose Quelle 13
2.2 Diskrete gedächtnisbehaftete Quelle 20

3. Redundanzreduzierende Codierung 25
3.1 Codierung diskreter gedächtnisloser Quellen 25
3.2 Theorem der Quellencodierung 29
3.3 Codierungsalgorithmus von Huffman 32
3.4 Codierung diskreter gedächtnisbehafteter Quellen 35
3.5 Anwendungsbeispiel 37

4. Kanäle 41
4.1 Fehlerstrukturen realer Kanäle 41
4.2 Diskreter, gedächtnisloser Kanal 45
4.3 Wechselseitiger Informationsgehalt und Transinformationsgehalt diskreter Ensembles 47
4.4 Wechselseitiger Informationsgehalt und Transinformationsgehalt kontinuierlicher Ensembles 49
4.5 Kanalkapazität diskreter, gedächtnisloser Kanäle 55
4.6 Kanalkapazität kontinuierlicher, gedächtnisloser Kanäle 63
4.7 Anwendungsbeispiel 66

5. Kanalcodierung 73
5.1 Allgemeine Grundbegriffe 73
5.2 Prinzipien der Blockcodierung 75
5.3 Kanaldecodierung 78
5.4 Theorem der Kanalcodierung 82

6. Irrelevanzreduzierende Codierung 87
6.1 Grundlagen der Rate–Distortion–Theorie 87
6.2 Rate–Distortion–Funktion diskreter Quellen 95
6.3 Rate–Distortion–Funktion kontinuierlicher Quellen 99

6.4 Anwendungsbeispiele 111
6.4.1 Amplitudenquantisierung mit minimaler Fehlervarianz 111
6.4.2 Quantisierung mit minimaler Ausgangsentropie 120
6.4.3 Pulscodemodulation (PCM) 124
6.4.4 Differenz–Pulscodemodulation (DPCM) 128

Literaturverzeichnis 135
Stichwortverzeichnis 137

1. Einführung

Die Informationstheorie dient der mathematischen Behandlung von Nachrichtenübertragungssystemen. Sie gestattet es, verschiedene Übertragungsverfahren zu vergleichen und die einzelnen Komponenten eines Übertragungssystems zu optimieren. Ein Nachrichtenübertragungssystem kann ohne wesentliche Einschränkung durch das Blockschaltbild gemäß Bild 1.1 dargestellt werden.

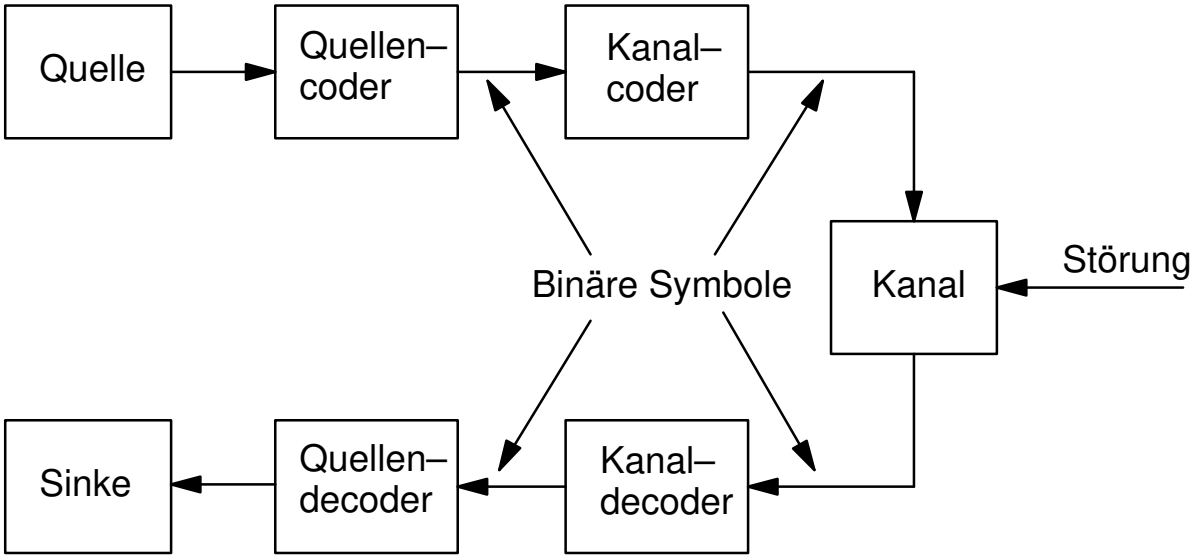


Bild 1.1 Blockschaltbild eines digitalen Nachrichtenübertragungssystems

Der Quelle wird die Erzeugung der Nachricht zugesprochen. Dabei können alle Kombinationen von wert– oder zeitkontinuierlichen bzw. wert– und zeitdiskreten Quellensignalen erzeugt werden. Liegen die Quellensignale in wertdiskreter und zeitdiskreter Form vor, z.B. als PCM–Signal, spricht man von einer diskreten Quelle.

Der Quellencoder wandelt das Quellensignal in eine Folge binärer Zeichen um. Dabei kann die Bedingung gestellt werden, das Eingangssignal mit einer möglichst geringen Anzahl binärer Zeichen je Zeiteinheit darzustellen.

Der Kanalcoder hat die Aufgabe, den binären Nachrichtenfluß gegen Kanalstörungen zu sichern, damit der binäre Nachrichtenfluß beim Kanaldecoder möglichst fehlerfrei rekonstruiert werden kann. Dazu können fehlererkennende und fehlerkorrigierende Codes angewendet werden.

Zur Übertragung der ebenfalls binären Codesymbole des Kanalcoders muß ein diskreter, binärer Kanal zur Verfügung stehen. Kontinuierliche Kanäle können mit Hilfe einer Zusatzeinrichtung, dem Modulator und Demodulator, zu einem diskreten Kanal ergänzt werden. Der Modulator setzt die Binärsymbole in kontinuierliche Signale um, die den physikalischen Eigenschaften des kontinuierlichen Kanals angepaßt sind, damit eine möglichst wirtschaftliche Ausnutzung des Kanals erreicht wird. Auf den kontinuierlichen Kanal wirken Störungen ein, wie thermisches Rauschen, Nebensprechen oder Impulsstörungen. Das gestörte übertragene Signal wird am Kanalausgang demoduliert und damit wieder in eine Folge von Binärsymbolen umgewandelt. Infolge der Störungen auf dem Kanal treten am Ausgang des diskreten binären Kanals Übertragungsfehler auf. Im Kanaldecoder wird der gesicherte, fehlerbehaftete binäre Nachrichtenfluß in einen möglichst fehlerfreien, ungesicherten, binären Nachrichtenfluß decodiert.

Der Quellendecoder fügt im Quellencoder entfernte redundante Anteile des binären Nachrichtenflusses wieder hinzu und wandelt den binären Nachrichtenfluß in eine für die Sinke annehmbare Form um. Die Nachrichtensinke ist der Verbraucher der Nachricht; dafür kann ein Mensch, ein Rechner oder ein beliebiges Nachrichtenverarbeitungssystem in Frage kommen.

Zur theoretischen Behandlung eines Nachrichtenübertragungssystems müssen Quelle, Kanal und Sinke durch mathematische Modelle beschrieben werden. Die mathematisch zugänglichen Modelle sind im allgemeinen nur eine grobe Näherung der physikalischen Realität. Sie werden mit den Mitteln der Wahrscheinlichkeitslehre und Systemtechnik beschrieben. Ausgehend von diesen Modellen ermöglicht die Informationstheorie vor allem eine Optimierung der Codierung und Decodierung. Darüber hinaus zeigt sie die Grenzen des Gesamtverhaltens eines Nachrichtenübertragungssystems auf. Die grundlegenden Arbeiten der Informationstheorie gehen auf C.E. Shannon (1948), [7] zurück. Zur Vertiefung des Stoffes sei auf das Lehrbuch von R.G. Gallager (1968), [3] hingewiesen.

Es wird zunächst eine kurze Übersicht über die grundlegenden Konzepte der Informationstheorie und der bei einer Anwendung zu lösenden Probleme gegeben.

1.1 Quellenmodelle und Quellencodierung

Als Quellenmodelle der Informationstheorie dienen Zufallsprozesse. Das einfachste Modell ist die diskrete, gedächtnislose Quelle. Diese Quelle verfügt über einen Symbolvorrat von K Symbolen $a_1,\ ...,\ a_k,\ ...,\ a_K$, die mit den Wahrscheinlichkeiten $P(a_1),\ ...,\ P(a_K)$ auftreten. Sie erzeugt eine Nachricht, indem sie aus dem

Symbolvorrat zeitlich sequentiell Symbole auswählt und zu einer Symbolfolge statistisch unabhängiger Symbole zusammenstellt. Im Gegensatz zur gedächtnislosen Quelle sind die Symbole einer Symbolfolge der gedächtnisbehafteten Quelle statistisch abhängig. Dadurch wird die theoretische Behandlung erschwert. Die kontinuierlichen Signale einer nicht–diskreten Quelle müssen zur Umwandlung in binäre Symbole zunächst abgetastet und die Abtastwerte in ihrer Amplitude quantisiert werden. Diese Umsetzung verursacht immer einen Quantisierungsfehler. Der maximal zulässige Quantisierungsfehler wird im allgemeinen durch ein vom Empfänger der Nachricht vorgegebenes, einzuhaltendes Gütekriterium beschränkt. Die quantisierten Abtastwerte können anschließend mittels binärer Symbole codiert werden.

Eine Nachricht kann gemäß Bild 1.2 in Form eines Mengendiagramms dargestellt werden. Eine vertikale Gerade trennt den redundanten (bekannten) Teil vom nicht redundanten (unbekannten) Teil, eine horizontale Gerade trennt den relevanten (bedeutenden) Teil vom irrelevanten (nicht bedeutenden) Teil.

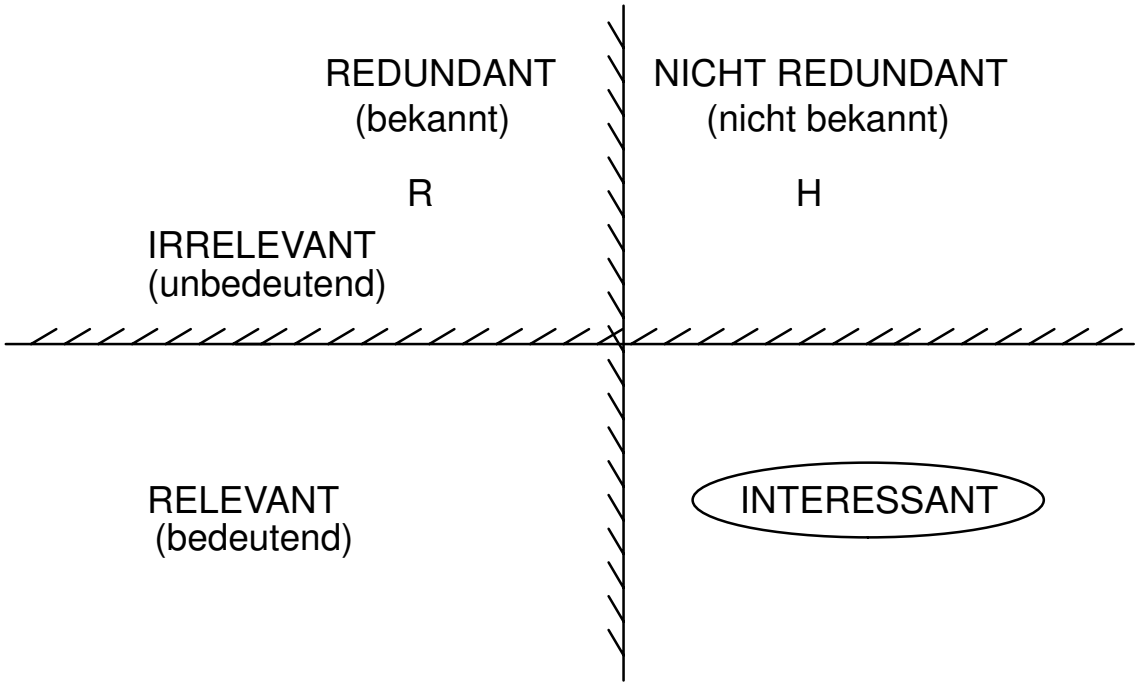


Bild 1.2 Nachrichtenebene nach Schouten

Nur der für den Nachrichtenempfänger interessante Teil braucht übertragen zu werden. Aufgabe der Quellencodierung ist es daher, die Redundanz und Irrelevanz einer Nachricht zu eliminieren. Die einzelnen Teile der Nachricht können mit Hilfe der Informationstheorie mathematisch erfaßt werden.

Jede diskrete Quelle besitzt einen Entscheidungsgehalt H_0. Bezogen auf ein einzelnes Quellensymbol a_k gibt dieser an, wie viele Binärentscheidungen zur Auswahl von a_k aus einem Vorrat von K Symbolen erforderlich sind. Er beträgt

$$H_0 = \log_2 K \ ,$$

wenn der Entscheidungsgehalt in bit je Symbol gemessen wird. Zur Beschreibung einer Nachricht sind maximal H_0 bit je Symbol erforderlich. Der Entscheidungsgehalt setzt sich aus dem Informationsgehalt H und der Redundanz R zusammen, d.h. $H_0 = H + R$. Der Informationsgehalt $I(a_k)$ eines einzelnen Symbols a_k ist durch die Beziehung

$$I(a_k) = - \log_2 \ P(a_k)$$

definiert und wird in der Einheit bit je Symbol gemessen. Je seltener a_k auftritt, umso größer wird nach dieser Definition sein Informationsgehalt. Ein Symbol a_k mit der Wahrscheinlichkeit $P(a_k) = 1$ besitzt demzufolge keinen Informationsgehalt, da $I(a_k) = 0$ wird. Der mittlere Informationsgehalt eines Symbols des Vorrates $a_1, \ldots, a_K$ beträgt somit

$$H = - \sum_{k=1}^{K} P(a_k) \cdot \log_2 \ P(a_k)$$

bit und wird als Entropie bezeichnet.

Das Theorem der Quellencodierung besagt, daß zur informationsverlustfreien Darstellung der Nachrichten einer diskreten, gedächtnislosen Quelle mindestens H bit je Symbol benötigt werden. Die Redundanz R kann durch eine sogenannte redundanzmindernde Codierung eliminiert werden. Dieser Codierungsprozeß ist reversibel und verursacht keinen Informationsverlust.

Wegen der unvermeidbaren Quantisierungsfehler bei der Codierung kontinuierlicher Quellen kann ein Nachrichtenempfänger in diesem Fall keine verlustlose Übertragung, sondern lediglich eine bestimmte Wiedergabegüte fordern, die er durch Angabe eines Gütekriteriums definiert. Dadurch wird neben der redundanzmindernden Codierung auch eine informationsreduzierende Codierung ermöglicht. Die Reduktion von Informationen bringt einen Informationsverlust, der irreversibel ist. Dadurch entsteht eine Verzerrung der übertragenen Nachricht. Dem Quellencoder muß deshalb das Gütekriterium oder ein Maß der zulässigen Verzerrung bekannt sein. Eine Informationsreduktion kann als Irrelevanzreduktion bezeichnet werden, solange die verursachte mittlere Verzerrung $\overline{d}$ die Bedingung

$$\overline{d} \leq d^*$$

einhält und das zulässige Maß d^* nicht überschreitet. Die Irrelevanzreduktion bzw. die informationsreduzierende Codierung wird dann als optimal bezeichnet, wenn sie die größtmögliche Reduktion des gedanklich bereits redundanzfreien Nachrichtenflusses erzielt, ohne das Gütekriterium zu verletzen. Die untere Grenze der Übertragungsrate kann mit der Rate–Distortion–Theorie bestimmt werden. Die Berechnung dieser Grenze ist relativ schwierig; Lösungen sind bisher nur für spezielle Gütekriterien, z.B. den mittleren quadratischen Fehler, bekannt. Der quadratische Fehler erweist sich aber in vielen Fällen als nicht ausreichend, insbesondere dann nicht, wenn die Nachricht in Form akustischer Signale oder Bildsignale einer physiologischen Wahrnehmung zugeführt wird. Redundanz– und Irrelevanzreduktion können auch in einem Codierungsprozeß zusammengefaßt werden.

1.2 Kanalmodelle und Kanalcodierung

Das einfachste Kanalmodell ist der diskrete, gedächtnislose Kanal. Wird dem Eingang eines diskreten Kanals eine Folge von Symbolen des Alphabetes $a_1, \ ..., \ a_K$ zugeführt und entsteht am Ausgang des Kanals eine Folge von Symbolen des gleichen oder eines anderen Alphabetes $b_1, \ ..., \ b_J$, so ist der diskrete, gedächtnislose Kanal dadurch definiert, daß jedes Ausgangssymbol b_j nur von einem entsprechenden Eingangssymbol a_k statistisch abhängig ist. Diese statistische Abhängigkeit wird durch die bedingte Wahrscheinlichkeit $P(b_j|a_k)$, daß b_j am Ausgang erscheint, wenn a_k am Eingang zugeführt wurde, beschrieben. In einem diskreten, gedächtnisbehafteten Kanal hingegen ist ein Symbol b_j der Ausgangsfolge statistisch nicht nur von einem entsprechenden sondern von mehreren Symbolen der Eingangsfolge abhängig.

Eine den physikalischen Kanälen näherkommende Klasse von Kanalmodellen sind die nicht–diskreten Kanäle, deren Eingangs– und Ausgangsinformationen in Form kontinuierlicher Zeitfunktionen vorliegen. Von besonderem theoretischen und praktischen Interesse ist dabei der nicht–diskrete Kanal mit additivem weißen Rauschen als Störung. Sollen über einen nicht–diskreten Kanal binäre Symbole übertragen werden, so ist ein zusätzlicher Modulator und Demodulator erforderlich, wie es in Bild 1.3 gezeigt wird. Mit Hilfe des Modulators und Demodulators kann ein nicht–diskreter Kanal in das Modell eines diskreten Kanals überführt werden.

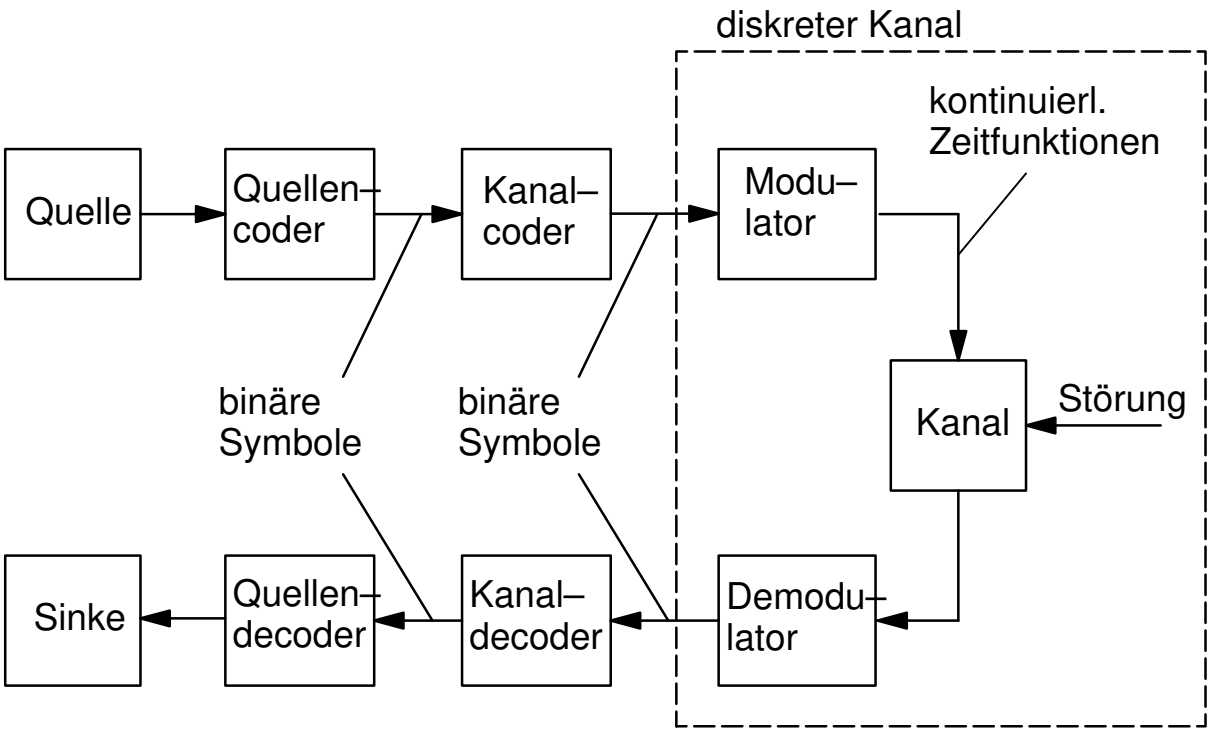


Bild 1.3 Darstellung eines nicht–diskreten Kanals durch einen diskreten Kanal

Die dem Modulator in zeitlicher Folge mit dem Zeitabstand τ zugeführten Symbole a_k werden Symbol für Symbol in Zeitfunktionen $s_k(t)$ der Dauer τ umgesetzt; so erzeugt a_1 die Funktion $s_1(t)$, a_2 erzeugt $s_2(t)$ usw. Der Demodulator wandelt die in einem Zeitintervall empfangene Zeitfunktion wieder in das entsprechende Symbol b_j um. Für den Fall, daß die Störung nur additives, weißes Rauschen ist, handelt es sich bei diesem Modell um einen diskreten, gedächtnislosen Kanal.

Die bedeutendste Größe zur Kennzeichnung eines Kanals ist die Kanalkapazität C. Die Kanalkapazität gibt an, wie groß der maximal fehlerfrei übertragbare Informationsfluß eines Kanals ist, wenn die Information nach der oben definierten Beziehung gemessen wird. Es läßt sich zeigen, daß die Kanalkapazität eines nicht–diskreten Kanals der Kanalkapazität des in Bild 1.3 dargestellten diskreten Kanals bei geeigneter Auslegung des Modulators und des Demodulators beliebig nahe kommt.

Das Theorem der Kanalcodierung besagt, daß es mit Hilfe einer geeigneten Kanalcodierung und Kanaldecodierung möglich ist, die Wahrscheinlichkeit für Übertragungsfehler beliebig klein zu halten, solange einem gestörten Kanal mit der Kanalkapazität C die Symbolfolge einer Quelle mit einer Bitrate $R < C$ zugeführt wird. Die zu übertragende Symbolfolge wird zu diesem Zweck in Blöcke segmentiert.

Jedem Block wird ein Codewort von N Codeelementen zugeordnet. Ist P_e die Wahrscheinlichkeit dafür, daß ein Block falsch decodiert wird, so kann eine Codierung im allgemeinen so gewählt werden, daß P_e die Bedingung

$$P_e \leq \exp[-N \cdot E(R)]$$

erfüllt. Dabei ist $E(R)$ eine Funktion der Bitrate R, abhängig von den physikalischen Eigenschaften des Kanals, aber unabhängig von N. $E(R)$ nimmt mit R ab und bleibt positiv, solange R kleiner als die Kanalkapazität C ist. Ein typischer Verlauf von $E(R)$ ist in Bild 1.4 skizziert.

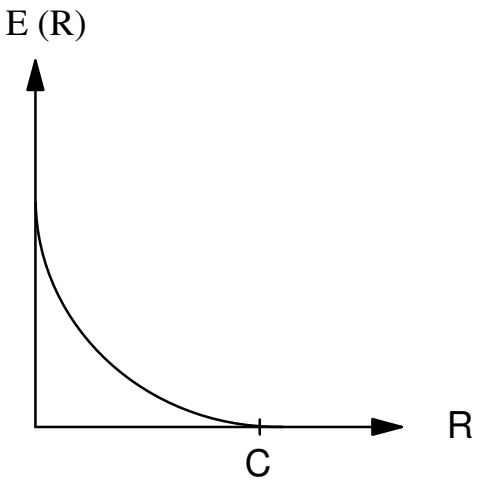


Bild 1.4 Beispiel für einen typischen Verlauf von *E(R)*

Um eine möglichst kleine Fehlerwahrscheinlichkeit P_e zu erreichen, muß das Produkt $N \cdot E(R)$ groß gehalten werden. Dies kann dadurch erreicht werden, daß entweder die Übertragungsrate R klein oder N groß gewählt werden. Mit größer werdendem N wächst der Aufwand der Kanalcodierung, weil die Blocklänge größer wird. Daher erfordert die Optimierung der Kanalcodierung einen Kompromiß zwischen Übertragungsrate, Fehlerwahrscheinlichkeit und Aufwand, der für verschiedene Übertragungssysteme je nach Bedeutung der genannten Faktoren unterschiedlich ausfallen kann.

2. Quellenmodelle

2.1 Diskrete gedächtnislose Quelle

Als Quellenmodelle der Informationstheorie dienen Zufallsprozesse. Das einfachste Modell ist die diskrete, gedächtnislose Quelle. Diese Quelle verfügt über einen Symbolvorrat von K Symbolen $a_1,\ ...,\ a_k,\ ...,\ a_K$, die mit den Wahrscheinlichkeiten $P(a_1),\ ...,\ P(a_K)$ auftreten. Sie erzeugt eine Nachricht, indem sie aus dem Symbolvorrat zeitlich sequentiell Symbole auswählt und zu einer Symbolfolge $u_1,\ ...,\ u_n,\ ...,\ u_N$ zusammenstellt. Jedes u_n dieser Folge steht für ein beliebiges Symbol des Symbolvorrats $a_1,\ ...,\ a_K$. Im Fall der gedächtnislosen Quelle sind die Symbole einer Symbolfolge statistisch unabhängig, d.h. es gilt

$$P(u_N \mid u_1,\ ...,\ u_{N-1}) = P(u_N) \tag{2.1}$$

Der Informationsgehalt eines Symbols einer Nachricht wird in der hier betrachteten statistischen Informationstheorie quantitativ an der Ungewißheit des Auftretens eines Symbols gemessen. Ein Symbol besitzt nur geringen Informationsgehalt, wenn sein Auftreten vorhersagbar ist; je größer die Ungewißheit bezüglich des Auftretens eines Symbols ist, um so größer ist sein Informationsgehalt. Mit Bekanntwerden des Symbols wird die Ungewißheit beseitigt. Aufgrund dieses Zusammenhanges erscheint es sinnvoll, den Informationsgehalt $I(a_k)$ eines Symbols a_k anhand der Wahrscheinlichkeit $P(a_k)$ des Auftretens desselben Symbols zu messen:

$$I(a_k) = -\log_2 P(a_k) \tag{2.2}$$

Mit einer solchen statistischen Größe als Maß für den Informationsgehalt wird der semantische Inhalt der Nachricht nicht erfaßt; beispielsweise sagt dieses Maß nichts darüber aus, ob die Nachricht gut oder schlecht ist. Dieses Maß der Information ermöglicht es aber, Informationserzeugung, Informationsflüsse und Informationsempfang quantitativ zu erfassen; es führt zu einer statistischen Informationstheorie.

Wird der Informationsgehalt $I(a_k)$ nach Gl. (2.2) mit dem Logarithmus zur Basis 2 bezeichnet, so ergibt sich die Einheit in "bit". Nach dieser Definition besitzt beispielsweise ein Symbol a_k mit der Wahrscheinlichkeit $P(a_k) = 1$ den Informationsgehalt $I(a_k) = 0$ bit. Es besitzt also keinen Informationsgehalt, da es mit Sicherheit auftritt und keine Ungewißheit bezüglich des Auftretens von a_k besteht. In Bild 2.1 ist die Beziehung (2.2) grafisch dargestellt.

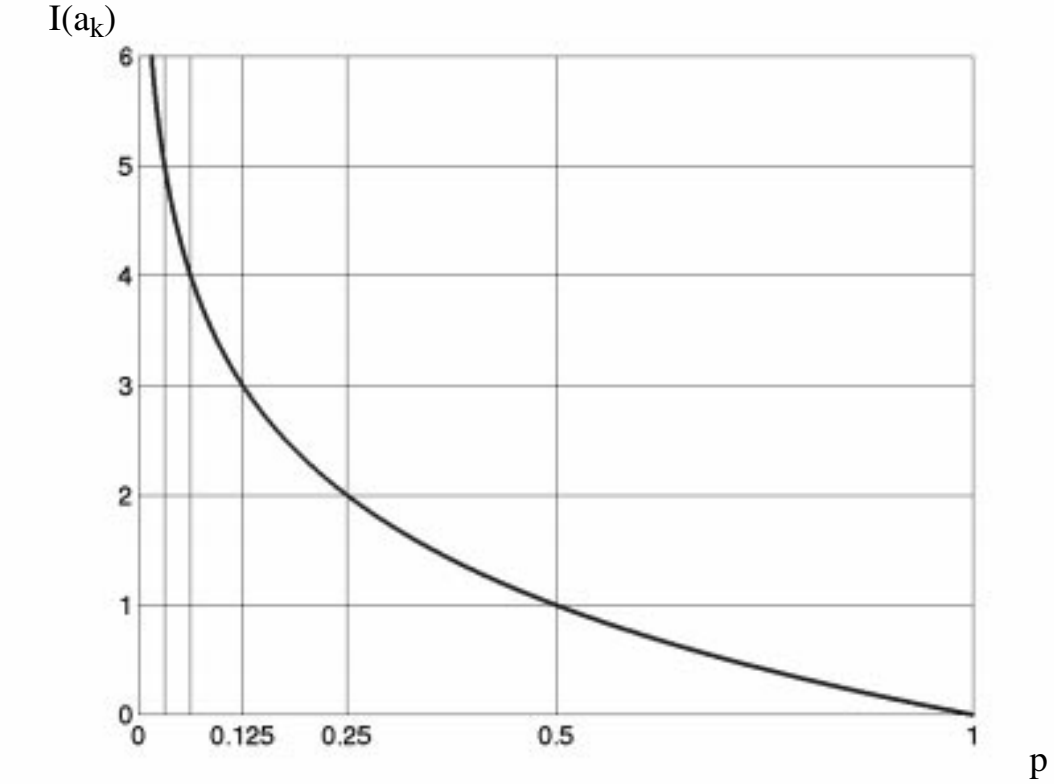


Bild 2.1 Informationsgehalt $I(a_k)$ eines Symbols a_k der Auftretenswahrscheinlichkeit $P(a_k) = p$

Der mittlere Informationsgehalt H in bit je Symbol der von einer diskreten, gedächtnislosen Quelle erzeugten Symbole beträgt demzufolge

$$H = \sum_{k=1}^{K} P(a_k) \cdot I(a_k) = -\sum_{k=1}^{K} P(a_k) \cdot \log_2 P(a_k) \tag{2.3}$$

und wird als Entropie bezeichnet. Bild 2.2 zeigt den von einem Symbol a_k mit der Wahrscheinlichkeit $P(a_k) = p$ zur Entropie im Mittel beigetragenen Informationsgehalt. Multipliziert man H mit der Anzahl der je Zeiteinheit von der Quelle gesendeten Symbole, so erhält man H', die von der Quelle je Zeiteinheit erzeugte Entropie.

Es wird mit dem Theorem der Quellencodierung in Abschnitt 3.2 gezeigt, daß der durch Gl. (2.3) definierte mittlere Informationsgehalt in bit je Symbol eine untere Grenze der zur Codierung eines Symbols erforderlichen Anzahl von Binärzeichen darstellt und somit eine sinnvolle Definition ist.

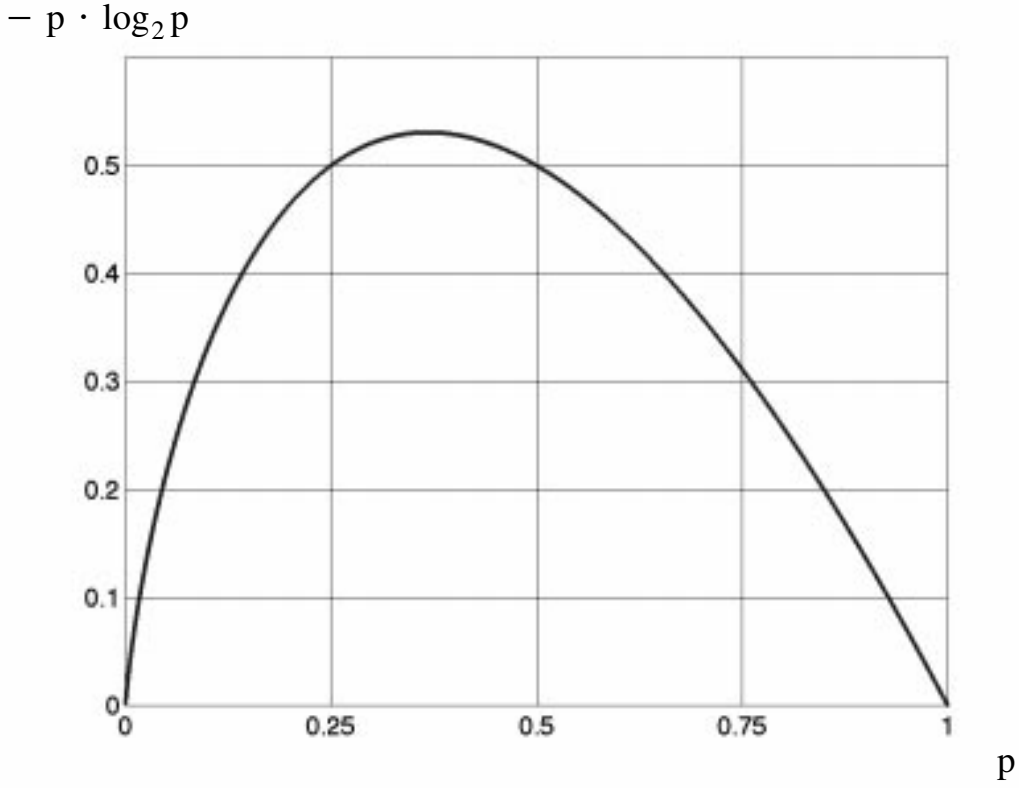


Bild 2.2 Beitrag $P(a_k)\cdot I(a_k)$ eines Symbols a_k der Auftretenswahrscheinlichkeit $P(a_k) = p$ zum mittleren Informationsgehalt H

Beispiel: Eine binäre, gedächtnislose Quelle besitzt einen Symbolvorrat, der nur aus $K = 2$ Symbolen 0 und 1 mit den Wahrscheinlichkeiten

$P(a_1 = 0) = p$ und $P(a_2 = 1) = 1 - p$

besteht.

Aus Gl. (2.3) folgt dann

$$H = -\left\{p \cdot \log_2(p) + (1 - p) \cdot \log_2(1 - p)\right\} .$$

Der Verlauf der Entropie $H(p)$ als Funktion von p ist in Bild 2.3 gezeigt.–

Bezeichnet u ein beliebiges Symbol des Symbolvorrates a_1, a_2, ..., a_K und $P(u)$ die Wahrscheinlichkeit des Symbols u, so bilden Symbole und Wahrscheinlichkeiten ein sogenanntes Nachrichtenensemble U, das die Nachrichtenquelle charakterisiert.

Die Entropie dieser Quelle kann dann anstatt in der Form von Gl. (2.3) auch in der vereinfachten Form geschrieben werden.

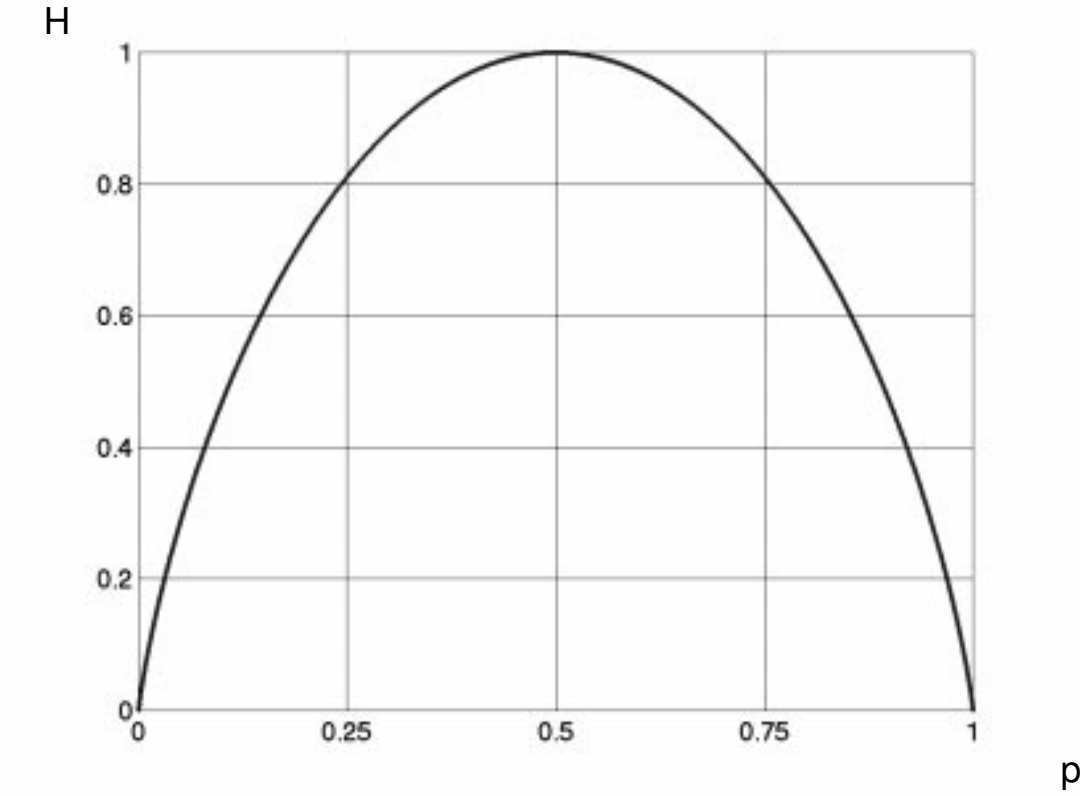


Bild 2.3 Entropie H je Symbol einer diskreten, binären, gedächtnislosen Quelle mit $P(0) = p$ und $P(1) = 1 - p$

$$\begin{aligned} H &= -\sum_{k=1}^{K} P(a_k) \cdot \log_2 P(a_k) \\ &= -\sum_{u} P(u) \cdot \log_2 P(u) \\ &= H(U) \end{aligned} \tag{2.4}$$

Zur Erläuterung dieses Begriffs sollen einige Eigenschaften der Entropie betrachtet werden.

Theorem 1.1:

Die Entropie $H(U)$ ist stets positiv, d.h.

$$H(U) = -\sum_{u} P(u) \cdot \log_2 P(u) \geq 0\,, \tag{2.5}$$

da jeder Summand positiv ist. Nur in einem Sonderfall wird sie Null, wenn ein Symbol mit der Wahrscheinlichkeit $P(a_1) = 1$ gesendet wird und alle anderen die Wahrscheinlichkeit Null aufweisen. In diesem Fall erzeugt die Quelle die Symbolfolge ..., a_1, a_1, a_1, ... ; die Folge dieser Symbole hat keinen Informationsgehalt.

Theorem 1.2:
Die Entropie $H(U)$ ist stets kleiner oder gleich $\log_2 K$, d.h.

$$H(U) \leq \log_2 K. \tag{2.6}$$

Beweis:

Der Beweis soll mit Hilfe folgender Beziehungen

$$\ln z < z - 1 \text{ für } z > 0, \; z \neq 1$$

$$\ln z = z - 1 \text{ für } z = 1 \tag{2.7}$$

sowie der Gleichung

$$\log z = \ln z \cdot \log e \tag{2.8}$$

erbracht werden.
Die Beziehungen (2.7) sind in Bild 2.4 grafisch dargestellt.
Anstelle der Ungleichung (2.6) wird die entsprechende Ungleichung

$$H(U) - \log_2 K \leq 0 \tag{2.9}$$

bewiesen. Es gilt mit (2.8)

$$\begin{aligned} H(U) - \log_2 K &= \sum_u P(u) \cdot \log_2 \frac{1}{P(u)} - \sum_u P(u) \cdot \log_2 K \\ &= \log_2 e \cdot \sum_u P(u) \cdot \ln \frac{1}{P(u) \cdot K} \end{aligned} \tag{2.10}$$

Wird die Summe über alle u mit $P(u) \neq 0$ gebildet und werden für jeden Summanden die Beziehungen (2.7) angewendet, so folgt der Beweis

$$H(U) - \log_2 K \leq \log_2 e \cdot \sum_u P(u) \left[\frac{1}{P(u) \cdot K} - 1 \right]$$

$$H(U) - \log_2 K \leq 0 \qquad .$$

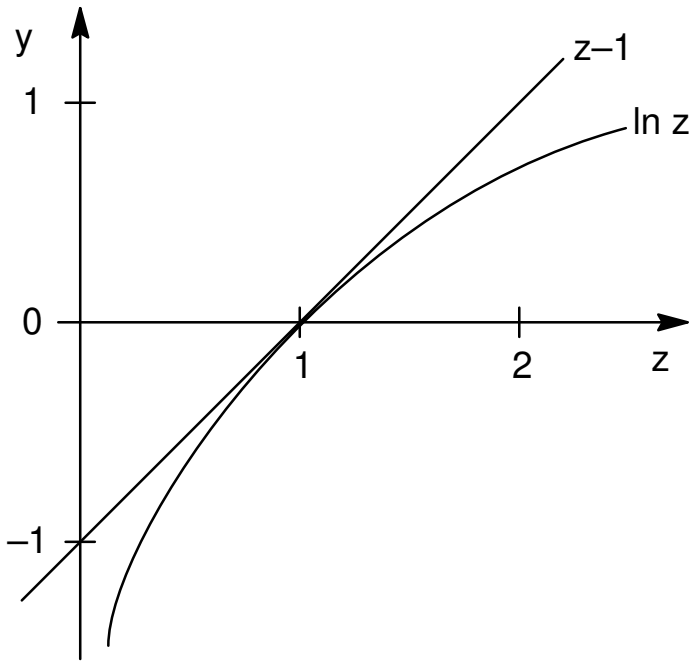


Bild 2.4 Skizze der Funktionen $y = z - 1$ und $y = \ln z$

Das Gleichheitszeichen dieser Ungleichung gilt aufgrund von (2.7) dann und nur dann, wenn für alle u

$$\begin{aligned} &\frac{1}{P(u) \cdot K} = 1 \\ \text{bzw.} \qquad &P(u) = \frac{1}{K} \end{aligned} \tag{2.11}$$

gilt. Somit besagt die Ungleichung (2.6), daß die Entropie je Symbol eines Nachrichtenensembles am größten ist, wenn alle Symbole des Symbolvorrats gleich wahrscheinlich sind. Dieser Maximalwert wird auch als Entscheidungsgehalt H_0 bezeichnet.

$$H_{\max}(U) = -\sum_{k=1}^{K} \frac{1}{K} \cdot \log_2 \frac{1}{K} = \log_2 K = H_0 \tag{2.12}$$

Wie im folgenden gezeigt wird, kann man die Auswahl eines Symbols aus einem Symbolvorrat von K Symbolen in einen Auswahlprozess von H_0 Binärentscheidungen zerlegen. Das Verzweigungsdiagramm nennt man Entscheidungsbaum oder auch Codebaum.

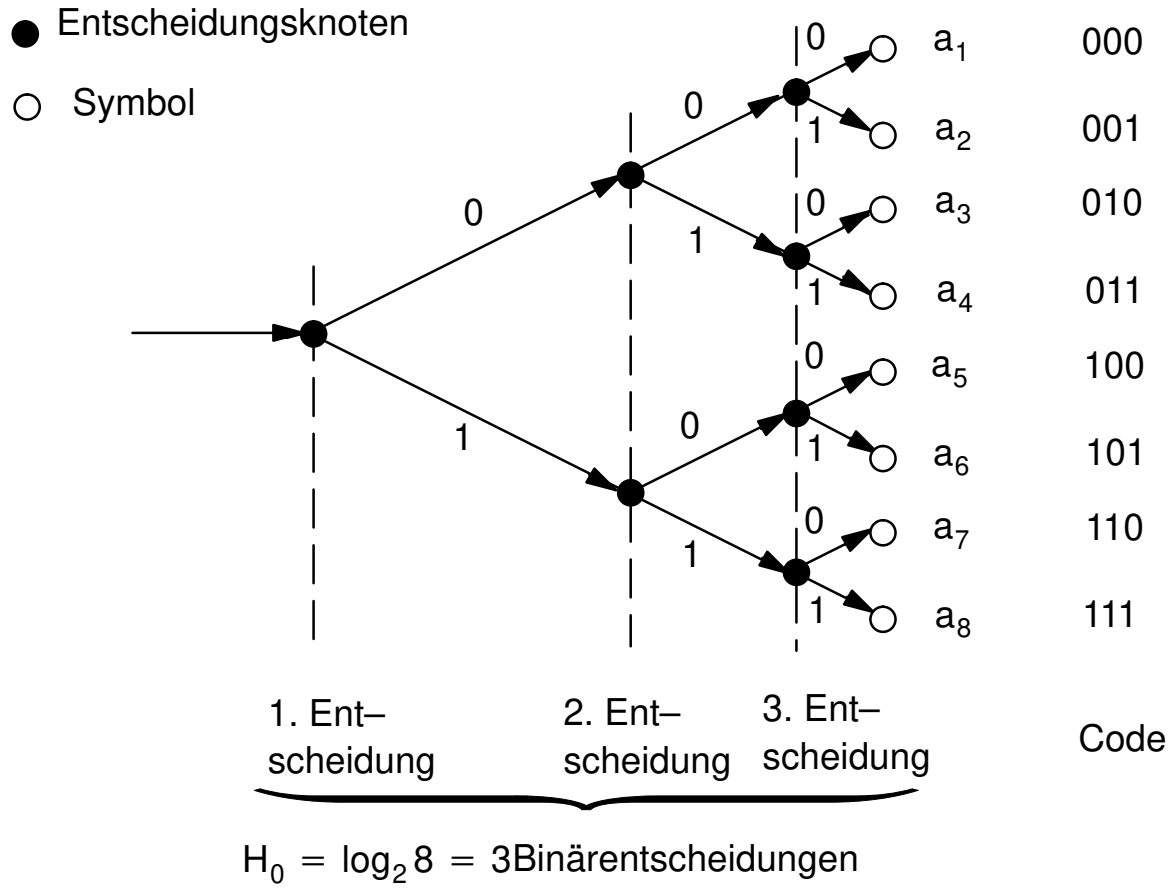


Bild 2.5 Codebaum für 8 Symbole

Jedes Symbol des in Bild 2.5 als Beispiel angeführten Symbolvorrats läßt sich durch drei Binärentscheidungen auswählen und entsprechend durch drei Binärzeichen eindeutig kennzeichnen. Die Reihenfolge der Binärzeichen ergibt sich aus der Folge der Binärzeichen entlang der Verzweigungen des Codebaumes, beginnend bei der Wurzel bis zum ausgewählten Symbol a_k. Allgemein erfordert die Auswahl eines Symbols aus K Symbolen maximal $H_0 = \log_2 K$ Binärentscheidungen.

2.2 Diskrete gedächtnisbehaftete Quelle

Reale Nachrichtenquellen sind im allgemeinen gedächtnisbehaftet. Ein sehr nützliches Modell der gedächtnisbehafteten Quelle ist die diskrete, gedächtnisbehaftete Markoff–Quelle. Diese Quelle erzeugt durch zeitlich sequentielle Auswahl der Symbole u aus dem Symbolvorrat a_1, ..., a_K einen Zufallsprozeß u_1, ..., u_N statistisch abhängiger Symbole. Dabei wird angenommen, daß das Symbol u_n nur von $N - 1$ unmittelbar vorausgehenden Symbolen statistisch abhängig ist, d.h.

$$P(u_N|\, u_1,\ ...,\ u_{N-1}) = P(u_N|\, u_{-m},\ ...,\ u_1,\ ...,\ u_{N-1}). \tag{2.13}$$

Werden die $N-1$ vorausgehenden Symbole als bekannt vorausgesetzt, so beträgt der sog. bedingte Informationsgehalt des N–ten Symbols dieser speziellen gedächtnisbehafteten Quelle

$$I(u_N|\, u_1,\ ...,\ u_{N-1}) = -\log_2 P(u_N|\, u_1,\ ...,\ u_{N-1}). \tag{2.14}$$

Der mittlere bedingte Informationsgehalt ergibt sich aus dem statistischen Mittelwert für $I(u_N|\, u_1,\ ...,\ u_{N-1})$. Der Mittelwert wird als bedingte Entropie $H(U_N|\, U_1,\ ...,\ U_{N-1})$ bezeichnet.

$$H(U_N|\, U_1,\ ...,\ U_{N-1}) = \\ -\sum_{u_1} ... \sum_{u_N} P(u_1,\ ...,\ u_N)\ \log_2 P(u_N|\, u_1,\ ...,\ u_{N-1}) \tag{2.15}$$

Im Sonderfall statistisch unabhängiger Symbole folgt mit

$$P(u_N|\ u_1,\ ...,\ u_{N-1}) = P(u_N) \tag{2.16}$$

bzw. $$P(u_1,\ ...,\ u_N) = P(u_1) \cdot P(u_2) \cdot ... \cdot P(u_N) \tag{2.17}$$

aus Gleichung (2.15)

$$H(U_N|\, U_1,\ ...,\ U_{N-1}) = -\sum_{u} P(u_N) \cdot \log_2 P(u_N) \\ = H(U_N)\ , \tag{2.18}$$

d.h. die bedingte Entropie geht in die bereits bekannte Form der Entropie nach Gleichung (2.4) über. Damit entartet die gedächtnisbehaftete Quelle zu einer gedächtnislosen Quelle.

Theorem 1.3:

Die bedingte Entropie genügt der Ungleichung

$$H(U_N|U_1,\ ...,\ U_{N-1}) \le H(U_N)\ , \tag{2.19}$$

wobei das Gleichheitszeichen dann und nur dann gilt, wenn die Ensembles U_N und $U_1,\ ...,\ U_{N-1}$ statistisch unabhängig sind.

Beweis:

Der Beweis wird mit Hilfe der Beziehungen (2.7) und (2.8) erbracht. Dazu wird die Differenz

$$H(U_N|U_1,\ ...,\ U_{N-1}) - H(U_N) = \sum_{u_1} ... \sum_{u_N} P(u_1,\ ...,\ u_N) \cdot \log_2 \frac{P(u_N)}{P(u_N|u_1,\ ...,\ u_{N-1})} \tag{2.20}$$

betrachtet. Indem für jeden Summanden (2.7) und (2.8) angewendet wird, folgt

$$H(U_N|\ U_1,\ ...,\ U_{N-1}) - H(U_N) \leq \log_2 e \cdot \sum_{u_1} ... \sum_{u_N} P(u_1,\ ...,\ u_N) \cdot \left[\frac{P(u_N)}{P(u_N|u_1,\ ...,\ u_{N-1})} - 1 \right]$$

$$H(U_N|\ U_1,\ ....,\ U_{N-1}) - H(U_N) \leq \log_2 e \cdot \left[\sum_{u_1} ... \sum_{N} P(u_N) \cdot P(u_1,\ ...,\ u_{N-1}) \dot{-} 1 \right] \leq 0 \tag{2.21}$$

Das Gleichheitszeichen gilt dann und nur dann, wenn $P(u_N) = P(u_N|u_1,\ ...,\ u_{N-1})$ gilt, d.h. die Ensembles U_N und $U_1,\ ...,\ U_{N-1}$ statistisch unabhängig sind. Damit wird auch der Fall eingeschlossen, daß alle Ensembles $U_1,\ U_2,\ ...,\ U_N$ statistisch unabhängig sind.–

Aufgrund von Theorem 1.3 ist der mittlere Informationsgehalt $H(U_N|U_1,\ ...,\ U_{N-1})$ einer physikalischen Quelle, berechnet mit dem Modell der gedächtnisbehafteten Quelle, stets kleiner oder gleich dem mittleren Informationsgehalt $H(U_N)$, berechnet mit dem Modell der gedächtnislosen Quelle.

Betrachtet man ein Verbundereignis $u_1,\ ...,\ u_N$ von N Symbolen, die sowohl gleichzeitig als auch zeitlich nacheinander auftreten können, als ein Ereignis, so kann die Verbundentropie

$$H(U_1,\ ...,\ U_N) = -\sum ... \sum P(u_1,\ ...,\ u_N) \cdot \log_2 P(u_1,\ ...,\ u_N) \tag{2.22}$$

als der mittlere Informationsgehalt eines Verbundereignisses interpretiert werden.

Mit Hilfe der Beziehung

$$P(u_1,\ ...,\ u_N) = P(u_N|u_1,\ ...,\ u_{N-1}) \cdot P(u_1,\ ...,\ u_{N-1}) \tag{2.23}$$

kann folgender mathematischer Zusammenhang zwischen den definierten Entropien abgeleitet werden:

$$H(U_1,\ ...,\ U_N) = H(U_N|U_1,\ ...,\ U_{N-1}) + H(U_1,\ ...,\ U_{N-1}). \tag{2.24}$$

Die wiederholte Anwendung dieser Beziehung auf die Verbundentropie der rechten Seite der Gleichung (2.24) führt auf die sogenannte Kettenregel:

$$H(U_1,\ ...,\ U_N) = H(U_1) + H(U_2|U_1) + ... + H(U_N|U_1,\ ...,\ U_{N-1}) \qquad (2.25)$$

Da für jeden Summanden das Theorem 1.3 gilt, erhält man weiterhin die Ungleichung

$$H(U_1,\ ...,\ U_N) \leq \sum_{n=1}^{N} H(U_n)\ , \qquad (2.26)$$

in der das Gleichheitszeichen dann und nur dann gilt, wenn alle Ensembles $U_1,\ U_2,\ ...,\ U_N$ statistisch unabhängig sind.

Theorem 1.4:

Es sei

$$H_N(U) = \frac{1}{N} H(U_1, ..., U_N) \qquad (2.27)$$

dann gilt

a) $H\,(\,U_N|(U_1, ..., U_{N-1})$ verringert sich mit wachsendem N

b) $H_N\,(U) \geq H\,(U_N|U_1,\ldots, U_{N-1})$ (2.28)

c) $H_N\,(U)$ verringert sich mit wachsendem N

d) $\lim\limits_{N\to\infty} H_N\,(U) = \lim\limits_{N\to\infty} H\,(U_N|U_1,\ldots, U_{N-1})$ (2.29)

Beweis:

a) Da eine zusätzliche Bedingung die Entropie verkleinert, folgt aus

$$H\,(U_N|U_1,\ldots, U_{N-1}) \leq H\,(U_N|U_2,\ldots, U_{N-1})$$
$$H\,(U_N|\,U_1,\ldots, U_{N-1}) \leq H\,(U_{N-1}|\,U_1,\ldots, U_{N-2}) \qquad (2.30)$$

der Beweis für (a).

b) Mit der Kettenregel Gl. (2.25) wird

$$H_N(U) = \frac{1}{N}[\,H\,(U_1) + H\,(U_2|\,U_1) + ...\ H\,(U_N|\,U_1, ..., U_{N-1})] \qquad (2.31)$$

Aufgrund von (a) sind alle Terme der Summe größer als der letzte Term, d.h.

$$H_N(U) \geq \frac{1}{N}[\,N \cdot H(U_N|\,U_1, ..., U_{N-1)}]$$

woraus (b) folgt.

c) $H_N(U)$ kann in der Form

$$H_N(U) = \frac{1}{N} H(U_1, ..., U_{N-1}) + \frac{1}{N} H(U_N | U_1, ..., U_{N-1})$$
$$H_N(U) = \frac{N-1}{N} H_{N-1}(U) + \frac{1}{N} H(U_N | U_1, ..., U_{N-1}) \quad (2.32)$$

geschrieben werden. Aufgrund von (b) gilt

$$H_N(U) \leq \frac{N-1}{N} H_{N-1}(U) + \frac{1}{N} H_N(U)$$
$$H_N(U) \leq H_{N-1}(U)$$

womit (c) bewiesen ist.

d) Unter Verwendung der Kettenregel Gl. (2.25) und Gl. (2.31) wird

$$\begin{aligned} H_{N+j}(U) &= \frac{1}{N+j} H(U_1,\ldots,U_{N-1}) \\ &+ \frac{1}{N+j} [\, H(U_N | U_1,\ldots,U_{N-1)} + H(U_{N+1} | U_1,\ldots,U_N) \\ &\qquad \ldots + H(U_{N+j} | U_1,\ldots,U_{N+j-1})] \\ &\leq \frac{1}{N+j} H(U_1,\ldots,U_{N-1}) + \frac{j+1}{N+j} H(U_N | U_1,\ldots,U_{N-1}) \end{aligned} \quad (2.33)$$

Mit $j \to \infty$ ergibt sich aus Gl. (2.33)

$$H_\infty(U) \leq H(U_N | U_1,\ldots,U_{N-1}) \quad (2.34)$$

für alle N, d.h.

$$H_\infty(U) \leq \lim_{N\to\infty} H(U_N | U_1,\ldots,U_{N-1}) \quad (2.35)$$

Aus Gl. (2.35) und Gl. (2.28) folgt der Beweis für (d).

3. Redundanzreduzierende Codierung

In diesem Abschnitt werden Möglichkeiten der verlustlosen Codierung diskreter, gedächtnisloser und diskreter gedächtnisbehafteter Quellen behandelt. Es wird nachgewiesen, daß zur eindeutigen Codierung und Decodierung eines Symbols im Mittel H bit je Quellensymbol benötigt werden, wobei H der mittlere Informationsgehalt je Quellensymbol ist. Im Fall, daß die Symbole einer Quelle mit $H_0 = \log_2 k$ bit dargestellt werden, kann eine noch zu definierende Redundanz R in bit je Symbol mittels einer Quellencodierung eliminiert werden, ohne daß ein Informationsverlust auftritt. Diese Codierung wird daher als redundanzmindernde Codierung bezeichnet.

3.1 Codierung diskreter gedächtnisloser Quellen

Es sei angenommen, daß eine diskrete, gedächtnislose Quelle eine zeitliche Folge $u_1, \ldots, u_N$ statistisch unabhängiger Quellensymbole erzeugt, wobei jedes Symbol u_n aus einem Symbolvorrat von K Quellensymbolen $a_1, \ldots, a_k, \ldots, a_K$ ausgewählt wird. Die Quellensymbole treten mit den Wahrscheinlichkeiten $P(a_1), \ldots, P(a_k), \ldots, P(a_K)$ auf. Jedes Quellensymbol soll durch ein Codewort dargestellt werden, das aus mehreren Codeelementen bestehen kann. Die Codeelemente werden aus einem Codesymbolvorrat von D Codesymbolen ausgewählt. Dabei werde das einzelne Quellensymbol a_k stets durch ein Codewort mit n_k Codeelementen codiert. Die Codewörter der Quellensymbole können unterschiedlich viele Codeelemente aufweisen.

Im folgenden soll die untere Grenze für die mittlere Codewortlänge $\overline{n}$ ermittelt werden, wobei

$$\overline{n} = \sum_{k=1}^{K} P(a_k) \cdot n_k \tag{3.1}$$

gilt.

Tabelle 3.1 zeigt einige Beispiele für binäre Codes mit Codewörtern unterschiedlicher Länge. Wenn eine Codierung keinen Informationsverlust verursachen darf, muß die Codierung eindeutig umkehrbar sein.

Quellen–symbol u	Wahrschein–lichkeit P(u)	Code I	Code II	Code III	Code IV
a_1	0,5	0	0	0	0
a_2	0,25	0	1	1 0	0 1
a_3	0,125	1	0 0	1 1 0	0 1 1
a_4	0,125	1 0	1 1	1 1 1	0 1 1 1

Tab. 3.1: Beispiel für binäre Codes mit ungleichlangen Codewörtern, $K = 4$, $D = 2$

Anhand der Beispiele soll gezeigt werden, daß für eine eindeutig umkehrbare Codierung gewisse Nebenbedingungen vom Code erfüllt werden müssen.

Code I ist nicht eindeutig zu decodieren, da für a_1 und a_2 das gleiche Codesymbol 0 verwendet wird. Dieser Code erfüllt somit nicht die gestellten Bedingungen.

Code II ist ebenfalls nicht eindeutig zu decodieren, da für die Folge a_1a_1 und das Quellensymbol a_3 die gleiche Folge von Codesymbolen 00 verwendet wird. Auch dieser Code erfüllt somit nicht die gestellten Bedingungen.

Damit auch eine Folge von Codewörtern eindeutig decodiert werden kann, müssen Anfang oder Ende der Codewörter erkennbar sein. Diese Forderung erfüllen sog. Präfixcodes. Ein Präfixcode ist ein Code mit der Eigenschaft, daß kein Codewort mit dem ersten Teil (Präfix) eines anderen Codewortes identisch ist.

Eine anschauliche grafische Darstellung für Codes ist der Codebaum. Bild 3.1 zeigt den Codebaum für den Präfixcode entsprechend Code III nach Tabelle 3.1.

Ein Präfixcode ist dadurch gekennzeichnet, daß den Quellensymbolen nur Endknoten und keine Entscheidungsknoten zugeordnet sind. Die Zuordnung der Codewörter zu den Quellensymbolen bzw. zu den Endknoten ergibt sich aus der Aneinanderreihung der Codesymbole, die am Verbindungsweg von der Wurzel bis zum Endknoten liegen.

Code III ist ein Präfixcode. Die zufällige Folge 0111100 kann eindeutig in die Symbolfolge $a_1a_4a_2a_1$ decodiert werden, da das Ende der Codewörter aufgrund der Präfixeigenschaften stets eindeutig erkennbar ist.

Code IV ist ebenfalls eindeutig zu decodieren, obwohl er kein Präfixcode ist. Bei diesem Code ist der Anfang eines Codewortes jeweils durch Auftreten des Symbols 0 zu erkennen.

Um eine eindeutig umkehrbare Codierung zu gewährleisten, werden zunächst nur Präfixcodes betrachtet. Die zur Konstruktion von Präfixcodes erforderlichen Vor-

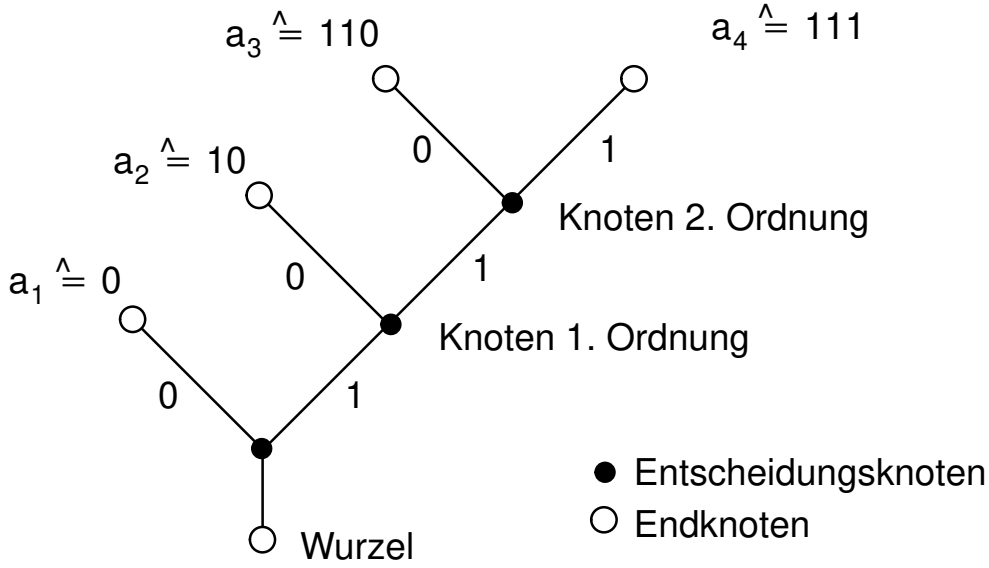


Bild 3.1 Codebaum des Präfixcode III von Tabelle 3.1

aussetzungen sollen im folgenden noch näher erörtert werden. Um den K verschiedenen Quellensymbolen je ein Codewort und je einen Endknoten zuordnen zu können, müssen die Codewortlängen $n_1,\ n_2,\ ...,\ n_k$ die sog. Kraft'sche Ungleichung erfüllen:

Theorem 1.5:

Eine notwendige Bedingung für die Konstruktion eines Präfixcodes mit D Codesymbolen ist, daß die Codewortlängen $n_1,\ ...,\ n_k,\ ...,\ n_K$ die Ungleichung

$$\sum_{k=1}^{K} D^{-n_k} \leq 1 \tag{3.2}$$

erfüllen.

Beweis:

Jeder Knoten des Codebaumes eines Präfixcodes mit D Codesymbolen besitzt D Zweige. Ist n_K die größte Codewortlänge, so gibt es maximal D^{n_K} Endknoten. Besitzt ein Präfixcode einen Endknoten der Ordnung n_k, wobei $n_k \leq n_K$, so vermindert dieser die Anzahl der verfügbaren Endknoten n_K–ter Ordnung um $D^{n_K - n_k}$, wie es in Bild 3.2 dargestellt ist.–

Die Summe aller nicht verfügbaren Endknoten n_K–ter Ordnung und verwendeter Endknoten n_K–ter Ordnung muß immer kleiner oder gleich D^{n_K} sein, d.h.

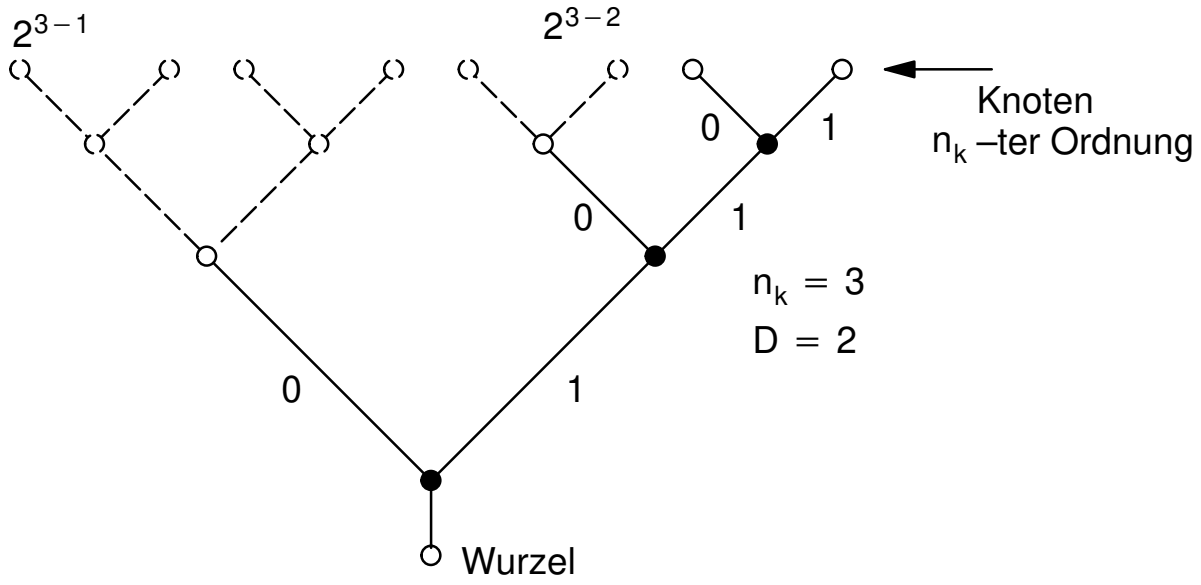


Bild 3.2 Ergänzter Codebaum des Code III

$$\sum_{k=1}^{K} D^{n_K - n_k} \leq D^{n_K} . \tag{3.3}$$

Wird diese Ungleichung durch D^{n_K} dividiert, so erhält man Theorem 1.5.–

Es läßt sich nachweisen, daß jeder eindeutig umkehrbare Code die Kraft'sche Ungleichung erfüllen muß. Somit ist jeder eindeutig umkehrbare Code auch durch einen Präfixcode darstellbar. Das folgende Theorem der Quellencodierung gilt somit allgemein für alle eindeutig umkehrbaren Codes wie auch für die besondere Gruppe der Präfixcodes.

3.2 Theorem der Quellencodierung

Theorem 1.6:

Gegeben sind eine diskrete, gedächtnislose Quelle U mit der Entropie $H(U)$ und ein Codesymbolvorrat von D Codesymbolen. Es ist möglich, den Quellensymbolen a_1, ..., a_K derart Codewörter eines Präfixcodes zuzuordnen, daß die mittlere Codewortlänge $\bar{n}$ je Quellensymbol der Bedingung

$$\frac{H(U)}{\log_2 D} \leq \overline{n} < \frac{H(U)}{\log_2 D} + 1 \tag{3.4}$$

genügt.

Beweis:

Es seien $P(a_1)$, ..., $P(a_K)$ die Wahrscheinlichkeiten der Quellensymbole a_1, ..., a_K und n_1, ..., n_K die zugehörigen Codewortlängen. Es wird zunächst die linke Ungleichung, d.h. die untere Grenze des Theorems 1.6 bewiesen:

$$H(U) - \overline{n}\ \log_2 D \leq 0 \tag{3.5}$$

$$H(U) - \overline{n}\ \log_2 D = \sum_{k=1}^{K} P(a_k)\ \log_2 \frac{1}{P(a_k)} - \sum_{k=1}^{K} P(a_k)\ n_k\ \log_2 D \tag{3.6}$$

$$H(U) - \overline{n}\ \log_2 D = \sum_{k=1}^{K} P(a_k)\ \log_2 \frac{D^{-n_k}}{P(a_k)}\ . \tag{3.7}$$

Mit den Beziehungen (2.7) und (2.8) erhält man:

$$H(U) - \overline{n}\log_2 D \leq \sum_{k=1}^{K} P(a_k)\left[\frac{D^{-n_k}}{P(a_k)} - 1\right]\log e \tag{3.8}$$

$$H(U) - \overline{n}\log_2 D \leq \left[\sum_{k=1}^{K} D^{-n_k} - 1\right]\log e \leq 0\ . \tag{3.9}$$

Die Ungleichung (3.9) folgt aus der Kraft'schen Ungleichung (3.2). In der Ungleichung (3.9) gelten die Gleichheitszeichen aufgrund der Beziehung (2.7) dann und nur dann, wenn für alle k

$$P(a_k) = D^{-n_k} \quad ; \quad 1 \leq k \leq K \tag{3.10}$$

gilt. Mit Gleichung (3.10) ist die Codewortlänge für ein Quellensymbol der Wahrscheinlichkeit $P(a_k)$ festgelegt:

$$n_k = -\frac{\log_2 P(a_k)}{\log_2 D}\ . \tag{3.11}$$

Die häufigen Quellensymbole werden demzufolge in kurze und die seltenen Quellensymbole in lange Codeworte codiert. Aufgrund der Ungleichung (3.5) liefert die Codierung mit Codewortlängen gemäß Gl. (3.11) den Code mit der kleinsten mittleren Codewortlänge.

Da nur Codierungen mit ganzzahligen Codewortlängen n_k realisiert werden können, ist die Codierungsvorschrift (3.11) nur anwendbar, wenn alle $P(a_k)$ in ganzzahligen Potenzen von D vorliegen. In allen anderen Fällen kann die Codierungsvorschrift (3.11) nur näherungsweise erfüllt werden, indem n_k als ganze Zahl entsprechend

$$-\frac{\log_2 P(a_k)}{\log_2 D} \leq n_k < -\frac{\log_2 P(a_k)}{\log_2 D} + 1 \quad , \quad 1 \leq k \leq K \tag{3.12}$$

gewählt wird. Durch Multiplikation mit $P(a_k)$ und Aufsummierung über k erhält man aus der Ungleichung (3.12) den Beweis für das Theorem 1.6. Da die Kraft'sche Ungleichung durch die linke Ungleichung des Theorems erfüllt wird, gibt es immer einen Präfixcode, der das Theorem 1.6. erfüllt.–

Als Redundanz R einer Quelle wird die Differenz zwischen Entscheidungsgehalt H_0 und Entropie H definiert. Diese Definition ist nur für diskrete Quellen anwendbar, da der Entscheidungsgehalt nur für diskrete Quellen definiert ist.

Die Redundanz R eines Symbols einer diskreten Quelle beträgt somit im Mittel

$$R = H_0 - H\,, \tag{3.13}$$

wobei $H_0 = \log_2 K$ und H der mittlere Informationsgehalt je Quellensymbol der diskreten gedächtnislosen oder gedächtnisbehafteten Quelle ist. Die Redundanz ist stets positiv und aufgrund von Theorem 1.3 für die gedächtnisbehaftete Quelle stets größer oder gleich der einer gedächtnislosen Quelle.

Bei einer Codierung mit gleichlangen Codewörtern der Länge n und dem Umfang D des Codesymbolvorrats ist der Entscheidungsgehalt der Codewörter $H_0 = n \cdot \log_2 D$ und die Redundanz

$$R = n \cdot \log_2 D - H\,. \tag{3.14}$$

Bei einer Codierung mit ungleich langen Codewörtern der mittleren Länge $\overline{n}$ wird

$$R = \overline{n} \cdot \log_2 D - H. \tag{3.15}$$

Dabei kann $\overline{n} \cdot \log_2 D$ als mittlerer Entscheidungsgehalt der Codewörter interpretiert werden.

Der Code III in Tab. 3.1 ist gemäß Gl. (3.15) ein optimaler, redundanzfreier Code. Es liegt der Sonderfall vor, daß die Codierungsvorschrift (3.11) ganzzahlige n_k liefert, weil die Wahrscheinlichkeiten der Quellensymbole zufällig ganzzahlige Potenzen von $D = 2$ sind. Aus Gl. (3.11) folgt mit $\log_2 D = 1$

$$\overline{n} = \sum_{k=1}^{K} P(a_k)\; n_k = -\sum_{k=1}^{K} P(a_k)\; \log_2 P(a_k) = H(U).$$

Damit erreicht der Code III die untere Grenze $\overline{n} = H(U) = 1.75$ bit. Bezogen auf eine Codierung mit gleichlangen Codewörtern gemäß Gl. (3.14) eliminiert er eine Redundanz von

$$R = n \log_2 D - H = 2 - 1,75 = 0,25 \text{ bit / Symbol.}$$

Redundanzmindernde Präfixcodes, die das Theorem 1.6 erfüllen, wurden von Shannon, [8] Fano und Huffman angegeben. Unter diesen Codes führt der von Huffman (1952) [4] vorgeschlagene Codierungsalgorithmus nachweisbar auf den Code mit der geringsten Redundanz bzw. geringsten mittleren Codewortlänge $\overline{n}$.

3.3 Codierungsalgorithmus von Huffman

Der von Huffman (1952) [4] entwickelte Codierungsalgorithmus zur Berechnung von Codes mit minimaler Redundanz lautet im Fall einer Binärcodierung:

1. Schritt: Die K Quellensymbole werden in der Reihenfolge abnehmender Wahrscheinlichkeit geordnet, wie es im Beispiel von Bild 3.3 gezeigt wird.

2. Schritt: Man fasse die zwei kleinstwahrscheinlichen Quellensymbole zu einem Hilfssymbol zusammen und berechne dessen Wahrscheinlichkeit.

3. Schritt: Man ordne das Hilfssymbol und die übrigen Quellensymbole wieder nach abnehmender Wahrscheinlichkeit, fasse dann die zwei kleinstwahrscheinlichen Symbole wieder zu einem neuen Hilfssymbol zusammen und berechne dessen Wahrscheinlichkeit.

4. Schritt: Der 3. Schritt ist so oft zu wiederholen, bis ein Symbol mit der Wahrscheinlichkeit 1 auftritt.

5. Schritt: Kennzeichnet man die Zusammenfassungen von Symbolen wie in Bild 3.3 durch Linien, so erhält man den Codebaum eines Präfixcodes, dessen K Endknoten den K Quellensymbolen zugeordnet sind. Jeder Knoten besitzt zwei Zweige. Jedem Zweig eines Knotens ist eines der zwei Codessymbole 0 oder 1 des Codesymbolvorrats zugeordnet; aus dieser Zuordnung ergibt sich der Code für die Quellensymbole.

Quellensymbol

u	P(u)	Code	Codebaum
a_1	0,25	10	
a_2	0,25	01	
a_3	0,20	00	
a_4	0,15	111	
a_5	0,10	1101	
a_6	0,05	1100	

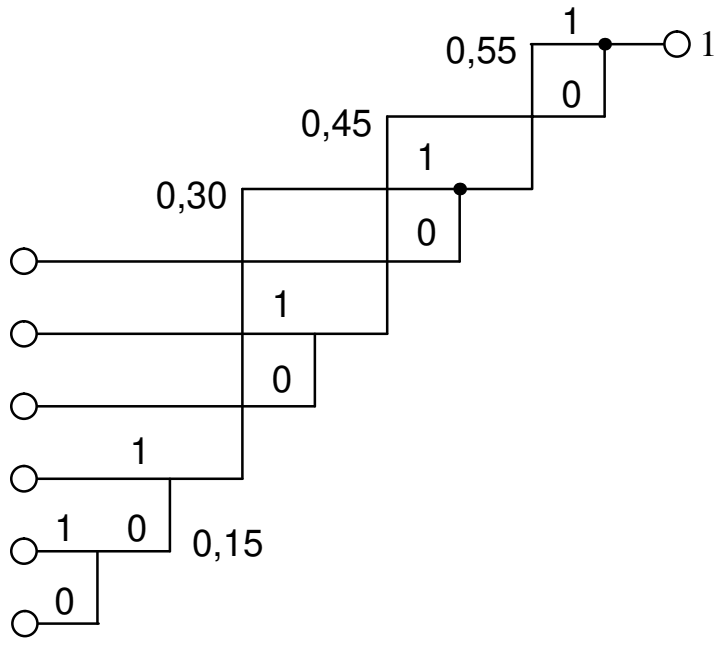


Bild 3.3 Beispiel für die Berechnung eines binären Huffman–Code

Beispiel: Für die in Bild 3.3 angegebene Quelle und Codierung berechnet sich

$$\begin{aligned} H(U) = & -[0,25 \cdot \log_2(0,25) + 0,25 \cdot \log_2(0,25) \\ & + 0,20 \cdot \log_2(0,20) + 0,15 \cdot \log_2(0,15) \\ & + 0,10 \cdot \log_2(0,10) + 0,05 \cdot \log_2(0,05)] \end{aligned}$$

$$H(U) = 2,42 \text{ bit / Symbol}$$

$$\begin{aligned} \overline{n} = & 0,25 \cdot 2 + 0,25 \cdot 2 + 0,20 \cdot 2 \\ & 0,15 \cdot 3 + 0,10 \cdot 4 + 0,05 \cdot 4 \end{aligned}$$

$$\overline{n} = 2,45 \text{ bit / Symbol}$$

Die verbleibende Redundanz der Codierung beträgt somit

$$R = \overline{n} - H(U) = 0,03 \text{ bit / Symbol.}$$

Ohne redundanzmindernde Codierung hätte die Redundanz bei einer Codierung mit 3 bit je Symbol

$$R = 3 - 2,42 = 0,58 \text{ bit / Symbol}$$

betragen.–

Die obere Grenze des Theorems der Quellencodierung kann noch eingeschränkt werden, indem nicht einzelne Quellensymbole, sondern jeweils Blöcke von L Quellensymbolen einer Symbolfolge mit einem Codewort codiert werden. Das Theorem der Quellencodierung geht dann in die Form

$$\frac{L \cdot H(U)}{\log_2 D} \leq L \cdot \overline{n} < \frac{L \cdot H(U)}{\log_2 D} + 1$$

bzw.

$$\frac{H(U)}{\log_2 D} \leq \overline{n} < \frac{H(U)}{\log_2 D} + \frac{1}{L} \tag{3.16}$$

über. Für hinreichend große L nähert sich damit die obere Grenze des Theorems 1.6 der unteren Grenze. Damit geht die je Symbol verbleibende Redundanz dieser redundanzmindernden Blockcodierung gegen Null.

Eine diskrete, gedächtnislose Quelle mit einem Symbolvorrat von K Quellensymbolen hat den mittleren Informationsgehalt $H(U)$ je Quellensymbol.

3.4 Codierung diskreter gedächtnisbehafteter Quellen

Es sei angenommen, daß eine diskrete, gedächtnisbehaftete Quelle eine zeitliche Folge von statistisch abhängigen Quellensymbolen $u_1, \ ..., \ u_N$ erzeugt, die alle einem Symbolvorrat von K Symbolen $a_1, \ ..., \ a_K$ entstammen. Es sei ferner angenommen, daß die betrachtete gedächtnisbehaftete Quelle eine diskrete Markoff–Quelle sei, bei der ein beliebiges Symbol u_N von jeweils nur $N - 1$ vorausgehenden Symbolen $u_1, \ ..., \ u_{N-1}$ statistisch abhängig ist. Jede Kombination von $N - 1$ vorausgehenden Symbolen $u_1, \ ..., \ u_{N-1}$ kennzeichnet einen Zustand s_j der Quelle. Mit der Erzeugung des Symbols u_N geht die Quelle in einen neuen Zustand über. Der bedingte mittlere Informationsgehalt je Quellensymbol sei $H(U_N|U_1, \ ..., \ U_{N-1})$.

Auf diese Markoff–Quelle können die bereits für gedächtnislose Quellen gewonnenen Ergebnisse angewendet werden, indem man jeden einzelnen Zustand s_j der Quelle für sich betrachtet. Zur Codierung der möglichen Symbole u_N wird in jedem Zustand s_j ein spezieller Präfixcode mit den Codewortlängen n_j gemäß

$$\frac{-\ log_2\ P(u_N|s_j)}{\log_2 D} \leq n_j < \frac{-\ log_2\ P(u_N|s_j)}{\log_2 D} + 1 \tag{3.17}$$

angewendet. Die mittlere Codewortlänge $\overline{n}_j$ je Quellensymbol für den im Zustand s_j angewandten Code ergibt sich dann aus dem Mittel über alle u_N mit der Grenze

$$\frac{H(U_N|s_j)}{\log_2(D)} \leq \overline{n}_j < \frac{H(U_N|s_j)}{\log_2(D)} + 1. \tag{3.18}$$

Mittelt man weiterhin über alle Zustände s_j, d.h. über alle Folgen $u_1,\ ...,\ u_{N-1}$, so erhält man für die Markoff–Quelle mit dem bedingten mittleren Informationsgehalt $H(U_N|U_1,\ ...,\ U_{N-1})$ je Symbol eine mittlere Codewortlänge $\overline{n}$ von

$$\frac{H(U_N|U_1,\ ...,\ U_{N-1})}{\log_2(D)} \leq \overline{n} < \frac{H(U_N|U_1,\ ...,\ U_{N-1})}{\log_2(D)} + 1. \tag{3.19}$$

Ein derartiger Code ist redundanzfrei, wenn die mittlere Codewortlänge $\overline{n} = H(U_N|U_1,\ ...,\ U_{N-1})/\log_2(D)$ ist, d.h. in Gl. (3.19) das Gleichheitszeichen gilt.

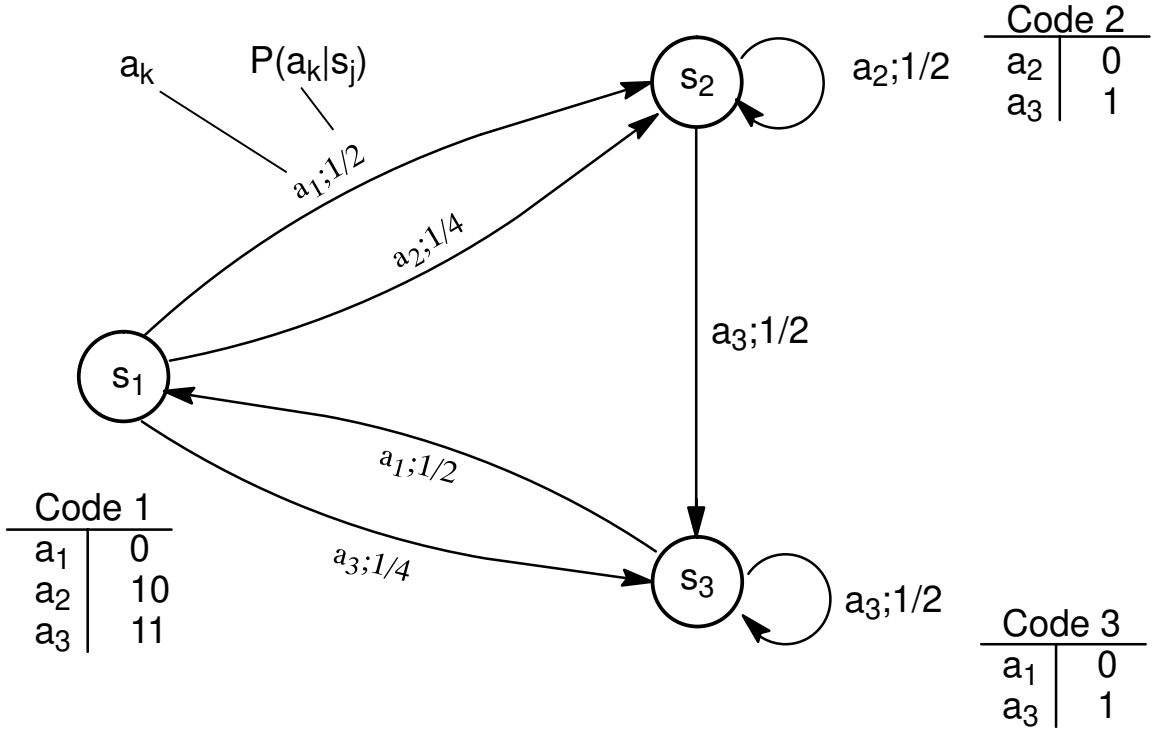


Bild 3.4 Graphische Darstellung einer Markoff–Quelle und ihrer Codierung

Beispiel: Bild 3.4 zeigt das Zustandsfolgediagramm einer Markoff–Quelle mit 3 Zuständen s_1, s_2, s_3, den Symbolwahrscheinlichkeiten $P(u_N|s_j)$ der Zustände und die zugehörigen Huffman–Codes.

Da die Markoff–Quelle bis zu $J = K^{N-1}$ verschiedene Zustände besitzt, werden zur Realisierung der erläuterten Codierung von Markoff–Prozessen bis zu J verschiedene Präfixcodes benötigt. Die Realisierung dieser Codierung ist wegen des damit verbundenen großen technischen Aufwandes nur in Sonderfällen durchführbar.

Die obere Grenze der Ungleichung (3.19) kann wiederum eingeschränkt werden, indem nicht einzelne Symbole u_N, sondern ganze Folgen $u_N,\ u_{N+1},\ ...,\ u_{N+L-1}$ von jeweils L Symbolen als ein Ereignis codiert werden. In Analogie zu Theorem 1.6 und der Blockcodierung (3.16) erhält man hier für die mittlere Codewortlänge die Grenzen

$$\frac{H(U_N|U_1,\ ...,\ U_{N-1})}{\log_2 D} \leq \overline{n} < \frac{H(U_N|U_1,\ ...,\ U_{N-1})}{\log_2 D} + \frac{1}{L}\,. \tag{3.20}$$

3.5 Anwendungsbeispiel

Für eine digitale Faksimile–Übertragung von Schwarz–Weiß–Bildvorlagen wird das durch zeilenweise Abtastung gewonnene Bildsignal in zwei Amplitudenstufen quantisiert. Jeder Abtastwert stellt einen schwarzen oder weißen Bildpunkt dar, der einen Entscheidungsgehalt von $H_0 = 1$ bit besitzt. Um die Übertragungszeit mittels einer redundanzreduzierenden Quellencodierung zu verkürzen, werden die Bildpunktfolgen in Abschnitte aufeinanderfolgender Bildpunkte gleicher Amplitude aufgeteilt und für jeden Abschnitt die sog. Lauflänge in codierter Form übertragen, siehe Bild 3.5. Dabei werden weiße und schwarze Lauflängen mit unterschiedlichen Codes codiert.

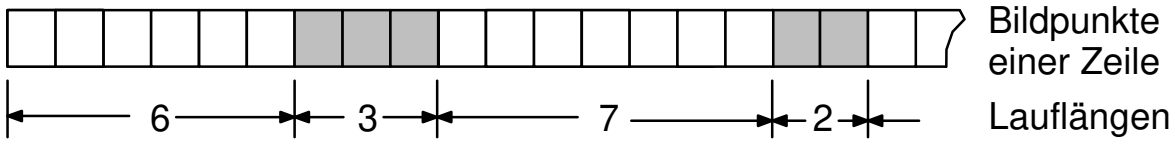


Bild 3.5 Lauflängencodierung (Run–length coding)

Legt man dieser Lauflängencodierung das Modell der gedächtnislosen Quelle zugrunde und geht dazu von der Annahme statistisch unabhängiger Lauflängen aus,

so beträgt der mittlere Informationsgehalt H_w und H_s der weißen und schwarzen Lauflängen

$$H_w = -\sum_{r=1}^{r_{\max}} P_w(r) \cdot \log_2 P_w(r)$$
$$H_s = -\sum_{r=1}^{r_{\max}} P_s(r) \cdot \log_2 P_s(r), \tag{3.21}$$

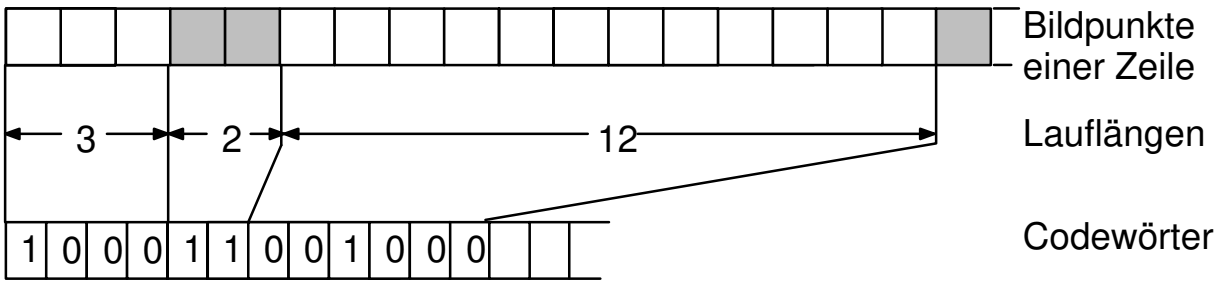
wobei r die jeweilige Lauflänge und $P(r)$ die Wahrscheinlichkeit der Lauflänge ist. Mit den mittleren Lauflängen $\overline{r}_w$ und $\overline{r}_s$ gemäß

$$\overline{r}_w = \sum_{r=1}^{r_{\max}} P_w(r) \cdot r$$
$$\overline{r}_s = \sum_{r=1}^{r_{\max}} P_s(r) \cdot r \tag{3.22}$$

folgt für die Entropie $H_{Bildpkt}$ eines Bildpunktes dieser Lauflängencodierung

$$H_{Bildpkt} = \frac{H_w + H_s}{\overline{r}_w + \overline{r}_s}. \tag{3.23}$$

Zur Codierung der weißen und schwarzen Lauflängen wendet man zwei Huffman–Codes an, deren Codewortlängen den Wahrscheinlichkeiten $P_w(r)$ und $P_s(r)$ angepaßt sind. Zur Verminderung des Realisierungsaufwandes werden jeweils nur die ersten 63 Lauflängen nach Huffman codiert und größere Lauflängen durch zwei Codewörter dargestellt, wobei das erste Codewort eine Lauflänge von $n \cdot 64$ und das zweite Codewort den verbleibenden Teil der Lauflänge darstellt. Die verbleibende Lauflänge kann auch Null sein. Bild 3.6 zeigt den international standardisierten sog. Modified Huffman–Code für den Telefax–Dienst. Eine Lauflänge von $r_w = 130$ wird demzufolge mittels der Binärzeichen 10010 0111 codiert. Die Decodierung ist aufgrund der Präfixeigenschaften der Codewörter eindeutig.



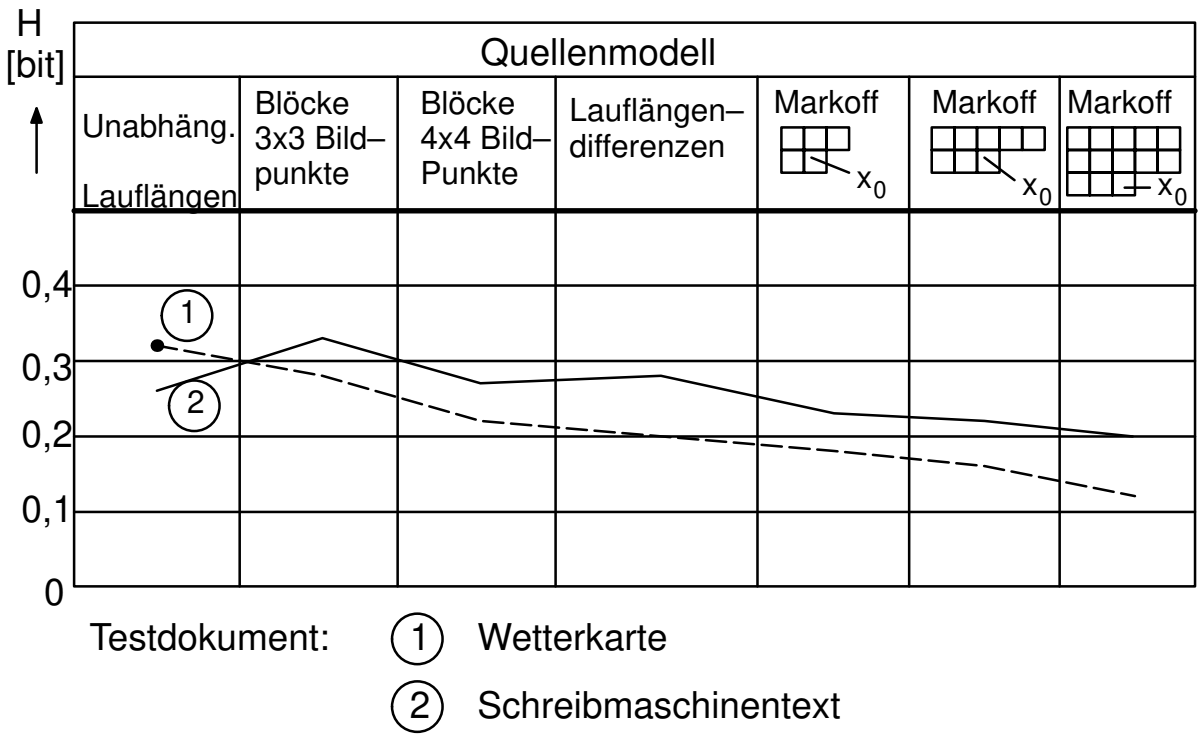
a)

b)

Lauflänge	Codewörter	
	Weiße Lauflänge	Schwarze Lauflänge
0	00110101	0000110111
1	000111	010
2	0111	11
3	1000	10
.	.	.
.	.	.
63	00110100	000001100111
64	11011	0000001111
128	10010	000011001000
192	010111	000011001001
.	.	.
.	.	.
1728	010011011	0000001100101

Bild 3.6 Modified Huffman–Code
a) Beispiel einer Codierung
b) Tabelle des Modified Huffman–Code

Weitere zur Codierung eines binären Faksimilesignals anwendbare Quellenmodelle sind in Bild 3.7 zusammen mit den zugehörigen Entropien je Bildpunkt dargestellt. Da die Entropien den unteren Grenzwerten für die erreichbare Übertragungsrate entsprechen, erlauben sie auch eine Bewertung unterschiedlicher Quellenmodelle.



Testdokument: ① Wetterkarte

② Schreibmaschinentext

Bildauflösung: 3,8 Zeilen / mm, 4 Bildpunkte / mm entlang der Zeile

Bild 3.7 Gemessene Entropie H je Bildpunkt für verschiedene Quellenmodelle

4. Kanäle

Um einen physikalischen Kanal einer mathematischen Behandlung zugänglich zu machen, muß der Kanal durch ein Kanalmodell dargestellt werden. Das Kanalmodell beschreibt die Störeinflüsse eines Kanals durch Zufallsprozesse. Zur Aufstellung eines Kanalmodells müssen die Störeinflüsse eines betrachteten physikalischen Kanals in Form von Fehlermustern gemessen werden.

Mit Hilfe eines Kanalmodells läßt sich die Kanalkapazität eines Kanals berechnen. Sie ist ein Maß für den mit einem Kanalsymbol übertragbaren Informationsgehalt. Das Kanalmodell kann ferner auch dazu verwendet werden, eine Kanalcodierung im Hinblick auf eine möglichst geringe Restfehlerwahrscheinlichkeit zu optimieren.

Nachfolgend werden zwei Kanalmodelle, der diskrete, gedächtnislose und der kontinuierliche, gedächtnislose Kanal behandelt. Zur Berechnung der Kanalkapazität muß zunächst der Begriff des wechselseitigen Informationsgehaltes erörtert werden.

4.1 Fehlerstrukturen realer Kanäle

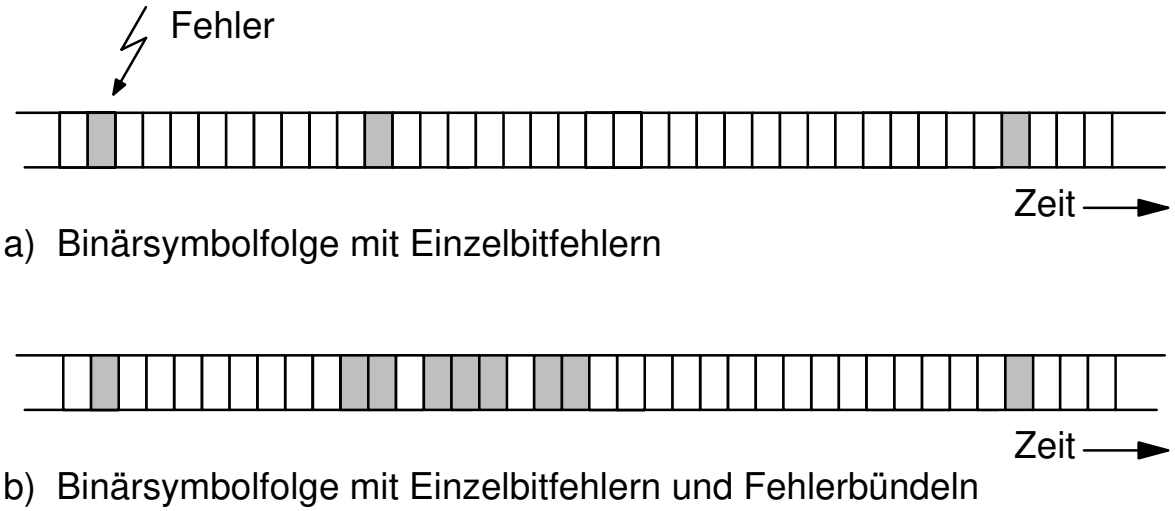


Bild 4.1 Beispiele für das Fehlerverhalten realer Kanäle

Bild 4.1 zeigt Beispiele typischen Fehlerverhaltens realer, diskreter, binärer Kanäle. Man unterscheidet zwischen Kanälen mit

a) Einzelbitfehlern
b) Bündelfehlern
c) gemischten Einzel– und Bündelfehlern

Die Fehlerursachen können verschiedener Art sein. Bild 4.2 enthält eine Zusammenstellung einiger für die Anwendung bedeutender Kanäle und ihres Fehlerverhaltens.

Kanal	Fehlerverhalten
Telefonleitung mit Modem und Vermittlung	Bündelfehler infolge mechanischer Wählvorgänge benachbarter Relais in der Vermittlung
Terrestrischer Richtfunk und Mobilfunk	Bündelfehler durch Fading und Dispersion
Satellitenstrecken	Einzelbitfehler durch thermisches Rauschen
Speichermedien, Magnetband, Magnetplatte	Einzelbitfehler und Bündelfehler infolge Fehler im Magnetmaterial und schwankenden Band–Kopf–Abstandes

Bild 4.2 Beispiele digitaler Übertragungs– und Aufzeichnungskanäle

Das infolge von Übertragungsfehlern entstehende Fehlermuster soll folgendermaßen definiert werden.
Es sei

$$\vec{x} = x_1,\ x_2,\ ...,\ x_N$$

eine Folge gesendeter Binärsymbole, wobei

$$x_n = 0 \text{ oder } 1$$

ist und

$$\vec{y} = y_1,\ y_2,\ ...,\ y_N$$

eine Folge empfangener, unter Umständen fehlerhafter Binärsymbole. Dann ist das Fehlermuster

$$\vec{e} = e_1,\ e_2,\ ...,\ e_N$$

definiert durch

$$\vec{e} = \vec{x} \oplus \vec{y}. \tag{4.1}$$

Die Operation "$\oplus$" entspricht der Modulo–2–Addition ↑ mit den Rechenregeln

$$\begin{aligned} 0 \oplus 0 &= 0 \\ 0 \oplus 1 &= 1 \\ 1 \oplus 0 &= 1 \\ 1 \oplus 1 &= 0. \end{aligned}$$

Beispiel:

$\vec{x} = 000101101$
$\vec{y} = 010101001$
$\vec{e} = 010000100$

Das Fehlermuster enthält an Stellen fehlerhafter Binärsymbole eine Eins und sonst Nullen. Bild 4.3 zeigt als Beispiel die mit einem Faksimilegerät aufgezeichneten Fehlermuster einer Datenübertragung mit Modem über eine Telefonleitung.–

Zur Unterscheidung von Einzelbitfehlern und Fehlerbündeln werden folgende Definitionen eingeführt:

Fehlerbündel:
Folge von Codeelementen, in der zwischen aufeinanderfolgenden fehlerhaften Codeelementen immer weniger als r korrekte Codeelemente liegen. Das letzte fehlerhafte Codeelement in einem Fehlerbündel und das erste fehlerhafte Codeelement eines folgenden Fehlerbündels sind demzufolge durch mindestens r richtige Codeelemente getrennt.

Fehlerbündellänge:
Die Anzahl der in einem Fehlerbündel enthaltenen Codeelemente.

a)

b)

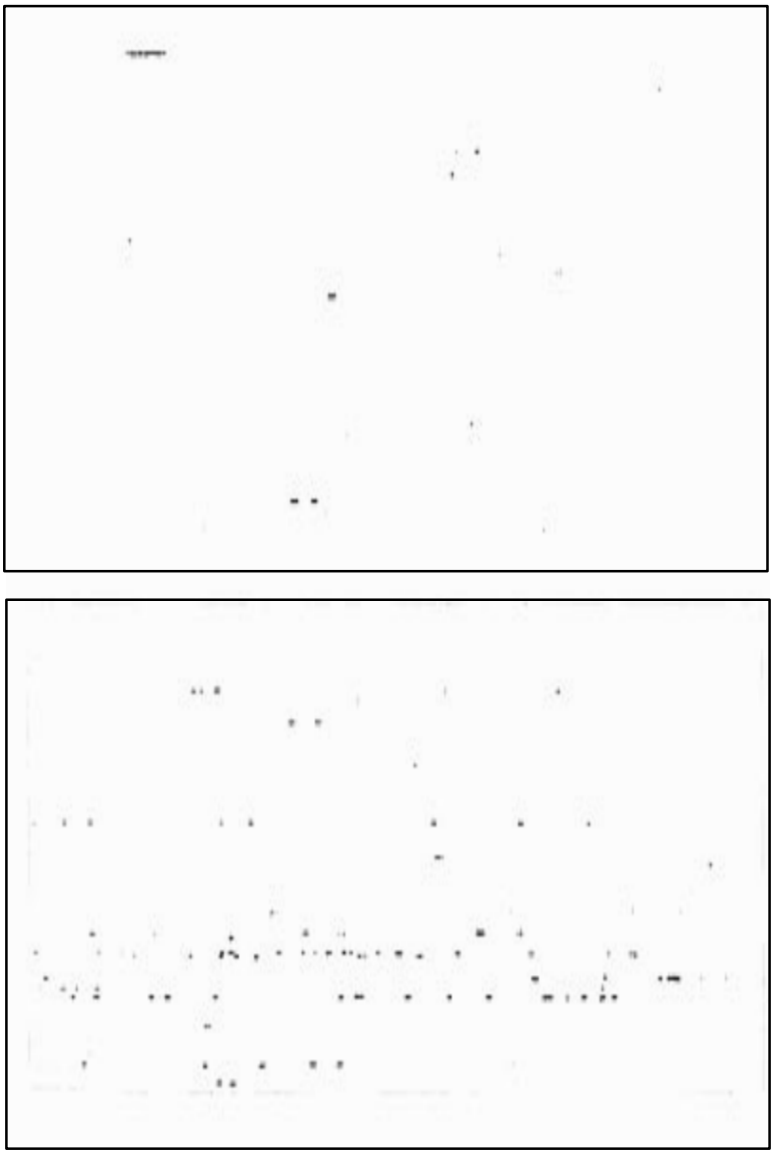


Bild 4.3 Beispiele gemessener Fehlermuster einer Datenübertragung mit Modem über einen Telefonkanal mit 4,8 kbit/s Übertragungsbitrate. Die Auflösung der gezeigten Muster beträgt 1728 bit/ Zeile und 7,7 Zeilen/mm.

a) Fehlerwahrscheinlichkeit $P_e = 5,8 \cdot 10^{-5}$
b) Fehlerwahrscheinlichkeit $P_e = 6,1 \cdot 10^{-4}$

Beispiel eines Fehlermusters: ...00000110011000100000... .
Dieses Fehlermuster enthält im Fall $r = 2$ insgesamt 2 Fehlerbündel der Länge 2 und einen Einzelbitfehler. Im Fall $r = 4$ enthält es nur noch ein Fehlerbündel der

Länge 10. Die genannte Definition des Fehlerbündels ermöglicht über die Festlegung der Größe r eine Anpassung an den jeweiligen Anwendungsfall.

4.2 Diskreter, gedächtnisloser Kanal

Ein relativ einfaches, aber dennoch sehr nützliches Modell zur Beschreibung eines Kanals mit statistisch unabhängigen Einzelfehlern ist das Modell des diskreten, gedächtnislosen Kanals. Dem Eingang eines diskreten, gedächtnislosen Kanals werden die Symbole $a_1,\ ...,\ a_k,\ ...,\ a_K$ zugeführt. Der Symbolvorrat am Kanalausgang sei $b_1,\ ...,\ b_j,\ ...,\ b_J$. Ein diskreter, gedächtnisloser Kanal ist dann durch die Übergangswahrscheinlichkeiten $P(b_j \mid a_k)$ definiert, die angeben, mit welcher Wahrscheinlichkeit ein gesendetes Symbol a_k in ein empfangenes Symbol b_j verändert wird.

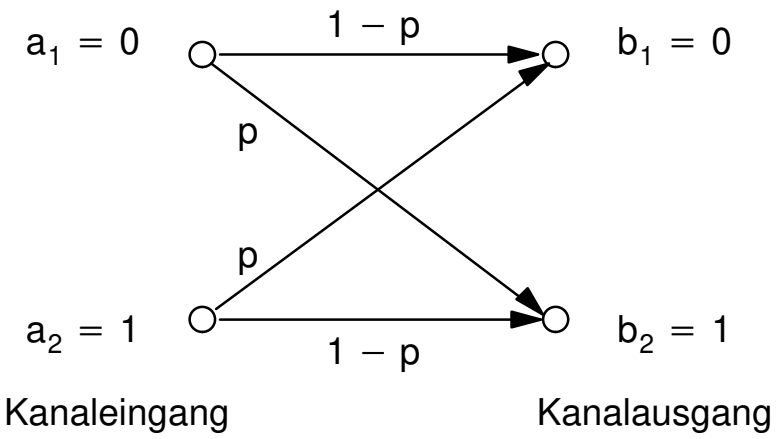
Es sei $\vec{x} = x_1,\ ...,\ x_n,\ ...,\ x_N$ eine Folge von Eingangssymbolen, wobei jedes x_n für ein Symbol des Alphabetes $a_1,\ ...,\ a_K$ steht und $\vec{y} = y_1,\ ...,\ y_n,\ ...,\ y_N$ die zugehörige Symbolfolge am Kanalausgang, wobei jedes y_n für ein Symbol des Alphabetes $b_1,\ ...,\ b_J$ steht.

Im Fall des gedächtnislosen Kanals ist jedes Ausgangssymbol einer Folge nur von dem entsprechenden Eingangssymbol statistisch abhängig. Somit gilt für eine N–stellige Symbolfolge:

$$P_N(\vec{y}|\vec{x}) = \prod_{n=1}^{N} P(y_n|x_n)\ . \tag{4.2}$$

Binärer, symmetrischer Kanal (BSK)

Der binäre, symmetrische Kanal kann als Sonderfall des diskreten, gedächtnislosen Kanals angesehen werden. Er wird durch das Flußdiagramm in Bild 4.4 beschrieben.



$$P(b_1|a_1) = P(b_2|a_2) = 1 - p$$
$$P(b_2|a_1) = P(b_1|a_2) = p$$

Bild 4.4 Flußdiagramm eines binären, symmetrischen Kanals

Die Fehlerwahrscheinlichkeit P_e des binären, symmetrischen Kanals beträgt demzufolge

$$\begin{aligned} P_e &= P(b_2|a_1) \cdot P(a_1) + P(b_1|a_2) \cdot P(a_2) \\ P_e &= p \cdot P(a_1) + p \cdot P(a_2) = p \end{aligned} \tag{4.3}$$

und es ist $1 - p$ die Wahrscheinlichkeit für richtige Übertragung eines Symbols. Im Fall des binären, symmetrischen Kanals entsprechen die bedingten Wahrscheinlichkeiten $P(\vec{y}|\vec{x})$ nach Gl. (4.2) den Wahrscheinlichkeiten der zugehörigen Fehlermuster $\vec{e}$, d.h.

$$P(\vec{y}|\vec{x}) = P(y_1, \ ..., \ y_N|x_1, \ ..., \ x_N) = P(e_1, \ ..., \ e_N) \, . \tag{4.4}$$

Beispiel:

$$\begin{aligned} P(001|011) &= P(e_1 = 0, \ e_2 = 1, \ e_3 = 0) \\ &= (1 - p) \cdot p \cdot (1 - p) \\ &= (1 - p)^2 \cdot p \end{aligned}$$

4.3 Wechselseitiger Informationsgehalt und Transinformationsgehalt diskreter Ensembles

Der wechselseitige Informationsgehalt ist ein Maß für den Informationsgehalt, der mit der Kenntnis eines Symbols b_j über ein Symbol a_k eines zweiten Ensembles vermittelt wird. Ist a_k beispielsweise ein gesendetes Symbol am Kanaleingang und b_j ein empfangenes Symbol am Kanalausgang, so kann der wechselseitige Informationsgehalt auch als der mit einem Symbol übertragene Informationsgehalt interpretiert werden. In diesem Fall bezeichnet man den wechselseitigen Informationsgehalt auch als Transinformationsgehalt.

Dieser Zusammenhang soll anhand des in Bild 4.5 dargestellten Übertragungssystems deutlich gemacht werden.

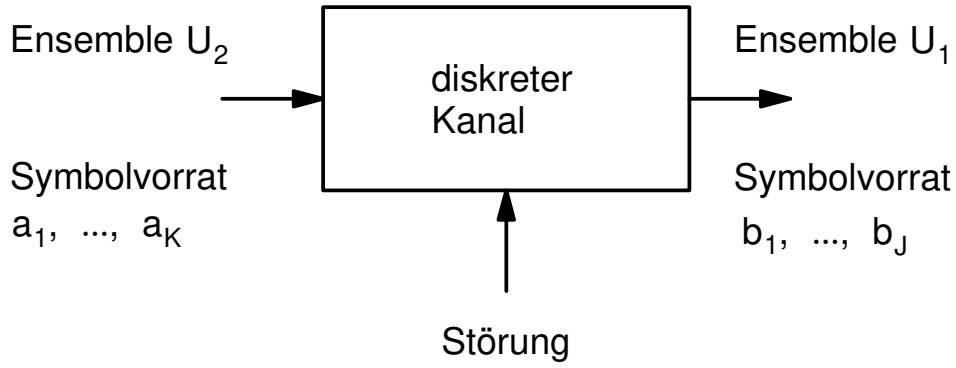


Bild 4.5 Übertragungskanal für diskrete Informationen

Es sei U_2 das Nachrichtenensemble am Eingang des Kanals mit den Symbolen a_1, ..., a_K und U_1 das Nachrichtenensemble am Ausgang des Kanals mit den Symbolen b_1, ..., b_J. Jedes dem Eingang des Kanals zugeführte Symbol u_2 aus dem Ensemble U_2 möge am Ausgang des Kanals irgendein Symbol u_1 aus dem Ensemble U_1 erzeugen. Im Fall eines ungestörten Übertragungskanals erzeugt ein gesendetes Symbol u_2 immer das gleiche Symbol u_1; im Fall eines gestörten Kanals kann das nicht vorausgesetzt werden, da das gesendete Symbol u_2 infolge einer Störung in ein beliebiges Symbol u_1 verfälscht worden sein kann.

Es soll nun der Informationsgehalt berechnet werden, den ein empfangenes Symbol $u_1 = b_j$ über ein gesendetes Symbol $u_2 = a_k$ vermittelt. Vor dem Empfang des Symbols $u_1 = b_j$ war die Wahrscheinlichkeit für das Auftreten von $u_2 = a_k$ durch $P(a_k)$ gegeben; mit dem Empfang von $u_1 = b_j$ hat sich diese Wahrscheinlichkeit in

$P(a_k|b_j)$ geändert. Damit hat sich die Ungewißheit über das gesendete Symbol $u_2 = a_k$ durch den Empfang von $u_1 = b_j$ um den Wert

$$\begin{aligned} T(a_k; b_j) &= -\log_2 P(a_k) + \log_2 P(a_k|b_j) \\ &= \log_2 \frac{P(a_k|b_j)}{P(a_k)} \end{aligned} \tag{4.5}$$

geändert. Diese Änderung der Ungewißheit wird durch den mit Empfang von $u_1 = b_j$ vermittelten Informationsgehalt erreicht. $T(a_k; b_j)$ ist der von b_j über a_k vermittelte Informationsgehalt, der sog. Transinformationsgehalt. Da nachzuweisen ist, daß

$$T(a_k; b_j) = T(b_j; a_k) \tag{4.6}$$

gilt, wird diese Größe auch allgemein als wechselseitiger Informationsgehalt bezeichnet.

Der im Mittel mit einem empfangenen Symbol u_1 über ein gesendetes Symbol u_2 übertragene Transinformationsgehalt beträgt

$$\begin{aligned} T(U_2; U_1) &= \sum_{k=1}^{K} \sum_{j=1}^{J} P(a_k, b_j) \cdot \log_2 \frac{P(a_k|b_j)}{P(a_k)} \\ T(U_2; U_1) &= \sum_{u_1} \sum_{u_2} P(u_1, u_2) \cdot \log_2 \frac{P(u_2|u_1)}{P(u_2)} \end{aligned} \tag{4.7}$$

Durch Umformung von Gl. (4.7) erhält man

$$\begin{aligned} T(U_2; U_1) &= -\sum_{u_2} P(u_2) \cdot \log_2 P(u_2) \\ &\quad + \sum_{u_1} \sum_{u_2} P(u_1, u_2) \cdot \log_2 P(u_2|u_1) \\ T(U_2; U_1) &= H(U_2) - H(U_2|U_1)\,. \end{aligned} \tag{4.8}$$

Mit den Beziehungen

$$\begin{aligned} H(U_1, U_2) &= H(U_2, U_1) \\ H(U_1|U_2) + H(U_2) &= H(U_2|U_1) + H(U_1) \end{aligned} \tag{4.9}$$

folgt aus Gl. (4.8) eine zweite Darstellungsform für den mittleren Transinformationsgehalt

$$T(U_2; U_1) = H(U_1) - H(U_1|U_2). \tag{4.10}$$

Sind die empfangenen Symbole u_1 statistisch unabhängig von den gesendeten Symbolen u_2, so ist $H(U_1|U_2) = H(U_1)$ und es folgt $T(U_2; U_1) = 0$. In diesem Fall beinhaltet das Ensemble U_1 keinen Informationsgehalt über das Ensemble U_2, d.h. es wird keine Information übertragen.

Theorem 1.7:

Der Transinformationsgehalt $T(U_2; U_1)$ genügt der Ungleichung

$$T(U_2; U_1) \geq 0 \tag{4.11}$$

in der das Gleichheitszeichen dann und nur dann gilt, wenn U_2 und U_1 statistisch unabhängig sind.

Dieses Theorem kann durch Einsetzen von Theorem 1.3 in Gl. (4.8) leicht nachgewiesen werden.

4.4 Wechselseitiger Informationsgehalt und Transinformationsgehalt kontinuierlicher Ensembles

Die bisherigen Betrachtungen bezogen sich auf Kanäle mit diskreten Ensembles U_2 und U_1 am Kanaleingang und Kanalausgang. Es sollen nun wechselseitiger Informationsgehalt und Transinformationsgehalt für Kanäle mit kontinuierlichen Zufallsprozessen am Kanaleingang und Kanalausgang betrachtet werden. Vereinfachend sei dazu angenommen, daß die Zufallsprozesse bandbegrenzt sind und die Frequenzbandgrenzen $\pm W$ aufweisen. Die Zufallsprozesse sind aufgrund des Abtasttheorems demzufolge durch die Abtastwerte im zeitlichen Abstand

$$T = \frac{1}{2W}$$

eindeutig beschrieben. Die kontinuierlichen Zufallsprozesse können in diesem Fall durch zeitdiskrete, wertkontinuierliche Zufallsprozesse dargestellt werden. Zum Hinweis auf den unendlichen Wertevorrat der so erzeugten Ensembles werden diese anstelle von U_2 und U_1 mit X und Y gekennzeichnet.

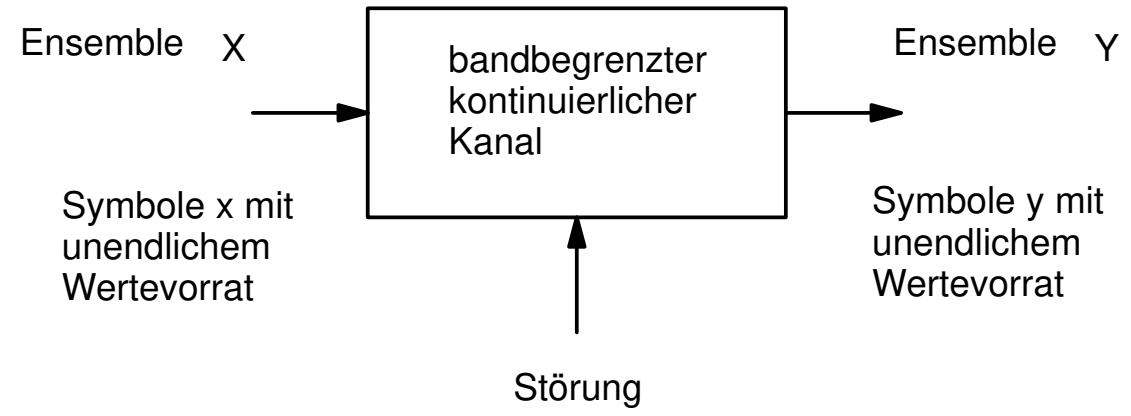


Bild 4.6 Übertragungskanal für bandbegrenzte, kontinuierliche Informationen

Es sei angenommen, daß am Kanaleingang ein Nachrichtenensemble X liegt, dessen einzelne Symbole x einem kontinuierlichen Wertevorrat entstammen, wobei jedes x einen Wert im Bereich $-\infty \leq x \leq +\infty$ annehmen kann. Jedem möglichen Wert x ist anstelle einer Wahrscheinlichkeit hier eine Wahrscheinlichkeitsdichte $p(x)$ zugeordnet.

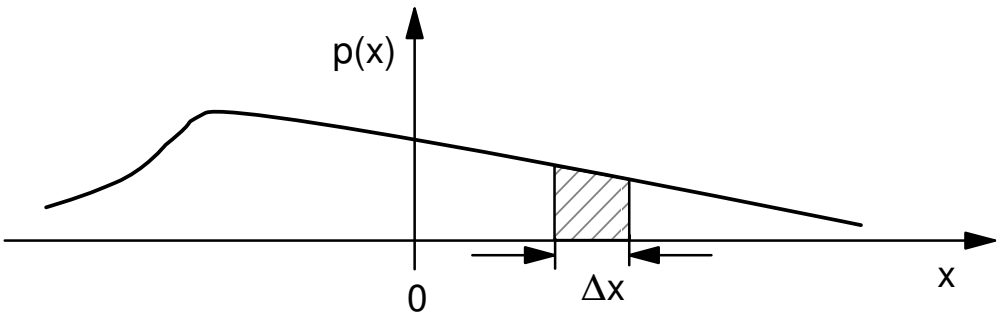


Bild 4.7 Wahrscheinlichkeitsdichte *p(x)*, dargestellt über einem quantisierten Wertebereich x

Zur Berechnung des Informationsgehaltes eines Symbols x sei zunächst angenommen, daß der kontinuierliche Wertevorrat von x, wie in Bild 4.7 dargestellt, in Stufen von Δx quantisiert und damit diskret ist. Die Intervalle seien so klein, daß $p(x)$ im Intervall Δx als konstant vorausgesetzt werden kann. Der Informationsgehalt $I(x)$ eines Symbols x aus einem Intervall Δx beträgt dann

$$I(x) = -\log_2[p(x) \cdot \Delta x]. \tag{4.12}$$

Im Grenzübergang $\Delta x \to 0$ folgt für den Informationsgehalt $I(x) \to \infty$. Dieses Ergebnis läßt sich wie folgt erklären: Wollte man einen beliebigen Zahlenwert x

beispielsweise in Dezimalzahlendarstellung angeben, so benötigt man zur exakten Angabe unendlich viele Dezimalstellen, was einem unendlichen Informationsgehalt von x entspricht. Aus physikalischen Gründen wird ein Wert x, z.B. der Abtastwert einer kontinuierlichen Zeitfunktion, immer nur mit endlicher Genauigkeit zur Verfügung stehen. Da $I(x)$ nach Gl. (4.12) einerseits als der Informationsgehalt eines mit Quantisierungsfehlern behafteten Symbols x eines kontinuierlichen Wertevorrates interpretiert werden kann, andererseits $I(x)$ der Informationsgehalt eines diskreten Wertevorrates mit K Quantisierungsstufen ist, kann das kontinuierliche Ensemble auch als Grenzfall eines diskreten Ensembles angesehen werden.

Berechnet man den mittleren Informationsgehalt $\overline{I(x)}$ eines Symbols x wie bei diskreten Ensembles und führt den Grenzübergang $\Delta x \to 0$ aus, so erhält man

$$\begin{aligned}
\lim_{\Delta x \to 0} \overline{I(x)} &= -\int_{-\infty}^{+\infty} p(x)dx \cdot \log_2[p(x)dx] \\
&= -\int_{-\infty}^{+\infty} p(x) \cdot \log_2 p(x)\, dx - \int_{-\infty}^{+\infty} p(x) \cdot \log_2 dx\, dx \qquad (4.13) \\
&= -\int_{-\infty}^{+\infty} p(x) \cdot \log_2 p(x)\, dx - \log_2 dx
\end{aligned}$$

Der zweite Summand nimmt unabhängig von der Wahrscheinlichkeitsdichte $p(x)$ stets den unbestimmten Wert $-\log_2 dx$ an. Dieses Resultat des mittleren Informationsgehaltes wäre unbrauchbar, wenn die in der Informationstheorie interessierenden Größen nicht Differenzen von Entropien wären. Bei der Bildung von Differenzen fällt der unbestimmte Teil heraus. Es erweist sich als zweckmäßig, die Entropie $H(X)$ eines kontinuierlichen Ensembles mit der Wahrscheinlichkeitsdichte $p(x)$ durch

$$H(X) = -\int_{-\infty}^{+\infty} p(x) \cdot \log_2 p(x)\, dx \qquad (4.14)$$

zu definieren. Die Entropie $H(X)$ ist im Gegensatz zu der Entropie diskreter Ensembles vom Koordinatensystem abhängig und somit nicht transformationsinvariant. Aufgrund der oben angeführten Erläuterungen ist diese Entropie des kontinuierlichen Ensembles kein absolutes Maß für den Informationsgehalt. Nur Differenzen

von Entropien kontinuierlicher Ensembles können als Informationsgehalt interpretiert werden.

Die bedingte Entropie $H(X|Y)$ der kontinuierlichen Ensembles X und Y wird auf ähnliche Weise durch

$$H(X|Y) = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x,y) \cdot \log_2 p(x|y)\, dx\, dy \tag{4.15}$$

und die Entropie $H(X,Y)$ des kontinuierlichen Verbundensembles durch

$$H(X,Y) = -\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} p(x,y) \cdot \log_2 p(x,y)\, dx\, dy \tag{4.16}$$

definiert. Es kann gezeigt werden, daß das Theorem auch für die Entropien kontinuierlicher Ensembles gültig ist.

Beispiel: Berechnung der Entropie $H(X)$ eines kontinuierlichen Ensembles X mit einer eindimensionalen Gaußschen Wahrscheinlichkeitsdichte $p_G(x)$ gemäß Bild 4.8 mit

$$p_G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp(-x^2 / 2\sigma^2), \tag{4.17}$$

dem linearen Mittelwert $m_x = 0$ und der Varianz σ^2.

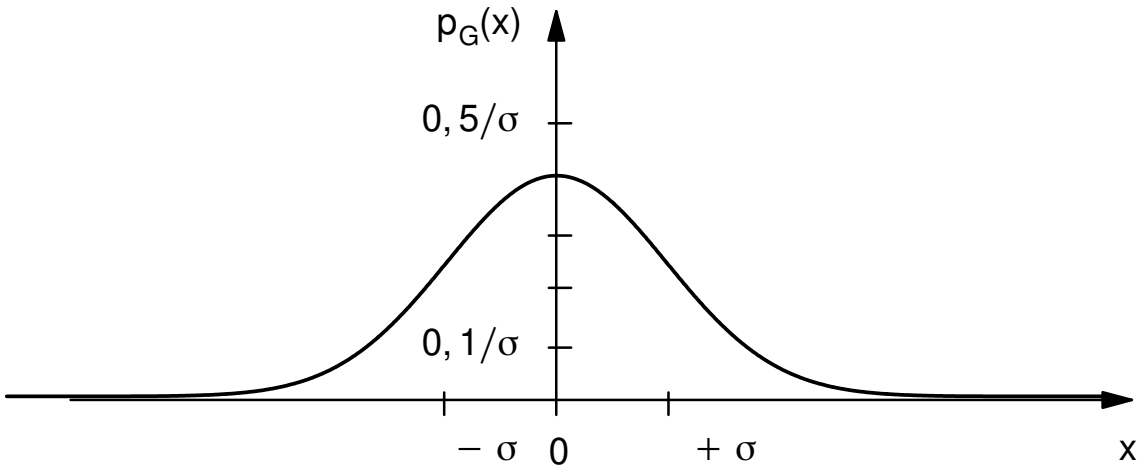


Bild 4.8 Graphische Darstellung der Gaußschen Wahrscheinlichkeits–dichte $p_G(x)$

Unter Verwendung der Gl. (2.8) und (4.17) wird

$$-\log_2 p_G(x) = \log_2\sqrt{2\pi}\cdot\sigma + \frac{x^2}{2\sigma^2}\log_2 e$$

$$H_G(X) = -\int_{-\infty}^{+\infty} p_G(x)\cdot\log_2 p_G(x)2dx$$

$$H_G(X) = -\int_{-\infty}^{+\infty} p_G(x)\left[\log_2\sqrt{2\pi}\cdot\sigma + \frac{x^2}{2\sigma^2}\log_2 e\right]dx \tag{4.18}$$

$$H_G(X) = \log_2\sqrt{2\pi}\cdot\sigma + \frac{\sigma^2}{2\sigma^2}\log_2 e$$

$$H_G(X) = \frac{1}{2}\log_2(2\pi e\sigma^2) \tag{4.19}$$

Die Auswertung von Gl. (4.18) führt für beliebige Wahrscheinlichkeitsdichten $p(x) \neq p_G(x)$ stets auf das gleiche Ergebnis (4.19). Diese Eigenschaft wird später noch verwendet.

Die Gaußsche Wahrscheinlichkeitsdichte $p_G(x)$ nach Gl. (4.17) besitzt die besondere Eigenschaft, daß sie unter allen Wahrscheinlichkeitsdichten mit $m_x = 0$ und vorgegebener Varianz σ^2 die maximale Entropie aufweist.

Theorem 1.8:

Die Entropie $H(X)$ eines kontinuierlichen Ensembles X mit dem Mittelwert $m_x = 0$ und der Varianz σ^2 genügt der Ungleichung

$$H(X) \leq \frac{1}{2}\log_2(2\pi e\sigma^2), \tag{4.20}$$

in der das Gleichheitszeichen dann und nur dann gilt, wenn $p(x) = p_G(x)$ eine eindimensionale Gaußsche Wahrscheinlichkeitsdichte ist.

Beweis:

Anstelle von Gl. (4.20) wird gezeigt, daß

$$H(X) - \frac{1}{2}\log_2(2\pi e\sigma^2) \leq 0 \tag{4.21}$$

gilt. Mit Hilfe der Beziehungen (2.7), (2.8) und (4.18) erhält man

$$
\begin{aligned}
H(X) - \frac{1}{2}\,\log_2\,(2\pi e\sigma^2) \;&=\; \int\limits_{-\infty}^{+\infty} p(x)\cdot \log_2\,\frac{p_G(x)}{p(x)}\,dx \\
&\leq \int\limits_{-\infty}^{+\infty} p(x)\cdot\left[\,\frac{p_G(x)}{p(x)} - 1\,\right]\log_2 e\;dx \\
&\leq\; 0\;.
\end{aligned}
\tag{4.22}
$$

Das Gleichheitszeichen dieser Ungleichung gilt dann und nur dann, wenn $p_G(x) = p(x)$ für alle x ist.

Die Verbundentropie eines N–dimensionalen kontinuierlichen Verbundensembles $X_1,\ ...,\ X_N$ mit einer N–dimensionalen Gaußschen Wahrscheinlichkeitsdichte

$$
P_G(x_1,\ ...,\ x_N) = \frac{1}{(2\pi)^{N/2}\,|\Lambda|^{1/2}}\cdot
\exp\left[-\frac{1}{2|\Lambda|}\sum_{n=1}^{N}\sum_{m=1}^{N}|\Lambda|_{nm}\,(x_n - m_n)\,(x_m - m_m)\right],
$$

den linearen Mittelwerten $m_n = m_m = 0$ und den vorgegebenen Kovarianzen $\lambda_{nm} = E[(x_n - m_n)\,(x_m - m_m)]$ besitzt die Entropie

$$
H(X_1,\ ...,\ X_N) = \frac{1}{2}\log_2(2\pi e|\Lambda|)\,, \tag{4.23}
$$

wobei Λ die Kovarianzmatrix ist. Der Beweis kann auf ähnliche Weise wie für die eindimensionale Gaußsche Wahrscheinlichkeitsdichte erbracht werden.–

Da sich der Transinformationsgehalt aus der Differenz zweier Entropien berechnet, entfällt die Problematik des unbestimmten Ausdrucks, und der Transinformationsgehalt des bandbegrenzten, kontinuierlichen Übertragungskanals gemäß Bild 4.6 kann entsprechend Gl. (4.8) und (4.10) mittels der Beziehung

$$
T(X;Y) = H(X) - H(X|Y) \tag{4.24}
$$

$$
= H(Y) - H(Y|X) \tag{4.25}
$$

berechnet werden.

4.5 Kanalkapazität diskreter, gedächtnisloser Kanäle

Im Abschnitt 4.3 wurde gezeigt, daß der mittlere Transinformationsgehalt ein Maß für den mit einem Kanalsymbol übertragenen Informationsgehalt ist. Mit dem Nachrichtenensemble U_2 am Kanaleingang und dem Nachrichtenensemble U_1 am Kanalausgang konnten für den mittleren Transinformationsgehalt die Beziehungen (4.8) und (4.10)

$$\begin{aligned} T(U_2;U_1) &= H(U_2) - H(U_2|U_1) \\ &= H(U_1) - H(U_1|U_2) \end{aligned} \tag{4.26}$$

abgeleitet werden. Die Informationsübertragung über einen gestörten Kanal und der Begriff des Transinformationsgehaltes sollen anhand des Bildes 4.9 veranschaulicht werden.

Eine diskrete, gedächtnislose Quelle mit der Entropie $H(U_2)$ und dem Symbolvorrat $a_1,\ ...,\ a_K$ ist mit dem Eingang eines gestörten Kanals verbunden. Infolge von Störungen auf dem Kanal werden nicht alle von der Quelle gesendeten Symbole unverfälscht übertragen und am Kanalausgang richtig empfangen. Auf diese Weise gelangt nur ein Teil des gesendeten Informationsgehaltes $H(U_2)$ je Symbol, nämlich der mittlere Transinformationsgehalt $T(U_2;U_1)$ zur Sinke. Die sog. Äquivokation $H(U_2|U_1)$ geht verloren. Dieser Zusammenhang wird durch den ersten Teil der Gl. (4.26) ausgedrückt.

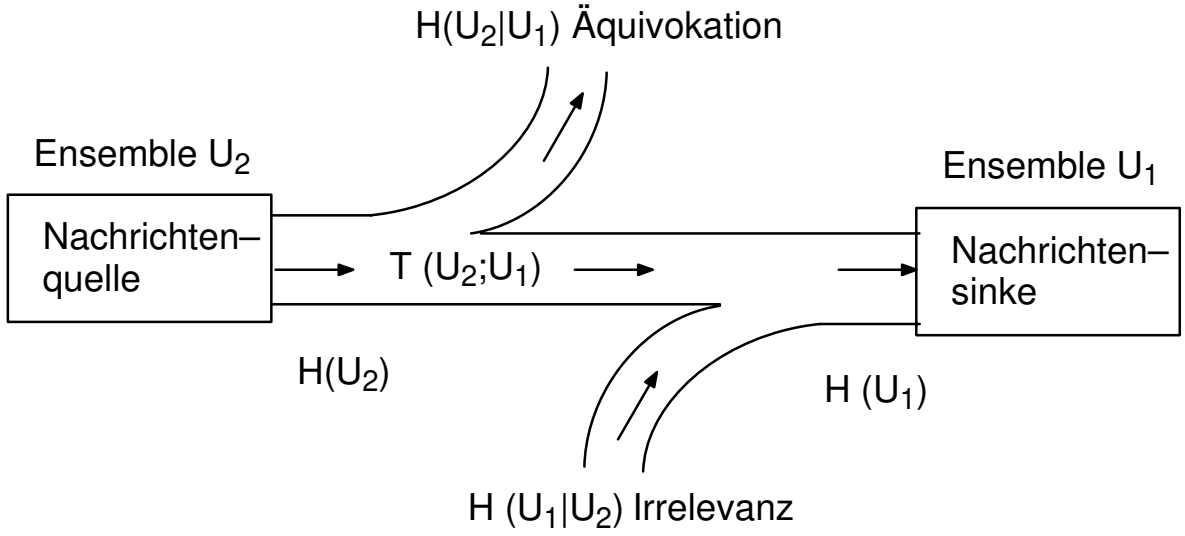


Bild 4.9 Schema eines gestörten Kanals

Für die Nachrichtensinke wirkt der Kanalausgang wie eine Quelle mit der Entropie $H(U_1)$. Diese setzt sich aus dem mittleren Transinformationsgehalt $T(U_2;U_1)$ und

der sog. Irrelevanz $H(U_1|U_2)$ zusammen, die nicht von der Nachrichtenquelle, sondern nur von den Störungen des Kanals herrühren kann. Dieser Zusammenhang kommt durch den zweiten Teil der Gl. (4.26) zum Ausdruck.

Der mittlere Transinformationsgehalt nach Gl. (4.8) hängt von den Übergangswahrscheinlichkeiten $P(u_1|u_2)$ des gegebenen Kanals sowie von der Wahrscheinlichkeitsverteilung $P(u_2)$ der gesendeten Symbole u_2 ab. Die Wahrscheinlichkeitsverteilung $P(u_2)$ kann nun so variiert werden, daß $T(U_2; U_1)$ den Maximalwert erreicht. Dieser Wert stellt den größtmöglichen im Mittel mit dem Symbol über einen gegebenen Kanal übertragbaren Informationsgehalt dar und wird als Kanalkapazität C bezeichnet.

Es gilt

$$C = \underset{P(u_2)}{Max}\ T(U_2; U_1). \tag{4.27}$$

Die Kanalkapazität C nach Gl. (4.27) hat die Einheit bit je Symbol. Multipliziert man C mit der Anzahl der je Sekunde übertragenen Symbole, so erhält man die Kanalkapazität C' in der Einheit bit je Sekunde.

Kanalkapazität des verlustfreien Kanals

Ein Kanal wird als verlustfrei bezeichnet, wenn der Informationsgehalt $H(U_2)$ der Quelle verlustfrei übertragen wird, d.h. $H(U_2|U_1) = 0$ gilt und Gl. (4.26) in der Form

$$T(U_2; U_1) = H(U_2) \tag{4.28}$$

geschrieben werden kann. Für den verlustfreien Kanal ergibt sich die Kanalkapazität

$$C = \underset{P(u_2)}{Max}\, T(U_2; U_2) = \underset{P(u_2)}{Max}\, H(U_2) = \log_2 K\,. \tag{4.29}$$

Das Maximum wird erreicht, wenn alle Symbole u_2 gleich wahrscheinlich sind.

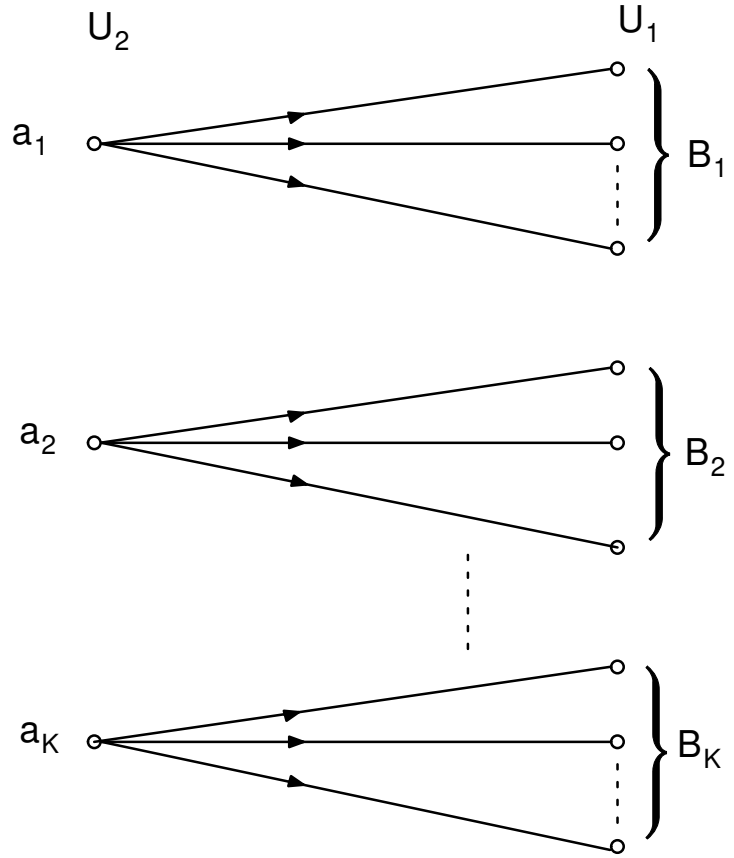


Bild 4.10 Schematische Darstellung eines verlustfreien Kanals

Der verlustfreie Kanal ist schematisch in Bild 4.10 dargestellt. Das Ensemble U_1 von J Symbolen $b_1, ..., b_J$ ist dabei in K disjunkte Untermengen B_k unterteilt, wobei jede Untermenge B_k eindeutig nur einem Symbol a_k des Quellenalphabets $a_1, ..., a_K$ zugeordnet ist. Die bedingten Wahrscheinlichkeiten $P(a_k|b_j)$ des verlustfreien Kanals sind aufgrund von $H(U_2|U_1) = 0$ durch

$$P(a_k|b_j) = \begin{cases} 1, & \text{wenn} \quad b_j \in B_k \\ 0, & \text{wenn} \quad b_j \notin B_k \end{cases} \tag{4.30}$$

gegeben.
Bei einem verlustfreien Kanal ist mit dem Empfang eines Symbols auch eindeutig entschieden, welches Symbol gesendet wurde.

Kanalkapazität des deterministischen Kanals

Ein Kanal wird als deterministisch bezeichnet, wenn $H(U_1|U_2) = 0$ für alle Wahrscheinlichkeitsverteilungen $P(u_2)$ gilt, so daß Gl. (4.26) in

$$T(U_2; U_1) = H(U_1) \tag{4.31}$$

übergeht. Für den deterministischen Kanal ergibt sich die Kanalkapazität

$$C = \underset{P(u_2)}{Max}\ H(U_1) = \log_2 J. \tag{4.32}$$

Bei der Berechnung von $H(U_1)$ ist dabei zu beachten, daß jedes $P(u_1)$ gemäß

$$P(u_1) = \sum_{u_2} P(u_2) \cdot P(u_1|u_2) \tag{4.33}$$

eine Funktion der Wahrscheinlichkeit $P(u_2)$ der Quellensymbole und der den Kanal charakterisierenden Übergangswahrscheinlichkeiten $P(u_1|u_2)$ ist. Das Maximum der Gl. (4.32) wird erreicht, wenn alle Symbole u_1 gleichwahrscheinlich sind.

Der deterministische Kanal ist schematisch in Bild 4.11 dargestellt.

Das Ensemble U_2 von K Symbolen $a_1,\ ...,\ a_K$ ist in diesem Fall in J disjunkte Untermengen A_j unterteilt, wobei jede Untermenge A_j eindeutig nur einem Symbol b_j des Alphabets $b_1,\ ...,\ b_J$ zugeordnet ist. Die bedingten Wahrscheinlichkeiten $P(b_j|a_k)$ des deterministischen Kanals sind durch

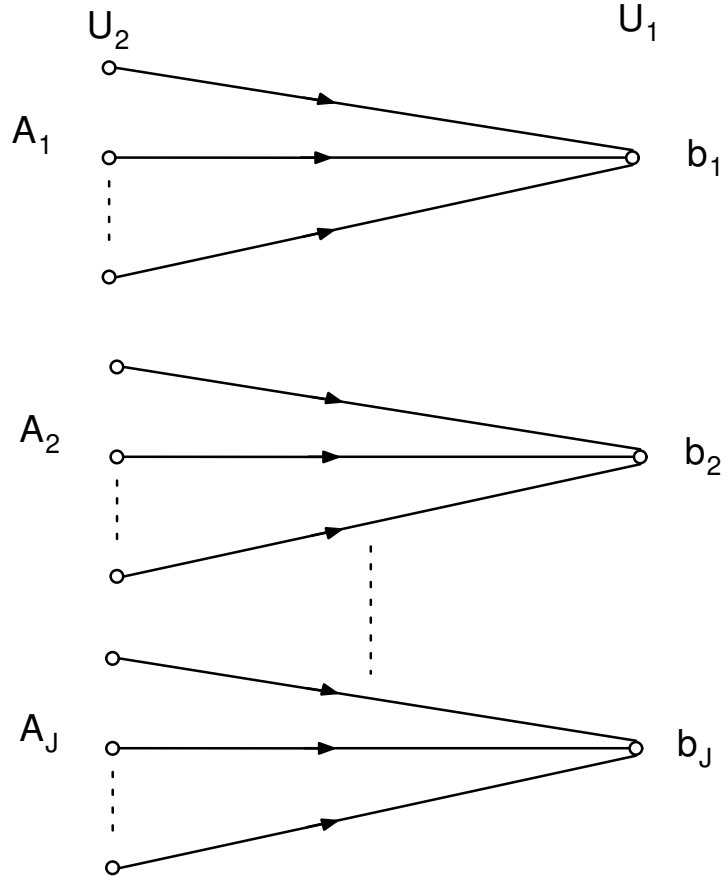


Bild 4.11 Schematische Darstellung eines deterministischen Kanals

$$P(b_j|a_k) = \begin{cases} 1, & \text{wenn} \quad a_k \in A_j \\ 0, & \text{wenn} \quad a_k \notin A_j \end{cases} \tag{4.34}$$

festgelegt. Bei einem deterministischen Kanal ist mit dem gesendeten Symbol auch eindeutig entschieden, welches Symbol empfangen wird.

Kanalkapazität des ungestörten Kanals

Ein Kanal wird als ungestört bezeichnet, wenn er sowohl verlustfrei als auch deterministisch ist und aus Gl. (4.26)

$$T(U_2; U_1) = H(U_2) = H(U_1) \tag{4.35}$$

folgt. Die Untermengen A und B bestehen in diesem Fall jeweils nur aus einem Symbol, die Ensembles am Kanaleingang und Kanalausgang sind gleich groß, d.h. $K = J$, und die Kanalkapazität beträgt

$$C = \log_2 K = \log_2 J. \tag{4.36}$$

Der ungestörte Kanal ist schematisch in Bild 4.12 dargestellt.

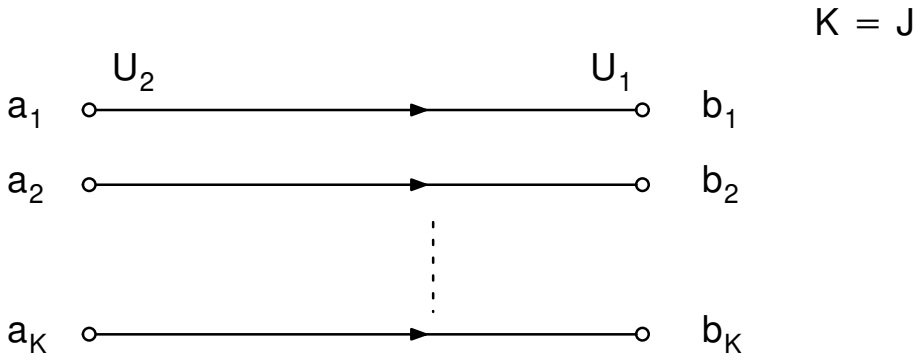


Bild 4.12 Schematische Darstellung eines ungestörten Kanals

Kanalkapazität des binären, symmetrischen Kanals (BSK)

Bild 4.13 zeigt das Flußdiagramm eines binären, symmetrischen Kanals mit der Fehlerwahrscheinlichkeit $P_e = p$.

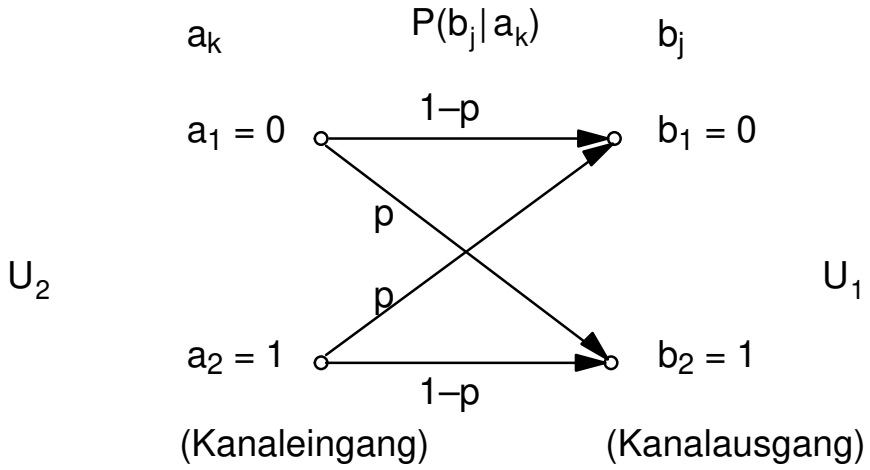


Bild 4.13 Flußdiagramm eines binären, symmetrischen Kanals

Die Kanalkapazität des BSK ergibt sich gemäß Gl. (4.10) zu

$$\begin{aligned} C_{BSK} &= \max_{P(u_2)} T(U_2; U_1) \\ &= \max_{P(u_2)} [\, H(U_1) - H(U_1|U_2) \,]. \end{aligned} \tag{4.37}$$

Bei einem symmetrischen Kanal gehen von jedem Symbol a_k Teil–Flußdiagramme gleicher Struktur aus. Aufgrund dieser Eigenschaft resultiert eine Irrelevanz

$H(U_1|U_2)$, die unabhängig von der Wahrscheinlichkeitsverteilung $P(u_2)$ der Eingangssymbole ist. Es gilt

$$\begin{aligned} H(U_1|U_2) &= -\sum_{k=1}^{K} P(a_k) \cdot \left[\sum_{j=1}^{J} P(b_j|a_k) \cdot \log_2 P(b_j|a_k) \right] \\ &= -\sum_{j=1}^{J} P(b_j|a_k) \cdot \log_2 P(b_j|a_k) \ , \end{aligned} \tag{4.38}$$

da die rechteckige Klammer für jedes a_k den gleichen Wert liefert. Somit folgt für die Kanalkapazität C_{BSK} des binären, symmetrischen Kanals

$$C_{BSK} = \underset{P(u_2)}{Max}\, H(U_1) - H(U_1|U_2) \tag{4.39}$$

$$C_{BSK} = 1 + (1-p) \cdot \log_2(1-p) + p \cdot \log_2 p \tag{4.40}$$

in bit je Symbol. Um eine Transinformation der Größe C_{BSK} zu erreichen, muß die Wahrscheinlichkeitsverteilung $P(u_2)$ die Bedingung $P(a_1) = P(a_2) = \frac{1}{2}$ erfüllen.
Im Fall $p = 0$ ist der Kanal ungestört, es wird $C_{BSK} = 1$ bit je Symbol; da $H(U_1) = 1$ bit je Symbol beträgt, muß die Irrelevanz $H(U_1|U_2) = 0$ sein. Im Fall $p = \frac{1}{2}$ ist der Kanal so stark gestört, daß $C_{BSK} = 0$ wird, folglich muß $H(U_1|U_2) = 1$ bit je Symbol sein. In diesem Fall wird vom Kanal kein Informationsgehalt übertragen. Die Ausgangssymbole u_1 sind statistisch unabhängig von den Eingangssymbolen u_2, da $H(U_1|U_2) = H(U_1)$ ist.
In Bild 4.14 ist die Beziehung (4.40) graphisch dargestellt.

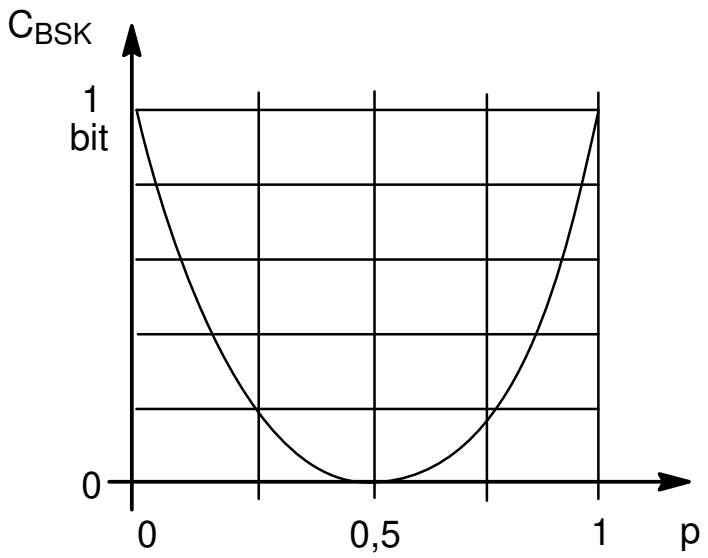


Bild 4.14 Kanalkapazität C_{BSK} in bit je Symbol eines binären, symmetrischen Kanals der Fehlerwahrscheinlichkeit p

Für den Fall, daß über einen binären, symmetrischen Kanal beispielsweise 2400 Binärsymbole je Sekunde übertragen werden, ergibt sich die Kanalkapazität C'_{BSK} gemäß

$$C'_{BSK} = 2400 \cdot [1 + (1 - p) \cdot \log_2(1 - p) + p \cdot \log_2 p]$$

in bit je Sekunde.

In der nachfolgenden Tabelle 4.1 ist die Kanalkapazität C'_{BSK} in der Einheit bit je Sekunde für verschiedene Bitfehlerwahrscheinlichkeiten p angegeben.

p	C'_{BSK} bit/s
10^{-4}	2396,5
10^{-3}	2372,6
10^{-2}	2206,1
10^{-1}	1274,4
5×10^{-1}	0

Tab. 4.1: Kanalkapazität C'_{BSK} bei einer Übertragungsrate von 2400 bit/s in Abhängigkeit von der Bitfehlerwahrscheinlichkeit p

In Kapitel 5 wird das Theorem der Kanalcodierung behandelt. Es besagt, wenn H' die Entropie einer diskreten, gedächtnislosen Quelle und C' die Kanalkpazität eines

diskreten Kanals je Sekunde ist, daß im Fall $H' \leq C'$ immer eine Kanalcodierung existiert, die es ermöglicht, die Quellensymbole mit beliebig kleiner Fehlerwahrscheinlichkeit zu übertragen.

Die mit dem Theorem der Kanalcodierung gegebene Grenze kann jedoch nur mit einem großen schaltungstechnischen Aufwand annähernd erreicht werden. Beim Entwurf eines realisierbaren Code muß immer ein Kompromiß zwischen verbleibender Fehlerwahrscheinlichkeit und ungenutzter Kanalkapazität eingegangen werden.

4.6 Kanalkapazität kontinuierlicher, gedächtnisloser Kanäle

Es wird zunächst die Kanalkapazität C des zeitdiskreten, wertkontinuierlichen Kanals in Analogie zum diskreten Kanal berechnet. Anschließend werden diese Ergebnisse zur Berechnung der Kanalkapazität C' bandbegrenzter, kontinuierlicher Kanäle verwendet. Ist X das wertkontinuierliche Ensemble am Kanaleingang und Y das wertkontinuierliche Ensemble am Kanalausgang, so folgt aus Gl. (4.27) für die Kanalkapazität eines zeitdiskreten, wertkontinuierlichen Kanals

$$\begin{aligned} C &= \underset{p(x)}{Max}\, T(X;Y) \\ C &= \underset{p(x)}{Max}\, [H(Y) - H(Y|X)]. \end{aligned} \tag{4.41}$$

Die Symbole x, y haben einen unendlichen Wertevorrat $-\infty \leq x,\ y \leq +\infty$ und können beispielsweise als Abtastwerte eines kontinuierlichen Signals interpretiert werden.

Kanal mit additivem, weißen Gaußschen Rauschen (AWGN)

Es sei ferner angenommen, daß die Störungen auf dem Kanal durch ein additives, weißes Gaußsches Rauschen der Augenblicksamplitude z, mit einer Wahrscheinlichkeitsdichte $p_G(z)$ entsprechend Gl. (4.17), dem Mittelwert $m_z = 0$ und der Varianz σ_z^2 erzeugt werden.

Für eine additive Störung gilt zu jedem Zeitpunkt

$$y = x + z \tag{4.42}$$

und

$$p(y \mid x) = p(y - x) = p(z). \tag{4.43}$$

Da z hier eine Gaußsche Wahrscheinlichkeitsdichte aufweist, folgt für $H(Y \mid X)$ mit (4.19) und (4.43)

$$H(Y \mid X) = \frac{1}{2}\log_2 \left(2\pi\, e\, \sigma_z^2 \right). \tag{4.44}$$

Die Kanalkapazität beträgt somit

$$C = \underset{p(x)}{Max}\, H(Y) - \frac{1}{2}\log_2 \left(2\pi\, e\, \sigma_z^2 \right). \tag{4.45}$$

Es sei weiterhin angenommen, daß die Signalleistung am Eingang des Kanals σ_x^2 begrenzt ist. Damit wird die Leistung am Ausgang des Kanals $\sigma_y^2 = \sigma_x^2 + \sigma_z^2$. Bei gegebener Varianz σ_y^2 hat die Gaußsche Wahrscheinlichkeitsdichte die größte Entropie. Folglich wird $H(Y)$ in Gl. (4.45) seinen Maximalwert erreichen, wenn nicht nur die Störung z, sondern auch die Ausgangssymbole y eine Folge statistisch unabhängiger Symbole mit Gaußscher Wahrscheinlichkeitsdichte bilden, d.h.

$$\underset{p(x)}{Max}\, H(Y) = \frac{1}{2}\, \log_2\left[2\pi\, e\, (\sigma_x^2 + \sigma_z^2)\right]. \tag{4.46}$$

Aufgrund von Gleichung (4.46) muß dann auch x eine Gaußsche Wahrscheinlichkeitsdichte aufweisen. Die Kanalkapazität eines zeitdiskreten, wertkontinuierlichen Kanals mit additivem, weißen Gaußschen Rauschen ergibt sich mit (4.45) und (4.46) zu

$$C = \frac{1}{2}\log_2\left[1 + \frac{\sigma_x^2}{\sigma_z^2}\right] \tag{4.47}$$

in bit je Abtastwert, wobei σ_x^2/σ_z^2 das Signal–Rauschleistungs–Verhältnis ist.

Bandbegrenzter Kanal mit AWGN

Da im kontinuierlichen Kanal mit den Bandgrenzen $\pm\ W$ je Zeiteinheit 2W Abtastwerte übertragen werden können, folgt für die Kanalkapazität C' des zeitkontinuierlichen, wertkontinuierlichen, bandbegrenzten Kanals

$$C' = 2W \cdot C = W \cdot \log_2\left[1 + \sigma_x^2/\sigma_z^2\right] \tag{4.48}$$

in bit je Sekunde.

Für die Gl. (4.48) kann im Fall $\sigma_x^2/\sigma_z^2 \gg 1$ eine Näherung angegeben werden, die eine grobe Abschätzung der Kanalkapazität bei gegebenen Kanaleigenschaften erlaubt.

$$[C']_{bit/s} \approx \frac{1}{3}[W]_{Hz} \cdot \left[\ \frac{P}{N}\ \right]_{dB} \tag{4.49}$$

wobei

$$\frac{P}{N} = 10 \log_{10} \frac{\sigma_x^2}{\sigma_z^2} \tag{4.50}$$

das Signal–Rauschleistungs–Verhältnis in der Einheit Dezibel ist.

Unter Verwendung dieser Näherung erhält man für die Kanalkapazität eines kontinuierlichen Fernsprechkanals bzw. Fernsehkanals folgende Werte:

Kanaltyp	W	P/N	C’
Fernsprechkanal	3,1 kHz	50 dB	50 kbit/s
Fernsehkanal	5,0 MHz	45dB	75 Mbit/s

Gl. (4.48) zeigt, daß zur Erzielung einer vorgegebenen Kanalkapazität C' die Bandbreite W gegen das Signal–Rauschleistungs–Verhältnis σ_x^2/σ_z^2 ausgetauscht werden kann. Zur Erläuterung dieses Zusammenhangs dient Bild 4.15. Das Produkt der Bandbreite $2W$ und des logarithmierten Signal–Rauschleistungs–Verhältnisses $1/2 \cdot \log_2(1 + \sigma_x^2/\sigma_z^2)$ entspricht der Kanalkapazität C' in bit je Sekunde gemäß Gl. (4.48). Die sich in der graphischen Darstellung nach Bild 4.15 für C' ergebende Fläche ist in ihrer Form kennzeichnend für die Übertragungseigenschaften des Kanals. Wird die Zeitachse als dritte Dimension in die Betrachtung mit einbezogen, so beschreibt der dargestellte Quader mit seinem Volumen die Informationsmenge in bit, die in einem Zeitintervall T über einen Kanal mit den gegebenen Eigenschaften σ_x^2/σ_z^2 und W übertragen werden kann.

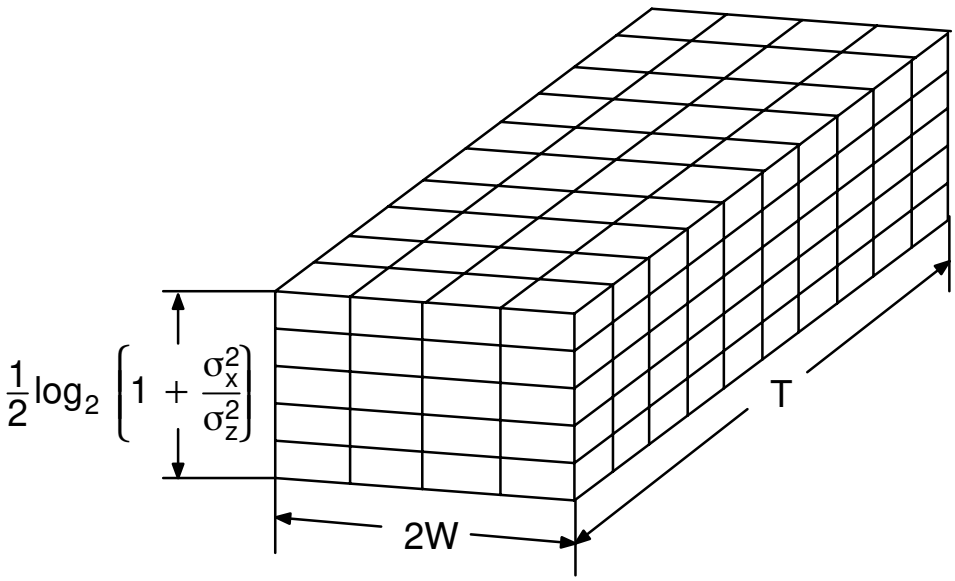


Bild 4.15 Darstellung des Informationsvolumens als Informationsquader

$$V = C' \cdot T = 2W \cdot 1/2 \cdot \log_2\left(1 + \sigma_x^2/\sigma_z^2\right) \cdot T$$

4.7 Anwendungsbeispiel

Als Beispiel soll die bipolare Basisbandübertragung von Binärsignalen über einen bandbegrenzten Kanal mit additivem weißen Gaußschen Rauschen (AWGN –Additive White Gaussian Noise) betrachtet werden. Zur Beurteilung dieses Modulationsverfahrens wird die Transinformation als Funktion des Signal–Rauschabstandes berechnet und anschließend mit der theoretischen Grenze, der sog. Shannon–Grenze, verglichen.

In einem bandbegrenzten Kanal der Bandbreite W können 2W Abtastwerte je Sekunde übertragen werden. Bei der bipolaren Basisbandübertragung stellt jeder Abtastwert ein Binärzeichen dar, siehe Bild 4.16. Wird der binäre, symmetrische Kanal als Kanalmodell zugrunde gelegt und angenommen, daß die Binärzeichen 0 und 1 mit gleicher Wahrscheinlichkeit auftreten, so beträgt die Transformation in der Einheit bit je Hertz Bandbreite und Sekunde gemäß Gl. (4.40)

$$\frac{T'}{W} = 2C_{BSK} \quad = 2\left\{1 + p \cdot \log_2(p) + (1-p) \cdot \log_2(1-p)\right\} . \tag{4.51}$$

In Gleichung (4.51) geht nur die Fehlerwahrscheinlichkeit $P_e = p$ des binären, symmetrischen Kanals ein. Diese Fehlerwahrscheinlichkeit kann mit Hilfe des Signal–Rauschleistungs–Verhältnisses P/N des Übertragungskanals bestimmt werden.

Im Fall additiven Rauschens $z(t)$ ergibt sich das Empfangssignal $y(t)$ aus dem Sendesignal $x(t)$ gemäß

$$y(t) = x(t) + z(t) . \tag{4.52}$$

Wird das Empfangssignal $y(t)$ zur Demodulation einer Entscheidungsschwelle gemäß Bild 4.17 zugeführt, so wird das Binärzeichen 1 verfälscht, wenn zum Zeitpunkt der Abtastung des übertragenen bandbegrenzten Signals eine Rauschspannung kleiner als $-\sigma_x$ auftritt und das Binärzeichen 0 verfälscht, wenn eine Rauschspannung größer als $+\sigma_x$ auftritt.

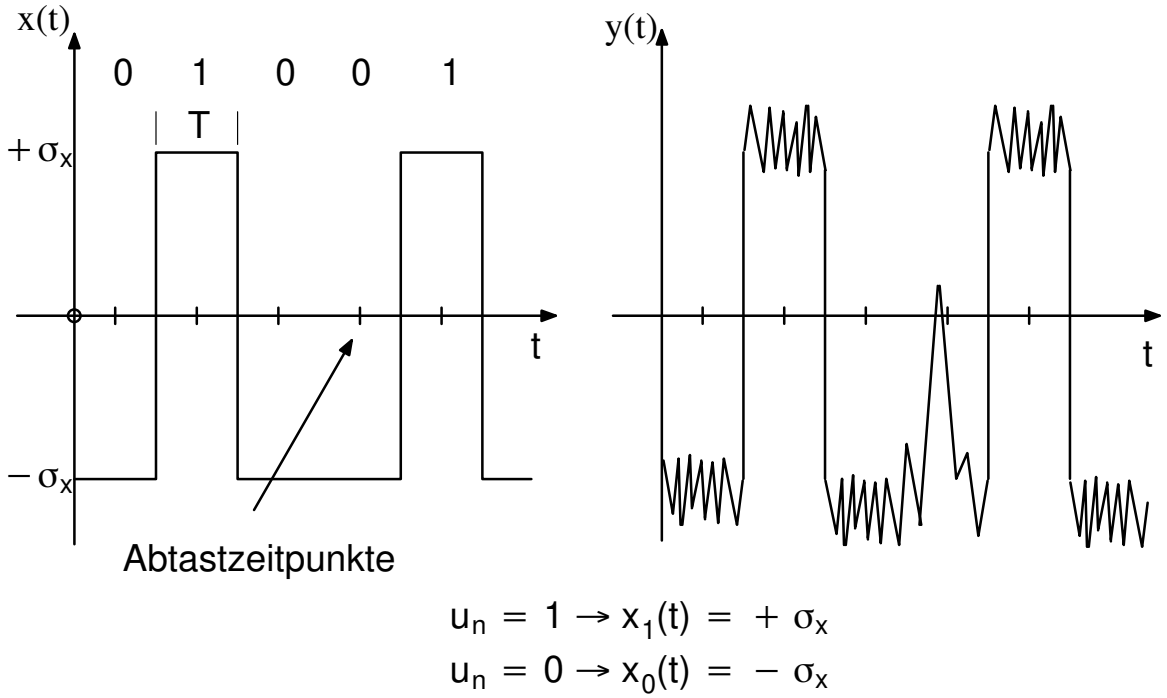


Bild 4.16 Sendesignal *x(t)* und Empfangssignal *y(t)* einer bipolaren Basisbandübertragung

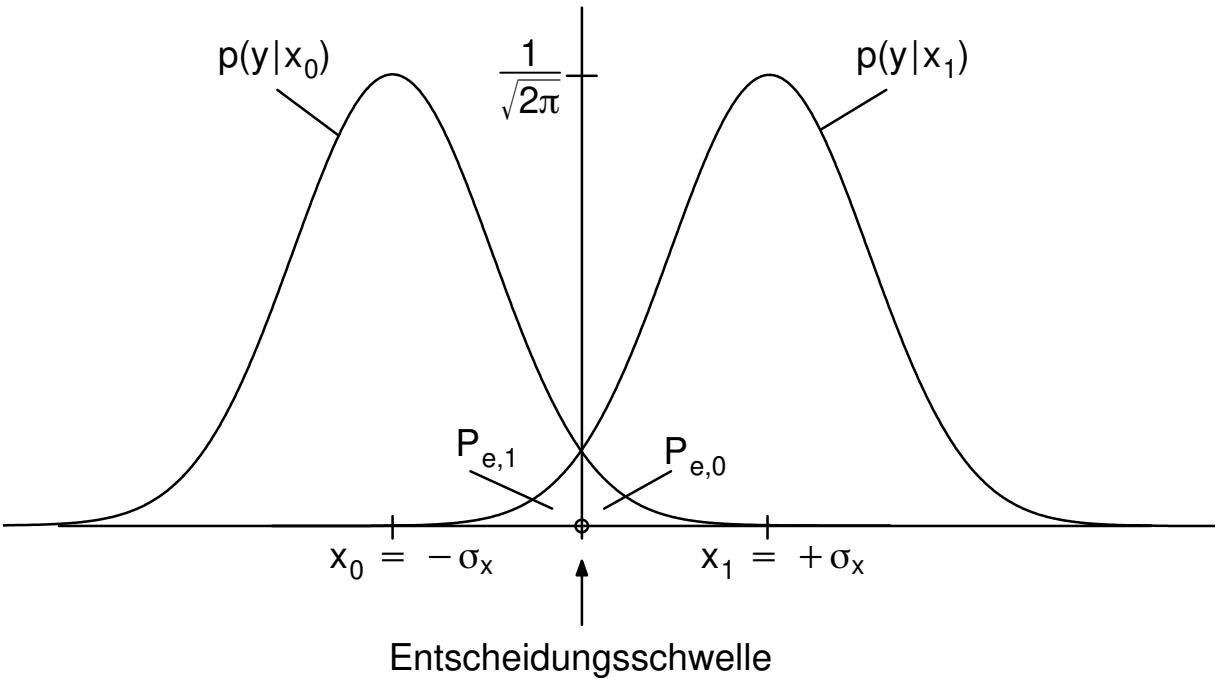


Bild 4.17 Wahrscheinlichkeitsdichte des Empfangssignals bei bipolarer Übertragung

Die Fehlerwahrscheinlichkeit P_e des Kanals beträgt

$$P_e = p = P(0) \cdot P_{e,0} + P(1) \cdot P_{e,1}$$

$$P_e = p = P(0) \cdot P(z > +\sigma_x) + P(1) \cdot P(z < -\sigma_x) \ . \tag{4.53}$$

Unter der Annahme, daß die Übertragungsfehler durch ein Rauschsignal mit Gaußscher Wahrscheinlichkeitsdichte, dem Mittelwert $m_z = 0$ und der Varianz ${\sigma_z}^2$ erzeugt werden und die Binärzeichen 0 und 1 mit gleicher Wahrscheinlichkeit $P(0) = P(1) = 1/2$ auftreten, gilt

$$P_e = \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma_z} \int_{\sigma_x}^{+\infty} e^{-\frac{z^2}{2{\sigma_z}^2}} dz + \frac{1}{2} \cdot \frac{1}{\sqrt{2\pi} \cdot \sigma_z} \int_{-\infty}^{-\sigma_x} e^{-\frac{z^2}{2{\sigma_z}^2}} dz \tag{4.54}$$

$$P_e = \frac{1}{2}\left[1 - 2\frac{1}{\sqrt{2\pi} \cdot \sigma_z} \int_{0}^{\sigma_x} e^{-\frac{z^2}{2\sigma_z^2}} dz\right]$$

Mit der Substitution

$$\xi^2 = \frac{z^2}{2{\sigma_z}^2} \tag{4.55}$$

und dem Signal–Rauschleistungs–Verhältnis v^2

$$v^2 = \frac{{\sigma_x}^2}{{\sigma_z}^2} \tag{4.56}$$

folgt

$$\begin{aligned} P_e &= \frac{1}{2}\left[1 - 2\frac{1}{\sqrt{\pi}}\int_0^{v/\sqrt{2}} e^{-\xi^2}\, d\xi\right] \\ P_e &= \frac{1}{2}\left[1 - erf\left[\frac{v}{\sqrt{2}}\right]\right] \end{aligned} \tag{4.57}$$

wobei $erf(x)$ das sog. Gaußsche Fehlerintegral oder auch sog. Fehlerfunktion (error-function) ist. Da das Integral nicht geschlossen lösbar ist, wird die Auswertung in Tabellenform angegeben.
Neben der Fehlerfunktion *erf(x)* definiert durch

$$erf(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-\xi^2} d\xi \tag{4.58}$$

gibt es eine sog. komplementäre Funktion $erfc(x)$ entsprechend

$$erfc(x) = 1 - erf(x). \tag{4.59}$$

Beide Funktionen sind in 4.2 angegeben.

x	erf(x)	x	erfc(x)
0	0	0	1,00
0,2	0,22270	0,2	$7{,}77 \cdot 10^{-1}$
0,4	0,42839	0,4	$5{,}71 \cdot 10^{-1}$
0,6	0,60386	0,6	$3{,}96 \cdot 10^{-1}$
0,8	0,74210	0,8	$2{,}57 \cdot 10^{-1}$
1,0	0,84270	1,0	$1{,}57 \cdot 10^{-1}$
1,2	0,91031	1,2	$8{,}97 \cdot 10^{-2}$
1,4	0,95229	1,4	$4{,}77 \cdot 10^{-2}$
1,6	0,97635	1,6	$2{,}37 \cdot 10^{-2}$
1,8	0,98909	1,8	$1{,}09 \cdot 10^{-2}$
2,0	0,995322	2,0	$4{,}68 \cdot 10^{-3}$
2,2	0,998137	2,2	$1{,}86 \cdot 10^{-3}$
2,4	0,999311	2,4	$6{,}89 \cdot 10^{-4}$
2,6	0,9997640	2,6	$2{,}36 \cdot 10^{-4}$
2,8	0,9999250	2,8	$7{,}50 \cdot 10^{-5}$
3,0	0,9999779	3,0	$2{,}21 \cdot 10^{-5}$
3,2	0,9999939	3,2	$6{,}03 \cdot 10^{-6}$
3,4	0,9999984	3,4	$1{,}52 \cdot 10^{-6}$

Tab. 4.2: Fehlerfunktion *erf(x)* und *erfc(x)*

Die Fehlerwahrscheinlichkeit gemäß Gl. (4.57) ist eine Funktion des Signal–Rauschleistungs–Verhältnisses

$$P/N = 10 \log_{10}(v^2) \tag{4.60}$$

in der Einheit dB. Dieser Zusammenhang ist in Bild 4.18 dargestellt.

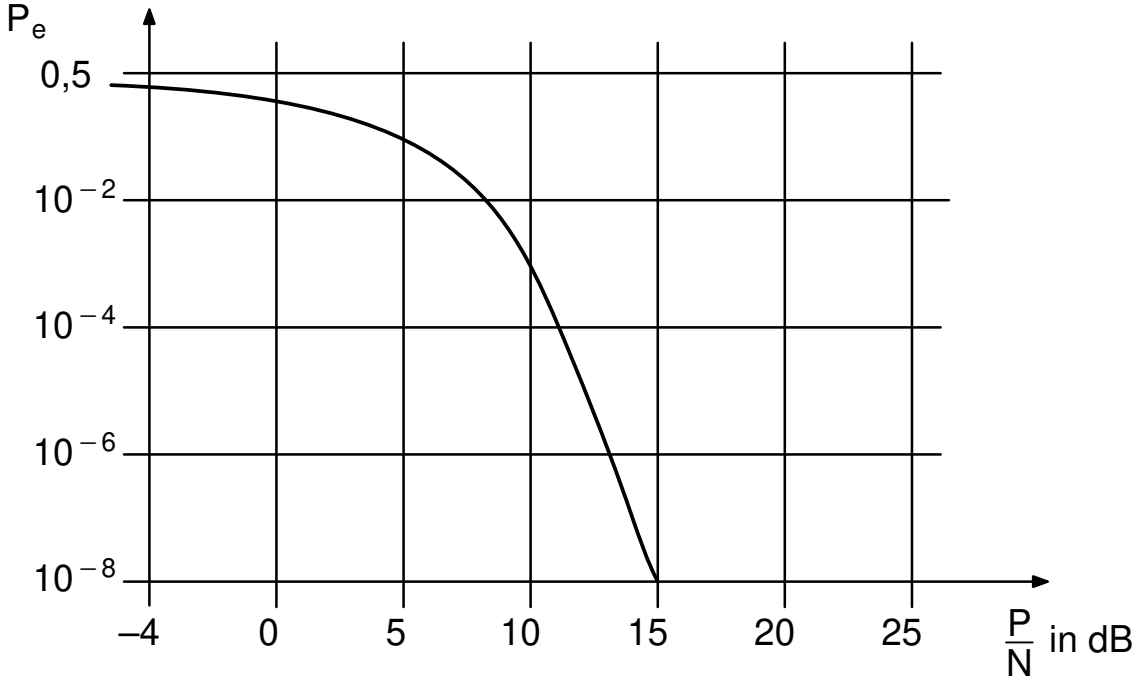


Bild 4.18 Fehlerwahrscheinlichkeit $P_e = \frac{1}{2}\ erfc\left[\frac{\nu}{\sqrt{2}}\right]$ der bipolaren Übertragung

Die Auswertung der Gleichung (4.57) zusammen mit Gl. (4.60) ergibt die in 4.3 zusammengestellten Kennwerte der bipolaren Basisbandübertragung. Aus Bild 4.19 ist zu erkennen, daß die Basisbandübertragung den Kanal bei großem P/N schlecht ausnutzt, da die theoretische Grenze nach Gl. (4.48) eine weit größere Transinformation zuläßt. Dies kann mittels einer mehrstufigen Basisbandübertragung verbessert werden.

P_e =p	erf ($v/2^{1/2}$)	v^2	P/N	T'/W	Fehlerrate für R=64 kbit/s
10^{-1}	0,8	1,66	2,2 dB	1,06	6400 /s
10^{-2}	0,98	5,44	7,3 dB	1,84	640 /s
10^{-3}	0,998	9,60	9,8 dB	2,0	64 /s
10^{-4}	0,9998	13,90	11,4 dB	2,0	6,4/s
10^{-6}	0,999998	22,80	13,6 dB	2,0	3,8/min
10^{-8}	0,99999998	31,60	15,0 dB	2,0	2,3/h

Tab. 4.3: Auswertung der Gleichungen (4.57) und (4.60)

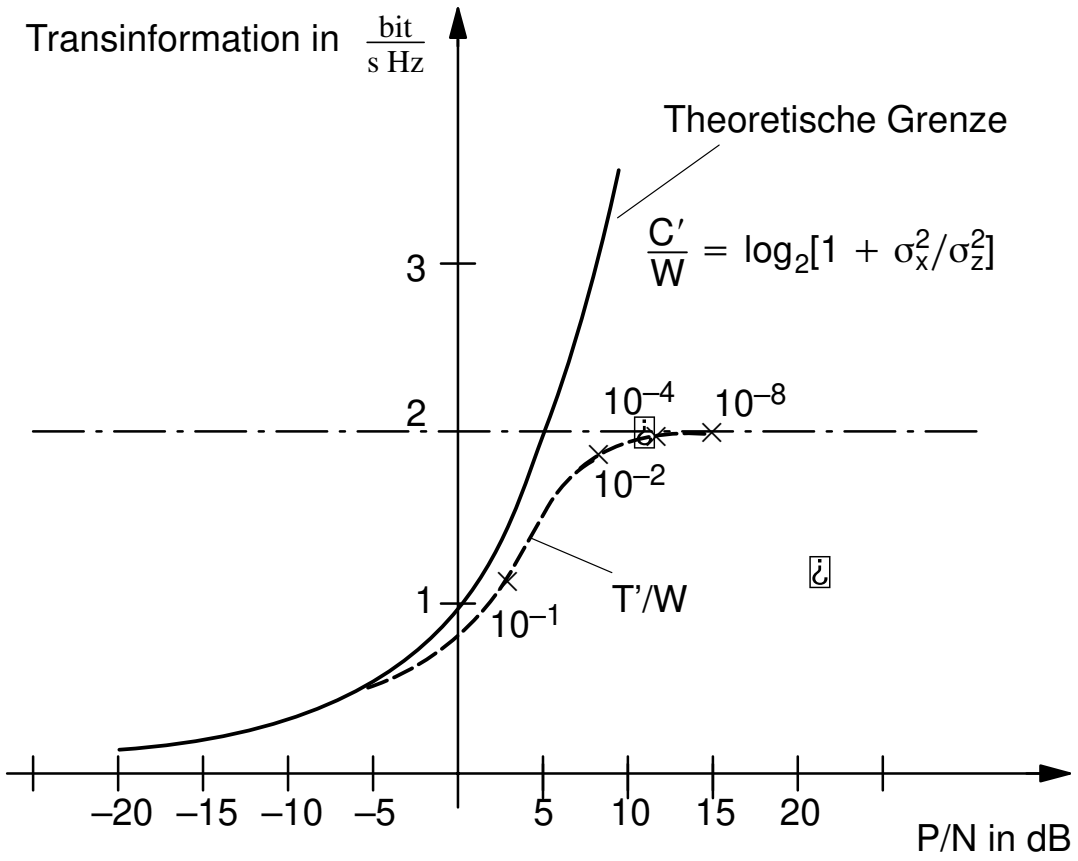


Bild 4.19 Transinformationsfluß je Hertz Bandbreite einer bipolaren Basisbandübertragung im Vergleich zur theoretischen Grenze für einen binären, symmetrischen Kanal mit additivem, weißen Gaußschen Rauschen

5. Kanalcodierung

5.1 Allgemeine Grundbegriffe

Die Kanalcodierung dient dem Ziel, die über einen Kanal zu übertragenden Informationen gegen den Einfluß von Übertragungsfehlern zu schützen. Dazu können sowohl fehlererkennende als auch fehlerkorrigierende Codes verwendet werden. Zur Fehlererkennung und zur Fehlerkorrektur muß ein Code Redundanz beinhalten. Bei einem binären Code mit N–stelligen Codewörtern gibt es insgesamt 2^N mögliche Codewörter. Gehören alle 2^N Codewörter zum Code, so wird durch Verfälschung eines Codeelementes stets ein anderes zum Code gehörendes Codewort entstehen. In diesem Fall kann ein Nachrichtenempfänger einen bei der Übertragung aufgetretenen Fehler nicht feststellen. Nutzt ein Code hingegen von den 2^N Kombinationen nur 2^K, wobei $K < N$, als Codewörter aus, so können die $2^N - 2^K$ redundanten Kombinationen zur Fehlererkennung und Fehlerkorrektur verwendet werden. Die relative, auf ein übertragenes, binäres Codesymbol bezogene Redundanz r beträgt

$$r = \frac{N - K}{N}. \tag{5.1}$$

Der mit einem Codesymbol übertragbare Informationsgehalt wird auch als Coderate R bezeichnet und ergibt sich aus

$$R = 1 - r = \frac{K}{N}. \tag{5.2}$$

Ein für die Behandlung fehlererkennender und fehlerkorrigierender Codes weiterer wichtiger Begriff ist die Hamming–Distanz D. Ihre Definition lautet:

Hamming–Distanz D: Die Anzahl der Codeelemente entsprechender Stellen, in denen sich gleich lange Codewörter eines Binärcodes unterscheiden.

Die Hamming–Distanz zweier Codewörter $\vec{x_i}$ und $\vec{x_j}$ läßt sich berechnen gemäß

$$D(\vec{x_i},\ \vec{x_j}) = w(\vec{x_i} \oplus \vec{x_j}) = w(\vec{x_R}), \tag{5.3}$$

wobei $w(\vec{x_R})$ das Gewicht (weight) von $\vec{x_R}$ bzw. die Anzahl der mit 1 belegten Stellen von $\vec{x_R}$ ist. Die Modulo–2–Summe $\vec{x_R} = \vec{x_i} \oplus \vec{x_j}$ enthält dabei genau an den Stellen eine 1, in denen sich $\vec{x_i}$ und $\vec{x_j}$ unterscheiden.

Beispiel: Die Hamming–Distanz D der beiden Codewörter $\vec{x_i} = 0111$ und $\vec{x_j} = 1011$ beträgt mit $\vec{x_R} = \vec{x_i} \oplus \vec{x_j} = 1100$
$D(\vec{x_i},\ \vec{x_j}) = w(\vec{x_R}) = 2.$

Die Minimaldistanz eines Code oder auch sog. Codedistanz d ist folgendermaßen definiert:

Codedistanz d: Die kleinste aller Hamming–Distanzen der Codewörter eines Codes.

$$d = \min_{\text{alle } i;\ \text{alle } j \neq i} D(\vec{x_i},\ \vec{x_j}) \tag{5.4}$$

Die fehlersichernden Eigenschaften fehlererkennender oder fehlerkorrigierender Codes sollen über die geometrische Darstellung der Codewörter in Bild 5.1 veranschaulicht werden.

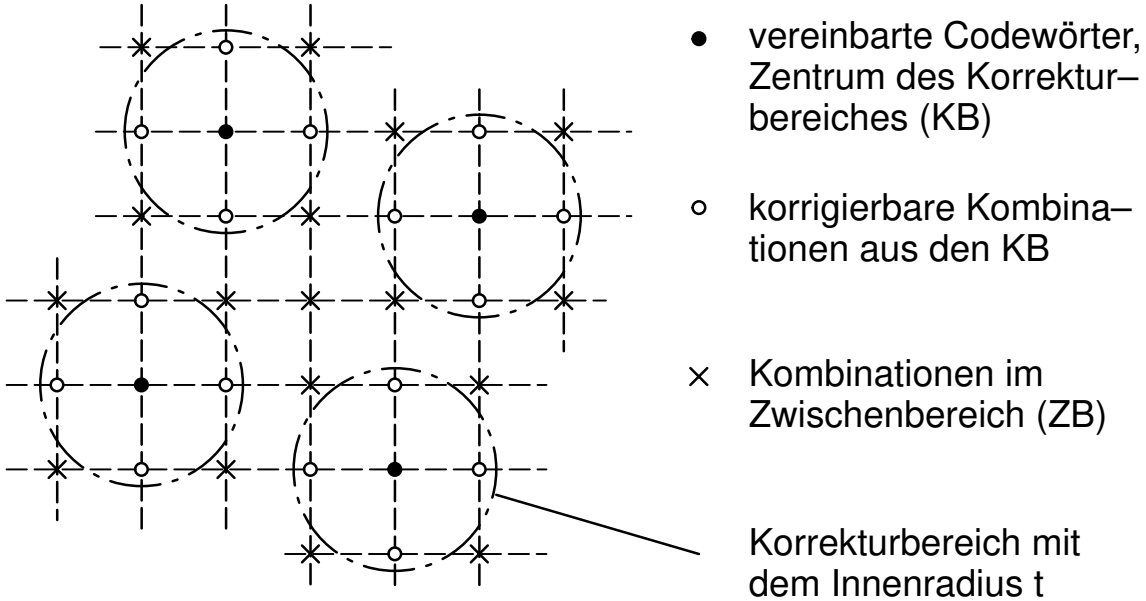


Bild 5.1 Geometrische Darstellung eines fehlerkorrigierenden Codes mit d=4 und t=1 (Ausschnitt)

Im Fall N–stelliger Codewörter enthält das Modell 2^N Raumpunkte. Jedem Raumpunkt ist eine N–stellige Kombination binärer Codesymbole zugeordnet und zwar derart, daß unmittelbar benachbarte Kombinationen jeweils die Hamming–Distanz $D = 1$ aufweisen. Jedes der 2^K zum Code gehörenden Codewörter ist von einem Korrekturbereich (KB) umgeben, in dem Kombinationen liegen, die nicht zum Code gehören, aber durch fehlerhafte Übertragung auftreten können.

Die Fehlerkorrektur erfolgt durch Ermittlung des Korrekturbereiches, zu dem eine empfangene Kombination gehört, und Ersetzen durch das Codewort im Zentrum des KB. Zwischen den Korrekturbereichen können auch Kombinationen liegen, die keinem KB zugeordnet sind. Diese bilden den sog. Zwischenbereich (ZB) des Code. Kombinationen des ZB werden als fehlerhafte Codewörter erkannt, können aber nicht korrigiert werden. Codes ohne ZB heißen dicht gepackt. Um t Fehler korrigieren zu können, muß der Korrekturbereich eines Codewortes alle Kombinationen $\vec{y}$ enthalten, die eine Hamming–Distanz $D(\vec{y},\ \vec{x_R}) \leq t$ zum Codewort $\vec{x_R}$ im Zentrum des KB aufweisen. Mit anderen Worten, der KB eines jeden Codewortes muß den Innenradius t haben. Die Korrekturbereiche der Codewörter dürfen sich nicht überschneiden. Die Codedistanz d eines t–fehlerkorrigierenden Code muß daher die Bedingung

$$d \geq 2t + 1 \tag{5.5}$$

erfüllen. Zur Erkennung von Fehlern genügt lediglich ein Zwischenbereich. Um e Fehler erkennen zu können, muß die Codedistanz

$$d \geq e + 1 \tag{5.6}$$

sein. Mit d wächst die Redundanz des Code.

5.2 Prinzipien der Blockcodierung

Die bekannten Codes können entsprechend ihrer Struktur in Blockcodes und Faltungscodes unterteilt werden. Im Fall der Blockcodierung wird eine Folge binärer Quellensymbole in Sequenzen von jeweils K Quellensymbolen $\vec{u} = u_1,\ ...,\ u_K$ unterteilt. Somit gibt es insgesamt $M = 2^K$ verschiedene Sequenzen. Zur Codierung dieser Sequenzen benötigt der Kanalcoder einen Vorrat von M Codewörtern. Jedes Codewort $\vec{x}$ besteht aus N binären Codeelementen $\vec{x} = x_1,\ ...,\ x_N$. Man bezeichnet N auch als Blocklänge des Blockcode. Beim Kanaldecoder wird die empfangene Folge binärer Symbole wieder in Blöcke $\vec{y} = y_1,\ ...,\ y_N$, entsprechend der Codewortlänge N, unterteilt. Vergleiche dazu Bild 5.2. Der Kanaldecoder decodiert $\vec{y}$ in eine Sequenz binärer Symbole $\vec{v} = v_1,\ ...,\ v_K$, die bei fehlerfreier Decodierung mit der gesendeten Sequenz $\vec{u} = u_1,\ ...,\ u_K$ übereinstimmt.

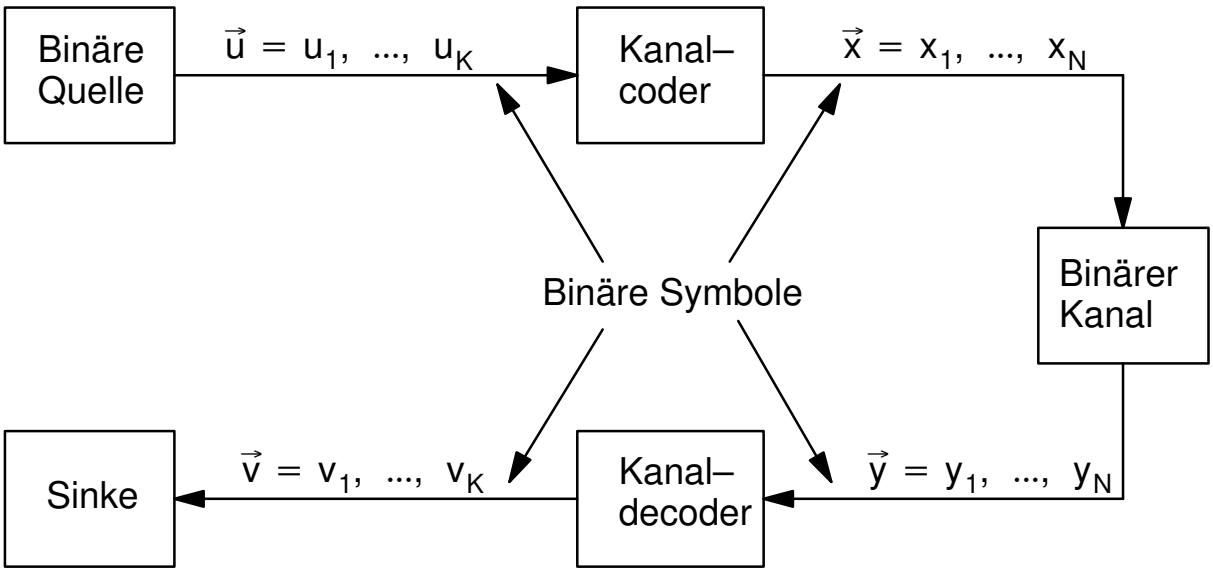


Bild 5.2 Übertragungssystem mit Blockcodierung der Blocklänge N

Als Beispiel eines Blockcode soll der Aufbau der Codewörter eines systematischen Parity–Check–Code betrachtet werden. Bei diesen Codes unterscheidet man zwischen den K Informationsstellen und $N - K$ Prüfstellen, siehe Bild 5.3.

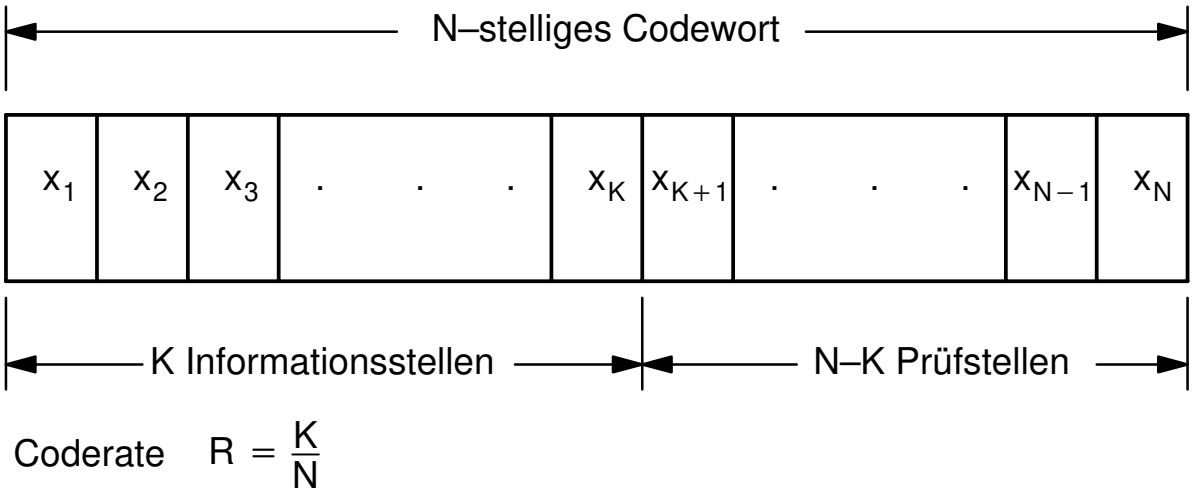


Bild 5.3 Codewort eines systematischen Parity–Check–Code

Im einfachsten Fall eines Parity–Check–Code mit nur einer Prüfstelle kann diese zur Erkennung eines Fehlers verwendet werden. Mit Hilfe der Prüfstelle wird die Parität des Codewortes überprüft. Die Parität ist folgendermaßen definiert:

Parität w_p: Bei einer vorgegebenen Anzahl von Binärelementen der Rest, der sich ergibt, wenn die Anzahl der mit 1 belegten Codeelemente durch zwei geteilt wird. Die Parität kann gerade (0) oder ungerade (1) sein.

Die Parität w_p einer empfangenen Codefolge $\vec{y}$ läßt sich sehr einfach durch Modulo–2–Addition aller Codeelemente berechnen.

$$w_p(\vec{y}) = y_1 \oplus y_2 \oplus ... \oplus y_{N-1} \oplus y_N \tag{5.7}$$

Ein Übertragungsfehler wird an der Veränderung der vereinbarten Parität erkannt. Zur Codierung wird die Prüfstelle x_N gemäß der Beziehung

$$x_N = x_1 \oplus x_2 \oplus ... \oplus x_{N-1} \oplus w_p \tag{5.8}$$

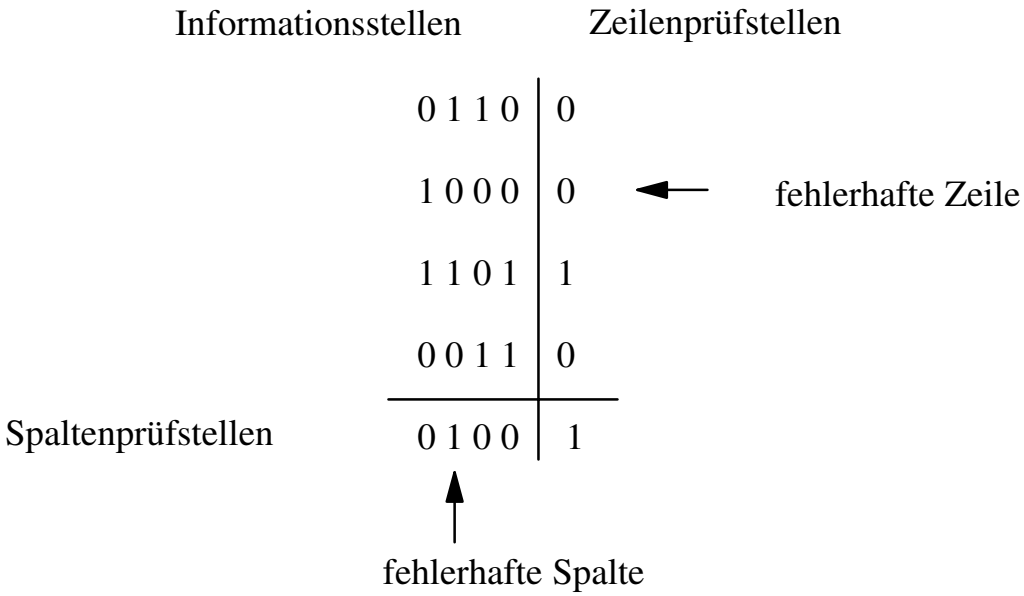
berechnet.

Beispiel: Codierung mit gerader Parität $w_p = 0$.
Gesendetes Codewort $\vec{x} = 01100$
Empfangenes Codewort $\vec{y} = 00100$
Parität $w_p(\vec{y}) = 1$, d.h.
der Übertragungsfehler wird erkannt.

Mit dieser speziellen Parity–Check–Codierung wird jede ungerade Anzahl von Übertragungsfehlern erkannt. Eine gerade Anzahl von Fehlern führt wieder auf die vereinbarte Parität; diese Fehlermuster werden nicht erkannt.

Für eine Fehlerkorrektur werden mehrere Prüfstellen je Codewort benötigt. Die einfachste Codierung dieser Art ist die zweidimensionale Anwendung von Paritätsstellen. Dazu werden die K Informationsstellen in Matrixform angeordnet und für jede Zeile und Spalte eine Prüfstelle ergänzt. Im Fall eines Übertragungsfehlers erkennt der Decoder eine Spalte und eine Zeile mit fehlerhafter Parität. Das fehlerhafte Binärsymbol im Schnittpunkt dieser Spalte und Zeile wird korrigiert.

Beispiel: Codierung und Decodierung von $K = 16$ Informationsstellen mit gerader Parität $w_p = 0$



Neben der Blockcodierung gibt es die sog. Faltungscodierung oder auch konvolutionelle Codierung. Bei diesen Codes werden keine Codewörter in Blockform gebildet. Die Prüfbits werden wie bei den Parity–Check–Codes gemäß Gl. (5.8) berechnet und fortlaufend in die Folge der Informationsbits eingefügt. Zur Berechnung der Prüfbits werden Informationsbits in ausgewählten Abständen verwendet.

5.3 Kanaldecodierung

Um die Fehlerwahrscheinlichkeit nach der Decodierung zu minimieren, wird ein empfangenes Codewort $\vec{y}$ dem Codewort $\vec{x_{m'}}$ mit der größten, bedingten Wahrscheinlichkeit $P(\vec{x_{m'}}|\vec{y})$ zugeordnet, d.h.

$$P(\vec{x_{m'}}|\vec{y}) \geq P(\vec{x_m}|\vec{y}) \quad \text{für alle } m \neq m' \tag{5.9}$$

wobei $m = 1,\ 2,\ 3,\ ...,\ M = 2^K$ gilt.
Die bedingte Wahrscheinlichkeit $P(\vec{x_m}|\vec{y})$ kann mit Hilfe des Kanalmodells ermittelt werden. Für den BSK gilt:

$$P(\vec{y}|\vec{x_m}) = \prod_{n=1}^{N} P(y_n|x_{m,\ n}) \tag{5.10}$$

und

$$P(\vec{x_m}|\vec{y}) = \frac{P(\vec{y}|\vec{x_m}) \cdot Q(\vec{x_m})}{P(\vec{y})} \tag{5.11}$$

wobei $Q(\vec{x_m})$ die Wahrscheinlichkeiten der Codewörter $\vec{x_m}$ sind.
Man erkennt aus der Gl. (5.11), daß die Decodierungsregel (5.9) auch in der Form

$$P(\vec{y}|\vec{x_{m'}}) \cdot Q(\vec{x_{m'}}) \geq P(\vec{y}|\vec{x_m}) \cdot Q(\vec{x_m}) \quad \text{für alle } m \neq m' \tag{5.12}$$

geschrieben werden kann. Die daraus abgeleitete, vereinfachte Decodierungsregel

$$P(\vec{y}|\vec{x_{m'}}) \geq P(\vec{y}|\vec{x_m}) \quad \text{für alle } m \neq m' \tag{5.13}$$

bezeichnet man als Maximum–Likelihood–Decodierung; sie ist nur dann der Regel (5.9) äquivalent, wenn alle Codewörter die gleiche Wahrscheinlichkeit aufweisen.
Die Maximum–Likelihood–Decodierung ordnet eine empfangene Symbolfolge $\vec{y}$ dem möglichen Codewort $\vec{x_{m'}}$ mit der größten bedingten Wahrscheinlichkeit $P(\vec{y}|\vec{x_{m'}})$ zu. Diese bedingte Wahrscheinlichkeit $P(\vec{y}|\vec{x_{m'}})$ ist im Fall gedächtnisloser Kanäle gemäß der Beziehung (5.10) nur von den Übergangswahrscheinlichkeiten $P(y_n|x_n)$ bzw. der Wahrscheinlichkeit der Fehlermuster abhängig. Eine Auswertung der Gl. (5.10) ergibt für den BSK

$P_0(\vec{y}|\vec{x_{m'}}) = (1-p)^N$ für ein Fehlermuster mit null Fehlern

$P_1(\vec{y}|\vec{x_{m'}}) = p(1-p)^{N-1}$ für ein Fehlermuster mit einem Fehler

$P_2(\vec{y}|\vec{x_{m'}}) = p^2(1-p)^{N-2}$ für ein Fehlermuster mit zwei Fehlern usw.

Da $P_0 > P_1 > P_2 > ... > P_N$ gilt, kann die Maximum–Likelihood–Decodierung auch derart interpretiert werden, daß eine empfangene Folge $\vec{y}$ dem möglichen Codewort $\vec{x_{m'}}$ mit der geringsten Hamming–Distanz $D(\vec{y}, \vec{x_{m'}})$ zugeordnet werden muß, d.h.

$$D(\vec{y}, \vec{x_{m'}}) \leq D(\vec{y}, \vec{x_m}) \quad \text{für alle} \quad m \neq m' \tag{5.14}$$

Die Beziehung (5.14) erklärt auch, warum der Korrekturbereich eines Codewortes in der geometrischen Darstellung der Codewörter in unmittelbarer Nachbarschaft um das Codewort herum angeordnet werden muß.
Die Maximum–Likelihood–Decodierung unterteilt die möglichen empfangenen Symbolfolgen $\vec{y}$ in M sog. Korrekturbereiche Y_m, wobei jeder Bereich einem vereinbarten Codewort $\vec{x_m}$ zugeordnet ist, das als Zentrum des Korrekturbereiches bezeichnet wird. Ein gesendetes Codewort $\vec{x_m}$ erzeugt am Kanalausgang eine empfangene Symbolfolge $\vec{y}$; diese wird fehlerfrei in $\vec{x_m}$ decodiert, wenn sie im Bereich

Y_m liegt, und fehlerhaft decodiert, wenn sie außerhalb von Y_m, im sog. Komplementärbereich Y_m^c, liegt. Die Fehlerwahrscheinlichkeit der Decodierung beträgt somit für jedes gesendete $\vec{x_m}$

$$P_{e,m} = \sum_{\vec{y} \in Y_m^c} P(\vec{y}|\vec{x_m}) \tag{5.15}$$

und gemittelt über alle m

$$P_e = \sum_{m=1}^{M} Q(\vec{x_m}) \cdot P_{e,m} \tag{5.16}$$

wobei $M = 2^K$ ist.

Während die Fehlerwahrscheinlichkeit eines Übertragungskanals sich immer auf ein übertragenes Symbol bezieht, wird die Restfehlerwahrscheinlichkeit P_e einer Kanalcodierung immer auf ein Codewort bezogen. Somit ist P_e die Wahrscheinlichkeit, daß die empfangene Symbolfolge $\vec{y}$ falsch decodiert wird.

Beispiel: Berechnung der Maximum–Likelihood–Decodierung für eine Blockcodierung nach Bild 5.4 im Fall eines binären, symmetrischen Kanals.

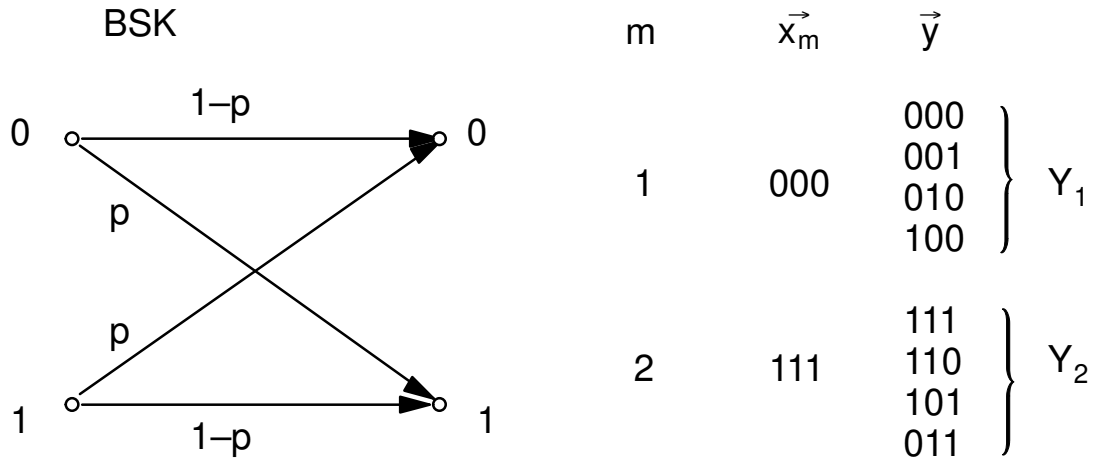


Bild 5.4 Blockcodierung mit N = 3, M = 2 für einen BSK

Die Maximum–Likelihood–Decodierung gemäß Gl. (5.13), zusammen mit Gl. (5.10), bildet für $\vec{x_1} = (000)$ den Korrekturbereich mit den Symbolfolgen (000), (001), (010), (100). Diese Symbolfolgen werden in $\vec{x_1}$ decodiert, die übrigen in $\vec{x_2}$. Es folgt für $P_{e,m}$ mit Gl. (5.15)

$$P_{e,1} = 3p^2 \cdot (1 - p) + p^3$$

$$P_{e,2} = 3p^2 \cdot (1 - p) + p^3$$

Treten die Codewörter $\vec{x_1}$ und $\vec{x_2}$ mit gleicher Wahrscheinlichkeit auf, gilt für den BSK ferner $P_{e,1} = P_{e,2} = P_e$.

Die Fehlerwahrscheinlichkeit P_e der Blockcodierung nach Bild 5.4 soll im folgenden noch für Codewortlängen $N > 3$ berechnet werden. Aufgrund von $M = 2$ gibt es nur zwei Codewörter $\vec{x_1}$ und $\vec{x_2}$ und zwei Korrekturbereiche Y_1 und Y_2. Diese sind in Bild 5.5 noch einmal räumlich für den Fall $N = 3$ dargestellt.

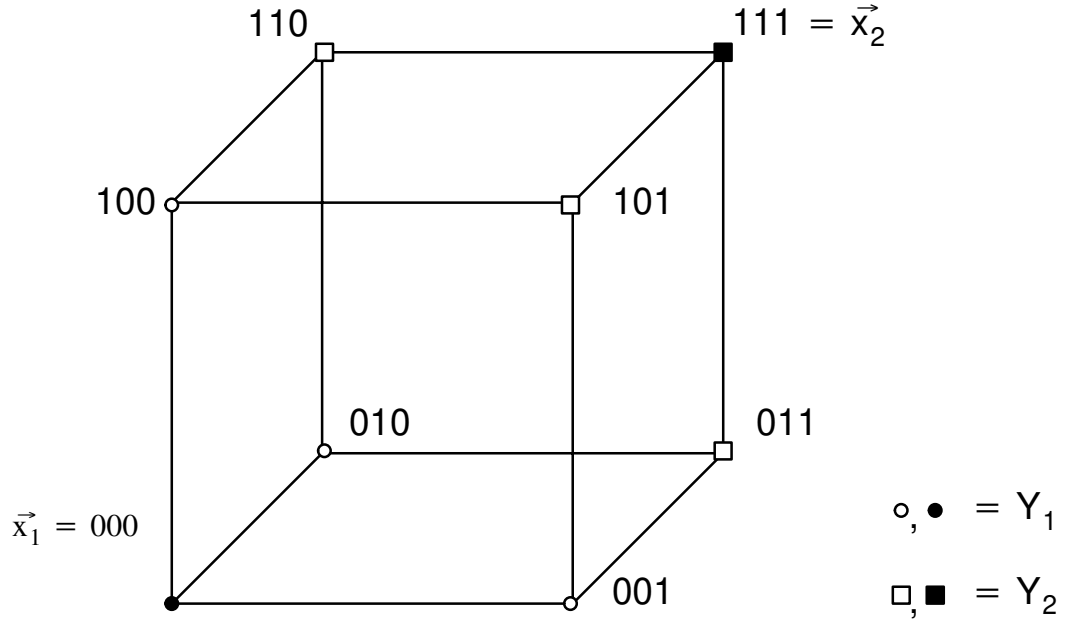


Bild 5.5 Räumliche Darstellung der Korrekturzentren und Korrekturbereiche für einen Code M = 2, N = 3

Wird die Codewortlänge N vergrößert, erweitern sich auch die beiden Korrekturbereiche. Die Wahrscheinlichkeiten $P_{e,m}$ sind dann nicht mehr so leicht zu bestimmen wie im oben angeführten Beispiel. Man behilft sich in diesen Fällen mit der Berechnung einer oberen Grenze für $P_{e,m}$. In diesem speziellen Fall mit nur zwei Korrekturbereichen wird immer dann ein Decodierungsfehler auftreten, wenn ein gesendetes Codewort mit N Stellen, wobei N ungerade angenommen werden soll, in $(N + 1)/2$ oder mehr Stellen verfälscht worden ist, d.h.

$$P_{e,m} = \sum_{i=(N+1)/2}^{N} \binom{N}{i} p^i (1 - p)^{N-i}. \tag{5.17}$$

Die größten Terme in dieser Summe werden von den Fehlermustern mit $(N+1)/2$ Binärfehlern, d.h. $i = (N+1)/2$, beigetragen. Diese Fehlermuster treten mit der Wahrscheinlichkeit $p^{(N+1)/2}(1-p)^{(N-1)/2}$ auf. Für $P_{e,m}$ kann nun eine obere Grenze angegeben werden, indem alle Wahrscheinlichkeiten der Summe Gl. (5.17) durch einen größeren Wert $p^{(N-1)/2}(1-p)^{(N-1)/2}$ ersetzt werden. Da jeder Komplementärbereich $2^N/2$ Symbolfolgen umfaßt, folgt

$$P_{e,m} \leq \frac{2^N}{2} p^{\frac{N-1}{2}} (1-p)^{\frac{N-1}{2}} \quad m = 1,2. \tag{5.18}$$

Die Ungleichung (5.18) gilt für beide m. Somit ist $P_{e,m}$ gleichzeitig eine obere Grenze für den mittleren Fehler P_e gemäß (5.16)

$$P_e \leq \left[2p^{\frac{1}{2}}(1-p)^{\frac{1}{2}}\right]^{N-1} . \tag{5.19}$$

Der Ausdruck in der rechteckigen Klammer ist im Fall $0 \leq p < \frac{1}{2}$ immer kleiner als 1. Damit ergibt sich aus der Ungleichung (5.19) die wesentliche Aussage, daß die Fehlerwahrscheinlichkeit von zwei Codewörtern $\vec{x_1}$, $\vec{x_2}$, die jeweils ein binäres Quellensymbol darstellen, mit wachsender Codewortlänge N gegen Null geht. Die Verringerung der Fehlerwahrscheinlichkeit P_e wird jedoch mit einer Abnahme der Übertragungsrate $R = 1/N$ erkauft. Dieser Nachteil der betrachteten Codierung kann nur durch eine Blockcodierung mit größerem Codewortvorrat $M > 2$ vermieden werden.

5.4 Theorem der Kanalcodierung

Theorem 1.9:

Es sei C die Kanalkapazität eines diskreten, gedächtnislosen Kanals und R die auf ein Kanalsymbol bezogene Entropie einer diskreten Quelle. Ist $R < C$, dann gibt es immer eine Codierung, die es ermöglicht, die Quellensymbole mit beliebig kleiner Fehlerwahrscheinlichkeit bzw. beliebig kleiner Äquivokation zu übertragen. Ist $R \geq C$, so beträgt die Äquivokation $R - C + \delta$, wobei δ beliebig klein ist. Es gibt dann aber keine Codierung mit einer Äquivokation kleiner als $R - C$! (5.20)

Der Beweis dieses Theorems der Kanalcodierung in allgemeingültiger Form für beliebige Kanäle ist relativ komplex. Während Shannon [7] für diskrete Kanäle zunächst einen heuristischen Beweis anführte, erbrachte Feinstein [2] erstmalig den mathematisch vollständigen Beweis. Zur Vereinfachung der Darstellung soll hier der Beweis lediglich für das Beispiel des binären, symmetrischen Kanals erbracht werden.

Für den Beweis wird ein Ensemble von Blockcodes betrachtet. Es wird gezeigt, daß der Mittelwert $\overline{P_e}$ der Fehlerwahrscheinlichkeiten dieser Codes mit wachsender Codewortlänge N exponentiell abnimmt, solange die auf ein Kanalsymbol bezogene Entropie der Quelle oder auch sog. Coderate $R = K/N$ kleiner als die Kanalkapazität C ist. Wenn das Ensemble von Codes im Mittel das Theorem erfüllt, muß zumindest ein Code im Ensemble enthalten sein, der das Theorem erfüllt.

Die Nachrichtenübertragung mit einem Ensemble von Codes oder einer sog. "Zufalls–Codierung" kann man sich auf folgende Weise vorstellen. Es wird eine große Anzahl von Blockcodes zusammengestellt, von denen jeder $M = 2^K$ Codewörter besitzt, die statistisch unabhängig und zufällig aus den 2^N verfügbaren Codewörtern ausgewählt werden. Die Blockcodes sind Coder und Decoder bekannt und werden in einer vereinbarten pseudostatistischen Reihenfolge angewendet. Bei dieser Nachrichtenübertragung stellt sich eine mittlere Fehlerwahrscheinlichkeit $\overline{P_e}$ ein, die für das Ensemble von Codes gilt.

Eine empfangene, gestörte Symbolfolge $\vec{y}$ kann als Modulo–2–Addition eines gesendeten Codewortes $\vec{x_m}$ und eines Fehlerwortes $\vec{e}$ dargestellt werden:

$$\vec{y} = \vec{x_m} \oplus \vec{e} \tag{5.21}$$

Das ebenfalls binäre Fehlerwort weist an ungestörten Binärstellen eine 0 und an gestörten Binärstellen eine 1 auf.

Beispiel: $\vec{x_m} = 0100...1110$

$\vec{e} = 0110...1100$

———————

$\vec{y} = 0010...0010$.

Bei großen Codewortlängen $N \to \infty$ wird das Fehlerwort infolge der Einwirkung eines binären, symmetrischen Kanals $N \cdot p$ Einsen und $N \cdot (1 - p)$ Nullen enthalten. Die Wahrscheinlichkeit für solch ein spezielles Fehlerwort beträgt aufgrund von Gl. (5.10)

$$P(\vec{e}) = P(\vec{y}|\vec{x_m}) = p^{N \cdot p} \cdot (1 - p)^{N \cdot (1-p)} \ . \tag{5.22}$$

Sie ist für alle Fehlerwörter gleich groß. Somit beträgt die Anzahl $Z(\vec{e})$ der möglichen Fehlerwörter

$$Z(\vec{e}) = \frac{1}{P(\vec{e})} = p^{-N \cdot p} \cdot (1 - p)^{-N \cdot (1-p)}. \tag{5.23}$$

Ein gesendetes Codewort $\vec{x_m}$ kann demzufolge $Z(\vec{e})$ verschiedene Symbolfolgen $\vec{y}$ beim Empfänger erzeugen. Sie bilden den Korrekturbereich Y_m des Codewortes $\vec{x_m}$. Aufgrund von Gl. (5.23) sind die Korrekturbereiche aller Codewörter gleich groß.

Ein Decodierungsfehler tritt auf, wenn eine durch $\vec{x_m}$ erzeugte Symbolfolge $\vec{y}$ in einen der anderen Korrekturbereiche fällt. Aufgrund der Zufalls–Codierung beträgt die Wahrscheinlichkeit, daß $\vec{y}$ innerhalb der 2^N möglichen Symbolfolgen mit einer Symbolfolge der $2^K - 1$ anderen Korrekturbereiche mit jeweils $Z(\vec{e})$ Symbolfolgen übereinstimmt

$$\begin{aligned} \overline{P_e} &= \frac{(2^K - 1)\,(Z(\vec{e})}{2^N} \\ &\simeq 2^{K-N} \cdot Z(\vec{e}) \quad . \end{aligned} \tag{5.24}$$

Der erste Faktor des Produktes (5.24) kann umgeformt werden in

$$2^{K-N} = 2^{-(N-K)} = 2^{-N(1-R)} \quad , \tag{5.25}$$

wobei R die Coderate gemäß Gl.(5.2) ist. Mit der Irrelevanz $H(U_1|U_2)$ des binären, symmetrischen Kanals gemäß Gl.(4.38)

$$H(U_1|U_2) = -\,[p \cdot \log_2 p + (1 - p) \cdot \log_2(1 - p)] \tag{5.26}$$

erhält man für den zweiten Faktor des Produktes (5.24)

$$\begin{aligned} Z(\vec{e}) &= p^{-N \cdot p} \cdot (1 - p)^{-N \cdot (1-p)} \\ &= 2^{\log_2 (p^{-N \cdot p} \cdot (1-p)^{-N \cdot (1-p)})} \\ &= 2^{-N[p \cdot \log_2(p) + (1-p) \cdot \log_2(1-p)]} \\ &= 2^{+N \cdot H(U_1|U_2)} \quad . \end{aligned} \tag{5.27}$$

Durch Einsetzen von Gl. (5.25) und (5.27) ergibt sich für die Fehlerwahrscheinlichkeit $\overline{P_e}$ gemäß Gl. (5.24)

$$\overline{P_e} \simeq 2^{-N \cdot [1 - R - H(U_1|U_2)]} \tag{5.28}$$

und mit der Kanalkapazität C_{BSK} nach Gl. (4.40)

$$\overline{P_e} \approx 2^{-N \cdot (C_{BSK} - R)} \quad . \tag{5.29}$$

Unter Verwendung der Hilfsbeziehung (2.8) kann die Fehlerwahrscheinlichkeit auch in der Form

$$\begin{aligned} \overline{P_e} &\approx \exp[-N \cdot (C_{BSK} - R) \cdot ln2] \\ \overline{P_e} &\approx \exp[-N \cdot E(R)] \end{aligned} \tag{5.30}$$

geschrieben werden, wobei $E(R)$ eine Funktion der Coderate und der Kanalkapazität, aber unabhängig von der Codewortlänge N ist.

Mit Gl. (5.30) ist das Theorem der Kanalcodierung für das Beispiel des binären, symmetrischen Kanals bewiesen. Die Fehlerwahrscheinlichkeit $\overline{P_e}$ kann durch Ver-

größerung der Codewortlänge N beliebig klein gewählt werden, solange die Differenz $C_{BSK} - R$ positiv ist, d.h.

$$R < C_{BSK} \tag{5.31}$$

gilt.

6. Irrelevanzreduzierende Codierung

6.1 Grundlagen der Rate–Distortion–Theorie

Das Ziel der folgenden Betrachtungen ist die Berechnung einer unteren Grenze der Übertragungsbitrate für eine gegebene Nachrichtenquelle, die eine Rekonstruktion der Nachricht unter Einhaltung eines vorgegebenen Gütekriteriums gewährleistet. Die theoretischen Grundlagen dieser Berechnungen liefert die Rate–Distortion–Theorie.

Mit Hilfe der Rate–Distortion–Theorie können neue und bekannte Quellencodierungen wie beispielsweise die Amplitudenquantisierung im Vergleich zur berechneten unteren Grenze der Übertragungsrate hinsichtlich ihrer Effektivität bewertet werden. Eine Schwierigkeit der Rate–Distortion–Theorie besteht in der Definition und Anwendung geeigneter Gütekriterien.

Es wird zunächst das Beispiel einer diskreten, gedächtnislosen Quelle U mit einem Symbolvorrat $(0,\ 1,\ ...,\ k,\ ...,\ K-1)$ betrachtet. Dabei sei K endlich und $Q(k)$ die Wahrscheinlichkeit der Quellensymbole k. Das Ausgangssignal dieser Quelle ist eine Symbolfolge $u_1,\ u_2,\ ...$ von statistisch unabhängigen Symbolen, wobei jedes u für ein beliebiges Symbol des Symbolvorrats $(0,\ 1,\ ...,\ K-1)$ steht. Nach Codierung, Übertragung und Decodierung möge der Sinke eine Symbolfolge $v_1,\ v_2,\ ...$ zugeführt werden, wobei jedes v für ein Symbol des Symbolvorrats $(0,\ 1,\ ...,\ j,\ ...,\ J-1)$ mit $J < \infty$ steht.

Mit Hilfe des Verzerrungsmaßes $d(k;\ j)$, das für alle $0 \le k \le K-1$ und $0 \le j \le J-1$ definiert ist, wird jeder möglichen Darstellung eines gesendeten Symbols k durch ein empfangenes Symbol j ein numerischer Wert als Maß für die Verzerrung oder Verfälschung beigemessen. Unter der Annahme fehlerfreier Übertragung ist die Verzerrung allein durch die Quellencodierung verursacht. Es sei vorausgesetzt, daß $d(k;\ j) \ge 0$ für alle $k,\ j$ gilt und für jedes k mindestens ein j existiert, für das $d(k;\ j) = 0$ gilt. Nach dieser Definition bedeutet ein großes $d(k;\ j)$ eine große Verfälschung des Symbols k. Die Verzerrungsmaße $d(k;\ j)$ können in einer $(K,\ J)$–Matrix zusammengestellt werden.

Zur Erläuterung seien zwei Beispiele angeführt:

Beispiel: 1. Beispiel: Annahme: $J = K$

$$d(k;\ j) = 0 \quad \text{für} \quad j = k$$

$$d(k;\ j) = 1 \quad \text{für} \quad j \neq k$$

Dieses Verzerrungsmaß gewichtet jede Verfälschung eines Symbols k gleich stark.–

Beispiel: 2. Beispiel: Annahme: $J = K$

$$d(k;\ j) = (j - k)^2 \quad \text{für alle } j \text{ und } k$$

Dieses Verzerrungsmaß ist für Symbole geeignet, die Amplitudenwerte darstellen, und entspricht dann dem quadratischen Fehler.–

Zur Berechnung der unteren Grenze der Übertragungsrate werden Quellencoder und Quellendecoder zusammen als ein gestörter Kanal betrachtet, wobei nur der Quellencoder Störungen erzeugt. Demzufolge kann die erforderliche Übertragungsrate am Ausgang des Quellencoders mittels der Transformation $F(Q;\ P)$ berechnet werden. Mit den Übergangswahrscheinlichkeiten $P(\,j|k)$ und den Wahrscheinlichkeiten $Q(k)$ der Quellensymbole ergibt sich für die mittlere Transinformation $F(Q;\ P) = H(V) - H(V/v)$.

$$F(Q;\ P) = \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} Q(k) \cdot P(j|k) \cdot \log_2 \frac{P(j|k)}{\sum_{i=0}^{K-1} Q(i) \cdot P(j|i)} \tag{6.1}$$

und die mittlere Verzerrung $\overline{d}$

$$\overline{d} = \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} Q(k) \cdot P(j|k) \cdot d(k;\ j). \tag{6.2}$$

Während bei der Berechnung der mittleren Transinformation eines Übertragungskanals die Übertragungswahrscheinlichkeiten $P(\,j|k)$ durch die physikalischen Eigenschaften des Kanals vorgegeben sind, können diese im vorliegenden Fall mit der Quellencodierung so gewählt werden, daß sich eine möglichst kleine Übertragungsrate ergibt. Dabei muß die als Gütekriterium vorgegebene maximale mittlere Verzerrung d^* eingehalten werden. Das Ergebnis dieser Optimierung ist die Rate–Distortion–Funktion $R(d^*)$

$$R(d^*) = \min_{P\,:\,\overline{d} \le d^*} F(Q;\ P). \tag{6.3}$$

Es kann nachgewiesen werden, daß eine Quelle stets mit einer Übertragungsbitrate entsprechend $R(d^*)$ codiert werden kann, ohne daß die vorgegebene zulässige Verzerrung d^* um mehr als ein beliebig kleines δ überschritten wird. Umgekehrt läßt sich zeigen, daß eine kleinere Übertragungsrate als $R(d^*)$ nur durch Überschreiten

der zulässigen Verzerrung d^* möglich ist. $R(d^*)$ kann deshalb als untere Grenze der für eine Quelle erforderlichen Übertragungsrate bei vorgegebenen Gütekriterium d^* angesehen werden.

Nachfolgend sollen einige Eigenschaften der Rate–Distortion–Funktion $R(d^*)$ erörtert werden.

Theorem 1.10:

$R(d^*)$ ist stets positiv, da die Transformation $F(Q;\ P)$ stets größer als Null ist. (6.4)

Theorem 1.11:

$R(d^*)$ ist eine mit wachsendem d^* fallende Funktion. Beweis: Die Menge der erlaubten Wahrscheinlichkeitsverteilungen $P(j|k)$ ist begrenzt und wächst mit d^*. Somit kann sich $R(d^*)$ mit wachsendem d^* nur vermindern. (6.5)

Theorem 1.12:

$R(d^*)$ ist eine konvexe U–Funktion, d.h. für jedes Paar von Verzerrungswerten d_1^* und d_2^* und beliebigen Wert $\lambda,\ 0 < \lambda < 1$, gilt die Ungleichung

$$R[\lambda \cdot d_1^* + (1 - \lambda) \cdot d_2^*\,] \leq \lambda \cdot R(d_1^*) + (1 - \lambda) \cdot R(d_2^*) \qquad (6.6)$$

dieser Zusammenhang wird in Bild 6.1 veranschaulicht und bedeutet, daß $R(d^*)$ zwischen den Abszissenwerten d_1^* und d_2^* immer unter der Sehne liegt.

Für die Sehne gilt dabei die Geradengleichung

$$y(d^*) = R(d_2^*) + \frac{R(d_1^*) - R(d_2^*)}{d_1^* - d_2^*} \cdot \lambda \cdot (d_1^* - d_2^*)$$

$$y(d^*) = \lambda \cdot R(d_1^*) + (1 - \lambda) \cdot R(d_2^*) \quad . \qquad (6.7)$$

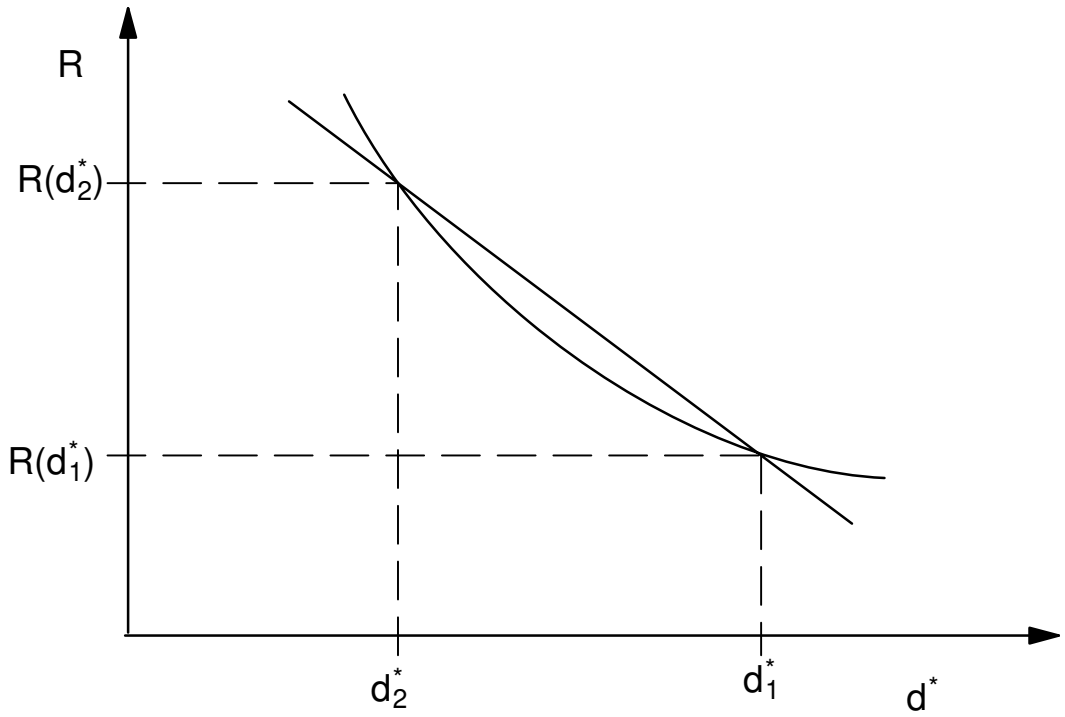


Bild 6.1 Eine konvexe U–Funktion

Beweis:

Beweis der Ungleichung (6.6) :
Es seien $P_1(j|k)$ und $P_2(j|k)$ die zu d_1^* und d_2^* gehörenden Verteilungen der Übergangswahrscheinlichkeiten, die Gl. (6.3) erfüllen, und $P'(j|k)$ eine Hilfsgröße, definiert durch

$$P'(j|k) = \lambda \cdot P_1(j|k) + (1 - \lambda) \cdot P_2(j|k). \tag{6.8}$$

Da $\overline{d}$ eine lineare Funktion von $P(j|k)$ ist, folgt

$$\begin{aligned} \overline{d}\,[\,P'(j|k)] &= \overline{d}\,[\,\lambda \cdot P_1(j|k) + (1 - \lambda) \cdot P_2(j|k)] \\ &= \overline{d}\,[\,\lambda \cdot P_1(j|k) + \overline{d}\,(1 - \lambda) \cdot P_2(j|k)] \\ &\lambda \cdot d_1^* + (1 - \lambda) \cdot d_2^*. \end{aligned} \tag{6.9}$$

Ferner gilt aufgrund von Gl. (6.3) und (6.9)

$$R[\lambda \cdot d_1^* + (1 - \lambda) \cdot d_2^*] \leq F(Q;\ P'). \tag{6.10}$$

Durch Einsetzen von Gl. (6.8) in $F(Q;P)$ und Erweiterung der Argumente der logarithmischen Funktionen erhält man

$$F(Q;\ P') = \lambda \sum_k \sum_j Q(k) \cdot P_1(j|k) \cdot \log_2\left[\frac{P'(j|k)}{P'(j)} \cdot \frac{P_1(j|k) \cdot P_1(j)}{P_1(j|k) \cdot P_1(j)}\right]$$
$$+ (1 - \lambda) \sum_k \sum_j Q(k) \cdot P_2(j|k) \cdot \log_2\left[\frac{P'(j|k)}{P'(j)} \cdot \frac{P_2(j|k) \cdot P_2(j)}{P_2(j|k) \cdot P_2(j)}\right] \tag{6.11}$$

Durch Zusammenfassen der Gleichungen (2.7), (2.8) erhält man

$$\log_2(x) \leq (x - 1) \cdot \log_2(e). \tag{6.12}$$

Damit kann Gl. (6.11) umgeformt werden in

$$F(Q;\ P') \leq \lambda \sum_k \sum_j Q(k) \cdot P_1(j|k) \cdot \left[\log_2 \frac{P_1(j|k)}{P_1(j)} + \left(\frac{P'(j|k) \cdot P_1(j)}{P'(j) \cdot P_1(j|k)} - 1\right) \cdot \log_2(e)\right]$$
$$+ (1 - \lambda) \sum_k \sum_j Q(k) \cdot P_2(j|k) \cdot \left[\log_2 \frac{P_2(j|k)}{P_2(j)} + \left(\frac{P'(j|k) \cdot P_2(j)}{P'(j) \cdot P_2(j|k)} - 1\right) \cdot \log_2(e)\right] \tag{6.13}$$

$$F(Q;\ P) \leq \lambda\left[F(Q;\ P_1) + \left[\sum_j \frac{P_1(j)}{P'(j)} \sum_k Q(k) \cdot P'(j|k) - 1\right] \cdot \log_2(e)\right]$$
$$+ (1 - \lambda)\left[F(Q;\ P_2) + \left[\sum_j \frac{P_2(j)}{P'(j)} \sum_k Q(k) \cdot P'(j|k) - 1\right] \cdot \log_2(e)\right]. \tag{6.14}$$

Da die Summation $\sum_k Q(k) \cdot P'(j|k) = P'(j)$ ergibt, folgt aus Gl. (6.14)

$$F(Q,\ P') \leq \lambda \cdot F(Q, P_1) + (1 - \lambda) \cdot F(Q,\ P_2)^{\wedge}$$

$$F(Q,\ P') \leq \lambda \cdot R(d_1^*) + (1 - \lambda) \cdot R(d_2^*). \tag{6.15}$$

Die Kombination von Gl. (6.15) und (6.10) liefert den Beweis des Theorems 1.12–

Bild 6.2 zeigt eine Skizze der Rate–Distortion–Funktion $R(d^*)$.

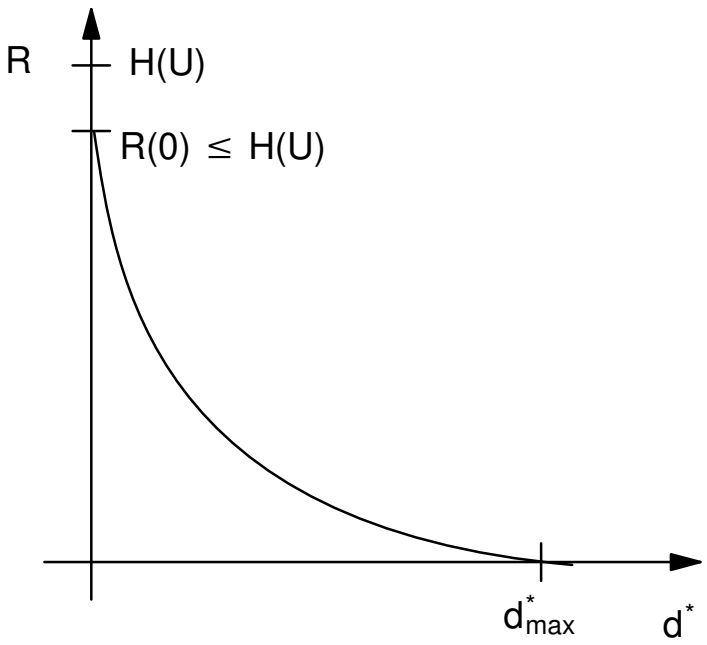


Bild 6.2 Typischer Verlauf der Rate–Distortion–Funktion $R(d^*)$ einer diskreten, gedächtnislosen Quelle

Der kleinste Wert der Verzerrung ist $d^* = 0$. In diesem Fall dürfen nur Symbolpaare $k,\ j$ auftreten, für die $d(k;\ j) = 0$ gilt. Der Maximalwert für $R(d^* = 0)$ ist $H(U)$. Es sei $d^*_{\max}$ als der kleinste Wert d^* definiert, für den $R(d^*) = 0$ wird, d.h.

$$d^*_{\max} = \min \sum_k \sum_j Q(k) \cdot P(j|k) \cdot d(k;\ j). \tag{6.16}$$

Die Nebenbedingung $R(d^*_{\max}) = 0$ bedeutet, daß die mittlere Transinformation Null ist bzw. die Symbole k und j statistisch unabhängig sind; somit folgt für Gl. (6.16)

$$d^*_{\max} = \min \sum_j P(j) \sum_k Q(k) \cdot d(k;j). \tag{6.17}$$

Das Minimum wird erzielt, indem die Summen $\sum_k Q(k) \cdot d(k;\ j)$ für alle j nacheinander überprüft werden und für das j mit der kleinsten Summe $P(j) = 1$ gewählt wird und Null in allen anderen Fällen. Die minimale Verzerrung $d^*_{\max}$ wird somit

$$d^*_{\max} = \min_j \sum_k Q(k) \cdot d(k;\ j). \tag{6.18}$$

Um für eine gegebene Übertragungsrate $R(d^*)$ eine möglichst geringe mittlere Verzerrung $\overline{d} \leq d^*$ zu erreichen, muß die Quellencodierung der Quellenstatistik und dem Verzerrungsmaß d^* angepaßt werden. Eine optimale Quellencodierung erfüllt Gl. (6.3).

Zur Codierung der diskreten, gedächtnislosen Quelle U wird eine Blockcodierung mit M Codewörtern angewendet, die eine Folge von L Quellensymbolen $\vec{u} = u_1,\ u_2,\ ...,\ u_L$ in eine Folge $\vec{v} = v_1,\ v_2,\ ...,\ v_L$ von Symbolen der Sinke abbildet, die hier als Codewort bezeichnet wird. Bild 6.3 zeigt ein Beispiel.

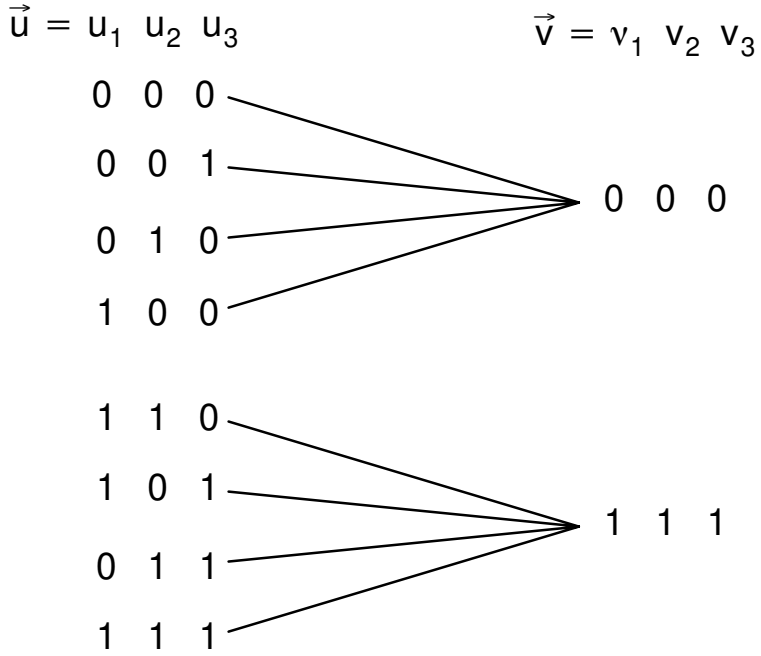


Bild 6.3 Codetabelle für eine Quellencodierung mit L=3 und M=2

Bei gegebenen Verzerrungsmaß $d(k;\ j)$ kann die mittlere Verzerrung je Symbol berechnet werden zu

$$\overline{d}_L = \frac{1}{L} \sum_{\vec{u}} Q(\vec{u}) \cdot D\left[\vec{u};\ \vec{v}(\vec{u})\right] \tag{6.19}$$

wobei $\vec{v}(\vec{u})$ das $\vec{u}$ entsprechende Codewort ist und

$$D\left[\vec{u};\ \vec{v}(\vec{u})\right] = \sum_{l=1}^{L} d(u_1;\ v_1) \tag{6.20}$$

gilt. Da $\overline{d}_L$ die auf ein Symbol bezogene mittlere Verzerrung ist, kann Gl. (6.19) auch in der bereits bekannten Form

$$\overline{d}_L = \sum_j \sum_k Q(k) \cdot P(j|k) \cdot d(k;\ j) \tag{6.21}$$

geschrieben werden.

Jedes der M Codewörter habe einen Informationsgehalt von $\log_2 M$ bit. Die Übertragungsbitrate je Codeelement beträgt dann $1/L \cdot \log_2 M$ und kann über einen Kanal der Kapazität $C = 1/L \cdot \log_2 M$ bit übertragen werden. Das zu lösende Problem besteht nun darin, bei vorgegebenen Werten für L und M einer Codierung die kleinstmögliche Verzerrung $\overline{d}_L$ zu erreichen.

Die Lösung dieses Problems ist mit den beiden folgenden Theoremen gegeben, die hier ohne Beweis angeführt werden.

Theorem der Quellencodierung bei Berücksichtigung eines Gütekriteriums

Theorem 1.13:

Es sei $R(d^*)$ die Rate–Distortion–Funktion einer diskreten, gedächtnislosen Quelle U. Dann existiert für jedes $d^* \geq 0,\ \delta > 0$ und hinreichend groß gewählter Blocklänge L ein Code mit M Codewörtern und der Übertragungsrate

$$\frac{1}{L}\log_2 M \leq R(d^*) + \delta, \tag{6.22}$$

der die Ungleichung

$$\overline{d} \leq d^* + \delta \tag{6.23}$$

erfüllt.–

Umgekehrtes Theorem der Quellencodierung bei Berücksichtigung eines Gütekriteriums

Theorem 1.14:

Ist $R(d^*)$ die Rate–Distortion–Funktion einer diskreten, gedächtislosen Quelle U, dann existiert kein Code, der das Gütekriterium $\overline{d} \leq d^*$ einhält und eine Übertragungsrate

$$\frac{1}{L}\log_2 M < R(d^*) \tag{6.24}$$

aufweist.–

Nach diesen Überlegungen ist ein Übertragungssystem dann optimal genutzt, wenn es aus einem Quellencoder besteht, der die Übertragungsrate von $R(d^*)$ bit je Quellensymbol liefert, und einen Kanalcoder aufweist, der diese Übertragungsrate bis auf eine beliebig kleine Restfehlerwahrscheinlichkeit gegen Übertragungsfehler des Kanals einer Kanalkapazität $C \geq R(d^*)$ sichert.

Die für eine optimale Codierung mit der Übertragungsrate $R(d^*)$ erforderlichen Übertragungswahrscheinlichkeiten $P(j \mid k)$ werden im Quellencoder durch eine Blockcodierung realisiert. Um eine eindeutige Decodierung zu gewährleisten, muß die Zuordnung zwischen den Symbolfolgen $\vec{u}$ und den Codewörtern $\vec{v}$ deterministisch sein, d.h. $P(\vec{v} \mid \vec{u}) = 0$ oder 1. Um gleichzeitig die gewünschten Übergangswahrscheinlichkeiten $P(j \mid k)$ zu erreichen, müssen die Zuordnungen dabei so gewählt werden. daß sich für die einzelnen Symbole die Übergangswahrscheinlichkeiten $P(j \mid k)$ des optimalen Code einstellen. Eine Codierung von Einzelsymbolen müßte immer deterministisch sein und kann beide Bedingungen daher nur in Sonderfällen erfüllen.

6.2 Rate–Distortion–Funktion $R(d^*)$ diskreter Quellen

Im folgenden soll ein Berechnungsverfahren zur Bestimmung der optimalen Übergangswahrscheinlichkeiten $P(j|k)$ abgeleitet werden, die auf eine Übertragungsrate $R(d^*)$ führen. Die Berechnung der Gl. (6.3)

$$R(d^*) = \min_{P(j|k)} F\,[\,P(j|k)] = \min_{P(j|} k) \sum_k \sum_j Q(k) \cdot P(j|k) \cdot \log \frac{P(j|k)}{P(j)} \tag{6.25}$$

mit

$$P(\,j) = \sum_i Q(i) \cdot P(j|i) \tag{6.26}$$

führt auf ein Extremwertproblem. Es ist das Minimum der Funktion $F[P(j|k)]$ unter den Nebenbedingungen $\overline{d} = d^*$, d.h.

$$\sum_k \sum_j Q(k) \cdot P(j \mid k) \cdot d(k;\ j) - d^* = 0 \tag{6.27}$$

und

$$\sum_j P(j|k) - 1 = 0 \qquad \text{für alle } k \tag{6.28}$$

bei positiven Wahrscheinlichkeiten $P(j|k)$ zu suchen. Die Nebenbedingung (6.27) folgt aus der Forderung $\overline{d} \leq d^*$. Aufgrund der Eigenschaft (6.5) muß das Minimum von $R(d^*)$ bei $\overline{d} = d^*$ liegen.

Zur Lösung des Extremwertproblems wird das Lagrange'sche Verfahren herangezogen. Dazu werden die Nebenbedingungen zusammen mit Lagrange'schen Multiplikatoren μ_k, s zur minimierenden Funktion $F[P(j|k)]$ addiert und das Minimum der Hilfsfunktion $\Phi[P(j|k),\ \mu_k,\ s]$ berechnet, wobei

$$\Phi\Big[P(j|k),\ \mu_k,\ s\Big] = F\Big[P(j|k)\Big] + \sum_k \mu_k \cdot \left[\sum_j P(j|k) - 1\right] + s \cdot \left[\sum_k \sum_j Q(k) \cdot P(j|k) \cdot d(k;\ j) - d^*\right]. \tag{6.29}$$

Zur Vereinfachung der Berechnung wird der in $F[P(j|k)]$ enthaltene Logarithmus zur Basis e gerechnet, d.h. $R(d^{*)}$ in der Einheit "nat" berechnet.

$$\Phi\Big[P(j|k),\ \mu_k,\ s\Big] = + \sum_k \sum_j Q(k) \cdot P(j|k) \cdot \ln \frac{P(j|k)}{\sum_i Q(i) \cdot P(j|i)} + \sum_k \mu_k \cdot \left[\sum_j P(j|k) \cdot - 1\right] + s \cdot \left[\sum_k \sum_j Q(k) \cdot P(j|k) \cdot d(k;\ j) - d^*\right]. \tag{6.30}$$

Die Extremwertberechnung von $\Phi[P(j|k),\ \mu_k,\ s]$ liefert

$$\frac{\partial \Phi}{\partial P(j|k)} = 0 \qquad \text{für alle } \ j,\ k \tag{6.31}$$

zunächst $K \cdot J$ Bestimmungsgleichungen und mit den Nebenbedingungen (6.27) und (6.28) zusätzlich $1 + K$ weitere Bestimmungsgleichungen zur Berechnung der insgesamt $K \cdot J + K + 1$ Unbekannten $P(j|k)$, μ_k und s.

Die Differentiation von GL. (6.30) ergibt für jedes spezielle Wertepaar $j,\ k$:

$$\frac{\partial \Phi}{\partial P(j|k)} = Q(k) \cdot \ln \frac{P(j|k)}{\sum_i Q(i) \cdot P(j|i)} + Q(k) - Q(k) \cdot P(j|k) \cdot \frac{Q(k)}{\sum_i Q(i) \cdot P(j|i)} - \frac{\sum_{i \neq k} Q(i) \cdot P(j|i) \cdot Q(k)}{\sum_i Q(i) \cdot P(j|i)} + \mu_k + s \cdot Q(k) \cdot d(k;j) \,. \tag{6.32}$$

Nach Zusammenfassung der Summanden folgt

$$\frac{\partial \Phi}{\partial P(j|k)} = Q(k) \cdot \left[\ln \frac{P(j|k)}{P(j)} + 1 - \frac{\sum_i Q(i) \cdot P(j|i)}{\sum_i Q(i) \cdot P(j|i)} \right] + \mu_k + s \cdot Q(k) \cdot d(k;j) \,. \tag{6.33}$$

Wird zur Vereinfachung der Rechnung ferner die Substitution

$$\mu_k = - Q(k) \cdot \ln \frac{f_k}{Q(k)} \tag{6.34}$$

eingeführt, so folgt für (6.33) mit (6.34)

$$\frac{\partial \Phi}{\partial P(j|k)} = Q(k) \cdot \ln \left[\frac{P(j|k) \cdot Q(k)}{P(j) \cdot f_k} + s \cdot d(k;\ j) \right] = 0 \,. \tag{6.35}$$

Die Auflösung nach $P(j|k)$ liefert

$$P(j|k) = f_k \cdot \frac{P(j)}{Q(k)} \cdot e^{-s \cdot d(k;\, j)} \tag{6.36}$$

für alle Wertepaare $j,\ k$. Dadurch, daß sich die optimalen $P(j|k)$ in Gl. (6.36) als Produkt von $f_k \cdot P(j)$ ergeben, werden anstelle von $K \cdot J + K + 1$ Bestimmungsgleichungen nur noch $J + K + 1$ Bestimmungsgleichungen benötigt, die mit (6.36), (6.27) und (6.28) berechnet werden können. Die Unbekannten sind $P(j)$, f_k und s.

Werden beide Seiten der Gl. (6.36) mit $Q(k)$ multipliziert und über k aufsummiert, erhält man

$$1 = \sum_k f_k \cdot e^{-s \cdot d(k;\, j)} \tag{6.37}$$

für alle j. Mit (6.36) und der Nebenbedingung $\sum_j P(j|k) = 1$ folgt

$$1 = \frac{f_k}{Q(k)} \sum_j P(j) \cdot e^{-s \cdot d(k;\, j)} \tag{6.38}$$

für alle k. Aus der Nebenbedingung $\overline{d} = d^*$ ergibt sich mit der Lösung (6.36) die mittlere Verzerrung

$$d^* = \sum_k \sum_j f_k \cdot P(j) \cdot e^{-s \cdot d(k;\, j)} \cdot d(k;\ j) \tag{6.39}$$

als Bestimmungsgleichung für s. Für die Rate–Distortion–Funktion folgt

$$\begin{aligned} R(d^*) &= \sum_k \sum_j Q(k) \cdot P(j|k) \cdot \ln \frac{P(j|k)}{P(j)} \\ &= \sum_k \sum_j Q(k) \cdot P(j|k) \cdot \left[\ln \frac{f_k}{Q(k)} - s \cdot d(k;\ j) \right] \\ &= \sum_k Q(k) \cdot \ln \frac{f_k}{Q(k)} - s \cdot d^* \\ &= \sum_k Q(k) \cdot \log_2 \frac{f_k}{Q(k)} - s \cdot d^* \cdot \log_2 e. \end{aligned} \tag{6.40}$$

Da $R(d^*)$ eine konvexe U–Funktion ist, muß das durch Gl. (6.40) dargestellte Extremum ein Minimum sein. Die Lösungen (6.39) und (6.40) werden auch als "Parameter–Form" der Rate–Distortion–Funktion bezeichnet. Sie gelten nur für Bereiche mit $P(j|k) \geq 0$.

Es kann nachgewiesen werden, daß für die Ableitung $R(d^*)$ nach d^*, vgl. dazu T. Berger (1971), [1].

$$\frac{d[R(d^*)]}{d(d^*)} = -s \tag{6.41}$$

gilt. Somit kann $(-s)$ als das Steigungsmaß der Funktion interpretiert werden. Dabei ist s selbst noch eine Funktion von d^*.

6.3 Rate–Distortion–Funktion $R(d^*)$ kontinuierlicher Quellen

Die Quellensignale zeit– und wertkontinuierlicher Quellen, wie Sprachquellen und Bildquellen, sind im allgemeinen bandbegrenzt und können gemäß dem Abtasttheo-

rem durch äquidistante Abtastwerte des Quellensignals dargestellt werden. Die Folge der Abtastwerte kann als ein zeitdiskretes, wertkontinuierliches Signal angesehen werden. Die folgenden Betrachtungen zur Rate–Distortion–Funktion kontinuierlicher Quellen beschränken sich auf diese für die Praxis sehr bedeutenden zeitdiskreten, wertkontinuierlichen Quellen.

Es sei $u_1,\ u_2,\ ...,\ u_L$ eine Folge von Quellensymbolen bzw. Abtastwerten, wobei jedes u für ein beliebiges Symbol x einer zeitdiskreten, wertkontinuierlichen Quelle steht. Die Quellensymbole x entstammen einem kontinuierlichen Wertevorrat $-\infty \leq x \leq +\infty$. Es sei $q(u)$ die Wahrscheinlichkeitsdichte der Quellensymbole u. Durch eine Quellencodierung möge die ursprüngliche Folge der Quellensymbole in eine Folge empfangener Codesymbole $v_1,\ v_2,\ ...,\ v_L$ abgebildet werden, wobei jedes v für ein beliebiges Symbol y steht.

Die Verzerrung wird mit dem Verzerrungsmaß $d(u;\ v)$ gemessen, das jedem möglichen Wertepaar $u,\ v$ von Quellensymbol und Codesymbol einen numerischen positiven Zahlenwert zuordnet. Zu jedem u möge es mindestens einen Wert v geben, gewöhnlich $u = v$, für den $d(u;\ v) = 0$ gilt. Als Gütekriterium soll hier vor allem die quadratische Abweichung $d(u;\ v) = (v - u)^2$ betrachtet werden.

Die Rate–Distortion–Funktion wird für die zeitdiskrete, wertkontinuierliche Quelle in gleicher Weise definiert wie für die diskrete Quelle. Anstelle der Summen treten hier lediglich Integrale auf

$$R(d^*) = \min_{p(v|u)} \int_u \int_v q(u) \cdot p(v|u) \cdot \log_2 \frac{p(v|u)}{p(v)}\, du\ dv. \tag{6.42}$$

Die Lösung der Gl. (6.42) unter der Nebenbedingung $\overline{d} \leq d^*$, wobei

$$\overline{d} = \int_u \int_v q(u) \cdot p(v|u) \cdot d(u;\ v)\, du\ dv \tag{6.43}$$

und der Nebenbedingung

$$\int_v p(v|u)\, dv = 1 \tag{6.44}$$

stellt ein Variationsproblem dar. Es kann gezeigt werden, daß $R(d^*)$ auch in diesem Fall eine konvexe Funktion ist. Im Unterschied zur diskreten Quelle kann

$$\lim_{d^* \to 0} R(d^*) = \infty$$

sein, wenn der Wertevorrat der Quellensymbole unendlich groß ist. Ferner kann nachgewiesen werden, vergl. dazu R. G. Gallager, S. 471 ff. [3], daß die Codierungs-

theoreme der diskreten Quelle auch für die zeitdiskrete, wertkontinuierliche Quelle gelten.
Die Berechnung von $R(d^*)$ führt auf folgende Lösungen, die hier an die Stelle der Gl. (6.36) bis (6.40) der diskreten Quelle treten:

$$p(v|u) = \frac{f(u) \cdot p(v)}{q(u)} \cdot e^{-s \cdot d(u;\, v)}, \tag{6.45}$$

wobei $f(u)$ die Bedingung

$$1 = \int_u f(u) \cdot e^{-s \cdot d(u;\, v)}\, du \tag{6.46}$$

für alle v erfüllt. Aus der Nebenbedingung (6.44) folgt hier

$$1 = \frac{f(u)}{q(u)} \int_v p(v) \cdot e^{-s \cdot d(u;\, v)} dv \tag{6.47}$$

für alle u. Für die Parameter–Darstellung der Rate–Distortion–Funktion ergibt sich mit diesen Lösungen

$$d^* = \int_u \int_v f(u) \cdot p(v) \cdot e^{-s \cdot d(u;\, v)} \cdot d(u;\ v)\, du\, dv \tag{6.48}$$

und

$$\begin{aligned} R(d^*) &= \int_u q(u) \cdot \ln \frac{f(u)}{q(u)}\, du - s \cdot d^* && \text{[nat]} \\ R(d^*) &= \int_u q(u) \cdot \log_2 \frac{f(u)}{q(u)}\, du - s \cdot d^* \cdot \log_2 e && \text{[bit]} \quad . \end{aligned} \tag{6.49}$$

Mit der Wahl eines Gütekriteriums $d(u;\ v)$ kann $R(d^*)$ noch näher spezifiziert werden. Für den Sonderfall

$$d(u;\ v) = (v - u)^2 \tag{6.50}$$

d.h. der quadratischen Abweichung als Verzerrungsmaß folgt aus (6.46)

$$1 = \int_{-\infty}^{+\infty} f(u) \cdot \exp[-s \cdot (v - u)^2]\, du \tag{6.51}$$

für alle v. Über die Substitution $z = u - v$ im Integral

$$\int f(z + v) \cdot \exp[-s \cdot z^2]\, dz$$

erkennt man, daß die Forderung (6.51) nur eingehalten werden kann, wenn $f(u)$ eine Konstante ist, gegeben durch

$$f(u) = \sqrt{\frac{s}{\pi}}. \tag{6.52}$$

Durch Einsetzen von (6.52) in (6.45) erhält man

$$\frac{p(u|v) \cdot q(u)}{p(v)} = \sqrt{\frac{s}{\pi}} \cdot e^{-s \cdot (v-u)^2} = p(u|v). \tag{6.53}$$

Die Konstante s wird mit Hilfe der Nebenbedingung (6.48) und der Substitution $z = u - v$ bestimmt

$$\begin{aligned} d^* &= \int_u \int_v \sqrt{\frac{s}{\pi}} \cdot p(v) \cdot e^{-s \cdot (v-u)^2} \cdot (v - u)^2\, du\, dv \\ &= \int_v p(v)\, dv \int_z \sqrt{\frac{s}{\pi}} \cdot e^{-s \cdot z^2} \cdot z^2\, dz = \frac{1}{2s}\ . \end{aligned} \tag{6.54}$$

Durch Umformung von Gl. (6.49) erhält man für $R(d^*)$ mit Gl. (6.52)

$$\begin{aligned} R(d^*) &= H(U) + \log_2 f(u) - s \cdot d^* \cdot \log_2(e) \\ &= H(U) + \frac{1}{2} \cdot \log_2 \frac{s}{\pi} - s \cdot d^* \cdot \log_2 e \end{aligned} \tag{6.55}$$

und nach Einsetzen der Gl. (6.54)

$$R(d^*) = H(U) - \frac{1}{2} \cdot \log_2 (2\pi \cdot e \cdot d^*). \tag{6.56}$$

Diese Beziehung für $R(d^*)$ gilt nur für das spezielle Gütekriterium $d(u;\ v) = (v - u)^2$. Der zweite Term auf der rechten Seite der Gl. (6.56) kann als die mögliche Entropiereduktion bei vorgegebenem Gütekriterium d^* interpretiert werden. Um $R(d^*)$ zu erreichen, muß die Verzerrung $v - u$ eine Gaußsche Wahrscheinlichkeitsdichte mit der Varianz d^* aufweisen.

Für eine zeitdiskrete, wertekontinuierliche, gedächtnislose Gaußschen Quelle mit der Entropie $H_G(U) = \frac{1}{2}\log_2(2\pi e\sigma_u^2)$ und der Varianz $\sigma_u^2 = \sigma_x^2$ wird

$$R_G(d^*) = \frac{1}{2}\log_2 (2\pi e{\sigma_x}^2) - \frac{1}{2}\log_2 (2\pi e d^*)$$
$$R_G(d^*) = \frac{1}{2}\log_2 \frac{{\sigma_x}^2}{d^*} \tag{6.57}$$

in bit je Quellensymbol bzw. Abtastwert.
Im Grenzfall, daß alle u in $v = 0$ abgebildet werden, beträgt die mittlere Verzerrung $\overline{d} = \sigma^2$. Daraus folgt

$$R(d^*) = \frac{1}{2}\log_2 \frac{\sigma^2}{d^*} \qquad d^* < \sigma^2$$
$$R(d^*) = 0 \qquad d^* \leq \sigma^2$$

Eine Umrechnung des Signal–Rauschleistungs–Verhältnisses σ_x^2/d^* in die Einheit "Dezibel" liefert

$$R_G(d^*) = \frac{1}{20} \cdot 10 \cdot \log_{10}\frac{\sigma_x^2}{d^*} \cdot \log_2(10) = 0,166 \cdot \left(\frac{\sigma_x^2}{d^*}\right)_{dB}.$$

Diese Funktion ist in Bild 6.4 dargestellt. Das Signal–Rauschleistungs–Verhältnis wird auch mit "Signal–to–Noise–Ratio" (SNR) bezeichnet. Damit kann Gl. (6.57) auch in der Form

$$SNR_{\mathrm{dB}} = 6,02 \cdot R_G(d^*) \tag{6.58}$$

geschrieben werden. Aus Gl. (6.58) ist zu entnehmen, daß jede Vergrößerung der Übertragungsrate um 1 bit je Abtastwert eine Verbesserung des Signal–Rauschleistungs–Verhältnisses von etwa 6 dB ergibt.

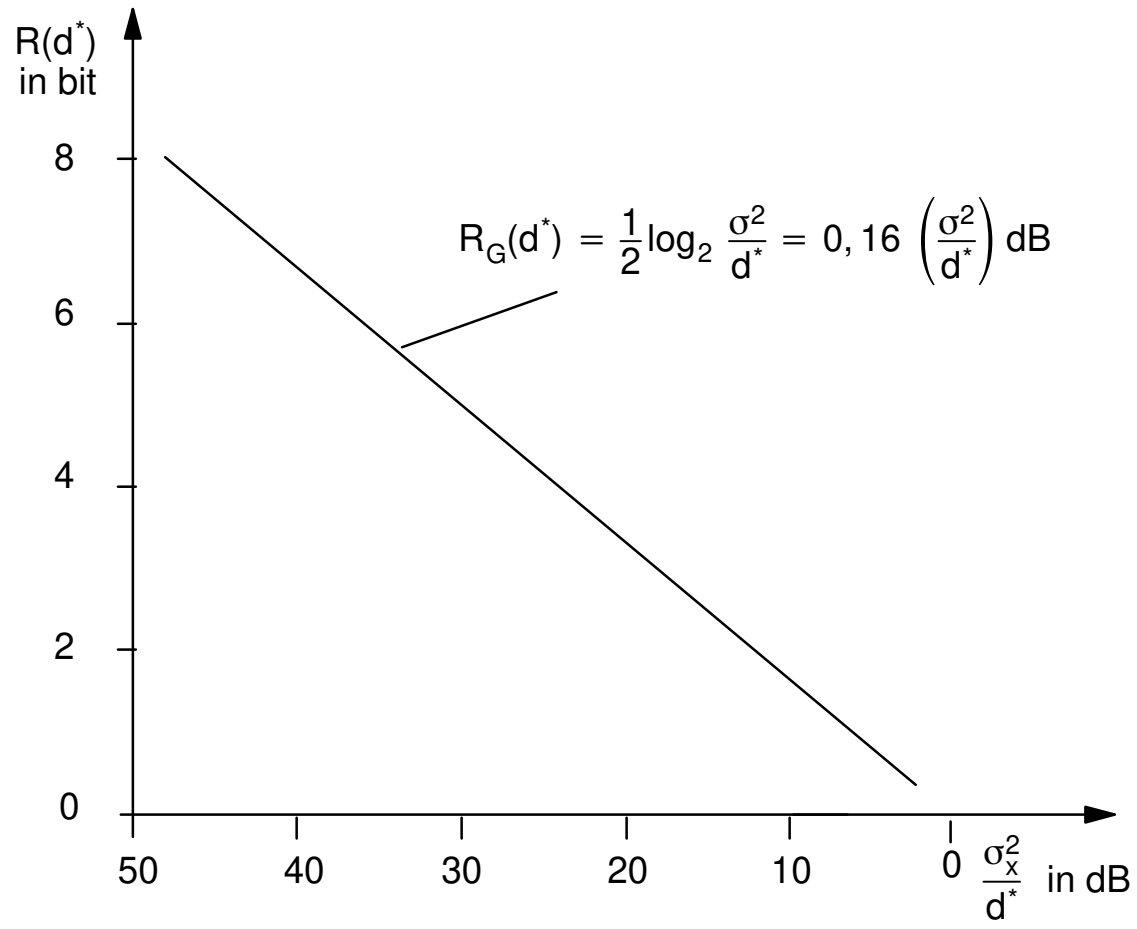


Bild 6.4 $R_G(d^*)$ für eine gedächtnislose Gaußsche Quelle mit quadratischer Abweichung als Verzerrungsmaß

Bei nicht Gaußschen gedächtnislosen Quellen mit gegebener Varianz σ_x^2 ist $H(U)$ kleiner als die Entropie $H_G(U)$ der Gaußschen Quelle. Nach Gl. (6.57) muß die Rate Distortion–Funktion dieser Quelle unter der von Bild 6.4 liegen. Gl. (6.57) stellt somit eine obere Grenze für $R(d^*)$ dar.

Zum Vergleich sei eine zeitdiskrete, gedächtnislose Quelle mit konstanter Wahrscheinlichkeitsdichte und gleicher Varianz σ_x^2 betrachtet. Eine solche Quelle besitzt die Entropie

$$H(U) = \frac{1}{2}\log_2 (12\ \sigma_x^2) \tag{6.59}$$

und liegt damit um die Differenz

$$\frac{1}{2}\log_2 (2\pi e\sigma_x^2) - \frac{1}{2}\log_2 (12\ \sigma_x^2) = 0,25 \text{ bit}$$

unter der Grenze für $R(d^*)$ gegeben durch die Gaußsche Quelle gemäß (6.57). Der relativ geringe Abstand von der Grenze deutet darauf hin, daß auch die übrigen Quellen gleicher Varianz relativ nahe an der Grenze liegen.

Bei 2W Abtastwerten je Sekunde einer kontinuierlichen, bandbegrenzten Gaußschen Quelle mit den Bandgrenzen ergibt sich aus Gl. (6.57) eine Übertragungsbitrate von

$$R(d^*) = W \cdot \log_2 \frac{\sigma_x^2}{d^*} \tag{6.60}$$

in der Einheit bit/s.

Berechnung der Rate–Distortion–Funktion einer Gaußschen Quelle mit dem Leistungsdichtespektrum $S(f)$

Für die Behandlung von Sprachsignal– und Bildsignalquellen ist das Modell der gedächtnisbehafteten Quelle von großer Bedeutung. Im folgenden soll für gedächtnisbehaftete Quellen, deren Leistungsdichtespektrum bekannt ist, eine obere Grenze für $R(d\,*)$ ermittelt werden. Dazu wird von einer gedächtnisbehafteten Gaußschen Quelle mit der spektralen Leistungsdichte $S(f)$ ausgegangen. Diese gedächtnisbehaftete Gaußsche Quelle kann man in eine Summe von gedächtnislosen Gaußschen Quellen mit jeweils weißem Spektrum und der Varianz $S(f)\,df$, wie in Bild 6.5 gezeigt zerlegen.

Ist $N(f)\,df$ die Verzerrung der einzelnen gedächtnislosen Gaußschen Quelle, so kann man mit Hilfe von Gl. (6.57) unter der Voraussetzung $S(f) > N(f)$ den Mittelwert der Übertragungsrate für die gedächtnislosen Gaußschen Quellen ermitteln.

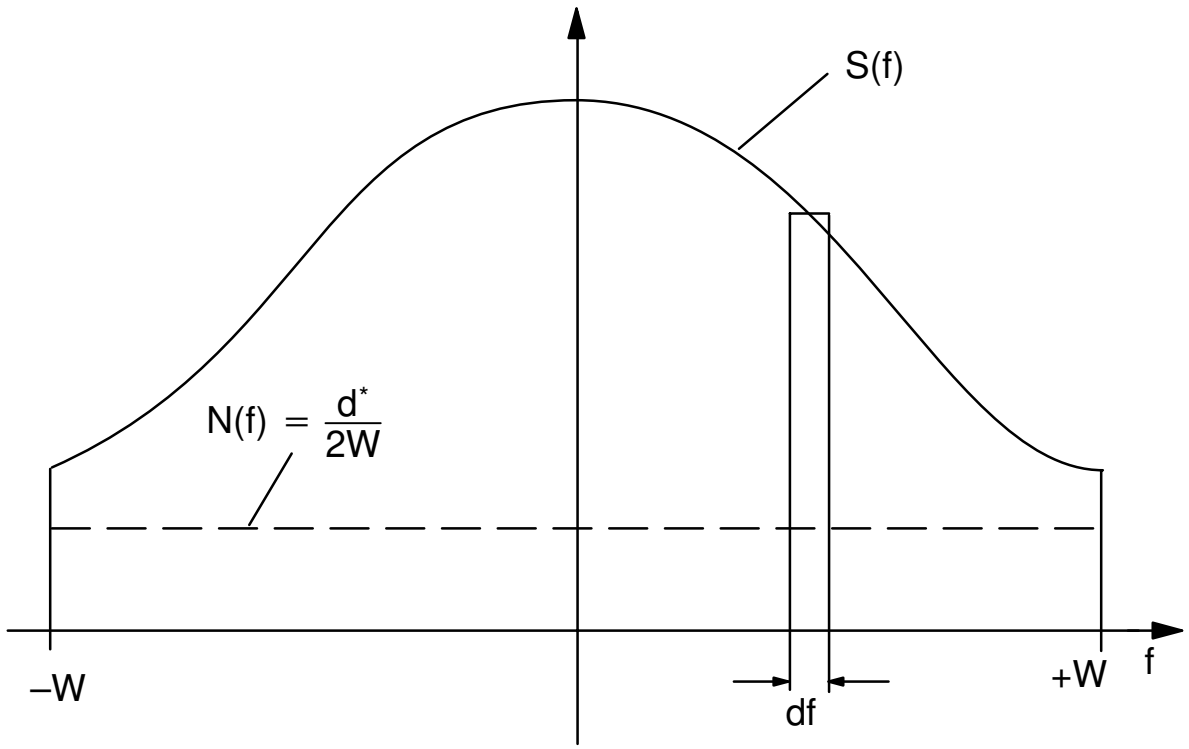


Bild 6.5 Zerlegung einer gedächtnisbehafteten Gaußschen Quelle in eine Summe gedächtnisloser Gaußscher Quellen

$$R(d^*) = \frac{1}{2W} \int_{-W}^{+W} \frac{1}{2} \log_2 \frac{S(f)\, df}{N(f)\, df}\, df$$

$$R(d^*) = \frac{1}{2W} \int_{-W}^{+W} \frac{1}{2} \log_2 \frac{S(f)}{N(f)}\, df \tag{6.61}$$

in bit je Quellensymbol bzw. Abtastwert. Zur Einhaltung des Gütekriteriums muß dabei

$$\int_{-W}^{+W} N(f)\, df = d^* \tag{6.62}$$

sein. Ferner muß die Verzerrung $N(f)$ eine konstante Leistungsdichte aufweisen, d.h. es muß jede einzelne Quelle gleich viel Verzerrung zur Gesamtverzerrung

beitragen, damit bei vorgegebener Verzerrung d^* die Übertragungsrate $R(d^*)$ das Minimum entsprechend Gl. (6.42) erreicht.

Beweis:

Die Minimierung von $R(d^*)$ als Funktion von $N(f)$ unter der Nebenbedingung (6.62) führt auf eine Variationsrechnung und Minimierung von

$$\int_{-W}^{+W} \left[\log_2 \frac{S(f)}{N(f)} + \lambda \cdot N(f) \right] df \,. \tag{6.63}$$

Die Bedingung dafür ist

$$-\frac{1}{N(f)} \log_2 e + \lambda = 0,$$

d.h. $N(f)$ ist eine Konstante gemäß

$$d^* = 2W \cdot N(f) = 2W \cdot \Theta \,. \tag{6.64}$$

Damit erhält man für $R(d\,^*)$

$$R(d^*) = \frac{1}{2W} \int_{-W}^{+W} \frac{1}{2} \log_2 \frac{S(f) \cdot 2W}{d^*} \, df. \tag{6.65}$$

Die Rate–Distortion–Funktion gemäß (6.65) stellt für nicht Gaußsche, gedächtnisbehaftete Quellen mit gleicher spektraler Leistungsdichte eine obere Grenze für $R(d^*)$ dar.–

Beispiel: Berechnung der oberen Grenze der Rate–Distortion–Funktion für Fernsprech– und Fernsehsignale. Die Leistungsdichtespektren von Fernsprech– und –Fernsehluminanzsignalen können durch die Näherung

$$S(f) = A \left[1 + \left(\frac{f}{f_0} \right)^n \right]^{-1} \tag{6.66}$$

approximiert werden, vergl. dazu Bild 6.6, wobei im Fall von Fernsprechsignalen die Bandbreite $W = 3.4\ kHz$ mit den Parametern $f_0 = 0,5\ kHz$, $n = 4$ und im Fall von Fernsehsignalen die Bandbreite $W = 5\ MHz$ und die Parameter $f_0 = 0,015\ MHz$, $n = 2$, angenommen werden.–

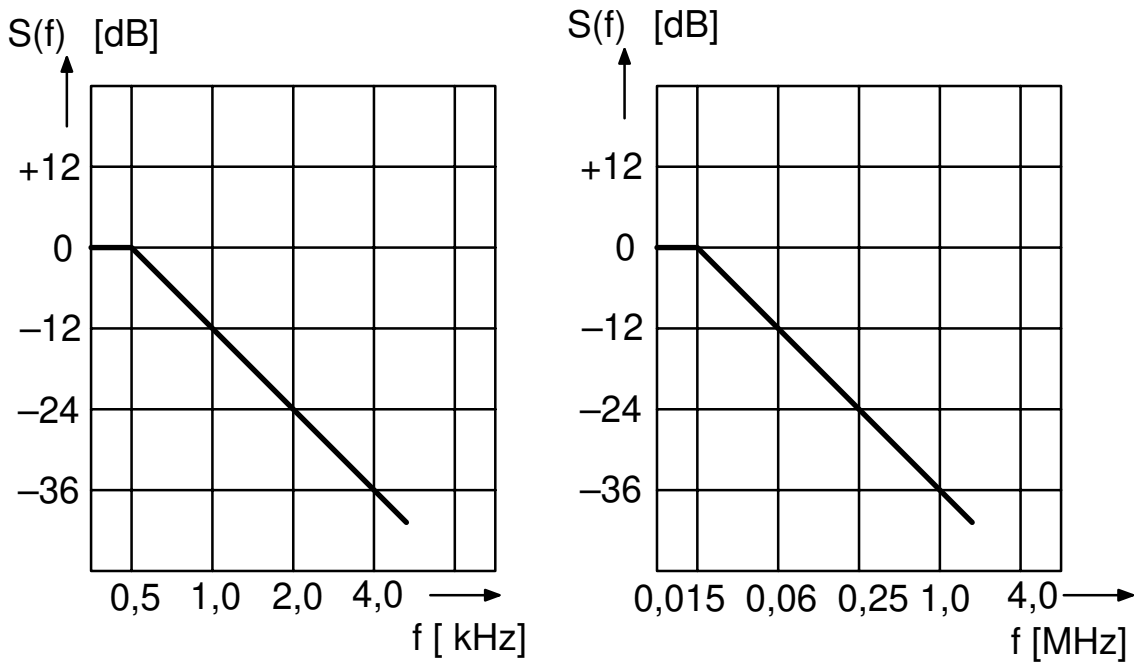


Bild 6.6 Leistungsdichtespektrum von Fernsprechsignalen (a) und Fernsehluminanzsignalen (b)

Die Auswertung der Gleichungen (6.65) und (6.64) liefert mit

$$\int_{-W}^{+W} S(f)\, df = \sigma_x^2 \tag{6.67}$$

für Fernsprechsignale

$$R_S(d^*) = \frac{1}{2}\log_2\frac{\sigma_x^2}{d^*} - 1,8 \text{ bit} \tag{6.68}$$

und für Fernsehsignale

$$R_{TV}(d^*) = \frac{1}{2}\log_2\frac{\sigma_x^2}{d^*} - 3,0 \text{ bit.} \tag{6.69}$$

Diese Rate–Distortion–Funktionen sind in Bild 6.7 im Vergleich zur Grenze gedächtnisloser Quellen dargestellt. Dabei ist das Signal–Rauschleistungs–Verhältnis unter Verwendung von Gl. (6.58) in die Einheit ”Dezibel” umgerechnet worden.

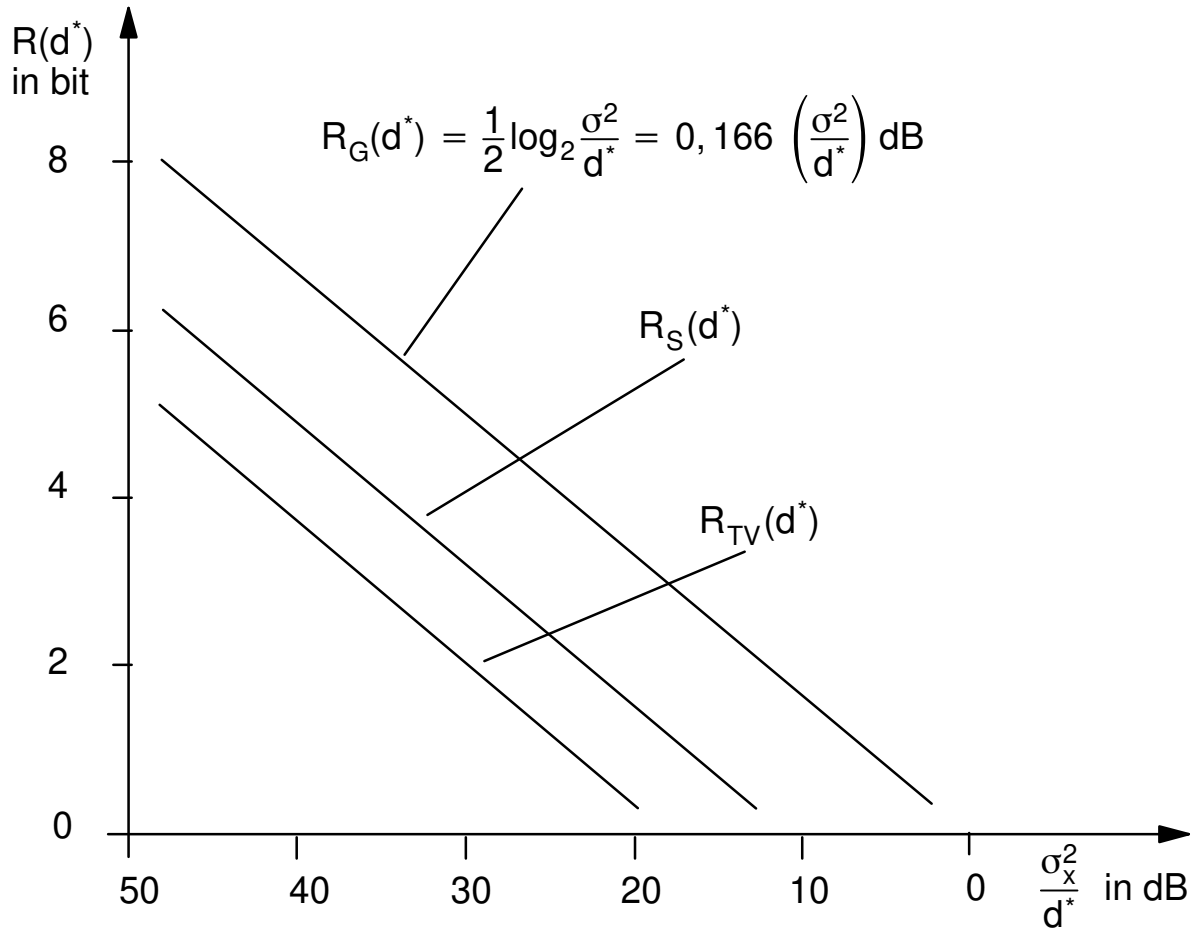


Bild 6.7 Obere Grenze der Übertragungsraten für eine gedächtnislose Quelle $R(d^*)$ sowie für Fernsprech– und Fernsehsignale $R_S(d^*)$, $R_{TV}(d^*)$

Im Vergleich zur gedächtnislosen Quelle ergibt sich für gedächtnisbehaftete Quellen eine geringere obere Grenze, die vom Verlauf der spektralen Leistungsdichte abhängt. Diese Verminderung der Übertragungsrate ist auf im Signal enthaltene Korrelation zurückzuführen, denn ein nicht weißes Leistungsdichtespektrum bedeutet gemäß dem Wiener–Khintchineschen Theorem, daß die Abtastwerte korreliert sind. Mit zunehmend weißer spektraler Leistungsdichte vermindert sich die Korrelation und $R(d^*)$ nähert sich der Grenze der Übertragungsrate für gedächtnislose Quellen. Es gilt

$$R(d^*) = \frac{1}{2W}\int_{-W}^{+W}\frac{1}{2}\log_2\frac{S(f)\cdot 2W}{d^*}\,df = \frac{1}{2}\log_2\frac{S(f)\cdot 2W}{d^*} = \frac{1}{2}\log_2\frac{\sigma_x^2}{d^*}, \qquad (6.70)$$

da bei konstanter Leistungsdichte

$$S(f) \cdot 2W = \sigma_x^2 \text{ ist.}$$

Die Verzerrung d^* wird im Codierungssystem durch den Quantisierungsfehler q bzw. dessen Varianz σ_q^2 bestimmt. Dabei gilt für eine gleichförmige Quantisierung mit der Stufenbreite Δx näherungsweise

$$d^* = 2W \cdot \Theta = \sigma_q^2 \approx \frac{\Delta x^2}{12} \tag{6.71}$$

Bei Quantisierung mit geringer Stufenzahl K, d.h. mit großer Stufenbreite Δx kann Θ größer als $S(f)$ werden (vergleiche Bild 6.8). In diesen Frequenzbereichen wird das Ausgangssignal $v = 0$ gesetzt. Auf diese Weise entsteht eine Verzerrung, die der spektralen Leistungsdichte $S(f)$ entspricht. Diese Frequenzbereiche tragen nicht zur Übertragungsrate bei, d.h.

$$d^* = \int\limits_{f:S(f)>\Theta} \Theta \, df + \int\limits_{f:S(f)\le\Theta} S(f) \, df \,, \tag{6.72}$$

$$R(d^*) = \frac{1}{2W} \int\limits_{f:S(f)>\Theta} \frac{1}{2} \log_2 \frac{S(f)}{\Theta} \, df \tag{6.73}$$

in bit je Abtastwert.

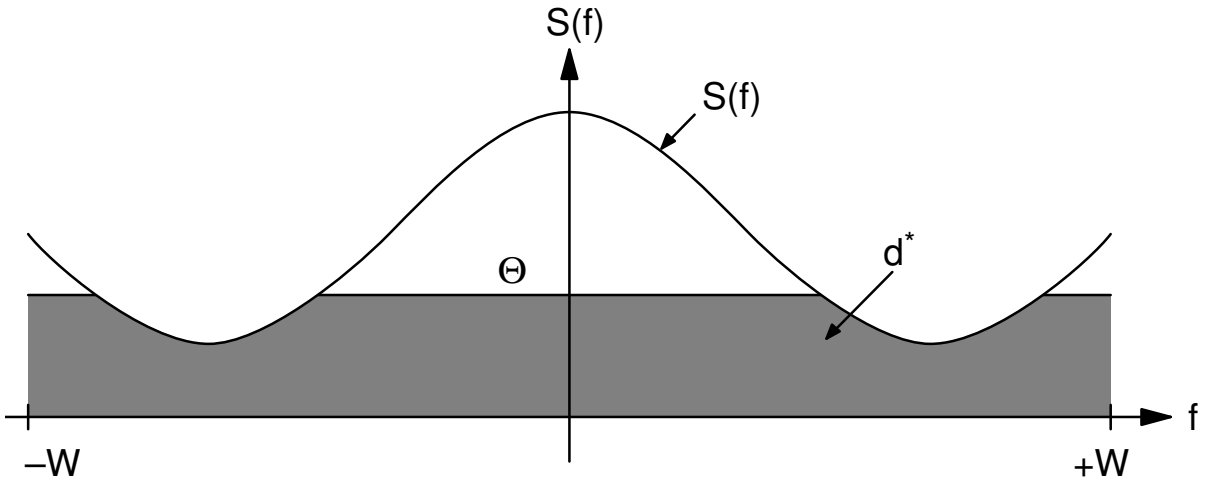


Bild 6.8 Grafische Darstellung der Integrale d^* und $R(d^*)$

6.4 Anwendungsbeispiele

6.4.1 Amplitudenquantisierung mit minimaler Fehlervarianz

6.4.1.1 Quantisierung mit großer Stufenzahl K

Eine zeitdiskrete, wertkontinuierliche Quelle X soll durch Quantisierung der Quellensymbole x in eine diskrete Quelle umgewandelt werden. Die Quellensymbole mögen einen begrenzten kontinuierlichen Wertevorrat $-1 \leq x \leq +1$ und eine Wahrscheinlichkeitsdichte $p(x)$ aufweisen. Dieser kontinuierliche Wertevorrat soll in K Quantisierungsstufen der Breite Δ_k mit einem jeweiligen Repräsentativwert x_k quantisiert werden. Es sei angenommen, daß die Quantisierung so fein ist, daß $p(x)$ als näherungsweise konstant im Intervall Δ_k angesehen werden kann, und daß der Repräsentativwert x_k jeweils in der Mitte des Intervalls Δ_k liegt. Die Quantisierung soll nun so gewählt werden, daß der mittlere quadratische Quantisierungsfehler $E[q^2]$ als Gütekriterium ein Minimum erreicht.

Der mittlere quadratische Quantisierungsfehler der k–ten Quantisierungsstufe beträgt

$$E_k[q^2] = \int_{x_k-\frac{1}{2}\Delta_k}^{x_k+\frac{1}{2}\Delta_k} (x - x_k)^2 \, p(x) \, dx \ , \tag{6.74}$$

wenn jeder in das Intervall Δ_k fallende Wert durch x_k dargestellt wird und somit $q = x - x_k$ der Quantisierungsfehler ist.

Mit der Annahme

$$p(x) \simeq p(x_k) \ \text{ für } \ x_k - \frac{1}{2} \cdot \Delta_k \leq x \leq x_k + \frac{1}{2} \cdot \Delta_k \tag{6.75}$$

wird $p(x)$ durch eine Treppenkurve dargestellt und es folgt

$$E_k[q^2] = \int_{x_k-\frac{1}{2}\Delta_k}^{x_k+\frac{1}{2}\Delta_k} (x - x_k)^2 \, dx = \frac{1}{12} \, p(x_k) \, \Delta_k{}^3 \ . \tag{6.76}$$

Die Wahrscheinlichkeit, daß ein Wert x in das Intervall Δ_k fällt, ist

$$P_k = p(x_k) \cdot \Delta_k \tag{6.77}$$

und für (6.76) folgt

$$E_k[q^2] = \frac{1}{12} P_k \Delta_k^{\,2} , \tag{6.78}$$

sowie für den gesamten mittleren quadratischen Quantisierungsfehler $E[q^2]$

$$E[q^2] = \sum_{k=1}^{K} E_k[q^2] = \frac{1}{12} \sum_{k=1}^{K} P_k \Delta_k^{\,2} . \tag{6.79}$$

Zur Minimierung von $E[q^2]$ muß im allgemeinen eine Quantisierung mit ungleich breiten Quantisierungsstufen angewendet werden. Eine solche ungleichförmige Quantisierung wird zur Veranschaulichung durch eine sog. Kompressorkennlinie dargestellt.

Im Diagramm des Bildes 6.9 wird die ungleichförmige Quantisierung der Eingangswerte x über die Kompressorkennlinie $y = f(x)$ auf eine gleichförmige Quantisierung der Ausgangswerte y abgebildet. Die ungleichförmige Quantisierung kann demgemäß gedanklich mit Hilfe eines Kompressors der Übertragungsfunktion $y = f(x)$ und einem gleichförmigen Quantisierer realisiert werden. Zur Rückwandlung wird dann ein Expander mit einer zur Kompression inversen Kennlinie erforderlich. Das Blockschaltbild eines solchen Systems ist in Bild 6.10 dargestellt.

Ersetzt man die Kompressorkennlinie durch eine Segmentkennlinie der Sehnen, wie in Bild 6.9 gezeigt, so gilt

$$\left(\frac{dx}{dy}\right)_{x=x_k} = \frac{\Delta_k}{2} K . \tag{6.80}$$

Damit erhält man für $E[q^2]$ die Form

$$E[q^2] = \frac{1}{12} \sum_{k=1}^{K} P_k \Delta_k^{\,2} = \frac{1}{3K^2} \sum_{k=1}^{K} P_k \left(\frac{dx}{dy}\right)_{x=x_k}^{2} . \tag{6.81}$$

Bei hinreichend feiner Quantisierung kann die Summe durch das Integral

$$E[q^2] = \frac{1}{3K^2} \int_{-1}^{+1} p(x) \left(\frac{dx}{dy}\right)^2 dx \tag{6.82}$$

dargestellt werden.

Die folgende Minimierung des mittleren quadratischen Quantisierungsfehlers $E[q^2]$ ist ein Problem der Variationsrechnung. Gesucht wird die Extremale $y' = dy/dx$ mit der Nebenbedingung

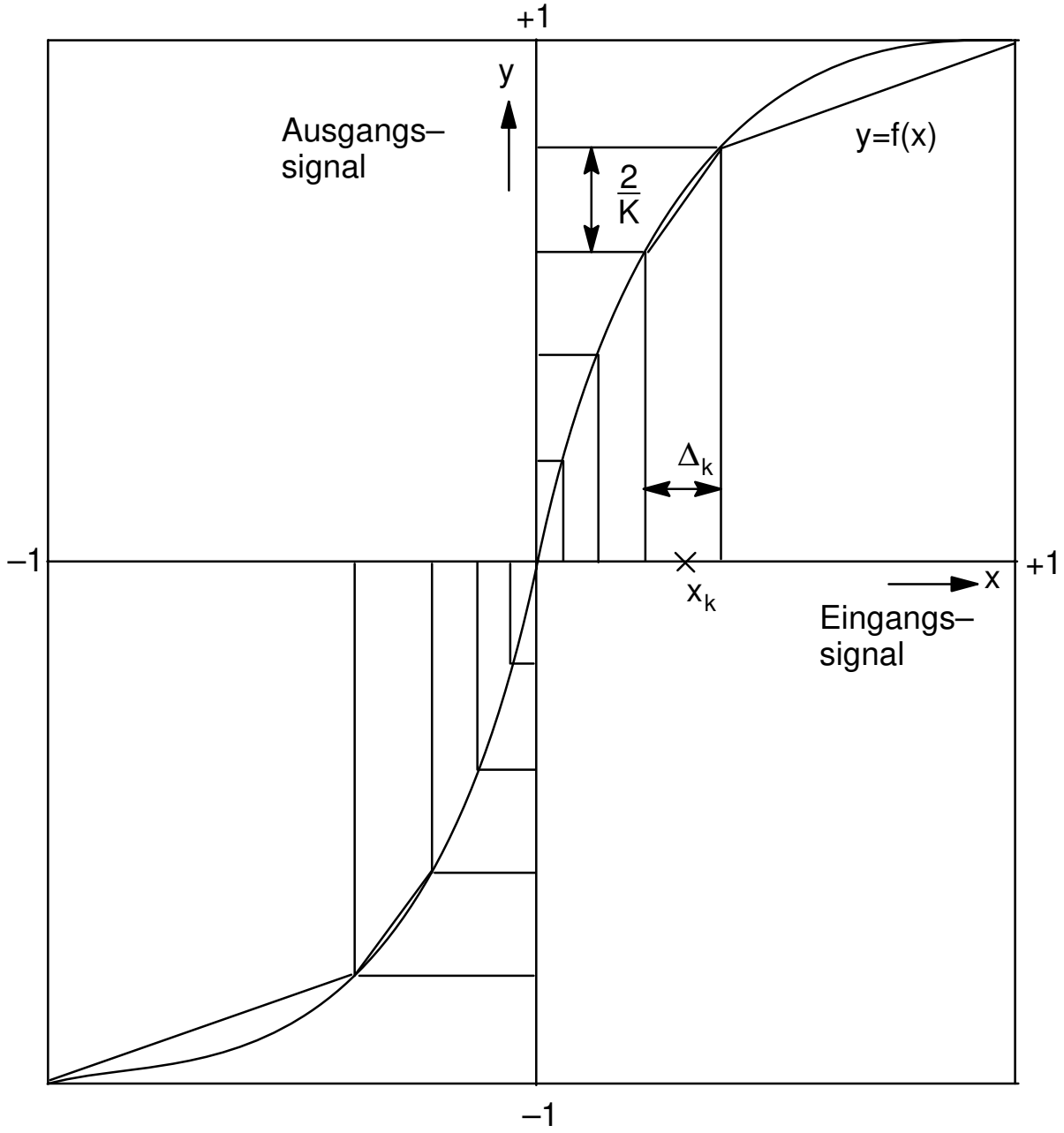


Bild 6.9 Beispiel einer Kompressorkennlinie

$$\int_{-1}^{+1} y' \, dx = 2 \, . \tag{6.83}$$

Die Variationsrechnung führt auf die Beziehung

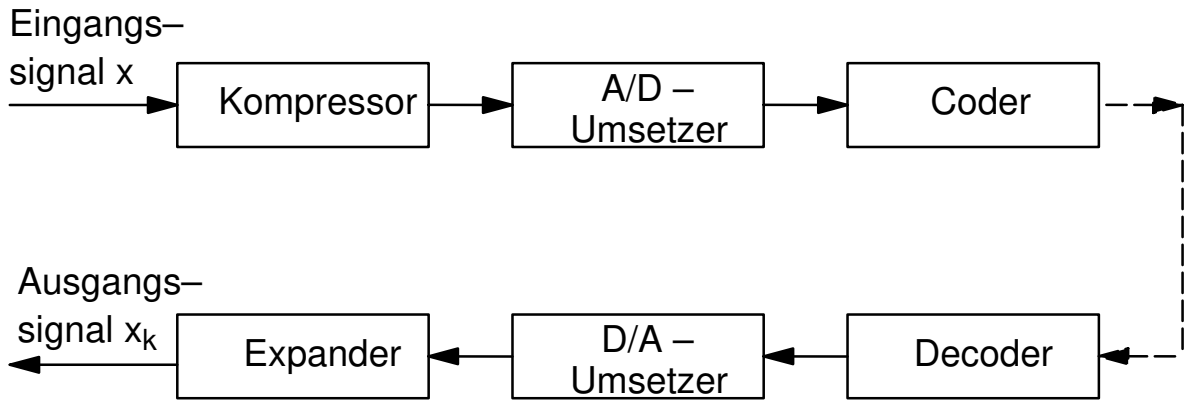


Bild 6.10 Übertragungssystem mit Kompandierung

$$\frac{d}{dy'}\left[p(x)\cdot y'^{-2}+\lambda\cdot y'\right] = -2\cdot p(x)\cdot y'^{-3}+\lambda = 0$$

$$y'_{opt} = \left(\frac{2}{\lambda}\right)^{1/3} p(x)^{1/3} \quad , \tag{6.84}$$

wobei der Lagrangesche Multiplikator λ mit Hilfe der Nebenbedingung (6.83) wie folgt bestimmt werden kann:

$$\left(\frac{2}{\lambda}\right)^{1/3} = \frac{2}{\int_{-1}^{+1} p(x)^{1/3}\,dx} = \frac{1}{C} \; . \tag{6.85}$$

Dabei ist C eine Konstante, die durch die Wahrscheinlichkeitsdichte der Quellensymbole bestimmt ist. Die gesuchte Kompressorkennlinie ergibt sich mit (6.84) und (6.85) zu

$$y'_{opt} = \frac{1}{C}\cdot p(x)^{1/3} \tag{6.86}$$

$$y_{opt} = \frac{1}{C}\cdot \int_{0}^{x} p(x)^{1/3}\,dx \; . \tag{6.87}$$

Der minimale mittlere quadratische Quantisierungsfehler beträgt mit (6.82) und (6.86)

$$E[q^2]_{\min} = \frac{1}{3K^2} C^2 \int_{-1}^{+1} p(x)^{1/3}\, dx = \frac{2}{3}\, \frac{C^3}{K^2} \ . \tag{6.88}$$

Bei optimaler Kompressorkennlinie zeigt sich, daß der mittlere quadratische Quantisierungsfehler der einzelnen Quantisierungsstufen (6.76) gleich groß ist.

$$E_k[q^2] = \frac{1}{12} p(x_k)\, \Delta_k{}^3 = \frac{1}{12} p(x_k)\, \frac{8}{K^3} \left(\frac{dx}{dy}\right)^3_{x=x_k}$$

$$E_k[q^2] = \frac{2}{3} \frac{C^3}{K^3} \ . \tag{6.89}$$

Im Sonderfall konstanter Wahrscheinlichkeitsdichte $p(x)$ folgt aus Gl. (6.87), daß die Kompressorkennlinie eine Gerade ist. Somit ergibt sich eine gleichförmige Quantisierung der Eingangswerte x. Alle Quantisierungsstufen sind gleich groß, $\Delta_k = \Delta = 2/K$. Aus der Beziehung (6.79) folgt

$$E_k[q^2] = \frac{1}{12}\, \Delta_k{}^2 \cdot \sum_{k=1}^{K} P_k = \frac{1}{12}\, \Delta_k{}^2 = \frac{1}{3K^2} \ . \tag{6.90}$$

Diese gleichförmige Quantisierung hat die besondere Eigenschaft, daß der mittlere quadratische Quantisierungsfehler auch bei nicht konstanter Wahrscheinlichkeitsdichte $p(x)$ des Eingangssignals stets den gleichen Wert liefert.

Die hier durchgeführte Berechnung des minimalen Quantisierungsfehlers geht auf Panter und Dite (1951), [6] zurück und beruht auf den eingangs erwähnten einschränkenden Annahmen, daß $p(x) \simeq p(x_k)$ im Intervall Δ_k und daß x_k in der Mitte von Δ_k liegt.

6.4.1.2 Quantisierung mit beliebiger Stufenzahl K

J. Max [5] gelang es, den Quantisierungsfehler auch ohne die einschränkenden Annahmen aus Kap. 6.4.1.1 zu berechnen.

Gegeben: $p(x)$ mit $-\infty \le x_n \le +\infty$

Quantisierer mit K Stufen, wobei

$$d_k \le x_k \le d_{k+1}$$

sowie $d_1 = -\infty$, $d_{K+1} = +\infty$

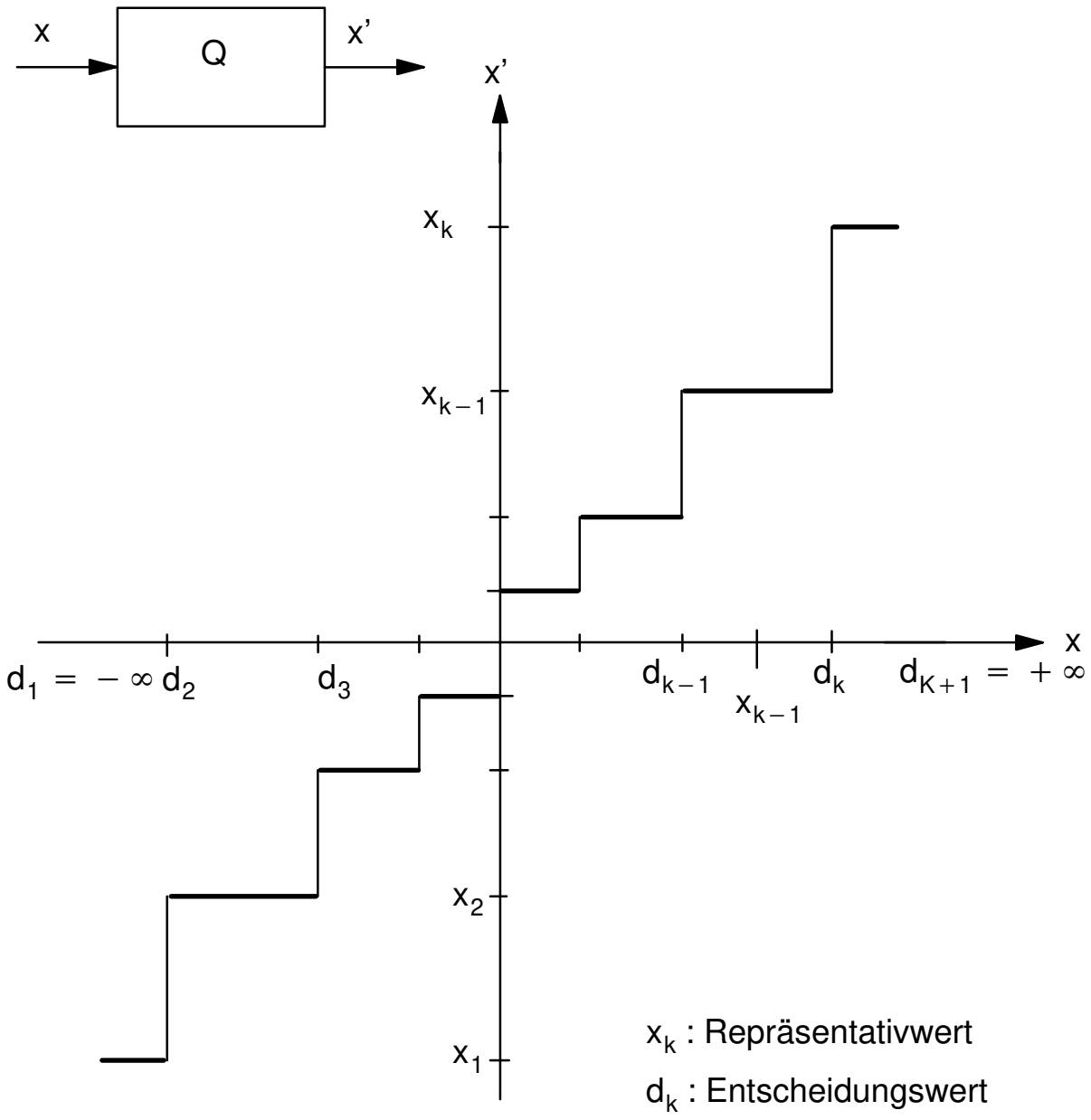


Bild 6.11 Charakteristik eines Quantisierers

$$D = \sum_{k=1}^{K} \int_{d_k}^{d_{k+1}} f(x - x_k)\, p(x)\, dx \tag{6.91}$$

Gesucht: d_k , x_k so daß $D \rightarrow \min$

Ansatz: Extremwertproblem

$$\frac{\partial D}{\partial d_k} = f(d_k - x_{k-1})\ p\,(d_k) - f(d_k - x_k)\ p\,(d_k) = 0$$

$$\text{für} \quad k = 2,\ ...,\ K \tag{6.92}$$

$$\frac{\partial D}{\partial x_k} = -\int_{d_k}^{d_{k+1}} f'(x - x_k)\ p\,(x)\,dx = 0$$

$$\text{für} \quad k = 1,\ ...,\ K \tag{6.93}$$

Annahme: $f(x - x_k) = (x - x_k)^2$

d.h. Gütekriterium d ist der mittlere quadratische Fehler.

Dann folgt aus Gl. (6.92)

$$(d_k - x_{k-1})^2 = (d_k - x_k)^2$$

mit $d_k < x_k$

$$d_k - x_{k-1} = -(d_k - x_k)$$

$$d_k = \frac{x_k + x_{k-1}}{2} \tag{6.94}$$

oder $x_k = 2d_k - x_{k-1}$

für $k = 2,\ ...,\ K$

sowie aus Gl. (6.93)

$$\int_{d_k}^{d_{k+1}} (x - x_k)\,p(x)\,dx = 0$$

$$x_k = \frac{\displaystyle\int_{d_k}^{d_{k+1}} x\,p(x)\,dx}{\displaystyle\int_{d_k}^{d_{k+1}} p(x)\,dx} \tag{6.95}$$

für $k = 1,\ ...,\ K$

1. Lösungsansatz: Iterative Lösung

1. Wähle x_1 und berechne d_2 mit Gl. (6.95)

$$\int_{d_1=-\infty}^{d_2} (x - x_1) \cdot p(x)\, dx = 0\,, \tag{6.96}$$

   so daß x_1 Schwerpunkt der Fläche von p(x) zwischen d_1 und d_2 ist.
2. Unter Verwendung von x_1 und d_2 berechne x_2 mit Gl. (6.94).

$$x_2 = 2d_2 - x_1\,, \tag{6.97}$$

   usw.
3. Überprüfe, ob x_K Schwerpunkt zwischen d_K und $+\infty$ ist, andernfalls muß mit einem anderen x_1 die Iteration wiederholt werden.

2. Lösungsansatz: LBG–Algorithmus

Linde Y., Buzo, A. and Gray, R.M. "An Algorithm for VQ Design", IEEE Transactions on Communications, COM–28 (1980), No. 1, January, pp. 84–95.

Beispiel: $p(x) = p_G(x) = \dfrac{1}{\sigma_x\sqrt{2\pi}}\, e^{-\frac{x^2}{2\sigma x^2}}\,,\quad \sigma_x^2 = 1$

Quantisierung mit $K = 2$ Stufen, $R = 1$ bit
Bei symmetrischer Wahrscheinlichkeitsdichte wird
$d_1 = -\infty\,,\ d_2 = 0\,,\ d_3 = +\infty$
$x_1 = -x_2$

Gesucht: x_1

Aufgrund von Gl. (6.95) wird

$$x_1 = \frac{\displaystyle\int_{-\infty}^{0} x\, p(x)\, dx}{\displaystyle\int_{-\infty}^{0} p(x)\, dx} = 2 \int_{-\infty}^{0} x\, p(x)\, dx \tag{6.98}$$

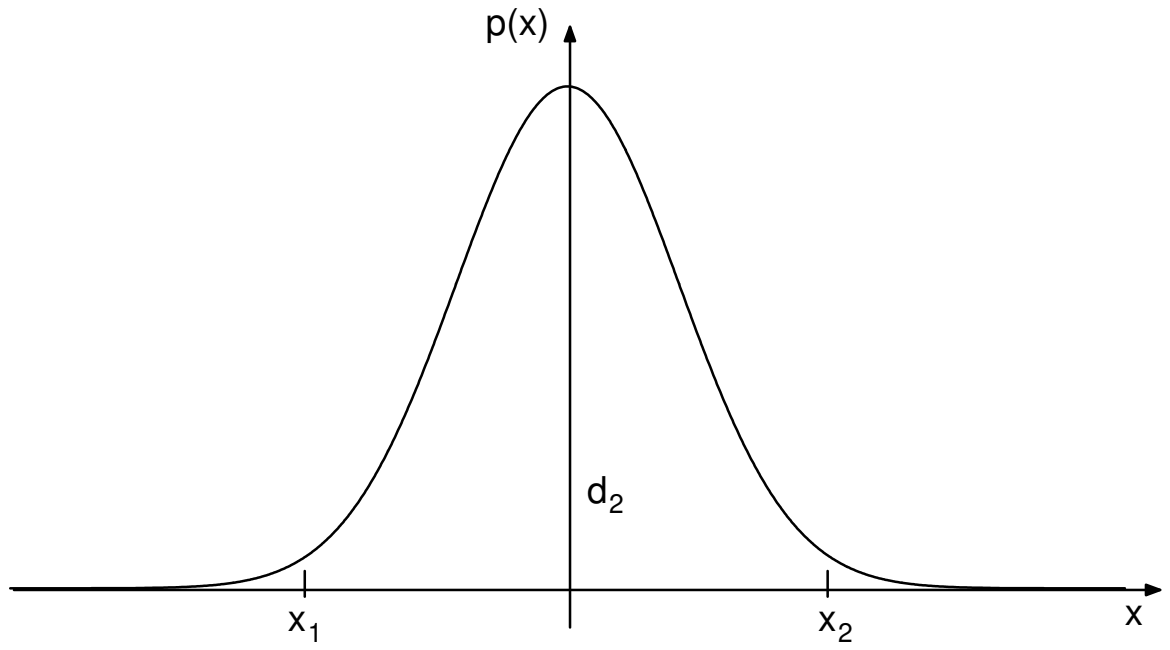


Bild 6.12 Normierte Gaußsche Wahrscheinlichkeitsdichte

$$x_1 = -0,798 \ , \ D = 0,363$$

$$SNR = 10 \log_{10} \frac{\sigma_x^2}{D} = 4,40 \text{ dB} \ .–$$

6.4.2 Quantisierung mit minimaler Ausgangsentropie

Die Entropie $H_Q(X)$ der Repräsentativwerte x_k eines beliebigen Quantisierers beträgt

$$\begin{aligned} H_Q(X) &= -\sum_{k=1}^{K} p(x_k)\ \Delta_k\ \log_2\big(p(x_k)\ \Delta_k\big) \\ &= -\int_{-\infty}^{+\infty} p(x)\ \log_2\ p(x)\, dx - \sum_{k=1}^{K} p(x_k)\, \Delta_k \cdot \log_2\ \Delta_k \\ &= H(X) - \sum_{k=1}^{K} p(x_k)\, \Delta_k \cdot \log_2\ \Delta_k \, , \end{aligned} \tag{6.99}$$

wobei Δ_k die Stufenbreite der k–ten Quantisiererstufe ist. Der mittlere quadratische Quantisierungsfehler q dieses Quantisierers ergibt sich bei einer Quantisierung mit hinreichend großer Stufenzahl K zu

$$E_Q[q^2] = \sum_{k=1}^{K} p(x_k) \cdot \Delta_k \cdot \frac{\Delta_k^2}{12} \ . \tag{6.100}$$

Zum Vergleich wird ein Quantisierer mit gleichförmiger Quantisierung betrachtet. Mit

$$\Delta_k = \Delta \qquad \text{für alle } k \tag{6.101}$$

erhält man für die gleichförmige Quantisierung

$$H_{GQ}(X) = H(X) - \log_2 \Delta \tag{6.102}$$

$$E_{GQ}[q^2] = \frac{\Delta x^2}{12} \ . \tag{6.103}$$

Es soll nun gezeigt werden, daß der gleichförmige Quantisierer bei gleichem mittleren Quantisierungsfehler die geringere Entropie hat. Dazu wird die Differenz der Entropie berechnet

$$\begin{aligned} H_{GQ}(X) - H_Q(X) &= \sum_{k=1}^{K} p(x_k)\Delta_k \cdot \log_2 \frac{\Delta_k}{\Delta} \\ &= \frac{1}{2} \sum_{k=1}^{K} p(x_k)\Delta_k \log_2 \left(\frac{\Delta_k}{\Delta} \right)^2 \ . \end{aligned} \tag{6.104}$$

Mit der Hilfsbeziehung $\ln z \leq z-1$ für $z > 0$ und $\ln z = z-1$ für $z = 1$ sowie $\log_2 z = \ln z \cdot \log_2 e$ wird

$$\begin{aligned} H_{GQ}(X) - H_Q(X) &\leq \frac{1}{2} \log_2 e \sum_{k=1}^{K} p(x_k)\, \Delta_k \left[\left(\frac{\Delta_k}{\Delta} \right)^2 - 1 \right] \\ &\leq \frac{1}{2} \log_2 e \left[\frac{E_Q[q^2]}{E_{GQ}[q^2]} - 1 \right] \ . \end{aligned} \tag{6.105}$$

Daraus folgt für

$$E_Q[q^2] \leq E_{GQ}[q^2] \ , \tag{6.106}$$

daß

$$H_{GQ}(X) - H_Q(X) \leq 0 \,. \tag{6.107}$$

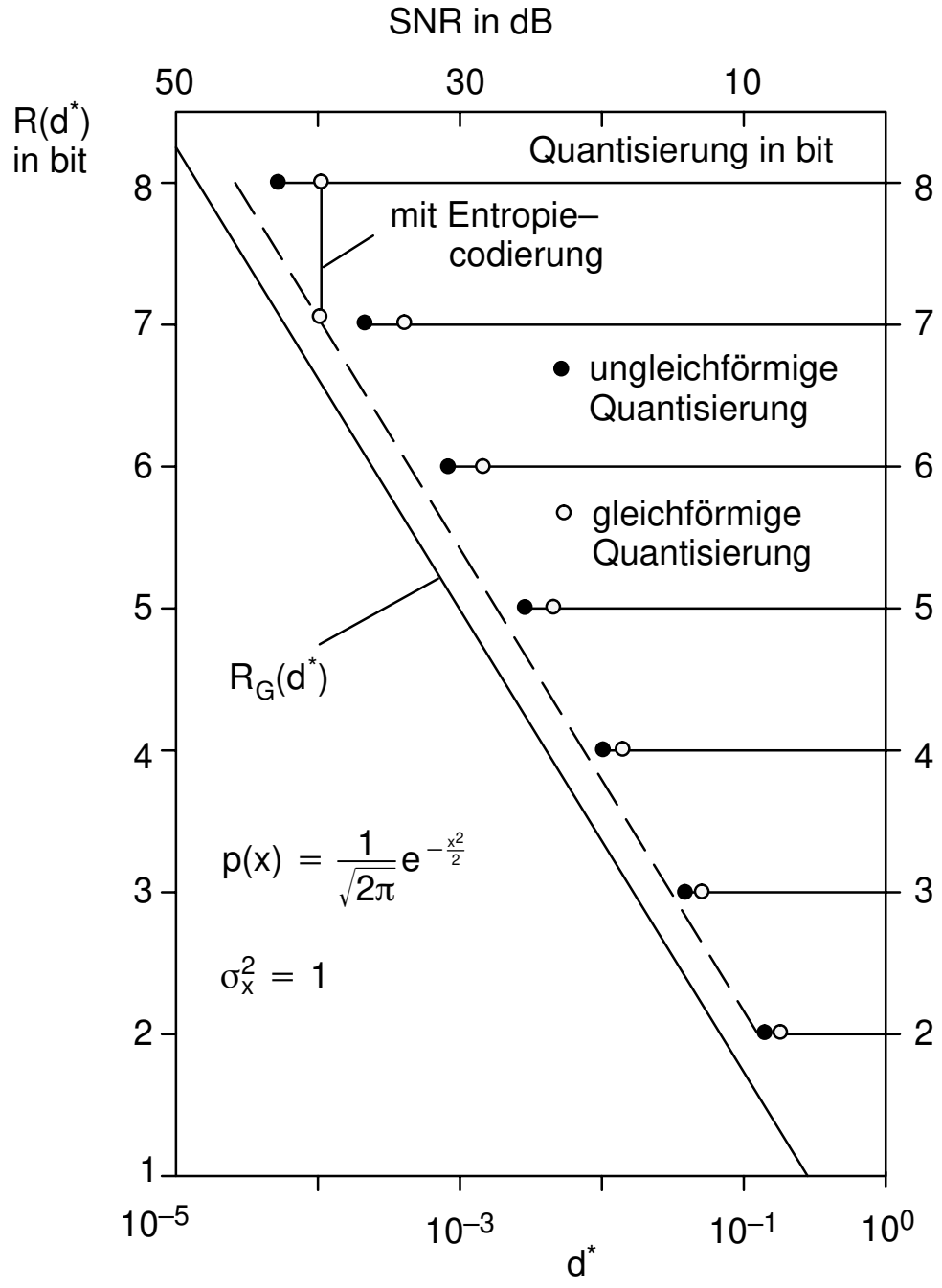


Bild 6.13 Übertragungsbitrate als Funktion des mittleren Quantisierungsfehlers d^* verschiedener Quantisierer im Vergleich zur Rate–Distortion–Funktion (RDF) der gedächtnislosen Gaußschen Quelle

Demzufolge gibt es unter den Quantisierern mit einem mittleren quadratischen Quantisierungsfehler, der kleiner oder gleich dem des gleichförmigen Quantisierers ist, keinen, dessen Entropie kleiner als die des gleichförmigen Quantisierers ist.

Bild 6.13 zeigt die Übertragungsbitrate als Funktion des mittleren Quantisierungsfehlers eines nach J. Max, [5] berechneten Quantisierers im Vergleich zu der eines gleichförmigen Quantisierers und zur Rate–Distortion–Funktion der gedächtnislosen Gaußschen Quelle.

Die mit gleichförmiger Quantisierung und anschließender redundanzmindernder oder auch sog. Entropie–Codierung erreichbaren Übertragungsbitraten müssen um 0,25 bit über der Rate–Distortion–Funktion liegen, da der Quantisierungsfehler näherungsweise eine konstante Wahrscheinlichkeitsdichte anstelle einer Gaußschen Wahrscheinlichkeitsdichte aufweist.

6.4.3 Pulscodemodulation (PCM)

Mit Hilfe der Pulscodemodulation können zeit– und wertkontinuierliche Quellensignale in eine binäre Darstellungsform umgewandelt werden, um beispielsweise über binäre Kanäle übertragen werden zu können. Die Pulscodemodulation kann somit als eine spezielle Quellencodierung angesehen werden. Ein PCM–Coder bzw. PCM–Decoder besteht dabei aus den in Bild 6.14 gezeigten Komponenten.

Das kontinuierliche Eingangssignal $x(t)$ wird mittels eines Tiefpasses der Bandbreite $\pm\ W$ bandbegrenzt und anschließend mit einer Abtastfrequenz $f_A = 2W$ abgetastet, gleichförmig oder ungleichförmig quantisiert und die Repräsentativwerte x_k des Quantisierers in binär codierter Form übertragen.

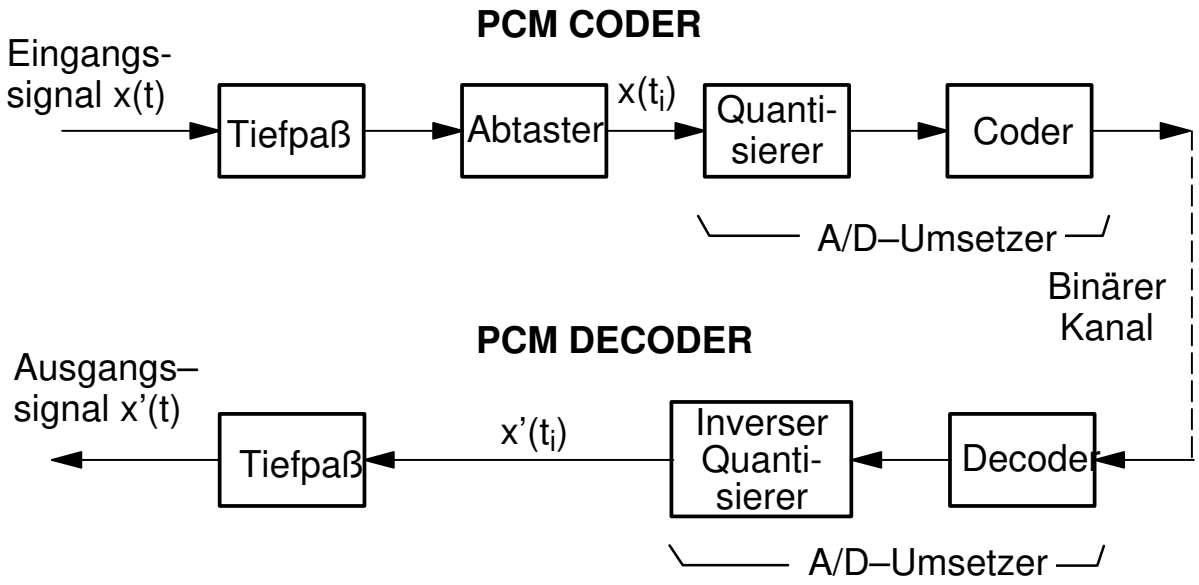


Bild 6.14 Komponenten eines PCM–Coders und –Decoders

Sofern keine besondere redundanzmindernde Codierung angewendet wird, beträgt die Übertragungsbitrate bei K Quantisierungsstufen $\log_2 K$ bit je Abtastwert bzw. $2W \cdot \log_2 K$ bit je Sekunde. Auf der Empfangsseite werden die binären Codeworte decodiert in Dualzahlen der Werte x_k , die der D/A Wandler in quantisierte Abtastwerte entsprechender Amplituden umwandelt. Zur Rückwandlung der quantisierten Abtastwerte in das bandbegrenzte Ausgangssignal ist entsprechend dem Abtasttheorem ein weiterer Tiefpaß der Bandbreite W zur Interpolation erforderlich.

Im folgenden soll die Übertragungsrate einer PCM als Funktion des Signal–Rauschleistungs–Verhältnisses als Gütekriterium berechnet werden und das Ergebnis mit der Rate–Distortion–Funktion verglichen werden. Als Beispiel soll dabei die PCM von Fernsprechsignalen behandelt werden.

Das Signal–Rauschleistungs–Verhältnis des Fernsprechsignals am Ausgang der Übertragungsstrecke wird im wesentlichen durch zwei Faktoren bestimmt, durch das Quantisierungsrauschen der Quellencodierung und durch das auf dem Übertragungskanal infolge von Übertragungsfehlern erzeugte Rauschen. Die Einflüsse des Übertragungskanals sollen hier unberücksichtigt bleiben und die Übertragungsbitrate als Funktion des Signal–Quantisierungs–Rauschleistungs–Verhältnisses bestimmt werden.

Bei der Fernsprechübertragung wird ein Signal–Rauschleistungs–Verhältnis angestrebt, das möglichst unabhängig ist von der Spracheigenschaft des Sprechers und der Zuleitungsdämpfung, d.h. möglichst unabhängig ist von der Wahrscheinlichkeits-

dichte $p(x)$ der Abtastwerte und ihrer Leistung σ_x^2 . Das Signal–Quantisierungs–Rauschleistungs–Verhältnis beträgt mit Gl. (6.82)

$$\frac{P}{E[q^2]} = 3K^2 \frac{\int p(x)\, x^2\, dx}{\int p(x) \left(\frac{dx}{dy}\right)^2 dx} \quad . \tag{6.108}$$

Es ist leicht einzusehen, daß jede Kompressorkennlinie $y = f(x)$, die der Bedingung

$$\frac{dx}{dy} = B \cdot x\,, \qquad \text{B = Konstante} \tag{6.109}$$

genügt, zu einem Signal–Rauschleistungs–Verhältnis gemäß

$$\frac{P}{E} = \frac{3K^2}{B^2} \tag{6.110}$$

führt, das die geforderte Eigenschaft besitzt. Die zugehörige Kompressorkennlinie erhält man durch Lösung der Differentialgleichung (6.109)

$$y = 1 + \frac{1}{B} \ln x \;, \tag{6.111}$$

wobei die Integrationskonstante aus der Anfangsbedingung $y = 1$, wenn $x = 1$, bestimmt wird. Da der Ausdruck $\ln x$ in Gl. (6.111) für $x \to 0$ unbestimmt wird, verwendet man bei der PCM–Sprachcodierung eine Approximation der Lösung (6.109). Für positive x gilt

$$\begin{aligned} y &= \frac{1 + \ln(A \cdot x)}{1 + \ln(A)} \qquad \text{für} \quad 1 \geq x \geq \frac{1}{A} \\ y &= \frac{A \cdot x}{1 + \ln(A)} \qquad \text{für} \quad \frac{1}{A} \geq x \\ A &= 87.56 \qquad\qquad\qquad\qquad , \end{aligned} \tag{6.112}$$

d.h. für kleine x wird y durch eine Gerade dargestellt. Diese sog. A–Kompressorkennlinie ist eine ungerade Funktion.

Aus Gründen der Schaltungsrealisierung wird sie nicht exakt, sondern in Form einer 13–Segment–Kennlinie realisiert. Die 13–Segment–Kennlinie berührt die A–Kennlinie an den Stellen $x = 1/2^n$ mit $n = 1,\ 2,\ 3,\ 4,\ 5,\ 6$, wie in Bild 6.15 dargestellt.

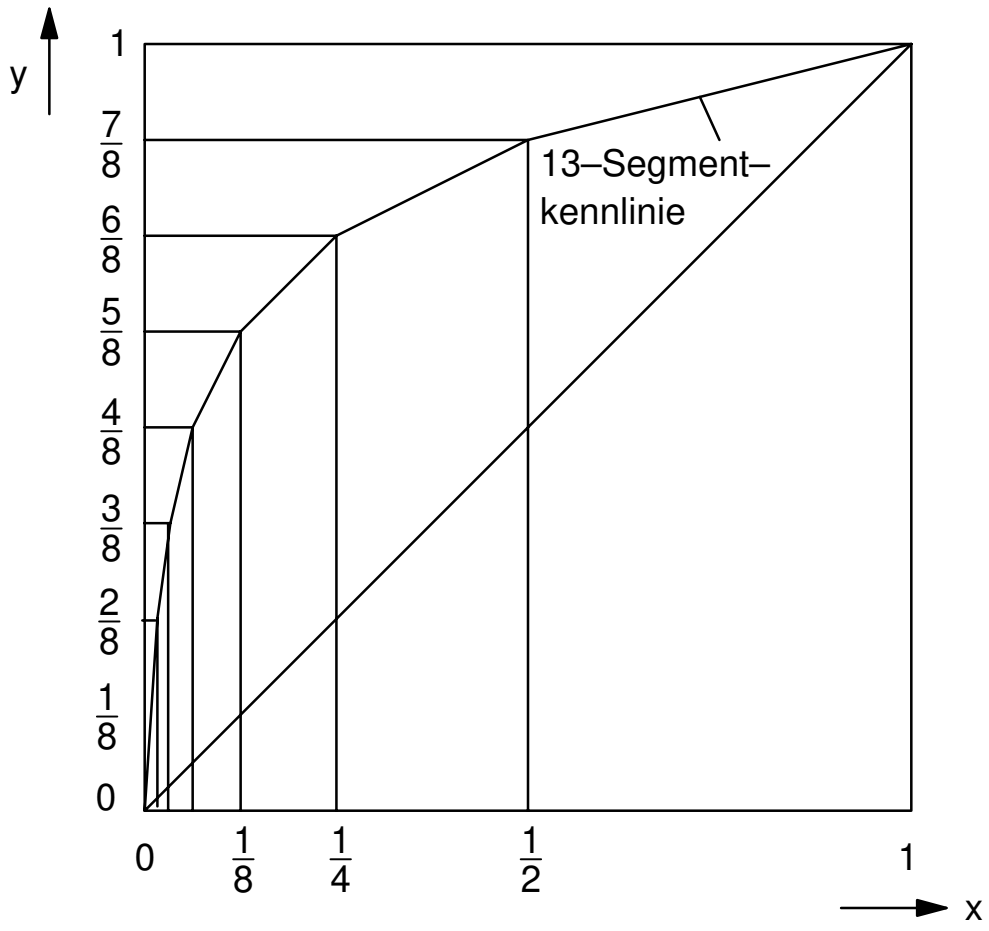


Bild 6.15 13–Segmentkennlinie der PCM–Sprachkompandierung

Das Signal–Rauschleistungs–Verhältnis für eine gleichförmige Quantisierung ergibt sich mit Gl. (6.90) zu

$$\left[\frac{P}{E}\right]_{\mathrm{dB}} = 10\log_{10}\frac{P_0 \cdot P}{E \cdot P_0} = 10\log_{10}\left(\frac{3}{2}\,K^2\,\frac{P}{P_0}\right), \tag{6.113}$$

wobei $P_0 = 1/2$ ein Bezugspegel der Eingangsleistung P ist, der einer Vollaussteuerung der Quantisierungskennlinie mit einer Sinusspannung der Amplitude 1 entspricht. Unter Verwendung der Beziehungen $\log_{10} x = \log_2 x \cdot \log_{10} 2$ und $\log_2 K = R$ wird

$$\left[\frac{P}{E}\right]_{\mathrm{dB}} = 6[R]_{\mathrm{bit}} + 1.8 + \left[\frac{P}{P_0}\right]_{\mathrm{dB}}. \tag{6.114}$$

Dieser Zusammenhang zwischen Signal–Rauschleistungs–Verhältnis und Eingangspegel einer gleichförmigen Quantisierung entspricht einer Geraden. Sie ist in

Bild 6.16 für das Beispiel einer Quantisierung mit $R = 8$ bit entsprechend $K = 256$ Quantisierungsstufen dargestellt.

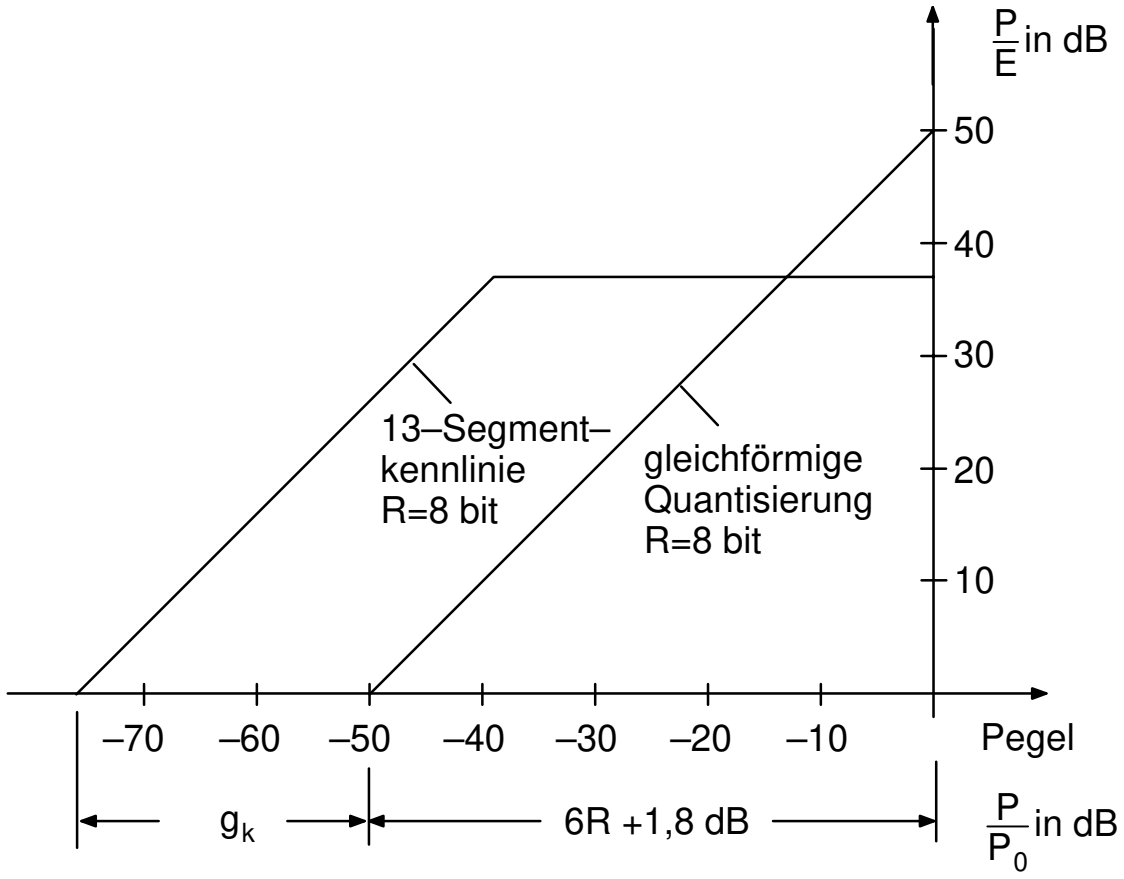


Bild 6.16 Signal–Rauschleistungs–Verhältnis *P/E* als Funktion der normierten Eingangsleistung P/P_0 für eine lineare und eine logarithmische Kompressorkennlinie

Das Signal–Rauschleistungs–Verhältnis der 13–Segment–Kennlinie wird für die beiden Bereiche der Kennlinie (6.112) getrennt ermittelt. Für $x \geq 1/A$ gilt mit (6.110)

$$\begin{aligned}\left[\frac{P}{E}\right]_{\mathrm{dB}} &= 10\log_{10}\frac{3K^2}{B^2} = 10\log_{10}\frac{3K^2}{[1+ln(A)]^2}\\ &= 10\log_{10}K^2 + 10\log_{10}\frac{3}{30} = 6\log_2 K - 10\\ &= 6[R]_{\mathrm{bit}} - 10\,.\end{aligned} \tag{6.115}$$

In diesem Bereich ist P/E unabhängig von der Eingangsleistung P. Für $x \leq 1/A$ ist die Kompressorkennlinie eine Gerade. Im Vergleich zur gleichförmigen Quantisierung in Bild 6.15 ist die Steilheit dieser Geraden jedoch um den Betrag $A/[1 + ln(A)]$ größer. Daraus resultieren kleinere Quantisierungsstufen und ein geringeres Quantisierungsrauschen. Die Verminderung des Quantisierungsrauschens wird als Kompandergewinn g_k

$$g_k = 20 \log \frac{A}{1 + ln(A)} = 24.1 \text{ dB} \tag{6.116}$$

bezeichnet. Das Signal–Rauschleistungs–Verhältnis als Funktion der normierten Eingangsleistung P/P_0 ergibt sich somit für $x \leq 1/A$ aus dem der gleichförmigen Quantisierung zuzüglich Kompandergewinn (6.114)

$$\left[\frac{P}{E}\right]_{\text{dB}} = 6[R] + 1.8 + \left[\frac{P}{P_0}\right]_{\text{dB}} + g_k \,. \tag{6.117}$$

Der Gesamtverlauf des Signal–Rauschleistungs–Verhältnisses als Funktion von P/P_0 ist für die 13–Segment–Kennlinie ebenfalls in Bild 6.15 dargestellt. Man erkennt, daß das konstante Signal–Rauschleistungs–Verhältnis über einen Dynamikbereich der Eingangsleistung P/P_0 von 35 dB erzielt werden kann. Die Kompandierung bringt im Vergleich zur gleichförmigen Quantisierung einen Gewinn an P/E bei kleinem Eingangspegel P/P_0 und dafür einen Verlust an P/E bei großem Pegel.

6.4.4 Differenz–Pulscodemodulation (DPCM)

Die Differenz–Pulscodemodulation ist im Vergleich zur PCM eine Quellencodierung, die neben der durch Quantisierung verursachten Irrelevanzreduktion eine Redundanzreduktion bewirkt und auf diese Weise bei gleichem Gütekriterium mit geringerer Übertragungsrate auskommt. Die zusätzliche Redundanzreduktion beruht im wesentlichen auf der Ausnutzung der in der Korrelation der Abtastwerte enthaltenen Redundanz. Die Wirkungsweise einer DPCM soll anhand von Bild 6.17 erläutert werden.

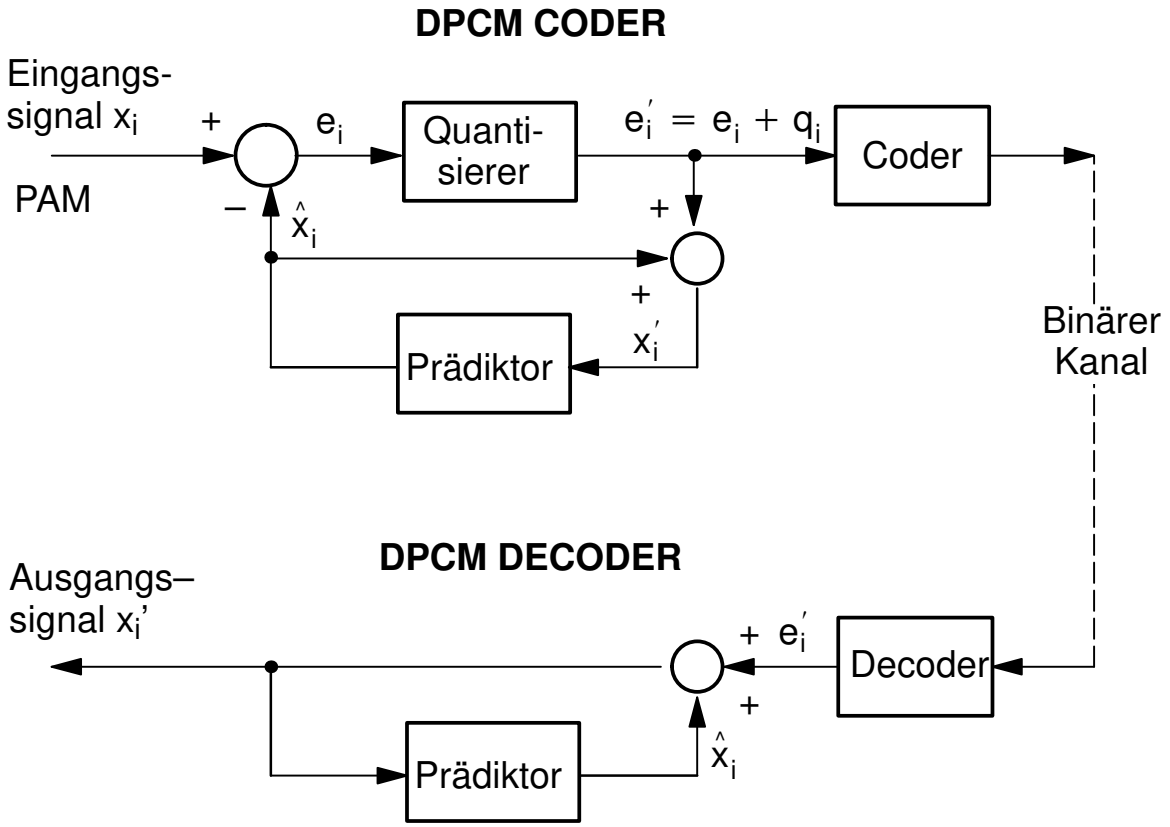


Bild 6.17 Komponenten eines DPCM–Coders und –Decoders

In einem DPCM–System wird von jedem Eingangsabtastwert $x_i = x(t_i)$ zum Zeitpunkt t_i ein Vorhersagewert $\hat{x}_i$ subtrahiert. Der Vorhersagefehler e_i

$$e_i = x_i - \hat{x}_i \tag{6.118}$$

wird quantisiert und der quantisierte Wert e_i' in binär codierter Form übertragen. Beim Empfänger wie auch beim Sender wird zu e_i' der Vorhersagewert $\hat{x}_i$ addiert. Die Summe

$$x_i' = e_i' + \hat{x}_i \tag{6.119}$$

ergibt das Ausgangssignal x_i'. Der Vorhersagewert $\hat{x}_i$ des Prädiktors wird aus einer Linerarkombination von N vorangegangenen rekonstruierten Abtastwerten x_{i-n}' gemäß der Beziehung

$$\hat{x}_i = \sum_{n=1}^{N} h_n \, x_{i-n}' = \underline{h}^T \underline{x}_i' \tag{6.120}$$

gebildet, wobei $\underline{x}_i^{'}$ ein Vektor mit den Komponenten $\underline{x}_i^{'}(x_{i-1}^{'},\ x_{i-2}^{'},\ ...,\ x_{i-N}^{'})$ und $\underline{h}^T$ der transponierte Vektor $\underline{h}\ (h_1,\ h_2,\ ...,\ h_N)$ der sog. Prädiktionskoeffizienten ist.

Bei störungsfreier Übertragung ist der Quantisierungsfehler des rekonstruierten Abtastwertes $x_i^{'}$

$$x_i^{'} - x_i = (\hat{x}_i + e_i^{'}) - x_i = e_i^{'} - e_i = q_i \tag{6.121}$$

gleich dem Quantisierungsfehler des entsprechenden Vorhersagefehlers $e_i^{'}$.

Für das DCPM–System gemäß Bild 6.17 soll nun das Signal–Quantisierungs–Rauschleistungs–Verhältnis P/E als Funktion der Übertragungsrate R ermittelt werden. Im Vergleich zur PCM kann bei der DPCM ein zusätzlicher Codierungsgewinn G dadurch erzielt werden, daß die Signalleistung σ_e^2 des zu quantisierenden Signals e im Vergleich zu $P = \sigma_x^2$ bei der PCM vermindert ist und damit die Quantisierungsstufen bei festgehaltener Stufenzahl kleiner werden. Im allgemeinen muß die Quantisierungskennlinie, so wie sie bei der PCM der Wahrscheinlichkeitsdichte $p(x_i)$ angepaßt wird, bei der DPCM der Wahrscheinlichkeitsdichte $p(e_i)$ angepaßt werden. Unterscheiden sich die beiden Wahrscheinlichkeitsdichten aber nur in der Varianz, wie es beispielsweise bei Sprachsignalen, nicht aber bei Bildsignalen zutrifft, so kann der Gewinn aus der Verminderung der zu quantisierenden Signalleistungen berechnet werden:

$$G = 10 \log \frac{\sigma_x^2}{\sigma_e^2} dB. \tag{6.122}$$

Dieser sog. Prädiktionsgewinn wird am größten, wenn σ_e^2 ein Minimum erreicht. Das Minimum kann durch Optimierung des Prädiktors $\underline{h}$ erzielt werden.

Es gilt unter der Annahme $E[e_i] = E[x_i] = 0$ mit Gl. (6.120)

$$\sigma_e^2 = E\Big[(x_i - \hat{x}_i)^2\Big] = E\Big[(x_i - \underline{h}^T \underline{x}_i^{'})^2\Big]. \tag{6.123}$$

Mit der Näherung

$$\hat{x}_i = \underline{h}^T \underline{x}_i^{'} \simeq \underline{h}^T \underline{x}_i \tag{6.124}$$

folgt

$$
\begin{aligned}
\sigma_e^2 &= E\left[(x_i - \underline{h}^T \underline{x}_i)^2\right] \\
\sigma_e^2 &= E\left[x_i^2\right] - 2E\left[x_i \underline{h}^T \underline{x}_i\right] + E\left[(\underline{h}^T \underline{x}_i)(\underline{h}^T \underline{x}_i)\right] \\
\sigma_e^2 &= E\left[x_i^2\right] - 2\,\underline{h}^T E\left[x_i \underline{x}_i\right] + E\left[(\underline{h}^T \underline{x}_i)(\underline{x}_i^T \underline{h})\right] \\
\sigma_e^2 &= E\left[x_i^2\right](1 - 2\underline{h}^T \underline{\varrho} + \underline{h}^T \wedge \underline{h}) \\
\sigma_e^2 &= \sigma_x^2 (1 - 2\underline{h}^T \underline{\varrho} + \underline{h}^T \wedge \underline{h})
\end{aligned}
\tag{6.125}
$$

wobei $\underline{\varrho}$ ein Vektor der AKF–Werte mit den Komponenten

$$
\varrho_n = \frac{E\left[x_i\, x_{i-n}\right]}{\sigma_x^2} \qquad n = 1,\ ...,\ N \tag{6.126}
$$

und Λ die Kovarianzmatrix mit den Elementen

$$
\lambda_{nm} = \frac{E[x_{i-n}\, x_{i-m}]}{\sigma_x^2} \tag{6.127}
$$

ist. Das Minimum für σ_e^2 gemäß Gl.(6.125) wird mit dem optimalen Prädiktor $\underline{h}_{opt}$

$$
\underline{h}_{opt} = \Lambda^{-1} \underline{\varrho} \tag{6.128}
$$

erreicht. Es wird

$$
\begin{aligned}
\sigma_{e^2_{\min}}^2 &= \sigma_x^2 (1 - \underline{\varrho}^T \cdot \underline{h}_{opt}) \\
&= \sigma_x^2 - h_1 E\left[x_i\, x_{i-1}\right] - h_2 E\left[x_i\, x_{i-2}\right] ... - h_N E\left[x_i\, x_{i-N}\right]
\end{aligned}
\tag{6.129}
$$

und der durch die DPCM im Vergleich zur PCM erzielbare zusätzliche Gewinn im Signal–Quantisierungs–Rauschleistungs–Verhältnis beträgt

$$
G_{opt} = 10 \log(1 - \underline{\varrho}^T \Lambda^{-1} \underline{\varrho})^{-1} \tag{6.130}
$$

Es kann gezeigt werden, daß der optimale Prädiktor unkorrelierte Vorhersagefehler e_i erzeugt.

Eine Auswertung der Gl. (6.130) für Fernsprechsignale gibt abhängig von der Anzahl N der Prädiktorkoeffizienten folgende Gewinne des Signal–Quantisierungs–Rauschleistungs–Verhältnisses einer DPCM im Vergleich zur PCM:

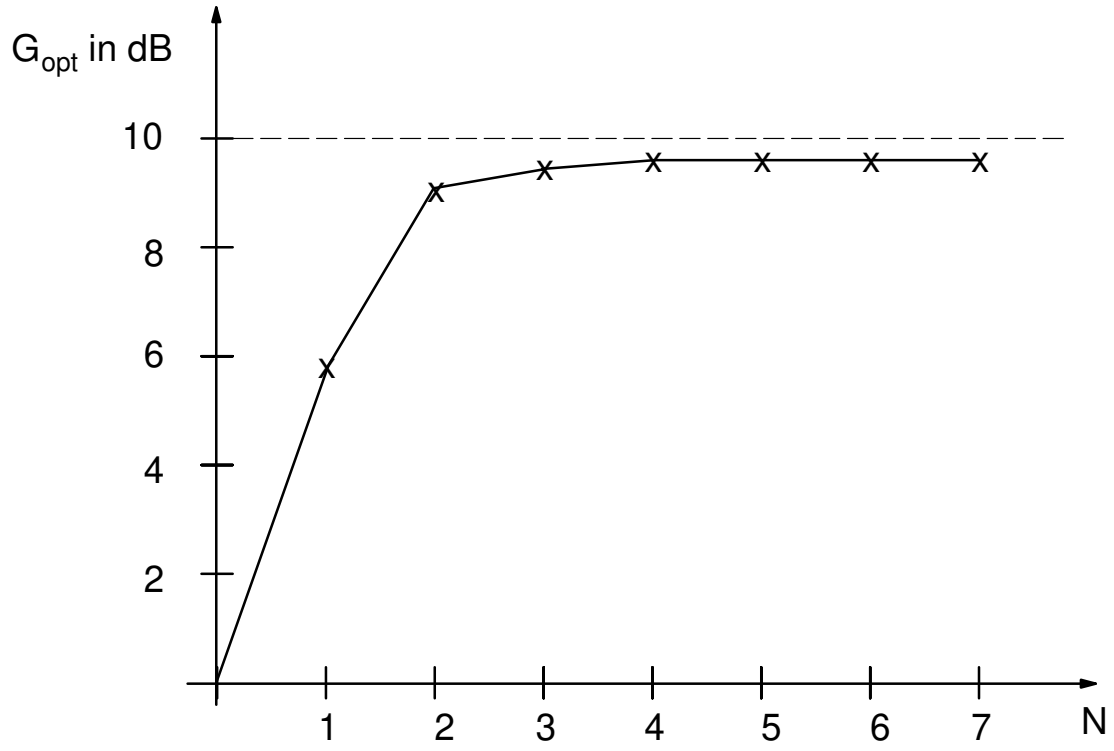


Bild 6.18 DPCM–Gewinn als Funktion der Anzahl N der Prädiktionskoeffizienten für Fernsprechsignale der Frequenzbandbreite W=3,4 kHz und Abtastfrequenz $f_A = 8$ kHz

Der Signal–Quantisierungs–Rauschabstand einer DPCM als Funktion der Übertragungsrate beträgt mit Gl. (6.115)

$$\begin{aligned} \left[\frac{P}{E}\right]_{DPCM} &= \left[\frac{P}{E}\right]_{PCM} + G_{opt} \\ \left[\frac{P}{E}\right]_{DPCM} &= 6[R]_{bit} - 10 + G_{opt} \end{aligned} \tag{6.131}$$

in dB für Fernsprechsignale. Für $R = 1$ bit entartet die DPCM zum Sonderfall der sog. Deltamodulation.

Bild 6.19 zeigt die Übertragungsbitrate einer PCM entsprechend Gl. (6.115) und einer DPCM entsprechend Gl. (6.131) als Funktion des Gütekriteriums P/E im Vergleich zur oberen Grenze der Rate–Distortion–Funktion R_S gemäß Gl. (6.68).

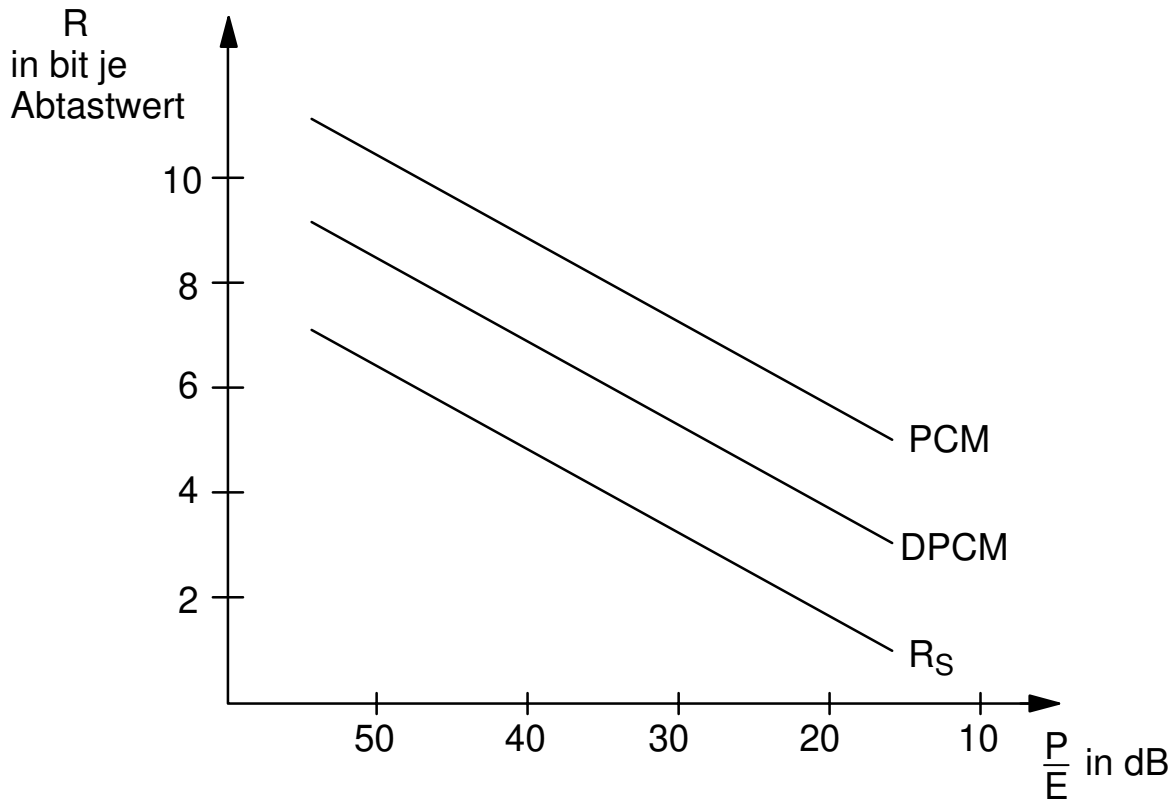


Bild 6.19 Übertragungsrate von PCM und DPCM bei Fernsprechsignalen als Funktion von P/E

Die Übertragungsbitrate der PCM und DPCM kann durch Anwendung einer Entropie–Codierung anstelle der angenommenen Codierung mit fester Codewortlänge von R bit je Abtastwert noch verringert werden.

Literaturverzeichnis

[1] Berger, T.: Rate Distortion Theory; Prentice Hall, Inc. Englewood Cliffs, New Jersey, 1971

[2] Feinstein, A.: Error Bounds in Noisy Channels without Memory; IRE Trans. Inform. Theory, Vol. IT–1, pp.13–14, 1955

[3] Gallager, R.G.: Information Theory and Reliable Communication; John Wiley and Sons, New York, 1968

[4] Huffman, D.A.: A Method for the Construction of Minimum Redundancy Codes; Proc. IRE, No. 40, pp. 1098–1101, 1952

[5] Max, J.: Quantizing for Minimum Distortion; IRE Trans. Inform. Theory, Vol. IT–6, pp. 7–12, 1960

[6] Panter, P.F., Dite, W.: Quantization Distortion in Pulse Count Modulation with nonuniform Spacing of Levels; Proc. IRE, Vol. 35, pp. 44–48, 1951

[7] Shannon, C.E.: A Mathematical Theory of Communication; Bell System Tech. J., Vol. 27, pp. 379–423 (Part I), pp. 623–656 (Part II), 1948.
Reprinted in book form by W.Weaver, Univ. of Illinois Press, Urbana, 1949

[8] Shannon, C.E.: Coding Theorems for a Discrete Source with a Fidelity Criterion IRE Nat'l. Conv. Rec., Part 4, pp. 142–163, 1959

Stichwortverzeichnis

A

Amplitudenquantisierung, 87

Äquivokation, 55

B

Basisbandübertragung, 66, 71

Blockcodierung, 34, 37, 75, 93

BSK (binärer, symmetrischer Kanal), 45, 60

C

Code
  fehlererkennender, 73
  fehlerkorrigierender, 73
  systematischer, 76

Codebaum, 19, 26

Codedistanz, 74

Coderate, 73, 83

Codesymbol, 25, 26, 27, 29

Codetabelle, 93

Codewort, 25, 30, 34, 38, 93

D

Decoder, 77, 124

DPCM, 128, 129

E

Entropie, 14, 29, 31, 38

Entscheidungsknoten, 27

F

Faltungscode, 75

Fehlerbündel, 43

Fehlererkennung, 73

Fehlerkorrektur, 73, 75

Fehlermuster, 41, 42, 77

Fehlerwahrscheinlichkeit, 82, 84

G

Gaußsche Quelle, 103, 104, 105

Gaußsche Wahrscheinlichkeitsdichte, 52

H

Hamming–Distanz, 73

Huffman–Codierungsalgorithmus, 32

I

Informationsgehalt, wechselseitiger, 41, 47, 49, 56

Irrelevanz, 56, 61, 87

K

Kanal
binär, symmetrisch, 45
diskreter, gedächtnisloser, 45

Kanaldecoder, 75

Kanalkapazität, 41, 56, 60, 95
des deterministischen Kanals, 58
des diskreten, gedächtnislosen Kanals, 55
des ungestörten Kanals, 59
des verlustfreien Kanals, 56
kontinuierlicher, gedächtnisloser Kanäle, 63

Kompressorkennlinie, 113, 115

Korrekturbereich, 81

L

Lauflängencodierung, 37

Leistung, 64

M

Markoff–Quelle, 35

Maximum–Likelihood–Decodierung, 79

N

Nachrichtenübertragungssystem, 5

P

Parität, 76, 77

Parity–Check–Code, 76

PCM, 124, 133

Prädiktion, 130

Präfixcode, 26, 27, 32, 35

Q

Quantisierungsfehler, 51

Quellencodierung, 14
optimale, 93

Quellenmodelle, 6

R

Rate–Distortion–Funktion, 88
für diskrete Quellen, 95
für gedächtnisbehaftete Gaußsche Quellen, 105
für kontinuierliche Quellen, 99

Rate–Distortion–Theorie, 87

Redundanz, 73, 129

Repräsentativwert, 111

Restfehlerwahrscheinlichkeit, 41, 95

T

Theorem
der Kanalcodierung, 10
der Quellencodierung, 8
der Quellencodierung bei Berücksichtigung eines Gütekriteriums, 94

Transinformationsgehalt, 47, 48, 49, 55

V

Verzerrungsmaß, 87, 93, 100

W

wechselseitiger Informationsgehalt, 48