Inhaltsverzeichnis

Kapitel 2

Informations- und Codierungstheorie

Die statistische **Informationstheorie** wurde durch **Claude E. Shannon** mit der Publikation grundlegender Artikel "A Mathematical Theory of Communications", "Communications in the Presence of Noise", "Communication Theory of Secrecy Systems" u.v.a.m. in den Jahren 1948/1949 begründet [1, 2, 3]. Information wird dabei messbar erfasst, und es werden grundlegende Beziehungen zwischen Information und deren Repräsentation durch physikalische Prozesse hergestellt. Die Informationstheorie umfasst zahlreiche Existenz- bzw. Nichtexistenzbeweise für technische Verfahren zur Informationsrepräsentation und -übertragung. Aus den Hauptsätzen zu Quellen- und Kanalcodierung entwickelte sich in der Folgezeit die **Codierungstheorie** sowohl bzgl. Datenkompression als auch störresistenter Datenübertragung. Aufgrund ihrer fundamentalen Bedeutung für die technische Entwicklung seit Mitte des 20. Jahrhunderts markieren die Arbeiten von C.E. Shannon zusammen mit der gleichzeitigen Erfindung des Transistors den Beginn des *Informationszeitalters*. Ausführliche Darstellungen von Informations– und Codierungstheorie sind z.B. in [4, 5] zu finden.

2.1 Grundlegende Definitionen

2.1.1 Diskrete gedächtnislose Informationsquelle

Eine **diskrete Informationsquelle** gibt zu diskreten Zeitpunkten $k \in \mathbb{Z}$ *Symbole* $X[k]$ ab, die einem bekannten *Symbolalphabet* $\mathcal{X} = \{x_1, \ldots, x_{M_x}\}$ mit endlicher Mächtigkeit M_x (*Symbolumfang*), entstammen. Bezüglich des Beobachters der Quelle stellt das Symbol $X[k]$ eine *Zufallsvariable* dar, wobei die M_x *a–priori* Wahrscheinlichkeiten für den Zeitpunkt k mit $\Pr(X[k] = x_i),\ i = 1(1)M_x$ bezeichnet werden.[1] Bei einer **gedächtnislosen, diskreten Informationsquelle** (**DMS**: discrete memoryless source) folgen die Symbole zudem *statistisch unabhängig* aufeinander. Die Wahrscheinlichkeit einer *Symbolsequenz* entspricht dem Produkt der einzelnen Symbolwahrscheinlichkeiten. Bei einer **zeitinvarianten**, diskreten Informationsquelle gilt außerdem $\Pr(X[k] = x_i) \stackrel{\triangle}{=} p_i,\ \forall k \in \mathbb{Z}$.[2] Die Quellensymbolsequenz ist in diesem Fall eine sog. **i.i.d.**–Folge (independent, identically distributed = Folge unabhängiger, identisch verteilter Zufallsvariablen).

2.1.2 Das Informationsmaß nach Shannon

Das Informationsmaß $I_S(x_i)$ nach Shannon für den Informationsgewinn (Verringerung der Unsicherheit über das Ergebnis eines Zufallsexperiments) durch Beobachtung des Symbols x_i am Ausgang einer Quelle folgt aus den Forderungen a) $I_S(x_i) \in \mathbb{R}, I_S(x_i) \geq 0$; b) $I_S(x_i)$ sei eine Funktion f der a–priori Wahrscheinlichkeit; c) bei einer *gedächtnislosen* Quelle (DMS) sei das Informationsmaß für eine Folge von N Symbolen die *Summe* der Informationsmaße für die Einzelsymbole: $I_S(x_{i_1}, \ldots, x_{i_N}) = \sum_{n=1}^{N} I_S(x_{i_n})$. Hieraus ergibt sich, dass für die Funktion f der negative Logarithmus zu wählen ist.

Def: Informationsmaß: $I_S(x_i) \stackrel{\triangle}{=} -\log_b(\Pr(X = x_i))$ [bit *oder* nat]

Bei Anwendung der Basis $b = 2$ (logarithmus dualis: $\log_2(\cdot) \stackrel{\triangle}{=} \mathrm{ld}(\cdot)$) ist die Hinweiseinheit **bit** (=binary digit), der Basis $b = \mathrm{e}$ ($\log_{\mathrm{e}}(\cdot) \stackrel{\triangle}{=} \ln(\cdot)$) die

[1] Durch die Notation $j = b(s)e$ werden der Variable j reelle Zahlen beginnend mit b bis e bei Schrittweite s zugewiesen.

[2] Das Symbol $\stackrel{\triangle}{=}$ weist auf eine Definition hin.

Hinweiseinheit **nat** (=natural digit) üblich; es gilt also: 1 bit = 0,693 nat. Nachfolgend wird o.B.d.A. ausschließlich die Basis 2, also das Informationsmaß bit, verwendet.

▶ *Hinweis:* Die Hinweiseinheit **bit** für das Shannon'sche Informationsmaß ist streng zu unterscheiden von **Bit** für ein *binäres Symbol* bzw. eine *zweiwertige Variable*, wobei Bit als populäre Verballhornung aus der ursprünglichen Bedeutung von bit entstanden ist, vgl. auch Quellencodierungstheorem.

Interpretation: Das Maß $-\log_b(\Pr(X[k] = x_i))$ für die *Unsicherheit* bzgl. des Auftretens des Quellensymbols x_i vor dessen Beobachtung (a–priori) ist zunächst zur a–priori Wahrscheinlichkeit $\Pr(X_k = x_i)$ völlig äquivalent, da $\log_b(x)$ für $x \in (0, 1]$, $b > 1$ bijektiv ist. Bei Verwendung des Logarithmus zur Basis 2 erfolgt ein Vergleich einer Wahrscheinlichkeit mit der Wahrscheinlichkeit für I_S Realisierungen des elementarsten Zufallsexperimentes, das denkbar ist, nämlich der Auswahl eines von zwei gleichwahrscheinlichen Ereignissen (z.B. Münzwurf), da gilt: $-\mathrm{ld}(0,5) = 1\,\mathrm{bit}$.

❑ *Beispiel:* Beim Lotto "6 aus 49" beträgt das Informationsmaß für das Ereignis "6 Richtige" $-\mathrm{ld}\left(1/\binom{49}{6}\right) = \mathrm{ld}(1,398 \cdot 10^7) = 23,7\,\mathrm{bit}$. Das bedeutet, das Ereignis "6 Richtige" ist etwa gleich (un)wahrscheinlich, wie 24 Münzwürfe nach vorgegebener Folge.

Aber nicht die Informationsmaße bzgl. einzelner Symbole, sondern *Mittelwerte* erweisen sich als fundamentale Größen. Erst durch die *Codierungstheoreme* der Informationstheorie wird die Relevanz des Shannon'schen Informationsmaßes bestätigt (vgl. Abschnitte 2.2.1 und 2.3.4).

2.1.3 Mittlerer Informationsgehalt, Entropie

Def: Entropie einer Informationsquelle:

$$H(X) \stackrel{\triangle}{=} \mathrm{E}\{I_S(X)\}, \quad \left[\frac{\mathrm{bit}}{\mathrm{Symbol}}\right] \text{ mit } \mathrm{E}\{\cdot\}: \text{ Erwartungswertoperator.} \quad (2.1)$$

Speziell für diskrete, zeitinvariante Quellen (DMS) gilt:

$$H(X) \stackrel{\triangle}{=} -\sum_{i=1}^{M_x} p_i \,\mathrm{ld}(p_i) \left[\frac{\mathrm{bit}}{\mathrm{Symbol}}\right] \text{ mit } p_i = \Pr(X = x_i) \qquad (2.2)$$

Die Bezeichnung *Entropie* für den *mittleren Informationsgehalt je Symbol* wurde von C.E. Shannon wegen der nahen Verwandtschaft dieser Größe zur Entropie in der Thermodynamik gewählt.

Eigenschaften der Entropie:

- *Unmögliche* Ereignisse liefern aufgrund $\lim_{p \to +0} p \ln(p) = 0$ keinen Beitrag.
- Für die Entropie $H(X)$ einer diskreten, gedächtnislosen Quelle (DMS) mit einem Symbolvorrat $\mathcal{X}$ mit dem Umfang M_x gelten folgende Schranken: $0 \leq H(X) \leq \mathrm{ld}\,(M_x)$.
- Die Entropie einer DMS mit dem Symbolumfang M_x wird durch und nur durch *gleichverteilte* Symbole maximiert: $\max_{\Pr(X)} H(X) = \mathrm{ld}(M_x)$ für $\Pr(X = x_i) = 1/M_x \;\forall\; x_i \in \mathcal{X}$.
- Die Entropie wird durch und nur durch ein a–priori sicheres Ereignis minimiert: $\min_{\Pr(X)} H(X) = 0$ falls $\exists i$ mit $\Pr(X = x_i) = 1$.

❑ *Beispiel:* Binäre Quelle $X \in \{A, B\}$ mit $\Pr(X = A) = p$, siehe Abbildung 2.1

$$H(X) = -p\,\mathrm{ld}(p) - (1-p)\,\mathrm{ld}(1-p) \stackrel{\triangle}{=} \mathrm{e}_2(p) \quad \text{binäre Entropiefunktion} \tag{2.3}$$

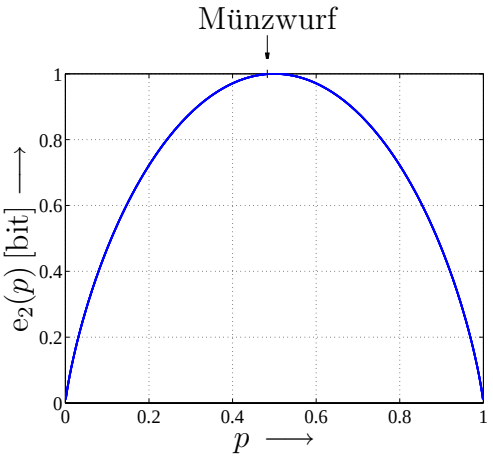


Abbildung 2.1: Binäre Entropiefunktion

2.1.4 Gekoppelte Informationsquellen

Zwei Zufallsvariablen X und Y seien infolge irgendwie gearteter Kopplungen zwischen den Informationsquellen statistisch abhängig, vgl. Abbildung 2.2:

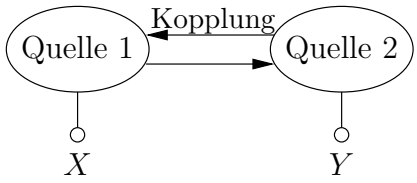


Abbildung 2.2: Gekoppelte Informationsquellen

Beide Quellen seien diskret, gedächtnislos, zeitinvariant (gekoppelte DMS's) und die a–priori Verbundwahrscheinlichkeiten $\Pr(X = x_i, Y = y_j)$ seien für alle $M_x M_y$ Symbolpaare bekannt. (Mit $X = X[k]$ und $Y = X[k+1]$ werden auch gedächtnisbehaftete Quellen erfasst.)

Def: Verbundentropie: Mittlerer Informationsgewinn bei Beobachtung beider Quellen (Hypersymbol XY).

$$H(XY) \stackrel{\triangle}{=} -\sum_{i=1}^{M_x}\sum_{j=1}^{M_y} \Pr(X{=}x_i, Y{=}y_j)\,\mathrm{ld}(\Pr(X{=}x_i, Y{=}y_j)) \left[\frac{\text{bit}}{\text{Symbolpaar}}\right]$$

Anmerkungen zu Verbund- und bedingten Wahrscheinlichkeiten:

$$\begin{aligned} \text{Bedingte Wahrscheinlichkeit:} \quad & \Pr(X\,|Y\,) \stackrel{\triangle}{=} \Pr(XY)\,/\Pr(Y) \qquad (2.4)\\ \text{Satz von Bayes:} \quad & \Pr(X\,|Y\,) = \Pr(Y\,|X\,)\cdot\Pr(X)\,/\Pr(Y)\\ \text{Randverteilung:} \quad & \Pr(X) = \sum_{Y}\Pr(XY)\,;\Pr(Y) = \sum_{X}\Pr(XY) \end{aligned}$$

Def: Bedingte Entropie:

$$H(X \mid Y) \stackrel{\triangle}{=} -\sum_{i=1}^{M_x}\sum_{j=1}^{M_y} \Pr(X = x_i, Y = y_j)\ \mathrm{ld}(\Pr(X = x_i \mid Y = y_j)) \qquad (2.5)$$

Satz: Eine zusätzliche Beobachtung kann die Unsicherheit bzgl. des Ergebnisses eines Zufallsexperimentes *im Mittel* niemals vergrößern, d.h.

$$H(X) \geq H(X \mid Y) \qquad (2.6)$$

Das Gleichheitszeichen gilt nur bei statistischer Unabhängigkeit.

Satz: Entwicklung der Verbundentropie (Kettenregel):

$$H(XY) = H(X) + H(Y \mid X) = H(Y) + H(X \mid Y) \tag{2.7}$$

Die statistische Abhängigkeit zweier Zufallsvariablen wird insbesondere dadurch gekennzeichnet, welcher Informationsgewinn bzgl. einer Variablen durch Beobachtung der anderen erzielt werden kann. Wird die Unsicherheit über die Variable X durch $-\mathrm{ld}\,(\Pr(X))$ vor der Beobachtung von Y (*a–priori*) charakterisiert, so verbleibt nach der Beobachtung von Y (*a–posteriori*) eine (Rest-) Unsicherheit über X von $-\mathrm{ld}\,(\Pr(X\,|Y\,))$; die Unsicherheit über X wird also durch die Beobachtung von Y um $\mathrm{ld}\,(\Pr(X\,|Y\,)/\Pr(X))$ bit verringert. Aus dem Bayes'schen Satz (2.4) folgt unmittelbar die **Symmetrie**: $\Pr(X\,|Y\,)/\Pr(X) = \Pr(Y\,|X\,)/\Pr(Y)$.

Def: Wechselseitige Information: Mittelwert des Informationsgewinns bzgl. der einen Quelle durch Beobachtung der anderen.

$$I(X;Y) \;\stackrel{\triangle}{=}\; \mathrm{E}\left\{\mathrm{ld}\left(\frac{\Pr(X \mid Y)}{\Pr(X)}\right)\right\} = \mathrm{E}\left\{\mathrm{ld}\left(\frac{\Pr(Y \mid X)}{\Pr(Y)}\right)\right\}$$

$$= \sum_{i=1}^{M_x}\sum_{j=1}^{M_y} \Pr(X = x_i, Y = y_j)\ \mathrm{ld}\left(\frac{\Pr(X = x_i \mid Y = y_j)}{\Pr(X = x_i)}\right)$$

$$I(X;Y) = H(X) - H(X \mid Y) = I(Y;X) = H(Y) - H(Y \mid X)$$

$$I(X;Y) = H(X) + H(Y) - H(XY),\ \text{(vgl. Gl. (2.7))}$$

Die wechselseitige Information ist von besonderer Bedeutung bei der digitalen Nachrichtenübertragung über einen gestörten Übertragungskanal, vgl. auch Abbildung 2.6: Wegen örtlicher oder zeitlicher Entfernung ist die Zufallsvariable X nicht direkt beobachtbar; nur eine Zufallsvariable Y am Ende eines gestörten Übertragungskanals kann auf *indirekte Weise* Information über die gewünschte Größe X vermitteln, wobei Y und X mehr (kleine Störung) oder weniger (große Störung) stark statistisch abhängig sind.[3]

[3]Falls $I(X;Y)$ die wechselseitige Information zwischen dem Eingang und dem Ausgang eines nicht zuverlässigen Kanals ausdrückt, wird auch die Bezeichnung **Transinformation** (für transportierte Information) benutzt.

Gemäß Gl. 2.6 ist die mittlere wechselseitige Information nicht negativ, $I(X;Y) \geq 0$, wobei $I(X;Y) = 0$ nur bei statistischer Unabhängigkeit gilt.

Die Beobachtung der einen Informationsquelle kann **im Mittel** die Unsicherheit bzgl. der anderen nicht vergrößern! **Konsequenz:** Die Verbreitung von „Falschinformation“ (hier Y) kann die Unsicherheit bzgl. eines tatsächlichen Ereignisses (hier X) **im Mittel** (also z.B. langfristig) nicht erhöhen. Je *gezielter* „Falschinformation“ eingesetzt wird (also je *mehr Kopplung* zwischen Y und X besteht), desto mehr Information über tatsächliche Ereignisse verrät „Falschinformation“. (*„Lügen haben kurze Beine“*)

Satz: Entwicklung der wechselseitigen Information (Kettenregel):

$$I((X_1, X_2); Y) = I(X_1; Y) + I(X_2; Y|X_1) \tag{2.8}$$

Satz 2.8 ist von grundsätzlicher Bedeutung für die digitale Übertragungstechnik: Mittels sukzessiver Vorgehensweise sind Detektionsverfahren aufwandsgünstig implementierbar und bleiben dennoch theoretisch optimal (z.B. Multilevel Codierung, entscheidungsrückgekoppelte Entzerrung, sukzessive Interferenzauslöschung u.v.a.m.)

2.2 Codierung für diskrete Quellen

Ziel der **Quellencodierung** ist eine möglichst kompakte Repräsentation von Quellensymbolsequenzen durch Codesymbolsequenzen ⇒ Codierung zur Datenkomprimierung. Hierzu werden Wörtern $\vec{X} = (X_1, \ldots, X_l)$ aus l Quellensymbolen Codewörter $\vec{C} = (C_1, \ldots, C_n)$ mit fester oder variabler Wortlänge n zugeordnet.

Grundlegende Begriffe: **Codesymbolalphabet** $\{c_1, c_2, \ldots, c_{M_c}\}$ mit Codeumfang M_c; **Code C**: Menge der verwendeten Codewörter $\vec{C}$ mit **Umfang des Codes** $K = |\mathbf{C}|$; **Codierung**: Zuordnungsvorschrift zwischen Quellen- und Codewörtern.

► *Hinweis:* Die Begriffe **Code** als Menge der Codewörter und **Codierung** als Zuordnungsvorschrift sind streng zu unterscheiden. Bei K Quellen- und Codewörtern gibt es $K!$ unterschiedliche Codierungen für einen Code.

Bei einer **verlustfreien Quellencodierung** erfolgt **redundanzarme**, **umkehrbare** Repräsentation der Information, durch im Mittel möglichst weni-

ge Codesymbole. Bei einer **verlustbehafteten Quellencodierung** werden hingegen unterschiedlichen Quellensymbolsequenzen mitunter auch gleiche Codesymbolsequenzen zugeordnet, z.B. weil für der Verbraucher Unterschiede nicht verwert- oder wahrnehmbar sind. Die Menge der vom Verbraucher unterscheidbaren Quellensymbolsequenzen kann mit Hilfe einer äquivalenten Informationsquelle beschrieben werden, wobei nicht unterscheidbare Sequenzen deren Redundanz bilden; diese wird als **Irrelevanz** bezeichnet. Verlustbehaftete Quellencodierungsverfahren nutzen häufig Irrelevanzreduktion zur Datenkompression (z.B. MP3 Audiosignalcodierung), da sie damit subjektiv als verlustlos (verlustarm) erscheinen.

Im Folgenden werden nur verlustlose Quellencodierverfahren behandelt, wobei die Quelle auch gedächtnisbehaftet sein mag. Es wird jedoch angenommen, dass Wörter $\vec{X}$ statistisch unabhängig aufeinander folgen bzw. deren Abhängigkeit nicht verwertet wird.

2.2.1 Codes mit variabler Wortlänge

Durch Zuweisung von kurzen Codewörtern zu häufigen Quellenwörtern und langen Codewörtern zu seltenen Quellenwörtern wird **im Mittel** eine Datenkompression erreicht. Codes mit variabler Wortlänge sind jedoch nur dann effizient, wenn kein Trennungssymbol zwischen den Codewörtern erforderlich ist (*kommafreier*[4] Code). Die Bestimmung von Wortgrenzen innerhalb einer Sequenz ist beispielsweise möglich, wenn kein kürzeres Codewort Teil eines längeren Codewortes in den ersten Stellen ist. Solche Codes werden als **präfixfreie Codes** bezeichnet (Präfix = Vorsilbe). Die Wörter eines Codes mit variabler Wortlänge über einem M_c–wertigen Alphabet werden durch **Pfade** von der **Wurzel** zu den **Endknoten** in einem M_c–**wertigen Baum** repräsentiert. (Bei einem M_c–wertigen Baum gehen von einem *Knoten* bis zu M_c *Zweige* aus, wobei jeder Zweig wiederum einen Knoten erzeugt. Die Wurzel bildet den *Urknoten* des Baumes; von Endknoten gehen keine Zweige aus.)

[4]Insofern ist der Morsecode trotz der Anpassung der Wortlänge an die Symbolwahrscheinlichkeit ein relativ ineffizientes Verfahren.

Def: Präfixfreier Code: Bei einem *präfixfreien Code* **C** mit variabler Wortlänge werden die $K = |\mathbf{C}|$ Codewörter nur durch Pfade von der Wurzel zu *Endknoten* repräsentiert. Zwischenknoten sind dagegen *keine* Codewörter zugeordnet.
⇒ Präfixfreie Codes sind durch Verfolgung von Pfaden von der Wurzel zu Endknoten eindeutig kommafrei decodierbar.

Satz: Kraft'sche Ungleichung: Ein präfixfreier Code **C** über einem M_c-wertigen Alphabet mit $K = |\mathbf{C}|$ Wörtern der Längen $n_1, n_2, \ldots, n_K$ existiert dann und nur dann, wenn die folgende Ungleichung erfüllt ist:

$$\sum_{i=1}^{K} M_c^{-n_i} \leq 1 \tag{2.9}$$

❑ *Beispiel:* Präfixfreier Binärcode mit $K = 6$ Wörtern, siehe Abbildung 2.3.

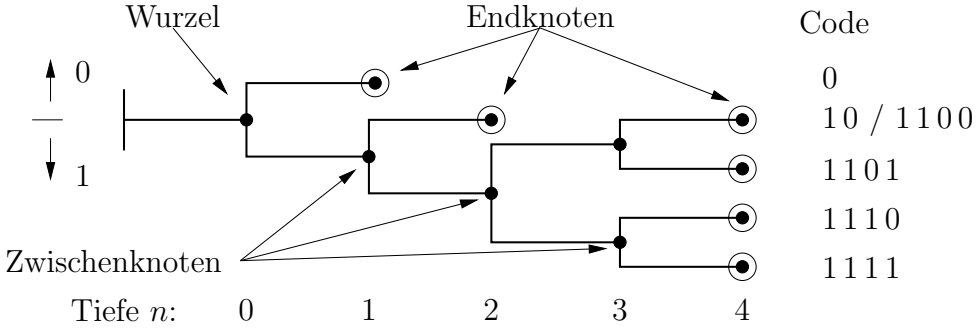


Abbildung 2.3: Baum eines präfixfreien Binärcodes

Der Beweis zur Kraft'schen Ungleichung erfolgt durch *ein Konstruktionsverfahren*, indem in einen zunächst bis zur Tiefe $\max\{n_i\}$ vollständigen Baum rekursiv Knoten bei den Tiefen n_i zu Endknoten erklärt und alle davon ausgehenden Zweige eliminiert werden. Der verbleibende Baum, der bei Gültigkeit der Kraft'schen Ungleichung immer, andernfalls niemals, konstruierbar ist, repräsentiert den präfixfreien Code.

Satz von McMillan: *Jeder* eindeutig decodierbare Code mit K Wörtern über einem M_c–wertigen Alphabet erfüllt die Kraft'sche Ungleichung Gl. (2.9).

Da präfixfreie Codes somit optimale Codes mit variabler Wortlänge sind,

interessieren evtl. existierende andere decodierbare Codes mit variabler Wortlänge nicht weiter.

Satz: Werden die Symbole, die von einer diskreten, gedächtnislosen Quelle (DMS) mit dem mittleren Informationsgehalt (Entropie) $H(X)$ abgegeben werden, umkehrbar eindeutig durch die Wörter eines eindeutig decodierbaren Codes $\mathbf{C}$ mit variabler Wortlänge über einem M_c–wertigen Alphabet effizient repräsentiert, so gilt für die *mittlere* Codewortlänge $\bar{n} \stackrel{\triangle}{=} \sum_{i=1}^{M_x} n_i \cdot \Pr(X = x_i)$:

$$\frac{H(X)}{\mathrm{ld}(M_c)} \leq \bar{n} < \frac{H(X)}{\mathrm{ld}(M_c)} + 1 \ , \tag{2.10}$$

d.h. es existiert eine solche Codierung mit einer mittleren Wortlänge kleiner als $H(X)/\mathrm{ld}(M_c) + 1$, es existiert jedoch keine solche Codierung mit einer mittleren Wortlänge kleiner als $H(X)/\mathrm{ld}(M_c)$.

Speziell für binäre Codesymbole gilt also: $H(X) \leq \bar{n} < H(X) + 1$.

Def: Optimale Wortlänge für ein Symbol x_i:

$$n_{\mathrm{o},i} \stackrel{\triangle}{=} -\log_{M_c}(\Pr(X = x_i)).$$

Kann für alle M_x Quellensymbole jeweils die optimale Codewortlänge gewählt werden, so wird die Kraft'sche Ungleichung mit Gleichheit erfüllt. Die mittlere Codewortlänge entspricht dann genau der (auf das M_c–wertige Alphabet umgerechneten) Entropie $H(X)/\mathrm{ld}(M_c)$. Bei kürzeren Wortlängen ist die Kraft'sche Ungleichung nicht mehr erfüllbar und es existiert damit kein eindeutig decodierbarer Code. Durch Aufrunden der "idealen" Wortlängen zur nächsten ganzen Zahl findet man Wortlängen, für die ein decodierbarer Code sicher konstruierbar ist. Die mittlere Wortlänge erfüllt dabei die rechte Ungleichung in (2.10).

Durch Zusammenfassen von jeweils L Quellensymbolen zu Wörtern, für die eine Quellencodierung bei einer mittleren Wortlänge $\leq L \cdot H_L(X)/\mathrm{ld}(M_c) + 1$ nach Gl. (2.10) sicher möglich ist, folgt unmittelbar das

Quellencodierungstheorem (fixe Quellenwort-, variable Codewortlänge): Für diskrete Informationsquellen mit dem mittleren Informationsgehalt (Entropie) $H(X)$ je Symbol existiert eine M_c–wertige umkehrbar eindeutige Codierung, für deren mittlere Wortlänge $\bar{n}$ gilt: $\bar{n} < H_L(X)/\mathrm{ld}(M_c) + 1/L$, $\forall L \in \mathbb{N}$. Es gibt keine solche Codierung mit $\bar{n} < H(X)/\mathrm{ld}(M_c)$.

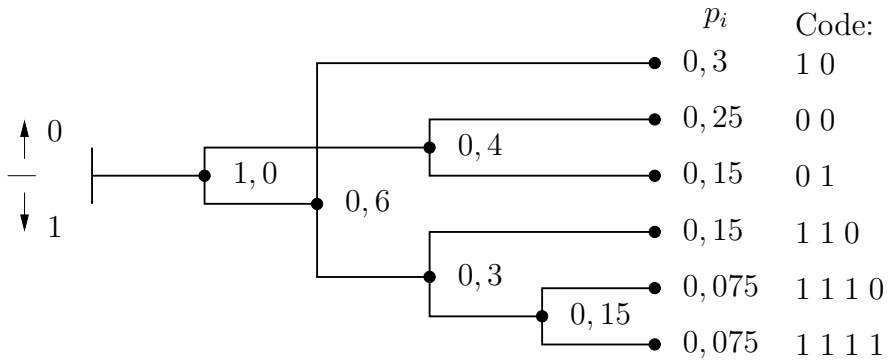
Interpretation: Dieser Satz zeigt, dass für jede DMS eine binäre Quellencodierung existiert, mit der die mittlere Zahl von Binärsymbolen je Quellensymbol beliebig nahe zur Entropie gesenkt werden kann; eine noch kompaktere Repräsentation ist aber prinzipiell nicht möglich. Damit ist gezeigt, dass das Shannon'sche Informationsmaß *tatsächlich* den Informationsgehalt eines Symbols genau trifft. Das axiomatisch eingeführte Informationsmaß wird also mit diesem Satz als praxisrelevant bestätigt. Je länger die Blöcke von Quellensymbolen sind, um so mehr nähern sich nach dem Gesetz der großen Zahlen die relativen Häufigkeiten für die einzelnen Symbole den a–priori Wahrscheinlichkeiten an; d.h. um so besser kann die Codierung der Quelle angepasst werden.

Optimaler Code, Huffman–Codierung

Aus dem Baum zu einem präfixfreien Code kann die mittlere Codewortlänge als *Summe aller Zwischenknotenwahrscheinlichkeiten* (einschliesslich der Wurzel) abgelesen werden, da eine Codewortwahrscheinlichkeit in jeder Zwischenknotenwahrscheinlichkeit längs des zugehörigen Pfades von der Wurzel zum Endknoten als Summand enthalten ist; die Codewortwahrscheinlichkeit wird also in dieser Summe unmittelbar mit der Codewortlänge multipliziert. Ein *optimaler, präfixfreier Code mit variabler Wortlänge* besitzt damit *minimale Zwischenknotenwahrscheinlichkeiten*. Deshalb wird im Konstruktionsverfahren nach **Huffman** [6] der Baum ausgehend von den Endknoten so bestimmt, dass jeweils die M_c Knoten (End- oder inzwischen gebildete Zwischenknoten) mit den geringsten Wahrscheinlichkeiten zur Bildung eines neuen Zwischenknoten zusammengefasst werden.[5]

❑ *Beispiel:* $M_x = 6, M_c = 2$, siehe Abbildung 2.4

[5]Damit ab dem zweiten Zusammenfassungsschritt jeweils genau M_c Zweige zusammengefasst werden können, sind für $M_c > 2$ im ersten Schritt nur $(M_x - M_c)\mathrm{mod}(M_c - 1) + 1$ Zweige zur Bildung des ersten Zwischenknotens zu verwenden.



$\bar{n} = 1 + 0,6 + 0,4 + 0,3 + 0,15 = 2,45$; vgl. Entropie: $H(X) = 2,403$

Abbildung 2.4: Beispiel für eine Huffman Codierung

2.2.2 Quellenwörter mit variabler Länge

Eine zu Abschnitt 2.2.1 duale Vorgehensweise besteht darin, mittels Quellenwörter unterschiedlicher Länge l_i im Mittel möglichst viele Quellensymbole auf Codewörter mit fester Länge N abzubilden. Diese Verfahren zeigen weniger ausgeprägte Fehlerfortpflanzungseffekte infolge Codesymbolfehler als Codes mit variabler Wortlänge. Das **Quellencodierungstheorem** kann auch für diese Form der verlustlosen Datenkompression nachgewiesen werden; d.h. es existiert eine Quellenwortformung, bei der die mittlere Zahl $N/\bar{l}$ von Codesymbolen je Quellensymbol durch $H(X)/\mathrm{ld}(M_c) \leq N/\bar{l} < H(X)/\mathrm{ld}(M_c) + \varepsilon$ mit $\varepsilon \to 0$ für $N \to \infty$ eingegrenzt werden kann, wobei $\bar{l}$ die mittlere Zahl von Quellensymbolen je Quellenwort bezeichnet. (Quellencodierungstheorem für variable Quellenwort-, fixe Codewortlänge.)

2.2.2.1 Tunstall–Codierung

Im Verfahren nach **Tunstall** [7] wird ein Baum für Quellenwörter dual zur Konstruktion nach Huffman (vgl. Abschitt 2.2.1) entwickelt. Ausgehend von der Wurzel (Wahrscheinlichkeit 1) wird jeweils der Endknoten mit der größten Wahrscheinlichkeit durch *alle* M_x Zweige expandiert, wobei jeder Endknoten durch die Wahrscheinlichkeit des zugehörigen Quellenwortes charakterisiert wird. Das Verfahren wird fortgesetzt, bis K Endknoten vorliegen mit $M_c^N - M_x + 2 \leq K \leq M_c^N$. Da jeder Knoten jeweils durch *alle* M_x Zwei-

ge expandiert wird, ist jede Quellensymbolfolge in die so bestimmten Quellenwörter partitionierbar.

❑ *Beispiel:* Gedächtnislose Binärquelle $X \in \{A, B\}$, $\Pr(X = A) = 0,3$; $H(X) = 0,8813$; $M_x = 2$; $M_c = 2$; $N = 3$ (8 Codewörter), siehe Abbildung 2.5

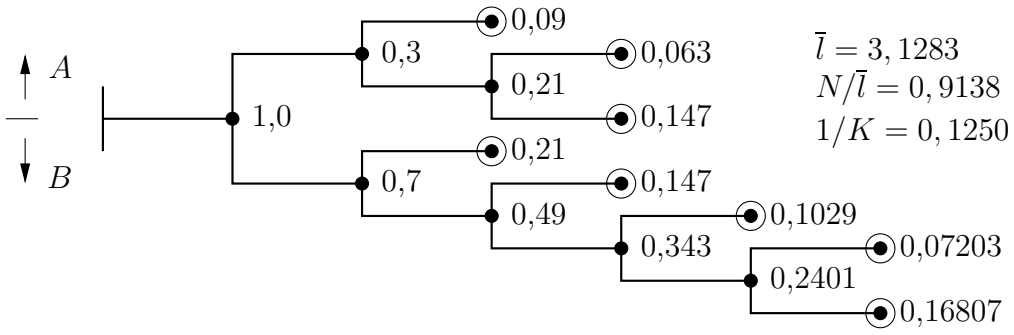


Abbildung 2.5: Baum für 8 Quellenwörter nach Tunstall

2.2.2.2 Universelle Quellencodierung nach Lempel–Ziv

Universelle Quellencodierverfahren zeichnen sich dadurch aus, dass keine a–priori Kenntnis über die Quelle benötigt wird; die Anpassung des Quellen-(bzw. Code-)wortbuches an die Quelle wird anhand der beobachteten Quellensymbolsequenz selbst vorgenommen. Beim Verfahren nach **Lempel** und **Ziv** [12] werden zu Beginn sende- und empfangsseitig Quellenwortbücher (für M_c^n Einträge) auf den ersten M_x Positionen mit den M_x Quellensymbolen initialisiert. Das aktuell längste Quellenwort im Quellenwortbuch, das dem nächsten Segment aus der Quellensymbolsequenz entspricht, wird anhand seiner Adresse übertragen und zugleich das nachfolgende Quellensymbol direkt dem Codewort angehängt, dessen Länge damit $N = n + \lceil \log_{M_c}(M_x) \rceil$ beträgt. Solange das Quellenwortbuch noch nicht vollständig gefüllt ist, wird das um das nachfolgende Symbol verlängerte Quellenwort anschließend beidseitig auf der nächsten freien Position eingetragen. Sind neben einem Quellenwort alle möglichen Erweiterungen um mindestens ein Symbol ebenfalls im Quellenwortbuch vorhanden, so kann es überschrieben werden. Es ist bewiesen, dass für $N \to \infty$ mit diesem universellen Verfahren (sowie mit zahlreichen anderen) die Entropie der Quelle von $N/\overline{l}$ beliebig nahe erreicht wird, also die

asymptotische Optimalität gegeben ist.

2.2.3 Weitere verlustlose Quellencodierverfahren

Bei der Zuordnung von Codewörtern mit fester Länge N zu Quellenwörtern mit fester Länge L (block–to–block encoding) werden nur für *typische Sequenzen* von Quellensymbolen Codewörter bereitgehalten. Andernfalls tritt ein *Codierversagen* $\mathcal{V}$ ein. Für die notwendige Zahl N/L von Codesymbolen je Quellensymbol gilt: $H(X)/\mathrm{ld}(M_c) \leq N/L < H(X)/\mathrm{ld}(M_c) + \varepsilon_1/L$, wobei für jedes $\varepsilon_1 > 0$ ein $\varepsilon_2 > 0$ existiert, mit $\Pr(\mathcal{V}) \leq \varepsilon_2/L$. (**Quellencodierungstheorem** für fixe Quellenwort-, fixe Codewortlänge) Mit wachsender Wortlänge L bzw. N wird die Entropie immer näher erreicht, wobei zugleich die Wahrscheinlichkeit $\Pr(\mathcal{V})$ für Codierversagen nach Null strebt; für sehr große Wortlängen treten also nur noch $2^{L \cdot H(X)}$ typische Sequenzen auf (vgl. Gesetz der großen Zahlen).

Neben den exemplarisch aufgeführten Basisverfahren existieren zahlreiche hochentwickelte Quellencodierverfahren, die viele praktische Anwendungen gefunden haben. Zum Beispiel erlaubt **arithmetische Codierung** (variable Codewortlänge) eine iterative Berechnung von Codewort und Quellenwort bei Codierung bzw. Decodierung. Es sind damit keine Codetabellen erforderlich, was die Anwendung sehr großer Wortlängen ermöglicht. Arithmetische Codierung dient zahlreichen, universellen Quellencodierverfahren als Basis; so z.B. dem **Context–Tree–Weighting Algorithm (CTW)** [8], bei dem eine Mittelung über sehr viele Modelle, welche eine gedächtnisbehaftete Quelle mehr oder weniger adäquat beschreiben, vorgenommen wird. Mittels CTW lassen sich insbesondere Texte extrem nahe an der Entropie komprimieren, vgl. [8]. Durch die **Burrows–Wheeler Transformation** [9] erfolgt eine Sortierung von Symbolsequenzen gemäß statistischer Abhängigkeiten innerhalb von Symbolsequenzen. Hieraus kann mittels **Differenzcodierung** das Verhalten einer nahezu gedächtnislosen Quelle mit stark unterschiedlichen Symbolwahrscheinlichkeiten erreicht werden, worauf wiederum übliche Quellencodierverfahren wie z.B. Huffman–Codierung erfolgreich anwendbar sind. Effiziente Quellencodierung erfolgt also häufig in mehreren Stufen.

2.3 Zuverlässige Informationsübertragung

Die Informationstheorie erlaubt grundsätzlich Aussagen zu Möglichkeiten und Grenzen der *zuverlässigen* Informationsübertragung über *gestörte Kanäle*. Als **Kanal** werden dabei die Teile eines Übertragungssystems definiert, die als unveränderlich vorgegeben sind.

2.3.1 Diskreter Übertragungskanal

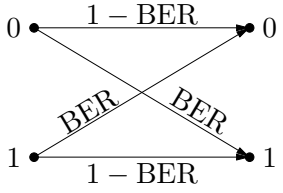
Übertragungskanäle werden als Quellen für zufällige Ausgangssymbole $Y[k]$ modelliert, wobei eine mehr (geringe Störung) oder weniger (starke Störung) statistische Abhängigkeit von der Sequenz $X[k]$, $k \in \mathbb{Z}$ von Kanaleingangssymbolen gegeben ist, vgl. Abschnitt 2.1.4 und Abbildung 2.6.

Abbildung 2.6: Diskreter Übertragungskanal

Ein Kanal wird als **gedächtnislos** bezeichnet, wenn diese statistische Abhängigkeit nur durch das gleichzeitige Eingangssymbol gegeben ist. Somit wird ein **diskreter gedächtnisloser Kanal** (discrete memoryless channel: **DMC**) vollständig durch die $M_x \times M_y$ **Übergangswahrscheinlichkeiten** $\Pr(Y[k] \mid X[k])$ spezifiziert. Liegt keine Abhängigkeit vom Zeitpunkt k vor, so ist der Kanal zudem **zeitinvariant**. Die Übergangswahrscheinlichkeiten eines zeitinvarianten DMC werden zu einer $M_x \times M_y$ Kanalmatrix $\mathbf{K} = (K_{ij})$ mit $K_{ij} = \Pr(Y = y_j \mid X = x_i), i = 1(1)M_x, j = 1(1)M_y$ zusammengefasst.

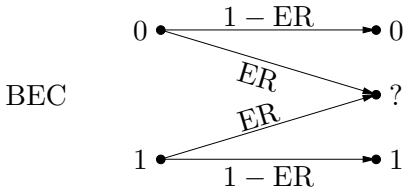
❑ *Beispiel:* Symmetrischer Binärkanal (binary symmetric channel (**BSC**)) $M_x = 2$, $M_y = 2$ mit Fehlerwahrscheinlichkeit (bit error ratio) **BER**:

BSC



$$\mathbf{K} = \begin{pmatrix} 1-\mathrm{BER} & \mathrm{BER} \\ \mathrm{BER} & 1-\mathrm{BER} \end{pmatrix}$$

❑ *Beispiel:* Symmetrischer binärer Auslöschungskanal (binary erasure channel (**BEC**))$M_x = 2$, $M_y = 3$ mit Auslöschungswahrscheinlichkeit (erasure ratio) **ER**:



$$\mathbf{K} = \begin{pmatrix} 1-\mathrm{ER} & \mathrm{ER} & 0 \\ 0 & \mathrm{ER} & 1-\mathrm{ER} \end{pmatrix}$$

2.3.2 Kanalkapazität, symmetrische Kanäle

Der mittlere ausgangsseitige Informationsgewinn bzgl. der Eingangsvariable X je Kanalbenutzung wird gemäß Abschnitt 2.1.4 durch die wechselseitige Information $I(X;Y)$ erfasst. Wird diese Größe über alle Einflussgrößen, die nicht durch den Kanal gegeben und somit veränderbar sind, maximiert, so erhält man die **Kapazität** C eines Kanals: Speziell bei der rückwirkungsfreien Übertragung über gedächtnislose Kanäle ist lediglich die Verteilung $\Pr(X)$ der Eingangssymbole veränderlich.

Def: Kanalkapazität des gedächtnislosen Kanals: $C \stackrel{\triangle}{=} \max\limits_{\Pr(X)} I(X;Y)$

❑ *Beispiel:*

$$\left.\begin{array}{ll} \text{BSC:} & C = 1 - \mathrm{e}_2(\mathrm{BER}) \\ \text{BEC:} & C = 1 - \mathrm{ER} \end{array}\right\} \text{für } \Pr(X=0) = \Pr(X=1) = 1/2$$

Ein Kanal wird als **symmetrisch** bezeichnet, wenn sich die M_y Spaltenvektoren von $\mathbf{K}$ so auf V Teilmatrizen $\mathbf{S}_v$, $v = 1(1)V$ verteilen lassen, dass innerhalb der Matrizen $\mathbf{S}_v$ sowohl Zeilen- als auch Spaltenvektoren jeweils durch Permutation der Elemente auseinander hervorgehen, die Matrizen $\mathbf{S}_v$ also *doppelt uniform* sind. Eine solche Aufteilung entspricht der Auswahl eines *streng symmetrischen Teilkanals* $\frac{1}{w_v}\mathbf{S}_v$ mit der Wahrscheinlichkeit w_v unabhängig vom Eingangssymbol X, wobei w_v irgendeine Zeilensumme aus $\mathbf{S}_v$ bezeichnet.

Satz: Bei symmetrischen Kanälen $\mathbf{K}$ wird die Kapazität durch gleichverteilte Eingangssymbole, d.h. $\Pr(X = x_i) = \frac{1}{M_x}$, $\forall i = 1(1)M_x$, erreicht. Diese Kapazität ist durch den Mittelwert $C = \sum_{i=1}^{V} w_i C_i$ der Kapazitäten C_i der streng symmetrischen Teilkanäle $\frac{1}{w_v}\mathbf{S}_v$, in die der Kanal aufteilbar ist, gegeben.

2.3.3 Prinzip der Kanalcodierung

Zuverlässige Informationsübertragung über gestörte Kanäle lässt sich durch **Kanalcodierung** erreichen, indem durch gezieltes Einbringen von Redundanz in den Datenstrom die Unterscheidbarkeit von Codewörtern erhöht wird; Kanalcodierung ist somit das zur Quellencodierung duale Prinzip der Informationsverarbeitung. Bei einer Kanalcodierung mit **Blockcodes** werden im Coder binären **Datenwörtern** der Länge k **Codewörter** der Länge n zugeordnet, siehe Abbildung 2.7

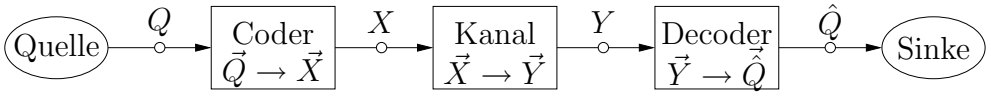


Abbildung 2.7: Prinzip der Kanalcodierung (Blockcodierung)

Datenwort:	$\vec{Q} = (Q_1, \ldots, Q_k)$	$\in \{0,1\}^k$
Codewort:	$\vec{X} = (X_1, \ldots, X_n)$	$\in \mathbf{C} \subset \mathcal{X}^n$
Beobachtetes Wort:	$\vec{Y} = (Y_1, \ldots, Y_n)$	$\in \mathcal{Y}^n$
Geschätztes Datenwort:	$\hat{\vec{Q}} = (\hat{Q}_1, \ldots, \hat{Q}_k)$	$\in \{0,1\}^k$

O.B.d.A. wird eine binäre i.i.d. Quellensymbolsequenz mit $H(Q) = 1$ vorausgesetzt.

Def: Kanalcode: Als Kanalcode $\mathbf{C}$ wird die Teilmenge von Wörtern $\vec{X}$ bezeichnet, $\mathbf{C} \subset \mathcal{X}^n$, zu denen bzgl. der Zuordnung im Codierungsvorgang Urbilder aus der Menge der 2^k Quellenwörter gehören. Der *mittlere Informationsgehalt je Codesymbol* wird als die **Rate** R des Codes bezeichnet.

Bei umkehrbar eindeutiger Codierung und für $H(Q) = 1$ sowie $n \geq k/\mathrm{ld}(M_x)$ gilt:

$$R = \frac{\mathrm{ld}\,(|\mathbf{C}|)}{n} = \frac{k}{n}, \text{ bzw. } |\mathbf{C}| = 2^k = 2^{nR} \tag{2.11}$$

Die *mittlere Redundanz* je Codesymbol beträgt damit $\rho \stackrel{\triangle}{=} \mathrm{ld}(M_x) - R$. Die Störresistenz der digitalen Übertragung wird dadurch erreicht, dass nur sehr wenige Wörter $\vec{X}$ aus $\mathcal{X}^n$ als zugelassene Codewörter für die Übertragung verwendet werden, wobei mit wachsender Codewortlänge n dieser relative Anteil $2^{nR}/M_x^n = 2^{n(R-\mathrm{ld}(M_x))} = 2^{-n\rho}$ für $\rho > 0$ exponentiell nach Null strebt.

- *Beispiel:* $M_x = 2$ (Binärcode); $R = 0,6$; $\rho = 0,4$
  $n{=}10$: $2^{-n\rho} = 0,0625$; $n{=}100$: $2^{-n\rho} = 9 \cdot 10^{-13}$; $n{=}1000$: $2^{-n\rho} = 4 \cdot 10^{-121}$

Bei konstanter Rate (Redundanz) nimmt mit wachsender Codewortlänge also die Wahrscheinlichkeit, dass durch den Kanal ein Codewort in ein anderes verfälscht wird, rasch ab; eine *Fehlererkennung* wird damit extrem sicher. Bei *Fehlerkorrekturverfahren* (**FEC**: Forward Error Correction) wird im Decoder in optimaler Weise dasjenige Datenwort bestimmt, dessen zugeordnetes Codewort nach der Beobachtung der Kanalausgangssymbolfolge $\vec{Y}$ (a–posteriori) die größte Wahrscheinlichkeit besitzt: *Maximum–A–Posteriori (MAP–) Decodierung* (vgl. Abbildung 2.7).

$$\mathrm{MAP} : \vec{\hat{X}} = \underset{\vec{x} \in \mathbf{C}}{\operatorname{argmax}} \left\{ \Pr\left(\vec{X} = \vec{x} \mid \vec{Y} \right) \right\}$$

Falls, wie hier vorausgesetzt, alle Codewörter a–priori gleichwahrscheinlich sind, ist dies einer *Maximum–Likelihood (ML-) Decodierung* gleichwertig:

$$\mathrm{ML} : \vec{\hat{X}} = \underset{\vec{x} \in \mathbf{C}}{\operatorname{argmax}} \left\{ \Pr\left(\vec{Y} \mid \vec{X} = \vec{x} \right) \right\}$$

Bei bekannter Kanalmatrix ist hierzu der Raum $\mathcal{Y}^n$ der möglichen Beobachtungsvektoren $\vec{Y}$ in 2^{nR} Entscheidungsgebiete $\mathbf{G}_i$ zu partitionieren mit

$$\mathbf{G}_i = \left\{ \vec{y} \,\middle|\, \Pr\left(\vec{Y} = \vec{y} \mid \vec{X} = \vec{x}_i \right) \geq \Pr\left(\vec{Y} = \vec{y} \mid \vec{X} = \vec{x}_j \right) \; \forall j \neq i \right\}$$

und die Entscheidungsregel: $\vec{\hat{X}} = \vec{x}_i$, falls $\vec{Y} \in \mathbf{G}_i$, anzuwenden. Die mittlere Wahrscheinlichkeit für eine Entscheidung zu einem falschen Codewort wird als **Wortfehlerwahrscheinlichkeit** $p_\mathrm{w} \triangleq \Pr(\vec{\hat{X}} \neq \vec{X})$ bezeichnet. Kanalcodierung dient infolge der dünnen Besetzung des Raumes $\mathcal{X}^n$ durch Codewörter der **Vermeidung von Entscheidungsfehlern**. Nur im Spezialfall gleicher Symbolalphabete an Kanalein- und ausgang ($\mathcal{X} = \mathcal{Y}$) liegt eigentlich eine Kanalcodierung **zur Fehlerkorrektur** (FEC) vor.

2.3.4 Gallager–Schranke, Kanalcodierungstheorem

Die Wahrscheinlichkeit $p_\mathrm{w} = \Pr(\vec{\hat{X}} \neq \vec{X})$ für die Decodierung zu einem anderen als dem gesendeten Codewort kann für lange Codes kaum analytisch

berechnet werden, zumal gute Codes primär nicht bekannt sind. Es gelingt aber, Schranken für den Erwartungswert $\mathrm{E}\{p_\mathrm{w}\}$ bei *zufälliger Wahl* der Codewörter zu finden. Werden dabei die Codewörter aus wechselseitig statistisch unabhängigen Codesymbolen gemäß einer a–priori Verteilung $\Pr(X)$ gebildet, so ist eine solche obere Schranke bei gedächtnislosen, zeitinvarianten Kanälen abhängig von den Elementen der Kanalmatrix $\mathbf{K}$, der Codewortlänge n und der Coderate R nach Gallager [10] gegeben durch:

$$\mathrm{E}\{p_\mathrm{w}\} \leq 2^{-nE_\mathrm{G}(R)} \text{ mit dem } \textit{Fehlerexponenten} \tag{2.12}$$

$$E_\mathrm{G}(R) = \max_{\rho\in(0,1]} \max_{\Pr(X)} \left(E_0(\rho, \Pr(X)) - \rho R\right) \text{ und} \tag{2.13}$$

$$E_0(\rho, \Pr(X)) \stackrel{\triangle}{=} -\mathrm{ld}\left[\sum_{\ell=1}^{M_y}\left(\sum_{i=1}^{M_x}[\Pr(Y=y_\ell \mid X=x_i)]^{\frac{1}{1+\rho}} \cdot \Pr(X=x_i)\right)^{1+\rho}\right] \tag{2.14}$$

$$R_0 \stackrel{\triangle}{=} \max_{\Pr(X)} E_0(\Pr(X), \rho=1) \tag{2.15}$$

Für symmetrische Kanäle werden die *Gallager–Funktion* $E_0(\rho, \Pr(X))$ und die *Cut–Off Rate* R_0 wie die wechselseitige Information durch *gleichverteilte* Codesymbole maximiert. (In der Abschätzung (2.12) wird nicht ausgeschlossen, dass unterschiedlichen Quellenwörtern (zufällig) auch identische Codewörter zugewiesen werden.) Für jeden Parameter $\rho \in (0, 1]$ repräsentiert Gl. (2.13) eine Gerade; deren Einhüllende ergibt den **Fehlerexponenten**, vgl. Abbildung 2.8.

Die hohe Zuverlässigkeit bei großen Codewortlängen ist darauf zurückzuführen, dass sich die zufälligen Codewörter infolge der immer dünner werdenden Besetzung mit hoher Wahrscheinlichkeit zunehmend regelmäßig bei wechselseitig großen Abständen im Raum $\mathcal{X}^n$ anordnen. Da die Schranke (2.12) als Mittelwert über ein Ensemble von vielen Codes interpretiert werden kann, ist damit auch die Existenz zumindesten eines Kanalcodes bewiesen, für den diese Ungleichung erfüllt wird. Damit ist gezeigt, dass mittels redundanter Kanalcodierung mit hinreichend großer Wortlänge n eine

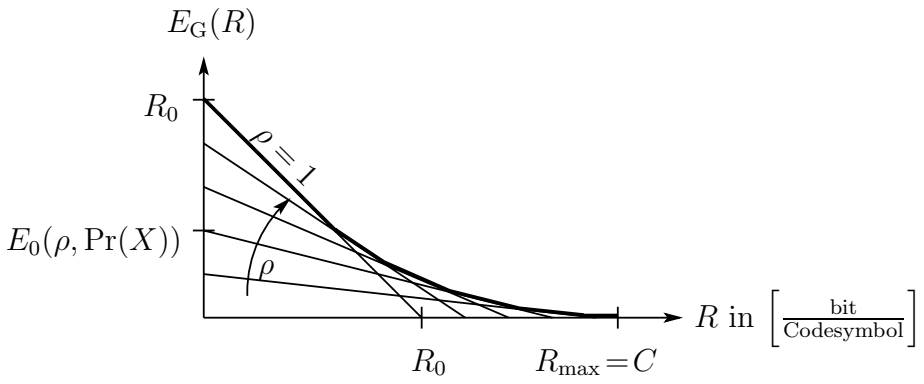


Abbildung 2.8: Gallager Fehlerexponent

hoch zuverlässige digitale Übertragung ($p_w \to 0$) möglich ist, solange gilt: $E_G(R) > 0$. Somit stellt die Coderate $R_{\max}$, bei welcher der Fehlerexponent mit horizontaler Tangente nach Null strebt, die Grenze dar, unterhalb derer eine zuverlässige digitale Übertragung über den gestörten Kanal prinzipiell möglich ist. Es gilt jedoch:

$$R_{\max} = \max_{\Pr(X)} \lim_{\rho \to 0} \frac{E_0(\rho, \Pr(X))}{\rho} = \max_{\Pr(X)} \lim_{\rho \to 0} \frac{\mathrm{d}}{\mathrm{d}\rho} E_0(\rho, \Pr(X)) = C \qquad (2.16)$$

Dieses Resultat wird im zweiten fundamentalen Theorem der Informationstheorie zusammengefasst:

Kanalcodierungstheorem: Über einen (gestörten) Übertragungskanal mit der Kapazität $C = \max_{\Pr(X)} I(X;Y)$ [bit/Kanalbenutzung] ist eine digitale Übertragung mit beliebig hoher Zuverlässigkeit mittels einer redundanten Kanalcodierung möglich, solange der mittlere Informationsgehalt je Codesymbol, die Coderate R [bit/Codesymbol], kleiner ist als die Kanalkapazität C und die Codewortlänge hinreichend groß gewählt wird ($R < C$). Es existieren ein solcher Code und eine geeignete Decodiermethode.

Der Existenzbeweis enthält allerdings keinerlei Hinweise auf Konstruktionsverfahren für praktikable Codes, da die dem Beweis zugrundeliegenden Tabellen für Codierung und Decodierung exponentiell mit der Wortlänge anwachsen würden. Es bedurfte etwa 45 Jahren wissenschaftlicher Entwicklung der Kanalcodierungstheorie, bis Codes gefunden waren, mit denen die Kanalkapazität auch in praktischen Anwendungen nahezu erreicht ist.

Aus der Schranke (2.12) kann darüberhinaus die maximale Rate bei endlicher Codewortlänge und tolerierbarer Wortfehlerwahrscheinlichkeit p_{w} abgeschätzt werden. Dabei erweist sich meist der Fehlerexponent $E_{\mathrm{G}}(R_0)$ als groß genug, um bei technisch gut handhabbaren Wortlängen ($n \lesssim 1000$) eine ausreichende Zuverlässigkeit ($p_{\mathrm{w}} \leq 10^{-4}$) zu ermöglichen. Deshalb wird die Cut–Off Rate R_0 oft auch als die *Kanalkapazität der Praxis* bezeichnet.

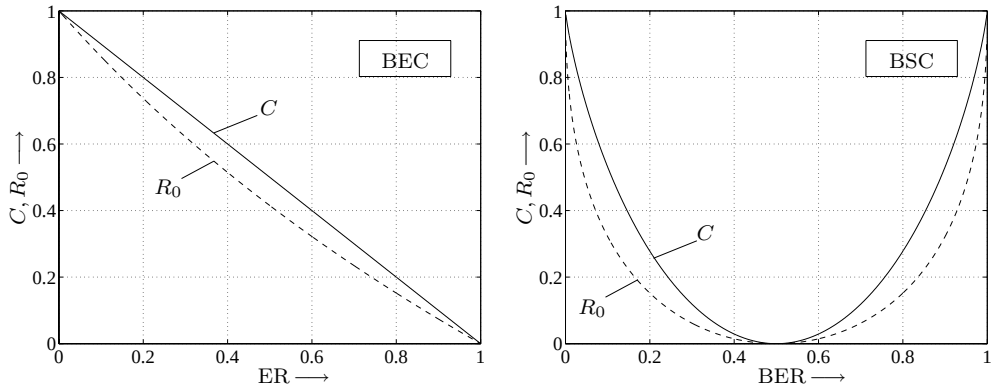


Abbildung 2.9: Cut–Off Rate und Kapazität für BEC und BSC

Umkehrung des Kanalcodierungstheorems:
Für einen Kanal mit der Kapazität C gibt es **kein** digitales Übertragungsverfahren mit einer Rate

$$R > C/(1 - \mathrm{e}_2(\varepsilon_{\mathrm{T}})) \quad [\text{bit/Kanalbenutzung}], \tag{2.17}$$

durch das vom Sendereingang zum Empfängerausgang eine geringere, mittlere Bitfehlerwahrscheinlichkeit als $\varepsilon_{\mathrm{T}} \leq 0,5$ erreichbar wäre.

Damit ist absolut fehlerfreie digitale Übertragung prinzipiell nicht möglich, wenn die Rate R der Eingangssequenz die Kanalkapazität C übersteigt, d.h. wenn dem Kanal im Mittel mehr Information zugeführt wird als dieser zu übertragen im Stande ist. Das Kanalcodierungstheorem und seine Umkehrung bestätigen erneut die Relevanz des Shannon'schen Informationsmaßes für die Praxis.

2.4 Kontinuierliche Zufallsvariablen

2.4.1 Differentielle Entropie

Die Definitionen für Information und Entropie für diskrete Symbolalphabete ist für eine kontinuierlich verteilte Zufallsvariable X mit einer a–priori Wahrscheinlichkeitsdichtefunktion (WDF) $f_X(x)$ nicht anwendbar, da diese Größen hier gegen ∞ streben. Es sind deshalb *Hilfsgrößen* gebräuchlich:

Def: Differentielle Entropie:

$$h(X) \stackrel{\triangle}{=} -\int_{-\infty}^{\infty} f_X(x) \cdot \mathrm{ld}\left(f_X(x)\right)\, \mathrm{d}x$$

Def: Bedingte differentielle Entropie:

$$h(X \mid Y) \stackrel{\triangle}{=} -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y) \cdot \mathrm{ld}\left(f_X(x \mid y)\right)\, \mathrm{d}x\, \mathrm{d}y$$

▶ *Hinweis:* Die differentielle Entropie ist kein Informationsmaß im eigentlichen Sinne; die Einheit bit darf hierfür nicht verwendet werden.

❑ *Beispiel:* Gleichverteilte Zufallsvariable $f_X(x) = 1/(2a)$ für $|x| \leq a$:

$$h(X) = \mathrm{ld}(2a) = \frac{1}{2}\mathrm{ld}(12\sigma_x^2) \tag{2.18}$$

Bei Begrenzung des Betrags einer Zufallsvariable, $|X| \leq a$, wird die maximale differentielle Entropie durch die (mittelwertfreie) *Gleichverteilung* erreicht.

❑ *Beispiel:* Gaußverteilte Zufallsvariable $f_X(x) = \mathrm{e}^{-x^2/(2\sigma_x^2)}/\sqrt{2\pi\sigma_x^2}$

$$h(X) = \frac{1}{2}\mathrm{ld}\left(2\pi\mathrm{e}\sigma_x^2\right) \tag{2.19}$$

Bei gegebener Varianz σ_x^2 wird die maximale differentielle Entropie durch die *Gaußverteilung* erreicht. Damit ist bei gegebener mittlerer Leistung eine mittelwertfreie Gauß'sche Zufallsvariable die "zufälligste". Weißes Gauß'sches

Rauschen ist der zufälligste aller Zufallsprozesse bzw. die "am meisten störende" Störung.

Für die wechselseitige Information zwischen zwei kontinuierlichen Zufallsvariablen X und Y gilt:

$$I(X;Y) = h(X) - h(X \mid Y) = h(Y) - h(Y \mid X) \quad \left[\frac{\text{bit}}{\text{Wertepaar}}\right] \tag{2.20}$$

Die *Differenz* zweier diff. Entropien ergibt eine wechselseitige Information im *üblichen Sinn* des Shannon'schen Informationsmaßes.

2.4.2 Kapazität des AWGN–Kanals

Bei einem Kanal, dessen Ausgangsvariable durch Addition einer von den Eingangsvariablen X unabhängigen Störvariablen Z entsteht, gilt gemäß Gl. (2.20) $I(X;Y) = h(Y) - h(Z)$. Bei gegebener Störvarianz $\sigma_z^2 \stackrel{\triangle}{=} N$ bewirkt somit eine Gauß'sche Störvariable Z eine Minimierung der wechselseitigen Information (*„worst–case"–Störung*). Dieses Modell wird als **diskreter AWGN–Kanal** (additive white Gaussian noise) bezeichnet. Bei vorgegebener Varianz $\sigma_x^2 \stackrel{\triangle}{=} S$ des Nutzsignals wird $I(X;Y)$ wiederum durch eine Gauß'sche Eingangsvariable X maximiert, da Y dann ebenfalls eine Gauß'sche Zufallsvariable darstellt und somit $h(Y)$ bei gegebener Varianz $S + N$ maximiert wird. Aus Gl. (2.19), (2.20) folgt damit

Satz: Kapazität des zeitdiskreten AWGN–Kanals:

$$C = \max_{f_X(x)} I(X;Y) = \frac{1}{2}\text{ld}\,(1 + S/N) \quad \left[\frac{\text{bit}}{\text{Kanalbenutzung}}\right] \tag{2.21}$$

Die Gleichungen (2.21) und (2.22) sind von grundlegender Bedeutung, da durch sie eine Verbindung zwischen der abstrakten Information und physikalischen Größen wie Leistung bzw. Energie, damit auch Materie, hergestellt wird. (vgl. auch den Beweis der Nichtexistenz des sog. Maxwellschen Dämons zur Umgehung des zweiten Hauptsatzes der Thermodynamik.)

Das Abtasttheorem der Systemtheorie besagt, dass ein *zeit- und wertkontinuierliches Signal* $x(t)$ mit einem (Fourier-)Spektrum $X(f) = \int_{-\infty}^{\infty} x(t)\text{e}^{-j2\pi ft}\,\text{d}t$, das nur im Intervall $(-B, +B)$ wesentlich von Null verschieden ist, vollständig

durch dessen (Abtast-)Werte $x(kT)$, $k \in \mathbb{Z}$ repräsentiert wird, solange für die *Abtastfrequenz* gilt: $1/T \geq 2B$. Umgekehrt sind damit bei Begrenzung eines informationstragenden Signals auf eine (einseitige Fourier-)Bandbreite B höchstens $2B$ Signalwerte je Sekunde frei wählbar. Somit existieren unter dieser Beschränkung max. $2B$ unabhängige Kanalbenutzungen gemäß dem zeitdiskreten Modell. Bei Störung durch (zeitkontinuierliches) weißes Gauß'sches Rauschen mit einer (einseitigen) Rauschleistungsdichte $\mathcal{N}_0$ ergibt sich infolge der Bandbegrenzung für die Störvarianz $N = \mathcal{N}_0 B$. Damit erhält man die berühmteste Gleichung der Informationstheorie:

Satz: Kapazität des zeitkontinuierlichen AWGN–Kanals:

$$C_{\mathrm{T}} = B \cdot \mathrm{ld}\left(1 + \frac{S}{N}\right) \left[\frac{\mathrm{bit}}{\mathrm{s}}\right], \text{ mit } S = \mathrm{E}\left\{|x(t)|^2\right\}, \; N = \mathcal{N}_0 B. \qquad (2.22)$$

2.4.3 Leistungs–Bandbreiten–Diagramm

Digitale Übertragungsverfahren werden hinsichtlich ihrer *Leistungseffizienz* dadurch gekennzeichnet, welcher Quotient $E_{\mathrm{b}}/\mathcal{N}_0$ beim AWGN–Kanal ausreicht, um eine hinreichende Datensicherheit zu erreichen. Dabei bezeichnen E_{b} die (mittlere) Energie des Nutzsignals am Empfängereingang je bit übertragener Information und $\mathcal{N}_0$ die (einseitige) Rauschleistungsdichte. Bei der Übertragung eines *Informationsflusses* (Nutzdatenrate) F [bit/s] gilt also $E_{\mathrm{b}} = S/F$, wobei S die mittlere Leistung des Empfangsnutzsignals darstellt. Die *Bandbreiteneffizienz* eines digitalen Übertragungsverfahrens wird durch den Quotienten F/B charakterisiert und gibt den je Hz (Fourier-)Signalbandbreite B übertragbaren Informationsfluss F an[6]. Insbesondere bei der drahtlosen Informationsübertragung mittels Funktechnik stellt die zur Verfügung stehende (Fourier-)Bandbreite eine äußerst knappe, nicht vermehrbare Ressource dar, so dass eine schnelle Datenübertragung eine hohe Bandbreiteneffizienz $F/B \gg 1$ des digitalen Übertragungsverfahrens erforderlich macht. Leistungs- und Bandbreiteneffizienz sind wechselseitig austauschbare Parameter, wobei für die digitale Kommunikation die

[6]Deshalb sollte man den Begriff „Bandbreite“ für den Informationsfluß, wie häufig populärwissenschaftlich üblich, tunlichst vermeiden!

Shannongrenze:

$$E_\mathrm{b}/\mathcal{N}_0 \geq \frac{B}{F}\left(2^{F/B}-1\right) \tag{2.23}$$

gilt. Gl. (2.23) folgt unmittelbar aus Gl. (2.22). Die Abbildung 2.10 zeigt das hieraus resultierende *Leistungs–Bandbreiten–Diagramm* der digitalen Übertragung. Im Bereich 1 können Übertragungsverfahren ohne Einsatz von Ka-

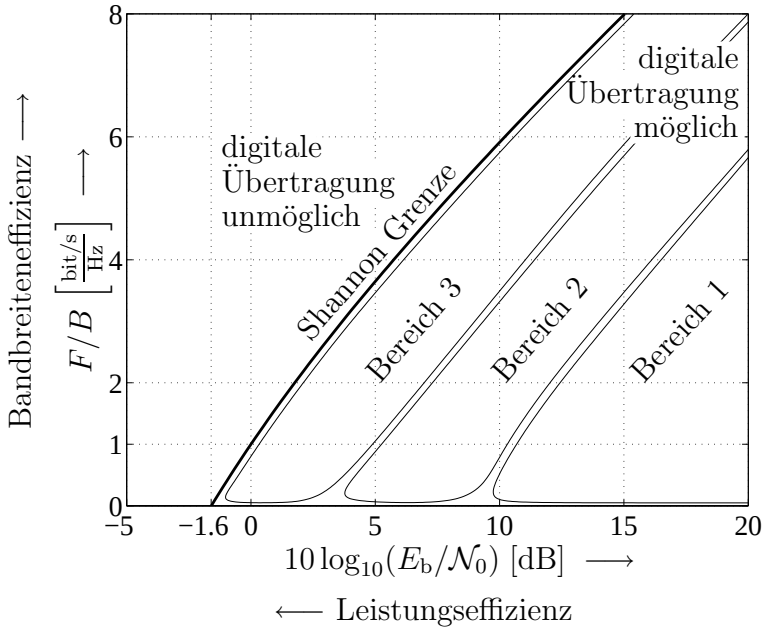


Abbildung 2.10: Leistungs–Bandbreiten–Diagramm der digit. Übertragung

nalcodierungsverfahren bei mäßiger Zuverlässigkeit und geringer Verzögerung des Datenstroms realisiert werden. Im Bereich 2 ist der Einsatz relativ einfacher Kanalcodierungsverfahren (z.B. Trelliscodierung, vgl. Abschnitt 2.5.3) bei mäßiger Verzögerung des Datenstroms (bis ca. 100 Symbole) erforderlich. Verfahren im Bereich 3 nahe an der Shannongrenze erfordern die Anwendung sehr langer Kanalcodes mit Decodierverfahren, bei denen a–posteriori Wahrscheinlichkeiten verrechnet werden. Zusätzlich müssen sog. Signalformungsmaßnahmen zur Erzeugung einer Gauß'schen Verteilung des Sendesignals ergriffen werden, die gemäß Abschnitt 2.4.2 zur Kapazität führt. Der Abstand

von ca. 10 bis 13 dB zwischen Bereich 1 und der Shannongrenze (Faktor 10 bis 20 in der erforderlichen Sendeleistung) ist durch komplexe Signalverarbeitung und die Hinnahme einer ansteigenden strukturellen Verzögerung des Datenstromes weitgehend überwindbar.

2.5 Einführung zur Kanalcodierung

Es stehen heute sehr leistungsfähige praktikable Kanalcodierungsverfahren zur Verfügung (vgl. [12]). Bei relativ geringer Verzögerung des Datenstroms (< 100 Symbole) zeigen Faltungs– bzw. Trelliscodes ganz hervorragende Eigenschaften. Kann sehr große Datenverzögerung zugelassen werden, so lassen sich mit Turbo– oder LDPC–Codes Existenzschranken erstaunlich nahe erreichen. Zudem ist bei vielen Anwendungen eine serielle Verkettung von Kanalcodierungsverfahren mit äußeren Reed–Solomon–Codes zu finden.

Leistungsfähige Kanalcodierungsverfahren erfordern algorithmische Verfahren für Codierung und Decodierung, da die Länge von Tabellen für Codewörter exponentiell mit der Codewortlänge anwachsen würde. Deshalb werden als Codesymbole Elemente von Mengen mit algebraischen Strukturen, vorzugsweise finiter Körper (Galoisfelder), verwendet, vgl. [14]. Besondere Bedeutung kommt dabei dem binären Körper $\mathbb{F}_2 = (\{0,1\};\ \oplus : \text{XOR};\ \odot : \text{AND})$ und dessen Erweiterungen $\mathbb{F}_{2^l},\ l \in \mathbb{N}$ zu. Die Elemente von $\mathbb{F}_{2^l}$ können als Reste von Polynomen bzgl. eines irreduziblen Polynoms $a(x)$ vom Grad l interpretiert und somit als l–Tupel über $\mathbb{F}_2$ notiert werden.

2.5.1 Lineare Blockcodes, zyklische Codes

Informations- und Codesymbole seien Elemente eines Körpers $\mathbb{F}$. Ein linearer Code $\mathbf{C}$ bildet einen k–dimensionalen Unterraum des Raumes $\mathbb{F}^n$:

$$\mathbf{C} = \{\, \vec{c} = (c_0, \ldots, c_{n-1}) = \vec{m} \odot \mathbf{G} |\ \forall \vec{m} = (m_0, \ldots, m_{k-1}), c_i, m_j, G_{j,i} \in \mathbb{F}\}$$

Die $k \times n$ Matrix $\mathbf{G}$ vom Rang k wird als *Generatormatrix* des Codes bezeichnet. Wird $\mathbf{G}$ mittels elementarer Transformationen auf Smith'sche Normalform gebracht, so werden die k *Informationssymbole* m_j direkt auf k Codesymbole abgebildet. Es liegt dann eine *systematische Codierung* zum

Code $\mathbf{C}$ vor. Eine Basis des Nullraumes zu $\mathbf{G}$ bildet die Zeilenvektoren einer $(n-k) \times n$ Matrix $\mathbf{H}$, die als Prüfmatrix zu $\mathbf{C}$ bezeichnet wird, da alle Codewörter der Bedingung

$$\vec{c} \odot \mathbf{H}^{\mathsf{T}} \stackrel{\triangle}{=} \vec{s} = \vec{0} \tag{2.24}$$

genügen. Verfahren zur *Fehlererkennung* beruhen auf diesem Test. Da der *Syndromvektor* $\vec{s}$ nur vom Fehlermuster $\vec{f}$ abhängt, das sich infolge Übertragungsfehler dem gesendeten Codewort $\vec{c}$ überlagert, $\vec{y} = \vec{c} \oplus \vec{f}$, wird bei algebraischen Verfahren zur *Fehlerkorrektur*[7] derjenige Fehlervektor $\hat{\vec{f}}$ vom *Empfangsvektor* $\vec{y}$ subtrahiert, der bei geringster Fehlerzahl zum Syndromvektor $\vec{s} = \vec{y} \odot \mathbf{H}^{\mathsf{T}}$ passt:

$$\hat{\vec{f}} = \underset{\forall \vec{x} \in \mathbb{F}^n | \vec{x} \odot \mathbf{H}^{\mathsf{T}} = \vec{s}}{\operatorname{argmin}} \{\text{weight}(\vec{x})\},$$

mit weight($\vec{x}$): Anzahl der Elemente $x_i \neq 0$ in $\vec{x}$ (*Gewicht* von $\vec{x}$). Falls $n-k$ hinreichend klein ist, können die Fehlermuster $\hat{\vec{f}}$ für alle Syndromvektoren $\vec{s}$ tabelliert werden.

Analog zur z–Transformation werden in der Codierungstheorie Wörter der Länge n synonym zur Vektordarstellung $(x_0, \ldots, x_{n-1})$ auch durch Polynome $x(D)$ über einer Dummy–Variablen D (delay) notiert: $x(D) \stackrel{\triangle}{=} \sum_{i=0}^{n-1} x_i D^i$ ($D \stackrel{\triangle}{=} z^{-1}$), wobei bzgl. der Koeffizienten die Operationen in $\mathbb{F}$ anzuwenden sind. Bei *zyklischen linearen Codes* sind alle zyklischen Verschiebungen eines Codewortes $c(D)$ ebenfalls Codewörter: $c(D) \odot D^i \bmod (D^n \ominus 1) \in \mathbf{C}$; $\forall i \in \mathbb{Z}$. Aufgrund der Gruppeneigenschaft linearer Codes sind die Codewörter zyklischer Codes als Vielfache eines *Generatorpolynoms* $g(D)$ vom Grad $n-k$ darstellbar:

$$\mathbf{C} = \{c(D) = m(D) \odot g(D) \mid \forall m(D) \text{ mit } \deg(m(D)) \leq k-1\}$$

Es gilt: Das Generatorpolynom $g(D)$ ist Teiler des *Hauptpolynoms* $D^n \ominus 1$ d.h. $D^n \ominus 1 = g(D) \odot h(D)$, wobei $h(D)$ als *Prüfpolynom* bezeichnet wird.

2.5.2 BCH– und RS–Codes

Für die zyklischen Blockcodes der Klasse der *Bose–Chauduri–Hocquenghem (BCH–) Codes* bzw. *Reed–Solomon (RS–) Codes* ist eine Korrektur von Feh-

[7] Es wird also auch $y_i \in \mathbb{F}$ vorausgesetzt.

lermustern mit einem maximalen Gewicht von $(d_{\min}-1)/2$ durch Lösen eines linearen Gleichungssystems mit Toeplitz–Struktur algebraisch sehr effizient möglich. Dabei bezeichnet $d_{\min}$ die minimale Hammingdistanz des Codes, d.h. die Zahl von Codesymbolen, in denen sich zwei verschiedene Codewörter mindestens unterscheiden. Aufgrund der Gruppeneigenschaft linearer Codes entspricht die Mindestdistanz dem *Mindestgewicht* aller Codewörter $\neq \vec{0}$.

$$d_{\min} = \min\{\text{weight}(\vec{c}_1 \ominus \vec{c}_2)\} = \min\{\text{weight}(\vec{c})\} \quad \forall \vec{c}_1 \ominus \vec{c}_2 = \vec{c} \in \mathbf{C}\backslash\{\vec{0}\}$$

Ein **binärer BCH–Code**, $c_i \in \mathbb{F}_2$, der Länge n kann mit Hilfe einer *diskreten Fouriertransformation*

$$(x_0, \ldots, x_{n-1}) \circ\!\!-\!\!\bullet\, (X_0, \ldots, X_{n-1}), \quad x_l \in \mathbb{F}_2,\ X_j \in \mathbb{F}_{2^l}$$
$$X_j \stackrel{\triangle}{=} \bigoplus\!\!\!\!\!\!\sum_{i=0}^{n-1} x_i w^{ij} = x(w^j); \quad x_i = \bigoplus\!\!\!\!\!\!\sum_{j=0}^{n-1} X_j w^{-ij} = X(w^{-i})$$

definiert werden, wobei w ein Element der Ordnung $n = \min_{k\in\mathbb{N}}\{k \mid w^k = 1\}$ im Erweiterungskörper $\mathbb{F}_{2^l}$ darstellt. Die Codewortlänge n ist also Teiler von $2^l - 1$; häufig wird für w ein *primitives Element* aus $\mathbb{F}_{2^l}$ gewählt, also $n = 2^l - 1$. Der Code mit Mindestdistanz $d_{\min}$ ist als Menge aller Vektoren über $\mathbb{F}_2$, deren *Spektren* $d_{\min} - 1$ aufeinanderfolgende Nullstellen aufweisen, darstellbar; z.B.

$$\mathbf{C} = \{(c_0, \ldots, c_{n-1}) \mid c_i \in \mathbb{F}_2, \text{ mit } C_{n-d_{\min}+j} = 0,\ j = 1(1)\mathrm{d}_{\min} - 1\} \quad (2.25)$$

Da ein Spektralpolynom $C(D) \neq 0$ somit maximal den Grad $n - d_{\min}$ besitzt, können in einem Vektor $\vec{c} \neq \vec{0}$ höchstens $n - d_{\min}$ Symbole zu Null werden (Nullstellen von $C(D)$); damit ist das Mindestgewicht $d_{\min}$ garantiert. Die Forderung (2.25) wird erfüllt, wenn für das Generatorpolynom gilt: $g(D = w^{n-j}) = 0$ für $j = 1(1)d_{\min} - 1$. Das Generatorpolynom besteht also aus *Minimalpolynomen* $v_q(D)$ bzgl. w^q mit Koeffizienten aus $\mathbb{F}_2$ (d.h. $v_q(w^q) = 0$): $g(D) = \mathrm{kgV}\{v_{n-1}(D), v_{n-2}(D), \ldots, v_{n-d_{\min}+1}(D)\}$.

Für $d_{\min} = 3$ erhält man die Klasse der perfekten 1–Fehler–korrigierenden Hammingcodes. Allgemein gilt für die Zahl der Prüfstellen bei einem t–Fehler–korrigierenden binären BCH–Code $n - k \leq t \cdot l$, d.h. $R \geq 1 - t \cdot l/n$.

Werden binäre l–Tupel als Codesymbole verwendet, also Elemente des Erweiterungskörpers $\mathbb{F}_{2^l}$, so ergibt sich die Klasse der **Reed–Solomon (RS–) Codes**. Das Generatorpolynom ist hier z.B. durch $g(D) = \prod_{j=1}^{d_{\min}-1}(D \ominus w^{n-j})$

gegeben, wobei w ein Element der Ordnung n bezeichnet. RS-Codes erreichen mit $d_{\min} = n-k+1$ die größtmögliche Minimaldistanz: je 2 Redundanzcodesymbole ist ein Codesymbol korrigierbar, wobei das binäre Fehlermuster innerhalb eines fehlerhaften l–Tupels (typischerweise Bytes) beliebig ist. Damit sind diese Codes insbesondere zur Korrektur kurzer Fehlerbündel geeignet. RS–Codes werden heute in nahezu allen Systemen zur Datenkommunikation eingesetzt (CD, DVD, Hard–Disc, Wireless LAN, u.v.a.m. vgl. auch Sicherungsprotokolle).

2.5.3 Faltungscodes und Viterbi–Algorithmus

Bei der **Faltungscodierung** werden aus k (im Prinzip unbegrenzten) Folgen $\langle m_\mu^{(i)} \rangle$; $i = 0(1)k-1$, $\mu \in \mathbb{Z}$, $m_\mu^{(i)} \in \mathbb{F}$ (meist $\mathbb{F}_2$: *binäre Faltungscodes*) mittels *Faltungsoperationen* n Codesymbolfolgen $\langle c_\mu^{(j)} \rangle$; $j = 0(1)n-1$, $c_\mu^{(j)} \in \mathbb{F}$ erzeugt. Da der Faltung die Multiplikation der zugehörigen Polynome entspricht, werden Faltungscodes vorzugsweise durch eine Generatormatrix

$$\mathbf{G}(D) = \begin{pmatrix} g_{00}(D) & \dots & g_{0,n-1}(D) \\ \vdots & \ddots & \vdots \\ g_{k-1,0}(D) & \dots & g_{k-1,n-1}(D) \end{pmatrix}$$

von *Subgenerator(polynom)en* $g_{ij}(D)$ spezifiziert. Werden die k Folgen $m^{(i)}(D)$ von Informationssymbolen bzw. die n Folgen $c^{(j)}(D)$ von Codesymbolen zu Zeilenvektoren zusammengefasst, so gilt:

$$\vec{c}(D) = \left(c^{(0)}(D) \dots c^{(n-1)}(D)\right) = \vec{m}(D) \odot \mathbf{G}(D)$$

Als Subgeneratoren $g_{ij}(D)$ sind sowohl ganze rationale als auch gebrochen rationale Übertragungsfunktionen (FIR– bzw. IIR–Systeme) gebräuchlich. Durch die Überführung von $\mathbf{G}(D)$ in Smith'sche Normalform erhält man eine *äquivalente systematische Codierung* für einen Faltungscode.

❑ *Beispiel:* Code der Rate 1/2; $k = 1$; $n = 2$; $\mathbb{F}_2$; $\mathbf{G}(D) = (1 \oplus D^2, 1 \oplus D \oplus D^2)$, siehe Abbildung 2.11.

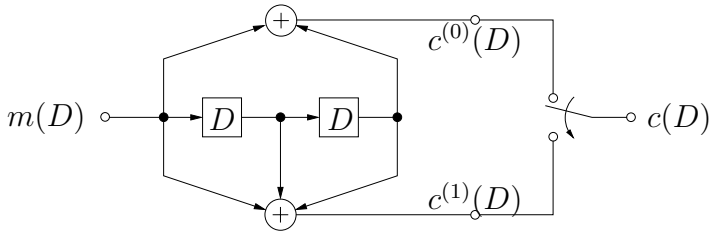


Abbildung 2.11: Codierung mit FIR System

Äquivalente systematische Codierung: $\mathbf{G}(D) = \left(1, \frac{1 \oplus D^2}{1 \oplus D \oplus D^2}\right)$, siehe Abbildung 2.12.

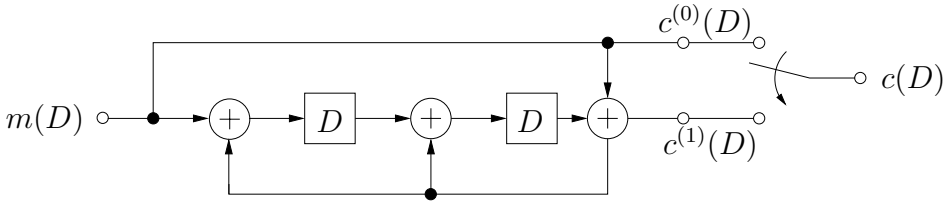


Abbildung 2.12: Codierung mit IIR System

Bezüglich der Decodierung wird der Faltungscoder als *endlicher Zustandsautomat* aufgefasst: Durch Zuführung eines Vektors $\vec{m}_\mu = \left(m_\mu^{(0)}, \ldots, m_\mu^{(k-1)}\right)$ im Schritt μ wird abhängig vom aktuellen Speicherzustand $S[\mu] \in \{0, 1, \ldots, Z-1\}$ ein Codesymbolvektor $\vec{c}_\mu = \left(c_\mu^{(0)}, c_\mu^{(1)}, \ldots, c_\mu^{(n-1)}\right)$ erzeugt; zugleich erfolgt ein Übergang in einen Folgezustand $S[\mu+1]$. Die zeitliche Abwicklung des Zustandsübergangsdiagramms des Automaten bildet das **Trellisdiagramm** des Faltungscoders, siehe Abbildung 2.13.

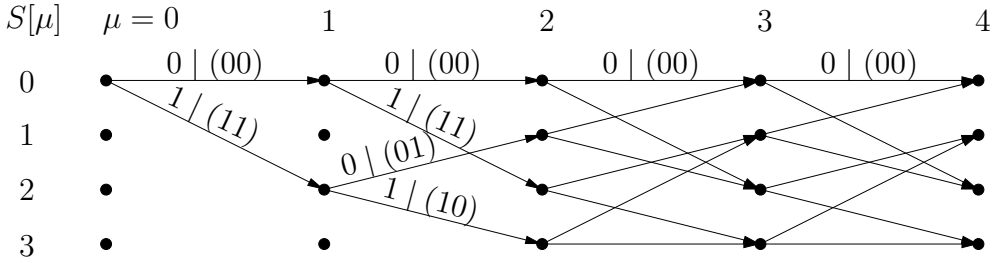


Abbildung 2.13: Trellisdiagramm zu Codierung gemäß Abbildung 2.11

Die möglichen Zustandsübergänge bilden **Zweige** im Trellisdiagramm, die jeweils durch erzeugenden Eingangsvektor $\vec{m}$ und erzeugten Codesymbolvektor $\vec{c}$ beschriftet werden.

Die Decodierung von Faltungscodes (allgemein Trelliscode) erfolgt vorzugsweise in diesem Trellisdiagramm, wobei die Folge von Zweigen (**Pfad**) bestimmt wird, die nach der Beobachtung der Sequenz von Empfangswerten r_μ, die am Kanalausgang aus den Vektoren $\vec{c}_\mu$ hervorgehen, die größte Wahrscheinlichkeit besitzt (Maximum–a–posteriori Decodierung). Wird ein gedächtnisloser Kanal vorausgesetzt, so können die Wahrscheinlichkeiten $\Pr(\vec{c}_\mu \mid r_\mu)$ multipliziert bzw. deren negative Logarithmen λ_μ (Metrik) addiert werden. Damit ist das Decodierproblem in ein Problem nach der Suche eines kürzesten Pfades überführt und es kann optimal mit Hilfe *linearer Programmierung* (Bellman–Algorithmus) sehr effizient gelöst werden, solange die Zahl Z der Zustände nicht sehr groß ist. Diese Zahl der Zustände steigt allerdings exponentiell mit dem Grad der Subgeneratoren von $G(D)$ an, was die Anwendbarkeit dieser Methode auf relativ einfache Codes begrenzt. Zu jedem Schritt μ wird für jeden Zustand S die Metriksumme für alle Zweige, die in einen Zustand führen, berechnet. Da nur der Pfad, der in einen aktuellen Zustand mit bester Metrik führt, eventuell auch der Pfad mit der insgesamt besten Metrik werden kann, können ohne Verlust der Optimalität alle anderen Pfade zu diesem Zustand endgültig ausgeschlossen werden. Auf diese Weise erhält man in jedem Schritt Z bedingt optimale Pfade. Da diese mit hoher Wahrscheinlichkeit einige Schritte rückwärts (Faustregel $\mu_0 = 5 \cdot \log_2(Z)$) diese alle aus einem gemeinsamen Pfad hervorgehen, kann aus dem *FIFO–Pfadregister* eines Zustandes der decodierte Vektor $\vec{m}_{\mu-\mu_0}$ entnommen werden. In der Codierungstheorie wird diese Vorgehensweise als **Viterbi–Algorithmus** bezeichnet. Gegenüber algebraischer Fehlerkorrekturverfahren bietet der Viterbi–Algorithmus den großen Vorteil, dass die anhand *analoger* Empfangswerte r_μ gewonnenene Zuverlässigkeitsinformation über die einzelnen Codesymbole ohne Mehraufwand im Decodierprozess verwertet werden kann. Aus diesem Grund ist mit Faltungscodes eine hohe Datensicherheit bei weit geringerer Verzögerung (kleinerer Codewortlänge) zu erreichen als mit algebraisch decodierten Blockcodes. Faltungscodes mit Viterbi–Decodierung werden bei vielen Kommunikationssystemen als *inneres* Kanalcodierungsverfahren von seriell verketteten Codierschemata verwendet, wobei Restfehler häufig durch einen *äußeren* RS–Code korrigiert werden. Vorzugsweise werden die Decodierschritte unter wechselseitiger Nutzung bis-

heriger Ergebnisse wiederholt ausgeführt, vgl. auch Abschnitt 2.5.4. Mit Hilfe des verwandten *BCJR–Algorithmus* [11] können neben der Informationssymbolsequenz selbst auch Zuverlässigkeitswerte für die einzelnen decodierten Informationssymbole in optimaler Weise bereitgestellt werden. Insbesondere bei einer Verkettung von Kanalcodierungsverfahren erlaubt die Weitergabe von Zuverlässigkeitsinformationen zwischen den einzelnen Decodern eine maßgebliche Steigerung der Leistungseffizienz der digitalen Übertragung.

2.5.4 Iterative Decodierung

Die Kanalkapazität wird mittels sehr langer Codes auch in der Praxis erstaunlich nahe erreicht, wenn beim Codeentwurf strukturelle Elemente mit Zufallscodierung so verknüpft werden, dass **iterative Decodierverfahren** anwendbar sind. Bei den sog. **Turbocodes** werden **Faltungsodes** über einen **Permutor** (oft auch als **Interleaver** bezeichnet) zur quasi–zufälligen Permutation langer Datenblöcke parallel oder seriell verkettet. Ein Low Density Parity Check **(LDPC-) Code** ist durch eine pseudo–zufällige, relativ dünn besetzte Prüfmatrix gegeben. In beiden Fällen werden Abhängigkeiten zwischen den Codesymbolen über mehrere Prüfgleichungen hinweg auf relativ lange Schleifen ausgedehnt, wodurch die statistische Unabhängigkeit von sog. *extrinsischer Information* über einzelne Codesymbole, die für die Anwendung iterativer Decodierverfahren eigentlich vorauszusetzen ist, leidlich gewahrt wird. O.B.d.A. werden hier ausschließlich lineare Binärcodes, $c_i \in \mathbb{F}_2$, betrachtet.

Mittels einer Prüfgleichung j (j–te Zeile der Prüfmatrix, vgl. Gl. (2.24))

$$c_{j_1} \oplus \cdots \oplus c_i \oplus \cdots \oplus c_{j_{d_c}} = 0 \Rightarrow c_i = c_{j_1} \oplus \cdots \oplus c_{j_{d_c}}$$

kann durch Auflösen nach c_i eine *extrinsische Wahrscheinlichkeit* $\varepsilon_{ij} \triangleq \Pr(c_i = 0 \mid \text{Prüfgleichung } i)$ aus den Wahrscheinlichkeiten $\gamma_l \triangleq \Pr(c_l = 0 \mid \ldots)$, $l = j_1 \ldots j_{d_c}$, $l \neq i$ für die übrigen Symbole in der Prüfgleichung berechnet werden. Ist das Codesymbol in d_v (degree of variable) Prüfgleichungen enthalten, enthält also die i–te Spalte der Prüfmatrix d_v Elemente 1, so liegen d_v solche extrinsischen Wahrscheinlichkeiten sowie die *intrinsische Wahrscheinlichkeit* $\gamma_i^{(0)} = \Pr(c_i = 0 \mid y_i)$ anhand des zugehörigen Kanalausgangs-

symbols y_i vor. Diese d_v+1 Aussagen zum Symbol c_i werden zur Wahrscheinlichkeit $\gamma_i^{(1)}$ kombiniert, wobei von statistischer Unabhängigkeit ausgegangen wird, um die Komplexität des Verfahrens zu begrenzen. Im nächsten Iterationsschritt wird die auf diese Weise *verbesserte Symbolwahrscheinlichkeit* über die Prüfgleichung wiederum zur Schätzung aller anderen Symbolwahrscheinlichkeiten nutzbar gemacht:

- Initialisierung: Vorbesetzung durch *intrinsische Information*:

$$\gamma_i^{(0)} = \theta_{ij}^{(0)} = \Pr\left(c_i = 0 \mid y_i\right) \quad i = 0, \ldots, n-1 \tag{2.26}$$

- Iterationsschritt μ:
  - Berechnung von *extrinsischen Wahrscheinlichkeiten*:

$$\varepsilon_{ij}^{(\mu)} = \theta_{ij_1}^{(\mu)} \boxtimes \theta_{ij_2}^{(\mu)} \boxtimes \cdots \boxtimes \theta_{ij_{d_c}}^{(\mu)} \quad i = 0(1)n-1; j = 0(1)n-k-1 \tag{2.27}$$

    mit $a \boxtimes b = a \cdot b + (1-a) \cdot (1-b); \quad a, b \in [0,1]$
  - Kombinieren der Informationen:

$$\gamma_i^{(\mu+1)} = \gamma_i^{(0)} \otimes \varepsilon_{i1}^{(\mu)} \otimes \cdots \otimes \varepsilon_{id_v}^{(\mu)} \tag{2.28}$$

    mit $a \otimes b = (a \cdot b)/(a \boxtimes b)$ und

$$\theta_{ij}^{(\mu+1)} = \gamma_i^{(\mu+1)} \otimes \left(1 - \varepsilon_{ij}^{(\mu)}\right) \tag{2.29}$$

In Gl. (2.28) werden intrinsische und extrinsische Wahrscheinlichkeiten kombiniert. Um die statistische Unabhängigkeit zu bewahren, darf Information über ein Symbol c_i, die über eine Prüfgleichung j zugeführt wurde, bei Auswertung dieser Prüfgleichung bzgl. anderer Symbole nicht wieder zurückgegeben werden. Deshalb wird in Gl. (2.29) die extrinsische Information, die über die Prüfgleichung j gewonnen wurde, wieder entfernt. (Das zu a inverse Element bzgl. $\otimes$ lautet $1-a$.)

Im Laufe der Iterationen breiten sich die a–posteriori Wahrscheinlichkeiten $\gamma_i^{(0)}$ für alle Codesymbole über die Prüfgleichungen zu allen anderen Symbolen aus (belief propagation). Solange dabei Wahrscheinlichkeiten nicht über Umwege auf sich selbst zurückwirken, bleibt die für Gl. (2.28) vorausgesetzt statistische Unabhängigkeit gewahrt. Über eine gewisse Anzahl von Iterationen läßt sich diese Bedingung für lange Codes mit quasi–zufälliger dünn besetzter Prüfmatrix aufrecht erhalten. Deshalb eignen sich LDPC–Codes für eine iterative Decodierung.

Bei Turbo–Codes werden für die einzelnen Komponentencodes mehrfach mittels des BCJR–Algorithmus Symbolwahrscheinlichkeiten ermittelt, die wiederum als extrinsische a–priori Wahrscheinlichkeiten zwischen den Decodern für die Komponentencodes ausgetauscht werden. Durch die zwischenliegende Permutation des Datenstroms erfolgt die Auflösung von statistischen Abhängigkeiten. Allerdings weisen Turbo–Codes insbesondere bei paralleler Verkettung relativ kleine minimale Hammingdistanzen auf, was zu seltenen Restfehlern auch bei höheren Störabständen führt (*Error–Floor–Effekt*).

Literaturverzeichnis

[1] C. E. Shannon. „A mathematical theory of communications“, *Bell System Technical Journal*, Vol. 27 (July and October 1948), S. 379-423 and 623-656.

[2] C. E. Shannon. „Communication in the presence of noise“, *Proceedings Instiute of Radio Engineers*, Vol. 37 (1949), S. 10-21.

[3] C. E. Shannon. „Communication theory of secrecy systems“, *Bell System Technical Journal*, Vol. 28 (1949), S. 656-715.

[4] S. Lin, D. J. Costello jr. . *Error control coding, Fundamentals and applications*, Pearson Prentice Hall, Englewood Cliffs, 2004.

[5] T. M. Cover, J. A. Thomas. *Elements of Information Theory*, John Wiley & Sons, New York, 1991.

[6] C. E. Shannon. „Prediction and entropy of printed english“, *Bell System Technical Journal*, Vol. 30 (1951), S. 50-64.

[7] D. A. Huffman. „A method for the construction of minimum redundancy codes“, *Proc IRE*, Vol. 40 (1952), S. 1098-1101.

[8] A. Tunstall. *Synthesis of noiseless compression codes*, PhD dissertation, Georgia Institute of Technology, Atlanta.

[9] F. M. J. Willems, Y. M. Shtarkov, T. J. Tjalkens. „The conext–tree weighting methods: basic properties“, *IEEE Transactions on Information Theory*, Vol. 41, No. 3 (1995), S. 653-664.

[10] M. Burrows, D. J. Wheeler. „A block–sorting lossless data compression algorithm“, *SRC Research report*, 1994.

[11] R. G. Gallager. „A simple derivation of the coding theorem and some applications“, *IEEE Transactions on Information Theory*, Vol. 19, No. 1 (1965), S. 3-18.

[12] L. Bahl, J. Cocke, F. Jelinek, J. Raviv. „Optimal decoding of linear codes for minimizing symbol error rate“, *IEEE Transactions on Information Theory*, Vol. 20, No. 2 (1974), S. 284-287.

[13] J. Ziv, A. Lempel. „A universal algorithm for sequential data compression“, *IEEE Transactions on Information Theory*, Vol. 23, No. 3 (1977), S. 337-343.

[14] R. Lidl, H. Niederreiter. „Finite Fields (Encyclopedia of Mathematics and its Applications)“, Addison-Wesley Pub(Sd), 1984.