Informations- und Codierungstheorie - eine Einführung

Ralph–Hardo Schulz

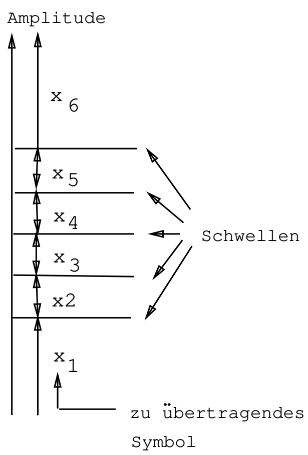
GLIEDERUNG

Einleitung

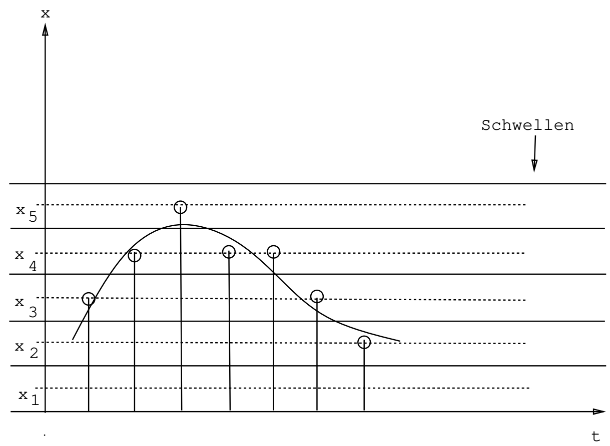
Im folgenden Artikel behandeln wir die Aufbereitung von Nachrichten für die digitale Übertragung. Dabei wollen wir untersuchen, wie die von einer Quelle ausgehenden Signale zunächst mit möglichst wenig Zeichen dargestellt (komprimiert) und dann gegen Fehler gesichert werden können.
Für die Beurteilung der Güte der erwähnten Kompression spielt eine theoretische Größe eine Rolle, nämlich der mittlere Informationsgehalt (die "Entropie") der Quelle; auf diesen Begriff wollen wir nach einigen einleitenden Beispielen von Nachrichtenquellen mit endlichem Signalvorrat eingehen.

1. Beispiele von (diskreten) Nachrichtenquellen

1.1 Beispiel 1
Wir stellen uns ein Meßgerät vor, zum Beispiel in einem Satelliten, dessen Meßergebnisse zu einem Empfänger gesendet werden sollen. Statt sich die genauen Meßdaten kontinuierlich übermitteln zu lassen, reicht es dem Forscher oft aus, nach einer festen Einteilung der Meßskala jeweils in kurz aufeinander folgenden Zeitintervallen zu erfahren, in welchem Bereich der Skala der aktuelle Meßwert



Figur 1: Quantisierung einer Messung



Figur 2: Abtastung und Quantisierung: $x_3, x_4, x_5, x_4, x_4, x_3, x_2$

liegt. (*Abtastung* und *Quantisierung* des analogen Signals). Damit wird die Meßanlage zu einer Quelle mit einem endlichen Signalvorrat (z.B. $x_1, x_2, \ldots, x_6$ in Figur 1). Unter Umständen kann man die Wahrscheinlichkeit des Auftretens der einzelnen Signale abschätzen.

1.2 Beispiel 2: **Tonverarbeitung**

Ähnlich lassen sich auch Tonquellen quantisieren: Man teilt die Amplitudenskala in Intervalle ein und übermittelt in sehr kurzen Zeitabständen den Mittelwert des Bereiches, in dem der aktuelle Meßwert liegt (Figur 2).
Bei geeigneter Wahl der Abtastfrequenz (z.Bsp. 44100 mal in der Sekunde) und der Quantisierungsintervalle ist eine sehr gute Rekonstruktion des ursprünglichen (analogen) Tonsignals möglich. (Die theoretische Grundlage bildet das sogenannte Abtasttheorem von SHANNON, auf das wir hier aber nicht eingehen können.)
Das beschriebene Vorgehen ist die Grundlage für die *Pulscodemodulation* (PCM), eines der wichtigen Verfahren der digitalen Tonverarbeitung; letztere setzt mit Compact Disc (CD) und Digital Audio Tape (DAT) neue Maßstäbe für Reinheit und Klangfarbe von Musikaufnahmen.

Literaturhinweise[1]: RUPPRECHT [14], SCHMIDT [20], SEYDEL & BURLISCH [16]

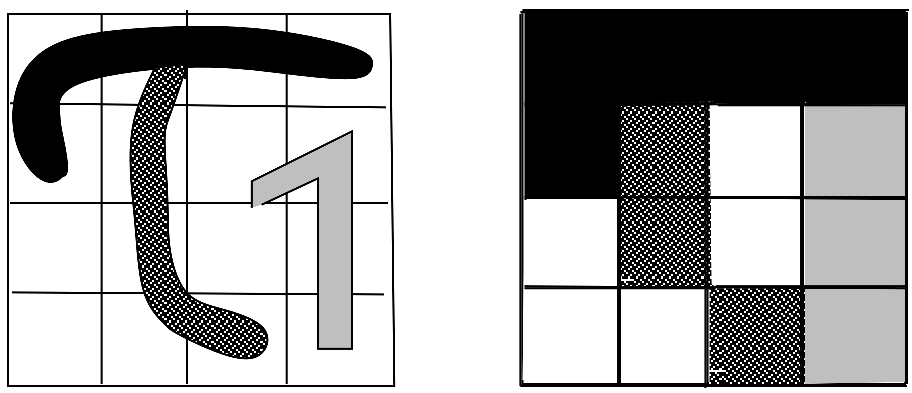
[1] Aus der Fülle der Literatur haben wir jeweils einige Bücher und Artikel ausgewählt; die genauen Zitate befinden sich am Ende des Artikels.

1.3 Beispiel 3: **Bildverarbeitung**

Bei der Bildverarbeitung (z.B. dem Fernsehen oder der Aufbereitung von Aufnahmen durch Raumsonden) wird ein einfarbiges Bild (in jeder der drei Komplementärfarben) durch Rasterung in einzelne Bildelemente (*"Pixels"*) aufgeteilt; für jedes dieser Pixels wird die Helligkeit (im Mittelpunkt oder als Mittelwert) gemessen; auch hier läßt sich die Intensitätsskala in verschiedene Stufen (*"Grauwerte"*) einteilen, deren aktueller Wert übertragen wird. In der Praxis sind heute mindestens 512×512 Bildelemente und $256 (= 2^8)$ Intensitätsstufen üblich. Bei der Vermessung mittels Videobild (Fotogrammetrie) z.Bsp. werden 2500×3500 Bildelemente pro Bild benutzt, beim hochauflösenden Fernsehbild (High Definition TV, HDTV) 1250 (bzw. 1125 beim japanischen Typ) Zeilen mit insgesamt 1,6 Millionen Pixels pro Bild. (Das heutige PAL-Fernsehbild hat 625 Zeilen und analoge Abtastung.)

Als ein stark vereinfachtes Beispiel haben wir das Bild von Figur 3a mit 4 Grauwerten ausgewählt und einer sehr groben Rasterung unterworfen. Der Fehler durch grobe Rasterung und Quantisierung ist in unserem Beispiel entsprechend groß (Figur 3b).

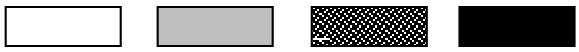
Verwenden wir dann die Bezeichnungen für die Grauwerte gemäß Tabelle 1, so erhalten wir als sogenanntes digitales Bild zu Figur 3 das Schema der Figur 4; dieses kann dann zeilenweise abgelesen werden. Die Übertragung des Bildes geht nun, so können wir es uns vorstellen, von einer Quelle mit den Signalen W, H, G, S aus. (Auf dieses Beispiel werden wir später mehrfach eingehen.)



(a) (b)

Figur 3: Ein Bild (a) vor und (b) nach der Rasterung (s. auch Tab.1 und Fig.4)

Wenn man nicht nur *ein* Bild, sondern, wie beim Fernsehen, eine Folge von Bildern übertragen will, so interessiert auch die generelle Wahrscheinlichkeitsverteilung der Signale. Durch längere Beobachtung der Quelle erhält man dann evtl. eine angenäherte Wahrscheinlichkeitsverteilung, von der im allgemeinen bei

Grauton				
Zeichen	W	H	G	S

S	S	S	S
S	G	W	H
W	G	W	H
W	W	G	H

Tabelle 1: Unsere Bezeichnungen für GrautöneFigur 4: Digitales Bild zu Figur 3

jedem einzelnen Bild die Häufigkeitsverteilung abweicht (s. Tabelle 2). Mögliche Abhängigkeiten in der Signalfolge vernachlässigen wir hier.

Zeichen	W	H	G	S
relative Häufigkeit in Fig.3b	$5/16 \approx 0{,}31$	$3/16 \approx 0{,}19$	$3/16 \approx 0{,}19$	$5/16 \approx 0{,}31$
Beispiel einer Wahrschein-lichkeits-Verteilung	0,38	0,13	0,19	0,3

Tabelle 2: Verteilung der Signale bei einer speziellen Bilddarstellung

Literaturhinweise: BAUER & GOOS [1] (insbes. 1.1-1.5), NEIDHARDT [13], OBERSCHELP [19], RUPPRECHT [14], SCHULZ [22].

1.4 Beispiel 4: **Geschriebener Text**

Eine weitere Nachrichtenquelleïst jeder geschriebene Text. Signale sind dabei die Buchstaben, Zwischenräume und die Interpunktion. Die Wahrscheinlichkeitsverteilung der Buchstaben der deutschen Sprache ist annähernd bekannt. Dabei sind die Signale nicht statistisch voneinander unabhängig: Manche Buchstabenpaare bzw. Buchstabentripel treten häufiger, andere seltener auf, als es der Wahrscheinlichkeit der einzelnen Buchstaben entspricht. Auch die Häufigkeit von Buchstaben-Paaren und -Tripeln sind an größeren Klartexten ermittelt worden. (In Abschnitt 5.5 gehen wir nochmals auf dieses Beispiel ein.)
Literaturhinweise: BAUER & GOOS [1], p.46-48, BEUTELSPACHER [2], DIFF [18], HOFFMANN [9], KAMEDA & WEIHRAUCH [10] (für deutschen Klartext), TOPSØE [17] (für englischen Text).

1.5 Beispiel 5: **Genetische Informationsquellen** *in der Molekularbiologie*
Betrachten wir, auf welche Weise die Gene in den Chromosomen eines Lebewesens Informationen liefern! Jedes DNS-Molekül enthält u.a. eine lange Sequenz von Molekülteilen, die aus den vier Basen (Nukleotiden) Adenin (A), Guanin (G), Cytosin (C) und Thymin (T) aufgebaut ist. Zur Herstellung von Proteinen wird aus einer solchen Folge eine Sequenz erzeugt (und an ein RNS-Molekül gebunden), bei der Thymin durch Uracil (U) ersetzt ist. Nacheinander wird jeweils ein Tripel dieser A,G,C,U-Sequenz "gelesen" und die entsprechende Aminosäure erzeugt. (Diese Zuordnung zu einer der 20 relevanten Aminosäuren erfolgt nach einem festen Schema, s. TOPSØE [17]). Die Tripel des RNS-Moleküls können so als (genetische) Signale einer Quelle verstanden werden.
In der Evolutionstheorie teilt man bei der Frage nach Entstehung des genetischen Codes die Molekülteilchen, die zum Aufbau eines t-RNS-Moleküls verwendet werden, in zwei Klassen R und Y ein (Purin-Pyrimidin-Codierung) und versucht dann, die verschiedenen Lebewesen nach der Ähnlichkeit dieser Sequenzen in einem Baum (phylogenetischer Baum) darzustellen und so Abstammungs-Verwandschaften aufzufinden.

Literaturhinweis: TOPSØE [17], WILEY [24], Arbeiten der Proffs. EIGEN, DRESS u.a.

Diese Beispiele sollen genügen, um einen Anwendungshintergrund für die folgende Einführung in die Informations- und Codierungstheorie zu geben.

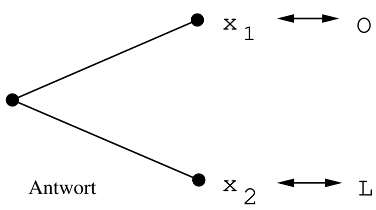
2. Der Begriff der Information

Teil A: *Information bei gleichwahrscheinlichen Ereignissen*

2.1 Wir stellen uns im folgenden eine *Nachrichtenquelle* X vor, die *Signale* der Form $x_1, \ldots, x_N$ (Zeichen, Symbole) produzieren kann (bzw. allgemein ein Experiment X mit Versuchsausgängen $x_1, \ldots, x_N$).
Bevor wir Kenntnis über die Aussendung eines Signals erhalten, sind wir über dieses im Unklaren. Diese Unsicherheit hängt unter anderem von der Wahrscheinlichkeit $\boldsymbol{p}(x_i)$ des möglichen Ereignisses x_i ab:
So besteht etwa beim sicheren Ereignis ($N = 1$) keine Ungewissheit über das Eintreffen. Und die Chancen, daß zum Beispiel bei einem Fußballpokalspiel eine Amateurmannschaft gewinnt, sind geringer als die Gewinnaussichten für die gegnerische haushoch favorisierte Bundesligamannschaft. Die Unsicherheit über einen Erfolg der Amateure ist also größer, die Überraschung über das Eintreffen des weniger zu erwartenden Ereignisses höher.
Eine Möglichkeit, Unsicherheit abzubauen, besteht darin, Antworten auf Fragen nach dem Ausgang zu erhalten. (Man kann sich ein *Orakel* vorstellen, das Alternativ-Fragen über den noch nicht bekannten Ausgang eines Experiments

mit Ja oder Nein beantwortet). Um auf diese Weise den Versuchsausgang zu erfahren, ist eine *Fragestrategie* dienlich.
Wir behandeln zunächst den Fall, daß alle N Quellenzeichen (Versuchsausgänge) gleich wahrscheinlich sind.

2.2 Bei genau zwei möglichen Signalen (Ausgängen) x_1, x_2 ist, unabhängig von der Wahrscheinlichkeit, die Frage tritt x_2 auf ?ëine sinnvolle Frage; ist die Antwort positiv, so kennzeichnen wir sie mit L (oder 1) andernfalls mit O (oh) (oder 0, Null) (Figur 5).[2]
Bei einer größeren Anzahl von Quellenzeichen ist allerdings die Frage nach x_2 als erste Frage höchstens dann geschickt, wenn das Eintreffen von x_2 eine größere oder ähnliche Wahrscheinlichkeit hat wie die übrigen Symbole zusammen. Dies ist bei etwa gleichwahrscheinlichen Ereignisssen nicht der Fall.

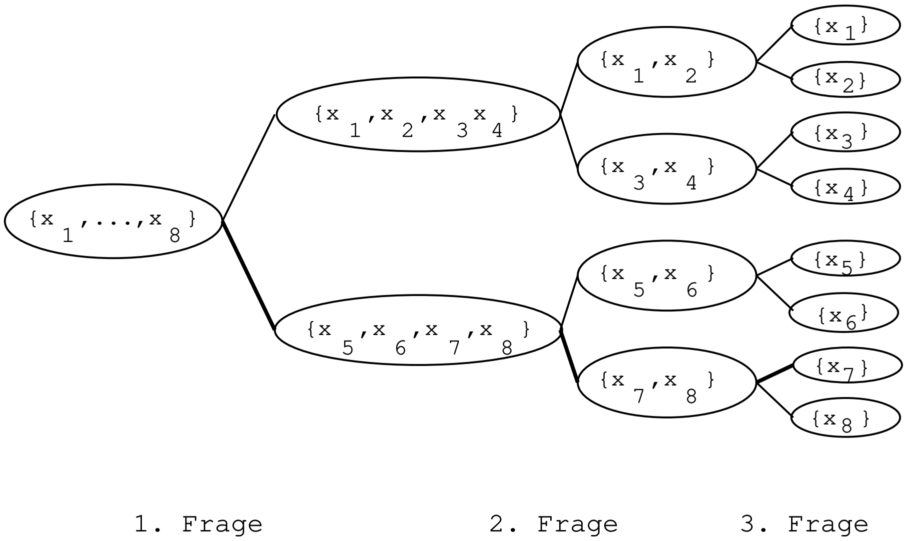


Figur 5: “Baum” (Verzweigung) zum Alternativ-Experiment

2.3 Für $N = 8$ z.B. ist es viel günstiger, zunächst zu fragen, ob das gesuchte Signal x in der Menge $\{x_5, \dots, x_8\}$ liegt oder nicht. Je nach Antwort stellen wir dann die 2. Frage. Ist die erste Frage positiv beantwortet, so fragen wir: “Gehört das fragliche Signal x zu $\{x_5, x_6\}$ oder nicht (d.h.also zu $\{x_7, x_8\}$) ?”; ist sie aber verneint worden, so wollen wir wissen: “Gehört x zu $\{x_1, x_2\}$ oder zu $\{x_3, x_4\}$?”. Durch eine 3. Frage klären wir schließlich, um welches Signal es sich handelt. In dem Diagramm von Figur 6 haben wir die Fragestrategie dargestellt. Verstärkt gezeichnet ist der Fall, daß $x = x_7$ ist.

In Figur 7 ist die Strategie nochmals formaler durch ein Baumdiagramm dargestellt und sind die Tripel der Anworten den Symbolen $x_1, \dots, x_8$ zugeordnet.
Jedes der 8 Quellenzeichen läßt sich also durch ein Wort der Länge 3 über dem Alphabet $\{O, L\}$ darstellen: $OOO, OOL, \dots, LLL$.
Dabei heißt bei einer endlichen nicht-leeren Menge A jedes Element $a_1 \dots a_n$ aus A^n ein **Wort der Länge** n **über dem Alphabet** A.
2.4 Es leuchtet ein , daß man den Prozeß der *sukzessiven Halbierung* der fraglichen Signal-Mengen bei größerer Anzahl N über die 3. Komponente hinaus weiterführen kann, sofern N eine 2-er Potenz ist.

[2]Bei der elektronischen Umsetzung kann “L” auch heißen “Strom fließt” oder “Lämpchen leuchtet” und “O” die Negation davon sein.



Figur 6:

Wissensstand über die Menge, in der das gesuchte Signal liegt, zu Beginn und nach der 1., 2. bzw. 3. Frage.

Hat man also z.B. $N = 2^m$ mögliche Signale $x_1, \ldots, x_N$ der Quelle (bzw. unabhängige Ausgänge - Elementarereignisse eines Versuchs), so lassen sich diese Ereignisse durch Wörter der Länge m über dem Alphabet $A = \{O, L\}$ darstellen; für deren Anzahl gilt ja

$$|A^m| = |A|^m = 2^m = N .$$

Allerdings gibt es auch andere Möglichkeiten der "Codierung" der Signal–Menge: Wir können Wörter unterschiedlicher Länge (s.u.) oder andere Alphabete nehmen. Will man (aus technischen Gründen) an $A = \{O, L\}$ festhalten, und sind alle Signale gleich wahrscheinlich, gilt also

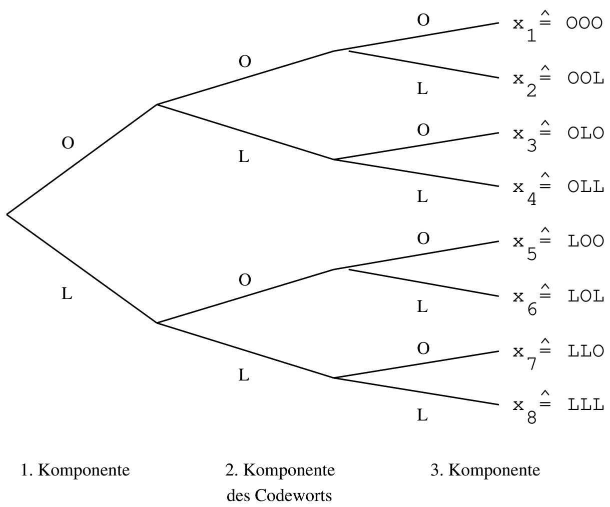
$$p := \boldsymbol{p}(x_i) = 1/N = \frac{1}{2^m} \quad (i = 1, \ldots, N),$$

so ist die Zuordnung zu Wörtern gleicher Länge m naheliegend (Figur 8). Für diese Wortlänge m gilt dann die Gleichung[3]

(2.4.1) $$m = \log_2 N = \log_2 (1/p) .$$

Die Zahl m bezeichnet man (nach HARTLEY) auch als die *Information* $I(x_i)$, die ein Signal x_i liefert. Als Maßeinheit für die Informationen ist "**bit**" gebräuchlich (wegen der Verbindung zu den **binären** Ziffern von $A = \{O, L\}$). Die Information des Signals x_i wird also im behandelten Fall gemessen durch die Anzahl der

[3]Zur Erinnerung: $\log_a x$ ist definiert als diejenige reelle Zahl y, für die $a^y = x$ gilt.



Figur 7: Baum zur Beschreibung von $x_1, \ldots, x_8$ durch Elemente aus $\{O, L\}^3$

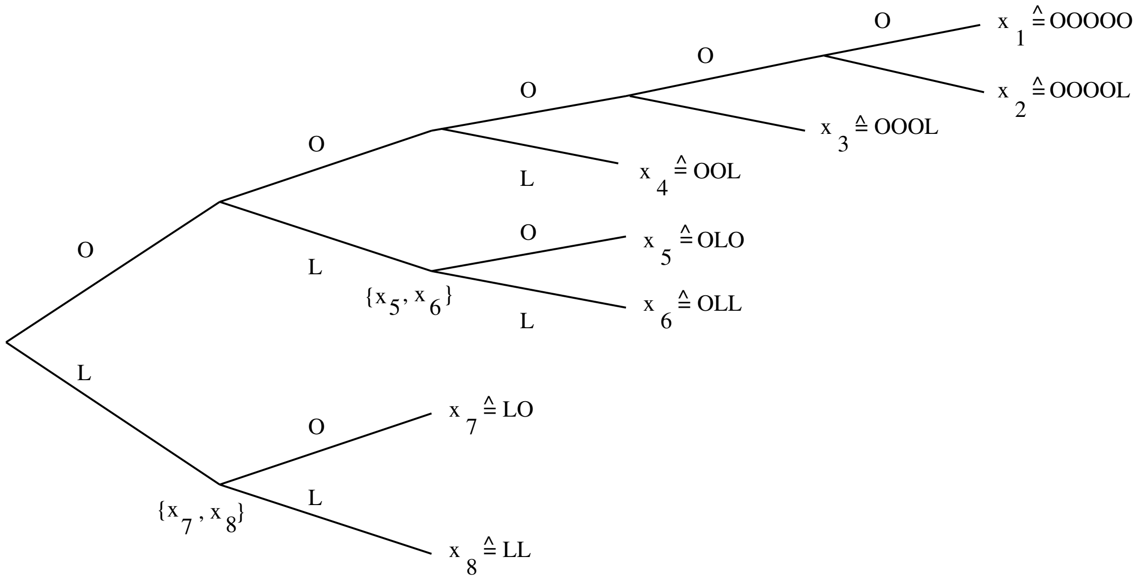
Fragen, die bei der betrachteten Fragestrategie nötig ist, x_i unter den anderen Signalen zu identifizieren.
Beispiel: Für $N = 8$ und $p = 1/8$ ist $I(x_i) = \log_2 8 = 3$ (vgl.2.3)

$$\begin{array}{ccc}
x_1 & \longleftrightarrow & OO\ldots OOO \\
x_2 & \longleftrightarrow & OO\ldots OOL \\
x_3 & \longleftrightarrow & OO\ldots OLO \\
x_4 & \longleftrightarrow & OO\ldots OLL \\
\cdot & & \cdot \\
\cdot & & \cdot \\
\cdot & \longleftrightarrow & LL\ldots LLO \\
x_{2^m} & \longleftrightarrow & LL\ldots LLL
\end{array}$$

Figur 8: Codierung von 2^m gleichwahrscheinlichen Signalen

2.5 Hier hat der Begriff *Information* eine Bedeutung, die gegenüber dem der Umgangssprache eingeengt ist. Insbesondere die subjektive Komponente, die bei letzterem gegeben ist und stark von dem Interesse an der Nachricht abhängt, bleibt hier völlig unberücksichtigt. Es ist allerdings sinnvoll, daß der Zahlenwert der Information von der Wahrscheinlichkeit der Quellensignale (der Versuchsausgänge) abhängt, wie es auch in (2.4.1) zum Ausdruck kommt: Der Informationszuwachs nach Kenntnis eines Quellensignals (des Ausgangs eines Versuchs) läßt sich als Abbau der vorher bestehenden Unsicherheit über das Eintreten des betreffenden Ereignisses (Signals) auffassen. Und diese Unsicherheit hängt von der Wahrscheinlichkeit des Ereignisses ab, wie wir sahen (vgl. 2.1).

2.6 Man kann zeigen (s.u. (4.13)), daß bei $N = 2^m$ Ausgängen gleicher Wahrscheinlichkeit die beschriebene sukzessive Halbierung auch eine *optimale Fragestrategie* ist, d.h. daß die über alle möglichen Ausgänge gemittelte Anzahl der Fragen, die hier ebenfalls m ist, minimal ist unter allen möglichen Fragestrategien nach $x_1, \dots, x_N$: Eine andere Darstellung durch binäre Wörter (bei anderer Fragestrategie) benötigt im Mittel eine größere oder höchstens gleiche Wortlänge (entsprechend der mittleren Fragenanzahl).
So gehört im Fall $N = 8$ zum Baum der Figur 9, mit dem sich x aus $x_1, \dots, x_8$ ebenfalls ermitteln läßt, eine mittlere Wortlänge von $\frac{1}{8}\,(5{+}5{+}4{+}3{+}3{+}3{+}2{+}2) = 3,375$ [bit/Symbol].
Die durch ein Signal gelieferte Information stimmt im beschriebenen Fall von 2^m gleichwahrscheinlichen Zeichen überein mit der kleinsten mittleren Wortlänge, die zu einer Darstellung der Signale über dem Alphabet $\{O, L\}$ nötig ist.



Figur 9: Beispiel einer Codierung mit unterschiedlichen Wortlängen

2.7 Bisher hatten wir vorausgesetzt, daß die Anzahl N der gleichwahrscheinlichen Ausgänge eine 2-er Potenz ist. Diese Einschränkung lassen wir nun fallen und definieren auch im allgemeinen Falle $\log_2(1/p)$ als das Maß der Information einer der Versuchsausgänge der Wahrscheinlichkeit p. Es handelt sich dann nicht mehr unbedingt um eine natürliche Zahl, also auch nicht mehr um eine konkrete Wortlänge.
Die Bedingung, daß die Ausgänge gleichwahrscheinlich sind, lassen wir im folgenden ebenfalls fallen.

Teil B: **Das Informationsmaß nach Shannon**

2.8 Im folgenden betrachten wir ein Experiment, dessen Ausgänge $x_1, \ldots, x_N$ mit Wahrscheinlichkeit $p_1, \ldots, p_N$ auftreten, bzw. (spezieller) eine Nachrichtenquelle mit möglichen Signalen (Zeichen, Symbolen) $x_1, \ldots, x_N$, die voneinander unabhängig mit der Wahrscheinlichkeit $\boldsymbol{p}(x_i) = p_i > 0$ gesendet werden. (D.h. also: $x_1, \ldots, x_N$ seien die Elementarereignisse eines geeigneten endlichen Wahrscheinlichkeitsraumes und $\boldsymbol{p} = (p_1, \ldots, p_N)$ die Verteilung dieser Ereignisse).
In Analogie zu (2.4.1) kommen wir zu folgender Definition (SHANNON):

Die **Information**, die durch das mit Wahrscheinlichkeit $p_i > 0$ eintretende Ereignis (Signal) x_i geliefert wird, ist definiert durch
(2.8.1) $$I(x_i) := \log_2(1/p_i) \quad [bit]$$

Anmerkung: Nach den Rechenregeln für Logarithmen gilt (mit $ln := \log_e$):

$$\log_2(1/p_i) \;=\; -\log_2 p_i = \; -ln\, p_i/ln\, 2\,.$$

2.9 Ausgehend von (2.4.1) kann man diese Definition der Information folgendermaßen plausibel machen: Bei ganzzahligem $1/p_i$ betrachtet man das Senden der Quelle (den Versuch) als Ziehen aus einem Hut, in dem das Symbol x_i in einer p_i entsprechenden relativen Häufigkeit vorkommt. Faßt man die übrigen Symbole so zu Gruppen zusammen, daß alle Gruppen möglichst die gleiche Häufigkeit p_i haben, so gibt es $1/p_i$ Gruppen gleicher relativer Häufigkeit. Erhält man nun x_i, so hat man eine dieser Gruppen gezogen und gewinnt so nach (2.4.1) eine Information vom Maß $\log_2(1/p_i)$ (s. MASSEY [12]). Das ist natürlich keine stichhaltige Begründung dafür, daß die Definition sinnvoll ist. Diese ergibt sich erst durch die Fülle der Resultate, die mit diesem Begriff gewonnen werden können.

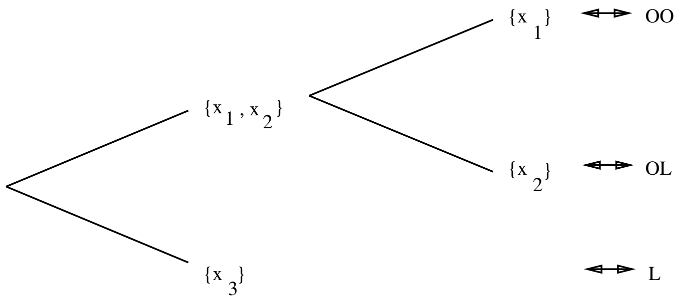
2.10 Als **Beispiel** betrachten wir eine Quelle mit Signalen x_1, x_2, x_3.

a) Es gelte $p_1 = \boldsymbol{p}(x_1) = 1/4\,,\ p_2 = 1/4\,,\ p_3 = 1/2$,
also $\boldsymbol{p} = (p_1, p_2, p_3) = (1/4, 1/4, 1/2)$.
Wir erhalten $I(x_1) = I(x_2) = \log_2 4 = 2[bit]$ und $I(x_3) = \log_2 2 = 1[bit]$.

An Figur 10 sehen wir, daß (bei Aufgabe der Forderung nach gleichen Wortlängen) jedes x_i sogar durch ein Wort der Länge $I(x_i)$ beschrieben werden kann, was allerdings an den speziellen Parametern liegt.

b) Ist hingegen $p = (1/3, 1/3, 1/3)$, so erhält man (bei gleichen Wortlängen wie in Beispiel a) als Maß der Information

$$I(x_i) = \log_2 3 = ln\, 3/ln\, 2 \approx 1,585\, [bit].$$



Figur 10: Binäre Codierung einer Quelle mit 3 Signalen (Beispiel)

2.11 Wir betrachten *Eigenschaften des Informationsmaßes* I.

Zunächst stellen wir fest, daß I lediglich eine Funktion f der Wahrscheinlichkeiten ist:[4]
(2.11.1) $I(x_i) = f(p_i)$ mit $p_i = \boldsymbol{p}(x_i)$ (für $i = 1, \ldots, N$).

Für diese Funktion f gilt:
(2.11.2) $f :]0,1[\to \mathbb{R}$ ist stetig und streng monoton fallend[5]
sowie
(2.11.3) $f(p_i \cdot p_j) = f(p_i) + f(p_j)$ für alle p_i, p_j aus $]0,1[$.

2.12 Auch unabhängig von der Definition (2.8.1) beschreiben (2.11.1) bis (2.11.3) Eigenschaften, die man von einem Informationsmaß sinnvollerweise erwarten kann: Zu der alleinigen Abhängigkeit von den Wahrscheinlichkeiten (objektives Maß) haben wir schon in 2.5 Stellung genommen, ebenso zur strengen Monotonie (- je unerwarteter der Ausgang, desto größer die Information als Beseitigung von Unsicherheit -). Und (2.11.3) kann man gemäß der Forderung erwarten, daß sich bei zwei voneinander unabhängig durch die Quelle gelieferten Signalen die Gesamtinformation additiv aus den Einzelinformationen zusammensetzt.
Man kann nun umgekehrt zeigen, daß die Funktionen f mit

$$f(p) = k \cdot ln\, p \quad (\text{für } k < 0)$$

die einzigen sind, die (2.11.1) bis (2.11.3) erfüllen.
(Beweisandeutung: $h := f \circ \exp$ ist stetig und erfüllt $h(x+y) = h(x) + h(y)$; daraus folgt $h(x) = k \cdot x$).
Bis auf einen konstanten Faktor ist damit I durch (2.11.1) bis (2.11.3) bestimmt. Wählt man diesen Faktor so, daß $I(1/2) = 1$ ist, so ergibt sich $I(x_i) = -\log_2 p_i$ - eine weitere Motivation für die Definition (2.8.1).

[4]nämlich f mit $f(x) = \log_2(1/x)$.

[5]Hierbei bedeutet $]0,1[$ die Menge aller reellen Zahlen zwischen 0 und 1 (außer 0 und 1 selbst), $\mathbb{R}$ die Menge der reellen Zahlen.

Literaturhinweise: KAMEDA & WEIHRAUCH [10], HOFMANN [9], MASSEY [12], SCHULZ [15], TOPSØE [17]

3. Entropie

3.1 Im allgemeinen ist man nicht allein an der Information nur eines einzelnen speziellen Quellensymbols interessiert. Ähnlich wie wir uns in 2.6 schon mit gemittelten Wortlängen beschäftigten, so behandeln wir nun die im Mittel pro Signal gelieferte Information, die sogenannte (ideelle) Entropie der Nachrichtenquelle. Manche Autoren benutzen das Wort Entropie auch als Maß für die Unbestimmtheit eines Versuchs vor dessen Ausführung; nach dem Versuch sinkt sie auf 0, während nun die Informationsmenge im Mittel von 0 auf den Wert der Entropie gestiegen ist.

3.2 Gegeben sei eine Quelle (bzw. ein Versuchsaufbau) X mit den möglichen Symbolen (Versuchsausgängen) $x_1, \ldots, x_N$ und Wahrscheinlichkeitsverteilung $\boldsymbol{p} = (p_1, \ldots, p_N)$,
(also $\boldsymbol{p}(x_i) = p_i$ mit $0 \leq p_i \leq 1$ für $i = 1, \ldots, N$ und $\sum\limits_{i=1}^{N} p_i = 1$).[6]
Dann heißt

$$H(X) := \sum_{i=1}^{N} p_i \cdot \log_2(1/p_i) \quad \text{[bit/Symbol]}$$

die (ideelle) **Entropie** von X.

Hierbei lassen wir im Gegensatz zu 2.8 auch $p_i = 0$ zu mit der Vereinbarung,für $p = 0$ dann $p \log(1/p) = -p \log p := 0$ zu setzen).
Bei $H(X)$ handelt es sich also um die gemäß der Wahrscheinlichkeit p_i des Auftretens von x_i über alle x_i gemittelten Informationen $I(x_i)$, d.h. **den mittleren Informationsgehalt** pro Quellenzeichen (bzw. Versuchausgang), (genauer: den Erwartungswert von I).

Wir beachten, daß $H(X)$ nicht von $x_1, \ldots, x_N$ abhängt, sondern allein von der Wahrscheinlichkeitsverteilung $\boldsymbol{p} = (p_1, \ldots, p_N)$; deshalb schreiben wir auch $H(\boldsymbol{p})$ oder $H(p_1, \ldots, p_N)$.
3.3 **Beispiele**

a) Sei $\boldsymbol{p} = (1/8, 1/8, \ldots, 1/8)$ (vgl. 2.3.). Dann ist
$H(\boldsymbol{p}) = \sum\limits_{i=1}^{8} 1/8 \cdot \log_2 8 = 8 \cdot 1/8 \cdot 3 = 3$ ($\hat{=} I(x_i)$ für $i = 1, \ldots, 8$).

[6] $\sum\limits_{i=1}^{N} a_i$ steht als Abkürzung von $a_1 + \ldots + a_N$.

Anmerkung:

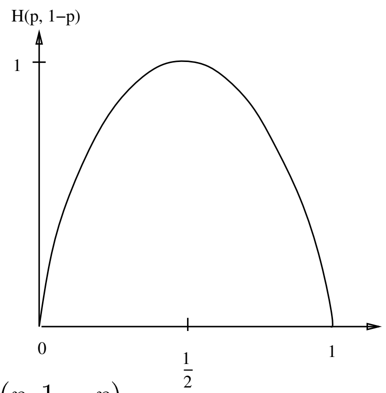
Es gilt allgemein: $H(1/N, \ldots, 1/N) = (\sum_{i=1}^{N} 1/N \log_2 N) = \log_2 N.$

b) (i) Sei $\boldsymbol{p} = (1/4, 1/4, 1/2)$ (vgl. 2.10.a). Dann gilt:
$H(\boldsymbol{p}) = 1/4 \cdot \log_2 4 + 1/4 \cdot \log_2 4 + 1/2 \cdot \log_2 2 = 2/4 + 2/4 + 1/2 = 1,5$ [bit/Symbol]
gegenüber
(ii) $H(1/3, 1/3, 1/3) = \log_2 3 = ln\, 3/ln\, 2 \approx 1,585$ [bit/Symbol] (vgl. 2.10.b).

c) Die Quelle mit den Signalen W, H, G, S und Verteilung $(0,38; 0,13; 0,19; 0,30)$ (vgl. Bsp. 1.3) hat die Entropie
$-1/ln\, 2 \cdot (0,38 \cdot ln\, 0,38 + 0,13 \cdot ln\, 0,13 + 0,19 \cdot ln\, 0,19 + 0,3 \cdot ln\, 0,3)$
$\approx 1,89$ [bit/Symbol] (d.h. weniger als $2 = \log_2 N$ [bit/Symbol] für $N = 4$).

d) Sei $\boldsymbol{p} = (p, 1-p)$. Es folgt $H(\boldsymbol{p}) = (-p \cdot \log_2 p - (1-p)\log_2(1-p))$. Der Graph der Funktion $[0,1] \to I\!R$ mit $p \mapsto H(p, 1-p)$ ist in Figur 11 skizziert.

$H(p, 1-p)$ ist maximal,
wenn
$p = 1 - p\ (= 1/2)$ ist.



Figur 11: Graph der Entropie-Funktion $H(p, 1-p)$

3.4 Auch allgemein kann man zeigen: Die Entropiefunktion $H(p_1, \ldots, p_N)$ nimmt (bei festgehaltenem N) ihr Maximum für $p_1 = \ldots = p_N = 1/N$ an.

Also: Unter allen Quellen mit N Signalen hat diejenige die größte Entropie, deren Signale alle die gleiche Wahrscheinlichkeit besitzen.

Die Kenntnis der Wahrscheinlichkeitsverteilung einer Quelle bedeutet dabei i.a. eine *Vorausinformation*, die die Entropie der Quelle unterschiedlich gegenüber der maximal möglichen Entropie mindert.
Literaturhinweise: HOFMANN [9]p.10ff, KAMEDA & WEIHRAUCH [10], TOPSØE [17], OBERSCHELP [19], SCHULZ [15]p.49ff.

4. Quellencodierung

4.1 In den vorigen Abschnitten haben wir oft von Darstellungen der Versuchsausgänge bzw. der Quellensignale durch binäre Wörter gesprochen. Wir wollen dies jetzt präzisieren und dann allgemein auf Codierungen einer Quelle eingehen. Wieder sei gegeben eine Quelle X mit verschiedenen möglichen Signalen $x_1, \ldots, x_N$ sowie evtl. einer Wahrscheinlichkeitsverteilung $\boldsymbol{p}$.

Unter einer **Codierung** von X über dem **Alphabet** A verstehen wir eine umkehrbar eindeutige Abbildung von $\{x_1, \ldots, x_N\}$ in die Menge aller Wörter der Form $c_1 \ldots c_n$ mit natürlicher Zahl n und $c_1, \ldots, c_n$ aus A. Jedes der verwendeten Wörter heißt dann **Codewort** dieser Codierung, die Menge aller Codewörter der verwendete **Code** $\boldsymbol{C}$.

4.2 **Erste Beispiele**

a) (vgl.2.4)
$$\begin{array}{lcc} x_1 & \mapsto & O \ldots OOO \\ x_2 & \mapsto & O \ldots OOL \\ \cdot & & \\ \cdot & & \\ x_N & \mapsto & \underbrace{L \ldots LLL}_{\text{jeweils m Zeichen}} \end{array}$$

ist eine mögliche Codierung von 2^m Signalen $x_1, \ldots, x_N$; das verwandte Alphabet ist $A = \{O, L\}$, der Code $\boldsymbol{C} = A^m$; (die Codewörter sind ja genau alle Wörter der Länge m über A).

b) (vgl.2.10)
$$\begin{array}{lcc} x_1 & \mapsto & OO \\ x_2 & \mapsto & OL \\ x_3 & \mapsto & L \end{array}$$

ist eine mögliche Codierung von $\{x_1, x_2, x_3\}$; das Alphabet ist wieder $\{O, L\}$, ferner $\boldsymbol{C} = \{OO, OL, L\}$ der verwendete Code.

c) Einige **übliche binäre Codierungen der Ziffern 0 bis 9** sind in der Tabelle 3 gelistet (vgl.BAUER & GOOS [1]S.36, s.auch DWORATSCHEK [4], SCHULZ[15]).
Hierbei sind die *direkte* und die *Aiken-Codierung* Stellenwert-Codierungen, d.h. es wird jedes L entsprechend seiner Position im Wort bewertet; der Wert der 1., 2., 3. bzw. 4. Komponente ist 8 - 4 - 2 - 1 bei der direkten Codierung (Dualdarstellung) bzw. 2 - 4 - 2 - 1 im Fall des Aiken- Codes. Der *Gray-Code* hat die Eigenschaft, daß sich zwei benachbarte Ziffern nur in einem Zeichen unterscheiden. Der *Exzess-3-Code* ist ein um "3" verschobener Code; er vermeidet so die "Tetraden" (4-Tupel) *OOOO* (Stromunterbrechung) und *LLLL* (Dauerstrom).

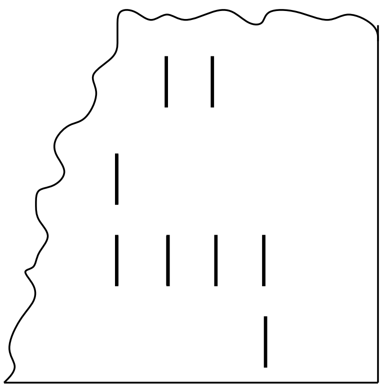
Codierung:

Signal	direkt	Gray-Code	Exzess-3	Aiken	2-aus-5	CCIT-2	ASCII
0	OOOO	OOOO	OOLL	OOOO	LLOOO	OLLOL	OOOOLLOO
1	OOOL	OOOL	OLOO	OOOL	OOOLL	LLLOL	LOOOLLOL
2	OOLO	OOLL	OLOL	OOLO	OOLOL	LLOOL	OLOOLLOL
3	OOLL	OOLO	OLLO	OOLL	OOLLO	LOOOO	LLOOLLOO
4	OLOO	OLLO	OLLL	OLOO	OLOOL	OLOLO	OOLOLLOL
5	OLOL	OLLL	LOOO	LOLL	OLOLO	OOOOL	LOLOLLOO
6	OLLO	OLOL	LOOL	LLOO	OLLOO	LOLOL	OLLOLLOO
7	OLLL	OLOO	LOLO	LLOL	LOOOL	LLLOO	LLLOLLOL
8	LOOO	LLOO	LOLL	LLLO	LOOLO	OLLOO	OOOLLLOL
9	LOOL	LLOL	LLOO	LLLL	LOLOO	OOOLL	LOOLLLOO

“*Tetraden−Codes*”

Tabelle 3: Einige übliche Codierungen der Ziffern 0 bis 9

Der *2-aus-5 - Code* war längere Zeit zur Codierung von Postleitzahlen in Benutzung (s . Figur 12). Bis auf die Ziffer 0 handelt es sich ebenfalls um eine Stellenwertcodierung mit den Werten $7-4-2-1-0$ und jeweils zwei mal “L”.



Figur 12: Beispiel einer 2-aus 5-Codierung

Der *CCIT-2-Code* ist der internationale Telegraphencode, der früher - auf Lochstreifen geschrieben - auch zur Computersteuerung benutzt wurde. (Zifferndarstellung nach Zeichenumschaltung LLOLL).

Der *ASCII-Code* (American standard code for information interchange) ibenutzt die ersten 7 Positionen jedes aus 8 Bits (= 1 Byte) bestehenden Codeworts und setzt die 8. Position O oder L, derart, daß die Gesamtzahl der mit L besetzten Positionen jeden Bytes gerade ist. (*Paritätskontrolle*)

d) Die *Codierung von Buchstaben* und anderen Zeichen ist (wie im Morsealphabet und im CCIT-Code) auch für den ASCII-Code standardisiert (z.B. durch DIN-Norm) und kann Tabellen entnommen werden.

4.3 Um (anders als beim Morsecode) ohne zusätzliches Pausensignal auszukommen, codieren wir eine Folge von Quellensymbolen, indem wir die entsprechenden Codewörter aneinander hängen. Wenn nicht alle Codewörter die gleiche Länge haben, so daß man nur die Stellen abzuzählen braucht, birgt dies trotz umkehrbar eindeutiger Zuordnung der Signale zu den Codewörtern Gefahren in sich: Ist etwa

$$\begin{array}{ccc} x_1 & \mapsto & O \\ x_2 & \mapsto & L \\ x_3 & \mapsto & OL \end{array}$$

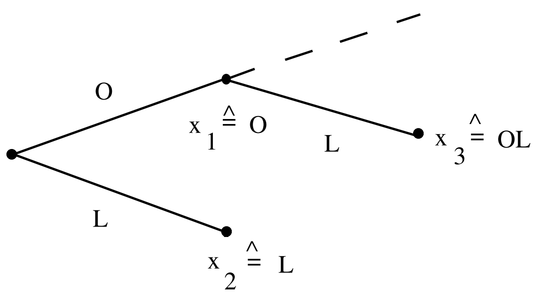
eine Codierung mit Code $\{O, L, OL\}$, so könnte OL sowohl x_3 als auch $x_1 x_2$ repräsentieren. Wir wollen aber, daß es zu jeder Folge $c_1 \ldots c_m$ von Elementen aus A höchstens eine Folge von Quellensignalen gibt, die zu $c_1 \ldots c_m$ codiert wird. Wir sprechen dann von *eindeutiger Decodierbarkeit* des Codes.

4.4 Wie muß nun der Code C aussehen, damit eine Folge von Codewörtern unzweifelhaft decodiert werden kann? Dazu sehen wir uns nochmals das Beispiel aus 4.3 an und stellen fest, daß dort ein Codewort (O) Anfangsstück eines weiteren (OL) ist. Dabei heißt $c_1 \ldots c_m$ ein Anfangsstück (*Präfix*) des Wortes $c_1 \ldots c_m c_{m+1} \ldots c_n$ und $c_{m+1} \ldots c_n$ Endstück (*Suffix*).

Eine Situtation wie oben beschrieben wollen wir ausschließen und kommen zu folgender

> *Definition:*
> Ein Code $\boldsymbol{C}$ heißt **Präfixcode** (irreduzibler Code), wenn kein Codewort aus $\boldsymbol{C}$ Präfix eines anderen Codeworts von $\boldsymbol{C}$ ist.

Beispiele sind sämtliche Codes, deren Codewörter alle die gleiche Länge haben, sogenannte **Block-Codes** (zum Beispiel die Codes von Figur 7, Figur 8 oder Tabelle 3), aber auch der Code von Figur 9.



Figur 13: Baum zum Code $\{O, L, OL\}$

4.5 Vergleicht man den Baum zum Code aus 4.3 mit den Bäumen der erwähnten Codes, so fällt auf (s.Figur 13), daß das Codewort O keiner End-Ecke (Spitze, Blatt) des Baumes entspricht.

Bei *Präfixcodes* gilt hingegen, daß die Codewörter durch End-Ecken im zugehörigen Codebaum repräsentiert sind.

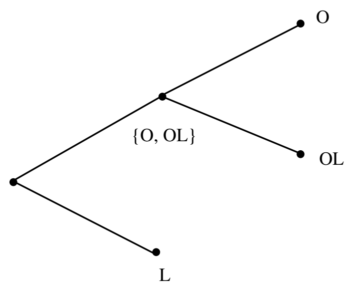
4.6 Wenn uns ein Präfixcode gegeben ist und eine Folge $a_1 \ldots a_m$ vorliegt, von der wir wissen, daß sie sich aus Codewörtern zusammensetzt, so können wir sukzessive $a_1, a_1a_2, a_1a_2a_3$ etc. bilden solange, bis wir ein Codewort $\boldsymbol{c}$ gefunden haben. Wegen der Präfix-Eigenschaft brauchen wir die weiteren Symbole zunächst nicht zu berücksichtigen, sondern dürfen $\boldsymbol{c}$ decodieren. Mit den verbleibenden Gliedern von $a_1 \ldots a_m$ können wir entsprechend verfahren. Damit sehen wir:

Jeder Präfixcode ist eindeutig decodierbar.

Präfixcodes sind aber nicht die einzigen Codes dieser Eigenschaft; man überlegt sich zum Beispiel, daß auch $\{0, 0L\}$ eindeutig decodierbar ist; ein Codewort kann hier aber nur decodiert werden, wenn man auch das nachfolgende Bit betrachtet. Präfixcodes sind hingegen *sofort decodierbar*, d.h. man kann decodieren, sobald ein Codewort erkannt ist, und die verbleibende Bitfolge läßt sich ebenfalls decodieren. (Diese Eigenschaft ist sogar kennzeichnend für Präfixcodes).

4.7 Wir wollen nochmals auf den Zusammenhang zwischen Codierungen und Fragestrategien eingehen. Dabei müssen wir beachten, daß nun Codewörter verschiedener Länge auftreten können. Auch wollen wir präzisieren, was wir unter einer **Fragestrategie** verstehen:
Es ist für ein unbekanntes x aus einer Menge $M = M_\circ$ durch Ja-Nein-Fragen zu klären, welches Element nun x ist - und zwar für jedes mögliche x aus M. Dazu sei nach dem i-ten Schritt einer solchen Strategie die Menge der noch möglichen Kandidaten für x auf M_i eingeschränkt; beim (i+1)-ten Schritt fragt man dann nach der Zugehörigkeit von x zu einer geeignet gewählten echten Teilmenge T von M_i und gelangt so zu einer kleineren Menge M_{i+1} der noch möglichen Elemente (je nach Antwort ist M_{i+1} gleich T oder $M_i \backslash T$); dies wiederholt man solange, bis man zur einelementigen Menge $\{x\}$ gelangt ist.
Im zugehörigen Baum entspricht jeder Ja-Nein-Frage eine Verzweigung und den erfragten Quellen-Symbolen die Endecken des Baumes.

Der Baum von Figur 13 zum Beispiel beschreibt keine Fragestrategie; denn nach der ersten Frage, die "Gehört x zu $\{O, OL\} \stackrel{\wedge}{=} \{x_1, x_3\}$?" lauten kann, ist zur Unterscheidung von $x_1 \stackrel{\wedge}{=} O$ und $x_3 \stackrel{\wedge}{=} OL$ noch eine weitere Frage nötig, sodaß der Baum von Figur 14 entsteht (und im Code OO statt O gewählt werden müßte).



Figur 14: Baum zur Fragestrategie nach $\{O, OL, L\}$

Man kann sich überzeugen, daß ähnliche Schwierigkeiten jedesmal auftreten, wenn kein Präfixcode vorliegt. Ja es gilt sogar:

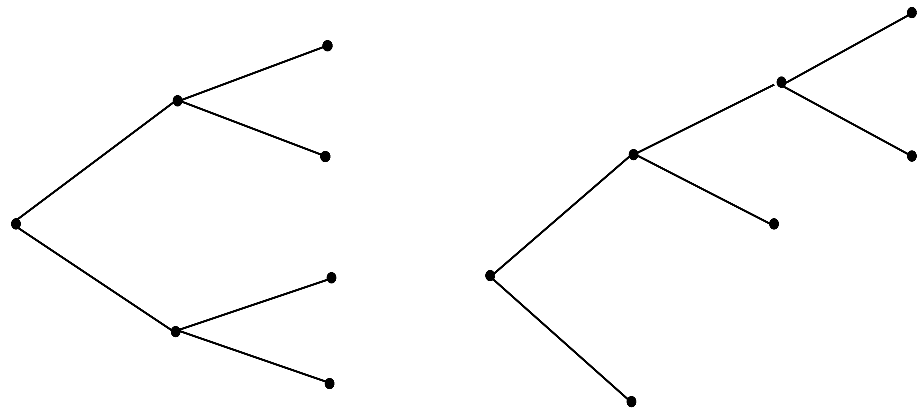
Jede Fragestrategie führt zu einem Präfixcode, und umgekehrt entspricht jedem Präfixcode eine Fragestrategie.

4.8 Bei gegebener Quelle X mit Symbolen $x_1, \ldots, x_N$ und Wahrscheinlichkeitsverteilung $\boldsymbol{p} = (p_1, \ldots, p_N)$ interessieren wir uns auch hier (ähnlich wie in 2.6) für optimale Fragestrategien bzw. optimale Codierungen mit Präfixcodes; um diesen Begriff präzisieren zu können, definieren wir noch die **mittlere Codewortlänge** (wirkliche Entropie, den *mittleren Codieraufwand*) einer Codierung als die (entsprechend den Wahrscheinlichkeiten des Auftretens der Symbole) über alle x_i gemittelte Codewortlänge (den Erwartungswert), und zwar durch die Festsetzung:

$$\boxed{(4.8.1) \quad \bar{l} := \sum_{i=1}^{N} \boldsymbol{p}(x_i) \cdot l(x_i) \text{ (mittlere Codewortlänge, mittlerer Codieraufwand)}}$$

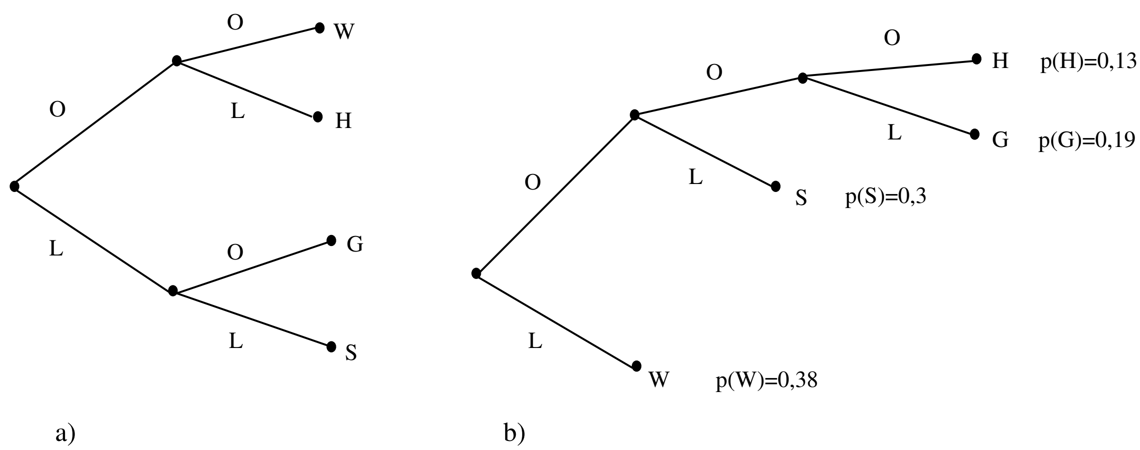
(dabei bezeichnet $l(x_i)$ der Länge des x_i zugeordneten Codeworts.)
Eine **optimale Codierung** ist nun eine Codierung der einzelnen Quellenzeichen mit minimaler mittlerer Codewortlänge $\bar{l}_{\min(X)}$ bzw. $\bar{l}_{\min(\boldsymbol{p})}$. (Für kurze Codewortlängen ist nämlich die erforderliche Geschwindigkeit bei der Sendung der Code-Bits und damit der technische Aufwand geringer als bei großen Längen, daher gelten in dieser Beziehung Codes kleinerer mittlerer Codewortlänge als besser).

4.9 Als theoretisches Beispiel betrachten wir bei der in 1.3 angeführten Quelle mit 4 Signalen die wichtigsten möglichen Codierungen mit Präfixcodes und berechnen die mittleren Codewortlängen. Zunächst stellen wir fest, daß es im wesentlichen nur zwei verschiedene (Code-) Bäume mit 4 Endecken (Blättern) gibt (s. Figur 15).



Figur 15: Codebäume mit 4 Endecken

Natürlich ist es angebracht, bei gegebenem Code die sehr häufig auftretenden Signale durch möglichst kurze Codewörter darzustellen. Beachten wir dies, so erhalten wir für die Quelle aus 1.3 die beiden Codierungen von Figur 16, von denen eine optimal sein muß.



Figur 16: Codierungen der Quelle aus 1.3.
a) $\bar{l} = 2$ [bit/Zeichen] *b)* $\bar{l} = 1,94$ [bit/Zeichen] $= \bar{l}_{\min}$

Da im Fall a) $\bar{l} = 2$ [bit/Zeichen] und im Fall b) $\bar{l} = 0,38 \cdot 1 + 0,3 \cdot 2 + (0,13 + 0,19)\,3 = 1,94$ [bit/Zeichen] ist, erweist sich die Codierung

$$\begin{array}{rcl} W & \mapsto & L \\ H & \mapsto & OOO \\ G & \mapsto & OOL \\ S & \mapsto & OL \end{array}$$

als optimal für die Quelle aus 1.3. Mit ihr wird das digitale Bild aus Figur 4 nun zum *codierten Bild* von Figur 17.

$$\begin{array}{ccccccccc}
O & L & O & L & O & L & O & L & \\
O & L & O & O & L & L & O & O & O \\
L & O & O & L & L & O & O & O & \\
L & L & O & O & L & O & O & O &
\end{array}$$

Figur 17: Codiertes Bild zu Figur 3

Die mittlere Codewortlänge ist ein Erwartungwert für die Signalübermittlung über längere Zeit hinweg; von ihr weicht i.a. die gemittelte Codewortlänge pro Symbol ab, wie sie bei Übertragung einer konkreten Signalfolge benötigt wird. So beträgt letztere bei der Übertragung des digitalen Bildes von Figur 4 mit dem Code aus Figur 16a) 2 [bit/Signal]; bei Verwendung des Codes aus Figur 16b) erhält man, vgl. Figur 17, sogar $33/16 \approx 2,06$ [bit/Signal]. Hier wäre di!Codierung nach Figur 16a) günstiger gewesen. Da die Codierung aber meist vor Bekanntwerden der Signalfolge festgelegt sein muß, sind solche Abweichungen unvermeidlich, gleichen sich jedoch bei weiteren Signalfolgen schließlich aus.

4.10 Wir behandeln nun *zwei Standard-Verfahren der Quellencodierung.* Dazu setzen wir voraus, daß die Symbole $x_1, \ldots, x_N$ der Quelle so geordnet sind, daß für die Wahrscheinlichkeiten $p_i = \boldsymbol{p}(x_i)$ gilt:

$$p_1 \geq p_2 \geq \ldots \geq p_N .$$

Zunächst gehen wir auf die **Codierung nach Fano** ein.
Bei dieser werden (analog zu unserem Vorgehen bei gleichwahrscheinlichen Signalen in 2.3) die Signale unter Beibehaltung der erwähnten Ordnung so in 2 Gruppen aufgeteilt, daß die beiden Teilmengen etwa gleiche Wahrscheinlichkeit besitzen. Mit den einzelnen entstandenen Teilmengen verfährt man dann ähnlich, solange bis man eine Codierung erhalten hat (also alle verbliebenen Teilmengen einelementig sind). Als Beispiel nehmen wir die Codierung der auch in 4.9 behandelten Quelle aus 1.3 (bzw. 3.3c). Das Codierverfahren ist in Figur 18 dargestellt; es führt zum Code von Figur 16b (mit L und O vertauscht).

<table>
<tr><td>Quellensymbole (geordnet nach Wahrscheinlkt.)</td><td>Wahrschein-lichkeiten</td><td>Summe d.Wkt'en (von unten nach oben)</td><td colspan="3"></td></tr>
<tr><td>W</td><td>0,38</td><td>1,00</td><td colspan="3">O</td></tr>
<tr><td>S</td><td>0,3</td><td>0,62</td><td rowspan="3">L</td><td colspan="2">O</td></tr>
<tr><td>G</td><td>0,19</td><td>0,32</td><td rowspan="2">L</td><td>O</td></tr>
<tr><td>H</td><td>0,13</td><td>0,13</td><td>L</td></tr>
</table>

Figur 18: Codierung nach Fano (Beispiel)

Im allgemeinen erzielt man durch dieses Verfahren der Codierung nach Fano eine gute, allerdings nicht immer eine optimale Codierung.

4.11 Das zweite Verfahren, das wir behandeln, ist die **Codierung nach Huffman**. Hier werden nicht wie beim Fano-Verfahren die Signalmengen sukzessive geteilt, sondern jeweils zwei Signale (bzw. zwei Endecken des zugehörigen Baumes) "verschmolzen".

Bei der *Huffman-Codierung* verfährt man folgendermaßen: Zunächst ordnet man die Quellenzeichen ihrer Wahrscheinlichkeit nach[7]. *Dann faßt man zwei Quellensymbole der geringsten Wahrscheinlichkeit zusammen* (hier hat man evtl. Wahlmöglichkeiten) und ersetzt sie durch ein neues Zeichen. (Dieses ordnet man gemäß seiner Wahrscheinlichkeit ein.)[8] Mit der so entstandenen gedachten neuen Quelle verfährt man analog, bis man zu einer Quelle mit 2 Signalen gelangt.

Wieder haben wir die Quelle aus 1.3 als Beispiel herangezogen, s. Figur 19 (auf der nächsten Seite). Ein weiteres Beispiel folgt in 5.2.
Wir erhalten (mit der unwesentlichen Vertauschung von G mit H) den Baum bzw. die Codierung aus Figur 16b.

4.12 Man kann nun allgemein zeigen:

Zu gegebener Quelle mit Wahrscheinlichkeitsverteilung hat keine Präfixcodierung der Einzelzeichen kleinere mittlere Codewortlänge als die Huffman-Codierung. Die Huffman-Codierung ist also eine *optimale Codierung*.

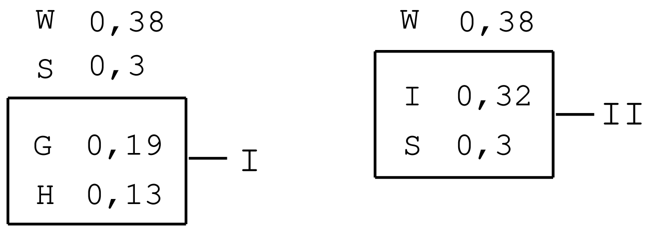
Beweis-Idee: Man zeigt durch vollständige Induktion, daß die beim i-ten Schritt entstandene Quelle optimal codiert ist. (Hierbei beginnt man bei der Quelle mit 2 Symbolen). (Ausführlicher Beweis s. z.B. SCHULZ [22] oder[15]p.44.)□

4.13 Wir vermerken noch, daß bei 2^m Signalen der Wahrscheinlichkeit $1/2^m$ die Huffman-Codierung zu Codewörtern der Länge m führt (vgl. 2.6).
Literaturhinweise: OBERSCHELP [19], KAMEDA & WEIHRAUCH [10], SCHULZ [15], [22], TOPSØE [17], WEFELSCHEID [23].

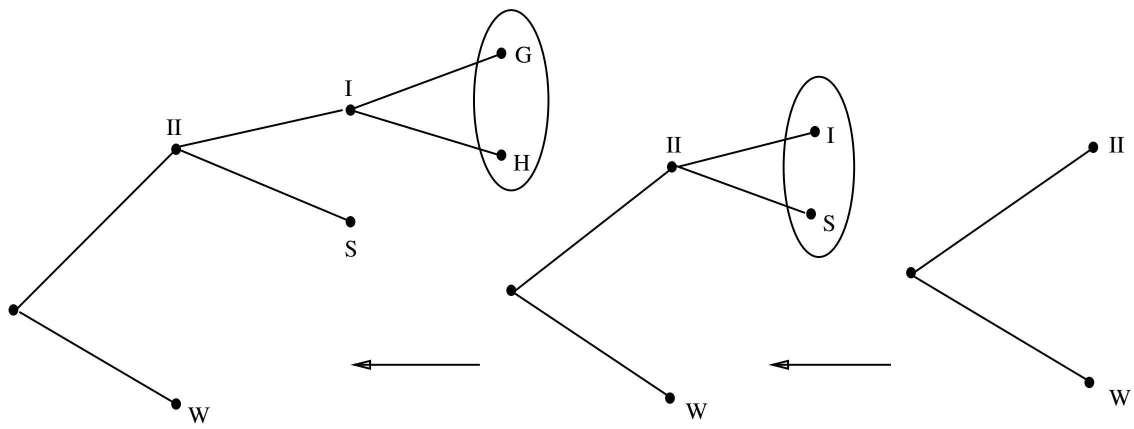
5. Mittlerer Codieraufwand und Entropie

5.1 Wie wir in 2.6 bemerkt hatten, stimmt bei einer Quelle mit 2^m Signalen gleicher Wahrscheinlichkeit die durch ein Signal gelieferte Information und damit die Entropie überein mit der minimalen mittleren *Codewortlänge*. Daß dies nicht allgemein gilt, haben wir an einigen Beispielen gesehen. Wir fragen uns aber,

[7] Dies hat hier lediglich schreibtechnische Gründe.
[8] Auch dies nur der besseren Übersicht wegen.



Figur 19a: Codierung nach Huffman (Beispiel)



Figur 19b: zugehörige Codebäume

ob vielleicht doch ein Zusammenhang besteht. In der Tat läßt sich zeigen, daß die mittlere Codewortlänge durch die Entropie der Quelle beschränkt ist. Es gilt sogar:

> Ist X eine Quelle mit Wahrscheinlichkeitsverteilung $(p_1, \ldots, p_N)$ und $p_i > 0$ für alle $i = 1, \ldots, N$, so gilt
> (5.1.1) $$H(X) \leq \bar{l}_{\min(X)} \; < \; H(X) + 1$$

Beweis-Andeutung (vgl. TOPSØE [17]p.24,25, s. auch SCHULZ [15]).
Man zeigt zunächst, daß ein aus N Wörtern der Längen $l_1, \ldots, l_N$ bestehender Präfix-Code genau dann existiert, wenn gilt:

$$\sum_{i=1}^{N} 1/2^{l_i} \leq 1 \; (= \sum_{i=1}^{N} p_i) \qquad (\textit{Kraftsche Ungleichung})\,.$$

Einerseits folgt aus dieser Gleichung mit Eigenschaften der log-Funktion ($ln\, x \leq x - 1$) nach geschickter Rechnung und Abschätzung

$$H(X) \leq \sum_{i=1}^{N} p_i \, l_i \quad \text{für jeden Präfixcode.}$$

Wählt man andererseits

$$-\log_2 p_i \leq l_i < -\log_2 p_i + 1 \quad (i = 1, \ldots, N),$$

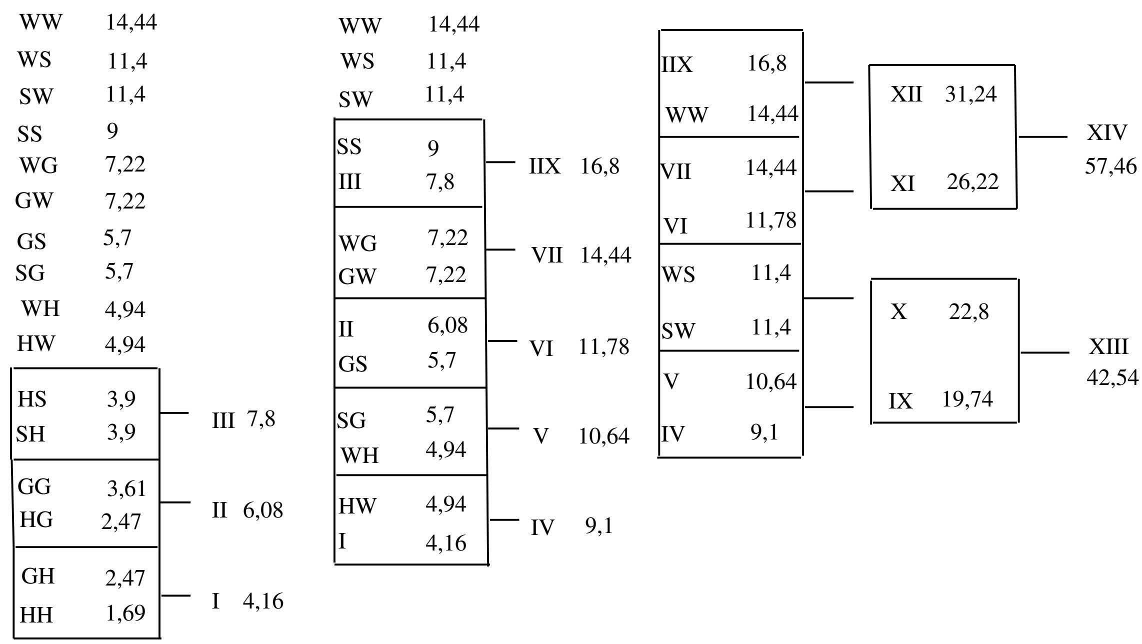
so ist für diese l_i die Kraftsche Ungleichung erfüllt, und somit existiert ein Präfixcode der entsprechenden Wortlängen.□

5.2 In 4.11 haben wir eine optimale Codierung der Quellenzeichen angeben können. Man muß dabei aber beachten, daß die so erreichte mittlere Codewortlänge $\bar{l}_{\min}$ minimal ist lediglich unter denjenigen Codierungen, die jedes Quellenzeichen gesondert codieren. Durch Zusammenfassen von Quellenzeichen zu Superzeichen (hier also Wörtern) und deren Codierung (*Wortcodierung*), läßt sich die mittlere Anzahl der Codesymbole pro Quellenzeichen (mittlerer Codieraufwand) noch weiter reduzieren.
Als **Beispiel** betrachten wir Paare von Signalen der Quelle aus 3.3c (bzw. 1.3), von der wir einschränkend annehmen, daß die Quellensignale unabhängig voneinander gesendet werden (s. Tabelle 4); (andernfalls müßten wir die Wahrscheinlichkeitsverteilung der Paare durch Beobachtung der Quelle herausfinden, könnten dadurch aber evtl. eine weitere Verkürzung der mittleren Codewortlänge erreichen).
Die Paare stellen wir uns als Zeichen einer neuen gedachten Quelle vor, die wir nun nach Huffman codieren (s. Figur 20). Als mittlere Codewortlänge für ein Paar ergibt sich $1/100 \cdot [(3,9+3,9+3,61+2,47+2,47+1,69) \cdot 5 + (9+7,22+7,22+5,7+5,7+4,94+4,94) \cdot 4 + (14,44+11,4+11,4) \cdot 3] = 3,808$, für das einzelne Zeichen der Quelle aus 3.3c also $1/2 \cdot 3,808 = 1,904$ [bit/Zeichen]. Dieser Wert ist kleiner als der Wert $1,94$ [bit/Zeichen], den wir bei der Einzelcodierung als mittlere Codewortlänge erhalten hatten (s. 4.9), aber immer noch größer als die Entropie der Quelle (1,89 [bit/Zeichen], s. 3.3c).

Paare von Quellenzeichen	WW	WH	WG	WS	HW	HH	HG	HS
Wahrscheinlichkeit in %[9]	$0,38^2 =$ 14,44	$0,38 \cdot 0,13 =$ 4,94	7,22	11,4	4,94	1,69	2,47	3,9

GW	GH	GG	GS	SW	SH	SG	SS
7,22	2,47	3,61	5,7	11,4	3,9	5,7	9

Tabelle 4: Wahrscheinlichkeitsverteilung der Paare von Zeichen (zur Quelle von Tab. 2)



Figur 20 a): Huffman-Codierung der "Quelle" aus Tabelle 4 (Fig. 20 b s. nächste Seite)

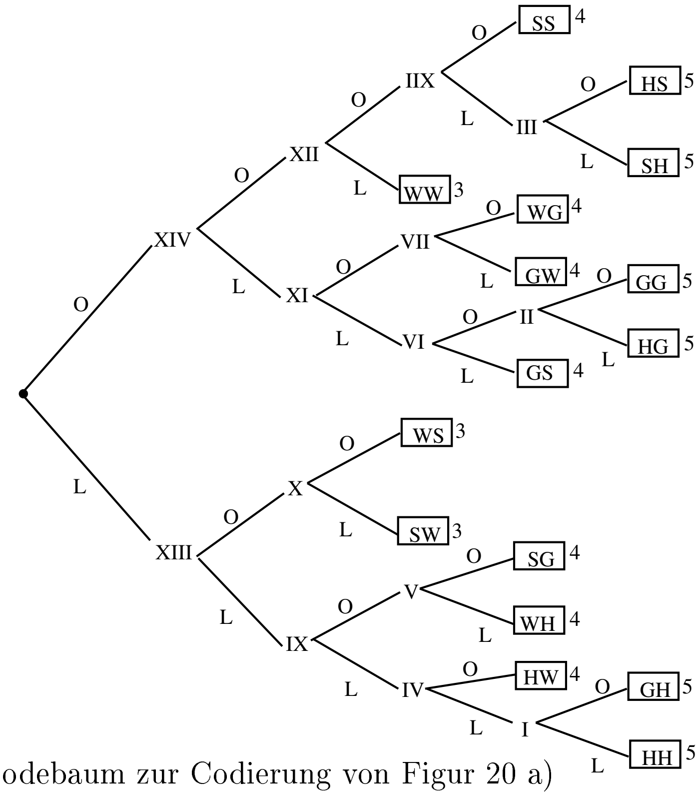
In Figur 21 ist das konkrete Bild aus 1.3 (zeilenweise abgetastet) mit der Codierung gemäß Figur 20 dargestellt. Die gemittelte Wortlänge pro Einzelsignal ist nun 2, ebenfalls eine Verbesserung gegenüber Einzelcodierung (- obwohl die Paare WS, SW geringer Wortlänge nicht auftraten).

5.3 Natürlich können wir nun versuchen, anstatt Einzelzeichen oder Paaren von Zeichen auch Tripel oder Quadrupel von Quellensignalen zu codieren. Der Aufwand zum Entwurf der Codierung würde allerdings immer größer: Bei den Quadrupeln zum Beispiel hätten wir schon $4^4 = 256$ Wörter im Code.

Es erhebt sich dabei die Frage, ob bei steigender Länge der gebildeten Quellensignal-Wörter auch allgemein eine Reduzierung der mittleren Codewortlänge pro Quellenzeichen möglich ist.

Eine Antwort darauf gibt der folgende 1. Hauptsatz der Informationstheorie (Fundamentalsatz über die Quellencodierung):

[9] Wir runden hier nicht, um die Paarcodierung besser mit der Einzelcodierung vergleichen zu können.



Figur 20 b): Codebaum zur Codierung von Figur 20 a)

$$\begin{array}{cccccccc}
O & O & O & O & O & O & O & O \\
L & L & O & O & L & L & O & L \\
O & L & O & O & L & L & O & L \\
O & O & L & L & L & L & L & O
\end{array}$$

Figur 21: Wortcodiertes Bild zu Figur 3

Fundamentalsatz über die Quellencodierung (SHANNON):
Es sei X eine Quelle ohne Gedächtnis (d.h. eine Quelle, die ihre Signale unabhängig voneinander sendet). Dann läßt sich eine Codierung durch einen binären Präfixcode erreichen, deren mittlerer *Codieraufwand* pro Quellenzeichen beliebig nahe bei der *Entropie* H(X) liegt. Dazu faßt man die Signale der Quelle zu Wörtern geeignet großer Länge k zusammen.

Die mittlere Codewortlänge pro Quellensignal ist also nicht nur durch die Entropie nach unten beschränkt, sondern läßt sich theoretisch auch beliebig nahe an diese Schranke drücken.
Beweis-Skizze: Statt der Quelle X mit Zeichen $x_1, \ldots, x_N$ betrachtet man die Super-Quelle X^k, deren Zeichen gerade die Wörter der Länge k über $\{x_1, \ldots, x_N\}$ sind; wegen der Unabhängigkeit der Signale gilt für diese Quelle

$$\boldsymbol{p}\,(a_1, \ldots, a_k) = \boldsymbol{p}\,(a_1) \cdots \boldsymbol{p}\,(a_k) \quad \text{mit} \quad a_1, \ldots, a_k \quad \text{aus} \quad \{x_1, \ldots, x_N\}\,.$$

Nach kurzer Rechnung ergibt sich $H(X^k) = k \cdot H(X)$; aus 5.1 folgt damit

$$H(X) = 1/k \cdot H(X^k) \leq 1/k \cdot \bar{l}_{\min(X^k)} < 1/k \cdot [H(X^k) + 1] = H(X) + 1/k.$$

Die Länge des Intervalls $[\,H(X), H(X) + 1/k\,[$ kann durch genügend großes k beliebig klein gemacht werden.□

5.4 Die Differenz zwischen dem mittleren Codieraufwand und der Entropie der codierten Quelle

$$\bar{l} - H(X)$$

heißt *(absolute) Redundanz der Codierung*; sie ist ein Maß dafür, wie komprimiert die Information der Quelle übertragen wird. Meist betrachtet man aber die auf die Codessymbole bezogene relative **Redundanz der Codierung**

(5.4.1) $$1 - H(X)/\bar{l} \quad ;$$

hierbei läßt sich $H(X)/\bar{l}$ deuten als die im Mittel pro Codesymbol übertragene Information. Von der erwähnten Redundanz ist die (relative) **Redundanz der Quelle** zu unterscheiden:

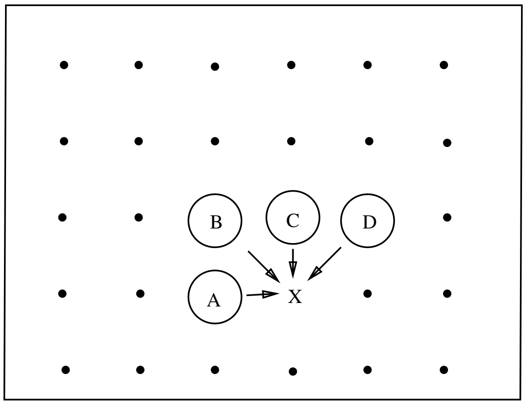
(5.4.2) $$1 - H(X)/H_{\max}$$ (oft in % ausgedrückt);

sie ist ein Maß dafür, wie die (unter Berücksichtigung von Abhängigkeiten in der Signalfolge) ermittelte Entropie pro Quellenzeichen von der Entropie einer Quelle mit N gleichwahrscheinlichen Symbolen abweicht; (sie mißt damit die Vorinformation, die man durch die Kenntnis hat, daß es sich um die Quelle X mit den entsprechenden Wahrscheinlichkeiten handelt).

5.5 Für die 30 Buchstaben (einschließlich Zwischenraum und Interpunktion) der **deutschen (Schrift-) Sprache** (vgl.1.4) ergibt sich z.B. $H_{\max} = \log_2 30 \approx 4,91$ [bit/Zeichen]; die Entropie unter Berücksichtigung der Häufigkeit der einzelnen Buchstaben ist aber kleiner gleich 4,11 , und nach Berücksichtigung der Häufigkeit von Buchstaben-Paaren, -Tripeln usw. erhält man $H \approx 1,6$ [bit/Zeichen] (vgl. BAUER & GOOS [1]p.46-48 bzw. RUPPRECHT [14]p.177ff). Es ergibt sich so eine relative Redundanz von $1 - 1,6/4,91 \approx 67$ %; dies läßt sich dahingehend interpretieren, daß in einem geschriebenen deutschen Text mehr als die Hälfte der Buchstaben überflüssig ist; und dabei ist die semantische Struktur (Grammatik und Sinn) noch nicht berücksichtigt. Diese Redunanz findet ihren Ausdruck darin, daß sich bekanntlich ein Text oft auch entziffern läßt, wenn er viele Druckfehler oder Lücken enthält.

5.6 Außer in §1 und 5.5 sind wir bisher immer von solchen Quellen ausgegangen, die ein Signal unabhängig von den zuvor aufgetretenen Signalen senden (Quelle ohne Gedächtnis). Bei vielen Anwendungen trifft dieses Modell jedoch nicht zu; als Beispiel seien neben der Sprache auch die Bildübertragung bei Fernsehen und Video genannt. Wegen großflächiger Bildteile treten dort Sprünge weniger auf, als sie bei Gleichverteilung zu erwarten wären. Man kann dann z.B. die Sprünge von einem Bild zum nächsten quantitativ erfassen und codieren (*Codierung der*

Differenzen). Üblich sind jedoch auch andere **Prädiktionen** (Vorhersagen) der zu erwartenden Quellensymbole innerhalb eines Bildes (z.Bsp. nach mehreren Nachbarstellen, s. Figur 22) oder von Bild zu Bild, bei bewegten oder unbewegten Zonen.



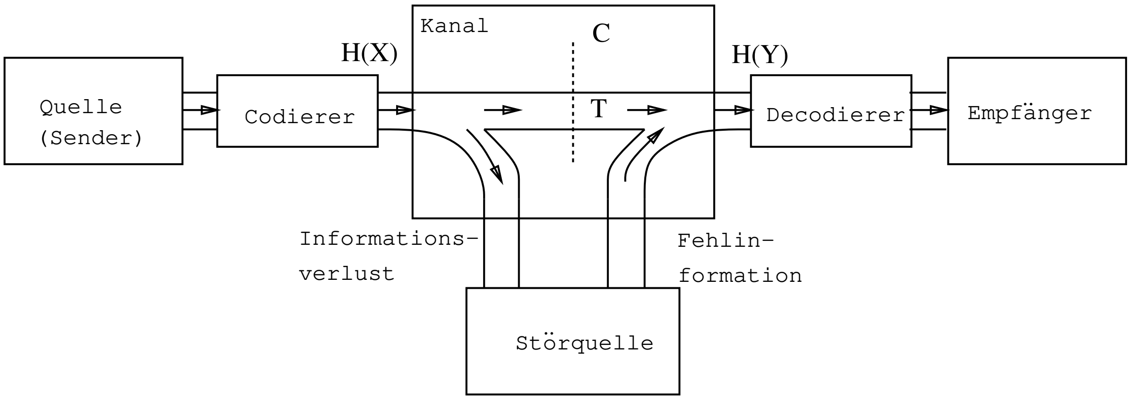
Figur 22: Prädiktion innerhalb eines Bildes

Es sei vermerkt, daß sich auch für Quellen mit statistischen Abhängigkeiten ein Entropie-Begriff bilden und eine entsprechende Theorie entwickeln läßt.

Literaturhinweise zu §5: BAUER & GOOS [1], HENZE & HOMUTH [7],[8], HOFMANN [9], KAMEDA & WEIHRAUCH [10], RUPPRECHT [14], SCHULZ [15],[22], TOPSØE [17], WEFELSCHEID [23].

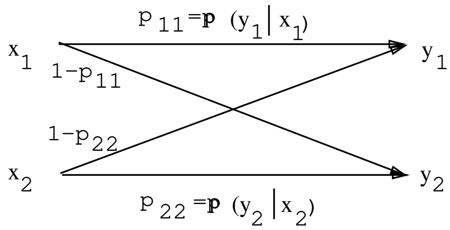
6. Übertragung durch einen diskreten Kanal

6.1 Wir haben Codierungen von Quellen behandelt, Codierungen, die zum Ziel haben, Nachrichten mit möglichst wenig Zeichen (‚die hohe Informationen tragen,) darzustellen (*Kompression durch Quellen-Codierung*). Diese Nachrichten sollen schließlich über einen **Kanal** (Telefon- oder andere Nachrichtenleitungen, z.B. Kabel oder Funkstrecken) zu einem Empfänger übertragen werden (Figur 23). Solche Kanäle sind i.a. durch atmosphärisches Rauschen oder Reflexionen gestört. Man stellt sich eine *Störquelle* als dafür verantwortlich vor. Wir betrachten nun Modelle dieser Kanäle.
Am Eingang des Kanals sei (schon nach Quellencodierung) eine Quelle X mit den Zeichen $x_1 = O$ und $x_2 = L$ (*Kanal–Eingabe–Alphabet*) und der Entropie $H(X)$ gegeben. Auch den *Ausgang des Kanals* können wir als Quelle Y auffassen, hier der Einfachheit halber mit dem gleichen Alphabet $y_1 = O$ und $y_2 = L$ und der Entropie $H(Y)$.



Figur 23: Schema eines Nachrichtenübertragungssystems mit gestörtem Kanal.

Infolge einer Störung sei ein Teil der empfangenen Signale von den gesendeten verschieden. Erneut vereinfachend wollen wir annehmen, daß die Sendung des Zeichens x_i das Zeichen y_j mit einer festen Wahrscheinlichkeit $p_{ij} = \boldsymbol{p}(y_j|x_i)$ (Wahrscheinlichkeit von y_j unter der Bedingung x_i) hervorruft (s. Figur 24)[10]. Außerdem habe der Kanal kein Gedächtnis, so



Figur 24: Binärer Kanal

daß also früher übertragene Signale diese Wahrscheinlichkeiten nicht beeinflussen. Als Beispiel betrachten wir den Fall, daß $p_{11} = p_{22}$ ist, (den sogenannten *binären symmetrischen Kanal*, BSC).

6.2 Durch die Störquelle geht im Kanal Information verloren; denn obwohl wir die Signale bei Y beobachten können, kennen wir die wirklich bei X gesendeten Signale nicht genau; die verlorene mittlere Information pro Zeichen heißt *Äquivokation* oder *Vieldeutigkeit des Kanals*; wir können diese auffassen als die Entropie einer gedachten Quelle des *Informationsverlustes*; um diesen Verlust zu quantifizieren, stellen wir uns eine Quelle mit Signalen (x_i, y_j) vor, bei der

[10] In der Praxis hingegen sind oft mehrere Bits hintereinander gestört; man spricht dann von "burst errors".

y_j jeweils bekannt und x_i unbekannt ist; die zugehörige (*bedingte) Entropie* wird mit $H(X|Y)$ bezeichnet.

Von Interesse ist nun die mittlere Information pro Zeichen, die als Teil der ursprünglichen durch den Kanal gelangt; sie ist gleich

(6.2.1)	$T(X) = H(X) - H(X|Y)$

und heißt *Transinformation* (oder *Synentropie* oder Übertragungsgeschwindigkeit der Information pro Symbol).
Da man zeigen kann, daß $H(X) + H(Y|X) = H(Y) + H(X|Y)$ gilt, folgt auch:

(6.2.2) $$T(X) = H(Y) - H(Y|X)$$

Zur Andeutung der Berechnung der Transinformation gehen wir davon aus, daß das Zeichen y_j am Ende des Kanals beobachtet wird; die Wahrscheinlichkeit dafür, daß x_i gesendet wurde ist $\boldsymbol{p}\,(x_i|y_j)$, die Information über diese Tatsache (bei bekanntem y_j) also $-\log_2 \boldsymbol{p}\,(x_i|y_j)$; diese Information geht im Kanal verloren (und dies auch, wenn $x_i = y_j$ ist, da man dies erst sicher weiß, wenn man das gesendete Signal erfährt). Da das beschriebene Ereignis (x_i, y_j) mit Wahrscheinlichkeit $\boldsymbol{p}\,(x_i, y_j) = \boldsymbol{p}\,(y_j) \cdot \boldsymbol{p}\,(x_i|y_j)$ eintritt, ergibt sich

(6.2.3)	$H(X|Y) = -\sum_{i,j=1}^{N} \boldsymbol{p}\,(y_j) \cdot \boldsymbol{p}\,(x_i|y_j) \log_2 \boldsymbol{p}\,(x_i|y_j)\,.$

Für das **Beispiel** des *binären symmetrischen Kanals* erhalten wir aus der (6.2.3) entsprechenden Formel für $H(Y|X)$ mit $p_{11} = p_{22} = p$ und $\boldsymbol{p}\,(x_i) = p_i$ die Gleichung

$$\begin{aligned} H(Y|X) =& -p_1 p_{11} \log p_{11} - p_1 p_{12} \log p_{12} - p_2 p_{21} \log p_{21} - p_2 p_{22} \log p_{22} = \\ & - (p_1 + p_2) p \log p - (p_1 + p_2)(1-p) \log(1-p) = \\ & - p \log_2 p - (1-p) \log_2 (1-p) \\ & = H(p, 1-p) \end{aligned}$$

und damit

$$T(X) = H(Y) + p \log_2 p + (1-p) \log_2 (1-p)\,.$$

(Speziell ist $T(X) = H(Y) = H(X)$, falls $p = 1$, der Kanal also ungestört ist, bzw. $T(X) = H(Y) - 1 \leq 0$, falls $p = 1/2$ ist und der Kanal so gestört, daß man, statt zu übertragen, eine Münze werfen könnte.)

Die Transinformation hängt aber nicht nur vom Kanal, sondern auch von der Wahrscheinlichkeitsverteilung der Quelle X ab. Bei gegebenem Kanal ist daher die maximal mögliche Transinformation (über alle möglichen binären Quellen ermittelt) von großer Bedeutung.

(6.2.4)	$C = \max\{T(X) \mid X$ binäre Quelle$\}$ heißt die **Kapazität des Kanals**.

In unserem *Beispiel* des BSC ist $H(Y|X)$ (ausnahmsweise) unabhängig von X (s. o.) und daher $T(X)$ maximal für maximales $H(Y)$; für eine gleichverteilte Quelle ($p_1 = p_2 = 1/2$) ergibt sich $\boldsymbol{p}\,(y_j) = \boldsymbol{p}\,(x_1) \cdot \boldsymbol{p}\,(y_j|x_1) + \boldsymbol{p}\,(x_2) \cdot \boldsymbol{p}\,(y_j|x_2) = p_1 p_{1j} + p_2 p_{2j} = 1/2 \cdot (p_{1j} + p_{2j}) = 1/2$ und damit $H(Y) = 1$, der maximal mögliche Wert. Folglich gilt für die Kapazität des binären symmetrischen Kanals:

(6.2.5) $$C_{BSC} = 1 + p \log_2 p + (1-p) \log_2 (1-p) .$$

Anmerkung: Die Größen $H(X)$, $T(X)$ und C werden oft auch auf eine Übertragungszeit-Einheit bezogen; man spricht dann von Informationsfluß, Transinformationsfluß und wieder von Kanalkapazität. Nimmt man o. B. d. A. an, daß pro Zeiteinheit ein Symbol übertragen wird, so sieht man, daß die Theorie dieser Größen sich nicht wesentlich von der geschilderten unterscheidet.

6.3 Wie der Name vermuten läßt, ist die Kanalkapazität eine Schranke für die Möglichkeit, mit diesem Kanal zugleich informationsreich und fehlerarm zu übertragen. Dies folgt u. a. aus dem folgenden 2. Hauptsatz der Informationstheorie:

Fundamentalsatz über Kanalcodierung (SHANNON):
Gegeben sei ein gestörter diskreter Kanal ohne Gedächtnis der Kapazität C (genauer: ein sogenannter "stationärer Kanal ohne Vorgriff mit endlichem Gedächtnis") und eine diskrete Informationsquelle X mit Entropie $H(X)$ (genauer eine "ergodische" Quelle).
Ist dann $H(X) < C$, so kann man die Nachrichten von X so codieren, daß bei Sendung durch den Kanal

1. der Informationsverlust $H(X|Y)$ beliebig klein wird (und damit die Übertragungsgeschwindigkeit beliebig nahe an die Entropie der Quelle kommt)
2. die übertragenen Nachrichten aus den am Kanalausgang ankommenden mit beliebig kleiner Fehlerwahrscheinlichkeit bestimmt werden können.

Anmerkungen:

1. Daß geringer Informationsverlust (im Kanal) (gemäß 1.) und damit hohe Transinformation auch eine gute Übertragungsgenauigkeit (gemäß 2.) bedeutet, ist nach unserer Interpretation nicht verwunderlich, aber beweisbedürftig.

2. Falls der Sender (bezogen auf den Kanal) nicht mit zu großer mittlerer Information pro Zeichen sendet, lassen sich also theoretisch Störungen des Kanals (egal wie diese aussehen, solange die Entropie noch kleiner als die Kapazität C ist) ausgleichen.

3. Für den binären symmetrischen Kanal bedeutet das z.B. für eine mit gleicher Wahrscheinlichkeit sendende Quelle, daß es zu $\varepsilon > 0$ und R mit

$0 < R < C$ sowie genügend großem n einen Code $\mathbf{C} \subseteq \{0,1\}^n$ mit einer Informationsrate (–diese ist definiert als der Wert $1/n \cdot \log_2 |\mathbf{C}|$) nahe bei R gibt derart, daß sich die am Kanalausgang empfangenen Wörter mit einer Wahrscheinlichkeit größer als $1 - \varepsilon$ richtig decodieren lassen.

Zum Beweis der Sätze verweisen wir z.Bsp. auf HENZE & HOMUTH [7], KAMEDA & WEIHRAUCH [10] und HOFMANN [9], vermerken hier nur, daß die erwähnte Codierung eine SZufallscodierungïst. Der Beweis liefert daher kein Verfahren für die Codierung im konkreten Fall.

Literaturhinweise zu 6.: HENZE & HOMUTH [7], HOFMAN [9], KAMEDA & WEIHRAUCH [10], RUPPRECHT [14], TOPSØE [17], WEFELSCHEID [23].

7. Kanalcodierung: Fehlerkorrigierende Codes

Wir wollen in diesem Paragraphen erläutern, wie durch "*gezieltes Hinzufügen von Redundanz*" die Fehleranfälligkeit der Übertragung über einen gestörten Kanal reduziert werden kann. Dabei beschränken wir uns auf binäre Codes, deren Wörter alle die gleiche Länge haben (binäre *Block-Codes*) und bei diesen auf einfache Fälle. (Die Behandlung der in der Praxis üblichen Codes würde den Rahmen dieses Artikels sprengen. Das Verständnis der in der Praxis üblichen Codes setzt zum Teil größere Vorkenntnisse in Mathematik voraus.)

7.1 Wir gehen davon aus, daß die Signale der Quelle bereits binär (quellen-) codiert sind. Den Strom der gelieferten Signale teilen wir zum Beispiel in Wörter der Länge 4 (*Tetraden*) auf. Diese seien ungefähr gleichwahrscheinlich. Der Kanal sei ein binärer symmetrischer Kanal mit $p_{11} > 1/2$ (- damit ist ein Fehler an einer einzigen Stelle eines übertragenen Wortes wahrscheinlicher als es Fehler an zwei oder mehreren gegebenen Stellen sind.)
Die Tetraden wollen wir nun erneut codieren und zwar so, daß *ein* Fehler pro Codewort am Ende des Kanals erkannt und korrigiert werden kann.

7.2 Eine einfache Codierung, die unsere Forderung nach Korrekturmöglichkeit von Einzelfehlern erfüllt, ist die mit einem *Wiederholungs-Code*. Senden wir z.B. statt

$$OOOL \quad \text{das Wort} \quad OOOL\,OOOL\,OOOL\,,$$

so kann man *einen* Fehler im Codewort erkennen:

$$O\underline{O}OL\,O\underline{\underline{L}}OL\,O\underline{O}OL$$

Da zwei Fehler unwahrscheinlicher sind als einer (s.o.), korrigieren wir durch Mehrheitsentscheidßum richtigen Codewort.

Allgemein korrigiert man zu dem Codewort, das sich vom empfangenen Wort in möglichst wenig Komponenten unterscheidet; diese Art der Korrektur führt jeweils zu einer der am wahrscheinlichsten gesendeten Nachrichten (**Maximum-Likelihood-Decodierung**).

7.3 Da der Wiederholungscode viel zu aufwendig ist (Verdreifachung der Bit-Zahl), suchen wir nach anderen Möglichkeiten.
Von Anwendungen her (Kontonummer, ISBN-Nummer, s.§8) kennen wir das Anhängen von *Prüfziffern*. Im binären Fall kann man je nach **Parität** (gerader oder ungerader Anzahl der Zeichen L) dem Wort ein O (bei gerader Parität) bzw. ein L (bei ungerader Parität) anhängen, also zum Beispiel

$$\begin{array}{lll} OOOL\mathbf{L} & \text{für} & OOOL \\ OLLO\mathbf{O} & \text{für} & OLLO \quad \text{usw.} \end{array}$$

senden.
So entsteht ein Code, dessen sämtliche Wörter gerade Parität haben. Eine ungerade Anzahl von Fehlern in einem Wort ändert die Parität, läßt sich also durch *Paritätskontrolle* erkennen. Eine Fehlerkorrektur ist aber nicht möglich, da zum Beispiel $O\,O\,O\,L\,O$ aus $O\,O\,O\,L\,\underline{L}$ oder aus $O\,O\,O\,\underline{O}\,O$ oder anderen Codewörtern entstanden sein kann. Die Codewörter liegen "zu nahe" beieinander.

7.4 Die letzte Bemerkung wollen wir etwas präzisieren. Dazu führen wir nach R. W. HAMMING einen *Abstandsbegriff* ein:

Seien $\boldsymbol{a} = a_1 \dots a_n$ und $\boldsymbol{b} = b_1 \dots b_n$ Wörter der Länge n über $\{O, L\}$; dann bezeichnet $d(\boldsymbol{a}, \boldsymbol{b})$ die Anzahl der unterschiedlichen Komponenten von $\boldsymbol{a}$ und $\boldsymbol{b}$. Diese Zahl $d(\boldsymbol{a}\ , \boldsymbol{b})$ heißt der **HAMMING-Abstand** zwischen $\boldsymbol{a}$ und $\boldsymbol{b}$.

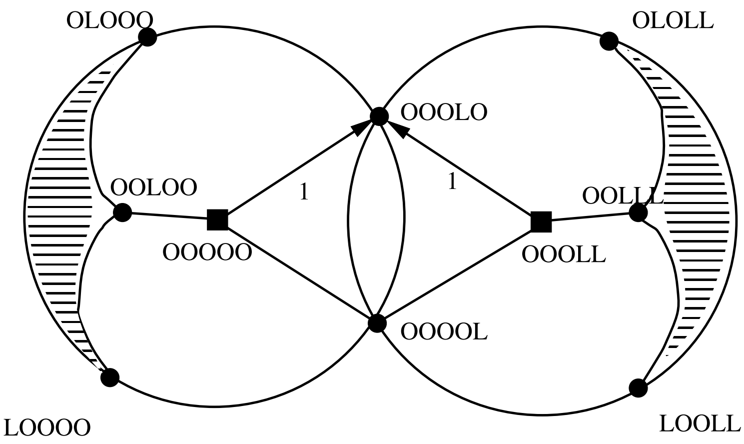
Beispiel: $d\,(OOOLO, OOOLL) = 1 = d\,(OOOLO, OOOOO)\,.$

Anmerkung: Die Menge $\{O, L\}^n$ zusammen mit dem Abstand d heißt *Sequenzraum*; dieser spielt u.a. auch bei der Theorie zu den Beispielen von 1.5 eine Rolle.

Wird ein Codewort $\boldsymbol{c}$ übertragen und ereignet sich in einer Komponente ein Fehler, so entsteht ein Wort $\boldsymbol{a}$, das von $\boldsymbol{c}$ den Abstand 1 hat. Alle solchen Wörter vom Abstand 1 von $\boldsymbol{c}$ bilden zusammen mit $\boldsymbol{c}$ eine Menge

$$K_1(\boldsymbol{c}) = \{\boldsymbol{a} \in \{O, L\}^n \,|\, d(\boldsymbol{a}, \boldsymbol{c}) \leq 1\}\,,$$

die wir uns als **Kugel** vom Radius 1 um $\boldsymbol{c}$ "vorstellen" sollten (und die auch so bezeichnet wird).
Im Beispiel liegt das empfangene Wort $O\,O\,O\,L\,O$ sowohl in der Kugel vom Radius 1 um $O\,O\,O\,O\,O$ als auch in der um $O\,O\,O\,L\,L$ (vgl. Figur 25), sodaß nicht klar ist, zu welchem Codewort korrigiert werden soll. Wie schon in 7.3 gesehen, liegen die Codewörter $O\,O\,O\,O\,O$ und $O\,O\,O\,O\,L\,L$ zu nahe beieinander: Sie haben nur Hamming-Abstand 2 voneinander.



Figur 25: "*Kugeln*" vom Radius 1 um zwei Codewörter

Wir suchen nun Codes, deren Wörter mindestens Abstand 3 voneinander haben.

7.5 Ausgangspunkt für die Konstruktion ist die Idee, die Tetraden im Quadrat anzuordnen und zeilen- und spaltenweise Paritätskontroll-Symbole anzuhängen (Figur 26).

a_1	$-$	a_2	$\longrightarrow$	a_5
$\vert$		$\vert$		
a_3	$-$	a_4	$\longrightarrow$	a_6
$\downarrow$		$\downarrow$		
a_7		a_8		

Figur 26: Beispiel mehrfacher Paritätskontrollzeichen

Aus $O\,O\,O\,L$ zum Beispiel erhält man so $O\,O\,O\,L\,O\,L\,O\,L$. Der entstandene Code hat die Länge 8 und läßt 1 Fehler erkennen und korrigieren: Fehlerhafte Übertragung der Tetrade verursacht ungerade Parität genau in der Zeile und der Spalte, in der sich der Fehler befindet. Zur Fehlererkennung und Korrektur sind dabei 4 **Paritätskontrollen** nach dem in Figur 27 angegebenen Schema vorzunehmen.

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
1. Kontrolle	X	X			X			
2. Kontrolle			X	X		X		
3. Kontrolle	X		X				X	
4. Kontrolle		X		X				X

Figur 27: Paritätskontrollen zum Code aus Figur 26.

7.6 Um das angegebene Verfahren verallgemeinern zu können, führen wir auf dem Alphabet gemäß dem möglichen Vorgehen bei den Paritätskontrollen eine "*Addition*" nach folgenden Regeln ein (Algebra der Paritäten oder Fehleraddition):

$$\begin{array}{ccccc} O & + & O & = & O \\ O & + & L & = & L \\ L & + & O & = & L \\ L & + & L & = & O \end{array}$$

In Übereinstimmung mit der Schreibweise in der Algebra schreiben wir im folgenden 1 anstatt L und 0 (Null) statt O (oh).
Nun fassen wir zusammen: Die *Codierung* der Tetrade $a_1a_2a_3a_4$ zu $a_1a_2a_3a_4a_5a_6a_7a_8$ läßt sich durch folgende Gleichungen beschreiben:

$$\begin{array}{ccccc} a_5 & = & a_1 & + & a_2 \\ a_6 & = & a_3 & + & a_4 \\ a_7 & = & a_1 & + & a_3 \\ a_8 & = & a_2 & + & a_4 \end{array}$$

Durch Umformungen sieht man: Der Code besteht aus allen Wörtern $a_1a_2 \ldots\ldots a_8$, die folgenden Gleichungen (**Kontrollgleichungen**) genügen:

$$\begin{array}{cccccccccccccccc} a_1 & + & a_2 & & & & & + & a_5 & & & & & & = & 0 \\ & & & & a_3 & + & a_4 & & & + & a_6 & & & & = & 0 \\ a_1 & + & & & a_3 & & & & & & & + & a_7 & & = & 0 \\ & & a_2 & + & & & a_4 & & & & & & & + \; a_8 & = & 0 \end{array}$$

Überprüft man bei einem empfangenen Wort die Gültigkeit dieser Gleichungen, so bedeutet dies eine *Paritätskontrolle* gemäß Figur 27. (Beachten Sie das Erscheinungsbild der linken Seite dieser Gleichungen und von Figur 27).

7.7 Das eben geschilderte Beispiel mit Wortlänge 8 ist aber noch nicht optimal: Wir können die Wortlänge noch weiter drücken, wie folgendes Beispiel eines Codes, des sogenannten (7,4)- **HAMMING-Codes** zeigt.

Bei der *Codierung* einer Tetrade $a_1a_2a_3a_4$ werden als Kontrollsymbole hinzugefügt:

$$\begin{array}{ccccccc} a_5 & = & a_1 & + & a_2 & + & a_3 \\ a_6 & = & a_2 & + & a_3 & + & a_4 \\ a_7 & = & a_1 & + & a_2 & + & a_4 \end{array}$$

Beispiel:

$$0\,0\,0\,1 \mapsto 0\,0\,0\,1\,0\,1\,1$$

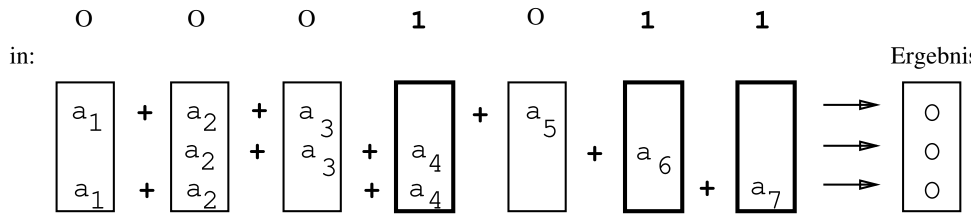
(Weitere Beispiele behandeln wir in 7.10).
Aus den Wörtern der Länge 4 entstehen Codewörter der Länge 7.

Die *Kontrollgleichungen* lauten nun

$$\begin{array}{ccccccccc} a_1 & +a_2 & +a_3 & & +a_5 & & & = 0 \\ & a_2 & +a_3 & +a_4 & & +a_6 & & = 0 \\ a_1 & +a_2 & & +a_4 & & & +a_7 & = 0 \end{array}$$

Genau die Codewörter erfüllen diese Gleichungen, s.z.Bsp. Figur 28.

Einsetzen:



Figur 28: Kontrolle eines Codewortes

7.8 Läßt sich nun *ein* Fehler in einem Codewort korrigieren? Und wenn ja, wie? Zur Beantwortung dieser Frage sind noch weitere Überlegungen erforderlich:
Ist der Fehler in der i-ten Komponente unterlaufen und sind alle anderen Komponenten richtig übertragen worden, so erhalten wir statt $a_1 \dots a_7$ ein Wort $a'_1 \dots a'_7$ mit $a'_i = a_i + 1$ und $a'_j = a_j$ sonst.

Wir setzen nun ein empfangenes fehlerhaftes Wort in die linke Seite der Kontrollgleichungen ein und erhalten:

$$\begin{array}{ccccccccc} a'_1 & +a'_2 & +a'_3 & & +a'_5 & & & = s_1 \\ & a'_2 & +a'_3 & +a'_4 & & +a'_6 & & = s_2 \\ a'_1 & +a'_2 & & +a'_4 & & & +a'_7 & = s_3 \end{array}$$

$s_1 s_2 s_3$ heißt dabei das **Syndrom** von $a'_1 a'_2 a'_3 a'_4 a'_5 a'_6 a'_7$.

Es ist allein aus dem empfangenen Wort ohne Kenntnis der gesendeten Nachricht berechenbar. Für ein Codewort wäre das Syndrom $0\,0\,0$; bei diesem geht man von unverfälschter Übertragung aus.

Setzen wir nun sukzessive für die (gemäß Voraussetzung einzige) Fehlerstelle i die Werte $1, 2, \ldots, 7$ ein und beachten, daß die Kontrollgleichungen für $a_1 \ldots a_7$ gelten, so erhalten wir folgende Syndrome:
Für $i = 1$:

$$\begin{aligned} s_1 &= a_1' + a_2' + a_3' + a_5' = & 1 + a_1 + a_2 + a_3 + a_5 &= 1 \\ s_2 &= a_2' + a_3' + a_4' + a_6' = & a_2 + a_3 + a_4 + a_6 &= 0 \\ s_3 &= a_1' + a_2' + a_4' + a_7' = & 1 + a_1 + a_2 + a_4 + a_7 &= 1 \end{aligned}$$

für $i = 2$:	für $i = 3$:	für $i = 4$:	für $i = 5$:	für $i = 6$:	für $i = 7$:
$s_1 = 1$	$s_1 = 1$	$s_1 = 0$	$s_1 = 1$	$s_1 = 0$	$s_1 = 0$
$s_2 = 1$	$s_2 = 1$	$s_2 = 1$	$s_2 = 0$	$s_2 = 1$	$s_2 = 0$
$s_3 = 1$	$s_3 = 0$	$s_3 = 1$	$s_3 = 0$	$s_3 = 0$	$s_3 = 1$

Dabei entspricht das Syndrom für ein Wort mit Fehler in der i-ten Komponente gerade den Koeffizienten von a_i in den Kontrollgleichungen!

Man stellt fest: Für $i = 1, \ldots, 7$ sind alle Syndrome $s_1 s_2 s_3$ verschieden und ungleich 0 0 0; daher gilt – immer vorausgesetzt, daß lediglich in der i-ten Komponente eine fehlerhafte Übertragung vorliegt: *Diese Fehler-Stellen-Nummer i ist eindeutig aus dem Syndrom des empfangenen Wortes ablesbar und damit korrigierbar.*

Wird am Kanal-Ende zum Beispiel 1 0 0 1 0 1 1 empfangen, so ermittelt der *Decodierer* - meist eine automatische Schaltung - das Syndrom 101, daraus die Fehlerstelle $i = 1$, korrigiert nun zu **0** 0 0 1 0 1 1 und meldet als empfangenes Quellen(code)- Signal 0 0 0 1 (*Fehlerkorrektur*).

7.9 Der behandelte *Hamming-Code* hat weitere interessante Eigenschaften:

a) Die (komponentenweise gebildete) *Summe zweier Codewörter* ist wieder ein Codewort. Damit ist der Code auffaßbar als ein Vektorraum über einem Körper mit 2 Elementen; für diesen besteht eine mathematische Theorie (cf. *Lineare Codes*).

b) Die Kugeln von Radius 1 um Codewörter sind nicht nur disjunkt, sondern füllen auch die Menge aller Wörter der Länge 7 aus, sodaß also jedes Wort decodiert werden kann. (Ein Code dieser Eigenschaft heißt *perfekt*).

c) Geht man von $a_1\ a_2\ a_3\ a_4\ a_5\ a_6\ a_7$ über zu $a_2\ a_3\ a_4\ a_5\ a_6\ a_7\ a_1$ (zyklische Vertauschung), so entsteht im vorliegenden Fall aus einem Codewort wieder ein Codewort. (Dies sieht man durch Betrachtung der Kontrollgleichungen). Ein Code mit dieser Eigenschaft heißt *zyklisch.* Er läßt sich meist durch ein Schieberegister auf technisch einfache Weise decodieren.

7.10 Zum Abschluß des Paragraphen codieren wir noch das Bild zu Figur 3 mit dem behandelten (7,4)-Hamming-Code. Dazu benutzen wir die Quellencodierung von Figur 21 mit 1 statt L und 0 statt O. Wir erhalten die Bitfolge aus Figur 29 ; benötigt werden 56

0 0 0 0 0 0 0	0 0 0 0 0 0 0
1 1 0 0 0 1 0	1 1 0 1 0 0 1
0 1 0 0 1 1 1	1 1 0 1 0 0 1
0 0 1 1 1 0 1	1 1 1 0 1 0 0

Figur 29: Figur 3 nach Quellen- und Kanalcodierung

Bit für die 16 Quellensignale, also 3,5 Bit pro Quellensignal; schon eine einfache Wiederholung jeden Bits nach der Wortcodierung (Figur 21) hätte 4 Bit pro Quellensignal benötigt. Nun können wir aber 1 Fehler pro Wort nicht nur erkennen, sondern auch korrigieren.

Die Bitrate wird günstiger, wenn man, wie technisch üblich, Codes wesentlich größerer Wortlängen benutzt; dann können auch mehrere Fehler pro Codewort korrigiert werden. Bekannt sind zum Beispiel andere *Hamming-Codes*, Bose-Chaudhuri-Hocquenghem-Codes (*BCH-Codes*), unter diesen die Reed-Solomon-Codes (*RS-Codes*), die Reed-Muller-Codes (*RM-Codes*), die Quadratischen-Rest-Codes (*QR-Codes*) und unter diesen die *Golay-Codes.*

Auf ein weiteres Codierungsverfahren (*Convolutional Codes*) können wir hier ebenfalls nicht eingehen.

Schließlich vermerken wir noch, daß der bei der *CD-Platte* verwandte Code CIRC (Cross Interleave Reed Solomon Code) sich aus zwei RS-Codes zusammensetzt und von der Platte mit einer Geschwindigkeit von 4,3218 Millionen Bits/sec abgetastet wird. Die maximale Länge eines vollständig korrigierbaren Bursts ist ungefähr 4000 Bits, die Länge eines interpolierbaren Bursts im schlimmsten Fall etwa 12.300 Bits ($\hat{=}$ 7,7 mm Spurlänge).

Literaturhinweise: BLAHUT[3], HEISE & QUADROCCHI[6], FURRER[5], HENZE & HOMUTH[8], KAMEDA & WEIHRAUCH[10], MACWILLIAMS & SLOANE[11], SCHULZ[15],[21].

8. Prüfzeichensysteme (Fehlererkennende Codes)

8.1 Auch bei dezimalen Kennummern $a_1 \dots a_n$ von Konten, Waren oder Personen wird oft eine Prüfziffer a_{n+1} derart angehängt, daß die Ziffern von $a_1 \dots a_{n+1}$ einer Prüfgleichung oder Prüfbedingung genügen. Fehlerhafte Übertragungen können dann entdeckt (und durch Rückfrage korrigiert werden).

8.2 **Europäische Artikel-Nummer**
Die Warenpackungen von Einzelhandelswaren sind meist mit der Europäischen

Artikelnummer (EAN) des Produkts versehen (und für opto-elektronische Lesegeräte, Scanner, durch den entsprechenden Strichcode gekennzeichnet.) Die EAN-Nummer $a_1a_2 \dots a_{13}$ setzt sich aus der Länderkennung a_1a_2, der Nummer $a_3 \dots a_7$ der produzierenden Firma, der Produktnummer $a_8 \dots a_{12}$ und der Prüfnummer a_{13} zusammen. Die letztere ist so gewählt, daß

$$a_1 + 3 \cdot a_2 + a_3 + 3a_4 + \dots + 3 \cdot a_{12} + a_{13}$$

eine Zehnerzahl (also durch 10 teilbar) ist.

Beispiel 4 0 2 5 7 0 0 0 0 1 0 2 3 führt zu
$4 + 2 + 3 \cdot 5 + 7 + 0 + 0 + 0 + 0 + 3 \cdot 1 + 0 + 3 \cdot 2 + 3 = 4 \cdot 10$

Einzelfehler (d.h. solche der Form: $a_1 \dots a_i \dots a_{13}$ wird als $a_1 \dots a_i' \dots a_{13}$ mit $a_i \neq a_i'$ übermittelt) lassen sich stets erkennen: Ist i ungerade, so ist $a_1 + \dots + a_i' + \dots + a_{13} = a_1 + \dots + a_i + \dots + a_{13}\ (-a_i + a_i')$ keine Zehnerzahl; ist i ungerade, so folgt dies für $a_1 + \dots + 3a_i' + \dots + a_{13} = a_1 + \dots + 3a_i + \dots a_{13} + 3(a_i' - a_i)$.

Zahlendreher (d.h. Fehler der Form: $a_1 \dots a_{i+1}a_i \dots a_{13}$ statt $a_1 \dots a_ia_{i+1} \dots a_{13}$) lassen sich nicht in allen Fällen erkennen. Für i ungerade ist zwar mit $a_1 + \dots + a_i + 3a_{i+1} + \dots + a_{13}$ und $a_1 + \dots + a_{i+1} + 3a_i + \dots + a_{13}$ auch die Differenz $3a_{i+1} - a_{i+1} + a_i - 3a_{i+1} = 2(a_{i+1} - a_i)$ Zehnerzahl; dies ist aber für $(a_{i+1} - a_i) = 5$ kein Widerspruch. (Entsprechendes gilt für i gerade.) Tatsächlich erfüllen 40 25700 00$\underline{\underline{16}}$27 und 40 25700 00$\underline{\underline{61}}$27 beide die Prüfbedingung.

8.3 **Internationale Standard Buchnummer**

Die *Internationale Standard Buchnummer (ISBN)* $b_1b_2 \dots b_{10}$, in der Herkunftsland, Verlag- und Buchtitel gekennzeichnet sind, ist nun durch Wahl der Prüfziffer b_{10} so konzipiert, daß

$$b_1 + 2b_2 + 3b_3 + \dots + 10\, b_{10} \qquad \text{durch 11 teilbar ist.}$$

Beispiel: $3 - 528 - 06419 - 6$ führt zu
$1 \cdot 3 + 2 \cdot 5 + 3 \cdot 2 + 4 \cdot 8 + 5 \cdot 0 + 6 \cdot 6 + 7 \cdot 4 + 8 \cdot 1 + 9 \cdot 9 + 10 \cdot 6 = 264 = 24 \cdot 11\,.$
Einzelfehler und Zahlendreher sind nun ausnahmslos erkennbar: Wären die richtige und die durch Einzelfehler gefälschte Zahl beide durch 11 teilbar, dann auch ihre Differenz
$(b_1 + \dots + ib_i + \dots + 10\, b_{10}) - (b_1 + \dots + ib_i' + \dots + 10\, b_{10}) = i\,(b_i - b_i')$; diese kann wegen $1 < i < 11$ nur für $b_i = b_i'$ Elferzahl sein. Ferner ist bei einem Zahlendreher
$(b_1 + \dots + ib_i + (i+1)b_{i+1} + \dots) - (b_1 + \dots + ib_{i+1} + (i+1)b_i + \dots) = b_{i+1} - b_i$ für $b_{i+1} \neq b_i$ ebenfalls keine Elferzahl. Also werden auch solche Fehler entdeckt.

Die besseren Möglichkeiten, Fehler zu entdecken, erkauft man bei diesem ISBN-System allerdings damit, daß man für b_{10} das Alphabet $\{0, \dots 9\}$ um ein (die

Zahl 10 repräsentierendes) Element X erweitert. Andernfalls könnten einige Zahlen $b_1 \dots b_9$ nicht durch eine der Prüfbedingung genügende Zahl b_{10} abgesichert werden.
Die behandelten Systeme haben also beide Nachteile; entweder kann man gewisse Fehlertypen nicht feststellen (die allerdings bei Scannern kaum vorkommen), oder man muß das Alphabet erweitern.

8.4 System der deutschen Geldscheine

Bei der Numerierung der Geldscheine der Deutschen Bundesbank hat man ein (schon vorher J. VERHOEFF bekanntes) anderes System verwandt. Dazu benutzt man auf der Menge $\{0, 1, 2, \dots, 9\}$ eine Verknüpfung $*$ mit der folgenden Verknüpfungstafel

*	0	1	2	3	4	5	6	7	8	9
0	0	1	2	3	4	5	6	7	8	9
1	1	2	3	4	0	6	7	8	9	5
2	2	3	4	0	1	7	8	9	5	6
3	3	4	0	1	2	8	9	5	6	7
4	4	0	1	2	3	9	5	6	7	8
5	5	9	8	7	6	0	4	3	2	1
6	6	5	9	8	7	1	0	4	3	2
7	7	6	5	9	8	2	1	0	4	3
8	8	7	6	5	9	3	2	1	0	4
9	9	8	7	6	5	4	3	2	1	0

Anmerkung: Nach eindeutiger Zuordnung der Ziffern $0, \dots, 9$ zu den 10 Deckabbildungen des regelmäßigen Fünfecks erhält man die Verknüpfung $*$ aus der Hintereinanderausführung der entsprechenden Abbildungen, vgl. [15].

Jeder Geldschein ist mit einer Kette von Ziffern und Buchstaben ("α-Zeichen") gekennzeichnet. Zur Prüfung werden die Buchstaben gemäß der folgenden Tabelle durch Zahlen ersetzt:

A	D	G	K	L	N	S	U	Y	Z
0	1	2	3	4	5	6	7	8	9

Es entsteht eine Nummer $a_1 a_2 \dots a_{11}$ mit $a_i \in \{0, \dots 9\}$. Die Prüfziffer ist nun so gewählt, daß diese Nummer der Prüfgleichung

$$T(a_1) * T^2(a_2) * \dots * T^{10}(a_{10}) * a_{11} = 0$$

genügt; dabei ist T die Abbildung mit der Zuordnungstabelle

i	0	1	2	3	4	5	6	7	8	9
$T(i)$	1	5	7	6	2	8	3	0	9	4

, meist als

$$T = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 1 & 5 & 7 & 6 & 2 & 8 & 3 & 0 & 9 & 4 \end{pmatrix}$$

geschrieben, und T^i die i-fache Hintereinanderausführung von T; zur Bequemlichkeit des Lesers geben wir T^i für $i = 2, \ldots 10$ an:

$$T^2 = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 5 & 8 & 0 & 3 & 7 & 9 & 6 & 1 & 4 & 2 \end{pmatrix} \quad T^3 = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 8 & 9 & 1 & 6 & 0 & 4 & 3 & 5 & 2 & 7 \end{pmatrix}$$
$$T^4 = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 9 & 4 & 5 & 3 & 1 & 2 & 6 & 8 & 7 & 0 \end{pmatrix} \quad T^5 = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 4 & 2 & 8 & 6 & 5 & 7 & 3 & 9 & 0 & 1 \end{pmatrix}$$
$$T^6 = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 7 & 9 & 3 & 8 & 0 & 6 & 4 & 1 & 5 \end{pmatrix} \quad T^7 = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 7 & 0 & 4 & 6 & 9 & 1 & 3 & 2 & 5 & 8 \end{pmatrix}$$
$$T^8 = \begin{pmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{pmatrix} \quad T^9 = T \qquad \text{und} \quad T^{10} = T^2.$$

Beispiel: Ein 50.–DM-Schein hat die Kennzeichnung AN9263843Z5. Durch Ersetzen der Buchstaben wird daraus 05926384395. Wir berechnen

$$P = T\,(0) * T^2(5) * T^3(9) * T^4(2) * T^5(6) * T^6(3) * T^7(8) * T^8(4) * T^9(3) * T^{10}(9) * 5$$

und erhalten

$$P = 1*9*7*5*3*3*5*4*6*2*5 = 5*2*1*6*9*5 = 8*7*4 = 1*4 = 0 \quad .$$

Beim Abschreiben der Kennzeichnung ist uns also kein Einzelfehler und (außer evtl. an den letzten beiden Stellen) auch kein Zahlendreher unterlaufen.

Literaturhinweis: SCHULZ [15] §8.

9. Literaturauswahl

Aus der Vielzahl der Bücher und Artikel, die die angesprochenen Themen behandeln, geben wir im folgenden eine (subjektive) Auswahl an.

a) *Fachwissenschaftliche Darstellungen*

[1] BAUER, F.L. u. G. GOOS: Informatik, 1. Teil. Springer Verlag, Berlin, etc. 1971.

[2] BEUTELSPACHER,A.: Kryptologie. Vieweg Verlag, Braunschweig 1987, 1991^2.

[3] BLAHUT, R.E.: Theory and Practice of Error Control Codes. Addison- Wesley Publ. Comp., Reading etc. 1983.

[4] DWORATSCHEK, S.: Einführung in die Datenverarbeitung. De Gruyter Verlag, Berlin 1970^3.

[5] FURRER, F.J.: Fehlerkorrigierende Block-Codierung für die Datenübertragung. Birkhäuser Verlag, Basel etc., 1981.

[6] HEISE, W. & P. QUATTROCCHI: Informations- und Codierungstheorie. Springer Verlag, Berlin etc. 1983.

[7] HENZE, E. & H.H. HOMUTH: Einführung in die Informationstheorie. Vieweg Verlag, Braunschweig 1974.

[8] HENZE, E. & H.H. HOMUTH: Einführung in die Codierungstheorie. Vieweg Verlag, Braunschweig 1974

[9] HOFMANN,K.-D.: Einführung in die Informationstheorie, Pädag. Verlag Schwann 1973.

[10] KAMEDA, T. & K. WEIHRAUCH: Einführung in die Codierungstheorie I. BI, Zürich 1973

[11] MACWILLIAMS, F.J. and N.J.A. SLOANE: The theory of error correcting codes. North Holland, Amsterdam etc. 1977.

[12] MASSEY, J.L.: Was ist ein Bit Information. Frequenz 37 (1983)5, 110-115.

[13] NEIDHARDT, P.: Informationstheorie und automatische Informationsverarbeitung. Berliner Union, Stuttgart 1964[2].

[14] RUPPRECHT, W.: Nachrichtenübertragung; Band II von K.STEINBUCH und W. RUPPRECHT: Nachrichtentechnik. Springer Verlag, Berlin etc. 1983[3].

[15] SCHULZ, R.-H.: Codierungstheorie. Eine Einführung. Vieweg Verlag, Braunschweig / Wiesbaden 1991.

[16] SEYDEL, R. & R. BULIRSCH: Vom Regenbogen zum Farbfernsehen, Springer Verlag, Berlin etc. 1986.

[17] TOPSØE, F.: Informationstheorie. Teubner Verlag, Stuttgart 1974.

b) *Schulnähere Veröffentlichungen*

[18] DIFF (Deutsches Institut für Fernstudien) Band CM1, Algorithmen der elementaren Zahlentheorie (C. Niederdrenk-Felgner), Tübingen 1988.

[19] OBERSCHELP, W.: Algorithmen und Computer im Unterricht, Kurseinheit 5. Fernuniv. GSH Hagen 1986.

[20] SCHMIDT, W.: Mathematikaufgaben, Anwendungen aus der modernen Technik und Arbeitswelt. Klett Verlag, Stuttgart 1984.

[21] SCHULZ, R.-H.: Wörterinterpretationen an Beispielen einfacher Codes. DdM 2 (1984) 113-131.

[22] SCHULZ, R.-H.: Übersetzen von Nachrichten für die digitale Übertragung. Ausgewählte Aspekte der Quellencodierung. MU 33/3 (1987) 23-44.

[23] WEFELSCHEID, H.: Einführung in die Informationstheorie. MU 2O/3 (1974) 5- 35.

c) *Zu Beispiel* 1.5:

[24] WILEY, E.O.: Phylogenetics. John Wiley, New York 1981.