1. Einführung: Codes und Kanäle

In diesem Kapitel wird dargestellt, welche Bedeutung und welchen Platz die Kanalcodierung in einem digitalen Kommunikationssystem einnimmt. Daneben werden wichtige Grundbegriffe wie das Prinzip der Blockcodierung, die Maximum-Likelihood-Decodierung und der asymptotische Codierungsgewinn eingeführt.

1.1 Was ist Kanalcodierung ?

Als Begründer der Informations- und Codierungstheorie gilt Claude E. Shannon mit den beiden berühmten Arbeiten *A Mathematical Theory of Communication* und *Communication Theory of Secrecy Systems*, die 1948 und 1949 im Bell Systems Technical Journal veröffentlicht wurden (Nachdruck in [65, 66]). Die Shannon'sche Theorie bildet auch heute noch die Grundlage für das Verständnis der digitalen Übertragung mit dem Ziel einer sicheren, zuverlässigen und effizienten Kommunikation.

Sowohl die Datenquellen wie die Übertragungskanäle werden mit stochastischen Modellen beschrieben. Mit einem mathematischen Informationsmaß (Entropie) wird jeder Nachricht ein Informationsgehalt zugeordnet. Damit kann die minimale Anzahl von Symbolen bestimmt werden, die zur fehlerfreien Darstellung einer Nachricht unbedingt erforderlich sind. Eine längere Nachricht mit dem gleichen Informationsgehalt weist dann eine Redundanz auf. Codierung bedeutet im allgemeinen Fall die Zuordnung von Nachrichten zu einer Menge oder Folge von Symbolen. In der Shannon'schen Theorie werden drei Arten von Codierung unterschieden:

Quellencodierung (source coding): Die Nachrichten werden so komprimiert, daß zwar keine Informationen verloren gehen und somit eine perfekte Wiedergewinnung der Nachrichten möglich ist, aber dafür wird die Anzahl der zu übertragenden Symbole reduziert. Durch Quellencodierung wird also überflüssige Redundanz eliminiert und das Übertragungssystem entlastet.

Kanalcodierung (error control coding): Diese stellt Methoden und Verfahren zur Verfügung, mit denen Informationen von einer Quelle zur Sinke mit einem Minimum an Fehlern übertragen werden können. Den eigentlichen Informationen wird sendeseitig kontrolliert Redundanz hinzugefügt, so daß

bei der Übertragung entstandene Fehler empfangsseitig erkannt und korrigiert werden können. Damit läßt sich eine extrem hohe Zuverlässigkeit der übertragenen Daten erreichen. Ferner sind Störungen kompensierbar, die durch andere Maßnahmen, wie beispielsweise durch eine Erhöhung der Sendeleistung, prinzipiell nicht zu verhindern wären.

Kryptographie: Darunter wird die Codierung zur Verschlüsselung verstanden, um Nachrichten für Unberechtigte unlesbar zu machen bzw. um zu verhindern, daß Nachrichten gefälscht oder vorgetäuscht werden können. Während durch Kanalcodierung Nachrichten auch im Fall von Störungen lesbar bleiben, sollen die verschlüsselten Nachrichten auch bei ungestörter Übertragung ohne Kenntnis des Schlüssels unlesbar sein.

In diesem Buch wird nur die Kanalcodierung behandelt, und zwar in enger Verbindung mit der digitalen Übertragungstechnik und modernen Modulationsverfahren, sowie unter Berücksichtigung der durch die Shannon'sche Informationstheorie aufgezeigten prinzipiellen Grenzen und Möglichkeiten einer zuverlässigen Informationsübertragung.

Seit 1948 hat sich die Kanalcodierung zu einer ausgedehnten anwendungsorientierten Wissenschaft entwickelt, die ihre wesentlichen Impulse der wechselseitigen Beeinflussung theoretisch und praktisch orientierter Arbeiten verdankt. Viele moderne digitale Systeme zur Nachrichtenübertragung erreichen ihre enorme Leistungsfähigkeit überhaupt nur durch den Einsatz von Kanalcodierung, dies gilt insbesondere bei stark gestörten Kanälen oder bei hohen Anforderungen an die Zuverlässigkeit. Eine spezielle Form der Übertragung (*von hier nach dort*) ist die Nachrichtenspeicherung (*von jetzt nach später*), da beim Ein- und Auslesen ebenso wie bei der Übertragung mit Störungen gerechnet werden muß, die den Einsatz von Fehlerschutzverfahren erfordern.

Neben der praktischen Bedeutung stellt die Kanalcodierung auch ein nachrichtentechnisch wie mathematisch hochinteressantes Gebiet dar, weil die Codes, die Algorithmen und die informationstheoretischen Grenzen auf einer anspruchsvollen, vielschichtigen und eleganten Theorie basieren. Über die Grundlagen der Übertragungstechnik und der Modulationsverfahren hinaus sind als Stichworte zu nennen: Wahrscheinlichkeitstheorie und Stochastik; Matrizen und lineare Algebra; endliche Körper und Polynomringe; Fouriertransformationen; spezielle Metriken; die Analyse, Synthese und Umformung von Schieberegistern; Zustandsautomaten; Trellisdiagramme; sowie effiziente Algorithmen und Strukturen.

Shannon konnte in der genannten Arbeit aus dem Jahr 1948 zeigen, daß jeder Übertragungskanal durch eine quantitative Größe namens *Kanalkapazität* beschrieben wird, so daß durch die Kanalcodierung eine beliebig kleine Fehlerrate erreicht werden kann, sofern die zu übertragende Datenrate kleiner als die Kanalkapazität ist und eine hinreichend aufwendige Verarbeitung in Sender und Empfänger möglich ist. Die Kanaleigenschaften begrenzen also nicht die Qualität der Übertragung, sondern nur den Durchsatz.

In der Shannon'schen Theorie wird die Existenz entsprechend leis-

tungsfähiger Codes jedoch nur theoretisch nachgewiesen, eine praktische Konstruktionsvorschrift fällt dabei nicht ab. In den Jahren nach 1948 gelang es nicht, die theoretisch vorhergesagten Codes auch tatsächlich praktisch zu finden – allerdings waren zur damaligen Zeit die entsprechenden Verfahren technisch ohnehin noch nicht realisierbar. Nach einer Phase der Ernüchterung über den Wert der theoretischen Erkenntnisse konzentrierte sich die weitere Entwicklung deshalb zunächst darauf, Codes zu finden, mit denen zwar nicht die informationstheoretischen Grenzen erreicht werden können, die aber dafür tatsächlich eine vernünftige Realisierung erlauben. Im Vordergrund stehen dabei die Verbesserungen durch die Codierung gegenüber der uncodierten Übertragung, wobei sich der Vergleich üblicherweise auf die Reduktion in der Fehlerrate oder auf die Einsparung an Sendeleistung bezieht.

Inzwischen wurde eine kaum noch überschaubare Menge von speziellen Codes gefunden und analysiert. Von überragender praktischer Bedeutung sind aber dennoch nur einige wenige Codeklassen, nämlich die RS- und BCH-Blockcodes sowie einige relativ einfache Faltungscodes, die deshalb hier auch vorrangig berücksichtigt werden. Mit dem Prinzip der Codeverkettung und leistungsfähigen Algorithmen zur Decodierung ist der Abstand praktisch realisierbarer Verfahren zu den informationstheoretischen Grenzen inzwischen aber ziemlich klein geworden.

Es gibt zwei prinzipiell unterschiedliche Codeklassen, nämlich Blockcodes (BC, Kapitel 3 bis 7) und Faltungscodes (FC, Kapitel 8 und 9), deren theoretische wie praktische Behandlung sich als ziemlich unterschiedlich herausstellen wird. Wie in fast allen anderen Büchern stehen auch hier die Blockcodes am Anfang, da sich einige grundsätzliche Fragen sowie die Shannon'sche Informationstheorie (Kapitel 2) damit einfacher erläutern lassen. Dazu reichen schon geringe Kenntnisse über Blockcodes (entsprechend Kapitel 1) aus, während die hochentwickelten RS- und BCH-Blockcodes eine sehr komplexe mathematische Struktur aufweisen. Die Einbettung der Kanalcodierung in ein Übertragungssystem wird mit den anschließend behandelten Faltungscodes deutlich.

Für Block- und Faltungscodes in Verbindung mit einfachen Modulationsverfahren wird hier der Begriff *klassische Kanalcodierung* verwendet. Die bei der Codierung hinzugefügte Redundanz erhöht die Datenrate, so daß der Übertragungkanal häufiger benutzt werden muß bzw. mehr Bandbreite benötigt wird. Auf diese Weise lassen sich nur leistungs- aber nicht bandbreiteneffiziente Übertragungsverfahren erzielen. Mit den bahnbrechenden Arbeiten von G.Ungerböck, die seit 1982 unter dem Begriff der trelliscodierten Modulation (TCM, Kapitel 10) bekannt sind, kann eine leistungseffiziente Übertragung auch ohne Expansion der Bandbreite erreicht werden. Durch TCM kann gleichzeitig die Fehlerrate und die notwendige Sendeleistung sowie im Extremfall sogar die notwendige Bandbreite reduziert werden. Die TCM-Verfahren basieren primär auf Faltungscodes, teilweise aber auch auf Blockcodes.

Ferner ist zwischen zwei grundsätzlichen Prinzipien bei der Kanalcodierung zu unterscheiden, die davon abhängen, ob eine Information sehr schnell übertra-

gen werden muß und ob ein Rückkanal verfügbar ist:

FEC-Verfahren (Forward Error Correction): Die bei der Kanalcodierung sendeseitig hinzugefügte Redundanz dient empfangsseitig zur *Korrektur* der Übertragungsfehler. Als Fehlerkorrekturcodes (error correction code) werden Blockcodes und Faltungscodes sowie die trelliscodierte Modulation verwendet. Allerdings wird sich später noch zeigen, daß man Faltungscodes und TCM-Verfahren eigentlich so nicht bezeichnen sollte, sondern besser als Übertragungscodes.

ARQ-Verfahren (Automatic Repeat Request): Hierbei werden die Übertragungsfehler nicht korrigiert, sondern es erfolgt empfangsseitig eine Beschränkung auf die *Erkennung* von Fehlern (error detection code). Dabei muß allerdings vorausgesetzt werden, daß die Übertragungszeit nicht extrem knapp vorgegeben wird und daß ein Rückkanal verfügbar ist. Bei erkannten Fehlern wird nämlich eine Wiederholung über diesen Rückkanal angefordert, um die Nachricht erneut zu senden oder um sie mit zusätzlicher Redundanz zu versehen. Zur Fehlererkennung werden fast ausschließlich Blockcodes verwendet.

Der Vorteil von ARQ liegt darin, daß zur Fehlererkennung weit weniger Redundanz als zur Fehlerkorrektur übertragen werden muß. Wenn allerdings Nachrichten wiederholt zu übertragen sind, kann es zu erheblichen Verzögerungen kommen. Der Durchsatz bei ARQ ist abhängig von der Kanalqualität, während die Fehlerrate davon unabhängig ist. Umgekehrt sind die Verhältnisse bei FEC: Die Kanalqualität bestimmt die Fehlerrate, aber nicht den Durchsatz. FEC- und ARQ-Verfahren können auch kombiniert werden, indem beispielsweise die Redundanz so dimensioniert wird, daß eine kleine Anzahl von Fehlern noch korrigierbar ist, aber bei vielen Fehlern eine Wiederholung angefordert wird. Aus Platzgründen werden in diesem Buch aber ausschließlich FEC-Verfahren behandelt.

Der Entwurf von leistungsfähigen Codierungsverfahren muß sich immer an den speziellen Randbedingungen des Übertragungssystems und inbesondere an den Eigenschaften des Übertragungskanals orientieren. Spezielle Anwendungen erfordern also spezielle Codes. Zu den wichtigsten Randbedingungen, die bei der Auswahl und Optimierung eines Übertragungssystems mit Kanalcodierung zu beachten sind, zählen die Eigenschaften des Übertragungskanals, insbesondere die verfügbare Bandbreite; die verfügbare Sendeleistung; das vorgegebene Modulationsverfahren, sofern nicht die Kanalcodierung und das Modulationsverfahren gemeinsam optimiert werden können; die Begrenzungen in der Verzögerungzeit bei der Übertragung; die Begrenzungen in der Komplexität der Signalverarbeitung in Sender und Empfänger; die Anforderungen an die Fehlerwahrscheinlichkeit und an die zulässigen Fehlerstrukturen nach der Decodierung; die Anforderungen an die Synchronisation; sowie Anforderungen sonstiger Art, wenn beispielsweise bei nichtlinearen Sendeverstärkern Modulationsverfahren mit konstanter Enveloppe erforderlich sind.

Daraus resultiert eine ziemlich komplexe Aufgabenstellung mit vielfältigen Lösungen je nach Gewichtung der einzelnen Randbedingungen. Die folgende Auflistung vermittelt einen Überblick zu den vielfältigen Anwendungen und Aufgaben der Kanalcodierung:

- Zur Leistungsersparnis, beispielsweise bei geostationären Kommunikationssatelliten („himmlischer" Kanal) und insbesondere bei erdfernen Forschungssatelliten (siehe Abschnitt 12.1). Hierbei liegt der Idealfall eines AWGN-Kanals vor (siehe Abschnitt 1.3), der zu statistisch unabhängigen Einzelfehlern führt.
- Zur Bandbreitenersparnis, indem die durch die Codierung erzeugte Redundanz nicht in zusätzlich zu übertragende Symbole mündet, sondern in höherstufige Codesymbole mit gleicher oder sogar reduzierter Datenrate. Als wichtige Anwendung sind beispielsweise die Telefonkanal-Modems zu nennen (siehe Abschnitt 12.2). Der digitale Mobilfunk (siehe Abschnitte 12.3, 12.4) ist ein typisches Beispiel für eine gleichermaßen leistungs- wie bandbreiteneffiziente Anwendung.
- Bei hohen Anforderungen an die Zuverlässigkeit wie beispielsweise im Rechnerverbund, bei der Kommunikation im Bankenbereich, bei sicherheitsrelevanten Diensten (z.B. Fernsteuerung im Bahnverkehr) sowie bei hochkomprimierten Daten. Hierbei sind oft Kombinationen mit kryptographischen Codes erforderlich.
- Bei zeitvariablen Störungen, wenn ein Codewort gute und schlechte Übertragungsabschnitte aufweist und der Empfänger die momentane Signalqualität gut schätzen kann. Insbesondere bei Mobilfunkkanälen treten Fadingeinbrüche auf, die kurz gegenüber der Codewortlänge und lang gegenüber der Symboldauer sind.
- Bei Kanälen mit Bündelfehlern, die typischerweise bei der Nachrichtenspeicherung auftreten, wie beispielsweise bei der Compact Disc (siehe Abschnitt 12.6).
- In Kombination mit der Quellencodierung, wenn die einzelnen Quellensymbole unterschiedlich wichtig sind und auch unterschiedlich gut geschützt werden sollen – aktuelle Beispiele sind die Sprachcodierung im GSM-Mobilfunk (siehe Abschnitte 8.3, 12.3, 12.4) und die Quellencodierung für den digitalen Hörrundfunk.
- Zur Erkennung von Fehlern anstelle der Korrektur. Im Zusammenhang mit der Quellencodierung von Sprache und Musik zählen hierzu auch Maßnahmen zur Fehlerverschleierung, so daß erkannte Fehler nicht zu subjektiven Auswirkungen führen (siehe Abschnitte 12.3, 12.4, 12.6).
- In Kombination mit ARQ-Verfahren wie vorangehend erläutert.
- Bei Störungen, die durch erhöhte Sendeleistung nicht zu unterdrücken sind wie beispielsweise Übersprechen, Knackgeräusche, Fading, Reflexionen, Mehrwegeausbreitung und Verzerrungen.

- Zur Verringerung der sogenannten Hintergrund-Fehlerrate, die bei nicht perfekten Sendern und Empfängern auftritt, beispielsweise aufgrund von Nichtlinearitäten (siehe Abschnitt 12.5).

1.2 Codierung in der Nachrichtenübertragung

Das Grundprinzip einer digitalen Nachrichtenübertragung mit Quellencodierung und Kanalcodierung zeigt Bild 1.1. Wie vorangehend dargestellt wird durch die Quellencodierung erst Redundanz eliminiert und durch die Kanalcodierung wird dann kontrolliert Redundanz hinzugefügt. Die in den Quellendaten eventuell vorhandene Redundanz ist für die Kanalcodierung unbrauchbar, da die Eigenschaften dieser Redundanz nicht genau kontrollierbar und steuerbar sind. Eventuell weisen die Quellendaten auch gar keine Redundanz auf.

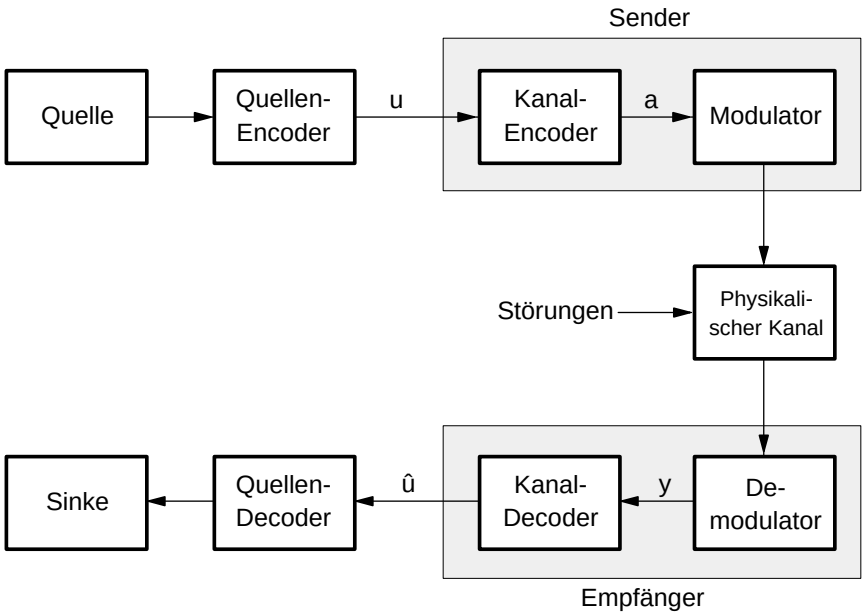

Bild 1.1. Digitale Nachrichtenübertragung mit Kanal- und Quellencodierung

Nach der Shannon'schen Theorie [39, 66] können wie in Bild 1.1 dargestellt die Quellencodierung und die Kanalcodierung getrennt ausgeführt und optimiert werden, was als *Separationsprinzip* bezeichnet wird. Allerdings kann es praktische Erfordernisse wie beispielsweise Beschränkungen in der Übertragungsverzögerung oder in der Komplexität geben, so daß ganz im Gegenteil Quellencodierung und Kanalcodierung aufeinander abgestimmt werden müssen (siehe dazu Abschnitt 12.3 und 12.4).

Jeder Encoder-Operation entspricht eine Decoder-Operation. Die Bezeichnung *Coder* wird nicht verwendet. Der in Bild 1.1 nicht dargestellte kryptographische Code ist normalerweise zwischen Quellencodierung und Kanalcodierung angeordnet. Nachfolgend wird die Quellencodierung und die Kryp-

tographie nicht mehr betrachtet, so daß die Kanalcodierung auch kurz als Codierung bezeichnet werden kann.

Die Daten u werden direkt als Quellendaten angesprochen mit der Bezeichnung *Infobits* im Binärfall bzw. *Infosymbole* im mehrstufigen Fall. Der *Encoder* überführt die Infosymbole bzw. Infobits u in die *Codesymbole* bzw. *Codebits* a. Dabei wird Redundanz hinzugefügt, so daß die Datenrate durch den Encoder erhöht wird.

Mit dem *Modulator* wird der Encoder an den *physikalischen Kanal* (waveform channel, continuous channel, transmission channel) angeschlossen. Der physikalische Kanal kann keine diskreten Symbole übertragen, sondern nur zeitkontinuierliche Signale. Somit ist es die Aufgabe des Modulators, den diskreten Werten a derartige Signale zuzuordnen, die über den physikalischen Kanal übertragbar sind. Darin enthalten ist die Anpassung des modulierten Signals an den Übertragungsbereich bzw. an das Spektrum des physikalischen Kanals, insbesondere also die Verschiebung des Basisbandsignals in die Bandpaßlage. Bei den in Kapitel 10 behandelten Verfahren der trelliscodierten Modulation ist die Aufteilung des Senders in einen Encoder und einen Modulator allerdings nicht mehr eindeutig und in der Form aus Bild 1.1 auch nicht sinnvoll.

Der *physikalische Kanal* ist prinzipiell nicht ideal, d.h. er verändert die Signale bei der Übertragung. Das gilt sowohl bei drahtgebundener Übertragung (z.B. Teilnehmeranschlußleitung, Koaxialkabel, Glasfaserkabel), bei terrestrischen Funkkanälen (z.B. Mobilfunk, Richtfunk, Rundfunk, Kurzwellenfunk), bei Satellitenstrecken, bei Speicherung (z.B. Magnetmedien, elektronische und optische Speicher) wie natürlich auch bei Kombinationen dieser Kanäle. Der physikalische Kanal ist beispielsweise charakterisiert durch nichtideale Amplituden- und Phasengänge, durch Verzerrungen, durch Störungen aufgrund von Rauschen oder Übersprechen oder aufgrund atmosphärischer Effekte oder verschiedener Ausbreitungswege sowie durch absichtliche Störungen.

Am *Demodulator* liegen also nicht exakt die zeitkontinuierlichen Sendesignale an, sondern nur eine gestörte und verfälschte Version davon. Dennoch sollte der Empfänger die zeitdiskreten Sendewerte bzw. die Infosymbole möglichst genau rekonstruieren. Dazu wird zunächst im Demodulator aus dem Bandpaßsignal das Basisbandsignal zurückgewonnen, was bei kohärenten Empfängern eine ideale Träger- und Phasensynchronisation einschließt. Daraus wird eine zeitdiskrete Wertefolge hergestellt, so daß jedem Codesymbol a ein *Empfangswert* y entspricht. Bei der sogenannten *Soft-Decision Demodulation* soll der Demodulator derartige Werte y herstellen, die für den Decoder möglichst viel Information enthalten – es muß dann nicht zwangsläufig das Ziel sein, daß y möglichst genau a entspricht.

Der *Decoder* arbeitet zeitdiskret: Aus der im Takt der Codesymbole anliegenden Folge der Empfangswerte y wird eine Schätzung $\hat{u}$ für die Infosymbole u abgeleitet, wobei diese Schätzung i.a. eine zeitliche Verzögerung aufweist. Im idealen Fall arbeitet der Decoder sogar sequenzweise: Erst nach dem Empfang

einer ganzen Sequenz von Empfangswerten wird auf einen Schlag die ganze Sequenz der Infosymbole geschätzt.

Für den Modulator sind die Codesymbole a nur Sendewerte ohne Kenntnis der Codierung. Um dies in besonderen Fällen hervorzuheben, werden in Analogie zu den Ausgangswerten y des Demodulators die Eingangswerte des Modulators auch mit x statt a bezeichnet.

1.3 Der Begriff des diskreten Kanals

Gemäß Bild 1.2 ist der *diskrete Kanal* (DC, discrete channel; auch digital channel, coding channel) die Zusammenfassung von Modulator, physikalischem Kanal und Demodulator. Der vom Modulationssystem erzeugte (zeit-)diskrete Kanal ist also in einem sehr allgemeinen Sinn zu verstehen, enthalten darin sind u.a. eventuell sehr komplexe Modulations- und Synchronisationsverfahren. In diesem Abschnitt werden idealisierte diskrete Kanäle formal beschrieben, während in Kapitel 12 einige sehr komplizierte Kanäle betrachtet werden, die bei verschiedenen Anwendungen entstehen.

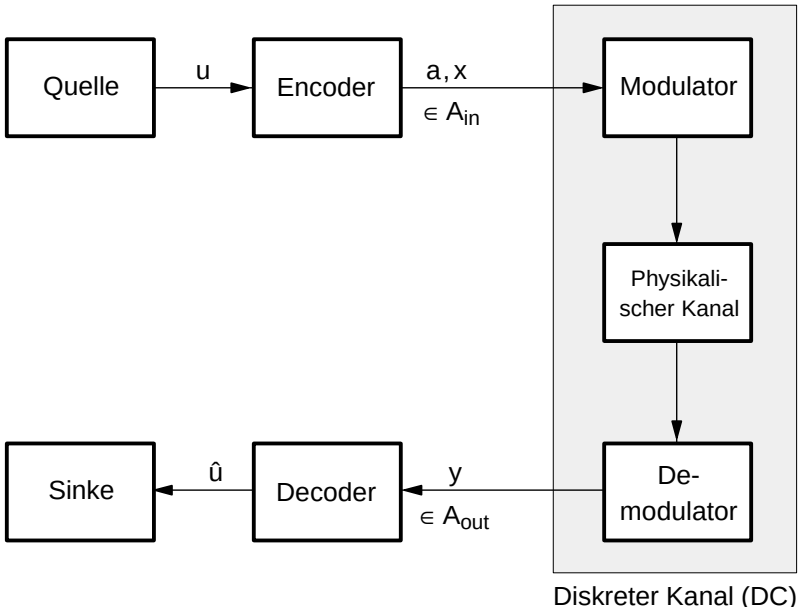

Bild 1.2. Erzeugung des diskreten Kanals durch das Modulationssystem

In der formalen Beschreibung wird ein diskreter Kanal charakterisiert durch das Tripel $(\mathcal{A}_{\text{in}}, \mathcal{A}_{\text{out}}, P_{y|x})$. Dabei bedeuten:

$\mathcal{A}_{\text{in}} =$ *Eingangsalphabet* mit q Werten. Dies ist der Wertebereich für die Infosymbole u sowie für die Codesymbole a sowie für die geschätzten Infosymbole $\hat{u}$, d.h. $u, a, \hat{u}$ sind jeweils q-stufig. Fallunterscheidungen:

Der einfachste Fall ist $q = 2$ für Binärcodes, wobei die Symbole lediglich Bits sind. Der allgemeine Fall ist $q = p^m$ mit p als Primzahl und m als

natürlicher Zahl. Der Normalfall für die meisten Codes ist $q = 2^m$, wobei den Symbole jetzt Bitgruppen (z.B. Bytes bei $m = 8$) entsprechen.

$\mathcal{A}_{\text{out}} =$ *Ausgangsalphabet*: Dies ist der Wertebereich für die Empfangswerte y. Fallunterscheidungen für den Demodulator:

Bei *Hard-Decision* gilt $\mathcal{A}_{\text{out}} = \mathcal{A}_{\text{in}}$, d.h. der Demodulator schätzt direkt die gesendeten Werte a bzw. x. Diese Situation liegt bei einfachen Blockcodes vor. Im binären Fall gilt dann $\mathcal{A}_{\text{in}} = \mathcal{A}_{\text{out}} = \{0, 1\}$.

Bei *Soft-Decision* umfaßt $\mathcal{A}_{\text{out}}$ mehr Werte als $\mathcal{A}_{\text{in}}$ – im Extremfall gilt sogar $\mathcal{A}_{\text{out}} = \mathbb{R}$ für einen wertkontinuierlichen Demodulatorausgang. In diesem Fall kann der Demodulator besonders viel Information über den Kanal (Zustand, Qualität) vermitteln. Beispielsweise teilt der Demodulator dem Decoder mit, mit welcher Sicherheit er seine Entscheidungen getroffen hat (sehr sicher oder gerade an der Grenze). Zwar kann prinzipiell jedes Codierungsverfahren diese Information ausnutzen, praktikabel ist das jedoch meistens nur bei Faltungscodes. Ein typischer Fall bei $\mathcal{A}_{\text{in}} = \{0, 1\}$ ist ein 8-stufiges $\mathcal{A}_{\text{out}}$, d.h. der Empfangswert wird mit 3-Bit quantisiert (siehe dazu auch Bild 1.4 und 1.5).

$P_{y|x} =$ *Übergangswahrscheinlichkeit* (Kanalstatistik; conditional, transition probability): Dabei ist $P_{y|x}(\eta|\xi)$ die bedingte Wahrscheinlichkeit dafür, daß $y = \eta$ empfangen wurde unter der Voraussetzung, daß $x = \xi$ gesendet wurde.

Input x und Output y des Kanals werden hier also als Zufallsgrößen angenommen, deren Werte mit $\xi \in \mathcal{A}_{\text{in}}$ und $\eta \in \mathcal{A}_{\text{out}}$ bezeichnet werden. Vereinfachend wird auch $P(y|x)$ geschrieben, wenn es auf die Unterscheidung zwischen den Zufallsgrößen und ihren Werten nicht ankommt. Für die Übergangswahrscheinlichkeit gilt generell:

$$\sum_{\eta \in \mathcal{A}_{\text{out}}} P_{y|x}(\eta|\xi) = 1 \quad \text{für alle } \xi \in \mathcal{A}_{\text{in}}. \tag{1.3.1}$$

Für den diskreten Kanal sind einige wichtige Fallunterscheidungen zu vermerken: Neben einer Hard-Decision oder Soft-Decision Demodulation können die Übertragungseigenschaften *zeitinvariant* oder auch *zeitvariant* sein. Ferner kann der diskrete Kanal ein *Gedächtnis* haben (d.h. der Empfangswert ist nicht nur vom zuletzt gesendeten Wert abhängig, sondern auch von den vorangehend gesendeten Werten) oder er ist *gedächtnislos* (d.h. der Empfangswert ist nur vom aktuell gesendeten Wert abhängig).

Definition 1.1 (DMC). *Als* diskreter gedächtnisloser Kanal *(DMC, Discrete Memoryless Channel) wird ein diskreter Kanal mit endlichen Alphabeten* $\mathcal{A}_{\text{in}}$ *und* $\mathcal{A}_{\text{out}}$ *bezeichnet, der zudem gedächtnislos und zeitinvariant sein soll. Die Gedächtnislosigkeit ist dadurch gekennzeichnet, daß die Übergangswahrscheinlichkeit für Sequenzen in ein Produkt der Übergangswahrscheinlichkeiten für*

Einzelsymbole übergeht:

$$P(y_0,\ldots,y_{n-1}|x_0,\ldots,x_{n-1}) = \prod_{i=0}^{n-1} P(y_i|x_i). \tag{1.3.2}$$

Wenn die Übergangswahrscheinlichkeiten bei Hard-Decision gewisse Symmetrien erfüllen, reicht zur Charakterisierung des DMC ein einziger Parameter aus:

Definition 1.2 (Symmetrischer Hard-Decision DMC). *Als ein* q-närer symmetrischer Kanal mit Hard-Decision *wird ein DMC mit* $\mathcal{A}_{\text{in}} = \mathcal{A}_{\text{out}}$ *und der Übergangswahrscheinlichkeit*

$$P(y|x) = \left\{ \begin{array}{ll} 1-p_e & y = x \\ p_e/(q-1) & y \neq x \end{array} \right\} \tag{1.3.3}$$

bezeichnet. Dieser Kanal ist eindeutig durch die Angabe der Symbol-Fehlerwahrscheinlichkeit p_e *bestimmt. Der binäre Fall mit* $\mathcal{A}_{\text{in}} = \mathcal{A}_{\text{out}} = \{0,1\}$ *wird als* binärer symmetrischer Kanal *(BSC, Binary Symmetric Channel) bezeichnet.*

Für $p_e = 0$ ist der Kanal fehlerfrei und für $p_e = 1/2$ bei $q = 2$ wird in Kapitel 2 noch gezeigt, daß eine zuverlässige Übertragung prinzipiell unmöglich ist. Bei einem veränderlichen p_e würde ein zeitvarianter Kanal vorliegen. Diese Situation wird noch in Abschnitt 9.7 behandelt. Ausgeschrieben lautet (1.3.3) für den BSC:

$$\begin{array}{rclcl} P_{y|x}(0|0) & = & P_{y|x}(1|1) & = & 1-p_e \\ P_{y|x}(1|0) & = & P_{y|x}(0|1) & = & p_e\,. \end{array} \tag{1.3.4}$$

Mit der Wahrscheinlichkeit p_e wird das Bit bei der Übertragung verfälscht und mit der Wahrscheinlichkeit $1-p_e$ ist die Übertragung korrekt:

$$\begin{array}{rcl} P(y = x) & = & 1-p_e \\ P(y \neq x) & = & p_e\,. \end{array} \tag{1.3.5}$$

Beispiel: Unter der Voraussetzung, daß 110 gesendet wurde, wird 101 mit der Wahrscheinlichkeit $P_{y|x}(101|110) = P_{y|x}(1|1)P_{y|x}(0|1)P_{y|x}(1|0) = (1-p_e)\cdot p_e \cdot p_e$ empfangen.

Das Prinzip des BSC zeigt Bild 1.3, wobei die Kanten von $x \in \mathcal{A}_{\text{in}}$ nach $y \in \mathcal{A}_{\text{out}}$ mit den Übergangswahrscheinlichkeiten $P(y|x)$ beschriftet sind.

Für den q-nären symmetrischen Hard-Decision DMC gelten einige wichtige allgemeine Formeln:

(1) Mit P_{ee} (ee = error event) wird die Wahrscheinlichkeit bezeichnet, daß bei der Übertragung einer Sequenz $\boldsymbol{x} = (x_0, x_1, \ldots, x_{n-1})$ der Länge n mindestens ein Fehler auftritt:

$$P_{ee} \;=\; P(\boldsymbol{y} \neq \boldsymbol{x})$$

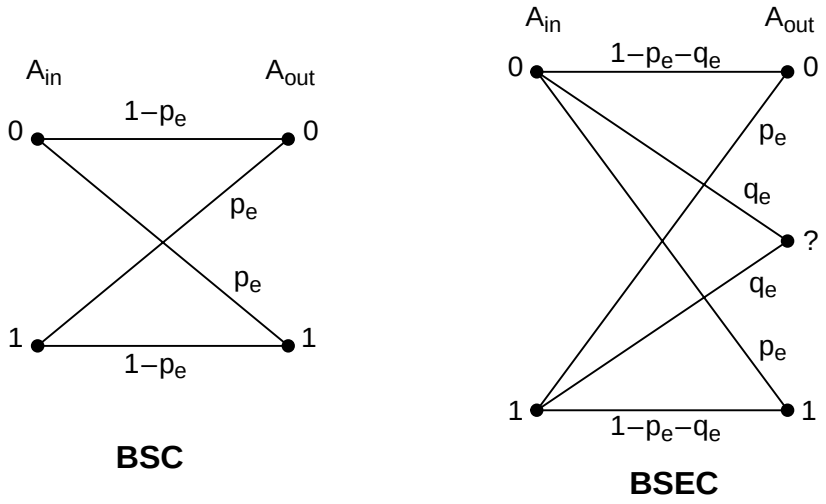

Bild 1.3. Modelle diskreter gedächtnisloser Kanäle (BSC, BSEC)

$$
\begin{aligned}
&= 1 - P(\boldsymbol{y} = \boldsymbol{x}) \\
&= 1 - P(y_0 = x_0, \ldots, y_{n-1} = x_{n-1}) \\
&= 1 - P(y_0 = x_0) \cdots P(y_{n-1} = x_{n-1}) \\
&= 1 - (1 - p_e)^n \qquad (1.3.6) \\
&\approx np_e \quad \text{bei } np_e \ll 1. \qquad (1.3.7)
\end{aligned}
$$

(1.3.7) folgt aus der Binomialentwicklung $(1 - p_e)^n = \sum_{i=0}^{n} \binom{n}{i} (-p_e)^i$.

(2) Die Wahrscheinlichkeit dafür, daß eine Sequenz von n Bits in eine andere bestimmte Sequenz verfälscht wird, wobei r Fehler auftreten, beträgt:

$$P(\text{von } n \text{ Bits sind } r \text{ bestimmte Bits falsch}) = p_e^r (1 - p_e)^{n-r}. \qquad (1.3.8)$$

(3) Die Wahrscheinlichkeit für r Fehler in einer Sequenz von n Bits beträgt nach der Binomialverteilung (siehe Anhang A.3):

$$P(\text{von } n \text{ Bits sind } r \text{ Bits falsch}) = \binom{n}{r} p_e^r (1 - p_e)^{n-r}. \qquad (1.3.9)$$

Eine Verallgemeinerung des BSC ist der in Bild 1.3 ebenfalls dargestellte *binäre symmetrische Kanal mit Ausfällen* (BSEC, Binary Symmetric Erasure Channel), bei dem der Output ternär ist: $\mathcal{A}_{\text{out}} = \{0, ?, 1\}$. Hierbei entscheidet der Demodulator auf den „Wert“ ?, wenn die Entscheidung auf 0 oder 1 sehr unsicher wäre. Für den Decoder ist es besser, über den Sendewert gar keine Information zu haben als eine Information, die in der Hälfte aller Fälle falsch ist. Der BSEC ist der einfachste Fall eines diskreten Kanals mit Soft-Decision mit

$$P(y|x) = \left\{ \begin{array}{ll} 1 - p_e - q_e & \text{für } y = x \\ q_e & \text{für } y = ? \\ p_e & \text{sonst} \end{array} \right\}. \qquad (1.3.10)$$

Natürlich gilt hierbei $P_{y|x}(0|x) + P_{y|x}(?|x) + P_{y|x}(1|x) = 1$ für $x \in \mathcal{A}_{\text{in}} = \{0, 1\}$. Für $q_e = 0$ wird der BSEC zum BSC und für $p_e = 0$ wird der BSEC zum reinen *Auslöschungskanal* (BEC, Binary Erasure Channel). Ein weiteres sehr wichtiges DMC-Modell ist:

Definition 1.3. *Als* AWGN-Kanal *(Additive White Gaussian Noise) wird ein Kanal mit binärem Input bezeichnet, bei dem weißes normalverteiltes (Gaußsches) Rauschen ν additiv überlagert wird:*

$$y = x + \nu.$$

Dabei sind x und ν statistisch unabhängig. Mit E_c wird die Energie pro Codebit und mit N_0 wird die einseitige Rauschleistungsdichte bezeichnet. Für die Alphabete gilt $\mathcal{A}_{\text{in}} = \{-\sqrt{E_c}, +\sqrt{E_c}\}$ und $\mathcal{A}_{\text{out}} = \mathbb{R}$ und die Übergangswahrscheinlichkeiten haben die Form von Verteilungsdichten:

$$f_{y|x}(\eta|\xi) = \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{(\eta - \xi)^2}{N_0}\right). \tag{1.3.11}$$

Also ist y bei gegebenem x normalverteilt mit dem Erwartungswert $x = \xi$ und der Varianz $\sigma^2 = N_0/2$, die der Varianz des Rauschens entspricht. Wenn der AWGN mit binärer Modulation (ASK, Amplitude Shift Keying) betrieben wird und im Demodulator binär quantisiert wird, so ergibt sich wieder ein BSC mit der Bit-Fehlerwahrscheinlichkeit

$$\begin{aligned} p_e &= P_{y|x}(y < 0 \mid x = +\sqrt{E_c}) \;=\; P_{y|x}(y > 0 \mid x = -\sqrt{E_c}) \\ &= \int_0^{+\infty} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{(\eta + \sqrt{E_c})^2}{N_0}\right) d\eta \\ &= Q\left(\sqrt{\frac{2E_c}{N_0}}\right). \end{aligned} \tag{1.3.12}$$

Dabei ist

$$Q(\alpha) = \frac{1}{\sqrt{2\pi}} \int_\alpha^\infty e^{-\eta^2/2}\, d\eta = \frac{1}{2}\text{erfc}\left(\frac{\alpha}{\sqrt{2}}\right) \tag{1.3.13}$$

$$= P(\nu > \alpha\sqrt{N_0/2}) \tag{1.3.14}$$

die *komplementäre Gaußsche Fehlerfunktion* (siehe Anhang A.3). Den numerischen Zusammenhang zwischen p_e und E_c/N_0 zeigt Tabelle 1.1 und der graphische Verlauf ist in den Bildern mit Fehlerwahrscheinlichkeits-Kurven dargestellt (siehe z.B. Bild 1.10, Kurve „uncodiert“, $E_c = E_b$).

Wenn beim AWGN im Demodulator nicht binär mit 1 Bit sondern oktal mit 3 Bit quantisiert wird, so ergibt sich ein DMC mit $\mathcal{A}_{\text{in}} = \{-\sqrt{E_c}, +\sqrt{E_c}\}$ und oktalem $\mathcal{A}_{\text{out}} = \{-1, -1', -1'', -1''', +1''', +1'', +1', +1\}$. Von einiger Bedeutung

Tabelle 1.1. BSC-Fehlerwahrscheinlichkeit

p_e	E_c/N_0 [dB]	p_e	E_c/N_0 [dB]
10^{-1}	$-0{,}86$	10^{-11}	13,52
10^{-2}	4,33	10^{-12}	13,93
10^{-3}	6,79	10^{-13}	14,31
10^{-4}	8,40	10^{-14}	14,66
10^{-5}	9,59	10^{-15}	14,99
10^{-6}	10,53	10^{-16}	15,29
10^{-7}	11,31	10^{-17}	15,57
10^{-8}	11,97	10^{-18}	15,84
10^{-9}	12,55	10^{-19}	16,09
10^{-10}	13,06	10^{-20}	16,32

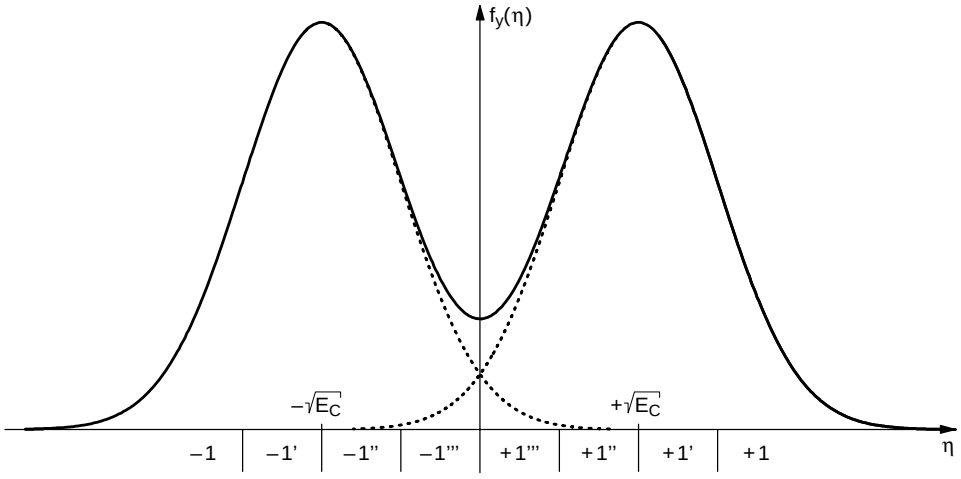

Bild 1.4. Oktale Quantisierung der AWGN-Empfangswerte

ist dabei die Wahl der 7 Sprungstellen in der *Quantisierungskennlinie*, was in [49] genauer analysiert wird.

In Bild 1.4 wird eine Kennlinie mit äquidistanten Sprungstellen angenommen, die genau auf die Sendewerte $-\sqrt{E_c}, +\sqrt{E_c}$ ausgerichtet ist, was im Demodulator natürlich eine Pegelregelung erfordert. Die Verteilungsdichtefunktion der Empfangswerte $f_y(\eta) = \frac{1}{2}\left(f_{y|x}(\eta| - \sqrt{E_c}) + f_{y|x}(\eta| + \sqrt{E_c})\right)$ ergibt sich durch Überlagerung von zwei Normalverteilungen, wobei für die Darstellung in Bild 1.4 $E_c/N_0 = 3$ dB angenommen wurde. Dabei sind die Übergangswahrscheinlichkeiten von $-\sqrt{E_c}$ nach $+1$ oder $+1'$ nahezu Null. In Bild 1.5 sind die Übergangswahrscheinlichkeiten des oktalen Kanals dagegen für $E_c/N_0 = -3$ dB angegeben. Die Werte in Bild 1.5 werden beispielsweise wie folgt berechnet:

$$\begin{aligned} P(-1'''|-\sqrt{E_c}) &= P(-0{,}5\sqrt{E_c} < y < 0 | x = -\sqrt{E_c}) \\ &= P(0{,}5\sqrt{E_c} < \nu < \sqrt{E_c}) \\ &= Q(0{,}5\sqrt{2E_c/N_0}) - Q(\sqrt{2E_c/N_0}) \end{aligned}$$

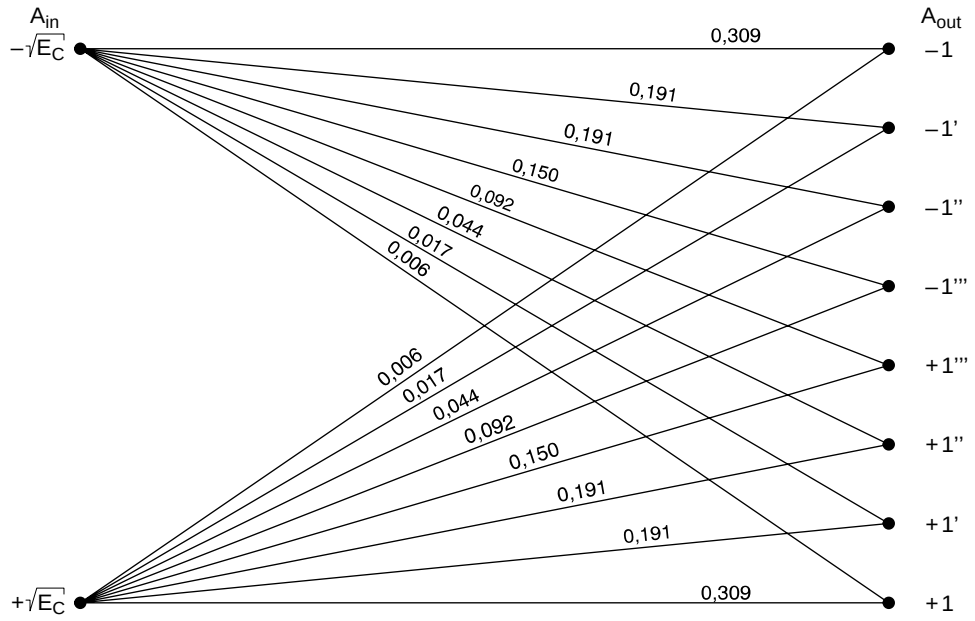

Bild 1.5. Übergangswahrscheinlichkeiten beim oktal quantisierten AWGN

$$= \quad Q(0{,}5) - Q(1) \;=\; 0{,}3085 - 0{,}1587 \;=\; 0{,}1498.$$

Für 2-dimensionale Modulationsverfahren bzw. Signalkonstellationen wird Definition 1.3 zum *2-dimensionalen AWGN* erweitert. Bei jeder Kanalbenutzung werden zwei Werte gesendet und zwei Werte empfangen, die in komplexer Schreibweise zu $x = x_I + jx_Q$ zusammengefaßt werden, wobei I für Inphase und Q für Quadraturphase steht. Für das überlagerte Rauschen gilt entsprechend $\nu = \nu_I + j\nu_Q$. Die Rauschenergie in jeder Komponente ist weiterhin $N_0/2$ und beide Komponenten sind statistisch unabhängig. Die Rauschenergie des 2-dimensionalen Rauschens ergibt sich über den Betrag einer komplexen Zahl:

$$E(|\nu|^2) \;=\; E(\nu_I^2 + \nu_Q^2) \;=\; \frac{N_0}{2} + \frac{N_0}{2} \;=\; N_0. \tag{1.3.15}$$

Die Übergangswahrscheinlichkeit entspricht der in Bild 1.6 dargestellten 2-dimensionalen Normalverteilung:

$$f_{y|x}(\eta\,|\xi) \;=\; \frac{1}{\pi N_0} \exp\left(-\frac{(\eta_I - \xi_I)^2 + (\eta_Q - \xi_Q)^2}{N_0}\right). \tag{1.3.16}$$

Bisher wurden nur zeitinvariante und gedächtnislose Kanäle betrachtet, bei denen statistisch unabhängige Einzelfehler auftreten. Die weiteren Kapitel führen noch verschiedene andere Kanaltypen ein: In den Abschnitten 5.6 bis 5.8, in Kapitel 7 und in Abschnitt 9.6 werden Kanäle betrachtet, bei denen Fehler in Bündeln auftreten. In Abschnitt 9.7 werden spezielle zeitvariante Kanäle diskutiert, die bei der Decodierung von Faltungscodes entstehen sowie Super-Kanäle

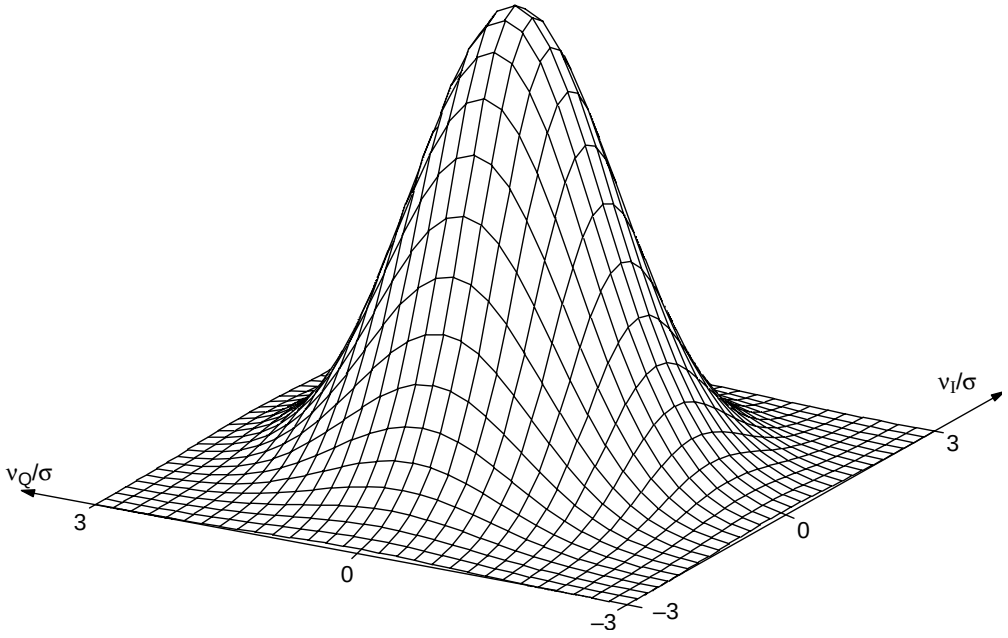

Bild 1.6. Verteilungsdichtefunktion der 2-dim. Normalverteilung ($\sigma = \sqrt{N_0/2}$)

bei verketteten Codes. Die für moderne Anwendungen sehr wichtigen Kanäle mit Fading sowie mit Verzerrungen werden in Kapitel 11 diskutiert. Weitere Kanaltypen werden in Kapitel 12 eingeführt.

1.4 Grundprinzip der Blockcodierung

Das Grundprinzip der Blockcodierung zeigt Bild 1.7: Der Datenstrom der Infosymbole bzw. Codesymbole wird unterteilt in Blöcke der Länge k bzw. n, die als *Infowörter* $\boldsymbol{u} = (u_0, \ldots, u_{k-1})$ bzw. *Codewörter* $\boldsymbol{a} = (a_0, \ldots, a_{n-1})$ bezeichnet werden. Dabei gilt $k < n$. Der Encoder ordnet jedem Infowort ein Codewort zu. Am Ausgang des diskreten Kanals entsteht das *Empfangswort* $\boldsymbol{y} = (y_0, \ldots, y_{n-1})$, aus dem der Decoder die Schätzung $\hat{\boldsymbol{u}} = (\hat{u}_0, \ldots, \hat{u}_{k-1})$ für das Infowort gewinnt.

Die Zuordnung der Codewörter zu den Infowörtern im Encoder ist (1) *eindeutig* und *umkehrbar*, indem zwei verschiedenen Infowörtern zwei verschiedene Codewörter zugeordnet werden, so daß zu jedem Codewort genau ein Infowort gehört; (2) *zeitinvariant*, indem die Zuordnungsvorschrift immer gleich bleibt; (3) *gedächtnislos*, indem jedes Infowort nur auf ein Codewort wirkt und jedes Codewort nur durch ein Infowort bestimmt wird.

Ohne die Forderung der Gedächtnislosigkeit würde sich ein Faltungscode ergeben (siehe Definition 8.1), bei dem neben dem aktuellen Infowort auch die vorangehenden Infowörter auf das Codewort einwirken. So gesehen sind Blockcodes also spezielle Faltungscodes.

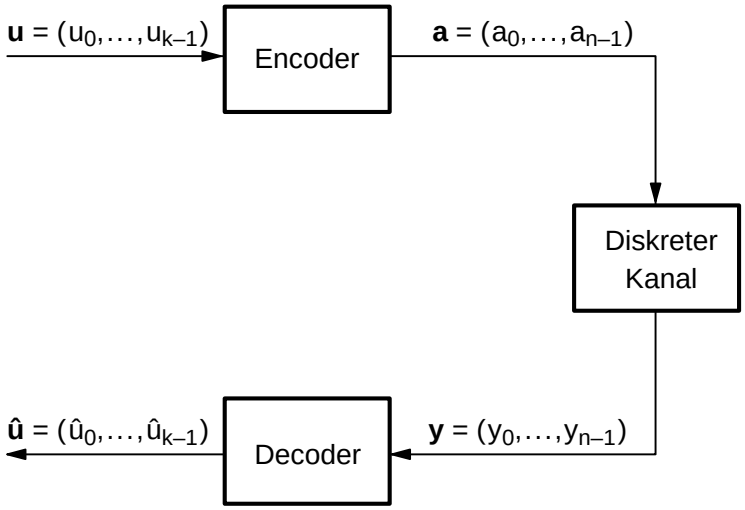

Bild 1.7. Prinzip des (n,k)-Blockcodes

Wegen der Gedächtnislosigkeit und Zeitinvarianz braucht die Folge der Infowörter bzw. Codewörter nicht durchnumeriert zu werden, da immer nur die Übertragung eines Wortes betrachtet wird.

Definition 1.4. *Durch die vorangehend beschriebene Methode wird ein* (n,k)-Blockcode *definiert, der in ausführlicherer Schreibweise auch als* $(n,k,d_{\min})_q$-*Blockcode bezeichnet wird, wobei* q *die Stufenzahl der Symbole* $u_i, a_i, \hat{u}_i$ *und* $d_{\min}$ *die Minimaldistanz (siehe Definition 1.8) bezeichnet. Als* Blocklänge *wird* n *bezeichnet und als* Coderate *wird das Verhältnis von* k *zu* n *bezeichnet:*

$$R = \frac{k}{n} < 1 \qquad \textit{Einheit: Infosymbol/Kanalbenutzung.} \tag{1.4.1}$$

Als Code Γ *wird die Menge aller Codewörter bezeichnet.*

Die Coderate $R \leq 1$ ist mit der eigentlich dimensionslosen Einheit Infosymbol/Codesymbol versehen. Da pro Codesymbol der diskrete Kanal genau einmal benutzt wird, ergibt sich die in (1.4.1) angegebene Einheit.

Die Datenraten werden immer auf Bit statt Symbol bezogen, d.h. die *Infobitrate* r_b hat die Einheit Infobit/s und die *Codebitrate* r_c hat die Einheit Codebit/s. Bei q-stufigen Symbolen entspricht ein Symbol genau $\log_2 q$ Bits, so daß $r_b/\log_2 q$ die Infosymbolrate und $r_c/\log_2 q$ die Codesymbolrate bzw. die Kanalbenutzungsrate ist. Pro Sekunde werden also $r_c/(n \cdot \log_2 q)$ Codeblöcke übertragen. Die Codierung bewirkt wegen

$$r_c \;=\; r_b \cdot \frac{1}{R} \;=\; r_b \cdot \frac{n}{k} \tag{1.4.2}$$

eine Erhöhung der Datenrate um den Faktor $1/R = n/k$, der deshalb zuweilen auch als *Bandbreitenexpansionsfaktor* bezeichnet wird. $R = 1$ bedeutet uncodierte Übertragung. Manchmal ist es erforderlich, die Coderate auf Bits statt

auf Symbole zu beziehen:

$$R_b \;=\; R \cdot \log_2 q \;=\; \frac{k}{n} \cdot \log_2 q \qquad \text{Einheit: Infobit/Kanalben.} \qquad (1.4.3)$$

Im binären Fall gilt natürlich $R_b = R$.

Die Anzahl der Infowörter der Länge k mit q-stufigen Symbolen beträgt offensichtlich q^k und somit gibt es auch $q^k = |\Gamma| = q^{nR} = 2^{nR_b}$ Codewörter. Die Anzahl der möglichen Sendewörter der Länge n beträgt jedoch q^n. Der Code Γ ist also eine Untermenge der Mächtigkeit q^k in der Menge aller q^n Wörter.

Der Begriff des Blockcodes wurde in Definition 1.4 etwas eingeschränkt festgelegt. In der allgemeinsten Form darf die Mächtigkeit $|\Gamma|$ eine beliebige ganze Zahl sein, d.h. nicht unbedingt eine q-Potenz. In diesem Fall resultiert dann $R_b = \frac{1}{n}\log_2 |\Gamma|$ und $k = nR = \frac{\log_2 |\Gamma|}{\log_2 q}$ ist nicht mehr notwendigerweise ganzzahlig. Durch diese Verallgemeinerung ergeben sich aber keine theoretischen Vorteile und in der Praxis wird immer $|\Gamma| = q^k$ mit ganzzahligem k gewählt.

Die *Güte eines Codes* wird ausschließlich dadurch geprägt, wie geschickt aus den q^n Wörtern die q^k Codewörter ausgewählt werden. Es wird darauf hinauslaufen, daß sich die Codewörter möglichst stark voneinander unterscheiden müssen.

Der Encoder trifft nur eine Zuordnung zwischen den q^k Infowörtern und den q^k Codewörtern. Wie diese Zuordnung über die Forderungen der Eindeutigkeit, Zeitinvarianz und Gedächtnislosigkeit hinaus organisiert ist, bleibt im Prinzip weitgehend belanglos. Insofern ist der Begriff *Güte eines Encoders* sinnlos. Allerdings werden in der Praxis fast ausschließlich systematische Encoder verwendet (siehe Definition 1.5) und zur Vereinfachung der Realisierung fast ausschließlich lineare Codes (siehe Kapitel 3) bzw. zyklische Codes (siehe Kapitel 5).

Definition 1.5. *Bei einem* systematischen Encoder *(auch: systematischer Code) erfolgt die Zuordnung zwischen Infowörtern und Codewörtern derart, daß das Infowort explizit Teil des Codewortes ist. Die restlichen $n-k$ Stellen heißen dann* Prüfstellen *(Prüfbits, parity bits).*

Beispiel 1.1. Die beiden Zuordnungen (Prüfstellen hinten bzw. vorn)

$00 \mapsto 000$	$00 \mapsto 000$
$01 \mapsto 011$	$01 \mapsto 101$
$10 \mapsto 101$	$10 \mapsto 110$
$11 \mapsto 110$	$11 \mapsto 011$

erzeugen den gleichen $(3,2)_2$ Code $\Gamma = \{000, 011, 101, 110\}$. Der Codemenge kann nicht angesehen werden, in welcher Weise encodiert wurde. Beide Encoder sind als gleichwertig anzusehen. □

Definition 1.6. *Zwei Blockcodes heißen* identisch, *wenn ihre Codemengen identisch sind. Zwei Blockcodes heißen* äquivalent, *wenn nach einer geeigneten Vertauschungsvorschrift für die Komponenten der Codewörter die Codemengen identisch sind.*

1.5 Hammingdistanz und Minimaldistanz

Es seien $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$ jeweils Wörter der Länge n mit q-stufigen Werten (beispielsweise Codewörter oder Empfangswörter).

Definition 1.7. *Die* Hammingdistanz $d_H(\boldsymbol{x}, \boldsymbol{y})$ *ist definiert als die Anzahl der Abweichungen zwischen den Komponenten von* $\boldsymbol{x}$ *und* $\boldsymbol{y}$. *Sofern eine Null definiert ist, wird als* Hamminggewicht $w_H(\boldsymbol{x})$ *die Anzahl der Komponenten von* $\boldsymbol{x}$ *bezeichnet, die ungleich Null sind.*

Für den Zusammenhang zwischen Abstand und Gewicht gilt:

$$w_H(\boldsymbol{x}) = d_H(\boldsymbol{x}, \boldsymbol{0}) \quad \text{mit} \quad \boldsymbol{0} = (0, \ldots, 0). \tag{1.5.1}$$

Falls im Wertebereich der Symbole eine „Subtraktion“ definiert ist, gilt weiter:

$$d_H(\boldsymbol{x}, \boldsymbol{y}) = w_H(\boldsymbol{x} - \boldsymbol{y}). \tag{1.5.2}$$

Satz 1.1. *Die Hammingdistanz ist eine* Metrik *im mathematischen Sinn, d.h. es gelten folgende Eigenschaften:*

$$d_H(\boldsymbol{x}, \boldsymbol{y}) = d_H(\boldsymbol{y}, \boldsymbol{x}) \tag{1.5.3}$$

$$0 \le d_H(\boldsymbol{x}, \boldsymbol{y}) \le n \tag{1.5.4}$$

$$d_H(\boldsymbol{x}, \boldsymbol{y}) = 0 \quad \Longleftrightarrow \quad \boldsymbol{x} = \boldsymbol{y} \tag{1.5.5}$$

$$d_H(\boldsymbol{x}, \boldsymbol{y}) \le d_H(\boldsymbol{x}, \boldsymbol{z}) + d_H(\boldsymbol{z}, \boldsymbol{y}). \tag{1.5.6}$$

Die Beziehung (1.5.6) wird als Dreiecksungleichung *bezeichnet und ist in Bild 1.8 veranschaulicht. Falls Addition und Subtraktion im Wertebereich der Symbole definiert sind, so gelten für das Hamminggewicht folgende Eigenschaften:*

$$w_H(\boldsymbol{x}) = w_H(-\boldsymbol{x}) \tag{1.5.7}$$

$$0 \le w_H(\boldsymbol{x}) \le n \tag{1.5.8}$$

$$w_H(\boldsymbol{x}) = 0 \quad \Longleftrightarrow \quad \boldsymbol{x} = \boldsymbol{0} \tag{1.5.9}$$

$$w_H(\boldsymbol{x} + \boldsymbol{y}) \le w_H(\boldsymbol{x}) + w_H(\boldsymbol{y}). \tag{1.5.10}$$

Definition 1.8. *Die* Minimaldistanz $d_{\min}$ *eines* $(n, k, d_{\min})$*-Blockcodes* Γ *ist definiert als die minimale Hammingdistanz zwischen allen Codewörtern:*

$$d_{\min} = \min\{d_H(\boldsymbol{a}, \boldsymbol{b}) \mid \boldsymbol{a}, \boldsymbol{b} \in \Gamma \ , \ \boldsymbol{a} \neq \boldsymbol{b}\}. \tag{1.5.11}$$

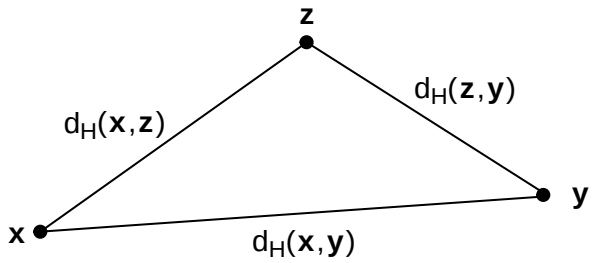

Bild 1.8. Veranschaulichung der Dreiecksungleichung für die Hammingdistanz

Die Minimaldistanz ist der wichtigste Parameter, der die Güte eines Codes bestimmt. Die vollständige Charakterisierung eines Codes wird später durch die Gewichtsverteilung gegeben (siehe Definition 3.7). Zur Bestimmung der Minimaldistanz müssen die Abstände aller Codewörterpaare betrachtet werden. Je größer die Minimaldistanz ist, je stärker sich also die Codewörter voneinander unterscheiden, desto besser ist der Code. Bei gegebener Coderate $R = k/n$ und gegebener Blocklänge n sollte derjenige Code gewählt werden, der zu großem $d_{\min}$ führt. Normalerweise wird $d_{\min}$ größer (d.h. besserer Code), wenn die Coderate kleiner wird (d.h. mehr Bandbreite erforderlich) oder wenn die Blocklänge größer wird (d.h. komplexere Verarbeitung erforderlich).

Beispiel 1.2. Der $(7,4)_2$-Hamming-Code besteht aus 16 Codewörtern der Länge 7:

$$\begin{aligned} \Gamma \;=\; \{\, & 0000000, \quad 1000011, \\ & 0001111, \quad 1001100, \\ & 0010110, \quad 1010101, \\ & 0011001, \quad 1011010, \\ & 0100101, \quad 1100110, \\ & 0101010, \quad 1101001, \\ & 0110011, \quad 1110000, \\ & 0111100, \quad 1111111 \,\}. \end{aligned}$$

Der Code ist hier durch eine Aufzählung der Codewörter gegeben und nicht durch die (unwesentliche) Art der Encodiervorschrift (systematisch, Infobits vorn). Der Vergleich der ersten beiden Codewörter ergibt $d_{\min} \leq 3$. Bei der Betrachtung aller Paare findet sich kein Paar mit der Hammingdistanz 2, so daß $d_{\min} = 3$ folgt. □

Klar ist schon jetzt, daß es besserer Methoden zur Beschreibung der Codemenge und zur Berechnung der Minimaldistanz bedarf – dazu wird später eine algebraische Struktur auf der Codemenge eingeführt.

1.6 Maximum-Likelihood-Decodierung

Die optimale Decodiervorschrift ist dadurch definiert, daß die Wort-Fehlerwahrscheinlichkeit $P_w = P(\hat{\boldsymbol{u}} \neq \boldsymbol{u})$ nach dem Decoder minimal wird:

Unterstellt wird also ein stochastischer Kanal (beispielsweise ein DMC), der zu Fehlern im Empfangswort führt, die möglicherweise durch den Decoder nicht korrigiert werden können und damit zu Fehlern im geschätzten Infowort führen. Derartige Fehler sollen während einer Übertragung von vielen Worten möglichst selten auftreten. Nicht berücksichtigt wird dabei, ob in einem falsch geschätzten Infowort nur ein Fehler oder mehrere Fehler enthalten sind. Eine solche Minimierung der Bit-Fehlerwahrscheinlichkeit $P_b = P(\hat{u}_i \neq u_i)$ ist wesentlich schwieriger.

Ziel ist also, daß das geschätzte Infowort möglichst oft exakt mit dem Infowort auf der Sendeseite übereinstimmt. Diese Forderung ist das Kriterium, nach dem der Decoder konstruiert werden soll. Es wird sich gleich zeigen, daß man diese Konstruktionsvorschrift für den Decoder ableiten kann, auch wenn das Kriterium P_w gar nicht explizit berechnet werden kann:

$$\begin{aligned} P_w &= P(\hat{\boldsymbol{u}} \neq \boldsymbol{u}) \quad \longrightarrow \text{ Minimum} \\ &= P(\hat{\boldsymbol{a}} \neq \boldsymbol{a}). \end{aligned} \tag{1.6.1}$$

Wegen der eindeutigen Zuordnung zwischen Infowörtern und Codewörtern kann anstelle des Infowortes auch das Codewort geschätzt werden. Die Schätzung für das Infowort ist genau dann korrekt, wenn die Schätzung für das Codewort korrekt ist. Deshalb kann die Wort-Fehlerwahrscheinlichkeit anstelle der Infowörter auch über die Codewörter definiert werden.

Bild 1.9 zeigt das Prinzip zur Herleitung der Decodiervorschrift. Der Decoder wird wie angegeben zerlegt in einen Codewortschätzer (hier mit δ bezeichnet) und ein Encoder-Inverses. Dieses Encoder-Inverse ist eine direkte Umkehrung des

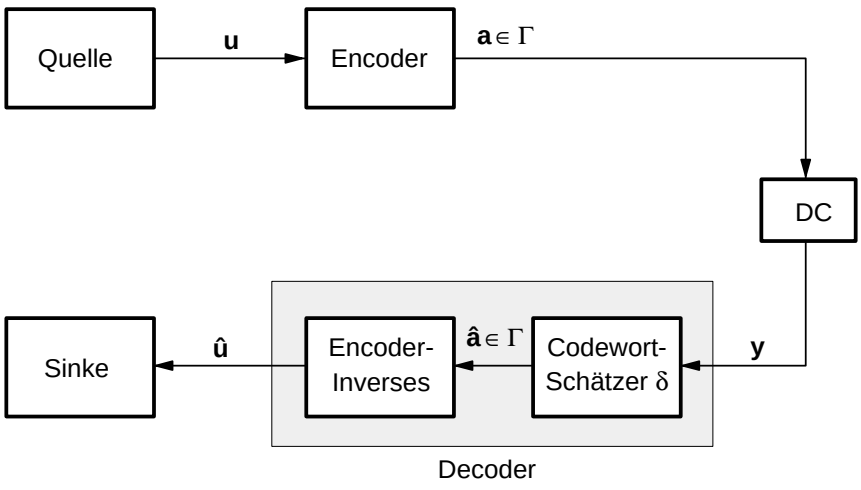

Bild 1.9. Zur Herleitung der Decodiervorschrift

Encoders und hat zu den geschätzten Codewörtern $\hat{\boldsymbol{a}}$ lediglich das zugehörige Infowort $\hat{\boldsymbol{u}}$ zu bestimmen. Das ist eine triviale Operation, beispielsweise sind bei systematischen Encodern lediglich die Prüfstellen auszublenden.

Die gesamte Intelligenz des Decoders steckt im Codewortschätzer, der im Gegensatz zum Encoder-Inversen nicht die Encodiervorschrift, sondern nur die Codemenge kennen muß. Zum Empfangswort $\boldsymbol{y}$ wird also im Codewortschätzer das geschätzte Codewort bestimmt – formal ist das eine Abbildung wie folgt:

$$\delta \;:\; \boldsymbol{y} \mapsto \delta(\boldsymbol{y}) = \hat{\boldsymbol{a}} \in \Gamma. \tag{1.6.2}$$

Die Funktion δ ist so zu konstruieren, daß die Wort-Fehlerwahrscheinlichkeit minimal wird:

$$P_w = P(\delta(\boldsymbol{y}) \neq \boldsymbol{a}) \tag{1.6.3}$$

Bei der Übertragung über einen diskreten Kanal mit Hard-Decision (also mit $\mathcal{A}_{\text{in}} = \mathcal{A}_{\text{out}}$) sind folgende Fälle zu unterscheiden:

$\boldsymbol{y} = \boldsymbol{a}$ Fehlerfreie Übertragung.

$\boldsymbol{y} \in \Gamma \setminus \{\boldsymbol{a}\}$ Verfälschung in ein anderes Codewort – dieser Fall kann niemals erkannt oder korrigiert werden.

$\boldsymbol{y} \notin \Gamma$ Die Verfälschung ist generell erkennbar und eventuell korrigierbar durch den Decoder. Bei $\delta(\boldsymbol{y}) = \boldsymbol{a}$ wird korrekt decodiert und bei $\delta(\boldsymbol{y}) \neq \boldsymbol{a}$ wird falsch decodiert. Der Fall $\delta(\boldsymbol{y}) =$ undefiniert tritt zwar beim idealen Decoder nicht auf, aber bei praktisch realisierten Decodern ist dieser Fall jedoch durchaus sinnvoll (siehe nachfolgende Erklärung).

In der formalen Beschreibung ordnet δ jedem der q^n möglichen Empfangsworte eines der q^k Codewörter zu. Später wird sich noch zeigen, daß man bei der Realisierung des Decoders teilweise darauf verzichtet, jedem möglichen Empfangswort die optimale Schätzung zuzuordnen – stattdessen wird die optimale Decodiervorschrift nur für die häufiger vorkommenden Empfangswörter realisiert. Mit dieser Methode können sich für den Decoder ganz erhebliche Vereinfachungen bei der Realisierung ergeben.

Wenn ein Empfangswort vollständig vorliegt, so kann (abgesehen von verarbeitungstechnischen Verzögerungen) sofort blockweise eine Schätzung des Codewortes bzw. Infowortes erfolgen. Im Normalfall kann also eine Schätzung für das erste Infosymbol empfangsseitig frühestens dann erfolgen, wenn das zuletzt gesendete Codesymbol empfangen wurde. Somit bestimmt die Blocklänge n eine untere Grenze für die prinzipiell mindestens auftretende Verzögerungszeit bei der codierten Übertragung.

Voraussetzung (gleiche Apriori-Wahrscheinlichkeiten) *zur Herleitung der Decodiervorschrift: Alle q^k Infowörter sollen mit der gleichen Wahrscheinlichkeit q^{-k} von der Quelle abgegeben werden.*

Mit dieser Voraussetzung treten auch alle Codewörter mit der gleichen Wahrscheinlichkeit q^{-k} auf. Die Wahrscheinlichkeit, daß ein Fehler bei der Decodierung auftritt unter der Voraussetzung, daß das Codewort $\boldsymbol{a}$ gesendet wurde, ergibt sich aus der Summation über diejenigen Empfangswörter, die zu einer Schätzung ungleich $\boldsymbol{a}$ führen:

$$\begin{aligned} P(\delta(\boldsymbol{y}) \neq \boldsymbol{a} \mid \boldsymbol{a} \text{ gesendet}) &= \sum_{\substack{\boldsymbol{y} \\ \delta(\boldsymbol{y}) \neq \boldsymbol{a}}} P(\boldsymbol{y} \text{ empfangen} \mid \boldsymbol{a} \text{ gesendet}) \\ &= \sum_{\substack{\boldsymbol{y} \\ \delta(\boldsymbol{y}) \neq \boldsymbol{a}}} P_{y|x}(\boldsymbol{y}|\boldsymbol{a}). \end{aligned} \tag{1.6.4}$$

Ferner gilt bei der Summation über alle Codewörter und alle Empfangswörter:

$$\begin{aligned} \sum_{\boldsymbol{a} \in \Gamma, \boldsymbol{y}} P_{y|x}(\boldsymbol{y}|\boldsymbol{a}) &= \sum_{\boldsymbol{a} \in \Gamma} P_{y|x}(\text{irgend ein } \boldsymbol{y} \text{ empfangen} \mid \boldsymbol{a}) \\ &= \sum_{\boldsymbol{a} \in \Gamma} 1 \;=\; q^k. \end{aligned} \tag{1.6.5}$$

Aus dem Satz von der vollständigen Wahrscheinlichkeit (A.3.1) folgt nun:

$$\begin{aligned} P_w &= P(\delta(\boldsymbol{y}) \neq \boldsymbol{a}) \\ &= \sum_{\boldsymbol{a} \in \Gamma} P(\delta(\boldsymbol{y}) \neq \boldsymbol{a} \mid \boldsymbol{a} \text{ gesendet}) \cdot P(\boldsymbol{a} \text{ gesendet}) \\ &= \sum_{\boldsymbol{a} \in \Gamma} \sum_{\substack{\boldsymbol{y} \\ \delta(\boldsymbol{y}) \neq \boldsymbol{a}}} P_{y|x}(\boldsymbol{y}|\boldsymbol{a}) \cdot q^{-k} \qquad \text{mit (1.6.4)} \\ &= q^{-k} \left(\sum_{\boldsymbol{a} \in \Gamma, \boldsymbol{y}} P_{y|x}(\boldsymbol{y}|\boldsymbol{a}) - \sum_{\substack{\boldsymbol{a} \in \Gamma, \boldsymbol{y} \\ \delta(\boldsymbol{y}) = \boldsymbol{a}}} P_{y|x}(\boldsymbol{y}|\boldsymbol{a}) \right) \\ &= 1 - q^{-k} \cdot \sum_{\substack{\boldsymbol{a} \in \Gamma, \boldsymbol{y} \\ \delta(\boldsymbol{y}) = \boldsymbol{a}}} P_{y|x}(\boldsymbol{y}|\boldsymbol{a}) \qquad \text{mit (1.6.5)} \\ &= 1 - q^{-k} \cdot \sum_{\boldsymbol{y}} P_{y|x}(\boldsymbol{y}|\delta(\boldsymbol{y})). \end{aligned}$$

Zur Minimierung von P_w sollte für jedes Empfangswort $\boldsymbol{y}$ also $\delta(\boldsymbol{y}) = \hat{\boldsymbol{a}}$ so gewählt werden, daß die Übergangswahrscheinlichkeit $P_{y|x}(\boldsymbol{y}|\hat{\boldsymbol{a}})$ maximal wird.

Satz 1.2 (Maximum-Likelihood-Decoder MLD). *Die Wort-Fehlerwahrscheinlichkeit P_w nach der Decodierung wird minimal, wenn wie folgt decodiert wird: Zum Empfangswort $\boldsymbol{y}$ wird als Schätzung $\hat{\boldsymbol{a}} \in \Gamma$ dasjenige Codewort gewählt, bei dem die Übergangswahrscheinlichkeit maximal wird:*

$$P_{y|x}(\boldsymbol{y}|\hat{\boldsymbol{a}}) \geq P_{y|x}(\boldsymbol{y}|\boldsymbol{b}) \quad \textit{für alle } \boldsymbol{b} \in \Gamma. \tag{1.6.6}$$

Die ML-Decodierung kann auch mehrdeutig sein – in diesem Fall wird dann irgendein Codewort mit maximaler Übergangswahrscheinlichkeit gewählt.

Dieses Ergebnis ist auch anschaulich einfach zu verstehen: Bei gegebenem Empfangswort wird dasjenige Sendewort bzw. Codewort gesucht, das am wahrscheinlichsten gesendet wurde. Die Berechnung der Wort-Fehlerwahrscheinlichkeit selbst erfolgt später in Kapitel 3.

Ohne die Voraussetzung gleicher Apriori-Wahrscheinlichkeiten ergibt sich allerdings ein anderer Decoder, nämlich der *Maximum Aposteriori Decoder* (MAP), der genau auf die Quellenstatistik angepaßt ist und bei nicht gleichwahrscheinlichen Sendeworten zu einer kleineren Wort-Fehlerwahrscheinlichkeit P_w führt (siehe Aufgabe 1.11). Wegen der Abhängigkeit von der Quellenstatistik wird der MAP-Decoder jedoch nur in Sonderfällen angewendet.

Noch anschaulicher werden diese Ergebnisse, wenn der q-näre symmetrische DMC betrachtet wird. Aus (1.3.2) und (1.3.3) folgt dann für $\boldsymbol{y} = (y_0, \ldots, y_{n-1})$ und $\hat{\boldsymbol{a}} = (\hat{a}_0, \ldots, \hat{a}_{n-1})$ mit $d = d_H(\boldsymbol{y}, \hat{\boldsymbol{a}})$:

$$\begin{aligned} P_{y|x}(\boldsymbol{y}|\hat{\boldsymbol{a}}) &= \prod_{i=0}^{n-1} P_{y|x}(y_i|\hat{a}_i) \\ &= \prod_{i=0}^{n-1} \left\{ \begin{array}{ll} 1-p_e & y_i = \hat{a}_i \\ p_e/(q-1) & y_i \neq \hat{a}_i \end{array} \right\} \\ &= (1-p_e)^{n-d} \cdot \left(\frac{p_e}{q-1} \right)^d \\ &= (1-p_e)^n \cdot \left(\frac{p_e}{(1-p_e)(q-1)} \right)^d . \end{aligned} \tag{1.6.7}$$

Der linke Faktor ist unabhängig von $\hat{\boldsymbol{a}}$ und somit muß nur der rechte Faktor durch geeignete Wahl von $\hat{\boldsymbol{a}}$ maximiert werden. Für $p_e < 0{,}5$ ist der Quotient kleiner als 1 und somit ergibt sich das Maximum, wenn $d = d_H(\boldsymbol{y}, \hat{\boldsymbol{a}})$ minimal wird:

Satz 1.3 (MLD für Hard-Decision-DMC). *Die Wort-Fehlerwahrscheinlichkeit P_w nach der Decodierung wird minimal, wenn wie folgt decodiert wird: Zum Empfangswort $\boldsymbol{y}$ wird als Schätzung $\hat{\boldsymbol{a}} \in \Gamma$ dasjenige Codewort gewählt, das vom Empfangswort den minimalen Hammingabstand hat:*

$$d_H(\boldsymbol{y}, \hat{\boldsymbol{a}}) \leq d_H(\boldsymbol{y}, \boldsymbol{b}) \quad \textit{für alle } \boldsymbol{b} \in \Gamma. \tag{1.6.8}$$

Beispiel 1.3. Es wird der $(5,2)_2$-Code $\Gamma = \{\underbrace{00000}_{\boldsymbol{a}_1}, \underbrace{11100}_{\boldsymbol{a}_2}, \underbrace{00111}_{\boldsymbol{a}_3}, \underbrace{11011}_{\boldsymbol{a}_4}\}$ betrachtet, für den offensichtlich $d_{\min} = 3$ gilt. In der nachfolgenden Tabelle sind für drei als beispielhaft gewählte Empfangswörter die Abstände zu allen Codewörtern angegeben und die daraus resultierende Entscheidung des Codewortschätzers:

$\boldsymbol{y}$	$d_H(\boldsymbol{y},\boldsymbol{a}_1)$	$d_H(\boldsymbol{y},\boldsymbol{a}_2)$	$d_H(\boldsymbol{y},\boldsymbol{a}_3)$	$d_H(\boldsymbol{y},\boldsymbol{a}_4)$	$\delta(\boldsymbol{y})$
10000	1	2	4	3	$\boldsymbol{a}_1$
11000	2	1	5	2	$\boldsymbol{a}_2$
10001	2	3	3	2	$\boldsymbol{a}_1$ oder $\boldsymbol{a}_4$

□

Schon beim $(7,4)$-Hamming-Code aus Beispiel 1.2 wird diese Methode ziemlich umständlich, so daß praktisch anwendbare Codes zwei Forderungen genügen sollten: (1) Die Minimaldistanz $d_{\min}$ soll möglichst groß sein. (2) Die Struktur der Codemenge Γ muß so sein, daß sich im Decoder die Suche nach dem minimalen Hammingabstand möglichst einfach organisieren läßt. Beide Forderungen setzen eine algebraische Struktur voraus, denn für (1) ist die Struktur notwendig, um überhaupt gute Codes konstruieren zu können und für (2), um realisierbare Decoder zu ermöglichen.

Satz 1.3 soll nun in entsprechender Form auch für den AWGN abgeleitet werden. Nach (1.3.11) gilt für die Verteilungsdichten:

$$\begin{aligned} f_{y|x}(\boldsymbol{y}|\hat{\boldsymbol{a}}) &= \prod_{i=0}^{n-1} f_{y|x}(y_i|\hat{a}_i) \\ &= \prod_{i=0}^{n-1} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{(y_i-\hat{a}_i)^2}{N_0}\right) \\ &= (\pi N_0)^{-n/2} \exp\left(-\frac{1}{N_0}\sum_{i=0}^{n-1}(y_i-\hat{a}_i)^2\right) \\ &= c \cdot \exp\left(-\frac{1}{N_0}\|\boldsymbol{y}-\hat{\boldsymbol{a}}\|^2\right). \end{aligned} \tag{1.6.9}$$

Dabei ist c eine Konstante und $\|\boldsymbol{y}-\hat{\boldsymbol{a}}\|^2$ die *euklidische Norm* von $\boldsymbol{y}-\hat{\boldsymbol{a}}$ bzw. der *euklidische Abstand* zwischen $\boldsymbol{y}$ und $\hat{\boldsymbol{a}}$. Der rechte Faktor muß durch geeignete Wahl von $\hat{\boldsymbol{a}}$ maximiert werden. Dies wird erreicht durch Minimierung der Norm und somit folgt:

Satz 1.4 (MLD für Soft-Decision-AWGN). *Die Wort-Fehlerwahrscheinlichkeit P_w nach der Decodierung wird minimal, wenn wie folgt decodiert wird: Zum Empfangswort $\boldsymbol{y}$ wird als Schätzung $\hat{\boldsymbol{a}} \in \Gamma$ dasjenige Codewort gewählt, das vom Empfangswort den minimalen euklidischen Abstand hat:*

$$\|\boldsymbol{y}-\hat{\boldsymbol{a}}\| \le \|\boldsymbol{y}-\boldsymbol{b}\| \quad \textit{für alle } \boldsymbol{b} \in \Gamma. \tag{1.6.10}$$

Ausgeschrieben bedeutet das $\sum_i (y_i - \hat{a}_i)^2 \leq \sum_i (y_i - b_i)^2$. Die y_i^2 können entfallen und somit vereinfacht sich das zu

$$\sum_{i=0}^{n-1} (\hat{a}_i^2 - 2y_i\hat{a}_i) \;\leq\; \sum_{i=0}^{n-1} (b_i^2 - 2y_i b_i) \quad \text{für alle } \boldsymbol{b} \in \Gamma$$

und das ist äquivalent zu

$$\sum_{i=0}^{n-1} y_i\hat{a}_i - \frac{1}{2}\sum_{i=0}^{n-1} \hat{a}_i^2 \;\geq\; \sum_{i=0}^{n-1} y_i b_i - \frac{1}{2}\sum_{i=0}^{n-1} b_i^2 \quad \text{für alle } \boldsymbol{b} \in \Gamma.$$

Speziell für binäre Codes gilt $\hat{a}_i, b_i \in \{+\sqrt{E_c}, -\sqrt{E_c}\}$ und somit entfallen die quadratischen Terme:

$$\sum_{i=0}^{n-1} y_i\hat{a}_i \;\geq\; \sum_{i=0}^{n-1} y_i b_i \quad \text{für alle } \boldsymbol{b} \in \Gamma. \tag{1.6.11}$$

Als Sendewort wird dasjenige Codewort geschätzt, bei dem die *Korrelation* mit dem Empfangswort maximal wird. Dennoch sind hier 2^k Skalarprodukte auszuführen und dies ist so aufwendig, daß Blockcodes normalerweise nur mit Hard-Decision decodiert werden können (siehe dazu auch Abschnitt 11.7).

1.7 Der Begriff des Codierungsgewinns

Die *Bit-Fehlerwahrscheinlichkeit* (BER, Bit Error Rate) bzw. *Symbol-Fehlerwahrscheinlichkeit* $P_b = P(\hat{a}_i \neq a_i)$ bezieht sich nur auf die Infosymbole und berücksichtigt nicht die Prüfsymbole. P_b und die Wort-Fehlerwahrscheinlichkeit $P_w = P(\hat{\boldsymbol{a}} \neq \boldsymbol{a})$ hängen in komplizierter Weise zusammen. Da die Anzahl der Fehler in einem fehlerhaft decodierten Wort zwischen 1 und k beträgt, gilt folglich

$$\frac{1}{k} \cdot P_w \;\leq\; P_b \;\leq\; P_w. \tag{1.7.1}$$

Normalerweise erweist sich die Näherung

$$P_b \;\approx\; \frac{d_{\min}}{k} \cdot P_w \tag{1.7.2}$$

als sinnvoll, die von $d_{\min}$ Fehlern pro falsch decodiertem Wort ausgeht. Eine ziemlich genaue Abschätzung von P_b und eine weitere Begründung für (1.7.2) erfolgt noch in Satz 3.15 – aber dieses Problem ist bei der praktischen Beurteilung von Codes von untergeordneter Bedeutung.

Der Vergleich von Codes untereinander und mit der uncodierten Übertragung erfolgt oft anhand des AWGN-Kanalmodells mit $q = 2$ gemäß Definition 1.3. Dazu sei N_0 die einseitige Rauschleistungsdichte und E_b die Energie pro Infobit. Dann ist

$$E_c = R \cdot E_b \tag{1.7.3}$$

die Energie pro Codebit, die also um den Faktor R (Coderate) kleiner ausfällt, sofern man gleiche Sendeleistung unterstellt. Die Signalleistung muß dann aufgeteilt werden auf die Infobits und auf die Prüfbits, so daß pro Codebit weniger Energie verfügbar ist. Damit wächst die Wahrscheinlichkeit, daß die Codebits im Demodulator falsch bestimmt werden. Ein Codierungsgewinn ergibt sich nur, wenn die Korrekturfähigkeit des Codes diesen negativen Effekt überwiegt.

In den Bildern 1.10 und 1.11 erfolgt nun anhand des AWGN-Modells ein quantitativer Vergleich zwischen codierter und uncodierter Übertragung, wobei P_w und P_b über E_b/N_0 aufgetragen werden. Die Kurve für die uncodierte Übertragung entspricht direkt Tabelle 1.1. Für die codierte Übertragung werden zwei *perfekte Codes* verwendet, bei denen als wesentlicher Vorteil die Wort-Fehlerwahrscheinlichkeit exakt berechnet werden kann (siehe dazu Satz 3.15).

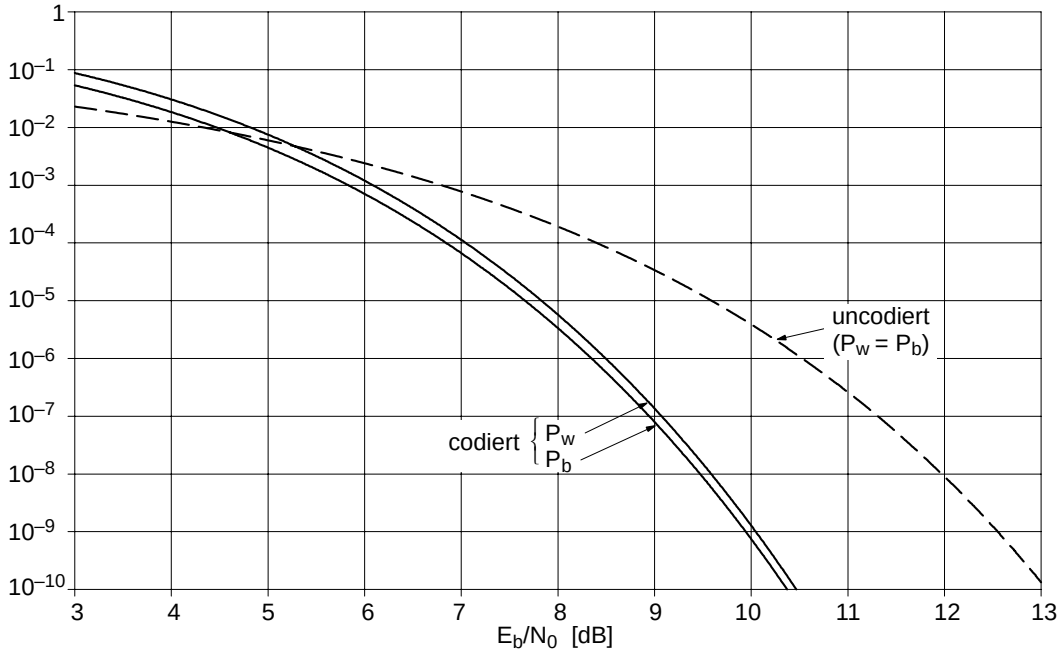

Bild 1.10. Fehlerwahrscheinlichkeit des $(23, 12)_2$-Golay-Codes (bei Hard-Decision)

Bild 1.10 zeigt am Beispiel des $(23, 12)_2$-Golay-Codes das prinzipielle Verhalten kanalcodierter Übertragungssysteme. Bei schlechten Kanälen ist die uncodierte Übertragung zunächst besser. Es gibt dann eine Schwelle (die hier bei etwa 5 dB liegt), von der an die codierte Übertragung besser wird und zu einem Codierungsgewinn (in dB) führt, der immer auf eine bestimmte Fehlerwahrscheinlichkeit P_w oder P_b bezogen wird. Bei $P_b = 10^{-6}$ beträgt dieser Codierungsgewinn etwa 2,0 dB. Unterhalb der Schwelle ergibt sich bei gleichem Signal/Rausch-Verhältnis eine kleinere Fehlerwahrscheinlichkeit bzw. bei gleicher Fehlerwahrscheinlichkeit kann die Sendeleistung reduziert werden. Bei einem sehr guten Kanal mit $E_b/N_0 \to \infty$ verlaufen die Kurven nahezu parallel

und werden gleichzeitig immer steiler. Der Unterschied zwischen P_w und P_b ist dabei unbedeutend.

Der Abstand zwischen codierter und uncodierter Übertragung wird nicht beliebig groß bei $E_b/N_0 \to \infty$, sondern konvergiert gegen einen Grenzwert, der jetzt berechnet werden soll. Zur Unterscheidung wird deshalb die Energie pro Infobit bei der uncodierten Übertragung mit E_b' und bei der codierten Übertragung mit E_b bezeichnet.

Für die uncodierte Übertragung ergibt sich P_b direkt als BSC-Bitfehlerwahrscheinlichkeit. Nach (1.3.12) und (A.3.18) gilt:

$$P_{b,\mathrm{unc}} \;=\; p_e \;=\; Q\left(\sqrt{\frac{2E_b'}{N_0}}\right) \;\approx\; \mathrm{const}\cdot e^{-E_b'/N_0}. \tag{1.7.4}$$

Wie später in Satz 3.15 noch gezeigt wird, gilt für die codierte Übertragung mit Hard-Decision für großes E_b/N_0 näherungsweise:

$$P_{w,\mathrm{cod}} \;\approx\; \mathrm{const}\cdot p_e^{t+1}. \tag{1.7.5}$$

Dabei ist const eine vom Code abhängige Konstante, p_e ist die BSC-Bitfehlerwahrscheinlichkeit der Codebits zu $E_c = RE_b$ und $t = \lfloor (d_{\min}-1)/2 \rfloor$ entspricht etwa der halben Minimaldistanz (mit $\lfloor\lambda\rfloor$ wird die größte ganze Zahl $\le \lambda$ bezeichnet). Nach (1.7.2) und (A.3.18) gilt also:

$$\begin{aligned} P_{b,\mathrm{cod}} &\approx \mathrm{const}\cdot P_{w,\mathrm{cod}} \\ &\approx \mathrm{const}\cdot p_e^{t+1} \\ &\approx \mathrm{const}\cdot \left[Q\left(\sqrt{\frac{2RE_b}{N_0}}\right)\right]^{t+1} \\ &\approx \mathrm{const}\cdot e^{-R(t+1)\cdot E_b/N_0}. \end{aligned} \tag{1.7.6}$$

Der Codierungsgewinn wird auf gleiche Bitfehlerraten bezogen: $P_{b,\mathrm{unc}} = P_{b,\mathrm{cod}}$. Hieraus folgt:

$$\mathrm{const}\cdot e^{-E_b'/N_0} \;=\; \mathrm{const}\cdot e^{-R(t+1)\cdot E_b/N_0}. \tag{1.7.7}$$

Die Konstanten sind allenfalls linear von E_b/N_0 bzw. t abhängig und können somit beim Logarithmieren für großes E_b/N_0 vernachlässigt werden:

$$\frac{E_b'}{N_0} \;\approx\; R(t+1)\cdot\frac{E_b}{N_0}. \tag{1.7.8}$$

Daraus ergibt sich der *asymptotische Codierungsgewinn* (asymptotic coding gain) für Hard-Decision als:

$$G_{\mathrm{a,hard}} \;=\; 10\cdot\log_{10}(R(t+1)) \quad \mathrm{dB}. \tag{1.7.9}$$

Für *Soft-Decision* wird später in Satz 3.17 gezeigt:

$$P_{w,\mathrm{cod}} \;\approx\; \mathrm{const}\cdot e^{-Rd_{\min}\cdot E_b/N_0}. \tag{1.7.10}$$

Der Vergleich mit (1.7.4) ergibt den asymptotischen Codierungsgewinn für Soft-Decision:

$$G_{\mathrm{a,soft}} \;=\; 10 \cdot \log_{10}(Rd_{\min}) \quad \mathrm{dB}. \tag{1.7.11}$$

Für großes E_b/N_0 ergibt sich die Fehlerwahrscheinlichkeit exponentiell aus E_b/N_0:

$$P_b \;=\; \mathrm{const} \cdot e^{-E_b/N_0 \cdot \mathrm{const}}. \tag{1.7.12}$$

Dies gilt unabhängig davon, ob codiert wird oder nicht. Für großes E_b/N_0 verlaufen die Kurven aus Bild 1.10 also parallel zueinander im Abstand $G_{\mathrm{a,hard}}$. Die Steigung der Kurven konvergiert gegen $-\infty$ für $E_b/N_0 \to \infty$. Deshalb spielen auch die Konstanten in (1.7.7) keine Rolle und auch der Zusammenhang zwischen Bit- und Wort-Fehlerwahrscheinlichkeit gemäß (1.7.1) bzw. (1.7.2) ist unwesentlich, denn der vertikale Abstand zwischen der P_b- und der P_w-Kurve bleibt zwar konstant, aber der horizontale Abstand konvergiert gegen Null.

Unmittelbar deutlich wird die enorme Bedeutung der Minimaldistanz: Je größer $d_{\min}$ wird bei gleicher Coderate (z.B. durch größere Blocklänge oder besseren Code), desto mehr kann E_b (codiert) gegenüber E_b' (uncodiert) vermindert werden. Durch Soft-Decision ergibt sich prinzipiell ein asymptotischer Gewinn von etwa 3 dB (mit $t+1 \approx d_{\min}/2$):

$$G_{\mathrm{a,soft}} \;\approx\; G_{\mathrm{a,hard}} + 3{,}01 \text{ dB}. \tag{1.7.13}$$

Bei „realistischen“ Werten von E_b/N_0 bzw. „mittleren“ Werten von P_b beträgt der Gewinn durch Soft-Decision allerdings üblicherweise nur rund 2 dB.

Für den $(23,12)_2$-Golay-Code aus Bild 1.10 mit $t=3$ ergibt sich ein asymptotischer Codierungsgewinn von $G_{\mathrm{a,hard}} = 10{\cdot}\log_{10}(12/23{\cdot}4) = 3{,}2$ dB, der allerdings aus dem Bild nicht direkt ablesbar ist, da hierfür ein größeres E_b/N_0 betrachtet werden müßte. Bei $P_w = 10^{-10}$ bzw. $E_b/N_0 = 10{,}5$ dB beträgt der Gewinn erst rund 2,5 dB. Insofern ist G_{a} eher ein Maß zum Vergleich verschiedener Codes als zur Berechnung des Codierungsgewinns bei moderaten Fehlerraten.

In Bild 1.11 ist die Fehlerwahrscheinlichkeit des $(7,4)_2$-Hamming-Codes mit $t=1$ dargestellt. Hier gilt nur $G_{\mathrm{a,hard}} = 10{\cdot}\log_{10}(4/7{\cdot}2) = 0{,}6$ dB und tatsächlich ist die codierte Übertragung nur geringfügig besser und das auch nur bei einem guten Kanal bzw. bei kleinen Fehlerwahrscheinlichkeiten. Hohe Codierungsgewinne können nur bei komplexen Codes erwartet werden, insbesondere nur bei großen Blocklängen, und dies ist auch die Aussage des Kanalcodierungstheorems (siehe Abschnitt 2.2).

Weitere Fehlerwahrscheinlichkeits-Kurven über E_b/N_0 sind in Abschnitt 3.8 (Vergleich von Hard- und Soft-Decision für den Hamming-Code) und insbesondere in Abschnitt 7.3 (BCH-Codes) sowie in Abschnitt 9.5 (Faltungscodes) angegeben. Die theoretische Berechnung der Fehlerwahrscheinlichkeit erfolgt für Blockcodes in den Abschnitten 3.7 und 3.8, für Faltungscodes in Abschnitt 9.5 und für die trelliscodierte Modulation in Abschnitt 10.7. Generell wird sich dabei zeigen, daß die Berechnung mit größerem E_b/N_0 immer einfacher wird. Andererseits können Fehlerwahrscheinlichkeiten bis herunter zur Größenordnung 10^{-6}

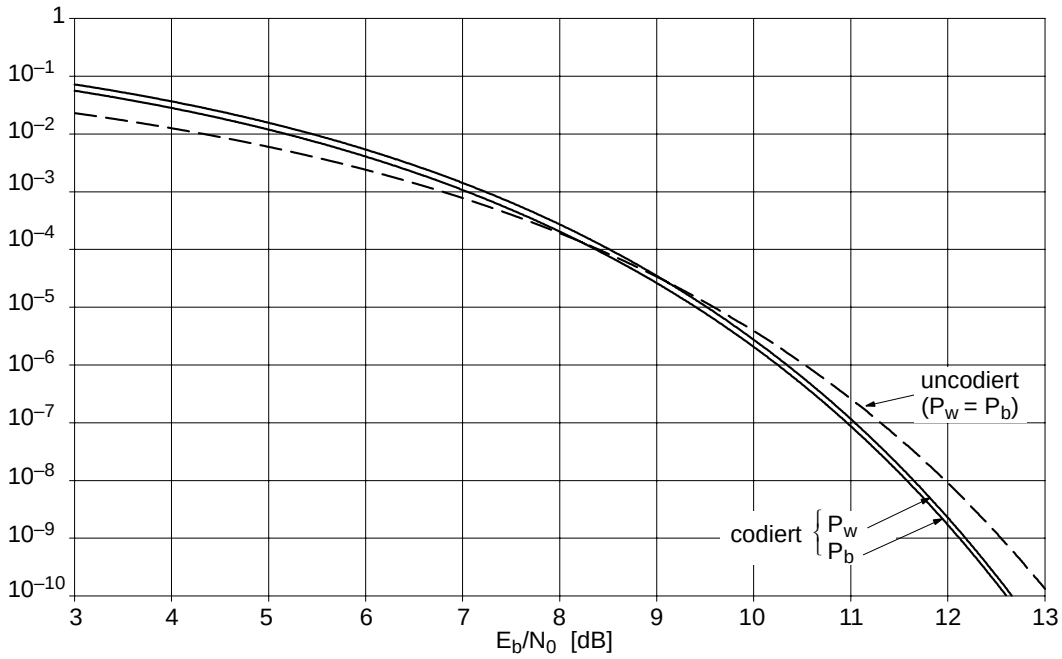

Bild 1.11. Fehlerwahrscheinlichkeit des $(7,4)_2$-Hamming-Codes (bei Hard-Decision)

noch mit vernünftigem Aufwand simuliert werden. Folglich ergänzen sich Theorie und Simulation bei den Fehlerwahrscheinlichkeits-Kurven nahezu perfekt.

Der bei einer konkreten Anwendung tatsächlich vorliegende Kanal ist oftmals so kompliziert, daß eine exakte Beschreibung nicht möglich ist. Deshalb werden Codes meistens in Bezug auf die einfachen Modelle BSC oder AWGN entworfen, womit gleichzeitig auch eine gewisse Robustheit gegen Kanaländerungen gegeben ist. Daneben werden allerdings auch noch Modelle für Kanäle mit Bündelfehlern (siehe Abschnitt 5.6, 5.7) sowie für Fadingkanäle (siehe Abschnitt 11.2, 11.3) verwendet.

Es ist also nicht überraschend, wenn der Codierungsgewinn in der Praxis nicht präzise den theoretischen Vorhersagen entspricht. Hinzu kommen immer sogenannte Implementierungsverluste sowie ein weiterer Effekt: Bei codierter Übertragung ist die Energie pro Codebit kleiner als bei der uncodierten Übertragung. Damit wird die Synchronisation (Frequenz, Takt, Rahmen) eventuell viel schwieriger, als die gesamten Operationen im Encoder und Decoder.

Ganz anders sind allerdings die Verhältnisse, wenn beim Einsatz von Codierung das Modulationssystem und die Bandbreite nicht geändert werden und stattdessen die Datenrate der Infobits vermindert wird. In diesem Fall beträgt der Gewinn $G_{\mathrm{a,hard}} = 10 \cdot \log_{10}(t+1)$ dB.

1.8 Grundgedanke der Kanalcodierung

Insbesondere mit Blockcodes wird unmittelbar der Grundgedanke deutlich, der hinter der Kanalcodierung steht. Es ist die gleiche Idee wie bei der Digitalisierung in der Nachrichtentechnik mit ähnlichen Vor- und Nachteilen:

Digitalisierung bedeutet Quantisierung der Symbole im Wertebereich: Wenn Δ die Quantisierungsbreite ist, so ist $\Delta/2$ der maximale Quantisierungsfehler. Daraus folgt:

Vorteil: Kleine Übertragungsfehler (kleiner als $\Delta/2$) werden völlig eliminiert, während bei der analogen Übertragung in jeder Verstärkerstufe das Signal/Rausch-Verhältnis prinzipiell immer schlechter wird.

Nachteil: Auch ohne Übertragungsfehler tritt immer ein mittlerer Quantisierungsfehler von $\sqrt{\Delta^2/12}$ auf. Große Übertragungsfehler (größer als $\Delta/2$) werden durch Zuordnung zur falschen Quantisierungsstufe noch vergrößert.

Fazit: Quantisierung nur dort, wo der Hauptanteil der Störungen kleiner als $\Delta/2$ ist.

Kanalcodierung bedeutet Quantisierung ganzer Symbolfolgen im Zeitbereich: Die $n-k$ Prüfstellen werden so gewählt, daß sich die verschiedenen Codewörter der Länge n an mindestens $d_{\min}$ Stellen unterscheiden. Das ist möglich, weil von den q^n möglichen Wörtern nur q^k Wörter auch tatsächlich Codewörter sind. Daraus folgt (wie in Abschnitt 3.2 noch detailliert gezeigt wird):

Vorteil: Bei weniger als $d_{\min}/2$ Übertragungsfehlern liegt das Empfangswort näher am gesendeten Wort als an allen anderen möglichen Codewörtern und kann somit richtig decodiert werden.

Nachteil: Bei mehr als $d_{\min}/2$ Übertragungsfehlern wird möglicherweise auf ein falsches Codewort entschieden und die Anzahl der Fehler erhöht sich durch die Codierung.

Fazit: Kanalcodierung nur dort, wo in der Mehrzahl weniger als $d_{\min}/2$ Fehler pro Codewort auftreten, d.h.: Kanalcodierung ist nur sinnvoll bei relativ guten Kanälen und bei hohen Anforderungen an die Zuverlässigkeit, während bei schlechten Kanälen eine Kanalcodierung nicht sinnvoll ist.

Die zentrale Idee der Kanalcodierung ist es, lange Infoblöcke in noch längere Codeblöcke zu transferieren. Das Ausmaß der hinzuzufügenden Redundanz ist abhängig von der Kanalqualität und der gewünschten Übertragungsqualität.

1.9 Aufgaben

1.1. Es wird ein BSC mit der Bit-Fehlerwahrscheinlichkeit $p_e = 0{,}01$ vorausgesetzt. Mit welcher Wahrscheinlichkeit wird das Wort 0110100 bei der Übertragung zu 0010101 verfälscht?

1.2. Es wird ein BSC mit $p_e = 0{,}1$ vorausgesetzt. Berechne für Wörter der Länge 7 die Wahrscheinlicheit für Fehlermuster aller möglichen Gewichte. Mit welcher Wahrscheinlichkeit treten höchstens 2 bzw. mehr als 2 Fehler auf?

1.3. Es wird ein BSC mit $p_e = 0{,}001$ vorausgesetzt. Berechne näherungsweise die Wahrscheinlicheit für mehr als 2 Fehler bei Wörtern der Länge 31.

1.4. Beschreibe den bei der Reihenschaltung zweier BSC's mit den Bit-Fehlerwahrscheinlichkeiten $p_{e,1}$ und $p_{e,2}$ entstehenden Kanal.

1.5. Zeige $w_H(\boldsymbol{x}+\boldsymbol{y}) \geq w_H(\boldsymbol{x}) - w_H(\boldsymbol{y})$.

1.6. Beweise und veranschauliche

$$d_H(\boldsymbol{x},\boldsymbol{y}) \geq |d_H(\boldsymbol{x},\boldsymbol{z}) - d_H(\boldsymbol{z},\boldsymbol{y})|. \tag{1.9.1}$$

1.7. Beweise die *Viereckungleichung*

$$|d_H(\boldsymbol{x},\boldsymbol{y}) - d_H(\boldsymbol{x}',\boldsymbol{y}')| \leq d_H(\boldsymbol{x},\boldsymbol{x}') + d_H(\boldsymbol{y},\boldsymbol{y}'). \tag{1.9.2}$$

1.8. Wieviele verschiedene $(3,2,2)_2$-Blockcodes gibt es?

1.9. Zeige für die Funktion $f(p_e) = p_e^r(1-p_e)^{n-r}$ das Resultat

$$\max_{p_e} f(p_e) = f\left(\frac{r}{n}\right) = 2^{-nH_2(r/n)}, \tag{1.9.3}$$

wobei $H_2(.)$ die binäre Entropiefunktion gemäß (A.2.3) bezeichnet. Interpretation?

1.10. Es wird der $(3,1,3)_2$-Code $\Gamma = \{(-1;-1;-1),(+1;+1;+1)\}$ und der AWGN vorausgesetzt. Berechne für das Empfangswort $\boldsymbol{y} = (+0{,}1;-1{,}0;+0{,}1)$ die ML-Schätzung bei Soft- und Hard-Decision.

1.11. Zeige, daß ohne die Voraussetzung gleicher Apriori-Wahrscheinlichkeiten die Wort-Fehlerwahrscheinlichkeit P_w durch den Maximum-Aposteriori-Decoder minimiert wird, der wie folgt definiert ist: Zum Empfangswort $\boldsymbol{y}$ wird als Schätzung $\hat{\boldsymbol{a}} \in \Gamma$ dasjenige Codewort gewählt, bei dem die Aposteriori-Wahrscheinlichkeit $P(x|y) = P(y|x) \cdot P(x)/P(y)$ maximal wird bzw. äquivalent

$$P_x(\hat{\boldsymbol{a}}) \cdot P_{y|x}(\boldsymbol{y}|\hat{\boldsymbol{a}}) \geq P_x(\boldsymbol{b}) \cdot P_{y|x}(\boldsymbol{y}|\boldsymbol{b}) \quad \text{für alle } \boldsymbol{b} \in \Gamma. \tag{1.9.4}$$