UNIVERSITÄT DER BUNDESWEHR MÜNCHEN
Fakultät für Elektrotechnik und Informationstechnik

# Soft-Decodierung für QAM-modulierte Signale

David Meintrup

UNIVERSITÄT DER BUNDESWEHR MÜNCHEN
Fakultät für Elektrotechnik und Informationstechnik

# Soft-Decodierung für QAM-modulierte Signale

David Meintrup

| | |
|---|---|
| Vorsitzender des Promotionsausschusses: | Prof.Dr.-Ing. K. Landes |
| 1. Berichterstatter: | Prof.Dr.rer.nat.Dr.-Ing. S. Schäffler |
| 2. Berichterstatter: | Prof.Dr.-Ing. K. Tröndle |

Tag der Prüfung: 22.01.2003

Mit der Promotion erlangter akademischer Grad:
Doktor-Ingenieur
(Dr.-Ing.)

Neubiberg, den 23. Januar 2003

# Inhaltsverzeichnis

# Notation

| | |
|---|---|
| $\mathbb{R}$ | Menge der reellen Zahlen |
| $\mathbb{R}_0^+$ | Menge der positiven reellen Zahlen mit Null |
| $\mathbb{N}$ | $\{1, 2, 3, \ldots\}$ |
| $\mathbb{N}_0$ | $\{0, 1, 2, 3, \ldots\}$ |
| n | Codelänge |
| k | Codedimension |
| $\mathbf{u} \in \{\pm 1\}^k$ | Ausgabe des Kryptocodierers |
| $\hat{\mathbf{u}} \in \{\pm 1\}^k$ | Ausgabe des Kanaldecodierers |
| $\mathbf{u}^T$ | der zu $\mathbf{u}$ transponierte Vektor |
| $\mathbf{c} \in \{\pm 1\}^n$ | Ausgabe des Kanalcodierers |
| $r : \mathbb{R} \to \mathbb{R}$ | Ergebnis der Modulation |
| $\tilde{r} : \mathbb{R} \to \mathbb{R}$ | Eingabe des Demodulators |
| $E_b$ | mittlere Energie pro Informationsbit |
| $\mathbf{I}_n$ | $n$-dimensionale Einheitsmatrix |
| $N_0$ | einseitige Rauschleistungsdichte |
| $\mathcal{N}(\mathbf{v},\mathbf{K})$ | $n$-dim. Normalverteilung mit Erwartungsvektor $\mathbf{v}$ und Kovarianzmatrix $\mathbf{K}$ |

| | |
|---|---|
| AWGN | Additive Gaussian White Noise |
| $(\{\pm 1\}, \oplus, \odot)$ | binärer Körper |
| $\mathcal{C}$ | Menge der Codewörter |
| $d_{min}$ | Minimaldistanz |
| $\mathbf{G} \in \{\pm 1\}^{k,n}$ | Generatormatrix |
| $g_{i,j}$ | $(i,j)$-tes Element einer Matrix |
| $g_{i,\cdot}$ | $i$-te Zeile einer Matrix |
| $g_{\cdot,j}$ | $j$-te Spalte einer Matrix |
| $(\Omega, \mathcal{S}, P)$ | Wahrscheinlichkeitsraum |
| $L(.)$ | L-Wert |
| $\mathbb{E}(.)$ | Erwartungswert |
| $d(\cdot,\cdot)$ | euklidische Metrik |
| $\|.\|_2$ | euklidische Norm |
| $\mathbb{V}(.)$ | Varianz |
| $\mathbb{K}(.,.)$ | Covarianz |
| MAP | Maximum A Posteriori |
| SNR | signal-to-noise ratio |
| ML | Maximum Likelihood |
| $A \dot{\cup} B, \quad \sum_{i \in I} A_i$ | disjunkte Vereinigung |
| $f : X \hookrightarrow Y$ | injektive Abbildung |
| $f \circ g$ | Kompositition der Abbildungen $f$ und $g$ |

# Einleitung

*"Suche das Einfache und misstraue ihm."*

A.N.Whitehead, Logiker und Philosoph, 1861-1947

Innerhalb der digitalen Nachrichtenübertragung hat die Decodierung die Aufgabe, aus dem demodulierten Signal die gesendete Information zu rekonstruieren. Dabei kann es zu Fehlern kommen, die letztlich von der Störung im physikalischen Kanal verursacht werden. Ziel einer jeden Decodierung ist es, die Fehlerrate so gering wie möglich zu halten. Dabei gibt es allerdings oft Nebenbedingungen, die erfüllt sein müssen, wie z.B. die zur Verfügung stehende Energie, die Komplexität der Decodierung etc. Ziel dieser Arbeit ist die Vorstellung eines Decodierverfahrens für binäre lineare Block-Codes und QAM-modulierte Signale. Um ein gutes Ergebnis, d.h. eine niedrige Fehlerrate zu erreichen, ist es allerdings wichtig, auch andere Komponenten der digitalen Nachrichtenstrecke in die Betrachtung miteinzubeziehen. Im einzelnen benötigen wir

(i) eine bestimmte Klasse von "Kombinationscodes",

(ii) eine bestimmte Modulatorfunktion,

(iii) ein geeignetes Kanalmodell,

um den Decodieralgorithmus möglichst effizient zu gestalten. Unsere Decodierung fällt in die Klasse der bitweisen Soft-Output-Decodierung. Das bedeutet im einzelnen, dass wir zum einen Bit für Bit decodieren, zum anderen für jedes Bit im Laufe der Decodierung noch ein Zuverlässigkeitsmaß, den sogenannten L-Wert (s. [HaOfPa96],[HaRu76]), berechnen. Wir werden zunächst eine genaue Formel für L-Werte jedes einzelnen Bits für QAM-modulierte Signale herleiten. Die genaue Berechnung dieser L-Werte ist im allgemeinen in der Praxis jedoch problematisch, da dazu $\min(2^k, 2^{n\mu-k})$ Additionen notwendig sind. Dabei ist $n\mu$ die Codelänge

und $k = \sum_{l=1}^{\mu} k_l$ die Anzahl der Informationsbits der von uns verwendeten Codeklasse, die sich additiv aus den Informationsbits $k_l$ einzelner Codes zusammensetzt. Daher stellen wir eine Alternative vor, die darin besteht, den zweidimensionalen QAM-Signalraum topologisch auf eine eindimensionale Situation abzubilden und bezüglich der so abgebildeten Punkte L-Werte zu bestimmen. Etwas vereinfacht erhalten wir dadurch eine Reduktion der Komplexität von

$$2^{\sum_{l=1}^{\mu} k_l} \quad \text{auf} \quad \sum_{l=1}^{\mu} 2^{k_l}. \tag{1}$$

Die QAM-Modulation gehört zu den höherstufigen Modulationsverfahren. Mehrere Bits werden zu einem Symbol zusammengefasst und in einen zweidimensionalen Raum abgebildet. Die höhere spektrale Effizienz bezahlt man durch einen Verlust an Übertragungsqualität, also durch eine höhere Fehlerrate bei gleicher Energie. Nach [Fr96] benötigt man zum Ausgleich dieses Effektes beispielsweise zwischen binärer Modulation und 16-QAM etwa 10 dB. Unsere numerischen Beispiele am Ende dieser Arbeit zeigen, dass es uns gelingt, durch das spezielle Zusammenspiel von Code, Modulatorfunktion und Decodierung diesen Verlust erheblich zu reduzieren.

Im ersten Kapitel gehen wir auf einige Grundlagen der digitalen Nachrichtenübertragung ein. Dabei legen wir einen Schwerpunkt auf die mathematische Formulierung des stochastischen Kanalmodells in Abschnitt 1.4. Im zweiten Kapitel formulieren wir zunächst die Grundlagen der bitweisen Soft-Decodierung, um sie dann am Beispiel BPSK-modulierter Signale durchzuführen (Abschnitte 2.1 und 2.2). Nach einigen Überlegungen zur Topologie können wir dann in Abschnitt 2.4 die Berechnung der L-Werte für QAM-modulierte Signale angehen. Das dritte Kapitel enthält den vollständigen Algorithmus sowie einige numerische Beispiele.

# Kapitel 1

# Grundlagen

## 1.1 Digitale Nachrichtenübertragung

Die Grundaufgabe der digitalen Nachrichtentechnik besteht darin, Information von einer Quelle zu einer Sinke zu übertragen. Quelle und Sinke sowie Sender und Empfänger können dabei viele verschiedene konkrete Ausprägungen haben. So kann es sich beispielsweise um Kommunikation zwischen zwei Mobiltelefonen handeln, um das Verschicken von Datenpaketen über Kabelverbindungen oder um eine Satellitenfunkstrecke. Je nach Anwendung sind die Anforderungen an die Übertragung sehr unterschiedlich. So kommt es beim Telefonieren in erster Linie auf eine gute Sprachqualität an, die von kleinen Fehlern bei der Übertragung unter Umständen nicht wesentlich beeinträchtigt wird. Bei der Datenübertragung ist hingegen die Fehlertoleranz in der Regel geringer. Für eine Satellitenfunkstrecke ist typischerweise die im All zur Verfügung stehende Energie eine knappe Ressource, so dass die Kommunikation mit einer möglichst geringen Energie sichergestellt werden muss. Für diese unterschiedlichen Anforderungen stehen zwischen Quelle und Sinke mehrere Komponenten zur Verfügung, die je nach Anwendung varieren können. Wir haben einen typischen Verlauf in Abbildung 1.1 skizziert. Im folgenden werden wir die einzelnen Komponenten kurz erläutern. Auf einige Bausteine, die für den weiteren Verlauf entscheidend sind, gehen wir in den folgenden Abschnitten ausführlicher ein.

**Quelle:** Wie bereits erwähnt, bezeichnet man den Ursprung der Nachrichten als Quelle. Dabei kann es sich im Mobilfunk um einen Sprecher bzw. Sprache handeln, genauso aber um einen Computer o.ä. Die genaue Form der Quelle ist für die weiteren Betrachtungen nicht relevant.

**Quellencodierer:** Die Quellencodierung ist die erste von drei Codierungsarten, die in der Shannon'schen Informationstheorie unterschieden werden. Ihre Aufgabe ist es, die Nachrichten der Quelle so in digitale Wertefolgen zu transformieren, dass dabei einerseits keine Information verloren geht, andererseits Redundanz

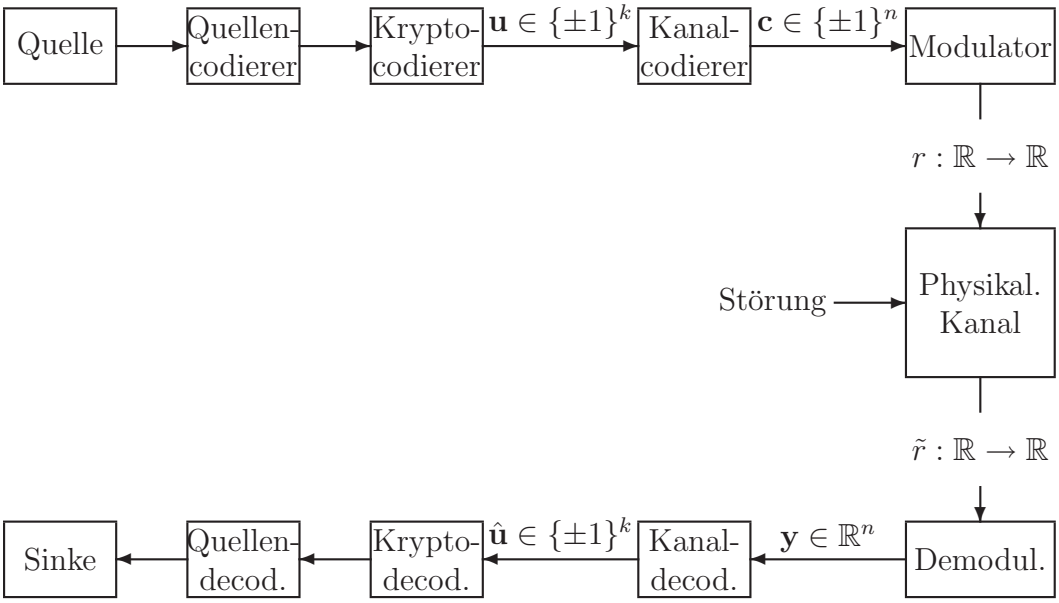


Abbildung 1.1: Digitale Nachrichtenübertragung

beseitigt wird. Der Vorteil der Quellencodierung besteht darin, dass ohne Informationsverlust Zeit und Energie bei der Übertragung gespart wird. Für eine ausführliche Darstellung verweisen wir auf [HeQu95] oder [Rot92]. Zur Realisierung der Datenreduktion können Datenkomprimierungsalgorithmen verwendet werden.

**Kryptocodierer:** Die Aufgabe des Kryptocodierers besteht darin, die Nachricht zu verschlüsseln und damit für Unbefugte nicht lesbar und nicht verfälschbar zu machen. Die Notwendigkeit einer Verschlüsselung ist nicht immer gegeben, daher ist diese Komponente optional. Ziel des Kryptocodierers ist es, auch bei ungestörter Übertragung die Nachricht so zu verändern, dass ein Empfänger ohne Kenntnis des Verschlüsselungsprinzips keine Möglichkeit erhält, an die Information zu gelangen. Die dazu benötigten theoretischen Grundlagen werden in der Kryptographie untersucht ([FuRi94], [Be93],[Ba94]).

**Kanalcodierer:** Im Kanalcodierer wird der Ausgabe des Kryptocodierers gezielt und kontrolliert Redundanz hinzugefügt. Dies ist notwendig, da im physikalischen Kanal, also bei der eigentlichen Übertragung, Störungen auftreten können, welche die Nachricht verfälschen können. Wir bezeichnen die Ausgabe des Kryptocodierers mit $\mathbf{u} \in \{\pm 1\}^k$, $k \in \mathbb{N}$. Dabei setzen wir für jedes $\mathbf{u}$ voraus, dass mit gleicher Wahrscheinlichkeit jede einzelne Komponente $\mathbf{u}_i, i = 1, \ldots, k$, den Wert $+1$ oder $-1$ annimmt. Diese Voraussetzung ist gerechtfertigt, da sie sich aus informationstheoretischer Sicht als Ziel der ersten beiden Codierungen ergibt. Durch

Hinzufügen kontrollierter Redundanz senderseitig kann auf der Empfängerseite im günstigsten Fall ein Fehler erkannt und korrigiert werden. Der Kanalcodieder bildet demnach ein Wort $\mathbf{u} \in \{\pm 1\}^k$ auf ein Codewort $\mathbf{c} \in \{\pm 1\}^n,\ n \geq k$, ab. Durch Kanalcodierung kann man daher hohe Übertragungszuverlässigkeit erreichen. Auf der anderen Seite kostet jedes hinzugefügte Redundanzbit Energie, so dass immer eine Abwägung zwischen Übertragungsqualität und Energiebedarf stattfinden muss. Wir werden im Abschnitt 1.2 auf einige Grundlagen der Kanalcodierung eingehen.

**Modulator:** Über den physikalischen Kanal können keine diskreten Werte, etwa Bits, übertragen werden, sondern nur zeitkontinuierliche Signale. Die Zuordnung der diskreten Codeworte $\mathbf{c} \in \{\pm 1\}^n$ auf kontinuierliche Signale $r : \mathbb{R} \to \mathbb{R}$ ist Aufgabe des Modulators. Dabei sind die Nebenbedingungen des physikalischen Kanals, z.B. sein Spektrum, zu beachten. Es ist naheliegend, dass der Modulator je nach gewählter Modulationsart einen entscheidenden Einfluss auf die Übertragungsqualität haben kann. Daher bilden Kanalcodierer und Modulator in manchen Szenarien eine Einheit, deren Komponenten nicht mehr scharf voneinander zu trennen sind.

**Physikalischer Kanal:** Das physikalische Medium, das der eigentlichen Übertragung dient, wird als physikalischer Kanal bezeichnet. Es kann sich dabei um leitergebundene Medien (z.B. Koaxialkabel, Glasfaserkabel) oder Funkkanäle (z.B. Mobilfunk, Rundfunk) oder auch Speichermedien (z.B. Magnetmedien, elektronische oder optische Speicher) handeln oder um eine beliebige Kombination dieser Kanäle. Charakteristisch für einen physikalischen Kanal ist, dass in ihm Störungen auftreten, d.h. er ist nicht ideal. Bei der Übertragung des Signals $r : \mathbb{R} \to \mathbb{R}$ wird man daher am Ausgang des physikalischen Kanals ein verfälschtes Signal $\tilde{r} : \mathbb{R} \to \mathbb{R}$ vorfinden. Die zufällige Störung in geeigneter Weise zu beschreiben, ist im allgemeinen eine schwierige Aufgabe der stochastischen Signaltheorie ([Boe93],[Ha91]). Wir gehen auf einige Aspekte im Abschnitt 1.4 über Kanalmodelle ein.

**Demodulator:** Der Demodulator ist das Gegenstück zum Modulator. Er wandelt das verfälschte Signal $\tilde{r} : \mathbb{R} \to \mathbb{R}$ wieder in einen diskreten Vektor $\mathbf{y} \in \mathbb{R}^n$ um. Dabei gilt im allgemeinen nicht $\mathbf{y} \in \{\pm 1\}^n$, d.h. die Ausgabe des Demodulators ist nicht notwendigerweise ein Vektor $\tilde{\mathbf{c}} \in \{\pm 1\}^n$. Dieser Unterschied erweist sich als sehr wertvoll, da die Absolutbeträge der Komponenten von $\mathbf{y}$ als Zuverlässigkeitsinformation für das entsprechende Vorzeichen Verwendung finden. Ist im Demodulator noch ein Entscheider eingebaut, also eine Abbildung von $\mathbf{y} \in \mathbb{R}^n$ auf $\tilde{\mathbf{c}} \in \{\pm 1\}^n$, so bedeutet dies zwar eine wesentliche Vereinfachung der nachfolgenden Schritte, sie geht aber mit einem erheblichen Informationsverlust einher, der sich wiederum negativ auf die Übertragungsqualität auswirkt.

**Kanaldecodierer:** Der Kanaldecodierer hat die Aufgabe, aus der Ausgabe des Demodulators die Information $\mathbf{u} \in \{\pm 1\}^k$ zu rekonstruieren. Dabei unterscheidet

man grundsätzlich zwei Arten der Decodierung. Liegt als Grundlage ein Vektor $\tilde{\mathbf{c}} \in \{\pm 1\}^n$ vor, so spricht man von Hard-Decision-Decodierung. Hat man hingegen zusätzlich Zuverlässigkeitsinformation in Form eines Vektors $\mathbf{y} \in \mathbb{R}^n$ zur Verfügung, so kann man einerseits aufgrund der zusätzlichen Information besser decodieren, andererseits auch das Ergebnis der Decodierung $\hat{\mathbf{u}} \in \{\pm 1\}^k$ mit einem Zuverlässigkeitsmaß versehen. In diesem Fall spricht man von Soft-Decision-Decodierung. Der Vorteil der Hard-Decision-Decodierung ist ihr geringerer Aufwand, den man in der Regel mit einem Verlust an Übertragungsqualität bezahlen muss. Ausführlicher gehen wir im Abschnitt 2.1 auf die verschiedenen Decodierungsmöglichkeiten ein. Ziel dieser Arbeit ist ein bestimmtes Verfahren der Soft-Decision-Decodierung vorzustellen.

**Kryptodecodierer:** Die Entschlüsselung der Ausgabe der Decodierung erfolgt im Kryptodecodierer. Er rekonstruiert also die quellencodierte Nachricht.

**Quellendecodierer:** Der Quellendecodierer verarbeitet die ankommende Information so, dass die Sinke sie verstehen kann. Im Mobilfunk z.B. erzeugt der Quellendecodierer im allgemeinen Sprache.

**Sinke:** Die Sinke ist der gewünschte Empfänger der Nachricht.

## 1.2 Kanalcodierung

Der physikalische Kanal stört das gesendete Signal $r : \mathbb{R} \to \mathbb{R}$ zu einem verfälschten Signal $\tilde{r} : \mathbb{R} \to \mathbb{R}$, wobei im allgemeinen $\tilde{r} \neq r$ gilt. Um dabei entstehende Fehler erkennen und eventuell sogar korrigieren zu können, fügt der Kanalcodierer gezielt Redundanz hinzu. Als Minimalforderung an eine solche Zuordnung wird man verlangen, dass Codewörter, die vor der Kanalcodierung verschieden waren, es auch nach der Codierung bleiben, es ist ja sogar das Ziel, sie so unterschiedlich wie möglich zu machen. Mit anderen Worten, jeder Kanalcodierer muss injektiv sein, und daher bezeichnet man ganz allgemein für zwei Mengen $X$ und $Y$ sowie zwei natürliche Zahlen $n, k$ mit $n \geq k$ jede

$$\text{injektive Abbildung } E : X^k \hookrightarrow Y^n \quad \text{als } (n,k)\text{-Block-Code.} \tag{1.1}$$

Für unsere Zwecke reichen viel speziellere Code-Klassen, die wir nun einführen wollen. Insbesondere beschränken wir uns auf binäre Codes, deren Zeichenvorrat also aus einem $n$-dimensionalen Vektorraum $\{\pm 1\}^n$, $n \in \mathbb{N}$, über dem Körper $\{\pm 1\}$ stammt. Dabei ist die Körperstruktur des zwei-elementigen Körpers $\{\pm 1\}$

mit Addition $\oplus$ und Multiplikation $\odot$ wie folgt definiert:

$$\begin{array}{rcrcr} -1 & \oplus & -1 & = & 1 \\ 1 & \oplus & 1 & = & 1 \\ 1 & \oplus & -1 & = & -1 \\ -1 & \odot & -1 & = & -1 \\ 1 & \odot & -1 & = & 1 \\ 1 & \odot & 1 & = & 1 \end{array} .$$

Nun können wir folgende Sprechweise vereinbaren. Für zwei natürliche Zahlen $n, k$ mit $n \geq k$ heißt jede

$$\begin{array}{rcl} \text{injektive Abbildung } E : \{\pm 1\}^k & \hookrightarrow & \{\pm 1\}^n \text{ binärer } (n,k)\text{-Block-Code.} \\ \mathbf{u} & \mapsto & \mathbf{c} := E(\mathbf{u}) \end{array}$$

Den Definitionsbereich eines Codes nennen wir Informationsraum, die Worte $\mathbf{u} \in \{\pm 1\}^k$ Infobits. Das Bild $\mathcal{C} := E(\{\pm 1\}^k)$ wird als Coderaum oder einfach als Code $\mathcal{C}$ bezeichnet. Denn oft kommt es nicht auf die Entstehung des Codes an, sondern lediglich auf die Menge der Codeworte $\mathcal{C}$. Dabei heißt ein binärer $(n,k)$-Block-Code

- *systematisch*, falls die ersten k Komponenten von $\mathbf{c}$ den Vektor $\mathbf{u}$ bilden,
- *linear*, falls $\mathcal{C}$ ein (k-dimensionaler) linearer Unterraum von $\{\pm 1\}^n$ ist.

Wir werden im folgenden immer systematische Codes betrachten, und lassen daher in der Regel das adjektiv "systematisch" weg. Wir wollen an dieser Stelle darauf hinweisen, dass ein linearer binärer $(n,k)$-Code $E$ *nicht* impliziert, dass die definierende Abbildung $E$ linear ist. So induziert z.B. jede Permutation $\pi : \{1, \ldots 2^k\} \to \{1, \ldots 2^k\}$ einen linearen $(k,k)$-Code

$$\begin{array}{rcl} E_\pi : \{\pm 1\}^k & \hookrightarrow & \{\pm 1\}^k \\ \mathbf{u}_j & \mapsto & E_\pi(\mathbf{u}_j) := \mathbf{u}_{\pi(j)}, \quad 1 \leq j \leq k. \end{array}$$

Dieser stellt jedoch im allgemeinen keine lineare Abbildung dar, da z.B. $E_\pi(\vec{0}) \neq \vec{0}$ sein wird. Der lineare Unterraum $\mathcal{C}$ hat jedoch eine $k$-dimensionale Basis, deren Vektoren eine $n \times k$ - Matrix $A$ bilden. Da wir über dem Körper $\{\pm 1\}$ arbeiten, gilt

$$\mathcal{C} = \{\mathbf{v} \in \{\pm 1\}^n | \mathbf{v} = A \cdot \mathbf{u},\ u \in \{\pm 1\}^k\}. \tag{1.2}$$

Mit anderen Worten, $\mathcal{C}$ ist das Bild der von $A$ induzierten linearen Abbildung

$$\begin{array}{rcl} f_A : \{\pm 1\}^k & \hookrightarrow & \{\pm 1\}^n \\ \mathbf{u} & \mapsto & f_A(\mathbf{u}) := A \cdot \mathbf{u}. \end{array}$$

Es gilt jedoch im allgemeinen nicht $f_A = E$. Vielmehr gibt es nach obigen Ausführungen zu jedem linearen $(n,k)$-Block-Code $E$ eine Matrix $A$ aus Basisvektoren von $\mathcal{C}$ sowie eine Permutation $\pi : \{1,\dots 2^k\} \to \{1,\dots 2^k\}$, so dass gilt:

$$\begin{aligned} E &= f_A \circ E_\pi, \\ E(\mathbf{u}_j) &= A \cdot \mathbf{u}_{\pi(j)} \quad 1 \le j \le k. \end{aligned}$$

Die transponierte Matrix zu $A$, die $k \times n$-Matrix $G := A^T$, wird als Generator-Matrix des linearen $(n,k)$-Block-Codes bezeichnet. Der Zusammenhang zwischem dem Coderaum $\mathcal{C}$ und der Generatormatrix $G$ ergibt sich aus (1.2):

$$\mathcal{C} = \{\mathbf{c} \in \{\pm 1\}^n | \mathbf{c}^T = \mathbf{u}^T G,\ \mathbf{u} \in \{\pm 1\}^k\}. \tag{1.3}$$

Sind zwei Codewörter $\mathbf{c}, \tilde{\mathbf{c}} \in \{\pm 1\}^n$ gegeben, so ist der sogenannte Hamming-Abstand definiert durch

$$d(\mathbf{c}, \tilde{\mathbf{c}}) := |\{i | \mathbf{c}_i \neq \tilde{\mathbf{c}}_i,\ i = 1, \dots, n\}|. \tag{1.4}$$

Der Hamming-Abstand bildet eine Metrik auf $\{\pm 1\}^n$. Die Minimaldistanz $d_{min}$ eines Codes $E$, definiert als

$$d_{min} := \min\{d(\mathbf{c}, \tilde{\mathbf{c}}) | \mathbf{c}, \tilde{\mathbf{c}} \in \mathcal{C}, \mathbf{c} \neq \tilde{\mathbf{c}}\} \tag{1.5}$$

ist ein Gütekriterium für einen Code. Sie gibt die Anzahl Bits an, in denen sich zwei verschiedene Codeworte mindestens unterscheiden. Je größer die Minimaldistanz, desto besser lassen sich Fehler erkennen und korrigieren. Aufgrund der nach Definition geltenden Injektivität eines Codes gilt $d_{min} \geq 1$. Andererseits gilt offensichtlich $d_{min} \leq n-k+1$. Wie vielleicht bisher schon ein wenig deutlich wurde, wird das Studium von Codes stark durch algebraische Methoden geprägt. Die Untersuchung linearer Codes kann man als Teil der Linearen Algebra über endlichen Körpern auffassen, allerdings mit einem entscheidenden Unterschied zur klassischen Linearen Algebra. Die Codes sind basisabhängig, so ist z.B. der Minimalabstand $d_{min}$ keine basisunabhängige Invariante. Für eine ausführliche Darstellung der algebraischen Codierungstheorie sei auf [Ju95] und [Ro95] verwiesen. Die wesentlichen Eigenschaften eines Codes werden durch das Tupel $(n,k)$ bzw. das Tripel $(n,k,d_{min})$ wiedergegeben. Insbesondere enthält es das Information-Redundanz-Verhältnis $R := \frac{k}{n}$, das als Coderate $R$ bezeichnet wird.

Wir wollen uns nun einige Beispiele linearer $(n,k)$-Block-Codes anschauen. Um die Lesbarkeit zu vereinfachen, werden wir dabei nicht streng zwischen einer Notation für Zeilen- und Spaltenvektoren unterscheiden. Die ersten beiden Beispiele stellen zwei Extremfälle dar, die aber als Testfälle oft sehr nützlich sind.

**Beispiel 1.2.1 (uncodierte Übertragung).** Rein formal können wir $n = k$ und als Code die Identität wählen:

$$E_{uncod} = \mathrm{id} : \{\pm 1\}^k \hookrightarrow \{\pm 1\}^k \tag{1.6}$$

$$\mathbf{u} \mapsto \mathrm{id}(\mathbf{u}) = \mathbf{u}. \tag{1.7}$$

In diesem Fall spricht man von uncodierter Übertragung. Es handelt sich formal um einen binären, linearen $(k, k)$-Block-Code mit dem schlechtest möglichen Minimalabstand $d_{min} = 1$ und der best möglichen Coderate $R = 1$.

**Beispiel 1.2.2 (Wiederholungscode).** In diesem Fall sei $k = 1$ und $n > 1$. Die Abbildungsvorschrift ist gegeben durch

$$E_{Rn} : \{\pm 1\} \hookrightarrow \{\pm 1\}^n \tag{1.8}$$

$$\mathbf{u} \mapsto \mathbf{u} \odot (-1, \ldots, -1). \tag{1.9}$$

Das Bit $\mathbf{u}$ wird also einfach $n$-mal wiederholt. In diesem Fall spricht man von einem $n$-fachen Wiederholungscode. Es handelt sich um einen linearen $(n, 1)$-Block-Code mit dem best möglichen Minimalabstand $d_{min} = n$ und der schlechtest möglichen Coderate $R = \frac{1}{n}$.

**Beispiel 1.2.3 (Der $(7, 4)$-Hamming Code).** Der Hamming-Code läßt sich am einfachsten durch seine $4 \times 7$ Generator-Matrix angeben:

$$G = \begin{pmatrix} -1 & 1 & 1 & 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{pmatrix}, \tag{1.10}$$

wobei die Operationen $\oplus$ und $\odot$ zugrundegelegt sind. Der Code ist in diesem Fall gegeben durch die lineare Abbildung

$$E_{Ham} : \{\pm 1\}^4 \hookrightarrow \{\pm 1\}^7 \tag{1.11}$$

$$\mathbf{u} \mapsto \mathbf{c} = G^T \mathbf{u}. \tag{1.12}$$

Der Hamming-Code ist ein binärer linearer Code mit Minimaldistanz $d_{min} = 3$ und einer Coderate von $\frac{4}{7}$.

**Beispiel 1.2.4 (Kombinationscode).** Natürlich kann man verschiedene Codes miteinander kombinieren, um neue zu erhalten. Schaltet man mehrere Codes hintereinander, spricht man von verketteten Codes. Wir wollen die uns nun schon bekannten Codes in folgender Weise kombinieren. Sei $\mu \geq 1$ eine natürliche Zahl und seien $\mu$ $(n, k_l)$-Block-Codes $E_l$, $l = 1, \ldots, \mu$ gegeben. Dann erhalten wir als Informationsraum $\prod_{l=1}^{\mu} \{\pm 1\}^{k_l}$ und unser Coderaum liegt in $(\{\pm 1\}^n)^\mu \simeq \{\pm 1\}^{\mu \times n}$.

Wir erhalten so einen matrix-wertigen Code, dessen Zeilen aus den $\mu$ einzelnen Codes besteht:

$$E_{Komb} : \prod_{l=1}^{\mu} \{\pm 1\}^{k_l} \quad \hookrightarrow \quad \{\pm 1\}^{\mu \times n} \tag{1.13}$$

$$\mathbf{u} = (\mathbf{u}_{l,\cdot})_{1 \le l \le \mu} \quad \mapsto \quad E_{Komb}(\mathbf{u}) = \left(E_l(\mathbf{u}_{l,\cdot})\right)_{1 \le l \le \mu}. \tag{1.14}$$

Wir betrachten also bei diesem Kombinationscode den Coderaum $\mathcal{C}$ als Teilraum der $\mu \times n$-Matrizen mit Einträgen im Körper $\{\pm 1\}$. In den Zeilen der Matrix stehen die einzelnen Codes. Dies ermöglicht, die Bits in den Zeilen unterschiedlich stark zu schützen, eine Tatsache, die wir uns im Zusammenhang mit der QAM-Modulation zu Nutze machen werden. Insgesamt bleibt aber auch der Kombinationscode ein linearer $(N, k)$-Block-Code, mit $k = \sum_{l=1}^{\mu} k_l$, $N = \mu n$ und der Minimaldistanz $D_{min} = \min\{d_l | l = 1, \ldots, \mu\}$.

Wir wollen an dieser Stelle noch ein explizites Beispiel für einen Kombinationscode angeben. Dazu setzen wir $\mu = 4$, $n = 7$ und wählen

| | |
|---|---|
| $E_1 := E_{R7}$ | 7-facher Wiederholungscode, |
| $E_2 := E_{Ham}$ | $(7, 4)$-Hamming-Code, |
| $E_3 := E_{Ham}$ | $(7, 4)$-Hamming-Code, |
| $E_4 := E_{uncod}$ | uncodierte Übertragung. |

Dann ensteht der folgende Kombinationscode, den wir als $E^I$ bezeichnen:

$$E^I : \{\pm 1\}^1 \times \{\pm 1\}^4 \times \{\pm 1\}^4 \times \{\pm 1\}^7 \quad \hookrightarrow \quad \{\pm 1\}^{4 \times 7} \tag{1.15}$$

$$\mathbf{u} \mapsto E^I(\mathbf{u}) = \begin{pmatrix} \mathbf{c}_{1,\cdot} \\ \mathbf{c}_{2,\cdot} \\ \mathbf{c}_{3,\cdot} \\ \mathbf{c}_{4,\cdot} \end{pmatrix} = \begin{pmatrix} E_{R7}(\mathbf{u}_{1,\cdot}) \\ E_{Ham}(\mathbf{u}_{2,\cdot}) \\ E_{Ham}(\mathbf{u}_{3,\cdot}) \\ E_{uncod}(\mathbf{u}_{4,\cdot}) \end{pmatrix}. \tag{1.16}$$

Hierbei ist das Konstruktionsprinzip derart, dass die Anzahl Redundanzbits von oben nach unter in der Matrix abnimmt, d.h. die oberen Informationsbits sind besser geschützt also die weiter unten liegenden Informationsbits. Dies werden wir an späterer Stelle mit einer Modulationsart verbinden, die genau für eine solche Situation ausgelegt ist. Insgesamt betrachtet handelt es sich bei $E^I$ um einen linearen $(28, 16)$-Block-Code mit der gleichen Coderate $\frac{16}{28} = \frac{4}{7}$ wie der $(7, 4)$-Hamming-Code. Allerdings ist aufgrund der letzten uncodierten Zeile die Minimaldistanz lediglich 1.

## 1.3 Modulation

Im nächsten Abschnitt wollen wir Kanalmodelle studieren. Da diese aber von der Modulation abhängen, werden wir zunächst kurz diejenigen Modulationsarten einführen, die wir im folgenden implizit voraussetzen werden. Explizit werden die Modulationen nicht benötigt, da sie im Kanalmodell integriert sind. Für eine vertiefte Darstellung der Modulation verweisen wir auf [Kam96] und [Pr00]. Die Aufgabe der Modulation ist es, Symbole in reellwertige Signale abzubilden, die über den physikalischen Kanal geschickt werden können. Grundsätzlich können sich Signale in Amplitude, Frequenz und Phase unterscheiden. Um einen Modulator formal einzuführen, betrachten wir einerseits einen

$$\text{Symbolraum } \mathcal{S} \tag{1.17}$$

und andererseits zu gegebenem $l \geq 1$ eine

$$\text{injektive Abbildung } \rho : \mathcal{S} \hookrightarrow \mathbb{R}^l. \tag{1.18}$$

Die Abbildung $\rho$ werden wir in diesem Zusammenhang immer als Modulatorfunktion bezeichnen. Es seien $r_1, \ldots, r_l : \mathbb{R} \to \mathbb{R}$ fest gewählte, linear unabhängige Signale gleicher endlicher Energie aus einem gegebenen Signalraum $\mathcal{R}$, dann heißt eine Abbildung

$$M : \mathcal{S} \;\to\; \mathcal{R} \tag{1.19}$$

$$\mathbf{s} \;\mapsto\; \sum_{i=1}^{l} \rho_i(\mathbf{s}) r_i \quad \text{Modulator .} \tag{1.20}$$

Die Signale $r_1, \ldots, r_l : \mathbb{R} \to \mathbb{R}$ heißen Trägersignale. Unter tatsächlichen Übertragungsbedingungen darf jedes Signal natürlich nur eine bestimmte Zeitdauer $T$, Übertragungsperiode genannt, gesendet werden. Diese zeitliche Abhängigkeit spielt für uns jedoch keine Rolle, genauso wenig wie die genaue Ausprägung der Trägersignale. Für die Kanalmodelle einzig entscheidend ist die Modulatorfunktion $\rho : \mathcal{S} \hookrightarrow \mathbb{R}^l$, welche den zu übertragenden Symbolen im Falle $l = 1$ Punkte auf der reellen Achse, im Falle $l = 2$ Punkte in der Ebene zuordnet. Der $\mathbb{R}^l$ kann als $l$-dimensionaler Signalraum aufgefasst werden, der von den Trägersignalen aufgespannt wird. Die nun folgenden zwei Beispiele werden uns immer wieder als Modellfälle begegnen.

**Beispiel 1.3.1 (BPSK-Modulation).** In der oben eingeführten Notation sei $l = 1$ und der Symbolraum $\mathcal{S}_{BPSK} = \{\pm 1\}$. Dann ist die Modulatorfunktion gegeben durch

$$\rho_{BPSK} = \mathrm{id}_{\mathcal{S}_{BPSK}} : \mathcal{S} \;\to\; \mathbb{R}, \tag{1.21}$$

$$\mathbf{s} \;\mapsto\; \mathbf{s}. \tag{1.22}$$

Entsprechend gilt für den Modulator

$$\begin{aligned} M_{BPSK} : \mathcal{S}_{BPSK} &\to \mathcal{R}, && (1.23)\\ +1 &\mapsto r_1, && (1.24)\\ -1 &\mapsto -r_1. && (1.25) \end{aligned}$$

Der Name BPSK steht für *"binary phase shift keying"*. Er erklärt sich durch die Wahl der Trägerfunktion, die üblicherweise durch

$$\begin{aligned} r_1 : \mathbb{R} &\to \mathbb{R}, && (1.26)\\ t &\mapsto g(t)\cos(2\pi f_0 t) && (1.27) \end{aligned}$$

gegeben ist. Nun gilt aber

$$\begin{aligned} M_{BPSK}(+1) = r_1(t) &= g(t)\cos(2\pi f_0 t), && (1.28)\\ M_{BPSK}(-1) = -r_1(t) &= -g(t)\cos(2\pi f_0 t) = g(t)\cos(2\pi f_0 t + \pi). && (1.29) \end{aligned}$$

Die Information über das Vorzeichen des Bits liegt also in einer Phasenverschiebung des Cosinus-Signals um $\pi$. Phasenverschiebungen um kleinere Winkel führen zu der Möglichkeit, mehr Bits pro Kanalbenutzung zu übertragen. Dieses Konzept führt zur $l$-PSK-Modulation (*$l$-phase shift keying*), s. [Pr00].

**Beispiel 1.3.2 ($m^2$-QAM-Modulation).** Die QAM-Modulation gehört zur Gruppe der höherdimensionalen Modulationsarten, bei der mehrere Bits zu einem Symbol zusammengefasst werden. Den Symbolen werden dann in einem quadratischen Muster Punkte im $\mathbb{R}^2$ zugeordnet. Genauer seien $m, \mu \geq 2$ ganzzahlig und $m^2 = 2^\mu$. Wir schließen damit einfachheitshalber die sogenannten Kreuzkonstellationen aus, die Idee unserer Decodierung ist jedoch sofort übertragbar. Dann ist der Symbolraum der $m^2$-QAM-Modulation $\mathcal{S}_\mu = \{\pm 1\}^\mu$. Die Modulatorfunktion $\rho_{m^2}$ mit $l = 2$ ist gegeben durch:

$$\begin{aligned} \rho_{m^2} : \mathcal{S}_\mu &\hookrightarrow \mathbb{R}^2, && (1.30)\\ \rho_{m^2}(\mathcal{S}_\mu) &= \left\{ \pm\frac{1}{\sqrt{2}(m-1)}, \pm\frac{3}{\sqrt{2}(m-1)}, \ldots, \pm\frac{m-1}{\sqrt{2}(m-1)} = \pm\frac{1}{\sqrt{2}} \right\}^2 \end{aligned}$$

Zur Veranschaulichung sind die ersten 3 Fälle, $m^2 = 4$, $m^2 = 16$ und $m^2 = 64$ in den Abbildungen 1.2, 1.3 und 1.4 dargestellt. In der Praxis wird QAM-Modulation bis $m = 1024$ (Richtfunk) eingesetzt. Wir weisen hier darauf hin, das wir bisher in (1.30) nur das Bild von $\rho$ angegeben haben und noch nicht die Abbildung $\rho$ selbst. Grundsätzlich bleiben uns bei $m^2$-QAM-Modulation noch $(m^2)!$ Möglichkeiten, die Symbole auf die Punkte im $\mathbb{R}^2$ zu verteilen. Welche Wahl getroffen wird, ist durchaus von Bedeutung. Wir wollen dies im folgenden am Beispiel 16-QAM erläutern. Dazu haben wir zwei konkrete Zuordnungen

ausgesucht. Die Abbildung 1.3 zeigt das sogenannte Gray-Coding ([Fr96]). Das Prinzip besteht darin, dass zwei benachbarte Punkte sich in ihrem Codewort jeweils um genau ein Bit unterscheiden. Wenn eine Störung des Signals zu einer falschen Entscheidung führt, so wird diese oft aus einem Nachbarpunkt bestehen. Da dieses sich aber nur in einem Bit vom richtigen Codewort unterscheidet, besteht eine große Wahrscheinlichkeit, diesen Fehler mit Hilfe der Kanalcodierung zu erkennen. Dies setzt allerdings ein bestimmtes Decodierverfahren voraus, auf das wir im Abschnitt 2.1 eingehen werden. Für eine andere Art der Decodierung ist die Zuordnung gemäß Abbildung 1.5 geeigneter (vgl. [Bo92][p. 440]). Ihr Entstehungsprinzip ist in Abbildung 1.6 dargestellt. Diese Form der Partitionierung der Signalpunkte geht auf Ungerböck [Ung82] zurück. Dabei ist der Grundgedanke der folgende. Entscheidet man sich in der Decodierung bitweise und fängt damit links an, so könnte die Entscheidung für das erste Bit beispielsweise "+1" sein. Anschließend sollen die potenziellen Konkurrenten, also die übrigen Punkte mit führendem Bit "+1" so weit weg sein wie möglich, daher hat man vorher die direkten Nachbarn mit einer "-1" versehen.

Welche Zuordnung tatsächlich besser geeignet ist, hängt wie schon erwähnt vom Decodierverfahren ab. Für das von uns in dieser Arbeit vorgestellte Decodierverfahren werden wir die spezielle Wahl nach Abbildung 1.6 treffen. Diese bezeichnen wir immer mit dem Index '$P$', was an die notwendige Partitionierung der Signalpunkte erinnern soll. Für die Modulatorfunktion schreiben wir demnach

$$\rho_{m^2}^P : \mathcal{S}_\mu \hookrightarrow \mathbb{R}^2. \tag{1.31}$$

Natürlich muss auch die Kanalcodierung entsprechend gewählt werden. Wir werden darauf an späterer Stelle nochmal zurückkommen. Im Zusammenhang mit der QAM-Modulation $\rho_{m^2}^P$ nach Abbildung 1.6 werden wir noch einige Notationen benötigen, die wir deshalb an dieser Stelle einführen wollen. Zunächst bezeichnen

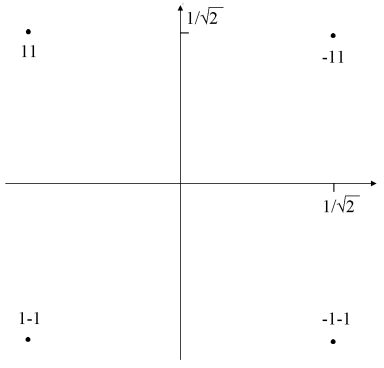


Abbildung 1.2: 4-QAM Modulation

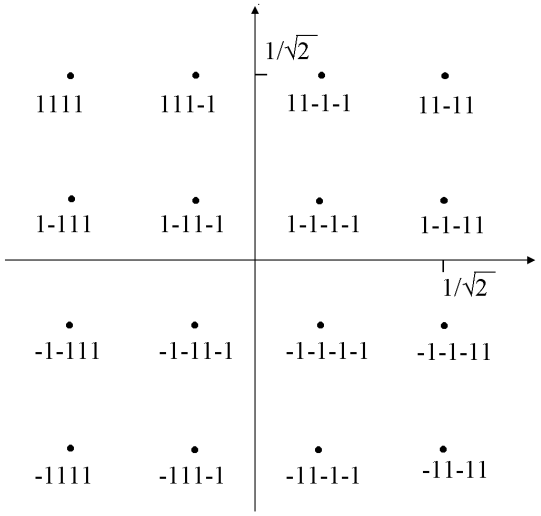


Abbildung 1.3: 16-QAM Modulation mit Gray-Codierung

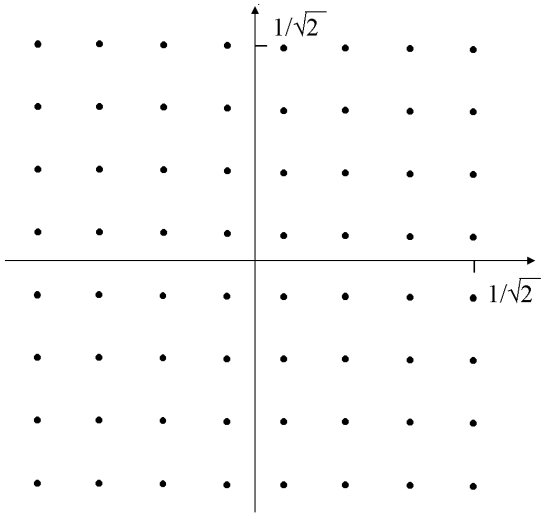


Abbildung 1.4: 64-QAM Modulation

wir die Bildpunkte der Modulatorfunktion als

$$P := \rho_m^P(\mathcal{S}_\mu) \tag{1.32}$$

und numerieren die $m^2$ Elemente gemäß ihrer binären Symbole:

$$P = \{p_1, \ldots, p_{m^2}\}. \tag{1.33}$$

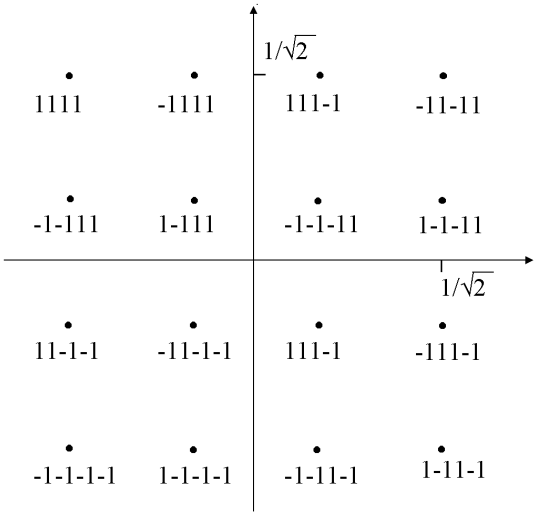


Abbildung 1.5: 16-QAM nach Partitionierung

Durch ein hochgestelltes $b$, das an binäre Darstellung erinnern soll, bezeichnen wir die Umkehrfunktion zu $\rho_{m^2}^P$:

$$( \ )^b : P \ \rightarrow \ \mathcal{S}_\mu \tag{1.34}$$

$$p_i \ \mapsto \ p_i^b \tag{1.35}$$

Mit

$$p_i^b(l), \quad l = 1, \ldots, \mu \tag{1.36}$$

bezeichnen wir die $l$-te Stelle in der Binärdarstellung. So gilt beispielsweise:

$$p_5^b \ = \ +1-1+1-1 \in \mathcal{S}_4 \tag{1.37}$$

$$\text{und } p_5^b(3) \ = \ +1. \tag{1.38}$$

Damit können wir jetzt für ein gegebens $f \in \{0, \ldots, \mu\}$ folgende Teilmengen der Signalpunkte $P$ definieren:

$$P_{c_1 \ldots c_f} = \{p_i \in P | p_i^b(l) = c_l \text{ für alle } 1 \leq l \leq f\}. \tag{1.39}$$

Mit anderen Worten besteht $P_{c_1 \ldots c_f}$ aus denjenigen Signalpunkten, deren erste $f$ Stellen der Binärdarstellung mit $c_1 \ldots c_f$ übereinstimmen. Um auch hier ein Beispiel anzugeben, betrachten wir für 16-QAM-Modulation die Menge $P_{1-1}$ und lesen aus Abbildung 1.5 und 1.6 ab, dass

$$P_{1-1} = \{p_4, p_5, p_6, p_7\}. \tag{1.40}$$

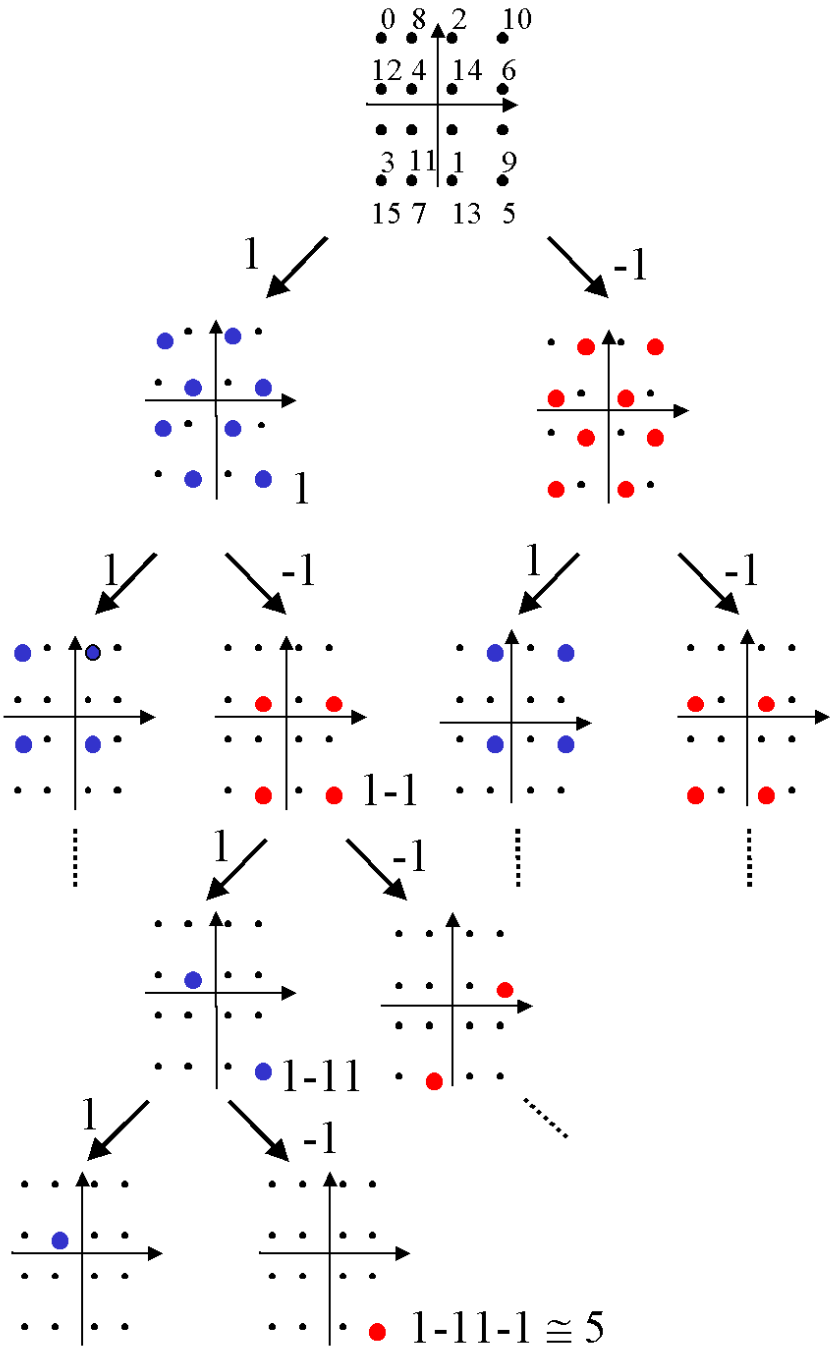


Abbildung 1.6: Partitionierung der Signalpunkte für $\rho^P_{m^2}$-Modulation

gilt. Der Sinn dieser Mengen ist der folgende: Wir werden ein Decodierverfahren vorstellen, bei dem ein Kombinationscode wie aus Beispiel 1.2.4 zeilenweise decodiert wird. Sind dabei z.B. die ersten beiden Zeilen schon decodiert und das Ergebnis in einer Spalte 1 und $-1$, so wird man sich sinnvollerweise für die Decodierung der nächsten Zeile auf diejenigen Punkte einschränken, deren erste zwei Bits 1 und $-1$ sind, also gerade die Menge $P_{1-1}$.

## 1.4 Kanalmodelle

Die Störungen, die bei der Übertragung eines Signals über einen physikalischen Kanal auftreten, sind nicht im einzelnen greifbar, genügen aber in der Regel dennoch gewissen Gesetzmäßigkeiten. Daher versucht man, durch ein stochastisches Modell, den Kanal möglichst genau wiederzugeben. Dies kann sich als sehr schwierig erweisen, wenn beispielsweise, wie im Mobilfunk, der Kanal zeitvariant ist, sich also die Übertragungsbedingungen mit der Zeit verändern. Aber auch zeitinvariante Kanäle können so viele verschiedene Effekte aufweisen, dass ein gutes stochastisches Kanalmodell sehr komplex wird. Wir werden in einem ersten Teil die allgemeine Theorie der Kanalmodellierung vorstellen. Anschließend werden wir uns auf zwei konkrete Beispiele konzentrieren, die an die bereits in Abschnitt 1.3 beschriebenen Situationen anknüpfen.

Im stochastischen Kanalmodell fasst man die Komponenten "Modulator", "physikalischer Kanal" und "Demodulator" zusammen. Dies ist in Abbildung 1.7 dargestellt. Dort ist auch eine weitere Komponente eingeführt, der Symbolisator (Mapper), den wir ebenfalls als Teil des Kanalmodells auffassen. Wir haben bereits dargestellt, dass der Modulator Symbole aus einem Symbolraum $\mathcal{S}$ in einen $l$-dimensionalen Signalraum abbildet, während der Kanalcodierer Codewörter $\mathbf{c} \in \mathcal{C} \subset \{\pm 1\}^n$ liefert. Es bedarf also noch einer Komponente, die aus den Codeworten $\mathbf{c} \in \mathcal{C}$ Symbole aus $\mathcal{S}$ macht. Diese Aufgabe übernimmt der Symbolisator. Dabei können aus einem Codewort mehrere Symbole enstehen. Jedes einzelne Symbol wird über den Kanal übertragen. Formal lässt sich der Symbolisator also zu gegebenem Symbolraum $\mathcal{S}$ und $p \geq 1$ folgendermaßen beschreiben:

$$b : \mathcal{C} \rightarrow \mathcal{S}^p, \tag{1.41}$$

$$\mathbf{c} \mapsto b(\mathbf{c}) = (b(\mathbf{c})_1, \ldots, b(\mathbf{c})_p). \tag{1.42}$$

Um das Konzept zu verdeutlichen, schauen wir uns Beispiele an.

**Beispiel 1.4.1 (Symbolisator BPSK).** Wir gehen von einem gegebenen $(n, k)$-Block-Code $\mathcal{C}$ aus. Der Symbolraum der BPSK-Modulation ist $\mathcal{S}_{BPSK} = \{\pm 1\}$.

Aus einem Codewort $\mathbf{c}$ werden also $p = n$ Symbole. Daher besteht der Symbolisator einfach aus der komponentenweisen Identität:

$$b_{BPSK} : \mathcal{C} \rightarrow \mathcal{S}_{BPSK}^n, \tag{1.43}$$

$$\mathbf{c} \mapsto \mathbf{c}. \tag{1.44}$$

Dieser Symbolisator scheint auf den ersten Blick künstlich, gilt doch $\mathcal{C} = \mathcal{S}_{BPSK}^n$ und besteht die Abbildung aus der Identität. Es gibt jedoch einen entscheidenden Punkt, der die Einführung auch in diesem Beispiel sinnvoll macht und in dem nachfolgenden Beispiel noch deutlicher wird: In unserer Vorstellung ist $\mathbf{c} \in \mathcal{C}$ **ein** Codewort, während $\mathbf{c} \in \mathcal{S}_{BPSK}^n$ aus $\mathbf{n}$ Symbolen besteht, die einzeln von der Modulatorfunktion $\rho : \mathcal{S} \rightarrow \mathbb{R}^l$ in den $l$-dimensionalen Signalraum abgebildet werden.

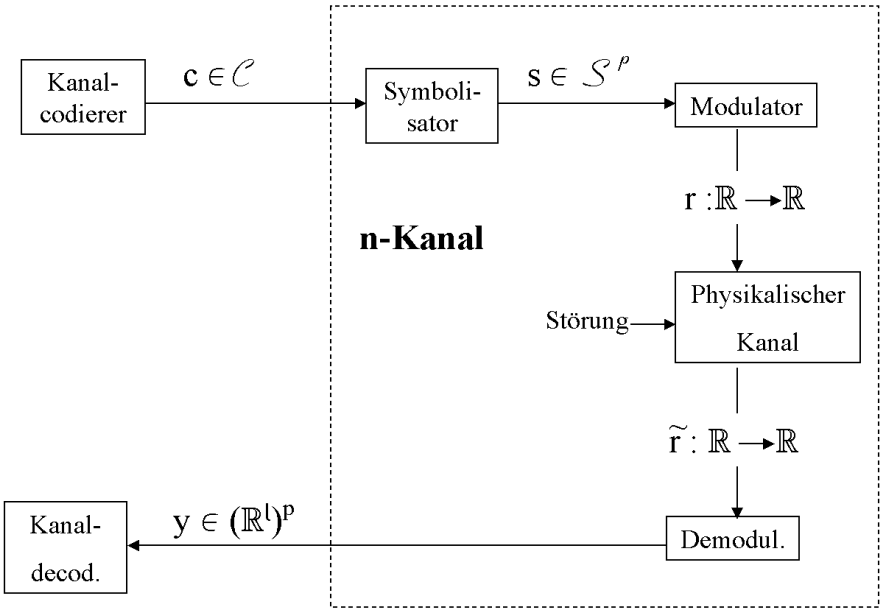


Abbildung 1.7: Kanal-Modell

**Beispiel 1.4.2 (Symbolisator QAM).** Der Symbolraum der $m^2$-QAM-Modulation ist $\mathcal{S}_\mu = \{\pm 1\}^\mu$, wobei $m^2 = 2^\mu$ gilt. Es werden also jedesmal $\mu$ Bit zu einem Symbol zusammengefasst:

$$b_{m^2} : \mathcal{C} \rightarrow \mathcal{S}_\mu^p. \tag{1.45}$$

Wie die Codeworte in jeweils $\mu$ Bit große Teilstücke eingeteilt werden, hängt von dem gegebenen Code ab. Wir wollen den in Beispiel 1.2.4 betrachteten Kombinationscode $E_{Komb}$ verwenden. Dessen Ausgabe besteht ja aus einer $\mu \times n$-Matrix, die wir spaltenweise zu Symbolen zusammenfassen. Aus einem Codewort werden so $n$ Symbole aus $\mathcal{S}_\mu$, je eines pro Spalte des Kombinationscodes. Formal können wir dies folgendermaßen beschreiben:

$$\begin{aligned} b_{m^2} : \mathcal{C} \subset \{\pm 1\}^{\mu \times n} &\rightarrow \mathcal{S}_\mu^n, && (1.46) \\ (\mathbf{c_{i,j}})_{\substack{1 \le i \le \mu \\ 1 \le j \le n}} &\mapsto \{\mathbf{c}_{.,1}, \ldots, \mathbf{c}_{.,n}\}. && (1.47) \end{aligned}$$

Wir benötigen für diesen speziellen Code also $n$ Symbole, um ein Codewort zu erfassen. Entsprechend wird der Kanal $n$ mal benutzt, um ein Codewort vollständig zu übertragen.

Jetzt haben wir alles bereit gestellt, um uns mit dem eigentlichen Kanalmodell auseinanderzusetzen. Das stochastische Kanalmodell macht aus einem Codewort $\mathbf{c} \in \mathcal{C}$ genau $p$ Vektoren $z_j \in \mathbb{R}^l,\ j = 1, \ldots, p$, im $l$-dimensionalen Signalraum des Modulators. Dabei ist für ein stochastisches Kanalmodell typisch, dass der Vektor $\mathbf{z} \in (\mathbb{R}^l)^p$ die Realisierung einer speziellen Zufallsvariablen darstellt. Die Gestalt der Zufallsvariablen bestimmt dann gerade die Eigenschaften des Kanals. Um dies formalisieren zu können, versehen wir den Coderaum $\mathcal{C} \subset \mathbb{R}^n$ mit der diskreten Topologie und betrachten einen gegebenen Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, \mathbb{P})$. Dann heißt eine ($\times$ bezeichnet das Kreuzprodukt)

$$\text{messbare Abbildung } \mathcal{K} : \ \mathcal{C} \times \Omega \rightarrow \mathbb{R}^n \ n\text{-Kanal}. \tag{1.48}$$

Da wir $\mathcal{C}$ mit der diskreten Topologie versehen haben, ist für einen $n$-Kanal gleichbedeutend, dass für jedes $\mathbf{c} \in \mathcal{C}$ die Abbildung

$$\mathcal{K}_\mathbf{c} : \Omega \rightarrow \mathbb{R}^n \text{ eine } n\text{-dimensionale reelle Zufallsvariable ist.} \tag{1.49}$$

Das stochastische Kanalmodell besteht also aus Zufallsvariablen, die über dem Coderaum $\mathcal{C}$ parametrisiert sind. Schauen wir uns Beispiele an:

**Beispiel 1.4.3 (ungestörter Kanal).** Der ungestörte Kanal ist ein theoretischer Idealfall. Im stochastischen Kanalmodell spiegelt sich dieser Fall durch das Auftreten der Einpunktverteilung wieder. Konkret ist das Kanalmodell $\mathcal{K}_{ungest}$ gegeben durch:

$$\begin{aligned} \mathcal{K}_{ungest} : \ \mathcal{C} \times \Omega &\rightarrow \mathbb{R}^n, && (1.50) \\ (\mathbf{c}, \omega) &\mapsto \mathbf{c}. && (1.51) \end{aligned}$$

Neben seinem Modell-Charakter ist der Nutzen des ungestörten Kanals insbesondere als Testfall für die Decodierung zu sehen. Offensichtlich sollte jede Decodierung im ungestörten Fall exakt das gesendete Wort liefern.

Bevor wir zum nächsten, fundamentalen Beispiel kommen, wollen wir noch ein paar grundlegende Notationen festlegen. Auf die Performance eines Kanals hat die zur Verfügung stehende Energie entscheidenden Einfluss. Um eine leichte Vergleichbarkeit für unterschiedliche Codes und Coderaten zu erreichen, wird im allgemeinen nicht die Energie pro Kanalbenutzung, sondern die Energie

$$E_b : \text{Energie pro Informationsbit} \tag{1.52}$$

betrachtet. Auf der anderen Seite ordnet man der Störung eine Größe $N_0$ zu,

$$N_0 : \text{die einseitige Rauschleistungsdichte.} \tag{1.53}$$

Entscheidend ist nun das Verhältnis dieser beiden Größen, also der Quotient

$$\frac{E_b}{N_0} : \text{SNR} = \text{Signal-to-Noise Ratio}, \tag{1.54}$$

der als Signal-Rausch-Verhältnis bekannt ist und üblicherweise in Dezibel angegeben wird. Desweiteren benötigen wir eine Notation für die

$$n\text{-dimensionale Einheitsmatrix: } \mathbf{I}_n. \tag{1.55}$$

**Beispiel 1.4.4 (allgemeiner AWGN-Kanal).** Wir gehen von einem gegebenen Coderaum $\mathcal{C}$ eines $(n,k)$-Block-Codes aus, sowie einem

$$\begin{aligned} \text{Symbolisator } b : \mathcal{C} &\rightarrow \mathcal{S}^p & (1.56)\\ \mathbf{c} &\mapsto b(\mathbf{c}) = (b(\mathbf{c})_j)_{1\le j\le p} & (1.57)\end{aligned}$$

und einer Modulatorfunktion $\rho : \mathcal{S} \to \mathbb{R}^l$. Dann ist der allgemeine AWGN-Kanal (AWGN = *additive gaussian white noise*) gegeben durch:

$$\begin{aligned} \mathcal{K} : \mathcal{C} \times \Omega &\rightarrow (\mathbb{R}^l)^p & (1.58)\\ \mathcal{K}_c : \Omega \to (\mathbb{R}^l)^p \quad &\text{ist} \quad \mathcal{N}(\rho(b(\mathbf{c})_j)_{1\le j\le p}, \mathbf{I}_{lp}\frac{N_0 n}{2E_b k\cdot l}) \text{ normalverteilt.}\end{aligned}$$

Fasst man $\mathcal{K}_c$ als $p$ einzelne Zufallsvariablen von $\Omega$ in den $\mathbb{R}^l$ auf,

$$\mathcal{K}_c = (\mathcal{K}_c^1, \ldots, \mathcal{K}_c^p), \quad \mathcal{K}_c^j : \Omega \to \mathbb{R}^l, \quad 1 \le j \le p, \tag{1.59}$$

was einer einzelnen Betrachtung der $p$ Kanalbenutzungen entspricht, so erhält man:

$$\mathcal{K}_c^j \text{ ist } \mathcal{N}\left(\rho(b(\mathbf{c})_j), \mathbf{I}_l \frac{N_0 n}{2E_b k\cdot l}\right) \text{ normalverteilt, } 1 \le j \le p. \tag{1.60}$$

Die normalverteilten Zufallsvariablen $\mathcal{K}_c^j$ haben also als Erwartungswert den Signalpunkt $\rho(\mathbf{s}_j)$, der dem Symbol $\mathbf{s}_j = b(\mathbf{c})_j$ zugeordnet wurde. Ihre Komponenten sind unkorreliert und aufgrund der Normalverteilung somit unabhängig. Die Varianz ist bis auf Normierungen gegeben durch $\frac{E_b k}{N_0 n}$, was die Energie pro Kanalbenutzung darstellt.

Der AWGN-Kanal, der genauer ein gedächtnisloser, zeitinvarianter AWGN-Kanal ist, ist sicherlich das wichtigste Kanalmodell der Nachrichtentechnik. Wir spezialisieren es nun auf die zwei uns begleitenden Modulationsarten, BPSK und QAM.

**Beispiel 1.4.5 (AWGN-Kanal, BPSK).** Wir bezeichnen wieder mit $\mathbf{c} \in \mathcal{C}$ die Ausgabe des $(n,k)$-Block-Codes des Kanalcodierers. Im Fall der im Beispiel 1.3.1 behandelten BPSK-Modulation ist der Symbolraum $\mathcal{S} = \{\pm 1\}$, der Symbolisator gemäß (1.43) die Identität

$$\begin{aligned} b_{BPSK} : \mathcal{C} &\rightarrow \mathcal{S}_{BPSK}^n, && (1.61)\\ \mathbf{c} &\mapsto \mathbf{c}, && (1.62) \end{aligned}$$

und die Modulatorfunktion gemäß (1.21) gegeben durch

$$\begin{aligned} \rho_{BPSK} = \mathrm{id}_{\mathcal{S}_{BPSK}} : \mathcal{S}_{BPSK} &\rightarrow \mathbb{R}, && (1.63)\\ \mathbf{s} &\mapsto \mathbf{s}. && (1.64) \end{aligned}$$

Wir erhalten damit

$$\rho_{BPSK}(b_{BPSK}(\mathbf{c}))_j = c_j, \text{ für jedes } c \in \mathcal{C},\ 1 \leq j \leq n. \tag{1.65}$$

Daraus ergibt sich durch Einsetzen in (1.58) als AWGN-Kanalmodell für BPSK-Modulation:

$$\begin{aligned} \mathcal{K}^{BPSK} : \mathcal{C} \times \Omega &\rightarrow \mathbb{R}^n && (1.66)\\ \mathcal{K}_c^{BPSK} &\text{ ist } \mathcal{N}(\mathbf{c}, \mathbf{I}_n \frac{N_0 n}{2E_b k}) \text{ normalverteilt.} && (1.67) \end{aligned}$$

Die normalverteilten Zufallsgrößen $\mathcal{K}_c^{BPSK}$ haben also als Erwartungswert das gesendete Codewort $\mathbf{c}$. Ihre Komponenten sind unkorreliert und aufgrund der Normalverteilung somit unabhängig. In die Varianz fließt wieder die Energie pro Kanalbenutzung $\frac{E_b k}{N_0 n}$ ein. Dieser AWGN-Kanal wird uns als Modellfall für die Entwicklung der Soft-Decodierung dienen.

**Beispiel 1.4.6 (AWGN-Kanal, QAM).** Genau wie im vorherigen Beispiel müssen wir auch hier lediglich die einzelnen Komponenten aus den vorherigen Abschnitten zusammentragen und in das allgemeine AWGN-Kanalmodell (1.58) einsetzen. Wir gehen wieder von einem gegebenen Kombinationscode

$$\begin{aligned} E_{Komb} : \prod_{l=1}^{\mu} \{\pm 1\}^{k_l} &\hookrightarrow \{\pm 1\}^{\mu \times n} && (1.68)\\ E_{Komb}\left(\prod_{l=1}^{\mu} \{\pm 1\}^{k_l}\right) &= \mathcal{C} && (1.69) \end{aligned}$$

gemäß Beispiel 1.2.4 aus. Die Anzahl Infobits ist $k = \sum_{l=1}^{\mu} k_l$. Die Modulatorfunktion ist nach 1.30 gegeben als

$$\begin{aligned} \rho_{m^2} : \mathcal{S}_\mu &\hookrightarrow \mathbb{R}^2, \qquad (1.70) \\ \rho_{m^2}(\mathcal{S}_\mu) &= \left\{ \pm\frac{1}{\sqrt{2}(m-1)}, \pm\frac{3}{\sqrt{2}(m-1)}, \ldots, \pm\frac{m-1}{\sqrt{2}(m-1)} = \pm\frac{1}{\sqrt{2}} \right\}^2 \end{aligned}$$

und der Symbolisator folgt aus (1.46)

$$\begin{aligned} b_{m^2} : \mathcal{C} \subset \{\pm 1\}^{\mu \times n} &\to \mathcal{S}_\mu^n, \qquad (1.71) \\ (\mathbf{c_{i,j}})_{\substack{1 \le i \le \mu \\ 1 \le j \le n}} &\mapsto \{\mathbf{c}_{\cdot,1}, \ldots, \mathbf{c}_{\cdot,n}\}. \qquad (1.72) \end{aligned}$$

Wir erhalten damit

$$\rho_{m^2}(b_{m^2}(\mathbf{c}))_j = \rho_{m^2}(\mathbf{c}_{\cdot,j}), \text{ für jedes } \mathbf{c} \in \mathcal{C},\ 1 \le j \le n. \qquad (1.73)$$

Die zur Verfügung stehende Energie pro Kanalbenutzung ist $\frac{E_b k}{N_0 n}$. Daraus ergibt sich durch Einsetzen in (1.60) als AWGN-Kanalmodell für $m^2$-QAM-Modulation:

$$\begin{aligned} \mathcal{K}^{m^2} :&\quad \mathcal{C} \times \Omega \to \mathbb{R}^{2 \times n} \qquad (1.74) \\ \mathcal{K}_c^{m^2,j} :&\quad \text{ist } \mathcal{N}(\rho_{m^2}(\mathbf{c}_{\cdot,j}), \mathbf{I}_2 \frac{N_0 n}{2E_b k \cdot 2}) \text{ normalverteilt, } 1 \le j \le n. \end{aligned}$$

Um dieses Modell etwas besser zu verstehen, wollen wir nochmal an einige Strukturen erinnern. Der Vektor $\mathbf{c}_{\cdot,j}$ ist die $j$-te Spalte im Kombinationscode (1.68). D.h., dass jeder Eintrag je nach Zeile zu einem anderen Code gehört. Zusammen ergeben sie eine $\mu$-stellige Binärzahl, von denen es $2^\mu = m^2$ verschiedene gibt. Die Modulatorfunktion $\rho_{m^2}$ ordnet dieser Zahl einen Punkt im quadratischen Muster der $m^2$-QAM-Modulation zu. Bezeichnen wir diesen Punkt kurz mal als $p_j \in \mathbb{R}^2$ und fixieren die Varianz $\sigma^2 := \frac{N_0 n}{2E_b k \cdot 2}$, so vereinfacht sich (1.74) zu:

$$\begin{aligned} \mathcal{K}^{m^2} :&\quad \mathcal{C} \times \Omega \to (\mathbb{R}^2)^n \qquad (1.75) \\ \mathcal{K}_c^{m^2,j} :&\quad \text{ist } \mathcal{N}(p_j, \mathbf{I}_2 \sigma^2) \text{ normalverteilt, } 1 \le j \le n. \end{aligned}$$

Jede Spalte des Kombinationscodewortes $\mathbf{c}$ ergibt also eine 2-dimensionale Normalverteilung mit gleicher Varianz und einem Punkt aus dem $m^2$-QAM-Schema als Erwartungswert.

# Kapitel 2

# Bitweise Soft-Decodierung

## 2.1 Grundlagen der Decodierung

Der Kanaldecodierer hat die Aufgabe, aus der Ausgabe des Demodulators die gesendeten Informationsbits $\mathbf{u} \in \{\pm 1\}^k$ zu rekonstruieren. Dies kann auf viele verschiedene Weisen geschehen, die sich in entscheidenden Merkmalen wie Fehlerwahrscheinlichkeit, mathematischer Aufwand, Optimalitätskriterium etc. unterscheiden. Für welche Decodierung man sich entscheidet, hängt also von der Anwendung sowie physikalischen Nebenbedingungen wie Echtzeitanforderungen, Komplexität oder Chipdimensionierung ab. Wir wollen zunächst einige Decodierprinzipien kurz vorstellen, um anschließend die bitweise Soft-Decodierung, die uns in diesem Kapitel beschäftigen wird, ausführlicher darzustellen.

**Hard-Decision:** Liegt als Grundlage der Decodierung ein Vektor $\tilde{\mathbf{c}} \in \{\pm 1\}^n$ vor, so spricht man von Hard-Decision-Decodierung. Der Vorteil der Hard-Decision-Decodierung ist ihr grundsätzlich geringerer Aufwand und einfachere Implementierung im Vergleich zur Soft-Decodierung. Diesen Vorteilen steht eine im allgemeinen schlechtere Qualität, also eine höhere Fehlerrate bei gleicher Energie pro Infobit gegenüber. Innerhalb der Hard-Decision-Decodierung gibt es wiederum verschiedene Verfahren, wie die Maximum-Likelihood Decodierung (ML) oder die Begrenzte-Minimaldistanz Decodierung (BMD). Für eine ausführliche Darstellung dieser Verfahren verweisen wir auf [Bo92].

**Soft-Decision:** Diese Decodiermethode entspricht dem von uns vorgestellten Konzept des stochastischen Kanalmodells. Ausgangspunkt ist somit eine Realisierung $\mathbf{y} \in (\mathbb{R}^l)^p$ einer Zufallsvariablenschar $\mathcal{K}$, die das stochastische Kanalmodell darstellt. Ziel ist es, die zusätzliche Information, die in dem reellen Vektor $\mathbf{y} \in (\mathbb{R}^l)^p$ enthalten ist, dahingehend zu interpretieren, dass sie als Zuverlässigkeitsinformation das Ergebnis der Decodierung stützt bzw. verbessert. Der Vorteil der Soft-Decodierung hat also zwei Aspekte: Einerseits wird die zusätzliche

Information zu einer besseren Decodierung, also weniger Fehlern verwendet. Andererseits kann die getroffene Entscheidung mit einem Sicherheitsmaß versehen werden. Letzteres ist insbesondere im Zusammenhang mit verketteten Codes von Bedeutung. Natürlich spiegeln sich die Vorteile der Hard-Decision Decodierung als Nachteile der Soft-Decision Decodierung wieder, insbesondere der höhere mathematische Aufwand verbunden mit einer komplexeren Implementierung.

Mathematisch besteht die Decodieraufgabe nun darin, aus der gegebenen Realisierung $\mathbf{y} \in (\mathbb{R}^l)^p$ des stochastischen Kanalmodells $\mathcal{K}$ die Informationsbits $\mathbf{u} \in \{\pm 1\}^k$ in einem noch zu definierenden Sinn "optimal" zurückzugewinnen. Dies ist eine klassische Aufgabe der mathematischen Statistik und der mathematischen Optimierung. Gesucht ist demnach eine Entscheidungsfunktion

$$\begin{aligned} \delta : (\mathbb{R}^l)^p &\rightarrow \{\pm 1\}^k && (2.1) \\ \mathbf{y} &\mapsto \hat{\mathbf{u}} = \delta(\mathbf{y}). && (2.2) \end{aligned}$$

Man kann sich eine Entscheidungsfunktion $\delta$ als Partition bzw. Äquivalenzrelation auf dem $(\mathbb{R}^l)^p$ vorstellen, deren Äquivalenzklassen $[\mathbf{u}] = \delta^{-1}(\mathbf{u})$ gerade durch die Informationsworte $\mathbf{u} \in \{\pm 1\}^k$ parametrisiert sind und als Menge aus den Urbildern $\delta^{-1}(\mathbf{u})$ bestehen. Zur Bestimmung von $\delta$ bieten sich zwei verschiedene Strategien an:

**Wortfehleroptimalität:** Hier besteht die Aufgabe darin, so zu decodieren, dass die Wahrscheinlichkeit, dass sich im decodierten Wort $\hat{\mathbf{u}} = \delta(\mathbf{y})$ kein falsches Bit befindet, maximal ist. Die Anzahl der Bitfehler in einem Wort $\hat{\mathbf{u}}$ spielt keine Rolle. Um dieses Optimalitätskriterium exakt formulieren zu können, gehen wir von einem gegebenen

$$(n,k)\text{-Block-Code } E : \{\pm 1\}^k \hookrightarrow \{\pm 1\}^n \tag{2.3}$$

aus, sowie von einem stochastischen Kanalmodell

$$\mathcal{K} : \mathcal{C} \times \Omega \rightarrow (\mathbb{R}^l)^p$$

auf einem gegebenen Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, \mathbb{P})$. Dann besteht die Optimierungsaufgabe darin,

$$\delta_{WO} : (\mathbb{R}^l)^p \rightarrow \{\pm 1\}^k \text{ so zu wählen, dass} \tag{2.4}$$

$$\sum_{\mathbf{u} \in \{\pm 1\}^k} P\left(\omega \in \Omega | \delta_{WO} \circ \mathcal{K}_{E(\mathbf{u})}(\omega) = \mathbf{u} = \mathrm{id}_{\{\pm 1\}^k}(\mathbf{u})\right) \text{ maximal wird.} \tag{2.5}$$

Die konkrete Bestimmung von $\delta_{WO}$ für ein bestimmtes Kanalmodell führt auf ein diskretes Optimierungsproblem, dessen Zielfunktion im allgemeinen nur numerisch auswertbar ist. In diesem Zusammenhang haben sich Branch-and-Bound Algorithmen sowie der Metropolis-Algorithmus bewährt.

**Bitfehleroptimalität:** Hier besteht die Aufgabe des Decoders darin, so zu decodieren, dass die Wahrscheinlichkeit für jedes einzelne Informationsbit, richtig zu sein, maximal wird. Zur Bestimmung der Partition

$$\delta_{BO} : (\mathbb{R}^l)^p \to \{\pm 1\}^k \tag{2.6}$$

stellen wir nun die sogenannte MAP-Decodierung (*maximum a posteriori*) mit Hilfe von L-Werten vor. Ausgangspunkt ist wieder ein gegebener (systematischer)

$$(n,k)\text{-Block-Code } E : \{\pm 1\}^k \hookrightarrow \{\pm 1\}^n \tag{2.7}$$

sowie ein stochastisches Kanalmodell

$$\mathcal{K} : \mathcal{C} \times \Omega \to (\mathbb{R}^l)^p$$

auf einem gegebenen Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, \mathbb{P})$. Sei $\mathbf{y} \in (\mathbb{R}^l)^p$ eine Realisierung des Kanalmodells und $\mathbf{u}_i, i = 1, \ldots, k$ das $i$-te Informationsbit in $\mathbf{u} \in \{\pm 1\}^k$. Dann lässt sich der Grundgedanke des L-Wertes folgendermaßen ausdrücken:

$$''L(\mathbf{u}_i|\mathbf{y}) = \ln \left[ \frac{P(\mathbf{u}_i = +1 | \mathcal{K} = \mathbf{y})}{P(\mathbf{u}_i = -1 | \mathcal{K} = \mathbf{y})} \right]'' \tag{2.8}$$

Wir weisen hier ausdrücklich darauf hin, dass es sich bei (2.8) nicht um einen mathematisch präzisen Ausdruck handelt, sondern lediglich um die Idee des L-Wertes. Der L-Wert setzt unter der Bedingung, dass $\mathbf{y}$ empfangen wurde (daher a posteriori) die Wahrscheinlichkeit, dass das $i$-te Infobit gleich 1 ist, ins Verhältnis dazu, dass das $i$-te Infobit gleich $-1$ ist. Der Logarithmus sorgt dafür, dass das Vorzeichen des L-Wertes angibt, welche der beiden Wahrscheinlichkeiten größer ist und sein Absolutbetrag den Unterschied zwischen den Wahrscheinlichkeiten misst, also als Zuverlässigkeit für die Entscheidung angesehen werden kann. Um den Ausdruck (2.8) in eine mathematische Form zu bringen, sind einige Schritte erforderlich. Zum einen betrachten wir aufgrund der Stetigkeit der vorkommenden Zufallsvariablen den Ball mit Radius $\varepsilon$ in der 1-Norm $\|\cdot\|_1$ um die Realisierung $\mathbf{y}$:

$$B(\mathbf{y}, \varepsilon) := \{z \in (\mathbb{R}^l)^p | \|z - \mathbf{y}\|_1 \leq \varepsilon\} \tag{2.9}$$

Jetzt können wir für $1 \leq i \leq k$ folgenden Ausdruck definieren:

$$L_\varepsilon(\mathbf{u}_i|\mathbf{y}) := \ln \left[ \frac{\sum\limits_{\substack{\mathbf{u}\in\{\pm1\}^k \\ \mathbf{u}_i=+1}} P(\omega \in \Omega | \mathcal{K}_{E(\mathbf{u})}(\omega) \in B(\mathbf{y}, \varepsilon))}{\sum\limits_{\substack{\mathbf{u}\in\{\pm1\}^k \\ \mathbf{u}_i=-1}} P(\omega \in \Omega | \mathcal{K}_{E(\mathbf{u})}(\omega) \in B(\mathbf{y}, \varepsilon))} \right] \tag{2.10}$$

$$= \ln \left[ \frac{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C} \\ \mathbf{c}_i=+1}} P(\omega \in \Omega | \mathcal{K}_{\mathbf{c}}(\omega) \in B(\mathbf{y}, \varepsilon))}{\sum\limits_{\substack{\mathbf{u}\in\mathcal{C} \\ \mathbf{c}_i=-1}} P(\omega \in \Omega | \mathcal{K}_{\mathbf{c}}(\omega) \in B(\mathbf{y}, \varepsilon))} \right] \tag{2.11}$$

Wir wollen nun weiter voraussetzen, dass die Zufallsvariablen $\mathcal{K}_{\mathbf{c}}$ für jedes $\mathbf{c} \in \mathcal{C}$ eine

$$\text{Dichtefunktion } d_{\mathbf{c}} : (\mathbb{R}^l)^p \to \mathbb{R}, \quad \mathbf{c} \in \mathcal{C}, \tag{2.12}$$

besitzen. Dadurch erhalten wir durch Einsetzen in (2.11):

$$L_\varepsilon(\mathbf{u}_i|\mathbf{y}) = \ln \left[ \frac{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C} \\ \mathbf{c}_i=+1}} \int\limits_{B(\mathbf{y},\varepsilon)} d_{\mathbf{c}}(x)\,\mathrm{dx}}{\sum\limits_{\substack{\mathbf{u}\in\mathcal{C} \\ \mathbf{c}_i=-1}} \int\limits_{B(\mathbf{y},\varepsilon)} d_{\mathbf{c}}(x)\,\mathrm{dx}} \right] \tag{2.13}$$

Betrachten wir nun durch wiederholte Anwendung der Regel von de L'Hospital den Grenzübergang $L_\varepsilon(\mathbf{u}_i|\mathbf{y})$ für $\varepsilon \to 0$ und bezeichnen den Grenzwert als $L(\mathbf{u}_i|\mathbf{y})$, so erhalten wir den

$$\text{L-Wert } L(\mathbf{u}_i|\mathbf{y}) = \ln \left[ \frac{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C} \\ \mathbf{c}_i=+1}} d_{\mathbf{c}}(\mathbf{y})}{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C} \\ \mathbf{c}_i=-1}} d_{\mathbf{c}}(\mathbf{y})} \right], \quad i = 1, \ldots, k. \tag{2.14}$$

Wir haben bereits erwähnt, dass das Vorzeichen des L-Werts zur Entscheidung für das entsprechende Bit herangezogen wird. Daher können wir nun als bitfehleroptimale Decodierfunktion

$$\delta_{BO} : (\mathbb{R}^l)^p \to \{\pm 1\}^k, \tag{2.15}$$

$$\mathbf{y} \mapsto (\text{signum}(L(\mathbf{u}_i|\mathbf{y}))_{(1\le i\le k)} \tag{2.16}$$

angeben. Der Logarithmus im L-Wert bewirkt lediglich eine Umskalierung des Intervalls $]0, \infty[$ auf das Intervall $]-\infty, \infty[$. Dadurch wird einerseits die Entscheidungsschwelle vom Punkt 1 auf den Punkt 0 abgebildet, was eine bequeme Entscheidung mittels des Vorzeichens ermöglicht. Andererseits erlaubt es, den Absolutbetrag des L-Werts direkt als Zuverlässigkeitsmaß für die Entscheidung zu interpretieren.

## 2.2 Soft-Decodierung BPSK-modulierter Signale

In diesem Abschnitt wollen wir die bitweise Softdecodierung, die wir im letzten Abschnitt vorgestellt haben, am Beispiel BPSK-modulierter Signale veranschaulichen. Wir gehen wieder von einem gegebenen Kanalcodierer mit einem $(n, k)$-Block-Code $\mathcal{C}$ aus. Das stochastische Kanalmodell für BPSK-modulierte Signale

in einem AWGN-Kanal ist gemäß (1.66) gegeben durch:

$$\mathcal{K}^{BPSK} : \mathcal{C} \times \Omega \quad \rightarrow \quad \mathbb{R}^n \tag{2.17}$$

$$\mathcal{K}_c^{BPSK} : \quad \text{ist} \quad \mathcal{N}(\mathbf{c}, \mathbf{I}_n \frac{N_0 n}{2E_b k}) \text{ normalverteilt.} \tag{2.18}$$

Die Dichten der $n$-dimensionalen Normalverteilungen mit diagonalen Kovarianzmatrizen sind bekanntlich:

$$d_{\mathbf{c}}^{BPSK} : \mathbb{R}^n \quad \rightarrow \quad \mathbb{R}, \tag{2.19}$$

$$\mathbf{y} \quad \mapsto \quad \frac{1}{\sqrt{2\pi}^n \sqrt{\frac{N_0 n}{2E_b k}}^n} \exp\left( -\frac{\|\mathbf{y} - \mathbf{c}\|^2}{2 \cdot \frac{N_0 n}{2E_b k}} \right). \tag{2.20}$$

Setzen wir diese Dichten in unsere L-Werte (2.14) ein, so erhalten wir, da sich die Normierungsfaktoren wegheben, die folgende Gestalt:

$$L^{BPSK}(\mathbf{u}_i|\mathbf{y}) = \ln \left[ \frac{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C} \\ \mathbf{c}_i=+1}} \exp\left( -\frac{\|\mathbf{y}-\mathbf{c}\|^2}{\frac{N_0 n}{E_b k}} \right)}{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C} \\ \mathbf{c}_i=-1}} \exp\left( -\frac{\|\mathbf{y}-\mathbf{c}\|^2}{\frac{N_0 n}{E_b k}} \right)} \right], \quad i = 1, \ldots, k. \tag{2.21}$$

Entsprechend lautet unsere Decodiervorschrift für BPSK-modulierte Signale gemäß (2.15):

$$\delta_{BO}^{BPSK} : \mathbb{R}^n \quad \rightarrow \quad \{\pm 1\}^k, \tag{2.22}$$

$$\mathbf{y} \quad \mapsto \quad \text{signum} \left( \ln \left[ \frac{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C} \\ \mathbf{c}_i=+1}} \exp\left( -\frac{\|\mathbf{y}-\mathbf{c}\|^2}{\frac{N_0 n}{E_b k}} \right)}{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C} \\ \mathbf{c}_i=-1}} \exp\left( -\frac{\|\mathbf{y}-\mathbf{c}\|^2}{\frac{N_0 n}{E_b k}} \right)} \right] \right)_{(1 \le i \le k)}. \tag{2.23}$$

Im Prinzip ist durch die Angabe der Decodiervorschrift $\delta_{BO}^{BPSK}$ die Decodieraufgabe für die bitweise Soft-Decodierung BPSK-modulierter Signale gelöst. Allerdings wird aus der Gleichung (2.21) deutlich, dass der Aufwand an Additionen zur Berechnung der L-Werte proportional zu $2^k$ ist. Durch Verwendung orthogonaler Codes kann man dies auf $\min(2^k, 2^{n-k})$ reduzieren. Trotzdem bleibt eine exakte Berechnung unter Umständen schwierig. Eine effiziente approximative Lösungsmethode zur Bestimmung dieser L-Werte wurde daher in [Sc97] entwickelt.

## 2.3 Topologische Überlegungen

Um die bitweise Softdecodierung QAM-modulierter Signale zu entwickeln, werden wir ein wenig Topologie auf dem Signalraum verwenden. Die Grundlagen der

von uns verwendeten mengentheoretischen Topologie haben wir im Anhang B zusammengefasst. Wir gehen einfach aus von $n \geq 2$ paarweise verschiedenen Punkten in der Ebene:

$$Q = \{q_1, \ldots, q_n\} \subset \mathbb{R}^2. \tag{2.24}$$

Diese definieren einen eindimensionalen Unterraum $N$, gegeben durch:

$$N := \{z \in \mathbb{R}^2 | \exists q_i, q_j \in Q, q_i \neq q_j \text{ mit } d(z, q_i) = d(z, q_j) \leq d(z, Q)\}. \tag{2.25}$$

Wir werden $N$ als "neutrale Zone" bezeichnen, da dort bei der Anwendung auf Decodierung keine eindeutige Hardbit-Entscheidung getroffen werden kann. Das Komplement von $N$ zerfällt disjunkt in $n$ Zusammenhangskomponenten $Z_q$, wobei jeweils $q \in Z_q$ gilt:

$$\mathbb{R}^2 \setminus N = \sum_{q \in Q} Z_q \tag{2.26}$$

Diese teilen wir wiederum in zwei Klassen ein, die wir mit 1 und $-1$ indizieren, um an die entsprechenden Bits zu erinnern:

$$Q = I_1 \dot{\cup} I_{-1}, \quad I_1 \neq \emptyset \neq I_{-1}, \tag{2.27}$$

$$B_1 = \sum_{q \in I_1} Z_q, \tag{2.28}$$

$$B_{-1} = \sum_{q \in I_{-1}} Z_q, \tag{2.29}$$

$$\mathbb{R}^2 \setminus N = B_1 \dot{\cup} B_{-1}. \tag{2.30}$$

Desweiteren benötigen wir noch die folgende Abbildung:

$$H : \mathbb{R}^2 \setminus N \rightarrow \{\pm 1\}, \tag{2.31}$$

$$y \mapsto \begin{cases} 1, \text{ falls } y \in B_1, \\ -1, \text{ falls } y \in B_{-1}. \end{cases} \tag{2.32}$$

Das $H$ soll an Hardbit-Entscheidung erinnern. Wir werden in der Anwendung auf die Decodierung die Mengen $B_1$ und $B_{-1}$ so definieren, dass folgendes gilt: ist $y$ eine Realisierung des Kanals, so ist $H(y)$ das entsprechende Bit des zu $y$ nächstgelegenen Signalpunktes. Schließlich betrachten wir noch die folgenden "minimalen Abstände" zur neutralen Zone der Zusammenhangskomponenten:

$$r_1 := \min_{q \in Q \cap B_1} \{d(q, N)\}, \tag{2.33}$$

$$r_{-1} := \min_{q \in Q \cap B_{-1}} \{d(q, N)\} \tag{2.34}$$

Zur besseren Orientierung sind alle bisher eingeführten Größen in Abbildung 2.1 beispielhaft dargestellt. Wir führen jetzt in folgender Weise eine Äquivalenzrela-

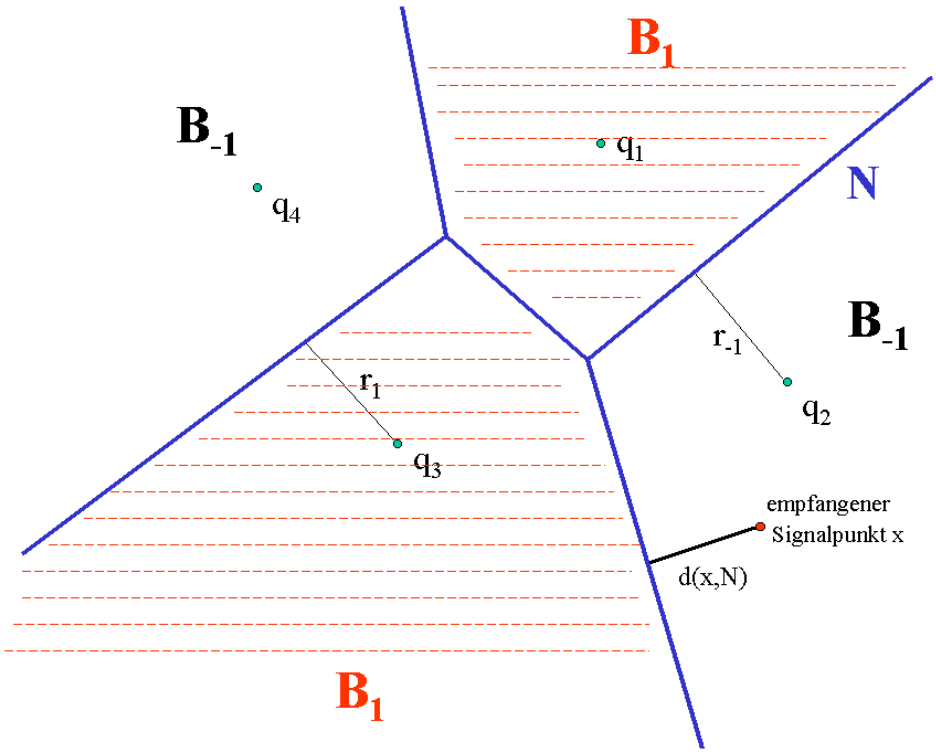


Abbildung 2.1: Beispiel mit $Q = \{q_1, q_2, q_3, q_4\}$

tion auf dem $\mathbb{R}^2$ ein. Für zwei Punkte $x, y \in \mathbb{R}^2$ gelte:

$$x \sim y :\Leftrightarrow d(x,N) = d(y,N) \text{ und} \tag{2.35}$$
$$(x,y \in N \text{ oder} \tag{2.36}$$
$$x,y \in B_1 \text{ oder} \tag{2.37}$$
$$x,y \in B_{-1}). \tag{2.38}$$

Betrachten wir den zugehörigen Quotienraum $T$ (s. Anhang B),

$$T := \mathbb{R}^2/\sim, \tag{2.39}$$

so ist er topologisch äquivalent zur reellen Achse, dargestellt in Abbildung 2.2. Der eingezeichnete Punkt $p$ beispielsweise repräsentiert die Äquivalenzklasse aller Punkte aus $B_1$ mit Abstand $a$ zur neutralen Zone.

$$[p] = \{z \in \mathbb{R}^2 : z \in B_1 \wedge d(z,N) = a\}. \tag{2.40}$$

Jetzt ist aber klar, weshalb wir diesen Quotientenraum eingeführt haben. Er versetzt uns genau in die Situation, die wir von der BPSK-Modulation her kennen. Es gibt einen neutralen Ursprung, und zwei Klassen von Punkten auf der

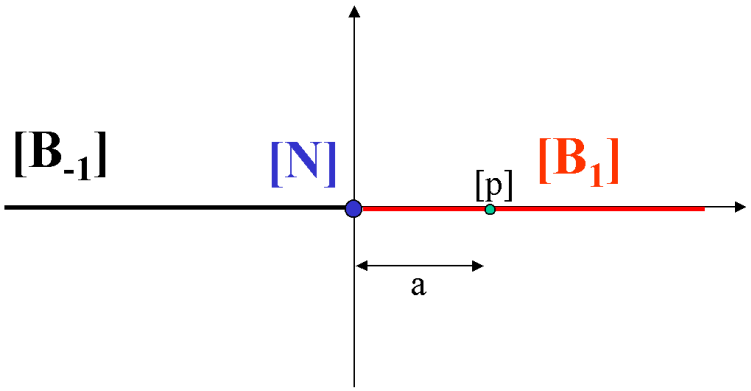


Abbildung 2.2: Der Quotientenraum $T$

rechten und linken reellen Halbachse, deren Abstand zum neutralen Punkt als Sicherheitsmaß interpretiert werden kann. Uns bleibt lediglich noch die Aufgabe, zu einem gegebenen $p \in \mathbb{R}^2$ den zugehörigen Wert $a$ im Quotientenraum $T$ zu bestimmen. Dazu müssen wir lediglich die kanonische Projektion

$$pr : \mathbb{R}^2 \rightarrow T, \tag{2.41}$$

$$y \mapsto [y] \tag{2.42}$$

mit einem offensichtlichen Homöomorphismus

$$\overline{\phi} : T \rightarrow \mathbb{R} \tag{2.43}$$

komponieren und erhalten so die Abbildung:

$$\phi(Q, B_1, B_{-1}) = \overline{\phi} \circ pr : \mathbb{R}^2 \rightarrow \mathbb{R}, \tag{2.44}$$

$$x \mapsto \begin{cases} 0, \text{ falls } x \in N, \\ H(x)\frac{d(x,N)}{r_{H(x)}}, \text{ falls } x \in \mathbb{R}^2 \setminus N. \end{cases} \tag{2.45}$$

Bei der Decodierung interpretieren wir den Quotienten $H(x)\frac{d(x,N)}{r_{H(x)}}$ als Softwert. Er besteht aus der Vorzeichen-Information $(H(x))$ und dem entscheidenden Abstand

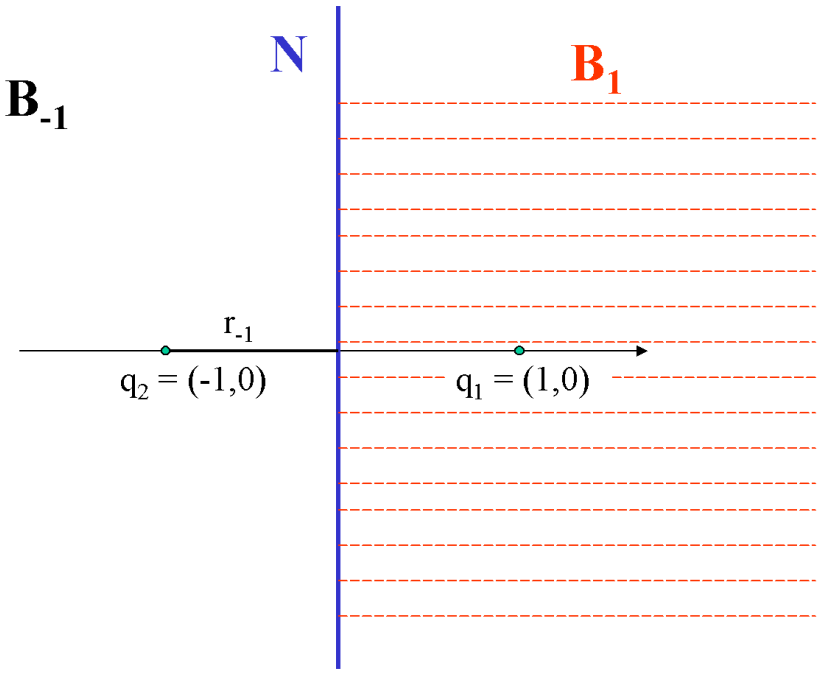


Abbildung 2.3: Die Topologie der BPSK-Modulation

zur neutralen Zone ($d(x, N)$), der durch $r_{H(x)}$ normiert wird (vgl. Abbildung 2.1). Fassen wir unsere bisherigen Überlegungen an dieser Stelle nochmal zusammen. Wir sind ausgegangen von einer ganz allgemeinen zwei-dimensionalen Situation, wie sie an einem Beispiel in Abbildung 2.1 gezeigt wird. Durch die Projektion auf den Quotientenraum $T$ haben wir eine Abbildung bestimmt, die jedem $x \in \mathbb{R}^2$ einen Wert $a = \phi(x) \in \mathbb{R}$ zuordnet, den wir ganz analog zur BPSK-Modulation als Sicherheitswert, nämlich als sein Abstand zur neutralen Zone zusammen mit einer Vorzeichen-Information interpretieren können. In der Anwendung auf QAM-modulierte Signale bleibt uns lediglich, die Mengen $Q$, $B_1$ und $B_{-1}$ zu bestimmen. Diese werden sich in völlig natürlicher Weise ergeben. Wir wollen dies schon einmal zur Veranschaulichung am Beispiel BPSK veranschaulichen (siehe Abbildung 2.3), wobei die Modulatorfunktion in den $\mathbb{R}^2$ abbildet, um die Notationen zu vereinfachen. Betrachten wir also die BPSK-Modulatorfunktion

$$\rho_{BPSK} : \{\pm 1\} \rightarrow \mathbb{R}^2, \tag{2.46}$$

$$j \mapsto (j, 0), \tag{2.47}$$

so definieren wir

$$\begin{aligned} Q &:= \rho_{BPSK}(\{\pm 1\}) = \{(-1,0),(1,0)\}, && (2.48)\\ B_1 &:= \{(x_1,x_2) \in \mathbb{R}^2 : x_1 > 0\}, && (2.49)\\ B_{-1} &:= \{(x_1,x_2) \in \mathbb{R}^2 : x_1 < 0\}, && (2.50)\\ & && (2.51) \end{aligned}$$

und es folgt:

$$\begin{aligned} N &= \{(x_1,x_2) \in \mathbb{R}^2 : x_1 = 0\}, && (2.52)\\ r_1 &= 1, && (2.53)\\ r_{-1} &= -1, && (2.54)\\ H((x_1,x_2)) &= \text{signum}(x_1). && (2.55) \end{aligned}$$

Die Abbildung $\phi$ ergibt sich daher als:

$$\begin{aligned} \phi : \mathbb{R}^2 &\rightarrow \mathbb{R}, && (2.56)\\ (x_1,x_2) &\mapsto x_1. && (2.57) \end{aligned}$$

An diesem einfachen Beispiel läßt sich schon die Funktion von $\phi$ erkennen. Da $\phi$ über die Äquivalenzklassen faktorisiert, bildet es Punkte gleicher Sicherheit, in diesem Fall bedeutet dies mit gleichem Abstand zur y-Achse, auf einen Punkt ab, und berücksichtigt dabei den Sicherheitswert und das Vorzeichen, also die Hartbit-Entscheidung. Genauso werden wir im nächsten Abschitt für die QAM-modulierten Signale vorgehen.

## 2.4 Soft-Decodierung QAM-modulierter Signale

In diesem Abschnitt wollen wir die bitweise Softdecodierung für QAM-modulierte Signale entwickeln. Dabei gehen wir zunächst vollkommen analog zur BPSK-Modulation vor, verfolgen also die in Abschnitt 2.1 vorgestellte Methode der L-Wert-Decodierung. Wie wir sehen werden, stellt sich bei höherdimensionalen Modulationen noch stärker als bei der eindimensionalen BPSK-Modulation das Problem der Komplexität der L-Werte. Daher werden wir eine Methode vorstellen, die zu einer Reduzierung der Komplexität von QAM-modulierten Signalen auf BPSK-modulierte Signale führt. Dementsprechend werden wir uns zur Entwicklung dieser Methode schon am BPSK-Fall orientieren.

Wir betrachten den Fall eines $m^2$-QAM-modulierten Signals mit dem Kombinationscode $\mathcal{C}$ aus Beispiel 1.2.4 und dem dazugehörigen stochastischen Kanalmodell

(1.74)

$$\begin{aligned} \mathcal{K}^{m^2}: &\quad \mathcal{C}\times\Omega\to\mathbb{R}^{2\times n} \\ \mathcal{K}_c^{m^2,j}: &\quad \text{ist } \mathcal{N}(\rho_{m^2}\left[(\mathbf{c}_{\cdot,j})\right],\mathbf{I}_2\frac{N_0 n}{2E_b k\cdot 2}) \text{ normalverteilt, } 1\le j\le n. \end{aligned} \tag{2.58}$$

Für die Dichten der Zufallsvariablen $\mathcal{K}_c^{m^2,j}$ gilt folglich:

$$\begin{aligned} d_{\mathbf{c}}^{m^2}:\mathbb{R}^2 &\to \mathbb{R}, \\ \mathbf{y} &\mapsto \frac{1}{2\pi\frac{N_0 n}{2E_b k}}\prod_{j=1}^{n}\exp\left(-\frac{\|\rho_{m^2}(\mathbf{c}_{\cdot,j})-\mathbf{y}_{\cdot,j}\|^2}{\frac{N_0 n}{2E_b k}}\right). \end{aligned} \tag{2.59}$$

Damit ergibt sich als L-Wert durch Einsetzen in Gleichung (2.14):

$$L^{m^2}(\mathbf{u}_{li}|\mathbf{y}) = \ln\left[\frac{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C}\\ \mathbf{c}_{li}=+1}}\prod\limits_{j=1}^{n}\exp\left(-\frac{\|\rho_{m^2}(\mathbf{c}_{\cdot,j})-\mathbf{y}_{\cdot,j}\|^2}{\frac{N_0 n}{2E_b k}}\right)}{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C}\\ \mathbf{c}_{li}=-1}}\prod\limits_{j=1}^{n}\exp\left(-\frac{\|\rho_{m^2}(\mathbf{c}_{\cdot,j})-\mathbf{y}_{\cdot,j}\|^2}{\frac{N_0 n}{2E_b k}}\right)}\right], \quad \begin{matrix} l=1,\dots,\mu, \\ i=1,\dots,k_l \end{matrix}. \tag{2.60}$$

Entsprechend lautet unsere Decodiervorschrift für QAM-modulierte Signale gemäß (2.15):

$$\delta_{BO}^{m^2}:\mathbb{R}^{2\times n} \to \prod_{l=1}^{\mu}\{\pm 1\}^{k_l}, \tag{2.61}$$

$$\mathbf{y} \mapsto \text{signum}\left(\ln\left[\frac{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C}\\ \mathbf{c}_{li}=+1}}\prod\limits_{j=1}^{n}\exp\left(-\frac{\|\rho_{m^2}(\mathbf{c}_{\cdot,j})-\mathbf{y}_{\cdot,j}\|^2}{\frac{N_0 n}{2E_b k}}\right)}{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C}\\ \mathbf{c}_{li}=-1}}\prod\limits_{j=1}^{n}\exp\left(-\frac{\|\rho_{m^2}(\mathbf{c}_{\cdot,j})-\mathbf{y}_{\cdot,j}\|^2}{\frac{N_0 n}{2E_b k}}\right)}\right]\right)_{\substack{l=1,\dots,\mu,\\ i=1,\dots,k_l}} \tag{2.62}$$

Im Prinzip könnte man hier vom theoretischen Standpunkt aus die bitweise Softdecodierung für QAM-modulierte Signale für beendet erklären. Es gibt jedoch mindestens zwei Gründe, dies nicht zu tun. Zum einen gilt, dass die Norm $\|\mathbf{y}_{\cdot,j}\|$ in keiner Weise als Zuverlässigkeit interpretiert werden kann. Dies ist ein Unterschied zur BPSK-modulierten Situation, bei der einfachheitshalber uncodiert betrachtet gemäß (2.21) gilt:

$$L^{BPSK}(\mathbf{u}_j|\mathbf{y}_j) = \ln\left[\frac{\exp\left(-\frac{\|\mathbf{y}_j-1\|^2}{\frac{N_0 n}{E_b k}}\right)}{\exp\left(-\frac{\|\mathbf{y}+1\|^2}{\frac{N_0 n}{E_b k}}\right)}\right] \sim \text{signum}(\mathbf{y}_j)|\mathbf{y}_j|. \tag{2.63}$$

Gewichtiger ist die Tatsache, dass zur Berechnung der L-Werte (2.60) $\prod_{l=1}^{\mu} 2^{k_l}$ Additionen erforderlich sind. Wir haben bereits erwähnt, dass es oft schon schwierig ist, $2^k$ Operationen, die für einen $(n,k)$-Code notwendig sind, auszuführen. Daher ist es im allgemeinen erst recht unmöglich, $\prod_{l=1}^{\mu} 2^{k_l}$ Operationen zu bewältigen.

Unser Ziel ist es daher im folgenden, die empfangenen Softwerte $\mathbf{y} \in \mathbb{R}^{2\times n}$ so in bitweise Softinformation zu transformieren, dass sie für die zeilenweise Decodierung der einzelnen im Kombinationscode vorkommenden Codes geeignet sind. Dabei werden wir uns grundlegend von den topologischen Gedanken des vorangegangenen Abschnitts leiten lassen, den QAM-modulierten Signalraum topologisch auf eine Situation abzubilden, die zum BPSK-modulierten Signalraum analog ist. Gelingt uns dies, so können wir zeilenweise L-Werte berechnen und haben somit die Komplexität auf die eines einzelnen Codes reduziert. Die zeilenweise Decodierung erbringt dabei einen weiteren Vorteil. Stehen in den Zeilen des Kombinationscodes bestimmte, einfach strukturierte Codes oder sogar uncodierte Zeilen, so kann ihre Decodierung erheblich schneller erfolgen, da die Berechnung der zugehörigen L-Werte sich nochmals vereinfachen läßt. Dies kann man jedoch nicht ausnutzen, wenn man über die gesamte Code-Matrix decodiert.

Bei der BPSK-Modulation entspricht ein Symbol gerade einem Bit. Da wir in natürlicher Weise Softinformation für jedes Symbol erhalten, hat man bei BPSK-modulierten Signalen sofort einen Softwert pro Bit zur Verfügung. Die auftretenden Schwierigkeiten bei QAM-modulierten Signalen kommen also gerade daher, dass wir den Kanal symbolweise benutzen, also auch nur einen Softwert pro Symbol erhalten, gleichzeitig aber bitweise decodieren wollen. Daher macht es Sinn, die Vorstufe, die wir jetzt zur Gewinnung bitweiser Softwerte vor den Decodierer einbauen, Desymbolisator zu nennen, wie in Abbildung 2.4 dargestellt. Formal suchen wir also eine Abbildung $D$, den

$$\text{Desymbolisator } D : \mathbb{R}^{2\times n} \;\rightarrow\; \mathbb{R}^n, \tag{2.64}$$

$$\mathbf{y} \mapsto D(\mathbf{y}), \tag{2.65}$$

um mit $D(\mathbf{y})$ zeilenweise den Kombinationscode (1.2.4) zu decodieren. Genauer gesagt benötigen wir $\mu$ solcher Funktionen,

$$D_l : \mathbb{R}^{2\times n} \;\rightarrow\; \mathbb{R}^n, \quad l = 1, \ldots, \mu \tag{2.66}$$

$$\tag{2.67}$$

da wir jede der $\mu$ Zeilen einzeln decodieren wollen. Wir führen diese Abbildungen zunächst formal ein und erläutern sie anschließend. Da wir von oben zeilenweise decodieren, stehen uns im $l$-ten Decodierschritt bereits die decodierten $(l-1)$ ersten Zeilen zur Verfügung und wir werden uns auf diejenigen Punkte bei der Decodierung beschränken, die mit der bereits decodierten Information übereinstimmen. Wir bezeichnen dazu mit $\hat{\mathbf{c}} \in \mathbb{R}^{m\times n}$ das aus der Decodierung entstehende Codewort, also das Bild $\hat{\mathbf{c}} = E_{Komb}(\hat{\mathbf{u}})$ der decodierten Information $\hat{\mathbf{u}}$. Im

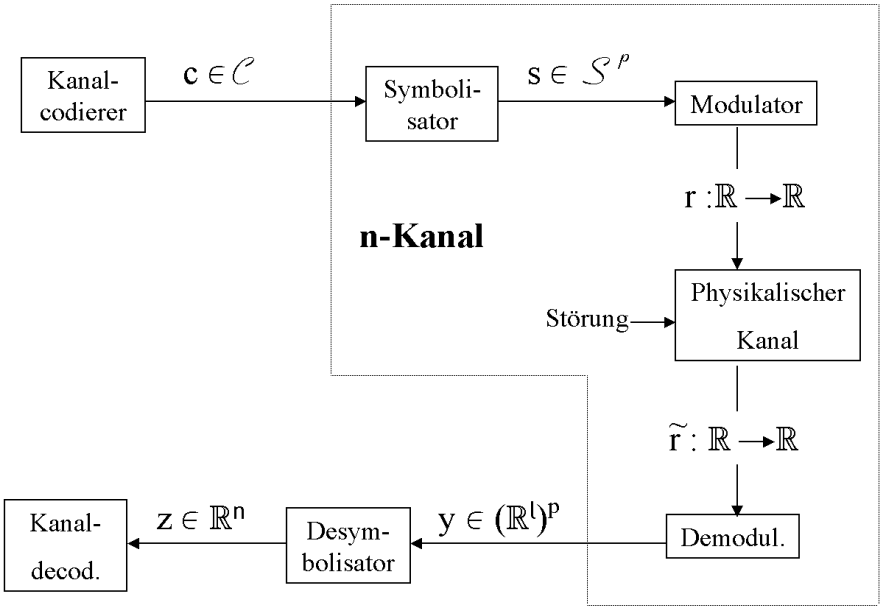


Abbildung 2.4: Kanal-Modell

$l$-ten Decodierschritt beschränken wir uns somit an der Stelle $j, 1 \leq j \leq n$, auf die Menge

$$Q^{lj} := P_{\hat{c}_{1j},\ldots,\hat{c}_{(l-1)j}}. \tag{2.68}$$

Das Komplement der zugehörigen neutralen Zone $N^{lj}$ zerfällt in Zusammenhangskomponenten $Z_p, p \in Q^{lj}$, die wir in ganz natürlicher Weise gemäß des harten Bits aufteilen:

$$B_1^{lj} := \bigcup_{\substack{p \in Q^{lj} \\ p^b(l)=1}} Z_p, \tag{2.69}$$

$$B_{-1}^{lj} := \bigcup_{\substack{p \in Q^{lj} \\ p^b(l)=-1}} Z_p. \tag{2.70}$$

Damit haben wir aber bereits alle Informationen zusammen, die wir für die Definition der Abbildung $\phi$ aus (2.44) gemäß unseren topologischen Überlegungen benötigen. Wir haben bereits im Abschnitt 2.3 erläutert, dass $\phi$ jedem Punkt aus der Ebene einen Sicherheitswert in $\mathbb{R}$ analog zur BPSK-Modulation zuordnet.

Daher definieren wir jetzt unseren Desymbolisator durch:

$$D_l : \mathbb{R}^{2\times n} \rightarrow \mathbb{R}^n, \tag{2.71}$$

$$\mathbf{y} \mapsto \mathbf{z}_{(l)} := D_l(\mathbf{y}) = (\phi(Q^{lj}, B_1^{lj}, B_{-1}^{lj})(\mathbf{y}_{\cdot,j}))_{1\le j\le n}. \tag{2.72}$$

Um die Abbildung weiter zu konkretisieren, erinnern wir noch einmal an die Hardbit-Entscheidungsfunktion:

$$H^{lj} : \mathbb{R}^2 \setminus N^{lj} \rightarrow \{\pm 1\}, \tag{2.73}$$

$$y \mapsto \begin{cases} 1, \text{ falls } y \in B_1^{lj}, \\ -1, \text{ falls } y \in B_{-1}^{lj}, \end{cases} \tag{2.74}$$

und die Minimalabstände

$$r_l^{m^2} = r_1^{lj} = \min_{p\in Q^{lj}\cap B_1^{lj}} \{d(p, N^{lj})\}, \tag{2.75}$$

$$= r_{-1}^{lj} = \min_{p\in Q^{lj}\cap B_{-1}^{lj}} \{d(p, N^{lj})\} \tag{2.76}$$

$$= \frac{(\sqrt{2})^{l-1}}{2(m-1)}, \tag{2.77}$$

die hier unabhängig von $j$ und der Hardbit-Entscheidung sind. Damit läßt sich $D_l$ gemäß (2.44) konkret angeben:

$$D_l : \mathbb{R}^{2\times n} \rightarrow \mathbb{R}^n, \tag{2.78}$$

$$\mathbf{y} \mapsto \mathbf{z}_{(l)} = D_l(\mathbf{y}) = \left( H^{lj}(\mathbf{y}_{\cdot,j}) \frac{d(N^{lj}, \mathbf{y}_{\cdot,j})}{r_l^{m^2}} \right)_{1\le j\le n}.$$

Wir wollen diese auf den ersten Blick vielleicht etwas abstrakte Definition nochmal erläutern. Nehmen wir an, wir wollten die oberste Bit-Ebene decodieren. Dann lassen sich zunächst 2 Beobachtungen machen. Zum einen gibt es Punkte in der Ebene, die keinerlei Entscheidungssicherheit bieten. Dies ist im BPSK-Fall der Nullpunkt, der genau zwischen den beiden Möglichkeiten $+1$ und $-1$ liegt. In allgemeinen höherdimensionalen Signalräumen handelt es sich genau um die von uns definierte neutrale Zone $N$. Sie ist für das Beispiel 16-QAM in Abbildung 2.5 blau dargestellt. Zum anderen ist es für die Entscheidungssicherheit, ob das erste Bit eine $+1$ oder $-1$ ist, irrelevant, in welcher Zusammenhangskomponente von $\mathbb{R}^2 \setminus N$ sich der Messpunkt befindet. Wir werden die Sicherheit solange als gleich ansehen, wie der orthogonale Abstand $d(x, N)$ zur neutralen Zone $N$ gleich ist. Wir haben in Abbildung 2.5 zwei rote Quadrate markiert, die bezüglich der Entscheidung "oberstes Bit $= -1$" die gleiche Sicherheit bieten. Daher ist es naheliegend, Bereiche gleicher Sicherheit zu identifizieren. Dies geschieht durch die Äquivalenzrelation (2.35). Sie identifiziert gerade Punkte gleicher Sicherheit, deren Hardbit-Entscheidung das gleiche Vorzeichen hat. Durch

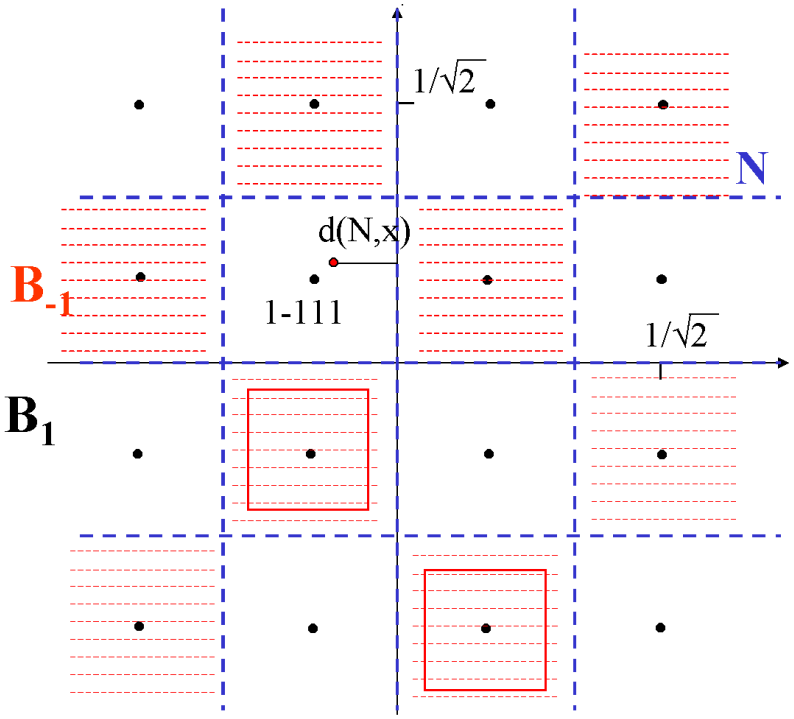


Abbildung 2.5: Beispiel 16-QAM, 1. Ebene

Betrachtung des Quotientenraums $T$ wie in Abbildung 2.2 kommen wir auf die Analogie zu BPSK-modulierten Signalen. Wir haben einen neutralen Nullpunkt, und nach rechts bzw. links mit größer werdendem Abstand zunehmende Sicherheit. Dementsprechend erhalten wir einen Sicherheitswert für $x \in \mathbb{R}^2$, indem wir das Bild von $x$ unter der kanonischen Projektion auf $T$ angucken, und genauso ist unser Desymbolisator entstanden. Die Werte $\mathbf{z}_{(l)} = D_l(\mathbf{y})$ können wir gemäß (2.21) zur L-Wert Decodierung verwenden:

$$L_{lj} := L_{lj}^{BPSK}(\mathbf{u}_{lj}|\mathbf{z}_{(l)}) = \ln \left[ \frac{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C}_l \\ \mathbf{c}_j=+1}} \exp\left(-\frac{\|\mathbf{z}_{(l)}-\mathbf{c}\|^2}{2\sigma(l)^2}\right)}{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C}_l \\ \mathbf{c}_j=-1}} \exp\left(-\frac{\|\mathbf{z}_{(l)}-\mathbf{c}\|^2}{2\sigma(l)^2}\right)} \right], \quad 1 \le j \le k_l. \tag{2.79}$$

Grundsätzlich wäre es möglich, die Softwerte (2.78) für $l = 1$ zu bestimmen und für die Decodierung aller Zeilen-Codes im gegebenen Kombinationscode zu verwenden. Jedoch kann man ein besseres Resultat erzielen, wenn man mit der ersten Zeile beginnt, die Softwerte bestimmt und decodiert. Für die nachfolgenden Zeilen werden dann nur noch diejenigen Punkte $Q^{lj} := P_{\hat{c}_{1j},\ldots,\hat{c}_{(l-1)j}}$ weiter

betrachtet, deren erste Bits mit den bereits decodierten übereinstimmen. Dazu passend verwenden wir die in (1.31) angegebene Modulatorfunktion $\rho^P_{m^2}$, die nach dem in Abbildung 1.6 dargestellten Prinzip der Partitionierung der Signalpunkte entstanden ist. Diese sorgt nämlich nun dafür, dass die noch übrig gebliebenen Punkte $Q^{lj}$ möglichst weit auseinander liegen. Durch dieses Vorgehen kann es auf den unterschiedlichen Bit-Ebenen zu verschiedenen harten Entscheidungen kommen und daher auch zu unterschiedlichen Softwerten, die Fehlerwahrscheinlichkeit wird nochmals verringert.

Fassen wir unser Vorgehen zusammen. Der Desymbolisator macht aus der Realisierung des QAM-Kanalmodells, $n$ Punkten $\mathbf{y}_{.,j} \in \mathbb{R}^2$ einen Vektor $\mathbf{z}_{(l)} \in \mathbb{R}^n$, der als Softwert in die L-Wert-Decodierung im BPSK-Schema eingeht. Dabei sind die Softwerte $D_l(\mathbf{y})$ für die verschiedenen Bitebenen unterschiedlich, da die bis zur $(l-1)$-ten Ebene erzielten Decodierergebnisse berücksichtigt werden. Zur Anwendung der L-Wert-Decodierung mit den L-Werten (2.79) fehlt uns noch eine Varianz. Diese erhalten wir, indem wir die Gesamtenergie $\frac{E_b k}{N_0 n}$ pro Kanalbenutzung, deren Wurzel man als Diagonale im QAM-Schema findet, ins Verhältnis setzt zu den gegebenen Radien $r_l^{m^2}$ auf der $l$-ten Ebene, die wiederum durch unsere kanonische Projektion auf 1 abgebildet werden. Mit anderen Worten, die Varianz $\sigma^2(l)$ ergibt sich folgendermaßen:

$$\frac{\sigma(l)}{\frac{E_b k}{N_0 n}^{-\frac{1}{2}}} = \frac{r_l^{m^2}}{1} \tag{2.80}$$

$$\Leftrightarrow \sigma^2(l) = (r_l^{m^2})^2 \cdot \frac{N_0 n}{E_b k}. \tag{2.81}$$

Mit $\mathbf{z}_{(l)} = D_l(\mathbf{y})$ und $\sigma^2(l)$ kann jetzt die L-Wert-Decodierung gemäß (2.79) stattfinden. Der prinzipielle Ablauf ist demnach der folgende. Gegeben ist die empfangene Matrix $\mathbf{y} \in \mathbb{R}^{2\times n}$. Mit $\mathbf{z}_{(1)} = D_1(\mathbf{y})$ und $\sigma^2(1)$ wird die erste Zeile des Kombinationscodes decodiert. Anschließend wird $\mathbf{z}_{(2)} = D_2(\mathbf{y})$ und $\sigma^2(2)$ bestimmt und die zweite Zeile decodiert etc. Wir geben im nächsten Abschnitt einen präzisen Algorithmus an und geben anschließend einige konkrete numerische Beispiele.

# Kapitel 3

# Algorithmus und numerische Ergebnisse

## 3.1 Ein Algorithmus zur Soft-Decodierung

In den vorherigen Abschnitten haben wir die Theorie soweit entwickelt, dass wir nun die einzelnen Bausteine zusammensetzen können. Dabei stellen wir den vollständigen Algorithmus zur Decodierung QAM-modulierter Signale über einem AWGN-Kanal vor. Bevor wir mit dem Algorithmus beginnen, stellen wir die Voraussetzungen, insbesondere zur Erinnerung an die Notationen, nochmal zusammen. Gegeben ist zunächst ein Kombinationscode

$$E_{Komb} : \prod_{l=1}^{\mu} \{\pm 1\}^{k_l} \hookrightarrow \{\pm 1\}^{\mu \times n} \tag{3.1}$$

mit Coderaum $\mathcal{C}$ und Anzahl $k = \sum_{l=1}^{\mu} k_l$ Infobits. Die Coderäume der einzelnen Zeilen bezeichnen wir naheliegenderweise mit $\mathcal{C}_l$, $l = 1, \ldots, \mu$. Weiterhin betrachten wir eine $m^2$-QAM-Modulatorfunktion mit $m^2 = 2^\mu$

$$\rho^P_{m^2} : \mathcal{S}_\mu = \{\pm 1\}^\mu \to \mathbb{R}^2 \tag{3.2}$$

mit der in Abschnitt 1.3 vorgestellten Partitionierung der Signalpunkte nach Abbildung 1.6. Wir erinnern nochmal an die Notationen

$$\begin{aligned} P &:= \rho^P_m(\mathcal{S}_\mu), && (3.3)\\ P &= \{p_1, \ldots, p_{m^2}\}, && (3.4)\\ P_{b_1 \ldots b_f} &= \{p_i \in P | p_i^b \text{ stimmt an den ersten } f \text{ Stellen mit } b_1 \ldots b_f \text{ überein}\}, \\ r_l^{m^2} &= \frac{\sqrt{2}^{l-1}}{2(m-1)}. && (3.5) \end{aligned}$$

Schließlich haben wir noch das Kanalmodell

$$\begin{aligned} \mathcal{K}^{m^2}: &\quad \mathcal{C}\times\Omega\to\mathbb{R}^{2\times n} \\ \mathcal{K}_c^{m^2,j}: &\quad \text{ist } \mathcal{N}(\rho_{m^2}\left[(\mathbf{c}_{.,j})\right],\mathbf{I}_2\frac{N_0 n}{2E_b k\cdot 2}) \text{ normalverteilt, } 1\le j\le n. \end{aligned} \tag{3.6}$$

Wie immer bezeichnen wir mit

$$\mathbf{y}\in\mathbb{R}^{2\times n} \tag{3.7}$$

eine gegebene Realisierung des stochastischen Kanals. Der Algorithmus zur Decodierung hat nun folgende Gestalt:

**Decodier-Algorithmus:** $\mathbf{y}\to\hat{\mathbf{u}}\in\prod_{l=1}^{\mu}\{\pm1\}^{k_l},\ L_{lj},\ 1\le l\le\mu, 1\le j\le k_l$

(i) Initialisierung $l=1$.

(ii) Bestimme für $1\le j\le k_j$ die Mengen:

$$Q^{lj} = P_{\hat{c}_{1j},\ldots,\hat{c}_{(l-1)j}}, \tag{3.8}$$

$$\mathbb{R}^2\setminus N^{lj} = \sum_{p\in Q^{lj}} Z_p, \tag{3.9}$$

$$B_1^{lj} = \bigcup_{\substack{p\in Q^{lj}\\ p^b(l)=1}} Z_p, \tag{3.10}$$

$$B_{-1}^{lj} = \bigcup_{\substack{p\in Q^{lj}\\ p^b(l)=-1}} Z_p, \tag{3.11}$$

und damit die Abbildung:

$$H^{lj}:\mathbb{R}^2\setminus N^{lj} \to \{\pm1\}, \tag{3.12}$$

$$y \mapsto \begin{cases} 1, \text{ falls } y\in B_1^{lj}, \\ -1, \text{ falls } y\in B_{-1}^{lj}. \end{cases} \tag{3.13}$$

(iii) Bestimme Softwerte:

$$\mathbf{z}_{(l)} = D_l(\mathbf{y}) = \left(H^{lj}(\mathbf{y}_{.,j})\frac{d(N^{lj},\mathbf{y}_{.,j})}{r_l^{m^2}}\right)_{1\le j\le n}. \tag{3.14}$$

(iv) Bestimme Varianz:

$$\sigma(l)^2 = (r_l^{m^2})^2\cdot\frac{N_0 n}{E_b k}. \tag{3.15}$$

(v) Bestimme (BPSK)-L-Werte:

$$L_{lj} := L_{lj}^{BPSK}(\mathbf{u}_{lj}|\mathbf{z}_{(l)}) = \ln \left[ \frac{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C}_l \\ \mathbf{c}_j=+1}} \exp\left(-\frac{\|\mathbf{z}_{(l)}-\mathbf{c}\|^2}{2\sigma(l)^2}\right)}{\sum\limits_{\substack{\mathbf{c}\in\mathcal{C}_l \\ \mathbf{c}_j=-1}} \exp\left(-\frac{\|\mathbf{z}_{(l)}-\mathbf{c}\|^2}{2\sigma(l)^2}\right)} \right], \quad 1 \le j \le k_l. \tag{3.16}$$

(vi)

$$\begin{aligned} \text{Setze } \hat{\mathbf{u}}_{lj} &:= \text{signum}(L_{lj}), \quad 1 \le j \le k_l, \\ \text{und } \hat{\mathbf{c}}_{l,\cdot} &:= E_l(\hat{\mathbf{u}}_{l,\cdot}). \end{aligned}$$

(vii) $l := l + 1$

(viii) Falls $l \le \mu$, gehe zu Punkt (ii).

(ix) Ausgabe: $\hat{\mathbf{u}} \in \prod\limits_{l=1}^{\mu} \{\pm 1\}^{k_l},\ L_{lj}, \quad 1 \le l \le \mu, 1 \le j \le k_l.$

Wir wollen im folgenden die einzelnen Schritte des Algorithmus etwas ausführlicher kommentieren.

(i) Initialisierung $l = 1$.
Die Variable $l$ durchläuft die $\mu$ Zeilen des Kombinationscodes. Wir beginnen mit der Decodierung der ersten Zeile.

(ii) Bestimme für $1 \le j \le k_j$ die Mengen:

$$Q^{lj} = P_{\hat{c}_{1j},\ldots,\hat{c}_{(l-1)j}}, \tag{3.17}$$

$$\mathbb{R}^2 \setminus N^{lj} = \sum_{p\in Q^{lj}} Z_p, \tag{3.18}$$

$$B_1^{lj} = \bigcup_{\substack{p\in Q^{lj} \\ p^b(l)=1}} Z_p, \tag{3.19}$$

$$B_{-1}^{lj} = \bigcup_{\substack{p\in Q^{lj} \\ p^b(l)=-1}} Z_p, \tag{3.20}$$

und damit die Abbildung:

$$H^{lj} : \mathbb{R}^2 \setminus N^{lj} \to \{\pm 1\}, \tag{3.21}$$

$$x \mapsto \begin{cases} 1, \text{ falls } x \in B_1^{lj}, \\ -1, \text{ falls } x \in B_{-1}^{lj}. \end{cases} \tag{3.22}$$

Die Menge $Q^{lj}$ enthält die Punkte, deren erste $(l-1)$ Bits mit den an der $j$-ten Stelle decodierten Bits übereinstimmen. Diese Punkte definieren ihre neutrale Zone $N^{lj}$. Die Mengen $B_1^{lj}$ bzw. $B_{-1}^{lj}$ sind so gebildet, dass diejenigen Punkte, die einem Punkt mit $l$-tem Bit gleich 1 bzw. $-1$ am nächsten liegen, in $B_1^{lj}$ bzw. $B_{-1}^{lj}$ liegen. Die Funktion $H^{lj}$ ordnet jedem Punkt diese Hardbit-Entscheidung zu, also 1 oder $-1$, je nachdem, ob der nächste Punkt in $Q^{lj}$ als $l$-tes Bit eine 1 oder $-1$ hat.

(iii) Bestimme Softwerte:

$$\mathbf{z}_{(l)} = D_l(\mathbf{y}) = \left( H^{lj}(\mathbf{y}_{.,j}) \frac{d(N^{lj}, \mathbf{y}_{.,j})}{r_l^{m^2}} \right)_{1 \le j \le n}. \tag{3.23}$$

Dies ist der entscheidende Desymbolisator-Schritt, der sich aus der topologischen Identifikation gleich sicherer Punkte ergibt. Im einzelnen besteht der Softwert $\mathbf{z}_{(l)}$ aus drei Teilen: aus dem Vorzeichen des $l$-ten Bits der Hardbit-Entscheidung $H^{lj}(\mathbf{y}_{.,j})$, aus dem orthogonalen Abstand zur relevanten neutralen Zone $d(N^{lj}, \mathbf{y}_{.,j})$ und aus der Normierung der Projektion durch den Faktor $(r_l^{m^2})^{-1}$.

(iv) Bestimme Varianz:

$$\sigma(l)^2 = (r_l^{m^2})^2 \cdot \frac{N_0 n}{E_b k}. \tag{3.24}$$

Die Varianz ergibt sich aus dem Kehrwert der Energie pro Kanalbenutzung $\frac{N_0 n}{E_b k}$, die anteilig entsprechend den Minimalradien $(r_l^{m^2})$ auf die $\mu$ Ebenen verteilt wird.

(v) Bestimme (BPSK)-L-Werte:

$$L_{lj} := L_{lj}^{BPSK}(\mathbf{u}_{lj}|\mathbf{z}_{(l)}) = \ln \left[ \frac{\sum\limits_{\substack{\mathbf{c} \in \mathcal{C}_l \\ \mathbf{c}_j = +1}} \exp\left( -\frac{\|\mathbf{z}_{(l)} - \mathbf{c}\|^2}{2\sigma(l)^2} \right)}{\sum\limits_{\substack{\mathbf{c} \in \mathcal{C}_l \\ \mathbf{c}_j = -1}} \exp\left( -\frac{\|\mathbf{z}_{(l)} - \mathbf{c}\|^2}{2\sigma(l)^2} \right)} \right], \quad 1 \le j \le k_l. \tag{3.25}$$

Die zuvor bestimmten Softwerte $\mathbf{z}_{(l)}$ für den $l$-ten Code $\mathcal{C}_l$ werden hier in die Formel für die BPSK L-Werte aus (2.21) eingesetzt. Dabei werden lediglich $2^{k_l}$ Additionen benötigt, im Vergleich zu $\prod_{l=1}^{m} 2^{k_l}$ für die direkte Bestimmung der L-Werte des Kombinationscodes nach (2.60). Wie bereits erwähnt besteht ein weiterer Vorteil darin, dass sich die L-Wert-Bestimmung bestimmter Codes erheblich vereinfachen läßt. Dies kann man aber erst nutzen, wenn man diese in den einzelnen Zeilen auftretenden Codes auch einzeln decodiert.

(vi)

$$\begin{aligned} \text{Setze } \hat{\mathbf{u}}_{lj} &:= \text{signum}(L_{lj}), \quad 1 \le j \le k_l, \\ \text{und } \hat{\mathbf{c}}_{l,\cdot} &:= E_l(\hat{\mathbf{u}}_{l,\cdot}). \end{aligned}$$

Das Vorzeichen des L-Werts $L_{lj}$ ergibt das Informationsbit $\mathbf{u}_{lj}$. Gleichzeitig ist der Absolutbetrag von $L_{lj}$ ein Sicherheitsmaß für die Entscheidung. Aus dem so gewonnenen Informationswort $\hat{\mathbf{u}}_{l,\cdot}$ entsteht umgekehrt wieder das Codewort $\hat{\mathbf{c}}_{l,\cdot} := E_l(\hat{\mathbf{u}}_{l,\cdot})$. Dies dient dazu, die Menge $P_{\hat{\mathbf{c}}_{1j}\ldots\hat{\mathbf{c}}_{(l-1)j}}$ festzulegen und somit die weiteren Decodierschritte auf diejenigen Signalpunkte zu reduzieren, die mit der bisherigen Decodierung übereinstimmen.

(vii) $l := l + 1$
Die Decodierung der $l$-ten Zeile ist abgeschlossen, also kann der Zeilen-Index $l$ um 1 erhöht werden.

(viii) Falls $l \le \mu$, gehe zu Punkt (ii).
Falls es noch weitere Zeilen im Kombinationscode gibt, werden diese nach dem gleichen Muster decodiert.

(ix) Ausgabe: $\hat{\mathbf{u}} \in \prod_{l=1}^{\mu} \{\pm 1\}^{k_l},\ L_{lj}, \quad 1 \le l \le \mu, 1 \le j \le k_l.$
Nach Beendigung der Decodierung kann die Ausgabe erfolgen. Diese besteht in der Angabe des vollständigen Codeworts $\mathbf{u}$ sowie eines Sicherheitsmaßes durch den Absolutbetrag von $L_{lj}$ für jedes einzelne Informationsbit.

## 3.2 Beispiele

In diesem Abschnitt zeigen wir die Ergebnisse des Decodier-Algorithmus aus Abschnitt 3.1 für verschiedene Kombinationscodes. Der AWGN-Kanal wurde dazu jeweils für verschiedene Signal-Rausch-Verhältnisse $\frac{E_b}{N_0}$ simuliert und gegen die Bitfehlerrate aufgetragen. Die Simulation wurde mit dem Programm Matlab [Mat] erstellt, mit jeweils $10^6$ Simulationen pro Energie-Niveau.

Wir erinnern an die allgemeine Form des von uns verwendeten Kombinationscodes. Sie lautet:

$$E_{Komb} : \prod_{l=1}^{\mu} \{\pm 1\}^{k_l} \hookrightarrow \{\pm 1\}^{\mu \times n} \tag{3.26}$$

$$\mathbf{u} = (\mathbf{u}_{l,\cdot})_{1 \le l \le \mu} \mapsto E_{Komb}(\mathbf{u}) = (E_l(\mathbf{u}_{l,\cdot}))_{1 \le l \le \mu}. \tag{3.27}$$

Wir haben verschiedene Codes getestet, die sich nicht nur in der Anzahl der Spalten unterscheiden, sondern auch in der Anzahl der Zeilen $\mu$. Für einen 4-zeiligen Code benötigen wir ein $m^2 = 2^4 = 16$-QAM-Kanalmodell, während wir

für 6-zeilige Codes bereits bei einem $2^6 = 64$-QAM-Kanalmodell ankommen. Wir bezeichnen die Codes auch, da keine Verwechslung möglich ist, nach der Anzahl ihrer Infobits als $(k_1, \ldots, k_\mu)$-Codes. Allen Beispielen ist folgendes gemeinsam: Das $\mu$-Tupel $(k_1, \ldots, k_\mu)$ ist monoton wachsend. Mit anderen Worten, je höher die Zeile, desto geringer ist die Anzahl Infobits des entsprechenden Codes. Dies ist abgestimmt auf die von uns verwendete Modulatorfunktion

$$\rho^P_{m^2} : \mathcal{S}_\mu \hookrightarrow \mathbb{R}^2, \tag{3.28}$$

an deren Konstruktionsprinzip gemäß Abbildung 1.6 wir nochmal erinnern. Mit jeder tieferen Zeile liegen die konkurrierenden Bits weiter auseinander, und damit nimmt die Notwendigkeit, sie durch Redundanzbits zu schützen, weiter ab.

Die nachfolgenden Abbildungen 3.1 bis 3.5 zeigen die Ergebnisse unserer Simulationen. Dabei ist auf der x-Achse jeweils das Signal-Rausch-Verhältnis $E_b/N_0$ in $dB$ aufgetragen. Die y-Achse zeigt die Bitfehlerrate an. Die Fehlerkurve unseres Algorithmus aus 3.1 ist dabei in den Legenden als "euklidisch soft" bezeichnet und blau dargestellt. Als Vergleichskurve dient in allen Beispielen ein Maximum-Likelihood-Hard Decision Decoder, der als "hard decision" bezeichnet wird und grün dargestellt ist. In den Abbildungen 3.1 und 3.3 ist jeweils eine weitere Vergleichskurve aufgeführt, die aus [Fr96], Abb. 7.1 und 7.2, entnommen ist. Diese binär modulierten BCH-Codes, die jeweils eine vergleichbare Coderate haben, sind rot dargestellt. Allgemein erwartet man bei höherer Modulation als Preis für die höhere spektrale Effizienz zusätzlich benötigte Signalenergie, um die gleiche Fehlerrate zu erreichen. So benötigt man nach [Fr96], Tabelle 10.2, z.B. bei 16-QAM-Modulation im Vergleich zu binärer Modulation asymptotisch eine zusätzliche Energie von etwa 10 dB. Der Vergleich der blauen mit der roten Fehlerkurve in den Abbildungen 3.1 und 3.3 zeigt jedoch, daß unser Algorithmus diesen Verlust an Übertragungsqualität nicht aufweist.

**Beispiel 3.2.1 ($(1,4,4,7)$-Code).** Für das erste Beispiel haben wir folgenden Code verwendet, den wir bereits im Beispiel 1.2.4 vorgestellt haben.

$$E^1 : \{\pm 1\}^1 \times \{\pm 1\}^4 \times \{\pm 1\}^4 \times \{\pm 1\}^7 \quad \hookrightarrow \quad \{\pm 1\}^{4\times 7} \tag{3.29}$$

$$\mathbf{u} \mapsto E^1(\mathbf{u}) \quad = \quad \begin{pmatrix} E_{R7}(\mathbf{u}_{1,\cdot}) \\ E_{Ham}(\mathbf{u}_{2,\cdot}) \\ E_{Ham}(\mathbf{u}_{3,\cdot}) \\ E_{uncod}(\mathbf{u}_{4,\cdot}) \end{pmatrix}. \tag{3.30}$$

Es handelt sich um einen vierzeiligen $(28, 16)$-Block-Code mit einer Coderate von $\frac{4}{7}$. Dementsprechend haben wir ein 16-QAM-Kanalmodell simuliert, jeweils $n = 7$ Kanalbenutzungen sind für die Übertragung eines vollständigen Codewortes nötig. Das Decodier-Ergebnis ist in Abbildung 3.1 dargestellt.

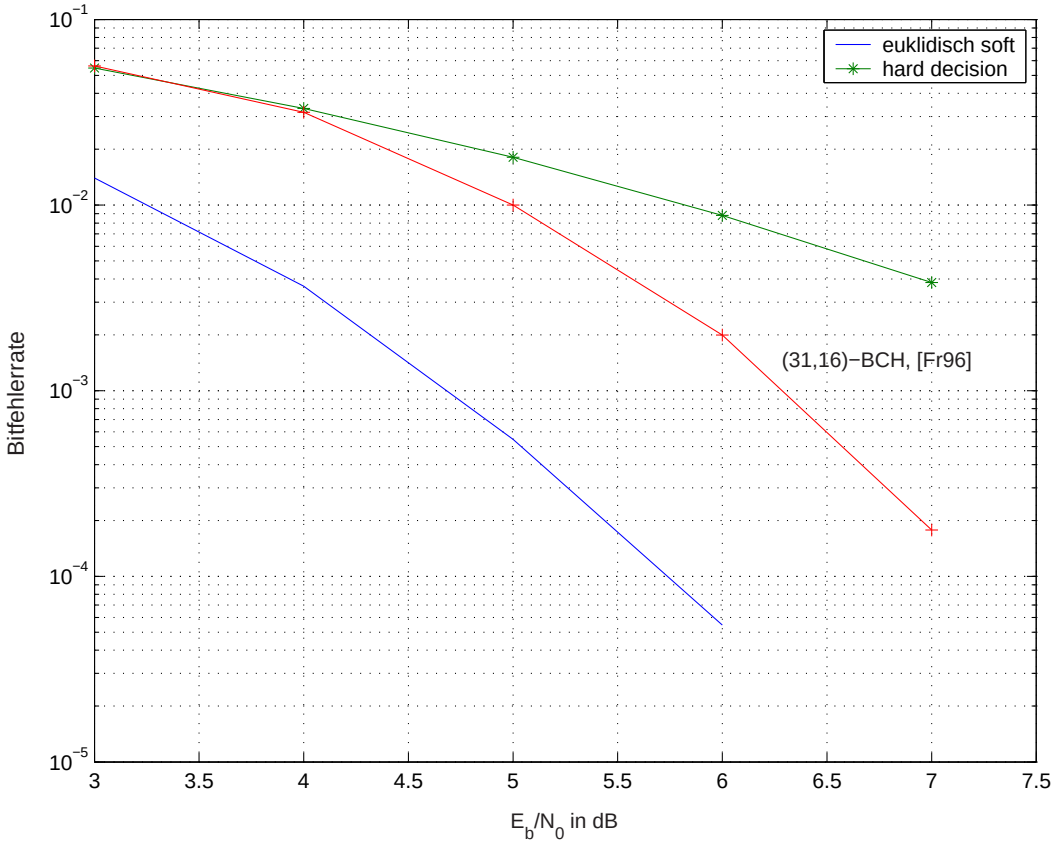


Abbildung 3.1: Beispiel 16-QAM, (1,4,4,7)-Kombinationscode

**Beispiel 3.2.2.** Wir wollen an dieser Stelle demonstrieren, wie wichtig es ist, dass die Code-Struktur mit höherer Zeile immer stärker geschützte Infobits hat. Dies ist nicht weiter überraschend, da auf den höheren Ebenen die Abstände zwischen den Signalpunkten viel geringer sind. Wir betrachten dazu einen Code, der auf allen vier Ebenen gleich ist:

$$E^{1'} : \{\pm 1\}^4 \times \{\pm 1\}^4 \times \{\pm 1\}^4 \times \{\pm 1\}^4 \quad \hookrightarrow \quad \{\pm 1\}^{4\times 7} \tag{3.31}$$

$$\mathbf{u} \mapsto E^{1'}(\mathbf{u}) \quad = \quad \begin{pmatrix} E_{Ham}(\mathbf{u}_{1,\cdot}) \\ E_{Ham}(\mathbf{u}_{2,\cdot}) \\ E_{Ham}(\mathbf{u}_{3,\cdot}) \\ E_{Ham}(\mathbf{u}_{4,\cdot}) \end{pmatrix}. \tag{3.32}$$

Dieser Code $E^{1'}$ ist ebenfalls ein $(28, 16)$-Block-Code mit einer Coderate von $\frac{4}{7}$ genau wie der Code $E^1$ aus obigem Beispiel. Das Decodier-Ergebnis ist in Abbildung 3.2 dargestellt. Der Vergleich der Abbildungen 3.1 und 3.2 zeigt deutlich, wie wichtig die richtige, an die Modulatorfunktion und die Decodiervorschrift angepasste Wahl der Code-Zeilen ist.

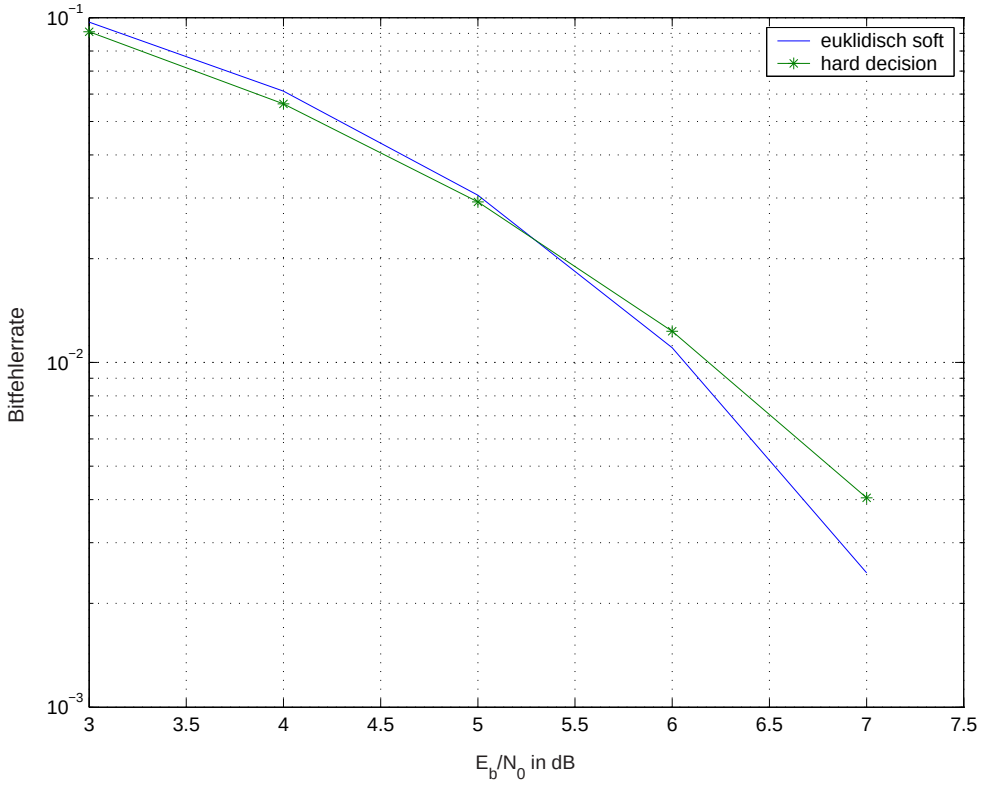


Abbildung 3.2: Beispiel 16-QAM, (4,4,4,4)-Kombinationscode

**Beispiel 3.2.3** ($(1,5,5,15)$**-Code).** Es folgt ein Beispiel mit $n = 15$ Spalten. Dazu bezeichnen wir mit

$$E_{15,5} : \{\pm 1\}^5 \hookrightarrow \{\pm 1\}^{15} \text{ den } (15,5)\text{-BCH-Code.} \tag{3.33}$$

Der zweite von uns getestete Kombinationscode hat die folgende Gestalt:

$$E^2 : \{\pm 1\}^1 \times \{\pm 1\}^5 \times \{\pm 1\}^5 \times \{\pm 1\}^{15} \hookrightarrow \{\pm 1\}^{4\times 15} \tag{3.34}$$

$$\mathbf{u} \mapsto E^2(\mathbf{u}) = \begin{pmatrix} E_{R15}(\mathbf{u}_{1,\cdot}) \\ E_{15,5}(\mathbf{u}_{2,\cdot}) \\ E_{15,5}(\mathbf{u}_{3,\cdot}) \\ E_{uncod}(\mathbf{u}_{4,\cdot}) \end{pmatrix}. \tag{3.35}$$

Es handelt sich um einen vierzeiligen $(60,26)$-Block-Code mit einer Coderate von $\frac{13}{30}$. Dementsprechend haben wir ein 16-QAM-Kanalmodell simuliert, jeweils $n = 15$ Kanalbenutzungen sind für die Übertragung eines vollständigen Codewortes nötig. Das Decodier-Ergebnis ist in Abbildung 3.3 dargestellt.

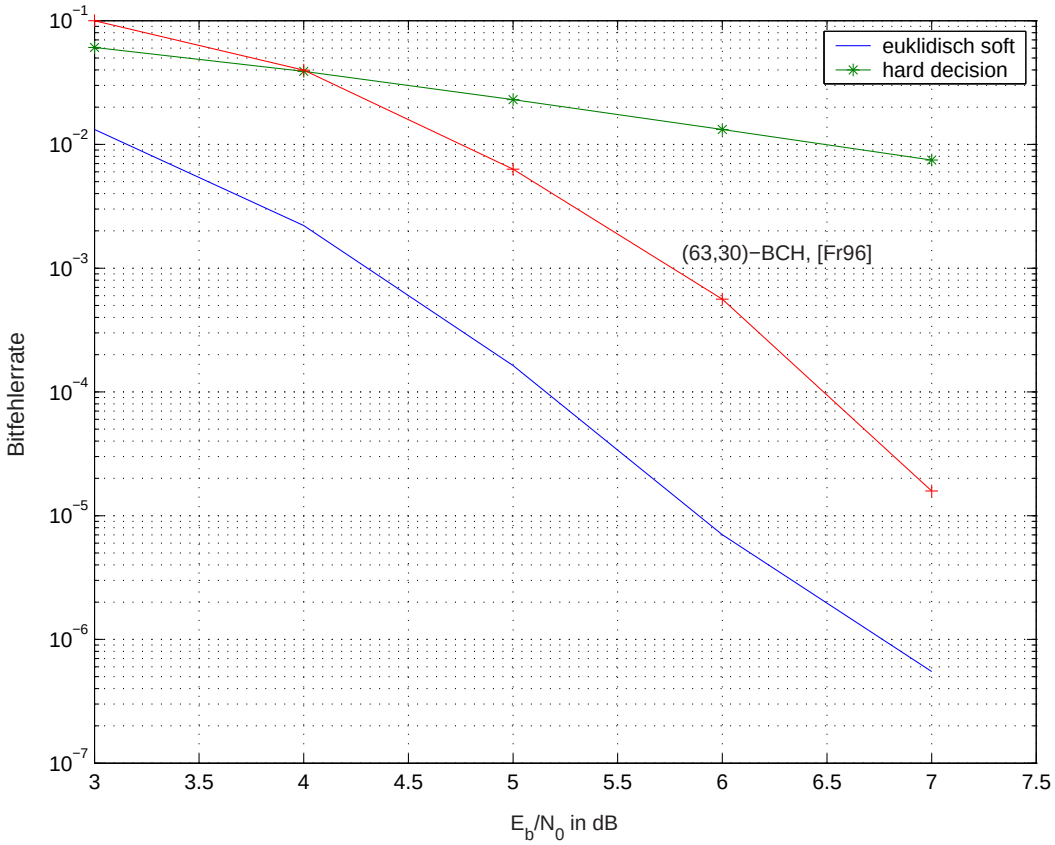


Abbildung 3.3: Beispiel 16-QAM, (1,5,5,15)-Kombinationscode

Die nächsten beiden Beispiele sind jeweils 6-zeilig, also für 64-QAM-modulierte Signale.

**Beispiel 3.2.4 ($(1,4,4,7,7,7)$-Code).** Wir haben die ersten beiden Codes für die 64-QAM-Beispiele jeweils um zwei uncodierte Zeilen erweitert. Dadurch verbessert sich die Coderate natürlich erheblich, aber wie zu erwarten, verringert sich der Abstand zur Vergleichskurve, da sich unsere Softdecodierung im wesentlichen bei codierten Zeilen als vorteilhaft erweist. Der Code hat nun folgende Gestalt:

$$\begin{aligned} E^3 : \{\pm 1\}^1 \times \{\pm 1\}^4 \quad \times \quad & \{\pm 1\}^4 \times \{\pm 1\}^7 \times \{\pm 1\}^7 \times \{\pm 1\}^7 \hookrightarrow \{\pm 1\}^{6\times 7} \\ \mathbf{u} \mapsto E^3(\mathbf{u}) \quad = \quad & \begin{pmatrix} E_{R7}(\mathbf{u}_{1,\cdot}) \\ E_{Ham}(\mathbf{u}_{2,\cdot}) \\ E_{Ham}(\mathbf{u}_{3,\cdot}) \\ E_{uncod}(\mathbf{u}_{4,\cdot}) \\ E_{uncod}(\mathbf{u}_{5,\cdot}) \\ E_{uncod}(\mathbf{u}_{6,\cdot}) \end{pmatrix}. \end{aligned} \tag{3.36}$$

Es handelt sich um einen 6-zeiligen $(42, 30)$-Block-Code mit einer Coderate von $\frac{5}{7}$. Das Decodier-Ergebnis ist in Abbildung 3.4 dargestellt.

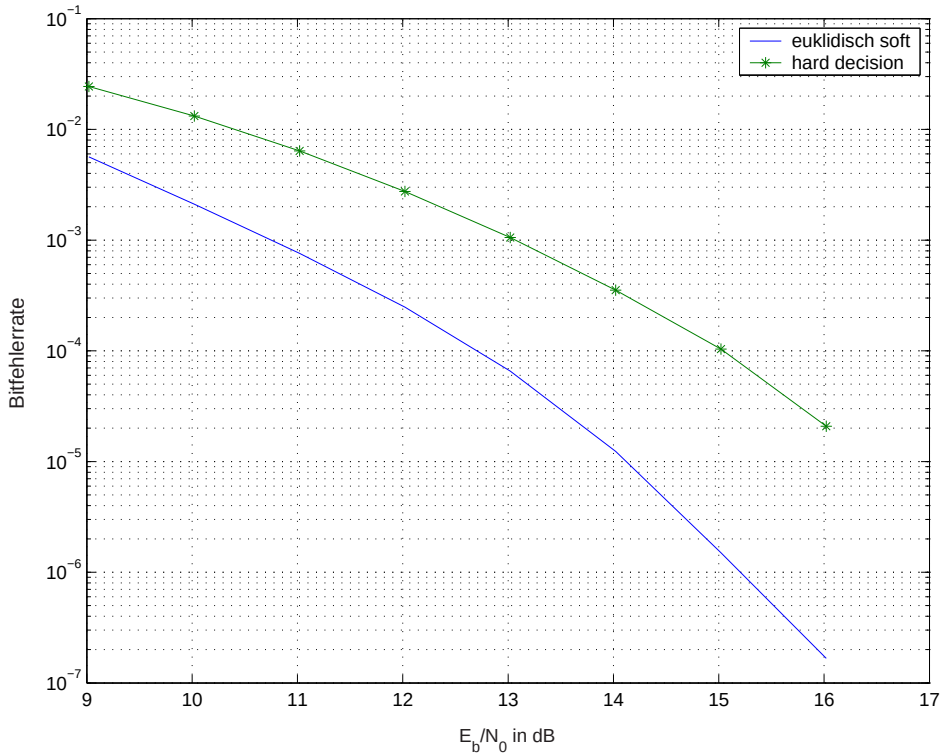


Abbildung 3.4: Beispiel 64-QAM, (1,4,4,7,7,7)-Kombinationscode

**Beispiel 3.2.5** ($(1,5,5,15,15,15)$**-Code).** Zu guter Letzt haben wir den 15-spaltigen Code $E^2$ um zwei uncodierte Zeilen erweitert, um ihn für ein 64-QAM-moduliertes Signal zu verwenden. Der Code hat also folgende Gestalt:

$$E^4 : \{\pm 1\}^1 \times \{\pm 1\}^5 \times \{\pm 1\}^5 \times \{\pm 1\}^{15} \times \{\pm 1\}^{15} \times \{\pm 1\}^{15} \hookrightarrow \{\pm 1\}^{6\times 15}$$

$$\mathbf{u} \mapsto E^4(\mathbf{u}) = \begin{pmatrix} E_{R15}(\mathbf{u}_{1,\cdot}) \\ E_{15,5}(\mathbf{u}_{2,\cdot}) \\ E_{15,5}(\mathbf{u}_{3,\cdot}) \\ E_{uncod}(\mathbf{u}_{4,\cdot}) \\ E_{uncod}(\mathbf{u}_{5,\cdot}) \\ E_{uncod}(\mathbf{u}_{6,\cdot}) \end{pmatrix}. \tag{3.37}$$

Es handelt sich um einen 6-zeiligen $(90, 56)$-Block-Code mit einer Coderate von $\frac{28}{45}$. Das Decodier-Ergebnis ist in Abbildung 3.5 dargestellt.

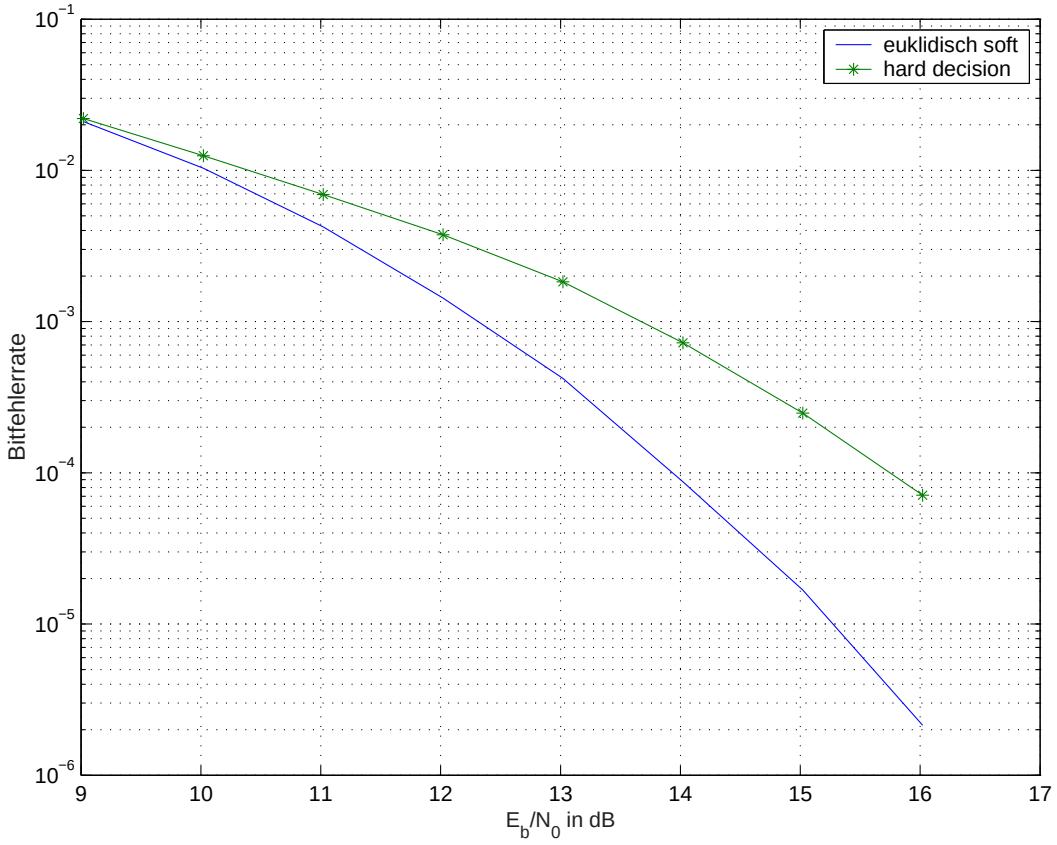


Abbildung 3.5: Beispiel 64-QAM, (1,5,5,15,15,15)-Kombinationscode

# Anhang A

# Wahrscheinlichkeitstheorie

In diesem Anhang geben wir einen kurzen Überblick über die Maßtheorie und Wahrscheinlichkeitstheorie. Für eine detailliertere Darstellung verweisen wir auf [Ma95], [Bau90], und für die stochastische Analysis auf [HaTh94], [KaSh91] und [ReYo91].

Wir betrachten eine beliebige nichtleere Basismenge $\Omega$. Eine Menge $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ wird als Mengensystem (über $\Omega$) bezeichnet, wobei $\mathcal{P}$ die Potenzmenge (Menge aller Teilmengen) von $\Omega$ darstellt. Mit $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}$ wird eine Erweiterung der Menge aller reellen Zahlen definiert. Die algebraische Struktur von $\mathbb{R}$ wird folgendermaßen auf $\bar{\mathbb{R}}$ erweitert: Für alle $a \in \mathbb{R}$ gilt:

$$a + (\pm\infty) = (\pm\infty) + a = (\pm\infty) + (\pm\infty) = (\pm\infty), \quad +\infty - (-\infty) = +\infty,$$

$$a \cdot (\pm\infty) = (\pm\infty) \cdot a = \begin{cases} (\pm\infty), & \text{für } a > 0, \\ 0, & \text{für } a = 0, \\ (\mp\infty), & \text{für } a < 0, \end{cases}$$

$$(\pm\infty) \cdot (\pm\infty) = +\infty, \quad (\pm\infty) \cdot (\mp\infty) = -\infty, \quad \frac{a}{\pm\infty} = 0.$$

Somit ist $\bar{\mathbb{R}}$ kein Körper. Die Vorzeichen bei $\pm\infty$ dürfen bei den obigen Formeln nicht kombiniert werden, denn der Ausdruck $+\infty - (+\infty)$ ist nicht definiert. Die Bedeutung der Festlegung $0 \cdot (\pm\infty) = (\pm\infty) \cdot 0 = 0$ wird später deutlich. Vorsicht ist allerdings bei den Grenzwertsätzen geboten:

$$\lim_{x \to +\infty} \left( x \cdot \frac{1}{x} \right) \neq (+\infty) \cdot 0 = 0.$$

Ergänzt man die Ordnungsstruktur von $\mathbb{R}$ durch $-\infty < a$, $a < +\infty$ für alle $a \in \mathbb{R}$ und $-\infty < +\infty$, so ist $(\bar{\mathbb{R}}, \leq)$ eine geordnete Menge. Aufgrund topologischer Überlegungen können wir unter Verzicht auf die entsprechenden Grenzwertsätze

vereinbaren, dass die Folge $\{n\}$, $n \in \mathbb{N}$, den Grenzwert $+\infty \in \bar{\mathbb{R}}$ besitzt. Für „$+\infty$“ schreiben wir oft „$\infty$“. In Analogie zur Berechnung von Volumina in der Geometrie versucht man, Mengen aus einem Mengensystem $\mathcal{F}$ über $\Omega$ Maße (Volumina) zuzuordnen. Zu diesem Zweck zeichnet man spezielle Funktionen aus.

**Definition A.1.1 ($\sigma$-endliches) Maß).** Sei $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, $\emptyset \in \mathcal{F}$. Eine Funktion $\mu : \mathcal{F} \to \bar{\mathbb{R}}$ heißt Maß auf $\mathcal{F}$, falls die folgenden Bedingungen erfüllt sind:

**(M1)** $\mu(A) \geq 0$ für alle $A \in \mathcal{F}$,

**(M2)** $\mu(\emptyset) = 0$,

**(M3)** Für jede Folge $\{A_i\}, i \in \mathbb{N}$, paarweise disjunkter Mengen mit $A_i \in \mathcal{F}$, $i \in \mathbb{N}$, und $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ gilt:

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \quad (\sigma\text{-Additivität}).$$

Besitzen für eine Folge $\{B_i\}, i \in \mathbb{N}$, mit $B_i \subseteq B_{i+1}$, $B_i \in \mathcal{F}$ und $\bigcup_{i=1}^{\infty} B_i = \Omega$ alle Mengen $B_i$, $i \in \mathbb{N}$, ein endliches Maß, so wird $\mu$ als $\sigma$-endlich bezeichnet.

Es wäre naheliegend, Maße auf der Potenzmenge von $\Omega$ zu betrachten. Allerdings ist diese Vorgehensweise problematisch, da es zum Beispiel nicht möglich ist, ein translationsinvariantes Maß $\mu$ auf der Potenzmenge des $\mathbb{R}^3$ mit $\mu(\mathbb{R}^3) = 1$ zu finden. Daher hat man sich im allgemeinen mit speziellen Mengensystemen über $\Omega$ (Teilmengen der Potenzmenge) zu begnügen. Dies führt auf den Begriff der $\sigma$-Algebra.

**Definition A.1.2 ($\sigma$-Algebra).** Ein Mengensystem $\mathcal{S} \subseteq \mathcal{P}(\Omega)$ heißt $\sigma$-Algebra über $\Omega$, falls die folgenden Axiome erfüllt sind:

**(S1)** $\Omega \in \mathcal{S}$,

**(S2)** Aus $A \in \mathcal{S}$ folgt $A^c := \Omega \setminus A \in \mathcal{S}$,

**(S3)** Aus $A_i \in \mathcal{S}$, $i \in \mathbb{N}$, folgt $\bigcup_{i=1}^{\infty} A_i \in \mathcal{S}$.

Die folgende Eigenschaft von $\sigma$-Algebren ist wichtig.

**Definition und Satz A.1.3 (Durchschnittsstabilität von $\sigma$-Algebren).** Sei $I$ eine beliebige nichtleere Menge und $\mathcal{S}_i$ für jedes $i \in I$ eine $\sigma$-Algebra über $\Omega$, so ist auch $\bigcap_{i\in I} \mathcal{S}_i$ eine $\sigma$-Algebra über $\Omega$. Diese Eigenschaft wird Durchschnittsstabilität von $\sigma$-Algebren genannt.

Wir können also von erzeugten $\sigma$-Algebren sprechen.

**Definition A.1.4 (erzeugte $\sigma$-Algebra).** Sei $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ und sei $\Sigma$ die Menge aller $\sigma$-Algebren über $\Omega$, die $\mathcal{F}$ enthalten, dann wird die $\sigma$-Algebra $\sigma(\mathcal{F}) := \bigcap_{\mathcal{S}\in\Sigma} \mathcal{S}$ als die von $\mathcal{F}$ erzeugte $\sigma$-Algebra bezeichnet.

Für $\Omega = \mathbb{R}^n$, $n \in \mathbb{N}$, betrachten wir die $\sigma$-Algebra

$$\mathcal{B}^n = \sigma(\{([a_1, b_1[\times \ldots \times [a_n, b_n[) \cap \mathbb{R}^n;\ -\infty \leq a_i \leq b_i \leq \infty,\ i = 1, \ldots, n\}),$$

wobei $[a_1, b_1[\times \ldots \times [a_n, b_n[ := \emptyset$, falls $a_j \geq b_j$ für mindestens ein $j \in \{1, \ldots, n\}$. Auf dieser $\sigma$-Algebra lässt sich nun ein eindeutiges Maß $\lambda$ durch

$$\lambda(([a_1, b_1[\times \ldots \times [a_n, b_n[) \cap \mathbb{R}^n) = \begin{cases} \prod_{i=1}^{n} (b_i - a_i), & \text{falls } b_i > a_i,\ i = 1, \ldots, n \\ 0, & \text{sonst} \end{cases}$$

festlegen. Dieses Maß heißt Lebesgue-Borel-Maß. Die $\sigma$-Algebra $\mathcal{B}^n$ wird als Borelsche $\sigma$-Algebra bezeichnet. Alle für die Praxis wichtigen Teilmengen des $\mathbb{R}^n$ (etwa alle offenen, abgeschlossenen und kompakten Teilmengen) sind in $\mathcal{B}^n$ enthalten. Das Maß $\lambda$ ist unter allen translationsinvarianten Maßen $\mu$ auf $\mathcal{B}^n$ das einzige Maß mit $\mu([0, 1[\times \ldots \times [0, 1[) = 1$. Sei nun $\mu$ ein Maß auf einer $\sigma$-Algebra $\mathcal{S}$ über $\Omega$, so heißt jede Menge $A \in \mathcal{S}$ mit $\mu(A) = 0$ eine $\mu$-Nullmenge. Es ist nun naheliegend, jeder Teilmenge $B \subseteq A$ einer $\mu$-Nullmenge ebenfalls das Maß $\mu(B) = 0$ zuzuordnen. Allerdings ist nicht gewährleistet, dass für jedes $B \subseteq A$ auch $B \in \mathcal{S}$ gilt. Das führt zum Begriff der Vervollständigung und des vollständigen Maßes.

**Definition A.1.5 (vollständiges Maß, Vervollständigung).** Ein Maß $\mu$ auf einer $\sigma$-Algebra $\mathcal{S}$ über $\Omega$ heißt vollständig, falls jede Teilmenge einer $\mu$-Nullmenge zu $\mathcal{S}$ gehört und damit eine $\mu$-Nullmenge ist. Ist $\mu$ nicht vollständig, so heißt die $\sigma$-Algebra

$$\mathcal{S}_0 := \{A \cup N;\ A \in \mathcal{S},\ N \text{ Teilmenge einer } \mu\text{-Nullmenge}\}$$

$\mu$-Vervollständigung von $\mathcal{S}$. Mit $\mu_0(A \cup N) := \mu(A)$ ist $\mu_0$ ein vollständiges Maß auf $\mathcal{S}_0$.

Die Mengen der $\sigma$-Algebra $\mathcal{B}_0^n$ heißen Lebesgue-messbare Mengen. Das Maß $\lambda_0^n$ auf $\mathcal{B}_0^n$ heißt Lebesgue-Maß. Die zugehörigen Nullmengen heißen Lebesguesche Nullmengen.
Betrachtet man eine Funktion $F : \mathbb{R} \to \mathbb{R}$ mit folgenden Eigenschaften:

- $F$ ist monoton steigend,
- $F$ ist stetig von links,

so ist durch

$$\mu^F([a,b[\cap\mathbb{R}) := \begin{cases} F(b)-F(a), & \text{falls } -\infty < a < b < \infty \\ \lim\limits_{b\to\infty} F(b) - F(a), & \text{falls } -\infty < a < b = \infty \\ F(b) - \lim\limits_{a\to-\infty} F(a), & \text{falls } -\infty = a < b < \infty \\ \lim\limits_{b\to\infty} F(b) - \lim\limits_{a\to-\infty} F(a), & \text{falls } -\infty = a,\ b = \infty \\ 0, & \text{falls } a \geq b \end{cases}$$

ein eindeutiges Maß $\mu^F$ auf $\mathcal{B}$ definiert. Dieser Sachverhalt führt zu folgender Definition.

**Definition A.1.6 (maßerzeugende Funktion, $\Omega = \mathbb{R}$).** Eine monoton steigende Funktion $F : \mathbb{R} \to \mathbb{R}$, die stetig von links ist, heißt maßerzeugende Funktion.

Das Maß $\mu^F$ heißt Lebesgue-Borel-Stieltjes-Maß. Das vollständige Maß $\mu_0^F$ auf der $\mu^F$-Vervollständigung $\mathcal{B}_0^F$ von $\mathcal{B}$ heißt Lebesgue-Stieltjes-Maß. Die Mengen $A \in \mathcal{B}_0^F$ heißen Lebesgue-Stieltjes-messbar. Durch analoge Vorgehensweise lassen sich maßerzeugende Funktionen auf $\Omega = \mathbb{R}^n$ definieren. Wir wollen darauf aber nicht näher eingehen.
Ist $\mathcal{S}$ eine $\sigma$-Algebra über $\Omega$, so bezeichnen wir im folgenden das Paar $(\Omega, \mathcal{S})$ als Messraum. Ist $\mu$ ein Maß auf $\mathcal{S}$, so heißt das Tripel $(\Omega, \mathcal{S}, \mu)$ Maßraum. Nun untersuchen wir spezielle Funktionen zwischen zwei Grundmengen $\Omega_1, \Omega_2 \neq \emptyset$.

**Definition A.1.7 (messbare Abbildung).** Seien $(\Omega_1, \mathcal{S}_1)$ und $(\Omega_2, \mathcal{S}_2)$ zwei Messräume.
Eine Abbildung $T : \Omega_1 \to \Omega_2$ mit $T^{-1}(A') := \{x \in \Omega_1; T(x) \in A'\} \in \mathcal{S}_1$ für alle $A' \in \mathcal{S}_2$ heißt $\mathcal{S}_1$-$\mathcal{S}_2$-messbar.

Messbare Abbildungen spielen in der Wahrscheinlichkeitstheorie bei der Definition von Zufallsvariablen eine wichtige Rolle. Der folgende Satz zeigt, dass für den Nachweis der Messbarkeit einer Abbildung nicht immer das Urbild $T^{-1}(A')$ für alle Mengen $A' \in \mathcal{S}_2$ untersucht werden muss.

**Satz A.1.8 (Messbarkeit bei einer erzeugten $\sigma$-Algebra $\mathcal{S}_2$).** *Seien $(\Omega_1, \mathcal{S}_1)$ und $(\Omega_2, \mathcal{S}_2)$ zwei Messräume, wobei $\mathcal{S}_2 = \sigma(\mathcal{F})$ von einem Mengensystem $\mathcal{F}$ erzeugt ist. Die Abbildung $T : \Omega_1 \to \Omega_2$ ist genau dann $\mathcal{S}_1$-$\mathcal{S}_2$-messbar, falls $T^{-1}(A') \in \mathcal{S}_1$ für alle $A' \in \mathcal{F}$.*

Sind drei Messräume $(\Omega_1, \mathcal{S}_1)$, $(\Omega_2, \mathcal{S}_2)$, $(\Omega_3, \mathcal{S}_3)$ und zwei Abbildungen
$T_1 :\ \Omega_1 \to \Omega_2$, $T_1$ $\mathcal{S}_1$-$\mathcal{S}_2$-messbar,
$T_2 :\ \Omega_2 \to \Omega_3$, $T_2$ $\mathcal{S}_2$-$\mathcal{S}_3$-messbar,
gegeben, so ist die Abbildung
$T_2 \circ T_1 :\ \Omega_1 \to \Omega_3$, $\omega \mapsto T_2(T_1(\omega))$, $\mathcal{S}_1$-$\mathcal{S}_3$-messbar.

**Definition und Satz A.1.9 (Bildmaß).** Seien $(\Omega_1, \mathcal{S}_1, \mu_1)$ ein Maßraum, $(\Omega_2, \mathcal{S}_2)$ ein Messraum und $T :\ \Omega_1 \to \Omega_2$ $\mathcal{S}_1$-$\mathcal{S}_2$-messbar, so ist durch

$$\mu_2(A') := \mu_1\left(T^{-1}(A')\right), \quad A' \in \mathcal{S}_2,$$

ein Maß $\mu_2$ auf $\mathcal{S}_2$ definiert.
Das Maß $\mu_2$ wird als Bildmaß von $\mu_1$ bezeichnet (Schreibweise: $\mu_2 = T(\mu_1)$).

Um Zufallsgrößen analysieren zu können, benötigt man einen Integralbegriff. Daher soll im folgenden kurz die Integrationstheorie für messbare Abbildungen zusammengefasst werden. Zunächst betrachten wir die Integration einer speziellen Klasse von Funktionen.

**Definition A.1.10 (elementare Funktion).** Sei $(\Omega, \mathcal{S})$ ein Messraum. Eine $\mathcal{S}$-$\mathcal{B}$-messbare Funktion $e :\ \Omega \to \mathbb{R}$ heißt elementare Funktion, falls sie nur endlich viele verschiedene Funktionswerte annimmt.

Eine spezielle elementare Funktion ist die Indikatorfunktion

$$I_A : \Omega \to \mathbb{R}, \quad \omega \mapsto \begin{cases} 1, & \text{falls } \omega \in A \\ 0, & \text{sonst} \end{cases},$$

die anzeigt, ob $\omega$ Element einer Menge $A \in \mathcal{S}$ ist. Mit Hilfe von Indikatorfunktionen lassen sich die elementaren Funktionen darstellen.

**Satz A.1.11 (Darstellung elementarer Funktionen).** *Sei $(\Omega, \mathcal{S})$ ein Messraum. Ist $e : \Omega \to \mathbb{R}$ eine elementare Funktion, so existieren eine natürliche Zahl $n$, paarweise disjunkte Mengen $A_1, \ldots, A_n \in \mathcal{S}$ und reelle Zahlen $\alpha_1, \ldots, \alpha_n$ mit:*

$$e = \sum_{i=1}^{n} \alpha_i I_{A_i}, \quad \sum_{i=1}^{n} A_i = \Omega.$$

Die eben betrachtete Darstellung von $e$ heißt eine Normaldarstellung von $e$. Sind alle $\alpha_i$ paarweise verschieden, so spricht man von einer kürzesten Normaldarstellung von $e$. Kürzeste Normaldarstellungen sind eindeutig. Aus der Normaldarstellung elementarer Funktionen folgt sofort: Summe, Differenz und Produkt elementarer Funktionen sind elementare Funktionen. Für alle $c \in \mathbb{R}$ ist auch $c \cdot e$ eine elementare Funktion, wenn $e$ eine elementare Funktion ist.
Nun betrachten wir nichtnegative elementare Funktionen auf einem Maßraum $(\Omega, \mathcal{S}, \mu)$ und definieren das ($\mu$-)Integral dieser Funktionen.

**Definition A.1.12 ($(\mu$-)Integral nichtnegativer elementarer Funktionen).** Sei $(\Omega, \mathcal{S}, \mu)$ ein Maßraum und $e : \Omega \to \mathbb{R}_0^+$, $e = \sum_{i=1}^{n} \alpha_i I_{A_i}$, $\alpha_i \geq 0$, $i = 1, \ldots, n$, eine nichtnegative elementare Funktion in Normaldarstellung, so wird

$$\int e \, d\mu := \int_{\Omega} e \, d\mu := \sum_{i=1}^{n} \alpha_i \cdot \mu(A_i)$$

als ($\mu$-)Integral von $e$ über $\Omega$ bezeichnet.

Damit $\int e \, d\mu$ wohldefiniert ist, ist natürlich zu zeigen, dass $\int e \, d\mu$ unabhängig von der Wahl der Normaldarstellung für $e$ ist.
Sei nun $E$ die Menge aller nichtnegativen elementaren Funktionen auf $(\Omega, \mathcal{S}, \mu)$, so erhalten wir eine Abbildung

$$\text{Int} : \; E \to \bar{\mathbb{R}}_0^+, \, e \mapsto \int e \, d\mu.$$

Die folgenden Eigenschaften von Int lassen sich leicht nachweisen:

- $\int I_A \, d\mu = \mu(A)$ für alle $A \in \mathcal{S}$.
- $\int (\alpha e) d\mu = \alpha \int e \, d\mu$ für alle $e \in E$, $\alpha \in \mathbb{R}_0^+$.
- $\int (u + v) d\mu = \int u \, d\mu + \int v \, d\mu$ für alle $u, v \in E$.
- Ist $u(\omega) \leq v(\omega)$ für alle $\omega \in \Omega$, so ist $\int u \, d\mu \leq \int v \, d\mu$ für alle $u, v \in E$.

Wählen wir $\Omega = \mathbb{R}^n$, $\mathcal{S} = \mathcal{B}^n$, $\mu = \lambda^n$ und $f : \Omega \to \mathbb{R}_0^+$, $x \mapsto 0$, so erhalten wir

$$\int f \, d\lambda^n = \int 0 \, d\lambda^n = 0 \cdot \lambda^n(\mathbb{R}^n) = 0 \cdot \infty = 0.$$

Unsere Vereinbarung $0 \cdot \infty = 0$ erlaubt uns somit, das ($\lambda^n$-) Integral über die Nullfunktion zu berechnen.
Betrachtet man die Menge $\bar{\mathbb{R}}$ der um $\{\pm\infty\}$ erweiterten reellen Zahlen, so bildet die Menge $\bar{\mathcal{B}} := \{A \in \mathcal{P}(\bar{\mathbb{R}}); \, A \cap \mathbb{R} \in \mathcal{B}\}$ eine $\sigma$-Algebra über $\bar{\mathbb{R}}$. Um nun den Integralbegriff auf eine größere Klasse von Funktionen fortzusetzen, benötigen wir die folgende Definition.

**Definition A.1.13 (numerische Funktion).** Eine auf einer nichtleeren Menge $A \subseteq \Omega$ definierte Funktion $f : \, A \to \bar{\mathbb{R}}$ heißt numerische Funktion.

Nun betrachten wir nichtnegative numerische Funktionen, die als Grenzwert einer Folge elementarer Funktionen gegeben sind.

**Satz A.1.14 (Grenzwerte spezieller Folgen elementarer Funktionen).** *Seien $(\Omega, \mathcal{S})$ ein Messraum und $f : \Omega \to \bar{\mathbb{R}}_0^+$ eine nichtnegative, $\mathcal{S}$-$\bar{\mathcal{B}}$-messbare numerische Funktion, so gibt es eine monoton steigende Folge $\{e_n\}$, $n \in \mathbb{N}$, von nichtnegativen elementaren Funktionen $e_n : \Omega \to \mathbb{R}_0^+$, $n \in \mathbb{N}$, die punktweise gegen $f$ konvergiert. Wir schreiben dafür: $e_n \uparrow f$.*

Nach diesen Vorbereitungen sind wir in der Lage, die ($\mu$-)Integration auf eine spezielle Klasse von Funktionen in naheliegender Weise fortzusetzen.

**Definition A.1.15 (($\mu$-)Integral für $\mathcal{S}$-$\bar{\mathcal{B}}$-messbare, nichtnegative numerische Funktionen).** Seien $(\Omega, \mathcal{S}, \mu)$ ein Maßraum und $f : \Omega \to \bar{\mathbb{R}}_0^+$ eine $\mathcal{S}$-$\bar{\mathcal{B}}$-messbare, nichtnegative numerische Funktion. Sei ferner $\{e_n\}$, $n \in \mathbb{N}$, eine monoton steigende Folge nichtnegativer elementarer Funktionen $e_n : \Omega \to \mathbb{R}_0^+$, $n \in \mathbb{N}$, mit $e_n \uparrow f$, so definieren wir durch

$$\int f \, d\mu := \int_\Omega f \, d\mu := \lim_{n\to\infty} \int e_n \, d\mu$$

das ($\mu$-)Integral von $f$ über $\Omega$.

Da die approximierende Folge elementarer Funktionen für $f$ nicht eindeutig ist, muss natürlich erwähnt werden, dass das eben definierte Integral wohldefiniert ist. Wir werden nun in einem letzten Schritt die Klasse der integrierbaren Funktionen erweitern. Dazu dient die folgende Definition.

**Definition A.1.16 (Positivteil, Negativteil einer numerischen Funktion).** Seien $(\Omega, \mathcal{S})$ ein Messraum und $f : \Omega \to \bar{\mathbb{R}}$ eine $\mathcal{S}$-$\bar{\mathcal{B}}$-messbare numerische Funktion, so wird die Funktion

$$f^+ : \Omega \to \bar{\mathbb{R}}_0^+, \, \omega \mapsto \begin{cases} f(\omega), & \text{falls } f(\omega) \geq 0 \\ 0, & \text{sonst} \end{cases}$$

Positivteil von $f$ und die Funktion

$$f^- : \Omega \to \bar{\mathbb{R}}_0^+, \, \omega \mapsto \begin{cases} -f(\omega), & \text{falls } f(\omega) \leq 0 \\ 0, & \text{sonst} \end{cases}$$

Negativteil von $f$ genannt.

Die folgenden Eigenschaften von $f^+$ und $f^-$ sind unmittelbar einzusehen:

- $f^+(\omega) \geq 0$, $f^-(\omega) \geq 0$ für alle $\omega \in \Omega$.
- $f^+$ und $f^-$ sind $\mathcal{S}$-$\bar{\mathcal{B}}$-messbare numerische Funktionen.
- $f = f^+ - f^-$.

Mit Hilfe des Positiv- und Negativteils einer messbaren numerischen Funktion $f : \Omega \to \bar{\mathbb{R}}$ können wir das ($\mu$-)Integral auf messbare numerische Funktionen erweitern.

**Definition A.1.17 (($\mu$-)integrierbar, ($\mu$-)quasiintegrierbar, ($\mu$-)Integral).** Seien $(\Omega, \mathcal{S}, \mu)$ ein Maßraum und $f : \Omega \to \bar{\mathbb{R}}$ eine $\mathcal{S}$-$\bar{\mathcal{B}}$-messbare numerische Funktion.
$f$ heißt ($\mu$-)integrierbar, falls $\int f^+ \, d\mu < \infty$ und $\int f^- \, d\mu < \infty$.
$f$ heißt ($\mu$-)quasiintegrierbar, falls $\int f^+ \, d\mu < \infty$ oder $\int f^- \, d\mu < \infty$.
Ist $f$ ($\mu$-)quasiintegrierbar, so ist durch

$$\int f \, d\mu := \int_\Omega f \, d\mu := \int f^+ \, d\mu - \int f^- \, d\mu$$

das ($\mu$-)Integral von $f$ über $\Omega$ definiert.

Als ($\mu$-)Integral über einer Menge $A \in \mathcal{S}$ definieren wir für ($\mu$-)quasiintegrierbares $f \cdot I_A$:

$$\int_A f \, d\mu := \int f \cdot I_A \, d\mu.$$

Betrachtet man speziell den Maßraum $(\mathbb{R}^n, \mathcal{B}^n, \lambda^n)$, so wird das ($\lambda^n$-)Integral als Lebesgue-Integral bezeichnet. Ist $f$ ($\lambda^n$-)integrierbar, so heißt $f$ Lebesgue-integrierbar. Ist ein Maß $\mu^F$ durch eine maßerzeugende Funktion $F : \mathbb{R}^n \to \mathbb{R}$ gegeben, so wird das ($\mu^F$-)Integral als Lebesgue-Stieltjes-Integral bezeichnet und in der Form

$$\int f \, dF := \int f \, d\mu^F$$

geschrieben. Lebesgue-Stieltjes-Integrale besitzen die wichtige Eigenschaft, dass sie häufig durch Riemann-Integrale berechnet werden können.
In der Wahrscheinlichkeitstheorie werden Methoden zur Beschreibung und Analyse von Zufallsexperimenten (Experimente mit nicht vorhersehbarem Ausgang) bereitgestellt (für Details sei auf [Bau90], [Bau91], [Bi86], [Ši88] und [ScSt94] verwiesen). Der umgangssprachliche Begriff 'Zufallsexperiment' wird durch einen Maßraum $(\Omega, \mathcal{S}, P)$ mit der Eigenschaft $P(\Omega) = 1$ mathematisch präzisiert. Wir definieren daher:

**Definition A.1.18 (Wahrscheinlichkeitsraum, Wahrscheinlichkeitsmaß, Ergebnis, Ereignis).** Ein Maßraum $(\Omega, \mathcal{S}, P)$ mit $P(\Omega) = 1$ wird als Wahrscheinlichkeitsraum bezeichnet. Die Punkte $\omega \in \Omega$ heißen Ergebnisse, die Mengen $A \in \mathcal{S}$ Ereignisse. Das Maß $P$ wird als Wahrscheinlichkeitsmaß bezeichnet. Für alle Ereignisse $A$ wird $P(A)$ die Wahrscheinlichkeit von $A$ genannt.

Wir werden im folgenden davon ausgehen, dass ein Zufallsexperiment durch einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ gegeben ist. Es ist in der Praxis oft nicht leicht, ein verbal formuliertes Zufallsexperiment durch einen Wahrscheinlichkeitsraum zu modellieren - insbesondere dann, wenn das Experiment ungenau formuliert ist. Die Elemente der Menge $\Omega$ stellen die möglichen Ergebnisse des Zufallsexperimentes dar.

**Roulette**
Beim Roulette wäre etwa $\Omega = \{0, \ldots, 36\}$. Häufig interessiert man sich weniger für die Frage, mit welcher Wahrscheinlichkeit ein spezielles Ergebnis eines Zufallsexperimentes eintrifft, sondern dafür, mit welcher Wahrscheinlichkeit das Ergebnis Element einer speziellen Teilmenge von $\Omega$ ist. Die in Frage kommenden Teilmengen werden Ereignisse genannt und in der $\sigma$-Algebra $\mathcal{S}$ zusammengefasst. Die Wahrscheinlichkeit, dass das Ergebnis eines durch $(\Omega, \mathcal{S}, P)$ repräsentierten Zufallsexperiments Element der Menge $A \in \mathcal{S}$ ist, ist durch $P(A)$ gegeben. Einen Roulettespieler, der auf 'gerade Zahl' setzt, interessiert es zum Beispiel nicht, mit welcher Wahrscheinlichkeit ein Ergebnis eintrifft, sondern mit welcher Wahrscheinlichkeit das Ergebnis Element der Menge $\{2, 4, 6, \ldots, 36\}$ ist. Gilt für ein $\omega \in \Omega$ auch $\{\omega\} \in \mathcal{S}$, so spricht man von einem Elementarereignis. Das Ereignis $\emptyset$ heißt unmögliches Ereignis, das Ereignis $\Omega$ heißt sicheres Ereignis . Da $P(\emptyset) = 0$ und $P(\Omega) = 1$, bezeichnet man eine Menge $A \in \mathcal{S}$, für die $P(A) = 0$ gilt (also eine $P$-Nullmenge) als ($P$-)fast unmögliches Ereignis, eine Menge $B \in \mathcal{S}$, für die $P(B) = 1$ gilt, als ($P$-)fast sicheres Ereignis. Es ist wichtig festzuhalten, dass die definierenden Eigenschaften eines Maßes mit den intuitiv einsichtigen Eigenschaften von Wahrscheinlichkeiten übereinstimmen. Für das Roulette ist es sicher sinnvoll, den folgenden Wahrscheinlichkeitsraum zu betrachten:
$\Omega = \{0, 1, 2, \ldots, 36\}$, $\mathcal{S} = \mathcal{P}(\Omega)$, $P : \mathcal{P}(\Omega) \to [0, 1]$,

$$A \mapsto \begin{cases} 0, & \text{falls } A = \emptyset \\ \frac{|A|}{37}, & \text{falls } A \neq \emptyset \end{cases}, \text{ wobei } |A| \text{ die Anzahl der Elemente von } A \text{ bezeichnet.}$$

**Scheibenschießen**
Nun betrachten wir das Schießen mit einem Gewehr auf eine kreisförmige Schießscheibe mit Radius $r = \frac{1}{\sqrt{\pi}}$ und dem Mittelpunkt $\mathbf{m} = (0, 0)^T$. Wir nehmen an, dass bei jedem Schuss die Scheibe getroffen wird. Als Ergebnis eines Schusses erhalten wir einen Punkt $\omega = (\omega^1, \omega^2)^T \in \Omega := K_{\frac{1}{\sqrt{\pi}},0} := \{\mathbf{x} \in \mathbb{R}^2;\ \|\mathbf{x}\|_2 \leq \frac{1}{\sqrt{\pi}}\}$. Wir wählen $\{A \cap K_{\frac{1}{\sqrt{\pi}},0};\ A \in \mathcal{B}^2\}$ als $\sigma$-Algebra und $P = \lambda^2|_{\mathcal{S}}$ als Wahrscheinlichkeitsmaß auf $\mathcal{S}$. Da der Schütze bei jedem Schuss umso mehr Punkte (Ringe) erhält, je kleiner der Abstand seines Schusses zum Mittelpunkt der Schießscheibe ist, interessiert als Ergebnis in erster Linie dieser Abstand zum Mittelpunkt.
Man betrachtet also eine Funktion

$$d : \Omega \to [0, \frac{1}{\sqrt{\pi}}] =: \Omega',\ \omega \mapsto \|\omega\|_2.$$

Kann man nun mit Hilfe der Funktion $d$ und des Wahrscheinlichkeitsraumes $(\Omega, \mathcal{S}, P)$ jeder Menge $A \in \mathcal{S}' := \{B \cap [0, \frac{1}{\sqrt{\pi}}];\ B \in \mathcal{B}\}$ eine Wahrscheinlichkeit zuordnen? Dies ist genau dann möglich, wenn $d$ $\mathcal{S}$-$\mathcal{S}'$-messbar ist. Daher definieren wir:

**Definition A.1.19 ((*n*-dimensionale reelle, numerische) Zufallsvariable).** Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $(\Omega', \mathcal{S}')$ ein Messraum, dann heißt eine $\mathcal{S}$-$\mathcal{S}'$-messbare Funktion $X : \Omega \to \Omega'$ Zufallsvariable.
Ist $\Omega' = \mathbb{R}^n$, $n \in \mathbb{N}$, und $\mathcal{S}' = \mathcal{B}^n$, so wird $X$ als $n$-dimensionale reelle Zufallsvariable bezeichnet. Ist $\Omega' = \bar{\mathbb{R}}$ und $\mathcal{S}' = \bar{\mathcal{B}}$, so wird $X$ als numerische Zufallsvariable bezeichnet. Eine eindimensionale reelle Zufallsvariable wird reelle Zufallsvariable genannt.

Als geeignetes Wahrscheinlichkeitsmaß $P'$ auf $\mathcal{S}'$ ergibt sich das Bildmaß von $X$. Somit erhalten wir für unser obiges Beispiel $P'(A') = P(d^{-1}(A'))$ für alle $A' \in \mathcal{S}'$. Die Tatsache, dass $\lambda^2(\{\omega\}) = 0$ für alle $\omega \in K_{\frac{1}{\sqrt{\pi}},0}$ verdeutlicht den Sinn der Verwendung von Ereignissen $A \in \mathcal{S}$.

**Definition A.1.20 (Verteilung einer Zufallsvariablen).** Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum, $(\Omega', \mathcal{S}')$ ein Messraum und $X : \Omega \to \Omega'$ eine Zufallsvariable, dann wird das Bildmaß $P_X$ von $X$ Verteilung von $X$ genannt.

Nach unserer Interpretation von Wahrscheinlichkeitsräumen ist der Wert $X(\omega)$ einer Zufallsvariablen an der Stelle $\omega$ vom Ergebnis eines Zufallsexperimentes abhängig. Wir fragen danach, welcher Wert von $X$ 'zu erwarten' ist.

**Definition A.1.21 (Erwartungswert einer numerischen Zufallsvariablen).** Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $X$ eine ($P$-)quasiintegrierbare numerische Zufallsvariable $X : \Omega \to \bar{\mathbb{R}}$, dann wird durch

$$\mathbb{E}(X) := \int X \, dP$$

der Erwartungswert von $X$ definiert.

Um eine Vorstellung vom Begriff des Erwartungswertes zu bekommen, betrachten wir die folgende reelle Zufallsvariable auf $(\Omega, \mathcal{S}, P)$: Seien $A_1, \ldots, A_n$ paarweise disjunkte Mengen aus $\mathcal{S}$ mit $\sum_{i=1}^{n} A_i = \Omega$ und $\alpha_1, \ldots, \alpha_n$ nichtnegative reelle Zahlen, dann ist

$$X : \Omega \to \mathbb{R},\ \omega \mapsto \sum_{i=1}^{n} \alpha_i I_{A_i}(\omega)$$

eine reelle Zufallsvariable. Für den Erwartungswert von $X$ erhalten wir

$$\mathbb{E}(X) = \sum_{i=1}^{n} \alpha_i P(A_i).$$

Der Erwartungswert ist in diesem Fall also eine gewichtete Summe der möglichen Werte von $X$, wobei die Gewichte gerade die Wahrscheinlichkeiten für das Auftreten dieser Werte sind. Gilt $P(A_i) = \frac{1}{n}$ für alle $i = 1, \ldots, n$, so erhalten wir als Erwartungswert das arithmetische Mittel der Werte von $X$.
Ist eine reelle Zufallsvariable ($P$-)integrierbar, so lässt sich der Erwartungswert von $X$ auch mit Hilfe des Bildmaßes $P_X$ berechnen:

$$\mathbb{E}(X) = \int x \, dP_X(x) := \int f \, dP_X \text{ mit } f : \mathbb{R} \to \mathbb{R},\ x \mapsto x.$$

Da wir auch an Erwartungswerten von speziellen Funktionen von $X$ interessiert sind, benötigen wir den folgenden Satz.

**Satz A.1.22 ($\mathcal{S}$-$\mathcal{B}$-Messbarkeit stetiger Funktionen reeller Zufallsvariablen).** *Seien $X$ eine reelle Zufallsvariable definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ und $g : \mathbb{R} \to \mathbb{R}$ eine stetige Funktion, dann ist $g \circ X : \Omega \to \mathbb{R}$, $\omega \to g(X(\omega))$ eine reelle Zufallsvariable auf $(\Omega, \mathcal{S}, P)$.*

Somit folgt sofort, dass für eine reelle Zufallsvariable $X$ auf $(\Omega, \mathcal{S}, P)$ und für jedes $k \in \mathbb{N}$ und jedes $\alpha \in \mathbb{R}$ auch $(X - \alpha)^k$ und $|X - \alpha|^k$ reelle Zufallsvariablen auf $(\Omega, \mathcal{S}, P)$ sind. Dies ermöglicht die folgende Definition.

**Definition A.1.23 (zentrierte (absolute) Momente $k$-ter Ordnung).** Sei $X$ eine auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ definierte reelle Zufallsvariable, dann heißt $\mathbb{E}(|X-\alpha|^k)$, $k \in \mathbb{N}$, das in $\alpha$ zentrierte absolute Moment $k$-ter Ordnung von $X$. Ist $(X-\alpha)^k$ ($P$-)quasiintegrierbar, so heißt $\mathbb{E}((X-\alpha)^k)$ das in $\alpha$ zentrierte Moment $k$-ter Ordnung. Ist $\alpha = 0$, so spricht man nur von absoluten Momenten bzw. Momenten $k$-ter Ordnung.

Besonders interessant ist der Fall $k = 2$.

**Definition A.1.24 (Varianz einer reellen Zufallsvariablen).** Sei $X$ eine auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ definierte, ($P$-)integrierbare reelle Zufallsvariable, dann heißt

$$\mathbb{V}(X) := \int (X - \mathbb{E}(X))^2 \, dP$$

die Varianz von $X$.

Die Zahl $\sigma = \sqrt{\mathbb{V}(X)}$ wird als Streuung oder Standardabweichung von $X$ bezeichnet. Oft schreibt man daher $\sigma^2$ für $\mathbb{V}(X)$. Die Varianz ist ein Maß für die zu erwartende Abweichung von $X$ und $\mathbb{E}(X)$.

**Standardisierung**
Für eine reelle Zufallsvariable $X$ mit Streuung $0 < \sigma < \infty$ folgt sofort, dass die reelle Zufallsvariable

$$Y := \frac{X - \mathbb{E}(X)}{\sigma}$$

den Erwartungswert $\mathbb{E}(Y) = 0$ und die Varianz $\mathbb{V}(Y) = 1$ besitzt. Den Übergang von $X$ zu $Y$ bezeichnet man als 'Standardisierung' von $X$. Den Erwartungswert einer $n$-dimensionalen reellen Zufallsvariablen definiert man durch komponentenweise Bildung des Erwartungswertes.
Im folgenden betrachten wir einige wichtige Begriffe der elementaren Wahrscheinlichkeitstheorie. Ausgangspunkt ist der Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ und zwei Mengen $A, B \in \mathcal{S}$ mit $P(B) > 0$. Auf $\mathcal{S}$ definieren wir nun ein Wahrscheinlichkeitsmaß $P^B : \mathcal{S} \to [0, 1]$ durch $A \mapsto \frac{P(A \cap B)}{P(B)}$. Durch den Übergang von $P$ zu $P^B$ erhält die Menge $B$ das Wahrscheinlichkeitsmaß 1. Wir interpretieren $P^B(A)$ als die Wahrscheinlichkeit von $A$ unter der Bedingung, dass das Ereignis $B$ ($P^B$-)fast sicher eintrifft.

**Formel von der totalen Wahrscheinlichkeit**
Betrachtet man nun eine Partition $\{D_i \subset \Omega;\, i \in \mathbb{N}\}$ von $\Omega$, sodass für alle $i \in \mathbb{N}$ $D_i \in \mathcal{S}$ und $P(D_i) > 0$ gilt, so lässt sich sehr leicht die folgende 'Formel von der totalen Wahrscheinlichkeit' nachweisen:

$$P(A) = \sum_{i=1}^{\infty} P(D_i) \cdot P^{D_i}(A) \quad \text{für alle } A \in \mathcal{S}.$$

**Satz von Bayes**
Gilt zusätzlich $P(A) > 0$, so folgt aus

$$P^A(D_i) = \frac{P(D_i \cap A)}{P(A)} = \frac{P^{D_i}(A) \cdot P(D_i)}{P(A)},\ i \in \mathbb{N},$$

der 'Satz von Bayes':

$$P^A(D_i) = \frac{P^{D_i}(A) \cdot P(D_i)}{\sum\limits_{i=1}^{\infty} P(D_i) \cdot P^{D_i}(A)} \quad \text{für alle } i \in \mathbb{N}.$$

**Dichte**
Im folgenden betrachten wir einen Messraum $(\Omega, \mathcal{S})$, ein Wahrscheinlichkeitsmaß $P$ auf $\mathcal{S}$ und ein Maß $\mu$ auf $\mathcal{S}$. Es soll zunächst die Frage untersucht werden, unter welchen Voraussetungen das Wahrscheinlichkeitsmaß $P$ in der folgenden Art und Weise durch das Maß $\mu$ dargestellt werden kann:
Es existiert eine nichtnegative, $\mathcal{S}$-$\bar{\mathcal{B}}$-messbare numerische Funktion $f : \Omega \to \bar{\mathbb{R}}$ mit

$$P(A) = \int_A f \, d\mu \text{ für alle } A \in \mathcal{S}. \tag{A.1}$$

Eine nichtnegative, $\mathcal{S}$-$\bar{\mathcal{B}}$-messbare numerische Funktion $f$, die die Bedingung (A.1) erfüllt, wird Dichte(funktion) des Wahrscheinlichkeitsmaßes $P$ bezüglich $\mu$ genannt. Man sagt auch, dass $P$ bezüglich $\mu$ eine Dichte $f$ besitzt.

**Satz A.1.25 (Beziehung zwischen ($P$-) und ($\mu$-)Nullmengen).** *Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum, $\mu$ ein Maß auf $\mathcal{S}$ und $f$ eine Dichte von $P$ bezüglich $\mu$, dann gilt für alle $A \in \mathcal{S}$ mit $\mu(A) = 0$: $P(A) = 0$.*

Die folgende Definition resultiert aus dem eben betrachteten Satz.

**Definition A.1.26 (absolute Stetigkeit von $P$ bez. $\mu$).** Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $\mu$ ein Maß auf $\mathcal{S}$. $P$ heißt absolutstetig bezüglich $\mu$, falls für alle $A \in \mathcal{S}$ mit $\mu(A) = 0$ gilt: $P(A) = 0$.

Wie der folgende Satz zeigt, ist die absolute Stetigkeit bezüglich eines $\sigma$-endlichen Maßes $\mu$ das entscheidende Kriterium für die Existenz einer Dichte.

**Satz A.1.27 (Radon-Nikodym).** *Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $\mu$ ein $\sigma$-endliches Maß auf $\mathcal{S}$, dann besitzt $P$ genau dann eine Dichte bezüglich $\mu$, wenn $P$ absolutstetig bezüglich $\mu$ ist.*

Nun betrachten wir eine spezielle Klasse von Wahrscheinlichkeitsmaßen. Mit $|A|$ wird die Anzahl der Elemente (Mächtigkeit) von $A$ bezeichnet.

**Definition A.1.28 (diskretes Wahrscheinlichkeitsmaß, diskrete Zufallsvariable).** Sei $(\mathbb{R}^n, \mathcal{B}^n, P)$, $n \in \mathbb{N}$, ein Wahrscheinlichkeitsraum. Das Wahrscheinlichkeitsmaß $P$ heißt diskret, falls eine Menge $B \in \mathcal{B}^n$ mit $|B| \leq |\mathbb{N}|$ und $P(B) = 1$ existiert. Eine $m$-dimensionale reelle Zufallsvariable $X$ definiert auf $(\mathbb{R}^n, \mathcal{B}^n, P)$, $m \in \mathbb{N}$, heißt diskret, falls das Bildmaß $P_X$ von $X$ ein diskretes Wahrscheinlichkeitsmaß auf $(\mathbb{R}^m, \mathcal{B}^m)$ ist.

Da für $m = n$ das Bildmaß $P_X$ der Zufallsvariable $X : \mathbb{R}^n \to \mathbb{R}^n$, $x \mapsto x$, gleich $P$ ist, wird oft der Begriff Verteilung statt Wahrscheinlichkeitsmaß verwendet. Um nun mit Hilfe des Satzes von Radon-Nikodym diskrete Verteilungen (Wahrscheinlichkeitsmaße) durch Dichtefunktionen darstellen zu können, benötigen wir ein spezielles Maß.

**Definition A.1.29 (Zählmaß).** Das auf einer $\sigma$-Algebra $\mathcal{S}$ über $\Omega$ definierte Maß

$$\zeta : \mathcal{S} \to \bar{\mathbb{R}},\ A \mapsto \begin{cases} |A|, & \text{falls } |A| \text{ endlich ist} \\ \infty, & \text{sonst} \end{cases}$$

wird als das Zählmaß auf $\mathcal{S}$ bezeichnet.

Sei nun $(\mathbb{R}^n, \mathcal{B}^n, P)$ ein Wahrscheinlichkeitsraum und $P$ eine diskrete Verteilung auf $\mathcal{B}^n$ mit $P(B) = 1$ für ein $B \in \mathcal{B}^n$ und $|B| \leq |\mathbb{N}|$, dann gilt für alle $C \in \mathcal{B}^n$:

$$P(C) = P(C \cap B) + P(C \cap B^c) = P(C \cap B).$$

Somit genügt es, den Wahrscheinlichkeitsraum $(B, \mathcal{B}_B^n, P)$ mit $\mathcal{B}_B^n := \{C \cap B;\ C \in \mathcal{B}^n\} = \mathcal{P}(B)$ zu betrachten. Da $\zeta$ ein $\sigma$-endliches Maß auf $\mathcal{P}(B)$ ist und $\zeta(A) = 0$ genau dann gilt, wenn $A = \emptyset$, ist jedes Wahrscheinlichkeitsmaß auf $\mathcal{P}(B)$ absolutstetig bezüglich $\zeta$. Somit existiert zu jedem Wahrscheinlichkeitsmaß $P$ auf $\mathcal{P}(B)$ eine Dichte $f : B \to \mathbb{R}_0^+$ mit

$$P(A) = \int_A f \, d\zeta = \sum_{\omega \in A} f(\omega) = \sum_{\omega \in A} P(\{\omega\}) \text{ für alle } A \in \mathcal{P}(B).$$

Es lässt sich also jede diskrete Verteilung auf $\mathcal{B}^n$ durch eine Folge $\{p_j\}$, $j \in \mathbb{N}_0$, nichtnegativer reeller Zahlen mit $\sum\limits_{j=0}^{\infty} p_j = 1$ darstellen.

**Poisson-Verteilung**
Ist

$$p_j = e^{-\lambda} \frac{\lambda^j}{j!},\ j \in \mathbb{N}_0,\ \lambda > 0,$$

so spricht man von einer Poisson-Verteilung mit Parameter $\lambda$ (D. Poisson (1781-1840)).

**Gleichverteilung und Laplace-Experiment**
Ist

$$p_j = \frac{1}{k+1},\ \text{für } j = 0, \ldots, k,\ \text{und } p_j = 0 \text{ für } j > k,\ k \in \mathbb{N}_0,$$

so wird diese Verteilung Gleichverteilung genannt. Ein Zufallsexperiment, das durch einen Wahrscheinlichkeitsraum mit Gleichverteilung repräsentiert wird, heißt nach P. S. de Laplace (1749-1827) Laplace-Experiment.

**Binomial-Verteilung und Bernoulli-Experiment**
Wählt man $p \in \mathbb{R}$, $0 < p < 1$, und $B = \{0, 1, 2, \ldots, s\}$, $s \in \mathbb{N}$, so wird (mit $\binom{s}{j} := \frac{s!}{(s-j)!j!}$) die durch

$$p_j = \binom{s}{j} p^j (1-p)^{s-j} \text{ für } j = 0, \ldots, s,\ \text{und } p_j = 0 \text{ für } j > s,$$

gegebene Verteilung Binomial-Verteilung $B(s, p)$ mit Parameter $s, p$ genannt. Ein Zufallsexperiment, das durch einen Wahrscheinlichkeitsraum mit Binomial-Verteilung mit Parameter $s, p$ repräsentiert wird, heißt nach J. Bernoulli (1654-1705) Bernoulli-Experiment mit Paramter $s, p$. Ein Bernoulli-Experiment kann

folgendermaßen interpretiert werden: Man betrachtet ein Zufallsexperiment, bei dem es nur zwei mögliche Ergebnisse gibt, nämlich mit Wahrscheinlichkeit $p$ das Ergebnis 'T' (Treffer) und mit Wahrscheinlichkeit $(1-p)$ das Ergebnis 'N' (Niete). Dieses Experiment führen wir $s$-mal durch, ohne dass sich die Ergebnisse gegenseitig beeinflussen. Die Wahrscheinlichkeit, dass nach diesen $s$ Versuchen genau $j$ Treffer auftreten, ist gegeben durch $\binom{s}{j}p^j(1-p)^{s-j}$, $0 \leq j \leq s$, $s \in \mathbb{N}$. Somit wird die $s$-malige Durchführung unseres Experimentes durch ein Bernoulli-Experiment beschrieben, falls die Ergebnisse sich nicht gegenseitig beeinflussen. Für sehr große $s$ und sehr kleine $p$ ist es möglich, eine Binomial-Verteilung durch die wesentlich einfacher zu berechnende Poisson-Verteilung mit Parameter $\lambda = s \cdot p$ zu approximieren.

Nun betrachten wir die folgende naheliegende Definition.

**Definition A.1.30 (absolutstetige Zufallsvariable).** Sei $(\mathbb{R}^n, \mathcal{B}^n, P)$, $n \in \mathbb{N}$, ein Wahrscheinlichkeitsraum. Eine $m$-dimensionale reelle Zufallsvariable $X$ definiert auf $(\mathbb{R}^n, \mathcal{B}^n, P)$, $m \in \mathbb{N}$, heißt absolutstetig, falls das Bildmaß $P_X$ von $X$ ein absolutstetiges Wahrscheinlichkeitsmaß auf $\mathcal{B}^m$ bezüglich $\lambda^m$ ist.

Nach dem Satz von Radon-Nikodym ist $P_X$ genau dann absolutstetig bezüglich $\lambda^m$, wenn $P_X$ eine Dichte bezüglich $\lambda^m$ besitzt. Mit Hilfe der beiden nächsten Definitionen ist es möglich, alle Wahrscheinlichkeitsmaße auf $\mathcal{B}$ zu klassifizieren.

**Definition A.1.31 (Verteilungsfunktion).** Sei $(\mathbb{R}^n, \mathcal{B}^n, P)$, $n \in \mathbb{N}$, ein Wahrscheinlichkeitsraum. Die Funktion

$$F : \mathbb{R}^n \to [0,1],\ (x_1, \ldots, x_n)^T \mapsto P\,]-\infty, x_1[\times \ldots \times ]-\infty, x_n[)$$

wird als Verteilungsfunktion von $P$ bezeichnet. Die Verteilungsfunktion des Bildmaßes $P_X$ einer $m$-dimensionalen reellen Zufallsvariable $X : \mathbb{R}^n \to \mathbb{R}^m$, $m \in \mathbb{N}$, wird auch Verteilungsfunktion von $X$ genannt.

**Definition A.1.32 (stetiges Wahrscheinlichkeitsmaß, stetige Zufallsvariable).** Sei $(\mathbb{R}^n, \mathcal{B}^n, P)$, $n \in \mathbb{N}$, ein Wahrscheinlichkeitsraum. Das Wahrscheinlichkeitsmaß (die Verteilung) $P$ heißt stetig, falls die Verteilungsfunktion von $P$ stetig ist. Eine $m$-dimensionale reelle Zufallsvariable $X$ definiert auf $(\mathbb{R}^n, \mathcal{B}^n, P)$, $n \in \mathbb{N}$, $m \in \mathbb{N}$, heißt stetig, falls die Verteilungsfunktion von $X$ stetig ist.

Die Verteilungsfunktion einer diskreten Verteilung ist eine Treppenfunktion und damit nicht stetig. Die Verteilungsfunktion einer bezüglich $\lambda^n$ absolutstetigen Verteilung auf $\mathcal{B}^n$ ist stetig (für $n = 1$ sogar absolut stetig im topologischen Sinne). Nun stellt sich natürliche die Frage, ob jede stetige Verteilung auch absolutstetig bezüglich $\lambda^n$ ist, und wir somit zwei verschiedene Namen für eine Klasse

von Verteilungen eingeführt haben. Diese Frage beantworten wir durch ein Gegenbeispiel.

**Cantor-Menge und ein singuläres Wahrscheinlichkeitsmaß**
Dazu betrachten wir das Intervall $C_0 := [0,1]$. Dieses Intervall wird nun in drei Teilintervalle gleicher Länge $[0,\frac{1}{3}]$, $]\frac{1}{3},\frac{2}{3}[$ und $[\frac{2}{3},1]$ aufgeteilt. Die Menge $C_1$ wird als $C_0$ ohne den mittleren Teil definiert:

$$A_1 := \left]\frac{1}{3},\frac{2}{3}\right[,\ C_1 := C_0 \setminus A_1 = \left[0,\frac{1}{3}\right] \cup \left[\frac{2}{3},1\right].$$

Mit diesen beiden Intervallen wird analog verfahren:

$$\left[0,\frac{1}{3}\right] = \left[0,\frac{1}{9}\right] \cup \left]\frac{1}{9},\frac{2}{9}\right[ \cup \left[\frac{2}{9},\frac{1}{3}\right] \text{ und } \left[\frac{2}{3},1\right] = \left[\frac{2}{3},\frac{7}{9}\right] \cup \left]\frac{7}{9},\frac{8}{9}\right[ \cup \left[\frac{8}{9},1\right].$$

Die Menge $C_2$ besteht nun aus $C_1$ ohne die beiden mittleren Intervalle:

$$C_2 := \left[0,\frac{1}{9}\right] \cup \left[\frac{2}{9},\frac{3}{9}\right] \cup \left[\frac{6}{9},\frac{7}{9}\right] \cup \left[\frac{8}{9},1\right],\ A_2 := C_1 \setminus C_2.$$

$C_3$ wird nun wieder durch Dreiteilung der vier Teilintervalle definiert, wobei die vier mittleren Teile entfernt werden. Setzt man diese Vorgehensweise fort, so erhält man zu jedem $n \in \mathbb{N}$ zwei Mengen $C_n$ und $A_n := C_{n-1} \setminus C_n$. Da $C_n \in \mathcal{B}$ für alle $n \in \mathbb{N}$, ist auch $C := \bigcap_{n=1}^{\infty} C_n$ in $\mathcal{B}$. Wir erhalten für $\lambda(C)$:

$$\lambda(C) = 1 - \left(\frac{1}{3} + 2\cdot\frac{1}{9} + 4\cdot\frac{1}{27} + \ldots\right) = 1 - \frac{1}{3}\sum_{k=0}^{\infty}\left(\frac{2}{3}\right)^k = 0.$$

Aufgrund der Konstruktion von $C$ können wir die Elemente von $C$ angeben:

$$C = \left\{\sum_{i=1}^{\infty} a_i 3^{-i};\ a_i = 0 \text{ oder } a_i = 2\right\}.$$

Die Menge $C$ wird Cantor-Menge genannt. $C$ ist eine überabzählbare Teilmenge des Intervalls $[0,1]$ mit Lebesgue-Maß 0. Mit Hilfe der Mengen $C_n$ und $A_n$, $n \in \mathbb{N}$, konstruieren wir nun eine spezielle Funktion $F : \mathbb{R} \to [0,1]$. Die Mengen $A_n$, $n \in \mathbb{N}$, bestehen aus $2^{n-1}$ offenen Intervallen, die paarweise disjunkt sind. Diese Intervalle numerieren wir mit 1 bis $2^{n-1}$ durch, wobei ein Intervall eine kleinere Nummer als ein zweites Intervall erhält, wenn die Elemente des ersten Intervalls kleiner sind als die Elemente des zweiten Intervalls. Sei nun $A := \bigcup_{n=1}^{\infty} A_n$, dann betrachten wir die folgende Funktion

$$\tilde{F} :\ A \to [0,1],\ x \mapsto \frac{2k-1}{2^n},$$

falls $x$ Element des $k$-ten Intervalls der Menge $A_n$ ist.
Die Funktion $F : \mathbb{R} \to [0,1]$ definieren wir durch

$$x \mapsto \begin{cases} 0, & \text{falls } x \leq 0 \\ \tilde{F}(x), & \text{falls } x \in A \\ \sup\limits_{y \in A, y<x} \left\{ \tilde{F}(y) \right\}, & \text{falls } x \in C \\ 1, & \text{falls } x \geq 1 \end{cases},$$

und erhalten folgende Eigenschaften:

- $F$ ist stetig.
- $F$ ist monoton steigend.
- $F'(x) = 0$ für alle $x \in A$.
- $\lim\limits_{x \to -\infty} F(x) = 0$, $\lim\limits_{x \to \infty} F(x) = 1$.

Offensichtlich erzeugt $F$ durch $P([a,b[) := F(b) - F(a)$ ein Maß auf $\mathcal{B}$ und offensichtlich ist $P$ ein Wahrscheinlichkeitsmaß auf $\mathcal{B}$ mit Verteilungsfunktion $F$. Da $F'(x) = 0$ für alle $x \in A$, ist $P(A) = 0$ und $P(C) = 1$. Das Wahrscheinlichkeitsmaß $P$ ist nicht absolutstetig bezüglich $\lambda$ und besitzt somit keine Dichte bezüglich $\lambda$, weil $\lambda(C) = 0$ und $P(C) = 1$ ist. Wir erhalten also ein stetiges Wahrscheinlichkeitsmaß $P$ auf $\mathcal{B}$, das nicht absolutstetig bezüglich $\lambda$ ist. $P$ wird als singulär bezüglich $\lambda$ bezeichnet, da eine Menge $B \in \mathcal{B}$ existiert (in unserem Fall die Cantor-Menge $C$) mit $P(B) = 1$ und $\lambda(B) = 0$.

Nun sind wir in der Lage, die Verteilungsfunktion einer reellen Zufallsvariable in drei Komponenten zu zerlegen.

**Satz A.1.33 (Zerlegungssatz von Lebesgue).** *Sei $X$ eine reelle Zufallsvariable mit Verteilungsfunktion $F$, die auf einem Wahrscheinlichkeitsraum $(\mathbb{R}, \mathcal{B}, P)$ definiert ist, dann gibt es nichtnegative reelle Zahlen $a_1, a_2, a_3$ mit $a_1 + a_2 + a_3 = 1$ und drei Funktionen $F_1, F_2$ und $F_3$, $F_i : \mathbb{R} \to \mathbb{R}$, $i = 1,2,3$, mit:*

- $F = a_1 F_1 + a_2 F_2 + a_3 F_3$.
- *$F_1$ ist Verteilungsfunktion einer diskreten Zufallsvariable auf $(\mathbb{R}, \mathcal{B}, P)$, $F_2$ ist Verteilungsfunktion einer bezüglich $\lambda$ absolutstetigen reellen Zufallsvariable auf $(\mathbb{R}, \mathcal{B}, P)$ und $F_3$ ist Verteilungsfunktion einer stetigen reellen Zufallsvariable auf $(\mathbb{R}, \mathcal{B}, P)$, deren Bildmaß singulär bezüglich $\lambda$ ist.*

**Riemann-Integration**
Ist $P_1$ ein bezüglich $\lambda$ absolutstetiges Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$, so existiert eine Dichte $f$ mit

$$P_1(A) = \int_A f \, d\lambda, \; A \in \mathcal{B}.$$

Die Funktion $f$ ist in einem Intervall $[a, b]$, $a < b$, Riemann-integrierbar, falls sie auf diesem Intervall beschränkt ist und die Menge der Unstetigkeitsstellen von $f$ auf $[a, b]$ das Lebesgue-Maß Null hat. Sind diese Voraussetzungen an $f$ erfüllt, so können wir für jede ($P_1$-)integrierbare Funktion $g : \; [a, b] \to \mathbb{R}$ das ($P_1$-)Integral von $g$ über dem Intervall $[a, b]$ durch ein Riemann-Integral berechnen, falls $g \cdot f$ Riemann-integrierbar über $[a, b]$ ist:

$$\int_{[a,b]} g \, dP_1 = \int_{[a,b]} g \cdot f \, d\lambda = \int_a^b g(x) \cdot f(x) \, dx.$$

**Dichtefunktion**
Sei $d : \mathbb{R}^n \to \mathbb{R}$, $n \in \mathbb{N}$, eine stetige Funktion mit folgenden Eigenschaften:

- $d(x) \geq 0$ für alle $x \in \mathbb{R}^n$,
- $\int_{-\infty}^{\infty} d(x) \, dx = 1$,

dann ist auch $\int d \, d\lambda^n = 1$ und wir können die Funktion $d$ als Dichte eines Wahrscheinlichkeitsmaßes bezüglich $\lambda^n$ auffassen. Nun betrachten wir für jeden Vektor $\mu \in \mathbb{R}^n$ und für jede positiv definite Matrix $\Sigma \in \mathbb{R}^{n,n}$ die Funktion

$$\nu_{\mu,\Sigma} : \mathbb{R}^n \to \mathbb{R}, \; x \mapsto \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \cdot \exp\left( -\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2} \right).$$

Offensichtlich ist $\nu_{\mu,\Sigma}(x) > 0$ für alle $\mu, x \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n,n}$, $\Sigma$ positiv definit.
Aus der Analysis (Substitutionsregel, Satz von Fubini) ist das Folgende bekannt:

$$\int_{\mathbb{R}^n} \exp\left( -\frac{(x-\mu)^T \Sigma^{-1} (x-\mu)}{2} \right) dx = \sqrt{(2\pi)^n \det(\Sigma)}$$

für alle $\mu \in \mathbb{R}^n$, $\Sigma \in \mathbb{R}^{n,n}$, $\Sigma$ positiv definit. Somit können wir $\nu_{\mu,\Sigma}$ als Dichte eines Wahrscheinlichkeitsmaßes bezüglich $\lambda^n$ auffassen.

**Definition A.1.34 (Normalverteilung).** Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum, $\mu \in \mathbb{R}^n$, $n \in \mathbb{N}$, und $\Sigma \in \mathbb{R}^{n,n}$, $\Sigma$ positiv definit. Die Zufallsvariable $X_{\mu,\Sigma} : \Omega \to \mathbb{R}^n$ heißt $\mathcal{N}(\mu, \Sigma)$ normalverteilt, falls ihr Bildmaß $P_{X_{\mu,\Sigma}}$ bezüglich $\lambda^n$ die folgende Dichte besitzt:

$$\nu_{\mu,\Sigma} : \mathbb{R}^n \to \mathbb{R},\ x \mapsto \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \cdot \exp\left(-\frac{(x-\mu)^T \Sigma^{-1}(x-\mu)}{2}\right)$$

Um die Parameter $\mu$ und $\Sigma$ einer Normalverteilung interpretieren zu können, benötigen wir die folgende Definition.

**Definition A.1.35 (Covarianz, unkorreliert).** Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$, $Y : \Omega \to \mathbb{R}$ zwei reelle, ($P$-)integrierbare Zufallsvariable mit ($P$-)integrierbarem Produkt $X \cdot Y$, dann heißt

$$\mathbb{K}(X, Y) := \mathbb{E}((X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$$

die Covarianz von $X$ und $Y$. $X$ und $Y$ heißen unkorreliert, falls $\mathbb{K}(X, Y) = 0$.

Besitzen die Zufallsvariablen $X$ bzw. $Y$ zudem endliche Varianzen $\mathbb{V}(X)$ bzw. $\mathbb{V}(Y)$, so wird die Größe

$$\rho(X, Y) := \frac{\mathbb{K}(X, Y)}{\sqrt{\mathbb{V}(X) \cdot \mathbb{V}(Y)}}$$

Korrelationskoffizient von $X$ und $Y$ genannt.
Normalverteilte Zufallsvariablen spielen in der Wahrscheinlichkeitstheorie eine bedeutende Rolle, auf die wir im Zusammenhang mit dem zentralen Grenzwertsatz noch zu sprechen kommen. Zunächst fassen wir einige Eigenschaften einer $\mathcal{N}(\mu, \Sigma)$ normalverteilten Zufallsvariablen $X_{\mu,\Sigma}$ zusammen. Dazu fassen wir die Funktion $X_{\mu,\Sigma} : \Omega \to \mathbb{R}^n$ als Abbildung

$$\omega \mapsto \left(X^1_{\mu,\Sigma}(\omega), \ldots, X^n_{\mu,\Sigma}(\omega)\right)^T$$

auf. Jede Funktion $X^i_{\mu,\Sigma} : \Omega \to \mathbb{R}$, $i = 1, \ldots, n$, ist eine reelle Zufallsvariable. Definiert man

$$\mathbb{E}(X_{\mu,\Sigma}) := \left(\mathbb{E}(X^1_{\mu,\Sigma}), \ldots, \mathbb{E}(X^n_{\mu,\Sigma})\right)^T,$$

so erhält man

$$\mathbb{E}(X_{\mu,\Sigma}) = \mu.$$

Ferner gilt mit $\Sigma = (\sigma_{i,j})_{i,j=1,\ldots,n}$:

$$\mathbb{K}(X^i_{\mu,\Sigma}, X^j_{\mu,\Sigma}) = \sigma_{i,j}, \quad i, j = 1, \ldots, n.$$

Daher heißt $\Sigma$ die Covarianzmatrix von $X_{\mu,\Sigma}$.
Auf der Basis eines Wahrscheinlichkeitsraumes $(\Omega, \mathcal{S}, P)$ haben wir für $A, B \in$

$\mathcal{S}$ und $P(B) > 0$ durch $P^B(A) = \frac{P(A \cap B)}{P(B)}$ ein Wahrscheinlichkeitsmaß auf $\mathcal{S}$ eingeführt. Wir interpretierten $P^B(A)$ als die Wahrscheinlichkeit von $A$ unter der Bedingung, dass $B$ ($P^B$-)fast sicher eintrifft. Nun stellt sich die Frage, wann diese Bedingung die Wahrscheinlichkeit für $A$ nicht ändert, wann also $P^B(A) = P(A)$ gilt. Wir erhalten:

$$P^B(A) = P(A) \iff P(A \cap B) = P(A) \cdot P(B).$$

**Definition A.1.36 (stochastisch unabhängige Ereignisse).** Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $A_1, \ldots, A_n \in \mathcal{S}$, $n \in \mathbb{N}$, dann heißen die Ereignisse $A_1, \ldots, A_n$ stochastisch unabhängig, falls für alle $k \in \mathbb{N}$, $k \leq n$, und für alle $i_j \in \mathbb{N}$, $1 \leq j \leq k$, mit $1 \leq i_1 < \ldots < i_k \leq n$ gilt:

$$P\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} P(A_{i_j}).$$

Die stochastische Unabhängigkeit einer Menge $\{A_i \in \mathcal{S};\, i \in I\}$, $I \neq \emptyset$, von Ereignissen führt man auf die stochastische Unabhängigkeit ihrer endlichen Teilmengen zurück.

**Definition A.1.37 (stochastische Unabhängigkeit einer Menge von Ereignissen).** Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $\{A_i \in \mathcal{S};\, i \in I\}$, $I \neq \emptyset$, eine Menge von Ereignissen, dann heißen diese Ereignisse stochastisch unabhängig, falls $A_{i_1}, \ldots, A_{i_n}$ für jedes $n \in \mathbb{N}$ mit $n \leq |I|$ und für jede Menge $\{i_1, \ldots, i_n\} \subseteq I$ stochastisch unabhängig sind.

Um stochastisch unabhängige Zufallsvariable definieren zu können, wird zunächst die stochastische Unabhängigkeit von Mengensystemen betrachtet.

**Definition A.1.38 (stochastische Unabhängigkeit von Mengensystemen).** Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $\{\mathcal{F}_i \subseteq \mathcal{S};\, i \in I\}$, $I \neq \emptyset$, eine Menge von Mengensystemen über $\Omega$, dann heißen diese Mengensysteme stochastisch unabhängig, falls für jedes $n \in \mathbb{N}$ mit $n \leq |I|$ und für jedes $\{i_1, \ldots, i_n\} \subseteq I$ die $n$ Ereignisse $A_{i_1}, \ldots, A_{i_n}$ für beliebige $A_{i_k} \in \mathcal{F}_{i_k}$, $i = 1, \ldots, n$, stochastisch unabhängig sind.

Nun betrachten wir einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$, einen Messraum $(\Omega', \mathcal{S}')$ und eine Zufallsvariable $X : \Omega \to \Omega'$. Mit $\mathcal{F}$ bezeichnen wir die Menge aller $\sigma$-Algebren über $\Omega$, für die gilt: $X$ ist $\mathcal{C}$-$\mathcal{S}'$-messbar genau dann, wenn $\mathcal{C} \in \mathcal{F}$. Die Menge $\sigma(X) := \bigcap_{\mathcal{C} \in \mathcal{F}} \mathcal{C}$ ist ebenfalls eine $\sigma$-Algebra und wird die von $X$ erzeugte $\sigma$-Algebra genannt. Unter allen $\sigma$-Algebren $\mathcal{A}$ über $\Omega$ ist $\sigma(X)$ die kleinste, für die $X$ $\mathcal{A}$-$\mathcal{S}'$-messbar ist. Somit sind wir in der Lage, die stochastische Unabhängigkeit von Zufallsvariablen in naheliegender Weise durch die stochastische Unabhängigkeit von speziellen Mengensystemen zu definieren.

**Definition A.1.39 (stochastische Unabhängigkeit von Zufallsvariablen).** Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum, $(\Omega', \mathcal{S}')$ ein Messraum und $\{X_i : \Omega \to \Omega';\, i \in I\}$, $I \neq \emptyset$, eine Menge von Zufallsvariablen, dann heißen diese Zufallsvariablen stochastisch unabhängig, falls die Mengensysteme $\{\sigma(X_i);\, i \in I\}$ stochastisch unabhängig sind.

Die stochastische Unabhängigkeit von Zufallsvariablen ist ein zentraler Begriff der Wahrscheinlichkeitstheorie und im wesentlichen Bestandteil der Modellierung zu untersuchender Vorgänge.
Da eine Folge von reellen Zufallsvariablen eine Funktionenfolge ist, betrachtet man - wie in der Analysis (z.B. gleichmäßige- und punktweise Konvergenz) - auch in der Wahrscheinlichkeitstheorie verschiedene Konvergenzbegriffe.

**Definition A.1.40 (verschiedene Konvergenzbegriffe für Folgen reeller Zufallsvariablen).** Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum, $\{X_i\}$, $i \in \mathbb{N}$, eine Folge reeller Zufallsvariable $X_i : \Omega \to \mathbb{R}$, $i \in \mathbb{N}$, und $X : \Omega \to \mathbb{R}$ ebenfalls eine reelle Zufallsvariable, dann konvergiert $\{X_i\}$, $i \in \mathbb{N}$, definitionsgemäß

- im $r$-ten Mittel ($r \in \mathbb{R}^+$) gegen $X$ genau dann, wenn

  $$\int |X_i|^r \, dP < \infty \text{ für alle } i \in \mathbb{N}, \int |X|^r \, dP < \infty \text{ und } \lim_{i\to\infty} \int |X_i - X|^r \, dP = 0,$$

- stochastisch gegen $X$ genau dann, wenn für alle $\epsilon > 0$

  $$\lim_{i\to\infty} P\left(\{\omega \in \Omega;\, |X_i(\omega) - X(\omega)| < \epsilon\}\right) = 1,$$

- mit Wahrscheinlichkeit 1 gegen $X$ genau dann, wenn

  $$P\left(\left\{\omega \in \Omega;\, \lim_{i\to\infty} X_i(\omega) = X(\omega)\right\}\right) = 1,$$

- in Verteilung gegen $X$ genau dann, wenn

  $$\lim_{i\to\infty} \int f \, dP_{X_i} = \int f \; dP_X$$

  für alle beliebig oft differenzierbaren Funktionen $f : \mathbb{R} \to \mathbb{R}$ mit kompaktem Träger.

Die stochastische Konvergenz von $\{X_i\}$, $i \in \mathbb{N}$, gegen $X$ wird oft durch

$$(P\text{-}) \lim_{i\to\infty} X_i = X,\ st\text{-} \lim_{i\to\infty} X_i = X \text{ oder } X_i \to X \text{ nach Wahrscheinlichkeit}$$

dargestellt. Die Konvergenz mit Wahrscheinlichkeit 1 von $\{X_i\}$, $i \in \mathbb{N}$, gegen $X$ heißt auch ($P$-)fast sichere Konvergenz und wird durch $X_i \to X$ ($P$-)f.s. dargestellt. Die Konvergenz nach Verteilung wird auch als schwache Konvergenz bezeichnet. Die folgenden Implikationen lassen sich leicht nachweisen.

| | | | | |
|---|---|---|---|---|
| schwache Konvergenz | $\Leftarrow$ | stochastische Konvergenz | $\Leftarrow$ | Konvergenz im $r$-ten Mittel |
| | | $\Uparrow$ | | |
| | | Konvergenz mit Wahrscheinlichkeit 1 | | |

Ausgehend von einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{S}, P)$ betrachten wir spezielle Folgen $\{X_i\}$, $i \in \mathbb{N}$, von reellen Zufallsvariablen $X_i : \Omega \to \mathbb{R}$, $i \in \mathbb{N}$, deren Quadrate $X_i^2 : \Omega \to \mathbb{R}$, $\omega \mapsto X_i^2(\omega)$ für alle $i \in \mathbb{N}$ ($P$-)integrierbar sind. Wegen

$$\int\limits_{\Omega} |X_i|\, dP = \int\limits_{\{\omega\in\Omega;\, |X_i(\omega)|\leq 1\}} |X_i|\, dP + \int\limits_{\{\omega\in\Omega;\, |X_i(\omega)|>1\}} |X_i|\, dP$$

$$\leq 1 + \int\limits_{\{\omega\in\Omega;\, |X_i(\omega)|>1\}} |X_i|\, dP \;\leq\; 1 + \int\limits_{\Omega} X_i^2\, dP \quad \text{für alle } i \in \mathbb{N}$$

besitzen die Zufallsvariablen $X_i$, $i \in \mathbb{N}$, endliche Erwartungswerte. Wir vereinbaren, dass für die Folge $\{X_i\}$, $i \in \mathbb{N}$, genau dann der zentrale Grenzwertsatz gilt, wenn die Folge $\{T_i\}$, $i \in \mathbb{N}$, standardisierter reeller Zufallsvariablen

$$T_i : \Omega \to \mathbb{R},\ \omega \mapsto \frac{\sum\limits_{j=1}^{i} (X_j - \mathbb{E}(X_j))}{\sqrt{\mathbb{V}(\sum\limits_{j=1}^{i} X_j)}}, \quad i \in \mathbb{N},$$

in Verteilung gegen eine $\mathcal{N}(0,1)$ normalverteilte Zufallsvariable konvergiert.

**Satz A.1.41 (Der zentrale Grenzwertsatz für stoch. unabh., identisch vert. Zufallsvariablen).** *Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $\{X_i\}$, $i \in \mathbb{N}$, eine Folge stochastisch unabhängiger, identisch verteilter (d.h. $P_{X_i} = P_{X_j}$ für alle $i, j \in \mathbb{N}$) reeller Zufallsvariablen $X_i : \Omega \to \mathbb{R}$ mit $0 < \mathbb{V}(X_i) < \infty$ für alle $i \in \mathbb{N}$, dann gilt für $\{X_i\}$, $i \in \mathbb{N}$, der zentrale Grenzwertsatz.*

**Satz von de Moivre-Laplace**
Besteht im obigen Satz die Folge $\{X_i\}$, $i \in \mathbb{N}$, aus stochastisch unabhängigen, $\mathrm{B}(1,p)$ binomialverteilten Zufallsvariablen, so wird die Gültigkeit des zentralen Grenzwertsatzes für $\{X_i\}$, $i \in \mathbb{N}$, als Satz von de Moivre-Laplace bezeichnet. In diesem Fall ist $X_1 + \ldots + X_n$, $n \in \mathbb{N}$, $\mathrm{B}(n,p)$ binomialverteilt und die für große $n$ aufwendig zu berechnende Binomial-Verteilung lässt sich somit durch die häufig tabellierte $\mathcal{N}(0,1)$ Normalverteilung approximieren.

Abschließend betrachten wir ein sehr hilfreiches Resultat

**Satz A.1.42 (Ungleichung von Chebyschev-Markov).** *Seien $(\Omega, \mathcal{S}, P)$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \bar{\mathbb{R}}$ eine numerische Zufallsvariable, dann gilt für jedes Paar reeller Zahlen $\alpha > 0$, $\kappa > 0$ die folgende Ungleichung von Chebyschev-Markov*

$$P\left(\{\omega \in \Omega;\, |X(\omega)| \geq \alpha\}\right) \leq \frac{1}{\alpha^{\kappa}} \int |X|^{\kappa}\, dP.$$

# Anhang B

# Topologie

In diesem Anhang stellen wir ein paar im Laufe der Arbeit benötigte Grundlagen aus der (mengentheoretischen) Topologie zusammen. Für eine ausführlichere Darstellung verweisen wir auf [StZi88], [Qu79], [Br93] und [May89], die auch als Quelle für diesen Anhang dienten.

In metrischen Räumen haben wir einen Abstandsbegriff zur Verfügung, den wir intuitiv verstehen. Dementsprechend haben wir auch eine Vorstellung davon, was Umgebungen, offene und abgeschlossene Teilmengen in metrischen Räumen sein sollen. Der Begriff der Topologie verallgemeinert dieses Konzept:

**Definition B.1.1 (Topologie, offen, abgeschlossen).** Eine Topologie $\mathcal{T}$ auf einer Menge $X$ ist eine Teilmenge der Potenzmenge von $X$ mit folgenden Eigenschaften:

(i) $\emptyset, X \in \mathcal{T}$.

(ii) $U_i \in \mathcal{T}, i \in I \Rightarrow \bigcup_{i \in I} U_i \in \mathcal{T}$.

(iii) $U, V \in \mathcal{T} \Rightarrow U \cap V \in \mathcal{T}$.

Die Elemente von $\mathcal{T}$ heißen offene Mengen. Eine Teilmenge $A$ heißt abgeschlossen, falls ihr Komplement $A^c$ offen ist.

Das Tupel $(X, \mathcal{T})$ heißt topologischer Raum. In der Regel wird die Topologie nicht explizit genannt, man spricht von dem topologischen Raum $X$.

**Beispiel B.1.2.** (i) Ist $X$ eine beliebige Menge, so ist die Potenzmenge $\mathcal{P}(X)$ von $X$ selbst eine Topologie. Diese Topologie heißt diskrete Topologie, und der topologische Raum $(X, \mathcal{P}(X))$ heißt diskreter Raum. In einem diskreten Raum ist jede Teilmenge offen und abgeschlossen.

(ii) Sei $X$ eine beliebige Menge. Dann ist durch $\mathcal{T} := \{\emptyset, X\}$ eine Topologie auf $X$ definiert. Diese Topologie heißt indiskrete Topologie.

(iii) Ist $(X, d)$ ein metrischer Raum mit Metrik $d : X \times X \to \mathbb{R}$, so definieren wir für jedes $\varepsilon > 0$ und $x \in X$ die $\varepsilon$-Umgebung von $x$ durch

$$B(x, \varepsilon) := \{y \in X : d(x, y) < \varepsilon\}. \tag{B.1}$$

Die Metrik induziert folgendermaßen die sogenannte metrische Topologie $\mathcal{T}_d$: Es gilt $U \in \mathcal{T}_d$ genau dann, wenn für jedes $x \in U$ ein $\varepsilon > 0$ existiert, so dass $B(x, \varepsilon) \subset U$.

(iv) Für $X = \mathbb{R}^n$ wird durch die euklidische Norm eine Metrik $d$ definiert. Explizit ist sie gegeben durch:

$$d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \tag{B.2}$$

$$(x, y) \mapsto \left( \sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}}. \tag{B.3}$$

Die mit dieser (euklidischen) Metrik definierte metrische Topologie auf dem $\mathbb{R}^n$ heißt Standardtopologie. Der $\mathbb{R}^n$ ist im folgenden immer mit der Standardtopologie versehen.

Der $\mathbb{R}^n$ wird für uns das wichtigste Beispiel eines topologischen Raumes sein. Wir wollen die gerade definierte euklidische Metrik noch nutzen, um den Abstand $d(x, N)$ eines Punktes $x \in \mathbb{R}^n$ zu einer nicht-leeren Teilmenge $N \subset \mathbb{R}^n$ zu definieren. Dieser ist naheliedender Weise gegeben durch

$$d(x, N) := \inf\{d(x, y) | y \in N\}. \tag{B.4}$$

Um weitere topologische Räume betrachten zu können, führen wir nun eine Klasse von Abbildungen ein, die im gewissen Sinne die Struktur erhalten.

**Definition B.1.3 (stetige Abbildung).** Seien $(X, \mathcal{T})$ und $(Y, \mathcal{S})$ topologische Räume. Eine Abbildung $f : (X, \mathcal{T}) \to (Y, \mathcal{S})$ heißt stetig, wenn für jede offene Teilmenge $U$ von $Y$ das Urbild $f^{-1}(U)$ offen ist in $X$.

Da es im allgemeinen klar ist, welche Topologien gemeint sind, schreibt man in der Regel einfach $f : X \to Y$. Auch ist die Eigenschaft der Stetigkeit für die Topologie so grundlegend, dass sie implizit immer vorausgesetzt wird. Unter einer Abbildung $f : X \to Y$ verstehen wir also fortan eine stetige Abbildung $f : (X, \mathcal{T}) \to (Y, \mathcal{S})$.

Bei stetigen Abbildungen sind Urbilder offener Mengen wieder offen. Die Umkehrung gilt im allgemeinen nicht, vielmehr wird dafür ein eigener Name verwendet:

**Definition B.1.4 (offene Abbildung).** Eine Abbildung $f : X \to Y$ heißt offen (abgeschlossen), falls das Bild $f(U)$ einer offenen (abgeschlossenen) Teilmenge von $X$ offen (abgeschlossen) in $Y$ ist.

Schließlich werden bijektive Abbildungen, die in beide Richtungen die topologische Struktur erhalten, ausgezeichnet:

**Definition B.1.5 (Homöomorphismus).** Sei $f : X \to Y$ eine Abbildung. $f$ heißt Homöomorphismus, falls $f$ bijektiv und sowohl $f$ als auch $f^{-1}$ stetig sind. Zwei topologische Räume $X$ und $Y$ heißen topologisch äquivalent oder homöomorph, falls es einen Homöomorphismus $f : X \to Y$ gibt.

Homöomorphie ist eine Äquivalenzrelation (s. Definition B.1.9) auf den topologischen Räumen. Im wesentlichen unterscheiden sich homöomorphe Räume nicht, es gibt eine bijektive Abbildung, die sowohl die Räume als auch die Strukturen aufeinander abbildet. Vom topologischen Standpunkt kennt man mit einem Repräsentanten einer Homöomorphieklasse alle übrigen Vertreter. Es stellt sich in natürlicher Weise das Problem, einerseits einen Überblick über alle Homöomorphieklassen zu gewinnen, andererseits von zwei gewählten Räumen zu entscheiden, ob sie homöomorph sind oder nicht. Dieses Problem heißt Homöomorphieproblem. So lässt sich z.B. die Frage nach der Homöomorphie von $\mathbb{R}^n$ und $\mathbb{R}^m$ für $m \neq n$ sehr elegant im Rahmen der algebraischen Topologie beantworten.

Die Zahl der von uns bisher betrachteten topologischen Räume ist klein. Wir wollen jetzt Methoden vorstellen, wie man aus topologischen Räumen neue generieren kann. Dazu verwendet man Konstruktionen aus der naiven Mengenlehre: Teilmengen, Summen, Produkte, Quotientenmengen. Es stellt sich dabei lediglich die Frage, wie man auf die mengentheoretische Konstruktion eine topologische Struktur induziert. Dies geschieht in der Regel durch eine natürlich gegebene Abbildung. Wir beginnen mit der Betrachtung von Teilmengen.

**Definition und Satz B.1.6 (Teilraumtopologie).** Sei $(X, \mathcal{T})$ ein topologischer Raum und $A \subset X$. Dann ist $\mathcal{T}_A := \{U \cap A | U \in \mathcal{T}\}$ eine Topologie auf $A$. $\mathcal{T}_A$ heißt Teilraumtopologie, $A$ Teilraum von $X$.

Immer wiederkehrend ist das folgende Prinzip: Zu einer Teilmenge gehört in der Kategorie der Mengen in natürlicher Weise die Inklusionsabbildung. Die Topologie auf dem Teilraum ist nun gerade so definiert, dass die natürliche Abbildung, in diesem Fall die Inklusion, stetig ist.

**Satz B.1.7.** *Sei $X$ ein topologischer Raum und $A$ ein Teilraum von $X$. Dann ist die Inklusion*

$$i : A \hookrightarrow X, \tag{B.5}$$

$$a \mapsto a, \tag{B.6}$$

*stetig.*

Analog geht man zur Bildung der topologischen Summe und des topologischen Produkts vor. Da wir diese Konstuktionen im weiteren Verlauf nicht brauchen, verzichten wir auf die technisch etwas aufwändigeren Details. Stattdessen wenden wir uns dem Begriff der Quotiententopologie zu.

**Definition und Satz B.1.8 (Quotiententopologie).** Sei $(X, \mathcal{T})$ ein topologischer Raum und $f : X \to Y$ eine surjektive Abbildung. Dann ist durch $\mathcal{S} := \{U \subset Y | f^{-1}(U) \in \mathcal{T}\}$ eine Topologie auf $Y$ definiert. Es ist die feinste Topologie auf $Y$, für die $f$ stetig ist. $\mathcal{S}$ heißt Quotiententopologie.

Oft wird die Quotiententopologie durch eine Äquivalenzrelation gegeben.

**Definition B.1.9 (Äquivalenzrelation).** Auf einer Menge $X$ ist eine Äquivalenzrelation $R$ gegeben durch eine Relation $\sim_R$ mit den folgenden Eigenschaften für jedes $x, y, z \in X$ :

(i) $x \sim_R x$ (Reflexivität)

(ii) $x \sim_R y \;\Rightarrow\; y \sim_R x$ (Symmetrie)

(iii) $x \sim_R y,\; y \sim_R z \;\Rightarrow\; x \sim_R z.$ (Transitivität)

Eine Äquivalenzklasse besteht aus paarweise in Relation stehenden Elementen von $X$. Die Äquivalenzklassen bilden eine Partition von $X$, d.h. eine disjunkte Zerlegung in nichtleere Teilmengen, deren Vereinigung die ganze Menge $X$ ist. Es gilt auch die Umkehrung: Jede Partition definiert in eindeutiger Weise eine Äquivalenzrelation. Die Menge der Äquivalenzklassen einer Äquivalenzrelation $R$ bezeichnet man mit $X/R$ oder $X/\sim_R$, die Äquivalenzklasse, die das Element $x$ enthält, mit $[x]$. Es gibt eine natürliche Abbildung,

$$\pi : X \;\to\; X/R, \tag{B.7}$$
$$x \;\mapsto\; [x], \tag{B.8}$$

die man als kanonische Projektion bezeichnet. Die kanonische Projektion ist offensichtlich immer surjektiv.

**Beispiel B.1.10.** Sei $X$ ein topologischer Raum und $R$ eine Äquivalenzrelation auf $X$. Durch die kanonische Projektion

$$\pi : X \;\to\; X/R, \tag{B.9}$$
$$x \;\mapsto\; [x], \tag{B.10}$$

ist eine surjektive Abbildung gegeben, die auf $X/R$ eine Quotiententopologie definiert. Der so entstandene topologische Raum $X/R$ heißt Quotientenraum von $X$ nach $R$.

Am häufigsten wird eine Äquivalenzrelation durch eine Gruppenaktion gegeben:

**Definition B.1.11 (Gruppenaktion).** Sei $X$ ein topologischer Raum und $G$ eine Gruppe. Eine Aktion von $G$ auf $X$ ist eine Abbildung $\alpha : G \times X \to X$, abkürzend schreibt man $gx$ für $\alpha(g, x)$ mit folgenden Eigenschaften:

(i) $h(gx) = (hg)x$ für alle $x \in X,\ g, h \in G$.

(ii) $ex = x$ für alle $x \in X$ und das neutrale Element $e \in G$.

Mit Hilfe der Gruppenaktion wird auf $X$ eine Relation $R$ definiert: $x \sim_R y :\Leftrightarrow$ es gibt ein $g \in G$ mit $gx = y$. Man überzeugt sich sofort, dass es sich dabei um eine Äquivalenzrelation handelt. Für den Quotientenraum schreibt man $X/G$.

**Beispiel B.1.12 (reelle projektive Räume).** Sei $S^n$ die Einheitssphäre im $\mathbb{R}^{n+1}$, also

$$S^n := \{x \in \mathbb{R}^{n+1} | \|x\|_2 = 1\}. \tag{B.11}$$

Die 0-dimensionale Einheitssphäre $S^0 = \{\pm 1\}$ ist bezüglich der Multiplikation eine Gruppe. Eine Gruppenaktion von $S^0$ auf $S^n$ ist gegeben durch:

$$\alpha : S^0 \times S^n \to S^n, \tag{B.12}$$

$$(g, x) \mapsto (gx_0, \ldots, gx_n). \tag{B.13}$$

Der Quotientenraum $S^n/S^0$ heißt $n$-dimensionaler projektiver Raum und wird mit $RP^n$ bezeichnet.

Abschließend wollen wir noch auf den Begriff des Zusammenhangs eingehen, der unmittelbar der Anschauung entnommen ist:

**Definition B.1.13 (Zusammenhang).** Ein topologischer Raum $X$ heißt zusammenhängend, falls aus $X = A \dot\cup B$, $A, B$ offen folgt, dass $A = \emptyset$ oder $B = \emptyset$. Eine Teilmenge $C \subset X$ heißt zusammenhängend, wenn der Teilraum $C$ zusammenhängend ist.

Mit anderen Worten ist ein topologischer Raum zusammenhängend, wenn er sich nicht disjunkt in zwei nichtleere offene Teilmengen zerlegen lässt.

**Beispiel B.1.14.** In $\mathbb{R}$ sind die zusammenhängenden Teilmengen genau die Intervalle.

So klar diese Aussage anschaulich sein mag, so erfordert ihr Beweis doch etwas Mühe. Zusammenhang verhält sich wohlwollend unter stetigen Abbildungen:

**Satz B.1.15.** *Ist $f : X \to Y$ stetig und $A \subset X$ zusammenhängend, so auch $f(A)$.*

**Definition B.1.16 (Zusammenhangskomponente).** Sei $X$ ein topologischer Raum und $x \in X$. Dann ist die Zusammenhangskomponente $K(x)$ von $x$ definiert als

$$K(x) := \bigcup_{\substack{x \in C \subset X \\ C \text{ zusammh.}}} C. \tag{B.14}$$

Ein topologischer Raum heißt total unzusammenhängend, wenn für alle $x \in X$ gilt: $K(x) = x$.

Die Zusammenhangskomponente $K(x)$ ist selbst zusammenhängend und damit die größte zusammenhängende Teilmenge von $X$, die $x$ enthält. Der Begriff des totalen Unzusammenhangs scheint zunächst nicht sehr anschaulich, jedoch sind z.B. $\mathbb{Q} \subset \mathbb{R}$ total unzusammenhängend oder auch die aus der Zahlentheorie bekannten p-adischen Zahlen.

**Satz B.1.17.** *Ist $f : X \to Y$ ein Homöomorphismus, so induziert $f$ eine bijektive Abbildung zwischen den Zusammenhangskomponenten.*

Dieser ganz einfache Satz ist schon in der Lage, einen ersten Beitrag zum Homöomorphie-Problem zu leisten. Sollen zwei Räume homöomorph sein, so muss zumindest die Zahl ihrer Zusammenhangskomponenten gleich sein. So können z.B. die Räume $\mathbb{R} \setminus \{0, 1\}$ und $\mathbb{R} \setminus \{2, 3, 4\}$ nicht homöomorph sein.

# Literaturverzeichnis


[Ba94] F. L. Bauer, *Kryptologie.* Springer-Verlag, 2. Auflage 1994.

[Bau90] H. Bauer, *Maß- und Integrationstheorie.* de Gruyter-Verlag, 1990.

[Bau91] H. Bauer, *Wahrscheinlichkeitstheorie.* de Gruyter-Verlag, 4. Auflage 1991.

[Be93] A. Beutelspacher, *Kryptologie.* Vieweg-Verlag, 3. Auflage 1993.

[Bi86] P. Billingsley, *Probability and Measure.* Wiley, 2. Auflage 1986.

[Boe93] J. F. Böhme, *Stochastische Signale.* Teubner-Verlag, 1993.

[Bo92] M. Bossert, *Kanalcodierung.* Teubner-Verlag, 1992.

[Br93] G. E. Bredon, *Topology and Geometry.* Springer-Verlag, 1993.

[Fr96] B. Friedrichs, *Kanalcodierung.* Springer-Verlag, 1996.

[FuRi94] W. Fumy und H. P. Rieß, *Kryptographie.* Oldenbourg-Verlag, 2. Auflage 1994.

[Ha91] E. Hänsler, *Statistische Signale.* Springer-Verlag, 1991.

[HaOfPa96] J. Hagenauer, E. Offer und L. Papke, *Iterative Decoding of Binary Block and Convolutional Codes.* IEEE Transactions on Information Theory **42**, 1996.

[HaRu76] C. R. P. Hartmann und L. D. Rudolph, *An Optimum Symbol-by-Symbol Decoding Rule for Linear Codes.* IEEE Transactions on Information Theory **22**, 1976.

[HaTh94] W. Hackenbroch und A. Thalmaier, *Stochastische Analysis.* Teubner-Verlag, 1994.

[HeQu95] W. Heise und P. Quattrocchi, *Informations- und Codierungstheorie.* Springer-Verlag, 3. Auflage 1995.

[Ju95] D. Jungnickel, *Codierungstheorie.* Spektrum Akademischer Verlag, 1995.

[Kam96] K. D. Kammeyer, *Nachrichtenübertragung.* Teubner-Verlag, 2. Auflage, 1996.

[KaSh91] I. Karatzas und S. E. Shreve, *Brownian Motion and Stochastic Calculus.* Springer-Verlag, 2. Auflage 1991.

[Lü95] H. D. Lüke, *Signalübertragung.* Springer-Verlag, 6. Auflage 1995.

[Ma95] P. Malliavin, *Integration and Probability.* Springer-Verlag, 1995.

[Mat] Matlab, Version 6.1, Release 12.1. The MathWorks, www.mathworks.com.

[May89] K. H. Mayer, *Algebraische Topologie.* Birkhäuser-Verlag, 1989.

[Pr00] J. G. Proakis, *Digital Communications.* McGraw-Hill, 4. Auflage 2000.

[Qu79] B. von Querenburg, *Mengentheoretische Topologie.* Springer-Verlag, 2. Auflage, 1979.

[ReYo91] D. Revuz und M. Yor, *Continuous Martingales and Brownian Motion.* Springer-Verlag, 1991.

[Ro95] H. Rohling, *Einführung in die Informations- und Codierungstheorie.* Teubner-Verlag, 1995.

[Rot92] S. Rotman, *Coding and Information Theory.* Springer-Verlag, 1992.

[Sc97] S. Schäffler, *Decodierung binärer linearer Blockcodes durch globale Optimierung.* Roderer-Verlag, 1997.

[ScSt94] S. Schäffler und T. F. Sturm, *Wahrscheinlichkeitstheorie und Statistik I für Mathematiker.* Skriptenreihe des IAMS, Band 5, TU München, 1994.

[ScSt95] S. Schäffler und T. F. Sturm, *Wahrscheinlichkeitstheorie und Statistik II für Mathematiker.* Skriptenreihe des IAMS, Band 6, TU München, 1995.

[Ši88] A. N. Širjaev, *Wahrscheinlichkeit.* VEB-Verlag, 1988.

[StZi88] R. Stöcker, H. Zieschang, *Algebraische Topologie.* Teubner-Verlag, 1988.

[Ung82] G. Ungerböck, *Channel coding with multilevel/phase signals.* IEEE Trans. on Inf. Theory, vol. IT-28, pp. 55-67, Januar 1982.