

2 Grundlagen

Lernziele:

1. Das Kapitel liefert die benötigten Grundlagen aus den Gebieten Kommunikations-, Nachrichtentechnik und Eingebettete Systeme.
2. Der Abschnitt „Kommunikationstechnik“ bespricht außer der Quellen- und Kanalcodierung Verfahren zur Verschlüsselung und die Netzwerktechnik mit Topologien, das ISO/OSI-Referenzmodell und die Kanalzugriffsverfahren.
3. Der Abschnitt „Nachrichtentechnik“ stellt die klassische Nachrichtenübertragungstechnik dar. Hierzu gehören die Modulationsverfahren, die Funktechnik und die Betriebsarten.
4. Der Abschnitt „Eingebettete Systeme“ erläutert deren Entwicklung und die hierfür benötigte Mikroprozessortechnik. Hierbei liegt der Fokus auf den Eingebetteten Funksystemen mit Energiesparkonzepten und Konzepten für die Echtzeit-Datenverarbeitung.

Das Kapitel Grundlagen stellt die drei „Säulen“ Kommunikations-, Nachrichtentechnik und Eingebettete Systeme der „Wireless-Netzwerke für den Nahbereich“ vor.

Es liefert das theoretische Basiswissen für die drahtlosen Verfahren (siehe Kapitel 3). Die „Stellschrauben“ (Parameter) für den anschließenden Vergleich der drahtlosen Verfahren (siehe Kapitel 4) werden ebenfalls in diesem Kapitel näher erläutert.

2.1 Einführung

Die Fähigkeit zu kommunizieren ist ein wesentliches Merkmal eines Lebewesens. Die Natur hat im Laufe der Evolution äußerst vielfältige und komplexe Systeme zur Kommunikation entwickelt.

2.1.1 Was heißt Kommunikation?

Auf den Menschen (und andere höher entwickelte Lebewesen) bezogen, ist das dort meist eingesetzte System die akustische Kommunikation. Der Grund dafür liegt sicher darin, dass das notwendige Übertragungsmedium Luft in der Regel im natürlichen Lebensraum überall vorhanden ist. Dieses System funktioniert auch bei Hindernissen, wenn andere, zum Beispiel optische Kommunikationswege, ausfallen (Nacht, Nebel, Hindernisse).

Definition: Kommunikation

(lat.: Austausch, Verständigung, Übermittlung und Vermittlung von Wissen) bedeutet im ursprünglichen Sinn gemeinschaftliche Teilhabe an Information. Im weiteren Sinne sind damit alle Prozesse der Übertragung von Nachrichten oder Informationen durch Lebewesen und/oder Maschinen gemeint.

Funktionen

Neben der eigentlichen Aufgabe, der Übermittlung von Nachrichten, erlaubt und realisiert ein Kommunikationssystem auf akustischer Basis eine Fülle anderer Funktionen. Möglich sind:

- die Identifizierung des Gesprächspartners
- eine Konferenz zwischen mehreren Gesprächspartnern
- ein Rundspruch an mehrere Empfänger
- die Ortung des Gesprächspartners
- die eigene Navigation
- eine adaptive Anpassung an einen vorhandenen Störpegel
- in gewissen Grenzen eine Resistenz gegen Störungen
- die Ausblendung von Störungen
- die Verschlüsselung von Informationen
- eine räumlich gerichtete Kommunikation
- die Übertragung auch feinsten Gefühlsregungen.

Diese Aufzählung ist sicher nicht vollständig. Sie zeigt jedoch die Komplexität natürlicher Kommunikationssysteme und ist zugleich auch ein Anforderungskatalog an moderne drahtlose Kommunikationssysteme.

2.1.2 Struktur eines natürlichen Kommunikationssystems

Jeder Kommunikationsprozess besteht aus mindestens drei Komponenten: einem Sender, einer Nachricht und einem Empfänger.

Abbildung 2.1 zeigt eine wohl jedem bekannte Form der Kommunikation: Ein Lehrer artikuliert einen mathematischen Lehrsatz, seine Schüler hören zu und verstehen ihn (hoffentlich). Dieses einfache Beispiel zeigt bereits die Komplexität der Abläufe in einem Kommunikationssystem. Abstrahiert man das oben gezeigte Beispiel einer zwischenmenschlichen Kommunikation, kommt man zu folgendem grundlegenden Schema eines in eine Richtung durchgeführten Kommunikationsprozesses (siehe Abbildung 2.2). Das Ziel dieses Kommunikationsprozesses ist es, dass die Idee (die Intention, eine Information) im Kopf des Lehrers dem Schüler übermittelt und in dessen Gedächtnis eingeprägt wird (die Orientierung). Damit das auch funktioniert, müssen eine Reihe von Voraussetzungen erfüllt werden (siehe Tabelle 2.1).

Kommunikationsprozess



Abbildung 2.1: Elementarer Kommunikationsprozess

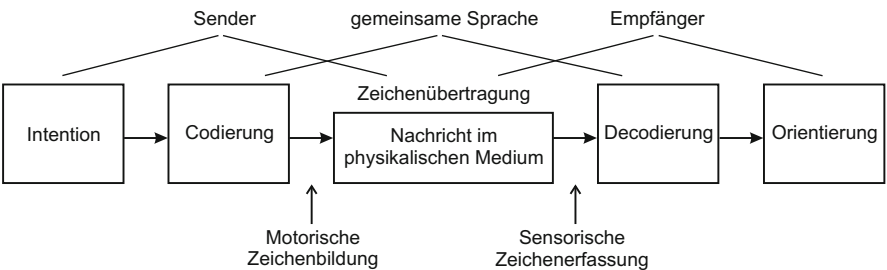


Abbildung 2.2: Grundlegendes Schema eines Kommunikationssystems

Bereich	Tätigkeit
	(A) Die Intention muss verbalisiert, das heißt in eine Sprache umgesetzt werden, das heißt, sie muss codiert werden.
Senderseite	(B) Die Sprache muss gesprochen werden. Im Falle eines akustischen Systems müssen die Stimmbänder aktiviert werden, damit Luftdruckschwankungen erzeugt werden, die im Mund zu Lauten geformt werden.
Übertragungsmedium	(C) Über das Medium „Luft“ werden die Luftdruckschwankungen weitergeleitet.
Empfangsseite	(D) Durch die empfangenen Luftdruckschwankungen werden im Ohr über das Trommelfell Nervenreize erzeugt, die im Gehirn ausgewertet werden.
	(E) Die empfangene Sprache muss in ein „Verstehen“ umgesetzt werden, d. h. sie führt zu einer Orientierung beim Empfänger.

Tabelle 2.1: Kommunikation

Man sieht hier, dass für die Sicherstellung einer funktionsfähigen Kommunikation eine Reihe von Vereinbarungen zwischen Sende- und Empfangsstelle getroffen werden muss.

Zu (A):

Der Vorgang der Verbalisierung dient dazu, aus einer Idee eine Nachricht zu erzeugen. Bei diesem gedanklichen Vorgang wird die ursprüngliche Intention auf die Inhalte reduziert, die zum Verständnis notwendig sind, ohne dass jedoch die Verständlichkeit leidet. Der Erfolg dieses Vorganges hängt ganz wesentlich von den intellektuellen Fähigkeiten des Absenders (und des späteren Nachrichtenempfängers) ab.

Zwischen- bemerkung

Besonders begabte Menschen, die es verstehen, diesen Vorgang zu perfektionieren, sind Künstler. Man bezeichnet sie als Dichter, weil sie es verstehen, eine Idee zu einem Text „zu verdichten“.

Zu (B):

Beim Vorgang der Codierung wird die bereits verbal ausformulierte Nachricht an das physikalische Medium des Nachrichtenkanals angepasst mit dem Ziel, die Übermittlung zu optimieren. Optimierung heißt hier, die Reichweite zu maximieren und die Störsicherheit und Übertragungsgeschwindigkeit der Nachricht zu erhöhen.

Die Erhöhung der Störsicherheit kann durch das Hinzufügen von Wiederholun-

gen (Redundanz) erreicht werden.

Die Erhöhung der Übertragungsgeschwindigkeit einer Nachricht erreicht man durch die Reduzierung des Textes auf das unbedingt Notwendige.

Es ist leicht erkennbar, dass sich diese letzten beiden Ziele in der Regel widersprechen.

Zu einer gewissen Perfektion haben es zum Beispiel Mathematiker gebracht, indem sie einen komplexen Sachverhalt auf eine Formel reduzieren. Die ursprüngliche Intention wird auf die wesentlichen Inhalte reduziert. In der dabei erzeugten Nachricht sind alle unwesentlichen Details weggelassen. Diese Reduktion auf das unbedingt Notwendige ist zwar im Sinne einer schnellen Informationsübertragung gerechtfertigt, sie wäre jedoch eine didaktisch völlig untaugliche Maßnahme, wenn der Empfänger ein Schüler ist, der etwas lernen soll.

**Zwischen-
bemerkung**

Zu (C):

Den Weg vom Sender einer Nachricht zum Empfänger, der durch einen Nachrichtenträger überbrückt wird, nennt man einen Übertragungskanal. Die Eigenschaften des Übertragungskanals werden ganz wesentlich von denen des physikalischen Mediums zwischen Sender und Empfänger bestimmt. Dazu gehören insbesondere die Übertragungsdauer (Ausbreitungsgeschwindigkeit) einer Nachricht und die Reichweite des Senders bzw. die Empfindlichkeit des Empfängers. Man muss hier zwischen der Ausbreitungsgeschwindigkeit im Medium (in der Regel eine Naturkonstante wie Schallgeschwindigkeit oder Lichtgeschwindigkeit) und Übertragungsgeschwindigkeit einer Nachricht unterscheiden. Letztere ist auch von der Länge der Nachricht und damit von der Effizienz der Codierung abhängig.

**Zwischen-
bemerkung**

Zu (D):

Auf der Empfangsseite geschieht mit der Decodierung unter (D) der umgekehrte Vorgang zu (B). Es ist leicht einzusehen, dass der Decodierer unter (D) kompatibel zum Codierer unter (B) sein muss. Wird zum Beispiel der Text einer Nachricht als Morsezeichen in hörbaren Tönen ausgesendet, kann ein Empfänger, der Sprachsignale erwartet, diesen Text nicht verstehen. Obwohl die Übertragung im gleichen Medium erfolgt (hier in der umgebenden Luft), ist eine Kommunikation nicht möglich.

Störungen im Übertragungskanal können häufig in einem gewissen Umfang durch einen gerichteten Empfang ausgeblendet werden.

Es ist zwar auch senderseitig eine gerichtete Aussendung möglich, jedoch muss dann die Position des Empfängers bekannt sein. Das ist beim Sendevorgang eher selten der Fall und bei einer gleichzeitigen Nachrichtenaussendung an mehrere Empfänger (Rundspruch) ist es sogar ausgeschlossen. Deshalb kommt diese Möglichkeit der Störausblendung häufiger auf der Empfängerseite zum Einsatz.

**Zwischen-
bemerkung**

Zu (E):

Das Verstehen hängt in hohem Maße von der Qualität des Vorganges der Ver-

balisierung in (A) ab. Auch hier müssen die Ebenen (A) und (E) zueinander kompatibel sein: Der in Abbildung 2.1 dargestellte Sachverhalt kann in verschiedenen Sprachen verbalisiert werden. Ein „Verstehen“ unter (E) ist nur möglich, wenn die Sprachen kompatibel sind.

Zwischen- bemerkung

Eventuelle Störungen im Übertragungskanal können durch eine Kontextanalyse erkannt und eliminiert werden. Dies setzt jedoch beim Empfänger bereits eine Analyse und ein tieferes Verständnis des Nachrichteninhaltes voraus.

2.1.3 Begriffserläuterungen

Aus dem oben gegebenen Beispiel lassen sich die folgenden Begriffsdefinitionen herleiten, die in den folgenden Kapiteln von besonderer Bedeutung sind. Informationen beinhalten das Wissen über Funktionen, Arbeitsweisen, Verfahren und Vorgänge.

Definition: **Information**

Informationen, vermittelt in textlicher, grafischer, akustischer, visueller und/oder audiovisueller Form, bilden den Inhalt einer Nachricht. Informationen können auf Daten abgebildet und als Daten gespeichert werden. Die auf Daten abgebildeten Informationen bezeichnet man als Nachrichten.

Daten

Daten sind in erkennungsfähiger (analoger und/oder digitaler) Form dargestellte Elemente einer Information, die in Systemen verarbeitet werden können.

Definition: **Daten**

Nach DIN 44 300 sind Daten als Zeichen oder kontinuierliche Funktionen definiert, die aufgrund von bekannten oder unterstellten Abmachungen dem Zwecke der Verarbeitung dienen.

Nachricht

Eine Nachricht ist eine Folge von Zeichen, die aufgrund von vereinbarten oder vorausgesetzten Abmachungen Informationen darstellen.

Definition: **Nachricht**

Nach DIN 44 300 sind Nachrichten „Gebilde aus Zeichen oder kontinuierlichen Funktionen, die aufgrund bekannter oder unterstellter Abmachungen Informationen darstellen und die zum Zweck der Weitergabe als zusammengehörig angesehen und deshalb als Einheit betrachtet werden“. Den Vorgang der Übermittlung einer Nachricht bezeichnet man als Kommunikation.

Nachrichten- übertragung

Eine Nachrichtenübertragung (Datenübertragung) benötigt Zeit. Als Träger kommen deshalb nur solche physikalischen Größen in Frage, die mit der Zeit veränderbar sind.

Definition: Signal, Signalparameter

Als Signal bezeichnet man den eine Nachricht übertragenden zeitlichen Verlauf einer physikalischen Größe. Die Kenngrößen eines Signals, die eine Nachricht darstellen, bezeichnet man als Signalparameter.

Definition: „digitales Signal“, „digitale Nachricht“

Ein Signal bezeichnet man als digitales Signal, wenn seine Signalparameter nur endlich viele Werte annehmen können, die nur zu endlich vielen Zeitpunkten relevant sind (zeitdiskret). Digitale Nachrichten sind Nachrichten, die durch digitale Signale übermittelt werden können.

Definition: Zeichen, Zeichenvorrat, Alphabet

Nach DIN 44 300 ist ein Zeichen ein Element aus einer vereinbarten endlichen Menge von verschiedenen Elementen. Diese Menge bezeichnet man als Zeichenvorrat. Einen Zeichenvorrat, in dem eine Reihenfolge für die Zeichen definiert ist, bezeichnet man als Alphabet.

2.1.4 Struktur eines technischen Kommunikationssystems

Seit Menschen denken können, sind sie bestrebt, die Reichweite ihrer natürlichen Kommunikationsorgane, das heißt ihrer Sinne, zu vergrößern. Dazu haben sie eine Reihe von Techniken entwickelt, die auch nur aufzuzählen den Rahmen dieses Buches sprengen würde. Stellvertretend seien hier nur genannt:

Signale

- Optische Signale: Feuerzeichen oder Rauchzeichen sind bereits aus der Antike bekannt.
- Akustische Signale: Trommeln kennt man vor allem aus Afrika. Jodeln, heute nur noch als Folklore bekannt, war einst jedoch eine wichtige Kommunikationstechnik von Almhirten, um sich über große Höhenunterschiede hinweg zu verständigen.

Geht die Kommunikation über die Hör- und Sichtweite hinaus, so spricht man von Telekommunikation. Kommunikation kann materiell erfolgen (z. B. Brief) oder immateriell (z. B. elektronisch wie beim Fernsprechen).

Abbildung 2.3 zeigt die prinzipielle Struktur eines natürlichen (biologischen) Kommunikationssystems. Es ist eine vereinfachte Darstellung der bereits ausführlich erläuterten Vorgänge (siehe Abschnitt 2.1.2).

Da Luft als Medium für Schallwellen nur eine begrenzte Reichweite zulässt, muss das Nachrichtensignal auf einem Wege übertragen werden, dessen physikalische Eigenschaften den gewünschten Anforderungen entsprechen.

Medium Luft

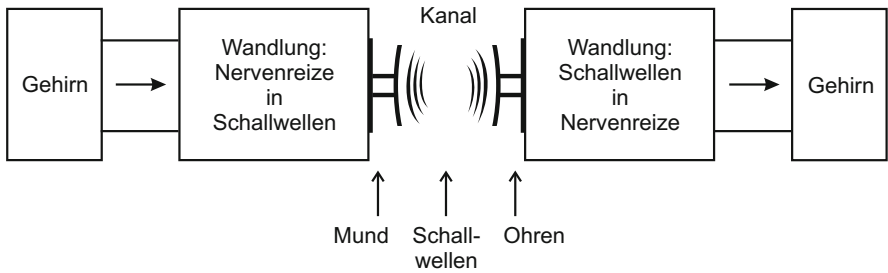


Abbildung 2.3: Struktur eines natürlichen (biologischen) Kommunikationssystems

Dazu kann es zum Beispiel auf einem Trägersignal aufmoduliert werden. Dieser Träger kann eine elektromagnetische Welle sein, wie Abbildung 2.4 zeigt, oder eine Lichtwelle, wie in Abbildung 2.5 dargestellt ist. Sowohl die Übertragung mittels elektromagnetischer Wellen als auch die über Lichtwellen kann durch Kabel (zum Beispiel Koaxialkabel oder Lichtwellenleiter) als auch drahtlos (Funk- oder Infrarotsignale) erfolgen.

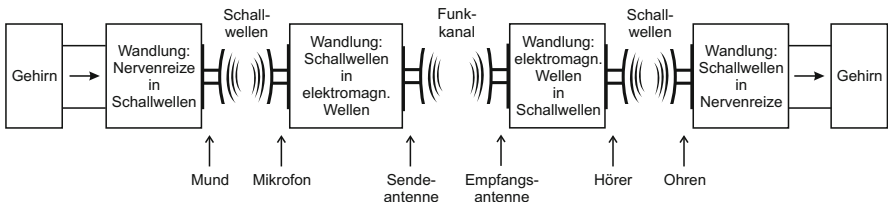


Abbildung 2.4: Grundlegendes Schema eines Kommunikationssystems auf der Basis elektromagnetischer Wellen.

Auf den Vorgang der hier meist notwendigen Modulation wird in Abschnitt 2.3.2 ausführlicher eingegangen.

Wie bei der natürlichen Kommunikation, müssen auch für eine funktionierende technische Kommunikation die verschiedenen Ebenen zueinander kompatibel sein. Man unterteilt dazu die Kommunikationsstruktur in verschiedene Schichten. So ist zum Beispiel die unterste Ebene die Transportschicht. Die sender- und empfangsseitig erfolgende Wandlung und Rückwandlung in das transportierende Medium müssen zueinander kompatibel sein.

Entsprechendes gilt für höhere Schichten. Ein und derselbe Sachverhalt kann zum Beispiel in verschiedenen Sprachen verbalisiert werden. Dass Sender und

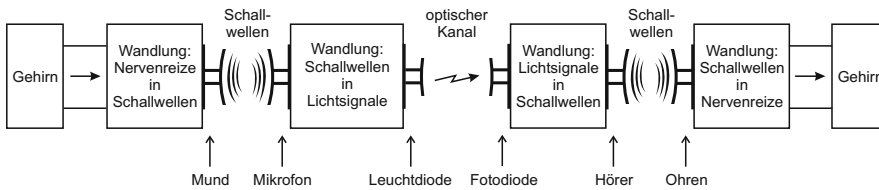


Abbildung 2.5: Grundlegendes Schema eines Kommunikationssystems auf optischem Wege.

Empfänger die gleiche Sprache benutzen, also kompatibel in der Sprachebene sind, ist eine der Voraussetzungen für eine funktionierende Kommunikation. Im Bereich der technischen Kommunikation wird dieses Prinzip im ISO/OSI-Siebenschichten-Referenzmodell realisiert. Eine ausführliche Erläuterung finden Sie dazu im Abschnitt 2.2.5.4.

ISO/OSI

2.1.5 Codierung und Decodierung

Beim Vorgang der Codierung wird einem Sachverhalt (einer Idee, einem Begriff, einem Symbol) ein Zeichen oder eine Zeichenfolge eineindeutig zugeordnet. Ein-eindeutig heißt, dass die Umkehrung der Zuordnung zu nur einem Begriff führen darf (Eindeutigkeit).

Dies ist in der Technik eine zwingende Voraussetzung für eine gültige Codierung. Bei einer natürlichen Codierung, wie sie zum Beispiel in einer natürlichen Sprache erfolgt, ist es dagegen nicht unüblich, dass gleiche Codes (Namen) für unterschiedliche Sachverhalte verwendet werden.

**natürliche
Codierung**

Beispiel: Der Begriff „Geldinstitut“ wird codiert als Zeichenfolge
B A N K

Auch der Begriff „Sitzgelegenheit“ wird codiert als Zeichenfolge
B A N K.

Eine eindeutige Rückabbildung ist hier nicht möglich. In natürlicher Sprache erfolgt die eindeutige Zuordnung durch eine Kontextanalyse beim Empfänger (siehe auch Abbildung 2.6).

Voraussetzung jeglicher gültigen Codierung ist es, dass jedem möglichen Datum ein anderes Codewort zugeordnet wird. Das heißt, jedes Codewort muss sich von jedem anderen Codewort an mindestens einer Stelle unterscheiden. Wird diese Bedingung nicht eingehalten, ist ein Code nicht eindeutig decodierbar (siehe

**gültige
Codierung**

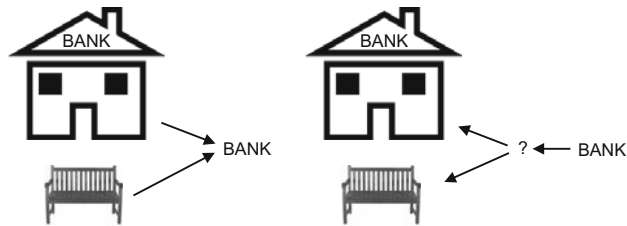


Abbildung 2.6: Eineindeutigkeit beim Vorgang der Codierung

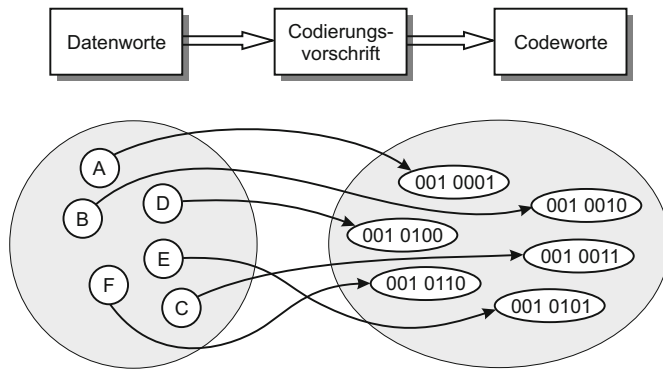


Abbildung 2.7: Prinzip des Codierungsvorganges

ASCII

Abbildung 2.7). Haben alle Codewörter die gleiche Länge (gleiche Anzahl von Bits), so spricht man von einem Blockcode.

Beispiel: Ein Beispiel für einen Blockcode ist der ASCII^a (siehe Abbildung 2.8).

^aASCII = American Standard for Code Information Interchange

2.1.6 Quellencodierung und Kanalcodierung

Nicht nur Kompatibilität und Eineindeutigkeit sind wichtige Anforderungen an die Codierung in einem Kommunikationssystem. Um eine Informationsübermittlung möglichst effizient zu gestalten, wird die Information zunächst auf das Wesentliche reduziert. Der Sachverhalt „Ich habe einen Autounfall gehabt.“ kann in gesprochener Form (und mehr noch bei Sichtverbindung) Informationen über

Quellen- codierung


A	001 0001
B	001 0010
C	001 0011
D	001 0100
E	001 0101
F	001 0110
G	001 0111
	
W	101 0111
X	101 1000
Y	101 1001
Z	101 1010

Abbildung 2.8: Aufbau eines Blockcodes, hier ASCII

den Schreck, über das Maß der Betroffenheit usw. vermitteln. In geschriebener (verbalisierter) Form werden dagegen diese Informationen nicht übermittelt. Das ist sicher auch sinnvoll, wenn zum Beispiel eine Versicherung über diesen Sachverhalt informiert werden soll.

Diesen Vorgang der Reduktion einer zu übermittelnden Information auf das Wesentliche nennt man Quellencodierung. Die Quellencodierung ist ein Codierungsvorgang, der unabhängig vom später verwendeten Übertragungsmedium weit vor der Übertragung stattfindet. Diese kanalunabhängige Optimierung von Daten erzeugt ein informationsverdichtetes Ausgangssignal.

Merksatz: Quellencodierung

Der Quellencodierer setzt seine Eingangsinformationen in eine Nachrichtendarstellung um, die in Abhängigkeit von der physikalischen Auslegung des Kanals (Kanalkapazität) eine zeitlich optimierte Übertragung gestattet. Jeder Nachrichtenkanal kann pro Zeiteinheit stets nur eine begrenzte Anzahl von Signalen übertragen. Um einer Nachrichtenquelle dennoch ein möglichst hohes Signalisierungstempo zu ermöglichen, werden im Quellencodierer die Eingangsinformationen derart codiert, dass relativ häufige Nachrichten weniger Übertragungszeit beanspruchen als seltene Nachrichten. Im Quellencodierer wird so die zu übermittelnde Information von unnötiger Redundanz befreit. Diese Funktion wird als „Informationsverdichtung“ oder auch als „Datenkompression“ bezeichnet.

Die Anpassung der im Quellencodierer aufbereiteten Daten an den Übertragungskanal heißt Kanalcodierung. Sie hat die optimale Anpassung der Daten an den Übertragungskanal zum Ziel. Optimale Anpassung heißt hier, dass:

Optimale Anpassung

- a) Übertragungsfehler erkannt und/oder korrigiert werden können.
- b) Störungen von Übertragungen in benachbarten Nachrichtenkanälen durch physikalische Koppel Effekte minimiert werden.
- c) besondere Leitungseigenschaften des Kanals berücksichtigt werden.

Zu a)

Daten, die vor ihrer Aussendung einer Quellencodierung unterzogen worden sind, reagieren empfindlich auf Übertragungsstörungen, da alle Informationen entfernt wurden, die im Fehlerfall eine nachträgliche Rekonstruktion (z. B. durch eine Kontextanalyse) möglich machen würden. Folglich stellt die Übertragung von quellencodierten (datenreduzierten) Informationen hohe Ansprüche an die Kanalcodierung. Auf die Methoden und Möglichkeiten des Fehlerschutzes im Kanalcodierer wird in Abschnitt 2.2.2 ausführlich eingegangen.

Zu b) und c)

Die Teile b) und c) sind dem Bereich „Leitungscodierung“ zuzuordnen. Obwohl es bei einer drahtlosen Kommunikation gerade keine Kabelverbindung zwischen Sender und Empfänger besteht, gibt es dennoch Anforderungen, die denen bei Kabelverbindung entsprechen. Im Abschnitt 2.2.3 werden diese Anforderungen erläutert.

**Leitungs-
codierung**

Merksatz: Kanalcodierung

Als Kanalcodierung (Encoder) bezeichnet man die Anpassung der zu übertragenden Nachrichten an die Eigenschaften des Übertragungskanals.

Bei aktiven Verfahren zur Minimierung von Übertragungsfehlern verwendet man Fehlererkennungs- und -korrekturmethode, die auf die im Übertragungskanal zu erwartenden Fehler hin optimiert sind.

Bei der passiven Fehlerminimierung erfolgt eine Begrenzung des bei der Aussendung erzeugten Störspektrums. Für die Bitsynchronisation werden entsprechende Mechanismen implementiert.

Ein Kanaldecodierer (Decoder) speichert die aus dem Übertragungskanal empfangenen Daten und versucht, in Kenntnis der im Kanalcodierer angewandten Verfahren eventuell aufgetretene Fehler zu korrigieren (Fehlerkorrektur). Ist eine Fehlerkorrektur nicht möglich, sollte die Tatsache, dass ein Fehler aufgetreten und nicht korrigierbar ist, als Fehlermeldung ausgegeben werden (Fehlererkennung).

Kombination

Quellen- und Kanalcodierung sind natürlich kombinierbar und daher nicht immer deutlich zu trennen. Das ist besonders bei natürlichen Kommunikationssystemen der Fall. Das nachfolgende Beispiel (siehe Abbildung 2.9) zeigt die Abbildung der Kommunikationsstruktur auf eine reale Datenübertragung über eine analoge Fernsprechleitung.

Oberer Teil

Der obere Teil der Abbildung 2.9 zeigt die generelle Struktur eines technischen Kommunikationssystems. Senderseitig werden die Informationen durch Quellen- und Kanalcodierer in eine maschinengerechte Form gewandelt und einem physikalischen Trägersignal aufgeprägt (Modulation).

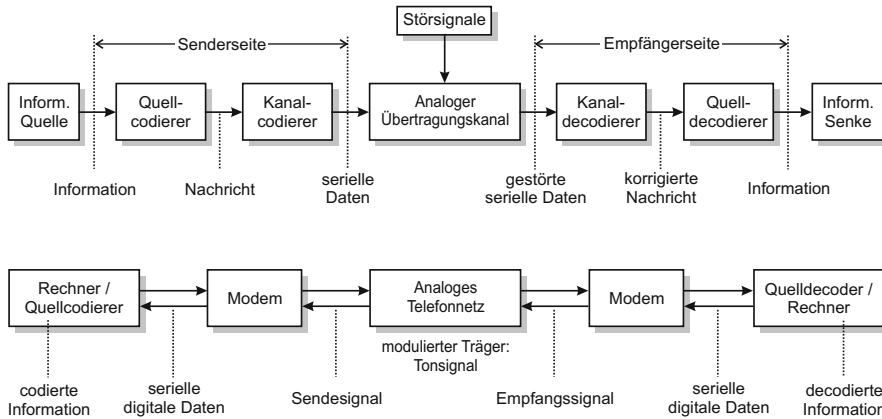


Abbildung 2.9: Abbildung der Kommunikationsstruktur auf ein reales Kommunikationssystem

Der Kanal überbrückt die räumliche Entfernung zwischen Sender und Empfänger, zwischen denen in diesem Beispiel eine Drahtverbindung besteht. Würde hier ein Funkmodem zum Einsatz kommen, wäre die drahtlose Verbindung bei sonst identischem Aufbau möglich. Ein Kanal ist stets ein analoges Medium, unabhängig davon, ob die zu übertragenden Signale in analoger oder digitaler Form vorliegen. Sind es digitale Signale, werden diese beim Vorgang der Modulation auf diskrete analoge Werte (-bereiche) von Amplitude, Frequenz, und/oder Phase abgebildet.

Auf der Empfangsseite erfolgt die Rückumwandlung (Demodulation und Decodierung) des ankommenden Signals in die primäre Information. Im Kanaldecodierer wird dabei die Korrektur eventuell aufgetretener Fehler durchgeführt. Der untere Teil der Abbildung 2.9 zeigt das Blockschaltbild einer Datenverbindung über das analoge Telefonnetz. Die Quellencodierung erfolgt im Rechner. Die als serielle digitale Daten aufbereitete Nachricht wird im Modem dem physikalischen Medium (hier das analoge Telefonnetz) angepasst.

Modem¹ ist eine „Wordmischung“ aus Modulator und Demodulator. Das Modem wandelt die digitalen Signale in Töne im Niederfrequenzbereich, die je nach Modulationsart in der Amplitude, Frequenz und/oder Phase moduliert werden. Wichtig ist, dass nur Signale im Sprachfrequenzbereich (0,3 bis 3,4 kHz) verwendet werden können, denn nur diese können im internationalen analogen Telefonsystem übertragen werden. Im Modem auf der Empfangsseite erfolgen die Demodulation und gegebenenfalls eine Fehlerkorrektur.

Unterer Teil

Modem

¹Modem = **M**odulator **D**emodulator

Auf die technische Realisierung von Quellen- und Kanalcodierern und -decodierern wird im folgenden Kapitel ausführlich eingegangen.

Aufgabe: Betrachten Sie den Ablauf bei der Aufnahme und der Wiedergabe einer CD^a: Was passiert hier bei der Quellen- und Kanalcodierung, was entspricht dem (Übertragungs-) Kanal, wo finden die Kanal- und Quellen-decodierung statt?

^aengl.: Compact Disc

2.1.7 Definitionen

Der Abschnitt definiert die drei „Säulen“ des Buches: Kommunikations-, Nachrichtentechnik und Eingebettete (Funk-)Systeme.

Definition: **Kommunikationstechnik**

beinhaltet die technischen Mittel zur Durchführung von Kommunikation. Kommunikationstechnik stellt ein Teilbereich der Nachrichten- bzw. Informationstechnik dar. Teilweise werden auch „Netzwerke“ der Kommunikationstechnik zugerechnet.

Die Informations- und Kommunikationstechnik wird oft unter dem Begriff „IuK-Technologie“ zusammengefasst.

Definition: **Nachrichtentechnik**

oder nach der klassischen Aufteilung der Elektrotechnik auch „Schwachstromtechnik“ genannt, ist ein Sammelbegriff für die gesamte Technik zur Übertragung, Vermittlung und Verarbeitung von Nachrichten.

Die Nachrichtentechnik bedient sich vorwiegend der Mittel der Elektrotechnik; aus diesem Grund umfasst der Begriff im eigentlichen Sinne die elektrische Nachrichtentechnik.

Die Aufgaben der Nachrichtentechnik sind: theoretische Beschreibung, Entwurf, Realisierung, Test und Betrieb von Nachrichtensystemen nach vorgegebenen Gütekriterien.

Die **Nachrichtenübertragungstechnik** ist ein Teilgebiet der Nachrichtentechnik. Sie beinhaltet die Übertragung von vorwiegend elektrischen Signalen zwischen zwei Kommunikationspartnern (z. B. Mensch und Computer) an zwei räumlich getrennten Orten.

Beispiel: In Abhängigkeit vom Übertragungsmedium unterscheidet man zwischen unterschiedlichen Übertragungssystemen. Beispielfhaft sind hierbei in der Kategorie „ohne“ Medium („drahtlos“) der Mobilfunk (z. B. GSM, DECT) und der Datenfunk (z. B. Bluetooth, ZigBee) zu nennen.

Die Grenze zwischen den beiden Gebieten Nachrichten- und Kommunikationstechnik ist fließend.

Die Abschnitte 2.2 und 2.3 stellen die Nachrichten- und Kommunikationstechnik vor.

Die Eingebetteten (Funk-)Systeme stellen die Plattform zur Implementierung der Wireless-Netzwerke für den Nahbereich dar (siehe Abschnitt 2.4).

Definition: Eingebettete (Funk-)Systeme

Der Begriff Eingebettetes System^a beschreibt einen elektronischen Rechner oder Computer, der in einen technischen Kontext eingebunden oder „eingebettet“ ist.

Eingebettete Systeme führen, für den Benutzer weitgehend unsichtbar, Dienste in einer Vielzahl von Anwendungsgebieten und Geräten durch. Beispielfhaft sind Waschmaschinen, Flugzeuge, Kraftfahrzeuge, Kühlschränke, Fernseher zu nennen. Diese Systeme sind meistens an eine bestimmte Aufgabe angepasst. Dies hat zur Folge, dass aus Kostengründen eine optimierte gemischte Hardware-Software-Implementierung zum Einsatz kommt.

Bei „Funksystemen“ steht die Vernetzung einer Vielzahl von ansonsten autonomen drahtlosen eingebetteten Systemen zu einem komplexen Gesamtsystem im Vordergrund.

^aengl.: Embedded System

2.2 Kommunikationstechnik

Dieses Kapitel stellt die Quellen- und Kanalcodierung sowie den Datenschutz in separaten Abschnitten vor. Hinzu kommt ein weiterer Abschnitt zum Thema Netzwerke.

2.2.1 Quellencodierung

Ein Kanal kann pro Zeiteinheit nur eine gewisse Anzahl von Signalen aufnehmen. Aufgabe des Quellcodierers ist es, die Nachricht, der physikalischen Auslegung des Kanals entsprechend, in Signalfolgen umzusetzen, die der Kanal übertragen kann.

Aufgabe

Daten- kompression

Um der Nachrichtenquelle eine möglichst effiziente (kurze) Übertragungsdauer im Kanal zu ermöglichen, werden durch den Quellencodierer die Eingangssignale komprimiert, ohne wesentliche Informationen zu verlieren. Dabei werden die Daten von unnötiger Redundanz befreit, denn bei der Nachrichtenübertragung kommt es nur auf den Transport des Informationsgehaltes der Daten an.

Relativ häufig auftretende Daten werden so codiert, dass sie nur wenig Zeit zur Übertragung im Kanal benötigen (im Gegensatz zu den selten auftretenden Daten). Diese Funktion eines Quellencodierers wird als „Informationsverdichtung der Nachrichten“ oder „Datenkompression“ bezeichnet.

Auftritts- wahrschein- lichkeit

Die nachfolgende Abbildung 2.10 zeigt die Auftrittswahrscheinlichkeit der Buchstaben (in Prozent) in einem englischen Standardtext (Spalte a): mit Zwischenraum (Leerzeichen); Spalte b): ohne Zwischenraum).

Die Spalten c) und d) zeigen die entsprechende Auftrittswahrscheinlichkeit im deutschen Standardtext.

Huffman- Code

Man erkennt hier leicht, dass im Englischen die Wörter signifikant kürzer sind. Das Datum „Zwischenraum“ tritt wesentlich häufiger auf als in einem deutschen Standardtext. Im unten abgebildeten Codebaum (siehe Abbildung 2.11) für einen Huffman-Code für deutschen Text fällt einem sofort die Asymmetrie auf. Eine Voraussetzung für diesen Code ist, dass kein Codezeichen aus dem Beginn eines anderen Codezeichens bestehen bleiben darf (Fano-Bedingung). Aufbaubedingt ist eine Huffman-Codierung sehr empfindlich gegen Übertragungsfehler, gegen die dann im Kanalcodierer entsprechende Maßnahmen getroffen werden müssen.

Aufgaben:

1. Codieren Sie das Wort „BERLIN“ im 7-Bit-„ASCII“-Code und als Huffman-Code. Wie groß ist der Komprimierungsfaktor K , den Sie bei diesem Wort erreichen? $K = (\text{Anzahl der Bits als Huffman-Code}) / (\text{Anzahl der Bits als „ASCII“-Code})$
2. Kann es bei einer Huffman-Codierung anstatt der zu erwartenden Komprimierung zu einer Vergrößerung der Bitanzahl kommen, im Vergleich zu einer Blockcodierung im 7-Bit- „ASCII“-Code (d. h. $K > 1$)? Geben Sie ein Beispiel an!

Einstieg: Quellencodierung

Weiterführende Literatur zum Thema findet man unter [Wer09].

	Englisch		Deutsch		Huffman-Code	..
	a) mit Leerzeichen	b) ohne Leerzeichen	c) mit Leerzeichen	d) ohne Leerzeichen	e) für das deutsche Alphabet mit Leerzeichen	f) hier nur für Großbuchstaben
LZ	19,25	-	15,15	-	001	010 0000
A	6,60	8,17	4,58	5,40	00001	100 0001
B	1,21	1,49	1,60	1,89	100010	100 0010
C	2,25	2,78	2,67	3,15	10110	100 0011
D	3,43	4,25	4,39	5,17	1101	100 0100
E	10,26	12,70	15,35	18,10	010	100 0101
F	1,80	2,23	1,36	1,60	100111	100 0110
G	1,63	2,02	2,67	3,15	10111	100 0111
H	4,92	6,09	4,36	5,14	00000	100 1000
I	5,63	6,97	6,38	7,52	0111	100 1001
J	0,12	0,15	0,16	0,19	0001110110	100 1010
K	0,62	0,77	0,96	1,13	0001011	100 1011
L	3,25	4,03	2,93	3,45	10010	100 1100
M	1,94	2,41	2,13	2,51	000100	100 1101
N	5,45	6,75	8,84	10,42	111	100 1110
O	6,06	7,51	1,90	2,24	000110	100 1111
P	1,56	1,93	0,50	0,59	00010101	101 0000
Q	0,08	0,10	0,01	0,01	000111011100	101 0001
R	4,84	5,99	6,86	8,08	0110	101 0010
S	5,11	6,33	5,39	6,35	1010	101 0011
T	7,31	9,06	4,73	5,57	1100	101 0100
U	2,23	2,76	3,48	4,10	10000	101 0101
V	0,79	0,98	0,74	0,87	0001111	101 0110
W	1,91	2,36	1,42	1,67	100110	101 0111
X	0,12	0,15	0,01	0,01	000111011101	101 1000
Y	1,59	1,97	0,02	0,02	00011101111	101 1001
Z	0,06	0,07	1,42	1,67	100011	101 1010

Abbildung 2.10: Auftrittswahrscheinlichkeit der Buchstaben (in Prozent) in einem englischen Standardtext

Merksatz: Datenkompression

Spricht man bei der Übertragung von Daten von Quellencodierung, so wird in der Datenverarbeitung eher der Begriff Datenkompression verwendet. Beispiele sind MPEG^a-Layer-3-Audiokomprimierung und JPEG^b-Komprimierung.

^aMPEG = Moving Picture Experts Group

^bJPEG = Joint Photographic Experts Group

2.2.2 Kanalcodierung

Durch Störungen in einem Übertragungskanal können Fehler bei der Datenübertragung verursacht werden. Primäre Aufgabe der Kanalcodierung ist es,

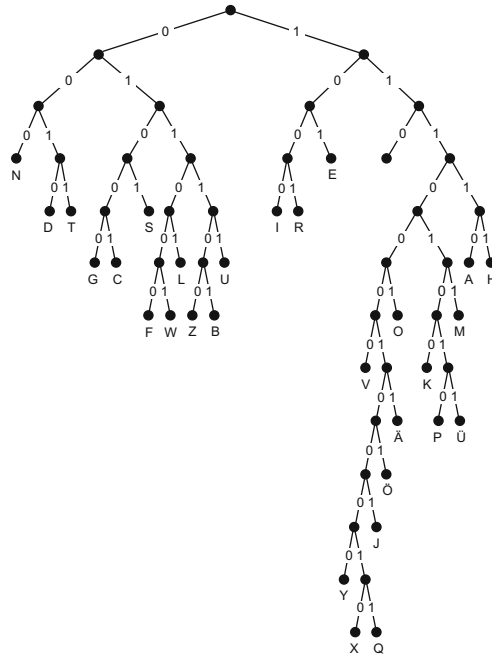


Abbildung 2.11: Codebaum eines Huffman-Codes für deutschsprachige Texte

durch geeignete Erkennungs- und Korrekturverfahren diese Fehler auszuschließen. Dazu wird unmittelbar vor und hinter dem Datenkanal die Kanalcodierung und -decodierung durchgeführt (siehe Abbildung 2.12).

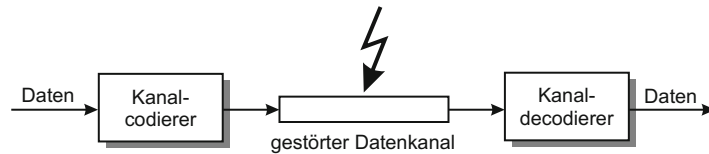


Abbildung 2.12: Anordnung von Kanalcodierer und -decoder in einem Kommunikationssystem

Die Kanalcodierung hat dabei drei verschiedene Aufgaben:

- Fehlerschutz
- Codeformung und
- Signalcodierung.

Fehlerschutz

Der Fehlerschutz ist der eigentliche Überbegriff für die Aufgaben der Kanalcodierung. Die zu übertragenden Daten werden derart verändert, dass es möglich wird, Fehler zu korrigieren oder Fehler zu erkennen und darauf entsprechend zu reagieren.

Die zu übertragenden Daten werden dabei um weitere Prüf- und Kontrollinformationen so ergänzt, dass sie trotz eventueller Störungen im Übertragungskanal auf der Empfangsseite regeneriert werden können (siehe Abbildung 2.13).

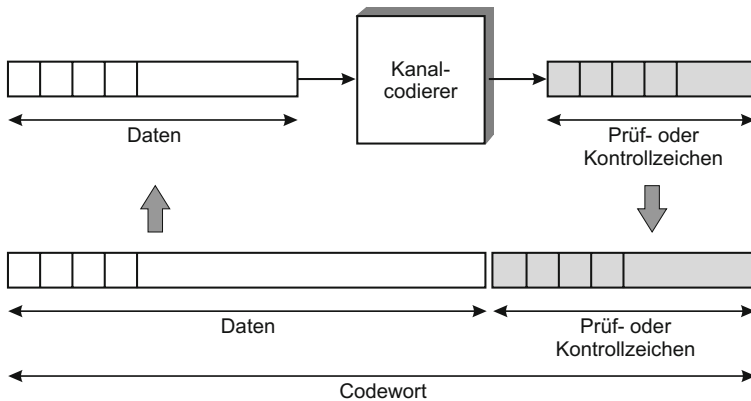


Abbildung 2.13: Prinzipielle Funktionsweise der Datenergänzung durch Prüf- oder Kontrollzeichen zum Zwecke der Fehlererkennung oder Fehlerkorrektur

Man unterscheidet dabei zwei prinzipielle Verfahren: Die Rückwärtskorrektur (REC²) und die Vorwärtskorrektur (FEC³), die in den nächsten beiden Abschnitten behandelt werden.

REC, FEC

Code- und Signalformung sind weitere Maßnahmen, um das eigene und benachbarte Datenübertragungssysteme robust gegenüber Störungen zu machen. Dabei wird eine auf den Übertragungskanal zugeschnittene Codierung eingesetzt. Dies entspräche bei einem kabelgebundenen Übertragungssystem der Leitungscodierung.

Code- und Signalformung

Hier geht es zum Beispiel darum, das Frequenzspektrum des Datenstroms zu verschmälern, um Störungen in Nachbarkanälen zu minimieren. Aber auch eine Verbreiterung des Frequenzspektrums bei gleichzeitiger Amplitudenreduktion wird eingesetzt. Durch entsprechende Codierungen wird es möglich, auf der Empfangsseite eine hohe Taktgüte zu erzielen, was der Fehlervermeidung zu Gute kommt.

²REC = Reverse Error Correction (deutsch: Rückwärtskorrektur)

³FEC = Forward Error Correction (deutsch: Vorwärtskorrektur)

Weiteres zu diesem Thema wird in den Abschnitten 2.2.3, 2.3.2.3 (Teil PSK-Modulation), 2.3.5.2 und 2.3.4.3 detailliert behandelt.

2.2.2.1 Rückwärtskorrektur

Bei der Rückwärtskorrektur muss im Kanalcodierer ein Fehlererkennungsmechanismus implementiert sein, der es dem Kanaldecodierer auf Empfangsseite erlaubt festzustellen, ob ein oder mehrere Fehler aufgetreten sind. Eine Fehleridentifizierung, die auch eine Fehlerkorrektur ermöglichen würde, ist nicht notwendig. Ein zwischen Sender und Empfänger vereinbartes Protokoll (Schicht 2 (Sicherungsschicht) im ISO/OSI-Schichtenmodell) initiiert im Falle eines erkannten Fehlers einen Quittungs- und Rückfragemechanismus, der ein erneutes Aussenden der bereits fehlerhaft übertragenen Nachricht einleitet.

Ein einfaches Beispiel für eine Ein-Bit-Fehler-Erkennung ist ein Prüfbit. Das zu sichernde Datum wird dabei um ein Prüfbit so ergänzt, dass die Anzahl der Einsen im Codewort (Datum um ein Prüfbit ergänzt) gerade ist (engl.: Even Parity, siehe Abbildung 2.14). Mehrbitfehler (genauer eine gerade Fehleranzahl) sind damit nicht erkennbar. Entsprechendes gilt auch für eine ungerade Parität⁴.

ISO/OSI

Paritäts-
prüfung

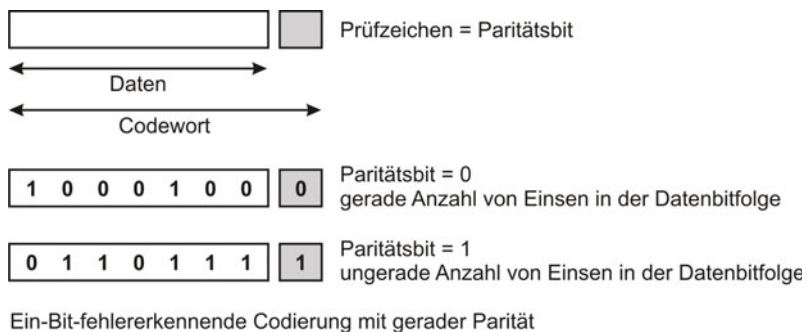


Abbildung 2.14: Ein-Bit-fehlererkennende Codierung mit gerader Parität

⁴engl.: Odd Parity

Beispiel: Die Datenbitfolge $[D_6, D_5, D_4, D_3, D_2, D_1, D_0] = „1010111“$ wird um ein gerades Paritätsbit ergänzt, das heißt die Datenbits D_6 bis D_0 werden Modulo-2-addiert. Die Modulo-2-Addition ist die Addition zweier Binärzahlen ohne Übertrag gemäß Tabelle 2.2.

$$P = D_6 \oplus D_5 \oplus D_4 \oplus D_3 \oplus D_2 \oplus D_1 \oplus D_0$$

$$P = 1 \oplus 0 \oplus 1 \oplus 0 \oplus 1 \oplus 1 \oplus 1$$

$$P = 1$$

Das Codewort (Datum mit Prüfbit) lautet dann:

$$C = [D_6, D_5, D_4, D_3, D_2, D_1, D_0, P]$$

$$C = 1\ 0\ 1\ 0\ 1\ 1\ 1\ 1$$

Kanaldecodierer

Im Kanaldecodierer wird das Prüfbit erneut berechnet und mit dem im Codewort übermittelten Prüfbit verglichen. Bei Ungleichheit zwischen dem empfangenen Prüfbit und dem im Kanaldecodierer berechneten Prüfbit wird eine Aktion eingeleitet, die zum Beispiel aus einer Rückfrage und dann einem erneuten Aussenden der bereits einmal fehlerhaft übertragenen Nachricht führt (siehe Abbildung 2.15). Bei Gleichheit des empfangenen und des berechneten (nachgerechneten) Prüfbits war die Übertragung fehlerfrei.

\oplus	0	1
0	0	1
1	1	0

Tabelle 2.2: Modulo-2-Addition

Aufgabe: Können mit einem Parity-Bit auch Fehler lokalisiert und somit korrigiert werden?

Zyklische Redundanzprüfung

Eine bessere Fehlererkennung bietet die zyklische Redundanzprüfung (CRC⁵). Dabei werden die zu sichernden Datenbits einer Nachricht um mehrere Prüfbits ergänzt und dann ausgesendet. Auf der Empfangsseite wird die Bitfolge aus Nachrichtenbits plus Prüfbits als eine große binäre Zahl betrachtet. Diese Binärzahl wird durch eine andere (vorher festgelegte oder sogar genormte)

CRC

⁵CRC = Cyclic Redundancy Check (deutsch: Zyklische Redundanzprüfung)

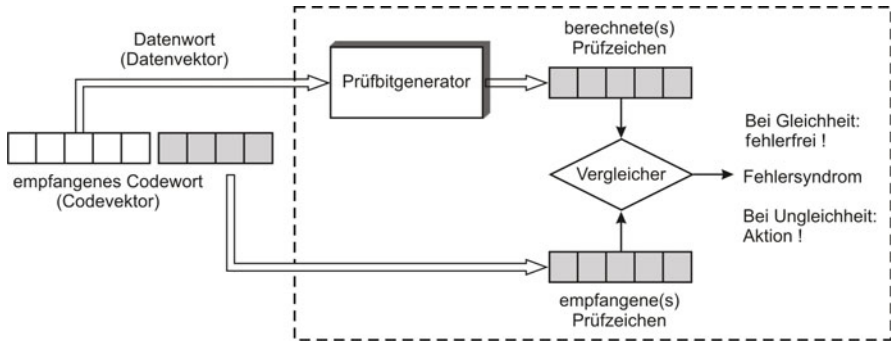


Abbildung 2.15: Erkennung von Übertragungsfehlern im Kanaldecodierer

Binärzahl dividiert. Das Ergebnis der Division wird weiter nicht beachtet, wohl aber der Rest. Man spricht hier von einer Modulo-Division.

Um nicht mit den langen Zahlenfolgen des binären Zahlensystems arbeiten zu müssen, sei hier ein Beispiel im dezimalen Zahlensystem gegeben. Die Zahl

4 3 2 1

soll um eine Prüfziffer derart ergänzt werden, dass sich bei der Division der neuen, um die Prüfziffer ergänzten Zahl „#“

4 3 2 1 #

durch 7 ein Rest von 3 ergibt. Im vorliegenden Fall muss die Prüfziffer 1 lauten, denn

$$43211 \bmod 7 = 3$$

Merksatz: Modulo-Division

Die Division liefert den ganzzahligen Rest einer Division.

Beispiel:

$$8 / 4 = 2 \text{ Rest } 0, \text{ damit ist } 8 \bmod 4 = 0$$

$$7 / 4 = 1 \text{ Rest } 3, \text{ damit ist } 7 \bmod 4 = 3$$

Die so aus der ursprünglichen Zahl (4321), ergänzt um die Prüfziffer (1) erhaltene Ziffernfolge (43211), wird ausgesandt. Der Empfänger kann mit der Durchführung der Modulo-7-Division überprüfen, ob das Ergebnis $43211 \bmod 7 = 1$ ist. Ist das nicht der Fall, liegt mit Sicherheit ein Übertragungsfehler vor.

Dieses Verfahren darf nicht mit dem Prüfsummenverfahren verwechselt werden. Bei letzterem wird lediglich die Quersumme der Ziffern einer Zahl als Prüfzahl hinzugefügt. Ein Tausch zweier Ziffern wäre damit zum Beispiel nicht erkennbar.

Überträgt man dieses Prinzip auf binäre Zahlenfolgen, entspricht dies einer zyklischen Redundanz-Prüfung (CRC).

Zum Verständnis der praktischen Realisierung müssen zuvor noch einige Begriffe erläutert werden:

Polynomdarstellung

Die Dezimalzahl „245“ kann man auch als die Summe der Produkte aus den Ziffern der Zahl mit ihrem Stellenwert darstellen.

$$245_{10} = 2 \cdot 10^2 + 4 \cdot 10^1 + 5 \cdot 10^0$$

Im binären Zahlensystem gilt für die Binärzahl „11110101“ entsprechend:

$$11110101_2 = 1 \cdot 2^7 + 1 \cdot 2^6 + 1 \cdot 2^5 + 1 \cdot 2^4 + 0 \cdot 2^3 + 1 \cdot 2^2 + 0 \cdot 2^1 + 1 \cdot 2^0$$

Eine Datenbitfolge der Länge K kann durch ein Polynom vom Grad K-1 dargestellt werden. Wenn die Basis des Zahlensystems durch X ersetzt wird, erhält man die Darstellung der Binärzahl „11110101“ als Polynom vom Grad 7.

$$\begin{aligned} 11110101_2 &= 1 \cdot X^7 + 1 \cdot X^6 + 1 \cdot X^5 + 1 \cdot X^4 + 0 \cdot X^3 + 1 \cdot X^2 + 0 \cdot X^1 + 1 \cdot X^0 \\ &= X^7 + X^6 + X^5 + X^4 + X^2 + X^0 \\ &= X^7 + X^6 + X^5 + X^4 + X^2 + 1 \end{aligned}$$

Polynomdivision

So wie man im dezimalen Zahlensystem den Quotienten aus Dividenden und Divisor berechnen kann, kann dies auch im binären Zahlensystem geschehen.

Dividiert man das Polynom

$$X^7 + X^6 + X^5 + X^4 + X^2 + 1 \text{ (Dividend)}$$

durch das Polynom

$$X^5 + X^4 + X^2 + 1 \text{ (Divisor)},$$

so erhält man den Quotienten :

$$\begin{aligned} &X^7 + X^6 + X^5 + X^4 + X^2 + 1 \text{ / } X^5 + X^4 + X^2 + 1 = X^2 + 1 \\ - &\underline{(X^7 + X^6 + X^4 + X^2)} \\ = &\quad X^5 + 1 \\ - &\underline{(X^5 + X^4 + X^2 + 1)} \\ = &\quad X^4 + X^2 \quad (= \text{Rest: } 10100_2) \end{aligned}$$

Division mit binären Zahlen

Bei der Division ist zu beachten, dass sie im Modulo-2-System durchgeführt wird und die Modulo-2-Arithmetik gilt (siehe auch oben Tabelle 2.2): **Modulo-2**

$$0+0=0 \quad 0+1=1 \quad 1+0=1 \quad 1+1=0 \quad -1=+1$$

$$11110101 \text{ / } 110101 = 101$$

```

110101
0010000
000000
100001
110101
010100 (=Rest: 101002)

```

Durchführung einer CRC-Generierung

Soll ein binäres Datenpaket mit einer Prüfzeichenfolge (FCS⁶) versehen werden, so muss zunächst das Generatorpolynom für die CRC-Generierung ausgewählt werden.

Es gibt dazu genormte Prüfzeichenverfahren, zum Beispiel DIN ISO 7064 oder für die HDLC⁷-Datenübertragungsnorm die ISO 3309, in denen festgelegt ist, welche Polynome zum Einsatz kommen und welche Fehlerabdeckung damit möglich ist.

<i>Beispiel:</i>				
	Polynomdarstellung	Duale Schreibweise	Grad	
CRC-16	$X^{16} + X^{15} + X^2 + 1$	1 1000 0000 0000 0101	16	
CRC-16 ^a	$X^{16} + X^{12} + X^5 + 1$	1 0001 0000 0010 0001	16	
CRC-12	$X^{12} + X^{11} + X^3 + X^2 + X + 1$	1 1000 0000 1111	12	
^a CRC-16-Polynom kommt bei FCS des HCLD- bzw. X25-Protokolls zum Einsatz				

Für das folgende Beispiel sei das Polynom

$$X^5 + X^4 + X^2 + 1$$

als Generatorpolynom (oder Prüfpolynom) ausgewählt. Dieses Polynom entspricht der sechsstelligen Binärzahl „1 1 0 1 0 1“. Es ist ein Polynom 5. Grades. Der durch eine Modulo-Division entstehende Rest ist daher maximal fünfstellig. Das Datenpaket, welches mit einer CRC-Prüfzeichenfolge versehen werden soll, lautet:

$$1\ 1\ 1\ 1\ 0\ 1\ 0\ 1 \quad (X^7 + X^6 + X^5 + X^4 + X^2 + 1)$$

Die Bitfolge im Datenpaket wird mit dem Grad des Prüfpolynoms multipliziert (X^5) d. h. es wird um fünf Nullen ergänzt. Die Bitfolge des Datenpakets (Dividend) lautet dann:

$$1\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 0\ 0.$$

Mit der Durchführung der Modulo-2-Division wird der Rest bestimmt:

$$1111010100000 / 110101 = 10111010$$

⁶FCS = Frame Check Sequence

⁷engl.: High Level Data Link Control


```

110101
0010000
000000
 100001
110101
0101000
110101
0111010
110101
0011110
000000
 111100
110101
0010010
000000
010010 Rest: 10010

```

Der errechnete Quotient „10111010“ wird nicht weiter beachtet.

Die Summe aus dem Dividenten und dem ermittelten Rest 10010 ergibt das mit dem CRC versehene Datenpaket.

```

1111010100000
+      10010
1111010110010

```

Durchführung der CRC-Prüfung bei einem empfangenen Datenpaket

Zur Prüfung eines empfangenen Datenpakets auf Fehler wird dieses durch das Generatorpolynom dividiert. Der Quotient wird nicht beachtet. Ist der ermittelte Rest wie im vorliegenden Fall Null, liegt kein Übertragungsfehler vor.

```

11110101110101 / 110101 = 10111010
110101
0010000
000000
 100001
110101
0101001
110101
0111000
110101
0011010

```

```

000000
110101
110101
0000000
000000
000000 Rest: 00000

```

Durch entsprechende Voreinstellung kann man jeden beliebigen Rest erzeugen, bzw. das fehlerfreie Datenpaket jeder beliebigen Restklasse (siehe unten unter Fehlerbetrachtung) zuordnen.

Technische Realisierung einer CRC-Erzeugung und -Prüfung

Nur in Ausnahmefällen wird eine Modulo-2-Polynomdivision algebraisch durch ein Softwareprogramm durchgeführt (z. B. bei der Fehlerprüfung von Dateien). Sie kann mit einem linear rückgekoppelten Schieberegister (LFSR⁸) leicht durch einen integrierten Schaltkreis (IC⁹) (siehe Abschnitt 2.4) realisiert werden. Der Divisor bestimmt dabei den Aufbau des Rückkopplungszweiges des Schieberegisters. Im vorliegenden Fall lautet der Divisor „110101“. Damit erhält man die in Abbildung 2.16 dargestellte Schieberegisterschaltung.

Integrierter Schaltkreis

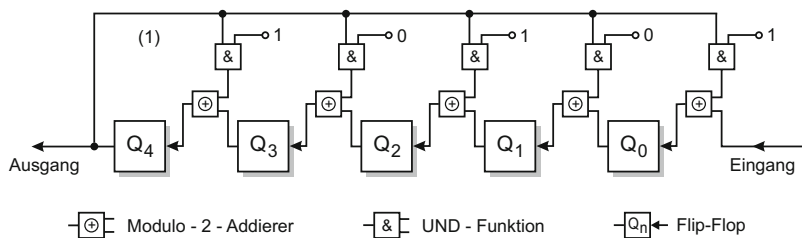


Abbildung 2.16: Linear rückgekoppeltes Schieberegister zur CRC-Generierung mit variabler Einstellung des Prüfpolyoms (hier $(X^5 + X^4 + X^2 + 1)$ bzw. "110101")

LFSR

Das Schieberegister hat den Anfangszustand $(Q_4Q_3Q_2Q_1Q_0) = „00000“$. Die Binärzahl: „11110101₂“ wird, beginnend mit der höchstwertigen Stelle (MSB¹⁰), von links nach rechts in das Schieberegister eingelesen. Die Zwischenzustände des Schieberegisters errechnen sich nach folgenden Funktionsgleichungen:

$$\begin{aligned} Q_0^{t+1} &= E^t \oplus Q_4^t \\ Q_1^{t+1} &= Q_0^t \\ Q_2^{t+1} &= Q_1^t \oplus Q_4^t \end{aligned}$$

⁸LFSR = **L**inear **F**eedback **S**hift **R**egister (*deutsch: linear rückgekoppeltes Schieberegister*)

⁹IC = Integrated Circuit (*deutsch: Integrierter Schaltkreis*)

¹⁰MSB = **M**ost **S**ignificant **B**it (deutsch: höchstwertiges Bit)

$$Q_3^{t+1} = Q_2^t$$

$$Q_4^{t+1} = Q_3^t \oplus Q_4^t$$

Damit ergibt sich folgende Tabelle 2.3 für die Zustände nach jedem Taktschritt:

	Q_4	Q_3	Q_2	Q_1	Q_0	E	Eingangs-Bitfolge
Start:	0	0	0	0	0	1	11110101
1. Takt:	0	0	0	0	1	1	1110101
2. Takt:	0	0	0	1	1	1	110101
3. Takt:	0	0	1	1	1	1	10101
4. Takt:	0	1	1	1	1	0	0101
5. Takt:	1	1	1	1	0	1	101
6. Takt:	0	1	0	0	0	0	01
7. Takt:	1	0	0	0	0	1	1
8. Takt:	1	0	1	0	0		← Rest: 10100

Tabelle 2.3: Zwischenzustände des Schieberegisters $X^5 + X^4 + X^2 + 1$

Wie man sieht, enthält das Schieberegister nach acht Taktschritten den Rest „10110“ als Ergebnis der Modulo-2-Division durch das Prüfpolyynom $(X^5 + X^4 + X^2 + 1)$. Dieser Rest wird dann der Datenbitfolge als Prüfsequenz hinzugefügt. Die schaltungstechnische Realisierung eines linear rückgekoppelten Schieberegisters vereinfacht sich erheblich, wenn das Prüfpolyynom fest eingestellt wird. In Abbildung 2.17 ist dies dargestellt.

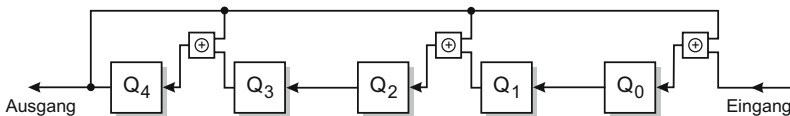


Abbildung 2.17: Linear rückgekoppeltes Schieberegister zur CRC-Generierung mit dem Polynom $X^5 + X^4 + X^2 + 1$

Fehlerbetrachtung

Das Prinzip beim Cyclic Redundancy Check beruht darauf, dass eine Datenbitfolge (Datenpaket) der Länge K durch eine konstante Prüfbitfolge der Länge r dividiert wird. Das Ergebnis dieser Division wird nicht weiter betrachtet, wohl aber deren Rest. Wenn die Prüfbitfolge $P(x)$ die Länge r hat, können bei Divisionen durch diese Prüfbitfolge 2^r verschiedene Reste auftreten. Diese verschiedenen möglichen Reste bezeichnet man auch als Restklassen.

Die Anzahl der möglichen Datenbitfolgen mit der Länge K beträgt 2^K . Bei der Division durch die Prüfbitfolge wird jeder Datenbitfolge eine der 2^r verschiedenen Restklassen zugeordnet. Da $K \gg r$ ist, müssen verschiedenen Datenbit-

folgen die gleichen Restklassen zugeordnet sein.

Von den 2^K möglichen verschiedenen Datenpaketen der Länge K ist nur das Datenpaket $M(x)$ korrekt, alle anderen sind fehlerhaft. Wird die Datenbitfolge $M(x)$ durch ein Prüfpolynom $P(x)$ vom Grade r dividiert, so ergibt das den Quotienten $Q(x)$ (der ja nicht weiter betrachtet wird) und den Rest $R(x)$. Die Beziehung zwischen $M(x)$, $P(x)$, $Q(x)$ und $R(x)$ lautet:

$$M(x) = Q(x) \cdot P(x) + R(x)$$

Für eine fehlerhafte Datenbitfolge $M'(x)$ gilt:

$$M'(x) = Q'(x) \cdot P(x) + R'(x)$$

Die Bedingung für einen nicht erkennbaren Fehler muss daher lauten:

$$M(x) = Q(x) \cdot P(x) + R(x) \text{ und}$$

$$M'(x) = Q'(x) \cdot P(x) + R(x),$$

denn $M(x)$ und $M'(x)$ erzielen bei einer Division durch $P(x)$ den gleichen Rest ($R(x) = R'(x)$).

Es gibt 2^K mögliche Datenpakete $M(x)$ der Länge K . Wenn ein Prüfpolynom $P(x)$ vom Grade r verwendet wird, beträgt die Anzahl der möglichen verschiedenen Restklassen 2^r . Bei der Polynomdivision $M(x) / P(x)$ fallen somit in jede Restklasse

Anzahl der möglichen Datenpakete der Länge K

Anzahl der möglichen Restklassen eines Prüfpolynoms des Grades r

$$= 2^K / 2^r$$

mögliche Datenpakete. Der Restklasse $R(x)$ des fehlerfreien Falles werden daher auch fehlerhafte Datenpakete zugeordnet. Von den $2^K / 2^r$ möglichen Datenpaketen dieser Restklasse $R(x)$ sind $(2^K / 2^r) - 1$ fehlerhaft, und ein Datenpaket ist korrekt. Die Anzahl der nichterkennbaren Fehler beträgt somit $2^{K-r} - 1$.

Von den 2^K möglichen Datenpaketen der Länge K enthalten $2^K - 1$ Fehler, und eines ist fehlerfrei. Die Anzahl der möglichen fehlerhaften Datenpakete beträgt daher $2^K - 1$. Die Wahrscheinlichkeit w , dass ein fehlerhaftes Datenpaket nicht erkannt wird, beträgt:

$$w = \frac{\text{Anzahl der nichterkennbaren Fehler}}{\text{Anzahl der möglichen Fehler}}$$

Hat das Datenpaket $M(x)$ die Länge K und ist Prüfpolynom $P(x)$ vom Grade r , so beträgt die Wahrscheinlichkeit w eines nichterkennbaren Fehlers:

$$w = \frac{2^{K-r} - 1}{2^K - 1}$$

Für lange Datenpakete mit $K \gg r$ wird für w der Grenzwert mit $k \rightarrow \infty$ ermittelt:

$$w = \lim_{k \rightarrow \infty} \frac{2^{K-r} - 1}{2^K - 1} = 2^{-r}$$

Die Wahrscheinlichkeit, dass ein fehlerhaftes Datenpaket als korrekt erkannt wird, beträgt $(0,5)^r$, wobei r der Grad des Polynoms ist, durch das man dividiert (r = Anzahl der Datenbits des Divisors (ohne führende Nullen) minus 1).

**nicht
erkennbare
Fehler**

**Wahrschein-
lichkeit**

Beispiel: Wird mit einem 16-Bit-Schieberegister die Polynomdivision durchgeführt, so ist das Prüfpolyynom vom Grad $r = 15$. Die Wahrscheinlichkeit w für einen nichterkennbaren Fehler ist in einem langen Datenpaket $w = 0,00003$

Diese geringe Wahrscheinlichkeit eines unentdeckten Fehlers wird häufig als Beleg für die Effizienz dieses Verfahrens zitiert. In die obige Berechnung geht jedoch nur der Grad des Prüfpolyoms mit ein, nicht jedoch dessen Zusammensetzung. Das heißt, wenn man beliebige 16 Bits aus dem Datenpaket $M(x)$ als Prüfzeichen betrachtet, kommt man auf die gleiche geringe Wahrscheinlichkeit eines unerkennbaren Fehlers.

Nur wenn die Fehlerauftretenswahrscheinlichkeit für jedes Bit des Datenpakets gleich groß ist und die Fehler unabhängig voneinander sind, ist diese Aussage über die Wahrscheinlichkeit eines unerkennbaren Fehlers korrekt. In der Praxis trifft diese Annahme jedoch eher selten zu. Wenn ein Datenbit fehlerhaft ist, ist die Wahrscheinlichkeit eines weiteren Fehlers meist höher, als wenn kein Datenbit fehlerhaft übertragen wurde. Man sollte daher Prüfpolynome auswählen, deren Fehlererkennungsmöglichkeiten mathematisch genau untersucht worden sind. Bei den in den Normen aufgeführten Polynomen ist das in der Regel der Fall.

Mit dem CRC-Verfahren kann auch die Integrität von Daten nicht sicher nachgewiesen werden, denn es ist verhältnismäßig leicht möglich, eine Datenbitfolge gezielt so zu modifizieren, dass die Division durch das Prüfpolyynom zu einer ganz bestimmten Restklasse (z. B. dem fehlerfreien Fall) führt.

2.2.2.2 Vorwärtskorrektur

Auch bei einer Vorwärtskorrektur werden die Datenbits durch zusätzliche Prüfbits ergänzt. Das Ziel ist hier jedoch eine Fehleridentifizierung, die eine Korrektur erlaubt, ohne dass eine Rückfrage beim Absender nötig ist. Nicht immer steht ein Rückkanal zur Verfügung (Simplex-Betrieb).

Grundsätzlich können bei einem entsprechenden Aufwand an Prüfbits Codierungen geschaffen werden, mit denen eine beliebige Anzahl von Fehlern erkannt und korrigiert werden kann.

Abbildung 2.18 zeigt als Beispiel das Blockschaltbild eines (7,4)-Hamming-codierers für einen Simplex-Nachrichtenkanal. Mit diesem Kanalcodierer können alle 1-Bit-Fehler in einem Blockcode, der aus vier Datenbits und drei Prüfbits, also sieben Codebits besteht, erkannt und korrigiert werden. Im Kanalcodierer werden die Prüfbits P_0 , P_1 , und P_2 aus dem Datenwort $[D_3, D_2, D_1, D_0]$ bestimmt, indem jeweils eine Modulo-2-Addition mit verschiedenen Kombinationen der Datenbits durchgeführt wird.

Hamming-codierer

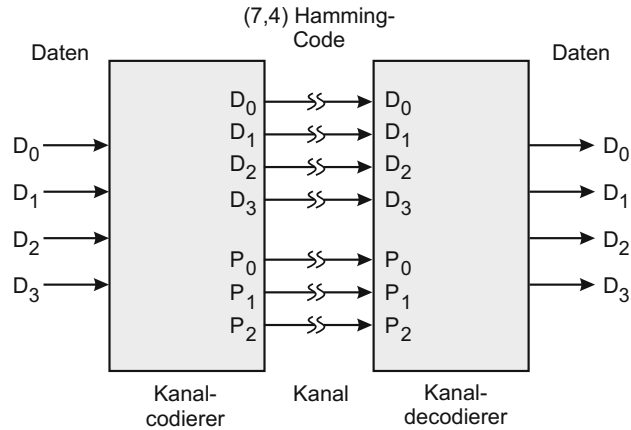


Abbildung 2.18: Grundaufbau einer fehlertoleranten Datenübertragung durch Kanalcodierung mittels Hamming-Code

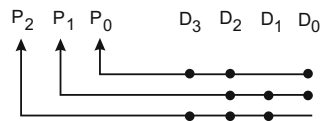


Abbildung 2.19: Zusammensetzung der Paritätsbits P_0 bis P_2 aus den Datenbits D_0 bis D_3

Die vier Datenbits werden um die berechneten drei zusätzlichen Prüfbits $[P_2, P_1, P_0]$ ergänzt. Das so erzeugte, aus sieben Bits bestehende Codewort wird dann auf dem Nachrichtenkanal ausgesendet. Aus Abbildung 2.19 ergeben sich die Werte der Prüfbits P_0, P_1 , und P_2 :

$$P_0 = D_3 \oplus D_2 \oplus D_0$$

$$P_1 = D_2 \oplus D_1 \oplus D_0$$

$$P_2 = D_3 \oplus D_2 \oplus D_1$$

Mathematisch wird hier der Datenvektor $D = [D_3, D_2, D_1, D_0]$ mit einer Generatormatrix $[G]$ multipliziert. Das Ergebnis der Multiplikation ist der Codevektor

$$C = [P_2, P_1, P_0, D_3, D_2, D_1, D_0].$$

$$C = D \cdot [G] \text{ mit}$$

$$G = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$C = [D_3, D_2, D_1, D_0] \cdot \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$C = [P_2, P_1, P_0, D_3, D_2, D_1, D_0]$$

Abbildung 2.20 zeigt eine technische Realisierung des Kanalcodierers, der den obigen 7-Bit-langen (7,4)-Hammingcode erzeugt. Dieser Kanalcodierer führt außer der Generierung der Prüfbits auch eine Serialisierung des Codewortes durch. In die Speicherzellen Q_3, Q_2, Q_1 , und Q_0 wird das zu codierende Datenwort geladen, zum Beispiel $[D_3, D_2, D_1, D_0] = [0101]$. Am Ausgang der rechten Speicherzelle (Q_0) liegt dann der Wert $D_0 = 1$ an. Mit jedem Taktschritt wird der Inhalt jeder Speicherzelle in die jeweils rechts daneben liegende Speicherzelle weitergeschoben.

Funktionsweise

In die linke Speicherzelle Q_3 wird zugleich die Summe einer Modulo-2-Addition aus den Inhalten der Speicherzellen Q_3, Q_2 , und Q_0 geladen. Der Inhalt der rechten Speicherzelle Q_0 geht als serieller Ausgang an den Übertragungskanal.

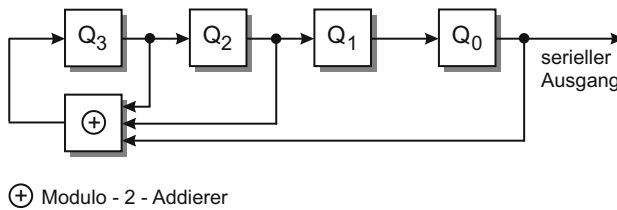


Abbildung 2.20: Technische Realisierung eines seriellen Hamming-Code-Generators mit einem rückgekoppelten Schieberegister

Beispiel: Abbildung 2.21 zeigt die Speicherzustandsfolge beim Startwert „0101“

Wie man hier leicht sieht, wird das Prüfbit P_0 aus den Werten der Speicherzellen Q_3, Q_2 und Q_0 (beim Start sind das die Datenbits D_3, D_2 und D_0) mit einer Modulo-2-Addition¹¹ berechnet. Nach dem ersten Taktschritt sind die Inhalte von $Q_3, Q_2, Q_0 = P_0, D_3, D_1$, nach dem zweiten Taktschritt lauten sie: $Q_3, Q_2, Q_0 = P_1, P_0, D_2$. Damit ergeben sich folgende Bestimmungsgleichungen für die Prüfbits:

¹¹Ex-Or-Funktion

	Q ₃	Q ₂	Q ₁	Q ₀	Ausgang
Startzustand:	0	1	0	1	1 (=D ₀)
n. d. 1. Takt:	0	0	1	0	0 (=D ₁)
n. d. 2. Takt:	0	0	0	1	1 (=D ₂)
n. d. 3. Takt:	1	0	0	0	0 (=D ₃)
n. d. 4. Takt:	1	1	0	0	0 (=P ₀)
n. d. 5. Takt:	0	1	1	0	0 (=P ₁)
n. d. 6. Takt:	1	0	1	1	1 (=P ₂)

Abbildung 2.21: Beispiel für die Speicherzustandsfolge beim Startwert „0101“

$$P_0 = D_3 \oplus D_2 \oplus D_0$$

$$P_1 = D_2 \oplus D_1 \oplus D_0$$

$$P_2 = D_3 \oplus D_2 \oplus D_1$$

Für andere Eingangsdaten kann die Ausgangsbitfolge der Abbildung 2.22 entnommen werden oder mittels der Generatormatrix berechnet werden:

Z.- Nr.	Eingangsdaten				Serieller Ausgang							
	D ₃	D ₂	D ₁	D ₀	P ₂	P ₁	P ₀	D ₃	D ₂	D ₁	D ₀	
0	0	0	0	0	0	0	0	0	0	0	0	
1	0	0	0	1	0	1	1	0	0	0	1	
2	0	0	1	0	1	1	0	0	0	1	0	
3	0	0	1	1	1	0	1	0	0	1	1	
4	0	1	0	0	1	1	1	0	1	0	0	
5	0	1	0	1	1	0	0	0	1	0	1	
6	0	1	1	0	0	0	1	0	1	1	0	
7	0	1	1	1	0	1	0	0	1	1	1	
8	1	0	0	0	1	0	1	1	0	0	0	
9	1	0	0	1	1	1	0	1	0	0	1	
10	1	0	1	0	0	1	1	1	0	1	0	
11	1	0	1	1	0	0	0	1	0	1	1	
12	1	1	0	0	0	1	0	1	1	0	0	
13	1	1	0	1	0	0	1	1	1	0	1	
14	1	1	1	0	1	0	0	1	1	1	0	
15	1	1	1	1	1	1	1	1	1	1	1	

Abbildung 2.22: Serieller Ausgang für verschiedene Folgen von Eingangsdaten

Kanaldecoder für Hammingcode

Auf der Empfangsseite im Kanaldecoder werden aus den empfangenen Da-

tenbits noch einmal die drei Prüfbits berechnet und diese mit den empfangenen Prüfbits verglichen (siehe Abbildung 2.15). Dies geschieht durch die Multiplikation des Codevektors $C = [P_2, P_1, P_0, D_3, D_2, D_1, D_0]$ mit der Prüfmatrix $[P]$. Das Ergebnis dieser Multiplikation ist das Fehlersyndrom $F = [F_2, F_1, F_0]$.

Fehler- syndrom

$F = C \cdot [P]$ mit

$$P = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

$$[P_2, P_1, P_0, D_3, D_2, D_1, D_0] \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = [F_2, F_1, F_0]$$

In der Fehlersyndromtabelle 2.4 findet man die Fehlerstelle bzw. bei einem Fehlersyndrom von $F = „000“$ den Hinweis, dass kein 1-Bit-Übertragungsfehler aufgetreten ist.

F_2	F_1	F_0	Fehler
0	0	0	–
0	0	1	P_0
0	1	0	P_1
0	1	1	D_0
1	0	0	P_2
1	0	1	D_3
1	1	0	D_1
1	1	1	D_2

Tabelle 2.4: Fehlersyndromtabelle

Beispiel: Am Eingang des obigen Kanalcodierers wird das Datenwort $[D_3, D_2, D_1, D_0] = [0101]$ angelegt. Durch Multiplikation mit der Generatormatrix erhält man das Codewort C :

$$C = D \cdot [G]$$

$$C = [P_2, P_1, P_0, D_3, D_2, D_1, D_0] = [1\ 0\ 0\ 0\ 1\ 0\ 1]$$

(siehe auch Zeile 5 in der Abbildung 2.22)

Beispiel: Durch einen Übertragungsfehler im Kanal wird anstatt des gesendeten Codewortes $[P_2, P_1, P_0, D_3, D_2, D_1, D_0] = [1000101]$ das Codewort $C = [1000001]$ empfangen. Im Kanaldecodierer erfolgt dessen Multiplikation mit der Prüfmatrix $[P]$, um das Fehlersyndrom F zu erhalten.

$$F = C \cdot [P]$$

$$F = [F_2, F_1, F_0] = [1\ 1\ 1]$$

Hinweis

Das Fehlersyndrom lautet: „111“. Anhand der Fehlersyndromtabelle kann man den Fehler lokalisieren und damit korrigieren. Im vorliegenden Fall wurde „ D_2 “ falsch übertragen.

Aufgabe: Sie haben für den oben angeführten Hammingcode die Codefolgen

a) „1001001“

b) „0001011“ und

c) „0011100“ empfangen.

Prüfen Sie die empfangenen Codewörter a), b) und c) auf Fehler und korrigieren Sie diese gegebenenfalls. Wie lauten die (korrigierten) Datenwörter?

Faltungscodierer

Ein anderes Beispiel für eine Kanalcodierung ist ein Faltungscodierer¹². Hier wird der Informationsgehalt eines Datenbits auf mehrere Prüfbits verteilt, also quasi „verschmiert“. In einem einfachen Faltungscodierer wird nur jeweils ein Prüfbit zwischen zwei Datenbits eingefügt.

Aus den zwei benachbarten Datenbits D_N und D_{N+1} wird durch eine Modulo-2-Addition¹³ (Addition der Bitwerte ohne Berücksichtigung eines entstehenden

¹²engl.: Convolutional Coder

¹³ \oplus : Exklusiv-Oder-Funktion

Übertrags) das Prüfbit P_N gebildet:

$$D_N = D_N \oplus D_{N+1}$$

$$\dots D_N P_N D_{N+1} P_{N+1} D_{N+2} P_{N+2} D_{N+3} \dots$$

Kanaldecodierer für Faltungscodes

Auf der Empfangsseite werden die Prüfbits nachgebildet. Sind zwei benachbarte Prüfbits nicht korrekt, so muss ein Übertragungsfehler bei dem dazwischen liegenden Datenbit aufgetreten sein. Ist dagegen nur ein Prüfbit nicht korrekt, so liegt der Übertragungsfehler beim Prüfbit selbst.

Beispiel: **Senderseite:**

Datenbitfolge: 0 0 0 1 1 0 1 1 0 0

Ausgesendete

Faltungscodefolge: 0 0 0 0 0 1 1 0 1 1 0 1 1 0 1 1 0 0 0

Beispiel: **Empfangsseite:**

empfangene

Bitfolge: 0 0 0 0 1 1 1 0 1 1 0 0 1 0 1 1 0 0 0

Berechnung der

Prüfbits: 0 0 1 0 1 1 0 1 0

Falsch erkannte

Prüfbits: 0 0 1 0 1 1 0 1 0

Korrigierte

Datenbitfolge: 0 0 0 1 1 0 1 1 0 0

Bei einer faltungscodierten Bitfolge kann eine beliebige Anzahl von Fehlern korrigiert werden, solange die Fehler nicht zu dicht benachbart sind (aufeinander folgen). Komplexere Faltungscodes erzeugen aus mehreren Datenbits mehrere Prüfbits, d.h. der Inhalt eines Datenbits wird auf viele Prüfbits „verschmiert“. Dafür können dann auch benachbarte Fehler korrigiert werden.

Die Dekodierung ist nicht trivial und erfordert einen gewissen Rechenaufwand. Es kommt meist der Viterbi-Algorithmus zum Einsatz, mit dem anhand eines Zustandübergangsdiagramms (Trellis-Diagramm) Decodierung und Fehlerkorrektur vorgenommen werden. Ein Faltungscode kommt z. B. beim Mobilfunkstandard GSM zum Einsatz.

2.2.2.3 Fehlererkennung und -korrektur

Hamming-Distanz

Die Hamming-Distanz „HD“ zwischen zwei Codewörtern ist gleich der Anzahl der unterschiedlichen Bits zwischen diesen beiden Codewörtern.

Beispiel:

a → „10100101“

b → „10010110“

Die Hamming-Distanz zwischen der Codierung für „a“ und „b“ beträgt HD = 4.

Die HD_{min} ist das Minimum aus der Menge aller möglichen Distanzen von Codewörtern, die in einem Codierungssystem (Abbildungsvorschrift) erzeugt werden.

Beispiel:

a → „10100101“

b → „10010110“

c → „11100001“

Damit ergeben sich folgende Distanzen:

$HD(a, b) = 4$

$HD(b, c) = 6$

$HD(a, c) = 2$

Die Hammingdistanz beträgt somit:

$HD_{min} = \text{Minimum } \{4, 6, 2\}$
 $= 2$

Codierungsvorschrift

Eine Codierungsvorschrift erzeugt einen gültigen Code, wenn jedem Symbol, welches codiert werden soll, ein Codewort derart zugeordnet wird, dass bei einer Decodierung dem Codewort das ursprüngliche Symbol auch wieder eindeutig zugeordnet werden kann.

Definition: Hamming-Distanz

Unter der HD^a versteht man die Anzahl von Bits, in denen sich zwei Codewörter unterscheiden. Die Mindest-Hamming-Distanz HD_{min} stellt die kleinste Hamming-Distanz dar.

Es gilt:

- e Fehler erkennen: $HD_{min} = e + 1$
- e Fehler korrigieren $HD_{min} = 2 \cdot e + 1$

^aHamming-Distanz

Es ist einsichtig, dass sich dazu jedes Codewort um mindestens ein Bit von jedem anderen Codewort unterscheiden muss. Die Hammingdistanz für eine gültige Codierung beträgt daher $HD_{min}=1$ (siehe Abbildung 2.23, Zeile a)).

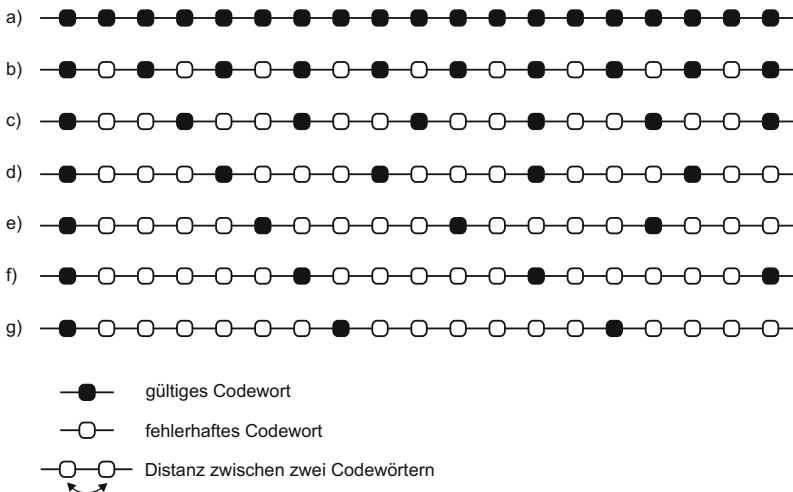


Abbildung 2.23: Schema für Codes mit den Hammingdistanzen $HD_{min} = 1$ a) bis $HD_{min} = 7$ g)

Bei einer 1-Bit-fehlererkennenden Codierung (Code mit Paritätsbit) beträgt die Hammingdistanz zwischen zwei gültigen Codewörtern (fehlerfreien Codewörtern) $HD_{min} = 2$. Damit führt jeder Ein-Bit-Fehler zu einem ungültigen Codewort. Eine Erkennung aller 1-Bit-Fehler ist so sichergestellt. Eine Fehlerkorrektur ist dagegen hier nicht so ohne Weiteres möglich, denn die Distanz des fehlerhaften Codewortes zu den beiden benachbarten gültigen Codewörtern ist im ungünstigsten Fall gerade $d = 1$. Damit ist, wie in Abbildung 2.23, Zeile b) zu sehen ist, keine Entscheidung möglich, welchem gültigen Codewort das fehlerhaft emp-

1-Bit-Fehler-erkennung

fangene Codewort beim Korrekturvorgang zugeordnet werden soll. Ein und dieselbe Codierungsvorschrift lässt auf der Empfängerseite eine unterschiedliche Interpretation des Fehlerbildes und damit des eingesetzten Erkennungs- und Korrekturvorgangs zu. Zur optimalen Interpretation des Fehlerbildes ist eine genaue Kenntnis des Gesamtsystems erforderlich („Fail-Safe“-Verhalten, Auswirkungen eines Fehlers, größter möglicher Schaden usw.). In Abbildung 2.24 werden die unterschiedlichen Fehlerinterpretationen exemplarisch für einen Code mit der Hammingdistanz von $HD_{min} = 6$ (siehe Zeile f) der Abbildung 2.23) durchgeführt.

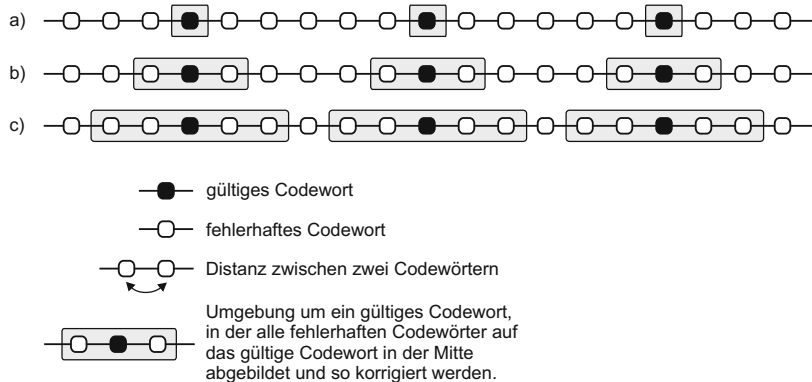


Abbildung 2.24: Korrekturumgebung eines Code mit $HD_{min} = 6$

Fehlerinterpretation A

Fehlerinterpretation A, Abbildung 2.24, Zeile a): Hier werden nur korrekte Codewörter ausgewertet, eine Fehlerkorrektur findet nicht statt. Mit $HD_{min} = 6$ ist daher eine Erkennung von bis zu fünf fehlerhaften Bits möglich.

Fehlerinterpretation B

Fehlerinterpretation B, Abbildung 2.24, Zeile b): Um ein gültiges Codewort wird eine Umgebung gelegt (hier grau unterlegt), in der alle fehlerhaften Codewörter dem korrekten Codewort in der Mitte zugeordnet werden (Entscheidung nach der maximalen Wahrscheinlichkeit¹⁴). Im vorliegenden Fall heißt dies, dass alle 1-Bit-Fehler erkannt und korrigiert werden. Alle 2-, 3- und 4-Bit-Fehler werden nur erkannt. Da sie außerhalb der Umgebung liegen, sind sie nicht korrigierbar.

Fehlerinterpretation C

Fehlerinterpretation C, Abbildung 2.24, Zeile c): Hier ist die Umgebung für die Korrektur vergrößert worden. Es werden alle 1- und 2-Bit-Fehler erkannt und korrigiert. Ein 3-Bit-Fehler fällt genau in die Mitte zwischen zwei Umgebungen und kann daher nicht eindeutig einem gültigen Codewort zugeordnet werden. Er ist somit nur erkennbar und nicht korrigierbar.

¹⁴engl.: Maximum Likelihood Method

Ein 4-Bit-Fehler würde in die Umgebung des benachbarten gültigen Codewortes fallen und daher „falsch“ korrigiert werden. 4-Bit-Fehler oder Fehler > 4 -Bit sind daher weder erkennbar noch korrigierbar.

In der Tabelle 2.5 sind alle möglichen Fehlerinterpretationen für die Hammingdistanzen $HD_{min} = 1$ (Zeile a)) bis $HD_{min} = 7$ (Zeile g)) aufgeführt (vergleiche auch Abbildung 2.23).

Weiterführende Literatur zum Thema Information und Codierung findet man unter [Ham86].

2.2.2.4 Kanalkapazität

Die Kanalkapazität stellt das Maximum der übertragbaren Transinformation pro Zeiteinheit (Informationsfluss F_{max}) dar. Hierbei ist die Transinformation der Informationsgehalt, der fehlerfrei von A nach B über einen gegebenen Kanal übertragen werden kann.

$$C = F_{max} = \left(\frac{H_T}{T}\right)_{max} [\text{Bit/s}]$$

Hierbei ist:

- C: Kanalkapazität
- F: Informationsfluss ($\frac{H_T}{T}$)
- H_T : mittlere Transinformation
- T: mittlere Zeiteinheit

Den mittleren Transinformationsgehalt eines einzelnen analogen Abtastwertes hat Shannon wie folgt bestimmt:

$$H_T(AW^{15}) = \text{Id} \sqrt{\frac{S+N}{N}}$$

Hierbei ist S die Signal- und N die Störleistung.

Des Weiteren gilt das Abtasttheorem mit: $T_{AW} \leq \frac{1}{2 \cdot B}$

oder $1/T_{AW} = 2 \cdot B$ Abtastwerte pro Sekunde bei gegebener Bandbreite B.

Hieraus ergibt sich der mittlere Transinformationsgehalt eines analogen Kanals:

$$H_T = 2 \cdot B \cdot T \cdot \underbrace{\text{Id} \sqrt{\frac{S+N}{N}}}_D = B \cdot T \cdot \text{Id} \left(1 + \frac{S}{N}\right)$$

Hierbei sind:

- $2 \cdot B \cdot T$: Anzahl der Abtastwerte in der Zeit T
- $\text{Id} \sqrt{\dots}$: Transinformation eines Abtastwertes (mit Id^{16})
- D: Dynamik

H_T kann als Nachrichtenquader (Kanten $D, T, (2 \cdot B)$) interpretiert werden.

**Trans-
informations-
gehalt**

**Nachrichten-
quader**

¹⁵Abkürzung: AbtastWert

¹⁶Logarithmus zur Basis 2

Zeile	HD_{min}	Fehlererkennung	Fehler-korrektur	Bemerkung
a)	1	keine	keine	Voraussetzung für eine gültige Codierung
b)	2	alle	1-Bit-Fehler	keine Paritätsprüfung
c1)	3	alle 1- und 2-Bit-Fehler	keine	
c2)	3	alle 1-Bit-Fehler	alle 1-Bit-Fehler	
d1)	4	alle 1-, 2- und 3-Bit-Fehler	keine	
d2)	4	alle 1- und 2-Bit-Fehler	alle 1-Bit-Fehler	
e1)	5	alle 1-, 2-, 3- und 4-Bit-Fehler	keine	
e2)	5	alle 1-, 2- und 3-Bit-Fehler	alle 1-Bit-Fehler	
e3)	5	alle 1- und 2-Bit-Fehler	alle 1- und 2-Bit-Fehler	
f1)	6	alle 1-, 2-, 3-, 4- und 5-Bit-Fehler	keine	
f2)	6	alle 1-, 2-, 3- und 4-Bit-Fehler	alle 1-Bit-Fehler	
f3)	6	alle 1-, 2- und 3-Bit-Fehler	alle 1- und 2-Bit-Fehler	
g1)	7	alle 1-, 2-, 3-, 4-, 5- und 6-Bit-Fehler	keine	
g2)	7	alle 1-, 2-, 3-, 4- und 5-Bit-Fehler	alle 1-Bit-Fehler	
g3)	7	alle 1-, 2-, 3- und 4-Bit-Fehler	alle 1- und 2-Bit-Fehler	
g4)	7	alle 1-, 2- und 3-Bit-Fehler	alle 1-, 2- und 3-Bit-Fehler	

Tabelle 2.5: Fehlererkennungs- und Korrekturmöglichkeiten nach dem Schema von Abbildung 2.23

Umcodierung

Abbildung 2.25 zeigt die redundanzfreie Umcodierung bei $H_{T1} = H_{T2}$ (gleiches Volumen der Quader 1 und 2). Ist die Übertragungszeit gleich ($T_1 = T_2 = T$), so kann Bandbreite „B“ gegen Dynamik „D“ ausgetauscht werden.

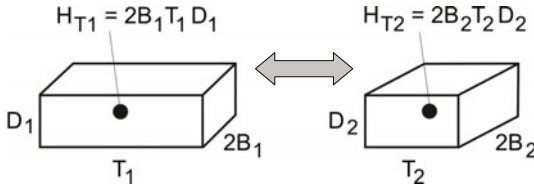


Abbildung 2.25: Nachrichtenquader und Umcodierung zwischen Quader 1 und Quader 2

Merksatz: Kanalkapazität

Um einen fehlerfreien Informationsfluss zu gewährleisten, muss der Kanal mindestens eine Kanalkapazität von $C = F_{max}$ haben. Hierbei ist die Kanalkapazität proportional zu Bandbreite B und Dynamik D.

Die Kanalkapazität nach Shannon ergibt sich dann wie folgt [Der99]:

$$C = \frac{H_T}{T} = B \cdot \lg\left(1 + \frac{S}{N}\right)$$

Beispiel: Aufgrund der hohen Bandbreite bei UWB^a werden hohe Datenraten erreicht (siehe Abschnitt 6.1).

^aUWB = Ultra Wide Band

2.2.3 Leitungscodierung

Bei der drahtlosen Datenkommunikation gibt es natürlich keine Leitungscodierung im eigentlichen Sinne. Das Äquivalent zur Leitung ist hier die Funkverbindung. In diesen Funkkanälen gibt es äquivalente Anforderungen zu denen einer Kabelverbindung.

Bitsynchronisation

Die seriell ausgesendeten Datenbits werden auf der Empfängerseite zeitsynchron abgefragt. Aus dem elektrischen Leitungszustand (Pegel, Frequenz, Phase) bei

Synchronisation

drahtgebundener Übertragung oder dem momentanen Zustand des elektromagnetischen Feldes (Pegel, Frequenz, Phase, je nach Modulationsart) wird die ursprünglich ausgesendete Datenbitfolge ermittelt.

Zur zeitlichen Synchronisation dieser Abfrage muss ein Taktsignal zur Verfügung stehen. Während bei der drahtgebundenen Übertragung in manchen Fällen dafür noch eine separate Taktleitung zur Verfügung steht, ist es unrentabel, einen eigenen Funkkanal für Bitsynchronisation zwischen Sender und Empfänger zu reservieren. Daher muss bei drahtloser Übertragung der Taktgenerator aus den Impulsfolgen der empfangenen Daten synchronisiert werden. Dies setzt eine gewisse Anzahl von Wechseln der Signalzustände voraus, die nicht unterschritten werden darf („Gleichspannungsfreiheit“ der Signale, NRZ¹⁷-Code).

Ein anderes Verfahren besteht darin, dass senderseitig nach längeren Folgen von Bits gleicher Polarität ein Bit gegenteiliger Polarität in den Datenstrom eingefügt wird. Damit ist sicher gestellt, dass Folgen gleicher Polarität nie eine bestimmte Länge überschreiten, unabhängig von den auszusendenden Nutzdaten. Dieses Bit-Stopfen¹⁸ wird natürlich auf der Empfängerseite rückgängig gemacht¹⁹.

Beispiel: Zeile a) zeigt eine originale Datenbitfolge mit längeren Folgen von Einsen (7x) und Nullen (6x). In Zeile b) ist durch das Bit-Stuffing-Verfahren die Länge der Bit-Folgen gleicher Polarität auf max. fünf begrenzt worden. Nach jeweils fünf Datenbits gleicher Polarität wurde hier je ein Datenbit gegenteiliger Polarität eingefügt. Damit stehen genügend Bitwechsel für eine Taktsynchronisation zur Verfügung.

a) 0 1 0 1 1 1 1 1 1 1 0 0 1 0 1 0 0 0 0 0 0 1 0 1 0

b) 0 1 0 1 1 1 1 1 0 1 1 0 0 1 0 1 0 0 0 0 0 1 0 1 0

Auf der Empfängerseite wird mit der empfangenen Bitfolge (Zeile b)) ein „De-Stuffing“ durchgeführt, um die originale Bitfolge wiederherzustellen.

Störungen in Nachbarkanälen

Bei drahtgebundener elektrischer Signalübertragung erfolgt stets auch eine Abstrahlung der Informationen durch elektrische oder magnetische Felder. Für benachbarte Leitungen sind diese Aussendungen Störquellen. Bei ungünstiger Konstellation, d. h. bei hohen Frequenzanteilen im Signal und zugleich ungenügender Abschirmung kann es zu störenden Beeinflussungen und somit zu Übertragungsfehlern auf den Nachbarleitungen kommen.

¹⁷NRZ = Non Return To Zero

¹⁸engl.: Bit Stuffing

¹⁹engl.: Bit Destuffing

Bei drahtloser Signalübertragung werden elektromagnetische Felder auf reservierten Kanälen (Frequenzbereichen) abgestrahlt. Außerhalb dieser Bereiche sind Abstrahlungen unerwünscht und in der Regel durch nationale Gesetze sogar verboten, denn es kann zu störenden Beeinflussungen und damit zu Übertragungsfehlern in den Nachbarkanälen kommen.

Durch geeignete Signalformen kann das Spektrum eines Signals so gestaltet werden, dass in nur geringem Maße hohe Frequenzanteile auftreten bzw. nur minimale Aussendungen in den Frequenzbereichen der Nachbarkanäle erfolgen. Die dabei angewandten Verfahren, z. B. GMSK²⁰, werden im Abschnitt 2.3.2 behandelt.

2.2.4 Datenschutz und Datensicherheit

Der Begriff „Datenschutz“ ist ein Oberbegriff für den Schutz personenbezogener Daten und bezieht sich nicht nur auf technische Schutzmaßnahmen. Zum Beispiel fallen darunter auch Gesetze über das Verbot der Weitergabe von Daten.

Wenn sich die Schutzmaßnahmen auf technische Systeme beschränken, spricht man von Datensicherheit oder Informationssicherheit. Damit bezeichnet man den Schutz von Daten hinsichtlich gegebener Anforderungen an deren Vertraulichkeit, Integrität und Verfügbarkeit.

Im Folgenden werden die Begriffe „Vertraulichkeit“, „Integrität“ und „Verfügbarkeit“ aus dem Blickwinkel der „drahtlosen Nahbereichs-Kommunikation“, also während einer Datenübertragung betrachtet.

Vertraulichkeit²¹ heißt, dass Daten lediglich von autorisierten Benutzern gelesen bzw. modifiziert werden dürfen. Dies gilt nicht nur beim Zugriff auf gespeicherte Daten, sondern insbesondere auch während der Datenübertragung. Während einer drahtlosen Nachrichtenübertragung sind die Daten besonders anfällig für Datenspionage, denn für den erfolgreichen Mitschnitt der übertragenen Daten sind in der Regel keine mechanischen Hindernisse (z. B. gesperrte Räume oder Firmenareale, Anzapfen von Leitungen usw.) zu überwinden.

Integrität²² heißt, dass Daten während einer Übertragung nicht unbemerkt verändert werden dürfen. Aufgetretene, meist zufällige Veränderungen der Daten durch Störungen im Übertragungskanal bezeichnet man als Übertragungsfehler. Gezielt durchgeführte Änderungen der Daten durch einen Angreifer sind Datensabotage. Beide Veränderungen müssen auf der Empfangsseite so erkennbar sein, dass sie korrigiert werden können. Ist eine Korrektur nicht möglich, muss die Tatsache, dass ein oder mehrere nicht korrigierbare Änderungen aufgetreten

Datenschutz

**Daten-
sicherheit**

**drahtlose
Kommunikation**

Vertraulichkeit

Integrität

²⁰GMSK = Gaussian Minimum Shift Keying

²¹engl.: Data Privacy

²²engl.: Data Security

sind, dem Empfänger und möglichst auch dem Absender der Nachricht mitgeteilt werden.

Verfügbarkeit

Verfügbarkeit²³ heißt, dass ein Nachrichtenkanal bei Erfüllung vorgegebener Randbedingungen, wie zum Beispiel räumliche Entfernung, Energieversorgung usw., für eine stabile Nachrichtenverbindung zeitunabhängig zur Verfügung steht. Stabil heißt hier, dass eine vorgegebene Fehleranzahl pro Zeiteinheit nicht überschritten wird bzw. dass aufgetretene Fehler erkannt und korrigiert werden können.

Der Vergleich von einem drahtlosen mit einem drahtgebundenen Übertragungskanal ist im Hinblick auf die Verfügbarkeit differenziert zu betrachten.

Grundsätzlich ist jede drahtlose Kommunikation anfälliger gegen Störungen auf dem Übertragungswege als eine drahtgebundene. Abschirmmaßnahmen gegen Störsignale sind nutzlos, denn sie würden ja gerade die drahtlose Kommunikation behindern oder gänzlich ausschließen.

Sabotage

Prinzipiell ist jede drahtlose Kommunikation nicht vollständig gegen Sabotage zu schützen, denn es müssen keine mechanischen Hindernisse (siehe oben unter Vertraulichkeit) überwunden werden, um einen Datenverkehr zu beeinträchtigen oder gänzlich auszuschließen. Es muss lediglich im Übertragungsband des Systems ein starker Sender ausstrahlen, der den Empfang blockiert. Dies verstößt zwar gegen jedes nationale Fernmelderecht, jedoch ist die Strafbarkeit einer Handlung nie eine Garantie dafür, dass sie nicht doch durchgeführt wird. Durch Verschleierung (siehe unten) oder Richtfunk kann in begrenztem Umfang eine Sabotage erschwert werden.

starker Sender

Es gibt jedoch Fälle, in denen drahtgebundene Kommunikation durch unsicheres Terrain geführt werden muss. Wenn kein hinreichender mechanischer oder personeller (Wachpersonal) Schutz der Nachrichtenkanal möglich ist, kann eine drahtlose Verbindung Vorteile beim Schutz gegen Sabotage bieten. Als Beispiel sei hier eine drahtgebundene Alarmanlage genannt: Diese kann durch Zerstörung der Leitungsverbindung außer Betrieb genommen werden. Der technische Aufwand, eine drahtlose Verbindung zu stören, ist in der Regel wesentlich höher als der Aufwand, ein Kabel zu kappen.

Zusammenfassend sei ausdrücklich darauf hingewiesen, dass der Einsatz einer drahtlosen Verbindung trotz höherer Sabotageanfälligkeit keinesfalls bedeutet, dass ein Angreifer die Kontrolle über das Nachrichtensystem übernehmen kann. Jede Beeinträchtigung der Verfügbarkeit kann auf der Sende- und der Empfängerseite erkannt werden und sollte bei einer fachgerechten Systemimplementierung das System wieder in einen gesicherten Zustand überführen (engl.: Fail-Safe-Verhalten).

²³engl.: Data Savety

2.2.4.1 Begriffe

Während die Codierung bzw. Verschlüsselung im Kanalcodierer ausschließlich dazu dient, die Nachrichtenübertragung gegen systembedingte Störungen (Rauschen usw.) im Kanal zu immunisieren, hat eine kryptographische Verschlüsselung mehrere Ziele (siehe a) bis e)):

a) Einem Dritten, dem „Angreifer“, soll es unmöglich gemacht werden, den Inhalt einer Nachricht zu verstehen (= Geheimhaltung einer Nachricht).

b) Dem Angreifer soll es unmöglich gemacht werden, den Inhalt einer Nachricht zu verändern (= Integrität einer Nachricht).

c) Es soll dem Empfänger möglich sein, den Absender, bzw. die Herkunft einer Nachricht eindeutig zu identifizieren (= Authentifizierung einer Nachricht).

d) Es kann erforderlich sein, dass der Empfänger einen Beweis haben muss, dass der Absender tatsächlich die beim Empfänger eingetroffene Nachricht versandt hat, der Absender also nicht leugnen kann, die Nachricht versandt zu haben (= Verbindlichkeit einer Nachricht).

e) In besonderen Fällen soll sogar die Tatsache, dass zwei Partner miteinander kommunizieren, geheim gehalten werden. (= Verschleierung einer Nachricht). Bei einer drahtlosen Nahbereichs-Kommunikation sind nicht alle der aufgelisteten Anforderungen relevant. Einige der obigen Anforderungen können zudem quasi als Nebeneffekt durch geeignete Modulationsverfahren erfüllt werden.

Geheimhaltung

Integrität

Authentifizierung

Verbindlichkeit

Verschleierung

2.2.4.2 Geheimhaltung durch Verschlüsselung

Das Thema Verschlüsselung (Kryptographie) ist ebenso umfangreich wie komplex. Es werden ständig neue Techniken und Verfahren gefunden und weiterentwickelt. Dieses Thema kann deshalb an dieser Stelle nur tangiert werden. Es wird stattdessen auf entsprechendes aktuelles Spezialschrifttum verwiesen.

Algorithmen zum Verschlüsseln einer Nachricht sind heutzutage leicht aus dem Internet ladbar und auf einem Personalcomputer ablauffähig. Sie erreichen bereits eine so hohe Sicherheit gegen Angriffe, dass sie nur mit einem großen professionellen Rechenaufwand geknackt werden können.

Prinzipiell sind diese Verschlüsselungsverfahren auch im drahtlosen Bereich einsetzbar. Die wichtigsten Kriterien zur Beurteilung der Verfahren unter dem Gesichtspunkt des Einsatzes bei einer drahtlosen Datenkommunikation sind:

Kryptographie

- Verschlüsselungsdauer
- Entschlüsselungsdauer
- Sicherheit gegenüber Angriffen

Die Verschlüsselungs- und Entschlüsselungsdauer sollten natürlich mit der Aktualitätsdauer einer Nachricht (Verfallsdatum) korrespondieren. So ist zum Beispiel der drahtlos übermittelte Messwert eines Temperatursensors meist zeitlich nicht so lange schützenswert wie eine Textdatei, die ein Strategiekonzept enthält oder sonstige Firmeninternas. Doch auch hier sei Vorsicht angebracht: Eine Einzelfallanalyse sollte unbedingt klären, ob nicht zum Beispiel auch durch Temperaturverläufe Rückschlüsse auf schützenswerte Eigenschaften industrieller Prozesse oder Produktionsabläufe gezogen werden können.

Eine häufig praktizierte Angriffsart ist der sog. Brute-Force-Angriff²⁴, auch Exhaustionsmethode genannt. Hier werden einfach alle möglichen Schlüssel bzw. Passwörter eines Verschlüsselungssystems durchprobiert. Es ist eine Frage der Zeit und der zur Verfügung stehenden Rechenleistung, bis ein Angreifer ein System geknackt hat. Durch Vergrößern der Schlüssellänge wird der zeitliche Aufwand rapide vergrößert, allerdings wird auch der Verschlüsselungsaufwand größer. Einen gewissen Schutz auch bei kleinen Schlüssellängen bietet der Einbau eines Zeitfaktors in die Schlüsselabfragefunktion. Wird ein Schlüssel mehrfach falsch eingegeben, werden weitere Eingabeveruche für eine bestimmte Zeit nicht mehr akzeptiert. Damit kann auch bei kurzen Schlüssellängen die Zeitdauer für einen Brute-Force-Angriff sehr hoch gesetzt werden.

Ohne weitere Schutzmaßnahmen besteht hier aber die Gefahr, dass ein Angreifer Systemstörungen auslösen kann. Er kann zwar nicht ins System gelangen, aber durch mehrfache fehlerhafte Schlüsseleingaben den Systemzugang für autorisierte Nutzer zeitweise blockieren.

Der Mitschnitt von Daten in einem Kommunikationssystem heißt nicht, dass der Angreifer diese auch lesen kann. Sind die verschlüsselten Daten jedoch gespeichert, kann zu einem späteren Zeitpunkt, wenn der Schlüssel einmal geknackt wurde, eine Dechiffrierung durchgeführt werden. Durch häufige Schlüsselwechsel, am besten automatisch, z. B. zeitabhängig, kann dies verhindert werden. Ein häufiger Schlüsselwechsel ist zu empfehlen, damit der Schaden durch einen einmal geknackten Schlüssel begrenzt wird.

2.2.4.3 Datenintegrität und Datenauthenzizität

Eine Methode, die Veränderung einer Nachricht durch eine zufällige Störung im Übertragungskanal (also nicht durch einen Angriff mit dem Ziel, die Nachricht zu verfälschen) zu erkennen, wurde oben bei der Behandlung des CRC²⁵ (siehe Abschnitt 2.2.2.1) erwähnt. Diese Methode ist jedoch als Integritäts- und Authentizitätsschutz ungeeignet. Wenn das Prüfpolynom bekannt ist (und es ist nicht allzu schwer, es zu ermitteln), kann jede Nachricht geändert (= Integritätsverletzung) oder gefälscht werden (= Authentizitätsverletzung, da ein

²⁴ engl.: Brute Force Attack

²⁵ CRC = Cyclic Redundancy Check (deutsch: Zyklische Redundanzprüfung)

anderer Absender vorgespiegelt wird).

Die Integrität und Authentizität einer Nachricht wird üblicherweise durch Ergänzung der Nachrichtenbitfolge um eine digitale Signatur nachgewiesen. Dazu wird auf der Senderseite aus den Nachrichtenbits eine Bitfolge (eine Signatur) komprimiert. Die ermittelten Bits der Signatur werden am Ende der Nachricht angefügt oder zwischen den Nachrichtenbits eingestreut.

Der Algorithmus zur Erzeugung dieser Signatur ist geheim und nur dem Absender bekannt. Man spricht hier auch von einer Hash-Funktion. Es müssen alle Nachrichtenbits Einfluss auf den Wert der Signatur haben. Die Nachricht selbst muss dabei nicht zwingend verschlüsselt sein, denn Ziel ist hier nicht deren Geheimhaltung, sondern deren Schutz vor Manipulation und Fälschung.

Die Unversehrtheit einer Nachricht (Korrektheit von Nachricht und Signatur) kann schnell und leicht mit einem öffentlich bekannten Algorithmus verifiziert werden.

Damit eine Signatur als sicher gilt, ist es erforderlich, dass man aus dem öffentlich bekannten Prüfalgorithmus keine Schlüsse auf den geheimen Algorithmus zur Erzeugung einer Signatur ziehen kann. Es gibt mathematische Verfahren, die dies sicherstellen. In der Praxis heißt das, dass man einen sehr großen Rechenaufwand (z. B. Einsatz von mehreren Jahren Rechenzeit) betreiben muss, um aus dem (öffentlichen) Signaturprüfalgorithmus den geheimen Algorithmus zur Signaturgenerierung zu ermitteln.

Es ist jedoch nicht gänzlich auszuschließen, dass neue Verfahren gefunden werden, die diesen Vorgang beschleunigen. Und letztlich nimmt die zur Verfügung stehende Rechenleistung stetig zu. Verfahren, die vor zehn Jahren noch als „sicher“ galten, können heute durch Vernetzung mehrerer Personalcomputer bereits in Stunden oder wenigen Tagen „geknackt“ werden. Aus diesem Grund ist die Zertifizierung von Signaturverfahren stets zeitlich limitiert.

Speziell in der drahtlosen Übertragungstechnik bestehen noch andere Möglichkeiten, in einem gewissen Umfang Integrität und Authentizität einer Nachricht zu garantieren. Wenn bei der Modulation Spreiztechniken, wie das Frequenzsprungverfahren (siehe Abschnitt 2.3.5.1), die direkte Spreizspektrumtechnik (siehe Abschnitt 2.3.5.2) oder Codemultiplexing (siehe Abschnitt 2.3.4.3) zur Anwendung kommen, sind in einem gewissen Maße Integrität und Authentizität garantiert, solange die zur Spreizung verwendeten Zufallssequenzen nicht ermittelt werden können.

**Integrität
und
Authentizität**

**drahtlose
Übertragungs-
technik**

2.2.4.4 Verbindlichkeit von Daten

Der Nachweis der Verbindlichkeit einer Nachricht ist im Falle der drahtlosen Nahbereichskommunikation selten erforderlich. In den meisten Fällen stehen Sender und Empfänger unter Kontrolle desselben Betreibers, deshalb reicht der

Nachweis der Authentizität der Nachricht auch als Nachweis der Verbindlichkeit.

In einfachen Anwendungen kann auch ein Mitschnitt des Nachrichtenverkehrs als Protokoll völlig ausreichend sein.

Aufwendiger ist dieser Nachweis durch eigens dafür implementierte Protokoll-elemente in einer höheren Ebene des ISO/OSI-Referenzmodells zu führen. So kann zum Beispiel durch Quittungsbetrieb, Zeitstempel, Durchnummerierung der Nachrichten o. ä. bewiesen werden, dass ein bestimmter Absender tatsächlich auch eine bestimmte Nachricht versandt hat.

2.2.4.5 Verschleierung von Datenverkehr

Die Verschleierung einer Nachricht bzw. eines drahtlosen Nachrichtenverkehrs, bedeutet, dass das Trägersignal von einem Angreifer nicht empfangbar sein darf. Eine Nachrichtenverbindung, die dem Angreifer nicht bekannt ist, wird meist auch keinen Sabotageversuch auslösen. Durch reine Chiffriertechnik ist jedoch die Verschleierung eines drahtlosen Nachrichtenverkehrs nicht möglich. In gewissem Umfang ist dies durch eine Richtfunkverbindungen mit einem kleinen Öffnungswinkel der Sendeantenne realisierbar. Allerdings kann z. B. durch metallische Objekte im Richtstrahl leicht eine Reflexion auftreten, die wiederum eine empfangbare Signalfeldstärke außerhalb des Richtstrahles erzeugen kann. Diese Effekte sind nur schwer vorhersagbar oder kontrollierbar.

Eine andere Möglichkeit besteht durch die Anwendung des Modulationsverfahrens der direkten Spreizspektrumtechnik (DSSS, siehe Abschnitt 2.3.5.2). Ein Nachrichtenverkehr erzeugt hier kein messbares Trägersignal auf einer bestimmten Sendefrequenz. Es ist stattdessen nur ein erhöhter Rauschpegel in einem (Übertragungs-) Frequenzband registrierbar. Wenn nicht mit besonderen Messmethoden in einem Frequenzband gezielt nach einem Nachrichtenverkehr gesucht wurde, bleibt der Nachrichtenaustausch weitgehend verschleiert.

2.2.4.6 Fazit

Durch Verwendung von proprietären Protokollen, Datenformaten, Modulations- und / oder Verschlüsselungsverfahren ist hier ein gewisser Schutz gewährleistet. Viele Hersteller von Wireless-Systemen klammern daher die Thematik „Schutz und Sicherheit“ aus.

Oft bieten Hersteller in ihrem System lediglich eine transparente Datenverbindung an. Transparent heißt hier, dass sie sicherstellen, dass die vom Absender dem Sender übergebenen Daten unverändert von der Empfangstation dem Empfänger übergeben werden. Ist dies durch Übertragungsfehler nicht möglich, erhält der Absender eine Fehlermeldung (negative Quittung).

Verschlüsselung und Entschlüsselung bleiben damit im Aufgaben- und Verant-

wortungsbereich des Betreibers der Datenverbindung. Er muss entsprechend seinen Anforderungen eine Ende-zu-Ende-Verschlüsselung selbst durchführen. Weiterführende Literatur zum Thema Datensicherheit findet man unter ([Tan03], S. 779 ff; [Sch06, Wob98]).

2.2.5 Netzwerke

Die folgenden Abschnitte zeigen zunächst eine mögliche Einteilung der Netzwerke in Kategorien und in die hierbei verwendeten Netzwerk-Topologien.

2.2.5.1 Kategorien

Die Funkübertragungsverfahren lassen sich zur besseren Übersicht für Hersteller und Anwender in Abhängigkeit von der Reichweite grob in sechs Klassen einteilen (siehe auch Abbildung 2.26): Unter WBAN²⁶ versteht man ein Netzwerk zum schnellen und spontanen Austausch von Daten zwischen kleinen tragbaren Systemen. Die Kommunikation findet zwischen Personen in einer Entfernung von wenigen Zentimetern statt. Die Datenrate ist mittel bis hoch bei z. B. Audio- und Videodaten und gering bei Kontaktdaten. Diese Netzwerke sind auch bei medizinischen Anwendungen zu finden, wo die Reichweite den menschlichen Körper umfasst. WBAN spielt bei „Ubiquitous Computing“ eine Rolle, steckt aber noch in den Anfängen.

WPAN²⁷ sind Netzwerke im Nahbereich – bis zu wenigen 10 Metern. Sie spielen

WBAN

WPAN

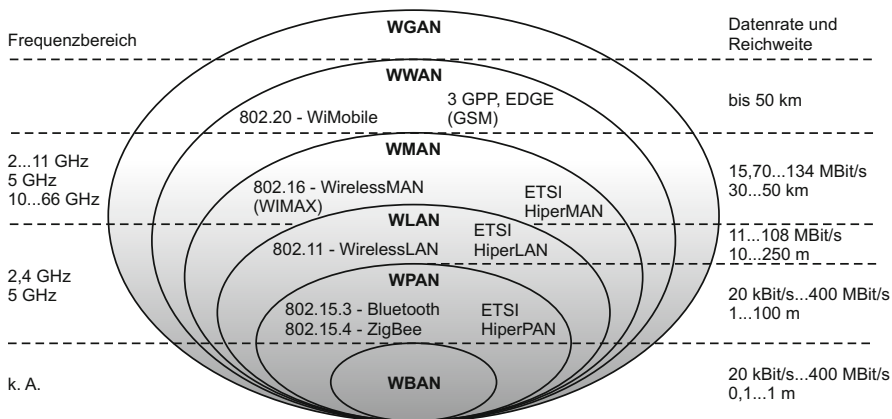


Abbildung 2.26: Kategorien der Funkübertragungsverfahren [Wol07b]

²⁶WBAN = Wireless Body Area Networks

²⁷WPAN = Wireless Personal Area Networks

bei den Ad-hoc-Netzwerken und bei der Kommunikation zwischen mobilen Endgeräten wie Mobiltelefonen und PDA²⁸ eine Rolle. In diesem Bereich findet man Verfahren wie Bluetooth, ZigBee, IEEE 802.15.4 und zukünftig auch UWB²⁹ (siehe hierzu auch Kapitel 6).

WLAN

WLAN³⁰ ist gleichbedeutend mit der Norm IEEE 802.11. Diese Arbeitsgruppe beschäftigt sich mit kabellosem Internet. Die Reichweiten betragen hierbei gut 100 Meter.

WMAN

WMAN³¹ sind Netze, die die kabelgebundenen Breitband-Anschlüsse ersetzen. Sie sind eine Art Wireless DSL. In diesem Zusammenhang steht WiMAX und speziell der Standard IEEE 802.16e/d.

WWAN

WWAN³² sind Netze für die Weitverkehrskommunikation. Hierunter versteht man in Europa GSM und UMTS.

WGAN

WGAN³³ ist die weltweite globale Vernetzung mit Satellitentechnik.

Das Buch beschäftigt sich im Schwerpunkt mit WPAN und WLAN als SRWN.

Definition: **Ad-hoc^a**

ein Hauptmerkmal von Ad-hoc-Netzwerken ist die Selbstkonfigurierbarkeit. Kommunikationspartner, die an einem bestehenden Netzwerk neu teilnehmen wollen, werden dynamisch in das Kommunikationsnetz integriert.

^aspontaner Aufbau von Kommunikationsnetzen

Merksatz: **Automatisierung**

WPAN und WLAN sind im industriellen Bereich (Automatisierung) von besonderer Bedeutung.

Merksatz: **Offenes System**

hierbei tauschen die Kommunikationspartner freizügig Informationen miteinander aus und nehmen gegenseitige Dienstleistungen in Anspruch

Begriffe

Weitere Begriffe in diesem Kontext sind:

- SRWN³⁴: Nahbereichsfunknetzwerke mit Reichweiten von einigen zehn Metern und typischen Datenraten von bis zu einigen 100 kBit/s.

²⁸engl.: Personal Data Assistant

²⁹UWB = Ultra Wide Band

³⁰WLAN = Wireless Local Area Networks

³¹WMAN = Wireless Metropolitan Area Networks

³²WWAN = Wireless Wide Area Networks

³³WGAN = Wireless Global Area Networks

³⁴SRWN = Short Range Wireless Networks

- L3-NET³⁵: Die Abkürzung steht für die drei Zielgrößen wie Low Power, Low Cost, Low Data Rate.
- SAN³⁶: beschreibt, eine häufige Anwendung in der Automatisierung, nämlich die intelligenten Netzwerke aus Sensoren (Dateneingängen) und Aktoren (Datenausgängen).
- M2M³⁷: Kommunikation zwischen Maschinen (siehe [Buf08])

Beispiel: Abbildung 2.27 zeigt die Verbindung von drahtlosen- und drahtgebundenen Netzen.



Abbildung 2.27: Drahtlose- und drahtgebundene Netze im Verbund

2.2.5.2 Abgrenzung

Der Abschnitt stellt die Generationen von Mobilfunksystemen vor und dient als Abgrenzung zu den Nahbereichs-Netzwerken. Mobilfunksysteme der x. Generation werden mit xG³⁸ abgekürzt.

Zu den Mobilfunksystemen der 1. Generation (1G) gehören das A-, B- und Teile des C-Netzes. Alle drei Netze verfügen über eine analoge Sprachübertragung (AMPS³⁹) auf der Funkschnittstelle zwischen Mobilfunktelefon und Basisstation. **1G**

Der GSM-Standard⁴⁰ setzte erstmals ein digitales Übertragungsverfahren zur Sprachübertragung ein. Hierdurch wurde die Kapazität der Funkschnittstelle besser ausgelastet. GSM gehört zu den Mobilfunksystemen der 2. Generation (2G). **2G**

Als Zwischenversionen auf dem Weg zum Mobilfunksystem der 3. Generation (3G) wurden für die Datenübertragung HSCSD⁴¹ und GPRS eingeführt. Beide **3G**

³⁵L3NET = Low Power Low Cost Low Datarate-**NET**work

³⁶SAN = Sensor **A**ctor **N**etworks

³⁷M2M = Machine **T**o Machine

³⁸xG = x**G**eneration

³⁹AMPS = **A**dvanced **M**obile **P**hone **S**ystem

⁴⁰GSM = **G**lobal **S**ystem **F**or **M**obile **C**ommunication

Verfahren wurden als 2.5G eingeordnet – als Zwischenschritt zwischen GSM (2G) und UMTS⁴² (3G). In vielen Dokumentationen wird EDGE⁴³ bereits zur 3. Generation zugeordnet. Der Datendienst reicht allerdings in Verbindung mit GSM nicht an UMTS heran. Aus diesem Grund wird EDGE eher als 2.75G eingeordnet. HSDPA⁴⁴ gehört ebenfalls zur 3. Generation. Hauptbestandteil von Mobilfunksystemen der 3. Generation (3G) sind Datendienste, wie zum Beispiel Videotelefonie und der mobile breitbandige Internet-Zugang.

4G Mobilfunksysteme der 4. Generation befinden sich im Aufbau (siehe auch [KK08]). Hierzu gehört auch WiMAX.

Tabelle 2.6 zeigt die verschiedenen Mobilfunk-Generationen.

Generation	Standard	Übertragung	Datenrate [Bit/s]
1G	AMPS	analog, leitungsvermittelt	k. A.
2G	GSM	digital, leitungsvermittelt	9,6 kBit/s
2.5G	HSCSD	digital, leitungsvermittelt	57,6 kBit/s
	GPRS	digital, paketvermittelt	115 kBit/s
2.75G	EDGE	digital, paketvermittelt	384 kBit/s
3G	UMTS	digital, paketvermittelt	217,6 kBit/s
	HSDPA	digital, paketvermittelt	7,2 MBit/s
4G	WiMAX	digital, paketvermittelt	20 MBit/s

Tabelle 2.6: Mobilfunk-Generationen. Die Abkürzung „k. A.“ steht für keine „Angabe“ [EK08].

Aufgabe: Was versteht man unter Leitungs- und was unter Paketvermittlung?

WiMAX

WiMAX⁴⁵ wird einerseits als stationäre Funkalternative zum Festnetz-DSL⁴⁶ gesehen [WiM08]. Gerne wird es als W-DSL⁴⁷ bezeichnet. Im Gegensatz zu

⁴¹HSCSD = High Speed Circuit Switched Data

⁴²UMTS = Universal Mobile Telecommunication System

⁴³EDGE = Enhanced Data Rate For The GSM Evolution

⁴⁴HSDPA = High Speed Downlink Packet Access

⁴⁵WiMAX = Worldwide Interoperability For Microwave Access

⁴⁶DSL = Digital Subscriber Line

⁴⁷engl.: Wireless-DSL

WLAN ermöglicht WiMAX einen erheblich größeren Durchmesser des Versorgungsbereichs einer Basisstation. Mehrere Kilometer Reichweite lassen die letzte Meile zwischen Netzbetreiber und Kunden schrumpfen. Vor allem in Gegenden, wo DSL oder Kabel keinen Internet-Zugang bieten können, ist WiMAX eine Alternative und wird dort auch schon vermarktet.

Weiterführende Literatur zum Thema Mobilfunk-Generationen findet man unter [Wol05, KK08].

Merksatz: LTE^a-Standard

zählt zur 4. Mobilfunk-Generation (4G^b). Die fünfte Mobilfunk- und Netzwerk-Generation befindet sich in Planung.

^aLTE = Long Term Evolution

^bFunk-Generation

2.2.5.3 Topologien

Der Duden erklärt das Wort Topologie als räumliche Ausdehnung. Bei Netzwerken versteht man darunter die Verbindungen zwischen zwei Knoten (Stationen). Die Netzwerk-Struktur ist abhängig von Kosten, Zuverlässigkeit sowie der Anzahl an benötigten Verbindungen.

Bei drahtlosen⁴⁸ Netzwerken unterscheidet man zwischen den Verbindungsarten ([BGF03], S. 148 ff.; [Jöc01], S. 20 ff.):

drahtlos

- Punkt-zu-Punkt⁴⁹: stellt die einfachste Form dar und ist die direkte Verbindung zwischen zwei Knoten⁵⁰. Die Sternnetze (siehe auch Abbildung 2.28: a) setzen sich aus mehreren P2P-Verbindungen zusammen.

Mehrpunkt⁵¹: Diese Netze verbinden eine Zentrale mit mehreren Außenstellen. Man unterscheidet zwischen Ansprechen einer definierten Gruppe⁵² und dem Rundspruch⁵³, wie z. B. beim Rundfunk.

- Stern: Das Sternnetz verbindet alle Knoten mit einer übergeordneten zentralen Station (siehe Abbildung 2.28: a)).
- Baum: Diese Netze haben einen zentralen Ursprung (siehe Abbildung 2.28: b)). Die Knoten am Ursprung sind wieder Ausgangspunkte für weitere Verzweigungen.

⁴⁸engl.: Wireless

⁴⁹P2P = Point To Point

⁵⁰engl.: Unicast

⁵¹engl.: Point-To-Multipoint

⁵²engl.: Multicast

⁵³engl.: Broadcast

- Maschen: stellen die Verbindung jedes Knotens mit jedem anderen dar (siehe Abbildung 2.29: a)). Diese Verbindungsart wird auch als vollständiger Graph bezeichnet. Maschennetze ermöglichen eine optimale Nachrichtenübermittlung, da im Falle eines Ausfalles Alternativwege zur Verfügung stehen.

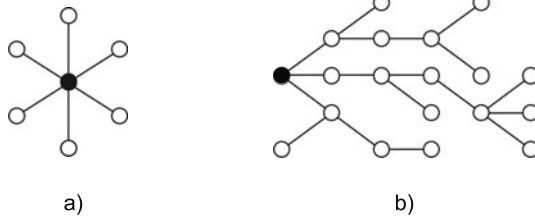


Abbildung 2.28: a) Stern- und b) Baumnetz

draht- gebunden

Bei kabelgebundenen Netzwerken kommen folgende Arten hinzu:

- Liniennetz und Busstruktur: verfügen über jeweils eine Verbindungen zwischen den Knoten. Die Kommunikation erfolgt hierbei von Knoten zu Knoten (siehe Abbildungen 2.30: a) und 2.30: b))
- Ring: Ringnetze bestehen aus P2P-Verbindungen, die einen geschlossenen Ring bilden (siehe Abbildung 2.29: b)). Sie können als Liniennetze mit verbundenen Endpunkten angesehen werden.

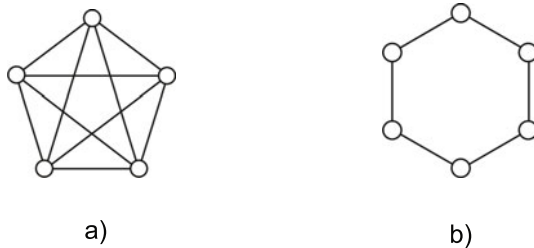


Abbildung 2.29: a) Maschen- und b) Ringnetz

Definition: Netzwerk-Topologie

beschreibt die Struktur der Übertragungswege zwischen den Knoten eines Netzes ([Jöc01], S. 148 ff.).



Abbildung 2.30: a) Linienetz und b) Busstruktur

Aufgaben:

1. Was sind die Vor- und Nachteile eines Maschennetzes gegenüber einem Sternnetz?

Tabelle 2.7 vergleicht die unterschiedlichen Arten von Netzwerken. Tabelle 2.8 vergleicht die drahtgebundenen mit den drahtlosen Standards und gibt hierfür Beispiele. Im Folgenden Erläuterungen zu folgenden Abkürzungen: PCB⁵⁴, P2P⁵⁵, GAN⁵⁶, RFID⁵⁷, LON⁵⁸, ATM⁵⁹, FDDI⁶⁰ und DQDB⁶¹

Netz	Vorteil	Nachteil
Linienetz	relativ geringe Anzahl von Leitungsbündeln notwendig	kein Ersatz bei Ausfall von Knoten oder von Leitungsbündeln
Sternnetz	übersichtliche Netzgestaltung, relativ geringe Anzahl an Leitungsbündeln	bei Ausfall des zentralen Knotens fällt das gesamte Netz aus
Maschennetz	bei Ausfall eines Leitungsbündels kann der Verkehr über andere Wege umgeleitet werden	umfangreiches (und unübersichtliches) Netzwerk

Tabelle 2.7: Vergleich der Netzwerke ([Göb99], S. 789 ff.)

⁵⁴PCB = **Printed Circuit Board** (deutsch: *Leiterkarte*)

⁵⁵P2P = **Point To Point**

⁵⁶GAN = **Global Area Networks**

⁵⁷RFID = **Radio Frequency Identification**

⁵⁸LON = **Local Operating Network**

⁵⁹ATM = **Asynchronous Transfer Mode**

⁶⁰FDDI = **Fiber Distributed Data Interface**

⁶¹DQDB = **Distributed Queue Dual Bus**

Entfernung	Drahtgebundene Netzwerke	Verfahren	Drahtlose Netzwerke	Verfahren
System: 1 m	P2P	seriell, USB	WBAN	RFID, proprietär
Raum: 10 m	PAN	RS 485	WPAN	Bluetooth, IEEE 802.15.4 & ZigBee, proprietär
Gebäude: 100 m; Campus 1 km	LAN	IEEE 802.3, LON	WLAN	IEEE 802.11, (HomeRF)
Stadt: 10 km	MAN	IEEE 802.3, (FDDI, DQDB)	WMAN	IEEE802.16 (WiMAX), proprietär
Stadt: 100 km; Kontinent: 1000 km	WAN	IEEE 802.3, ATM	WWAN	GSM, GPRS, UMTS
Planet: 10.000 km	GAN	Internet	WGAN	Satelliten

Tabelle 2.8: Vergleich von drahtgebundenen mit drahtlosen Standards [Sik05, KK05]

Merksatz: Client/Server-Modell

Bei diesem Modell werden die Daten in leistungsstarken Computern gespeichert – dem Server. Dieser ist oft an einem zentralen Ort angeordnet und wird vom Systemadministrator verwaltet. Die Mitarbeiter greifen mit einfachen Rechnern, den Clients, auf die entfernten Daten zu.

Simplex, Halbduplex und Duplex

Simplex⁶² Der Informationsaustausch erfolgt jeweils nur in eine Richtung. Die Endgeräte sind jeweils reine Empfänger oder Sender. Für einen zweikanaligen Gegenverkehrbetrieb werden zwei vollständige Systeme benötigt.

Simplex

Beim Halbduplex⁶³ findet abwechselnd eine zweiseitige Übertragung statt. Die Endgeräte sind sowohl als Sender wie auch als Empfänger mit Umschalter ausgestattet.

Halbduplex

(Voll)duplex⁶⁴ Für den zweiseitigen Austausch von Informationen wird nur ein Übertragungskanal benötigt. Die Endgeräte beinhalten Sender und Empfänger mit einer Weiche.

Duplex

2.2.5.4 ISO/OSI-Referenzmodell

Die Grundlage für heutige Netze bildet das von der ISO⁶⁵ publizierte OSI⁶⁶-Referenz-Modell. Es stellt das zentrale Ordnungselement für die in Kapitel 3 vorgestellten Verfahren und für den späteren besseren Vergleich in Kapitel 4 dar. Die einzelnen Schichten sind wie folgt definiert ([BGF03], S. 77):

- 7 - Anwendungsschicht⁶⁷: Anwendungsdienste, wie elektrische Post und Datentransfer
- 6 - Darstellungsschicht⁶⁸: Überwindung von Heterogenität bei kommunizierenden Rechnern
- 5 - Sitzungsschicht⁶⁹: Sicherstellung von kooperativen Beziehungen
- 4 - Transportschicht⁷⁰: medienunabhängige Steuerung und Überwachung der korrekten Übertragung von Dateneinheiten
- 3 - Vermittlungsschicht⁷¹: Vermittlung von Datenpaketen durch Netze
- 2 - Sicherungsschicht⁷²: Steuerung von Dateneinheiten auf Vermittlungsabschnitten zwischen Knoten; Fehlererkennung und -korrektur
- 1 - Physikalische Schicht⁷³: Übertragung von Bitströmen: Kanalcodierung, Modulation/Demodulation

⁶²Richtungsverkehr

⁶³Wechselverkehr

⁶⁴Gegenverkehr

⁶⁵ISO = International Standard Organisation

⁶⁶OSI = Open System Interconnection

⁶⁷engl.: Application Layer

⁶⁸engl.: Presentation Layer

⁶⁹engl.: Session Layer

⁷⁰engl.: Transport Layer

⁷¹engl.: Network Layer

Schicht	Aufgabe	Analogie
7	Anwendungen (Web-Browser, FTP)	Uhr senden
6	Datencode (ASCII, TIFF, MPEG)	Päckchen packen
5	Verbindungsaufbau/-abbau (SQL)	Paketkarte
4	Flusskontrolle (ACK, Window, Duplex)	Fahrplan
3	Routing, Protokolle (IP)	Container, Pakete
2	LLC: Übertragungsarten (LLC-Frames) MAC: Adressen + Netze (Ethernet)	LKW, Flugzeug
1	PHY: Kabel + Stecker, elektrische + optische Signale (Bits)	Straße

Tabelle 2.9: Aufgaben und Analogie der OSI-Schichten ([BGF03], S. 16)

Tabelle 2.9 (mit den Abkürzungen FTP⁷⁴ und SQL⁷⁵) zeigt die Aufgaben der unterschiedlichen Schichten und eine Analogie zum Postversand einer Uhr. Im Folgenden werden die einzelnen Schichten im Detail erläutert.

Merksatz: Modellvorstellung

Die Anwenderdaten kommen in einen Briefumschlag (Schicht 7). Dieser Briefumschlag kommt in den Briefumschlag der folgenden Schicht 6 etc. Nach der Bitübertragung wird der Umschlag ausgehend von Schicht 1 entpackt.

Physikalische Schicht

Die physikalische Schicht oder Bitübertragungsschicht (PHY⁷⁶) stellt die Aufbereitung und Übertragung von Bits über ein Übertragungsmedium⁷⁷ dar. Hierbei werden die mechanische und elektrische Beschreibung des Mediums, die Anschlüsse und die Randbedingungen spezifiziert. Randbedingungen sind zum

⁷²engl.: Datalink Layer

⁷³engl.: Physical Layer

⁷⁴FTP = File Transfer Protocol

⁷⁵SQL = Structured Query Language

⁷⁶PHY = PHYsikalische Schicht

⁷⁷Kanal

Beispiel Kanalkapazität (siehe Abschnitt 2.2.2), maximale Übertragungsrate und die Teilnehmer. In Funknetzen sind keine Stecker oder Kabel vorhanden. Deshalb definiert diese Schicht 1 die Luftschnittstelle.

Funknetze

Beispiele:

1. Drahtgebundene Übertragungen: Übertragung im Basisband
2. Drahtlose Übertragungen: Optische und Funk-Schnittstelle

Sicherungsschicht

Die Sicherungsschicht hat die Aufgabe, Datenrahmen von einem Ende eines Kanals⁷⁸ zum anderen zu transportieren. Zu den Aufgaben gehören: Zugriffssteuerung auf Kanal, Erstellen von Systemverbindungen und Datenrahmen, Sicherung der Datenpakete durch Prüfsummen, Dienste vom Senden und Empfangen von Paketen.

Im Basisband wird zwischen den beiden physikalischen Datenkanälen Leitungs- und Paketvermittlung unterschieden. Die Leitungsvermittlung wird üblicherweise bei Sprachübertragung eingesetzt (Telefonie). Ist ein leitungsvermittelter Datenkanal (synchroner Datenkanal) eingerichtet, so kann kein anderer Teilnehmer die Übertragung beeinflussen. Von Vorteil ist die definierte Bandbreite, die hierbei exklusiv verfügbar ist. Bei der Paketvermittlung werden Datenpakete mit Zieladresse versandt. Diese Übertragungsart setzt die Speicherung der Pakete bei den Übertragungsgeräten voraus. Das Internet ist ein bekanntes Netz mit Paketvermittlung. Bei der Paketvermittlung kann keine definierte Zeit für die Paketankunft beim Empfänger angegeben werden. Man spricht deshalb von asynchroner Kommunikation.

Leitungs- vermittlung

Paket- vermittlung

Vermittlungsschicht

In der Sicherungsschicht werden Daten von einem Ende der Leitung zum anderen Ende übertragen. Bei der Vermittlungsschicht geht es um die Vermittlung von Netzteilnehmern. Der Schwerpunkt liegt beim Routing⁷⁹.

Man unterscheidet hierbei zwischen den beiden Diensten:

- verbindungsorientiert
- verbindungslos

⁷⁸Übertragungsmedium

⁷⁹deutsch: Wegfindung

verbindungs-orientiert

Verbindungsorientierte Dienste verwenden einen festen Übertragungskanal während der Dauer einer Übertragung. Zu Beginn der Kommunikation wird die Dienstgüte zwischen den Teilnehmern vereinbart.

Beispiele:

1. TCP^a: verbindungsorientiertes Transportprotokoll
2. UDP^b: verbindungsloses Transportprotokoll

^aTCP = **T**ransmission **C**ontrol **P**rotocol

^bUDP = **U**ser **D**ata **P**rotocol

verbindungslos

Die verbindungslosen Dienste basieren auf der Annahme, dass nur Bit-Ströme über Netze transportiert werden und diese Übertragungen unzuverlässig sind. Die Hostrechner übernehmen die Fehlerüberwachung. Dies hat eine Verschiebung der Komplexität in Richtung der Transportschicht zur Folge.

Beispiele:

1. verbindungsorientiert: ist in Analogie zum Telefonsystem konzipiert. Um mit einer Person zu sprechen, nimmt man den Hörer ab, wählt die Nummer, spricht mit dem Teilnehmer und legt im Anschluss auf. Anwendungen sind die Echtzeitübertragung von Audio und Video.
2. verbindungslos: ist in Analogie zum Postsystem entworfen. Die Nachricht (Brief) enthält die vollständige Adresse und wird unabhängig von allen anderen Nachrichten durch das Kommunikationssystem versandt.

Transportschicht

Die Schicht liefert dem Nutzer Dienstleistungen, die eine bestimmte Qualität der Übertragung gewährleisten. Diese Dienste beziehen sich auf den Verbindungsaufbau, die Übertragungsrate und -verzögerung. Transport- und Vermittlungsschicht ähneln einander. Tabelle 2.10 zeigt die QoS⁸⁰ für Multimedia.

⁸⁰QoS = **Q**uality **O**f **S**ervice

Art	Datenrate [MBit/s]	Bitfehler- rate	Paketfehler- rate
Sprache	0,064	$< 10^{-1}$	$< 10^{-1}$
Video (TV-Qualität)	100	$< 10^{-2}$	$< 10^{-3}$
Kompr. Video	2 - 10	$< 10^{-6}$	$< 10^{-9}$
Datentransfer	2 - 100	0	0
Echtzeit	< 10	0	0
Festbild	2 - 10	$< 10^{-4}$	$< 10^{-9}$

Tabelle 2.10: „QoS-Anforderungen“ bei Multimedia ([Wo102], S. 190)

Beispiele:

1. verbindungslos: Verschicken von Postkarte: die Postkarte wird abgegeben und man geht davon aus, dass sie ankommt (keine Empfangsbestätigung)
2. verbindungsorientiert: Versenden von Einschreibbrief mit Rückmeldung. Der Empfänger bestätigt den fehlerfreien Erhalt ([Jöc01], S. 19 ff).

Sitzungsschicht

Die Steuerung der Kommunikation wird in Form einer Sitzung beschrieben. Die Schicht wird ebenfalls als Kommunikationsschicht bezeichnet.

Darstellungsschicht

Diese Schicht führt eine Formatkonvertierung, Verschlüsselung und Komprimierung von diversen Rechnern durch. Oft werden diese Dienste auch auf der Anwendungsschicht ermöglicht.

Die Sitzungs- und Darstellungsschicht entfallen beim TCP/IP-Modell.

**Verschlüsselung,
Komprimierung**

Anwendungsschicht

Das HTTP⁸¹ stellt das Basisprotokoll auf der Anwendungsschicht dar. Ein weiteres Protokoll in dieser Schicht speziell für drahtlose Kleingeräte, wie zum Beispiel Mobiltelefone oder PDAs, ist WAP. OBEX⁸² stellt einen weitgehend geräteunabhängigen Standard zum spontanen Austausch von Daten dar.

⁸¹HTTP = HyperText Transfer Protocol

⁸²OBEX = OBject EXchange Protocol

Definition: Angepasste Protokolle^a

Hierunter versteht man die aus der Informationstechnik bekannten Protokolle – Schichten 3,4 des ISO-OSI-Modells. Sie gewährleisten Interoperabilität und Integration in eine vorhandene Infrastruktur. Beispielhaft ist das Internetprotokoll zu nennen.

^aengl.: Adopted Protocols

Vorteilhaft bei der Verwendung der obigen „Angepassten Protokolle“ ist die optimale Verbindung mit der vorhandenen IT-Infrastruktur.

Das OSI-Modell basiert auf folgenden drei Konzepten [Tan03]:

- Dienste: definieren, was die jeweilige Schicht macht
- Schnittstellen: teilen den Zugriff den darüber liegenden Schichten mit.
- Protokolle: einer Schicht sind die private Angelegenheit der jeweiligen Schicht.

Zuordnung

Im Folgenden wird der Bezug zwischen den einzelnen Schichten und den nachrichten- und kommunikationstechnischen Methoden hergestellt:

- Physikalische Schicht: nachrichtentechnische Methoden wie Funktechnik, Modulation etc. (siehe Abschnitte 2.3.1.2, 2.3.2)
- Sicherungsschicht: kommunikationstechnische Methoden aus Kanal- und Quellencodierung (siehe Abschnitte 2.2.2, 2.2.1)
- Darstellungsschicht: kommunikationstechnische Methoden wie Quellencodierung und Verschlüsselung (siehe Abschnitte 2.2.1, 2.2.4)

TCP/IP-Referenzmodell

Im Folgenden ein Vergleich des ISO/OSI-Modells mit dem TCP/IP-Referenzmodell:

- Physikalische und Sicherungsschicht: zusammengefasst in der „Host-zu-Netz“-Schicht
- Vermittlungsschicht: Internet-Schicht
- Transport- und Anwendungsschicht: siehe ISO/OSI-Modell
- Sitzungs- und Darstellungsschicht: entfallen

Merksatz: Eingebettete Systeme

Die Entwicklung von Eingebetteten Systemen stellt eine Gratwanderung zwischen Miniaturisierung auf der einen und Leistungsfähigkeit auf der anderen Seite dar.

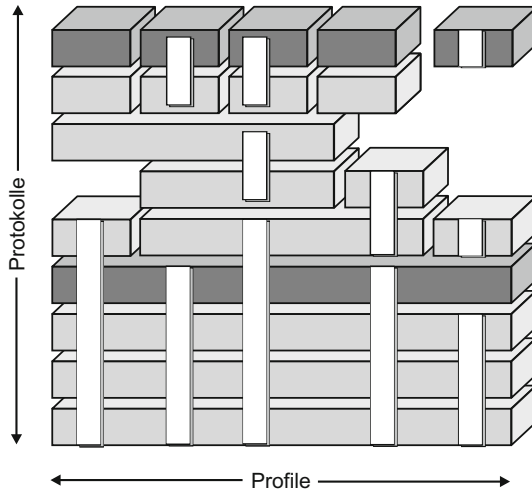


Abbildung 2.31: Protokolle und Profile ([Wol02], S. 223)

Abbildung 2.31 zeigt Protokolle und Profile.
Im Folgenden einige wichtige Definitionen [BGF03]:

Definition: Protokollstapel^a

Hierunter versteht man die Architektur für eine Kommunikation, zum Beispiel TCP/IP.

^aengl.: Stack

Definition: Protokoll

Menge von Vereinbarungen, mit denen eine Kommunikation ausgeführt wird.

Definition: Dienste

für die Erbringung eines Dienstes kann es mehrere Protokolle geben.

Definition: Profile

Kommunikationsarchitektur mit einer bestimmten Menge aufeinander abgestimmter Protokolle und Dienste. Bei „Eingebetteten Systemen“ macht es wenig Sinn, alle Protokolle in allen Geräten zu implementieren. Es kommen unterschiedliche Profile zum Einsatz.

Beispiele:

1. der Dienst File Transfer hat das Protokoll FTP und setzt auf dem TCP/IP-Protokollstapel auf.
2. der Dienst WWW hat das Protokoll HTTP und setzt auf dem TCP/IP-Protokollstapel auf ([BGF03], S. 87).

Aufgaben:

1. Nennen und beschreiben Sie die Schichten des ISO/OSI-Modells.
2. Was versteht man unter Protokollen, was unter Profilen?

2.2.5.5 Protokolle

Der Abschnitt stellt Protokolle der oberen Schichten (ab Schicht 3) des ISO/OSI-Modells vor.

Im folgenden werden die Internetprotokolle kurz erläutert:

- TCP⁸³: Transportschicht (ISO/OSI-Schicht 4) → stellt ein verbindungsorientiertes Protokoll dar.
- UDP⁸⁴: Transportschicht (ISO/OSI-Schicht 4) → verbindungsloses Protokoll; UDP hat weniger Overhead und erlaubt einen schnelleren Verbindungsaufbau als TCP. UDP ist jedoch nicht so zuverlässig und robust wie das TCP-Protokoll.
- IP⁸⁵: Vermittlungsschicht (ISO/OSI-Schicht 3) ermöglicht den netzübergreifenden Datenverkehr, das „routen“ der Daten und die globale Adressierung (erforderlich bei Internetworking). Die Hauptanwendung ist das Internet.
- PPP⁸⁶: Punkt-zu-Punkt-Verbindung (unterhalb IP) – zum Beispiel bei Internet-Zugang via Computer oder PDA⁸⁷.

⁸³TCP = Transmission Control Protocol

⁸⁴UDP = User Data Protocol

⁸⁵IP = Internet Protocol

⁸⁶PPP = Point To Point Protocol

⁸⁷engl.: Personal Data Assistant

Es folgt die Erläuterung einiger Protokolle für mobile Applikationen:

- WAP⁸⁸: Anwendungsprotokoll (ISO/OSI-Schicht 7) – Anbindung von mobilen Endgeräten an das Internet. WAP ermöglicht, spezielle Internetseiten auf mobile Geräte, wie Mobiltelefone oder PDAs zu übertragen.
- OBEX⁸⁹: Anwendungsprotokoll (ISO/OSI-Schicht 7) – Austausch von Daten und Objekten. OBEX stellt einen bewährten Standard dar, bekannt von IrDA.

**mobile
Protokolle**

Beispiel: Abbildung 2.32 zeigt beispielhaft den IP-Aufbau. Weitere Details und Erläuterungen siehe ([BGF03], S. 91)

0	8	16	24	31
Vers	Hlen	Type of Service	Total Length	
Identification			Flags	Fragment Offset
Time to live		Protocol	Header Checksum	
Source IP Address				
Destination IP Address				
Options				Padding
Payload				

Abbildung 2.32: Aufbau des Internet Protocols ([BGF03], S. 91)

Weiterführende Literatur zum Thema „Netzwerke“ findet man unter [Tan03].

2.2.5.6 Standards

Das Normungsgremium „802“ entwickelt Standards für LAN⁹⁰ und MAN⁹¹, hauptsächlich aber Ethernet-Techniken. Die Standards beschreiben im Wesentlichen die beiden Schichten physikalische (PHY: Schicht 1) und Sicherungsschicht (Schicht 2) im ISO/OSI-Modell (siehe Abschnitt 2.2.5.4).

IEEE 802

⁸⁸WAP = Wireless Application Protocol

⁸⁹OBEX = OBject EXchange Protocol

⁹⁰LAN = Local Area Networks

⁹¹MAN = Metropolitan Area Networks

Die Sicherungsschicht wird hierbei unterteilt in MAC⁹² und Logische Verbindungssteuerung (LLC⁹³). Die MAC-Teilschicht⁹⁴ umfasst die Steuerung des Zugriffs auf das Übertragungsmedium und ist somit für den fehlerfreien Transport der Daten verantwortlich. Die LLC-Teilschicht⁹⁵ hingegen ist für die Übertragung und den Zugriff auf die logische Schnittstelle zuständig (siehe Abbildung 2.33).

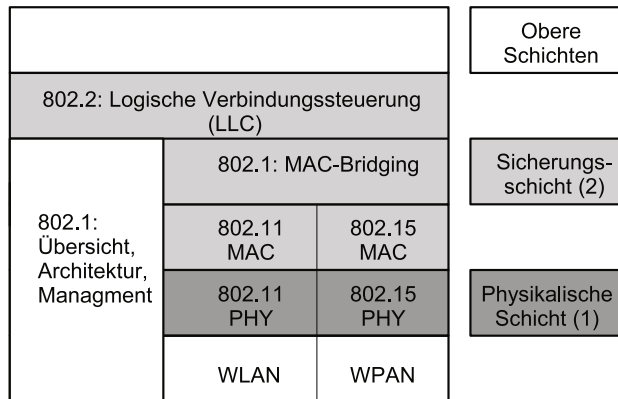


Abbildung 2.33: Übersicht „IEEE 802-Standards“ mit physikalischer Schicht, Sicherungsschicht und oberen Schichten (ISO-OSI-Modell) für WPAN und WLAN

Neben der Standardisierung neuer Übertragungstechniken hat das IEEE 802 die Aufgabe, bestehende Techniken und aufkommende Anwendungen zu optimieren. Einige Standards bauen deshalb aufeinander auf oder hängen voneinander ab. Momentan liegen die drahtlosen Standards, wie 802.11 und 802.15, im Trend.

Definition: IEEE

Das IEEE^a ist eine Organisation von Experten und Fachleuten aus der Elektrotechnik und den Ingenieurwissenschaften (siehe [IEE08b]). Sie ist vergleichbar mit dem deutschen VDE^b.

^aInstitute of Electrical and Electronics Engineers

^bVerband der Elektrotechnik Elektronik Informationstechnik e.V.

⁹²MAC = Media Access Control

⁹³LLC = Logical Link Control

⁹⁴Schicht 2a

⁹⁵Schicht 2b

Im Folgenden werden einige wichtige Standards des IEEE-Projektes 802 vorgestellt (siehe Tabelle 2.11 und Abbildung 2.33).

Bezeichnung	Beschreibung
802.1	Übersicht, Architektur, Management, MAC-Bridging
802.2	Logische Verbindungssteuerung (LLC)
802.3	Ethernet (10Base5) und CSMA/CD-Zugriffsverfahren
802.10	Sicherheit
802.11	Wireless LAN (WLAN) / Drahtlose Netze
802.15	Wireless Personal Area Network (WPAN)
802.15.1	Bluetooth
802.15.3a	UWB - Ultra Wideband Wireless
802.15.4	ZigBee
802.16	WiMAX

Tabelle 2.11: Einige wichtige „IEEE 802-Standards“ im Überblick

Merksatz: Standardisierung

Die wesentlichen Spezifikationen beziehen sich auf die beiden Schichten 1 und 2 des ISO/OSI-Modells.

Tabelle 2.12 zeigt einige wichtige nationale und internationale Normierungsbehörden. Im Folgenden Abschnitt als Fußnoten die Bedeutung der Abkürzungen ITU⁹⁶, ISO⁹⁷, IEC⁹⁸, ETSI⁹⁹, CENELEC¹⁰⁰, IEEE¹⁰¹, DKE¹⁰² und DIN¹⁰³. Die drahtlosen Verfahren im Nahbereich basieren hauptsächlich auf der IEEE-Norm.

**Normierungs-
behörden**

IEEE

2.2.5.7 Kanalzugriffs-Verfahren

Der Abschnitt stellt CSMA¹⁰⁴ im Allgemeinen und die beiden Ausprägungen CSMA/CD¹⁰⁵ und CSMA/CA¹⁰⁶ detailliert vor.

⁹⁶ITU = International Telecommunications Union

⁹⁷ISO = International Standard Organisation

⁹⁸IEC = International Electrotechnical Commission

⁹⁹ETSI = European Telecommunications Standards Institute

¹⁰⁰CENELEC = Comité Européen de Normalisation ELECTrotechnique

¹⁰¹IEEE = Institute of Electrical And Electronic Engineers

¹⁰²DKE = Deutsche Kommission Elektrotechnik

¹⁰³DIN = Deutsches Institut für Normung

¹⁰⁴CSMA = Carrier Sense Multiple Access

¹⁰⁵CSMA/CD = Carrier Sense Multiple Access/Collision Detection

¹⁰⁶CSMA/CA = Carrier Sense Multiple Access/Collision Avoidance

Normung Telekommunikation	Allgemeine Normung	Normung Elektrotechnik	Bereich
ITU	ISO	IEC	international
ETSI	CEN	CENELEC	regional (Europa)
		IEEE	national (USA)
DKE	DIN	DKE	Deutschland

Tabelle 2.12: Einige Normierungsbehörden ([BGF03], S. 569)

Definition: Kanalzugriffs-Verfahren

sind Vorschriften/Algorithmen zur Vergabe des Kanals an sendewillige Stationen. Sie sind Teil der OSI-Schicht zwei, genauer der MAC-Subschicht.

draht- gebunden

CSMA/CD-Verfahren

Dieses Verfahren wird in der drahtgebundenen Übertragungstechnik angewandt. Die Abkürzungen stehen für

- CS: Carrier Sense
- MA: Multiple Access
- CD: Collision Detection

Carrier Sense

Bevor sich eine Station als Sender auf einen Übertragungskanal schaltet, wird überprüft, ob dieser belegt ist (= Carrier Sense). Nur wenn der Kanal frei ist, wird der Sender eingeschaltet, denn sonst würde die auf dem belegten Kanal gerade durchgeführte Datenübertragung gestört werden.

Multiple Access

Ein Zugriffskonflikt kann trotz dieser „Carrier Sense“-Prüfung nicht ausgeschlossen werden. Wenn zwei oder weitere Stationen die „Carrier Sense“-Prüfung eines freien Kanals zeitgleich durchführen, erkennt jede Station, dass der Kanal frei ist und belegt ihn (= Multiple Access).

Collision Detection

Das Ergebnis ist eine Kollision, die zur Folge hat, dass die Datenübertragung aller beteiligten Stationen gestört bzw. unmöglich wird. Es ist deshalb erforderlich, dass eine Kollision erkannt wird (= Collision Detect), damit sie aufgelöst werden kann.

Nun könnte man meinen, dass eine absolut zeitgleiche Abprüfung eines Kanals durch zwei unterschiedliche Stationen extrem selten sei. Das ist wohl auch so. Doch wenn eine Station die Belegung des Übertragungskanals erkennt, wartet sie, dass der Kanals frei wird. Dazu wird die Kanalbelegung ständig abgefragt (Abbildung 2.34). Da dies auch von den anderen wartenden Stationen gemacht würde, wird das Freiwerden aller wartenden Stationen zeitgleich erkannt, und

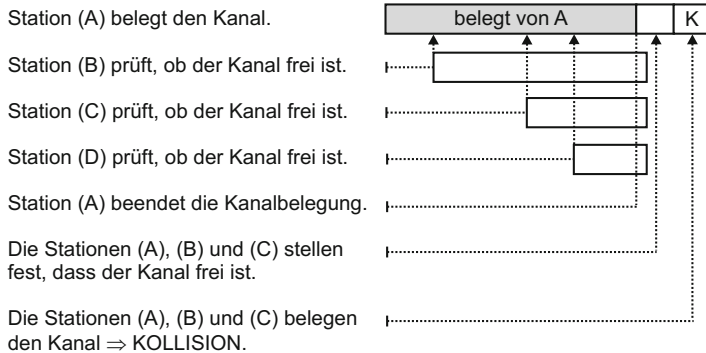


Abbildung 2.34: Kollisionsentstehung beim CSMA-Verfahren

eine Kollision ist die Folge. Es gibt in der drahtgebunden Übertragungstechnik mehrere Verfahren zur Kollisionsauflösung. Beim Ethernet hat eine erkannte Kollision zur Folge, dass sich alle beteiligten Stationen vom Übertragungskanal zurückziehen (abschalten) und erst nach einer von jeder Station für sich berechneten Zufallszeit (Aloha-Prinzip) wieder versuchen, auf den Kanal zuzugreifen. Eine andere Möglichkeit besteht darin, dass in der Schicht 1 (ISO/OSI-Referenzmodell, siehe Abschnitt 2.2.5.4) auf dem Kanal ein dominanter und ein rezessiver Pegel definiert sind. Bei einer Kollision setzt sich der dominante Pegel durch.

Aloha-Prinzip

In der Regel befindet sich bei den Übertragungsprotokollen im Kopf ein Adressfeld, in dem sich alle Stationen voneinander in mindestens einem Bit unterscheiden. Wenn es zu einer Kollision kommt, da zum Beispiel zwei Stationen gleichzeitig senden, so wird dies spätestens bei der Aussendung des Adressfeldes erkannt. Da jede Station ihre eigene Aussendung mitliest, erkennt sie durch die Nichtübereinstimmung von ausgesendetem und empfangenem Bit, dass eine Kollision vorliegt. Sie stellt dann sofort jede weitere Aussendung ein.

Eine mögliche Realisierung eines dominanten und rezessiven Pegels ist die „Open Collector“-Endstufe der TTL-Technik. Dort setzt sich bei einer Kollision stets der „Low“-Pegel durch.

CSMA/CA-Verfahren

Dieses Verfahren ist eine Weiterentwicklung des CSMA/CD-Verfahrens für die drahtlose Übertragungstechnik.

drahtlos

Merksatz: Erkennung Belegung

In der Funktechnik sind die Verhältnisse für die Erkennung der Belegung eines (Funk-)Kanals und für eine Kollisionserkennung völlig anders als in der drahtgebundenen Technik.

Carrier Sense

1) Vor jeder Aussendung prüft die Sendestation, ob der Kommunikationskanal (Sendefrequenz) belegt ist (= Carrier Sense). Diese Prüfung unterliegt aber stets einer gewissen Unsicherheit:

- a) Wenn zwei Stationen den gleichen Kanal benutzen wollen und sie dessen Belegung exakt zeitgleich abprüfen, erhalten beide als Prüfergebnis, dass der Kanal frei sei, und so beginnen beide mit ihren Aussendungen.
- b) Die Abprüfung eines Kanals auf Belegung durch eine Station (A), die an eine Station (B) senden möchte, erfolgt natürlich am Senderstandort von (A). Dort kann eventuell die Aussendung einer dritten Station (C) im gleichen Übertragungskanal durch die lokalen Gegebenheiten nicht empfangbar sein. Fälschlicherweise wird der Kanal als frei erkannt, und (A) beginnt mit ihren Aussendungen.

Während der Fall 1a) eher selten auftritt, kommt der Fall 1b) häufiger vor. Beide Fälle erzeugen nicht erkennbare Kollisionen, denn wenn eine Station sendet, ist deren gleichzeitiger Empfang auf der gleichen Frequenz (fast) unmöglich. Da dies jedoch Voraussetzung für eine Kollisionserkennung ist, ist diese damit ausgeschlossen.

2) Ohne besondere Maßnahmen steigt das Risiko einer unerwünschten Kollision auch dann, wenn ein Kanal korrekt als belegt erkannt wurde. Wenn mehrere Stationen auf einen bereits belegten Kanal zugreifen möchten, werden sie abwarten, bis dieser frei ist, ehe sie den Zugriff durchführen (siehe Abbildung 2.34). Da dies dann zeitgleich durch alle wartenden Stationen geschieht, entsteht eine Kollision (wie beim CSMA/CD-Verfahren bzw. wie oben unter 1a) beschrieben).

Beim CSMA/CA-Verfahren versucht man daher solche nicht erkennbaren Kollisionen zu vermeiden, indem auf einen als belegt erkannten Kanal nicht sofort zugegriffen wird, wenn dieser frei wird. Erst nach dem Ablauf einer für jede Station zufällig bestimmten Zeit (Aloha-Prinzip) wird erneut geprüft, ob der Kanal frei ist, worauf der Vorgang von Neuem beginnt (= CA: Collision Avoidance). In der Vorschrift IEEE 802.11 ist ein CSMA/CA-Verfahren für WLAN standardisiert (siehe Abschnitt 3.2.3).

Collision Avoidance

2.3 Nachrichtentechnik

In diesem Kapitel werden die Vorgänge und Prinzipien bei der drahtlosen Nachrichtenübertragung behandelt.

Die Übertragung analoger Daten (analoge Nachrichtentechnik) wird dabei nur tangiert. Das zentrale Thema der Nahbereichs-Kommunikation dieses Buches ist die Übertragung digitaler Daten. Das schließt jedoch einen Analogbetrieb nicht gänzlich aus, denn analoge Daten können digitalisiert werden, bevor sie einem digitalen Kommunikationssystem übergeben werden. Auf der Empfangsseite müssen sie dann mit einem Digital/Analog-Wandler wieder in analoge Signale rückgewandelt werden.

digitale Übertragung

2.3.1 Einführung und Überblick

Abbildung 2.35 zeigt das allgemeine Schema einer Kommunikation, unabhängig vom Medium im Übertragungskanal. Die von der Quelle erzeugte Nachricht wird in verschiedenen Codierungsebenen möglichst optimal den physikalischen Bedingungen des Nachrichtenkanals angepasst. Optimal heißt hier: fehlerfreie und schnelle Datenübertragung. Die Aufgaben von Quell-, Kanal- und Leitungscodierern und -decodern wurden im Abschnitt 2.2 behandelt.

allgemeines Schema

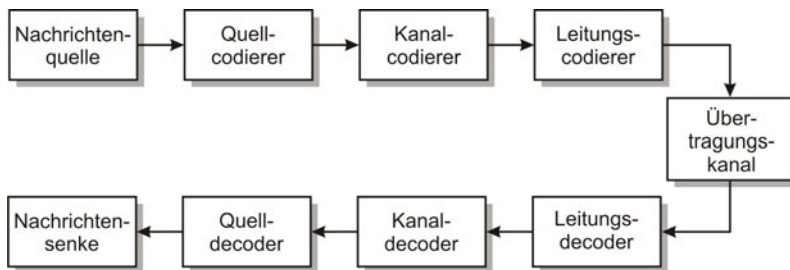


Abbildung 2.35: Allgemeine Struktur eines Kommunikationssystems

2.3.1.1 Drahtlose Sende- und Empfangstechniken

Drahtlose Sende- und Empfangstechniken sind schon aus dem Altertum bekannt. Die hochentwickelte Technik der Buschtrommeln diverser Naturvölker sei hier erwähnt. Signalfeuer, Rauchzeichen und die Zeigertelegraphen (Semaphore) aus der Neuzeit sind optische Verfahren, die eine drahtlose Kommunikation ermöglichen.

Geschichte

Beide Verfahren gibt es im Prinzip auch heute noch in der drahtlosen Nahbereichs-Kommunikation. Es sei hier nur die Ultraschallfernbedienung für Fernsehgeräte genannt. Diese ist allerdings heutzutage bei den audiovisuellen Gerätschaften im Heimbereich weitgehend durch Infrarotfernbedienungen abgelöst worden.

2.3.1.2 Funktechnik

ISO/OSI

Der folgende Abschnitt stellt die wichtigsten Elemente der Physikalischen Schicht (siehe Abschnitt 2.2.5.4) vor.

elektro- magnetische Wellen

Wenn man jedoch von drahtlosen Sende- und Empfangstechniken in der Nahbereichs-Kommunikation spricht, impliziert das in der Regel eine Funkkommunikation, d. h. ein Nachrichtenaustausch mittels modulierter elektromagnetischer Wellen (Felder).

Um eine solche Funkkommunikation zu realisieren, benötigt man bei der Datenquelle eine Sendeanlage¹⁰⁷ und bei der Datensenke eine Empfangsanlage¹⁰⁸. Wenn der Nachrichtenaustausch in beiden Richtungen stattfinden soll, muss an beiden Endpunkten des Kommunikationspfades jeweils eine Sende- und eine Empfangsanlage zur Verfügung stehen.

Die Abbildung 2.36 gibt einen Überblick über die Anordnung der Komponenten eines Funkkommunikationssystems für Daten von einer Quelle zu einer Senke. Die Signalaufbereitung von der Quelle bis zum Modulator erfolgt in einem Rechner und umfasst die Quell- und Kanalcodierung. Diese Themen wurden im Abschnitt 2.2 behandelt.

Basisband- signal

Das vom Kanalcodierer erzeugte Basisbandsignal wird der Sendeanlage zugeführt. Dort erfolgt zunächst die Anpassung der Daten an die Eigenschaften des Funkkanals bzw. an das Trägersignal. Diesem Vorgang entspricht bei der leitungsgebundenen Kommunikation der „Leitungscodierung“. Bei einer Funkkommunikation bezeichnet man es als Modulation. Beim Vorgang der Modulation werden einem im Sender erzeugten Hochfrequenzträger Basisbandsignale aufgeprägt. Es werden dabei die Betriebsparameter des Hochfrequenzträgers in Abhängigkeit vom Basisbandsignal variiert (= moduliert). Im Abschnitt 2.3.2 wird dieses Thema behandelt.

Modulation

Der prinzipielle Aufbau einer Sende- und Empfangsanlage wird im Abschnitt 2.3.3 behandelt.

Die Komponenten Quellcodierer und -decoder, Leitungscodierer und -decoder und Sende- und Empfangsanlage müssen aufeinander abgestimmt sein. Sie arbeiten aber letztlich separat voneinander ohne weiteren Datenverkehr untereinander und zusätzlich zu den Nutzdaten.

ISO/OSI

Beim Kanalcodierer und -decoder ist das anders. Wenn zur sicheren Datenübertragung eine Rückwärtskorrektur implementiert ist (ISO/OSI-Schicht 2), müssen neben den reinen Nutzdaten noch Protokollaten, wie Informationen über die Korrektheit der empfangenen Daten, Quittungen usw., ausgetauscht werden. Diese Zusatzinformationen werden vom Kanalcodierer erzeugt und zusammen mit den Ausgangsdaten des Quellcodierers als Basisbandsignal dem Modulator des Senders (Leitungscodierer) übergeben.

¹⁰⁷engl.: Transmitter

¹⁰⁸engl.: Receiver

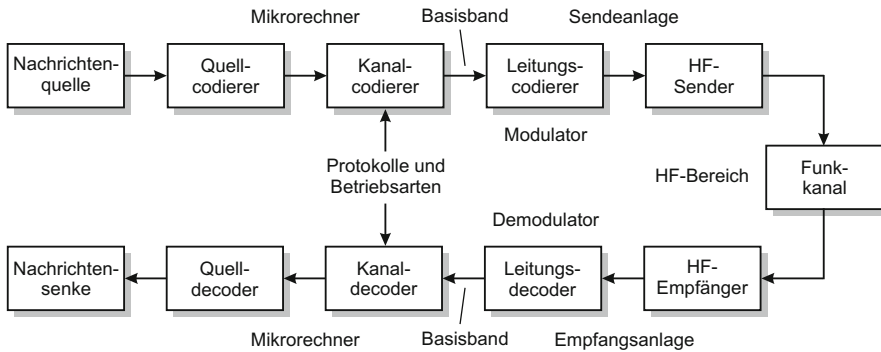


Abbildung 2.36: Struktur eines Funkkommunikationssystems

Auf der Empfängerseite werden diese Zusatzinformationen zur Protokollabwicklung im Kanaldecoder benötigt. Wenn die empfangenen Nutzdaten geprüft und als korrekt erkannt sind, werden sie (ohne Protokolldaten) an den Quelldekodierer weitergegeben.

2.3.1.3 ISM-Band

Das ISM¹⁰⁹-Band kann für Applikationen in den Bereichen Industrie, Wissenschaft und Medizin weltweit lizenzfrei genutzt werden. Im Frequenzspektrum 2,4 GHz findet ein Wettbewerb zwischen zahlreichen Standards und proprietären Funktechniken von unterschiedlichen Anwendungen und Herstellern statt. Dies gilt auch für Geräte des täglichen Gebrauchs, wie z. B. Mikrowellenherde und schnurlose Telefone. In der Praxis wird jedoch verhältnismäßig verantwortungsvoll mit den Frequenzbereichen umgegangen. Dies zeigt sich beim relativ störungsfreien Betrieb von Garagentüröffnern, Autozentralverriegelungen, kabellosen Kopfhörern etc.. Entscheidend beim ISM-Band ist die gleichmäßige Ausnutzung des Bandes. Mechanismen hierfür sind das Frequenzsprung-Verfahren oder ein frequenzwechselnder Code (siehe Abschnitt 2.3.5.1). Im 2,4-GHz-Band stehen die hier diskutierten Verfahren, wie WLAN, ZigBee, Bluetooth, HomeRF, im Wettbewerb. Weitere freie Frequenzbereiche sind 433 MHz, 868 MHz und 5,7 GHz (siehe auch Frequenzbänder in Tabellen 2.13 und 4.2 von Kapitel 4). Das Frequenzband um 13,56 MHz wird bevorzugt für passive RFID¹¹⁰ (Logistik) verwendet.

ohne Lizenz

Beispiel: Tabelle 2.13 zeigt ISM-Frequenzbereiche.

¹⁰⁹ISM = Industrial Scientific Medical¹¹⁰RFID = Radio Frequency Identification

Frequenzband	Mittenfrequenz	Wellenlänge
6765 bis 6795 kHz	6,78 MHz	44,25 m
13553 bis 13567 kHz	13,56 MHz	22,12 m
26957 bis 27283 kHz	27,12 MHz	11,06 m
40,66 bis 40,7 MHz	40,68 MHz	7,37 m
433,05 bis 434,79 MHz	433,92 MHz	69,13 cm
2,4 bis 2,5 GHz	2,45 GHz	12,24 cm
5,725 bis 5,875 GHz	5,8 GHz	5,17 cm
24,0 bis 24,25 GHz	24,125 GHz	1,24 cm
61,0 bis 61,5 GHz	61,25 GHz	4,9 mm
122 bis 123 GHz	122,5 GHz	2,45 mm
244 bis 246 GHz	245,0 GHz	1,22 mm

Tabelle 2.13: ISM-Frequenzbereiche

2.3.1.4 Leistungsbilanz

Die Leistungsbilanz¹¹¹ (siehe auch Abschnitt A.4) beschreibt die gesamte nachrichtentechnische Strecke vom Sender zum Empfänger (siehe Abbildung 2.37). Für die Empfängerleistung gilt:

$$P_R = P_T - C_T + G_T - L_{Fs} - L_{Div} + G_R - C_R$$

Hierbei sind:

- P_R, P_T : Empfangs¹¹²- und Sendeleistung¹¹³
- G_R, G_T : Gewinn¹¹⁴ der Empfangs- und Sendeantenne
- C_R, C_T : Antennenanbindung¹¹⁵ von Empfänger und Sender
- L_{Fs} : Dämpfung¹¹⁶ des freien Raumes (siehe Abschnit 2.3.1.5)
- L_{Div} : Diverse Dämpfungen wie Abschirmung, Regen etc.

Vorzeichen

Mit den Vorzeichen „+“ werden Zuwächse (Gewinne) und mit den Vorzeichen „-“ werden Verluste (Dämpfungen) gekennzeichnet. P_R wird in dB¹¹⁷ angegeben (siehe Anhang A.1). Zur Vereinfachung können $C_R = C_T = L_{Div} = 0$ gesetzt werden (siehe auch Abschnitt A.4).

¹¹¹engl.: Link Budget

¹¹²engl.: Receive

¹¹³engl.: Transmit

¹¹⁴engl.: Gain

¹¹⁵engl.: Connection

¹¹⁶engl.: Loss

¹¹⁷dB = DeziBel

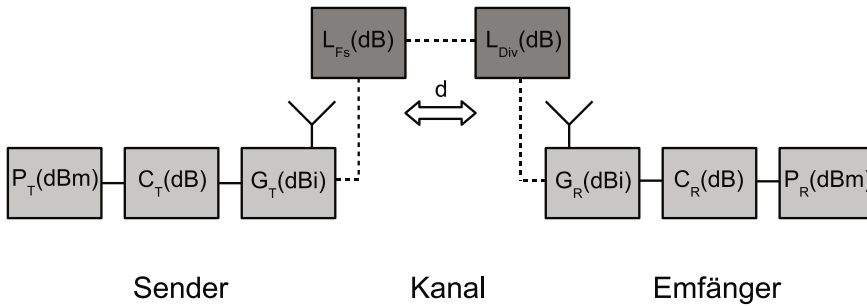


Abbildung 2.37: Leistungsbilanz: $P_R = P_T - C_T + G_T - L_{FS} - L_{Div} + G_R - C_R$

Definition: Empfänger-Empfindlichkeit

Die Empfindlichkeit gibt eine Schwelle (Grenze) für die Sendeleistung am Empfänger an, ab der die Leistung nicht mehr ausreicht. Diese Empfangsleistung (P_R) wird typischerweise mit einer Bit- (BER^a) oder Paketfehlerrate (PER^b) angegeben. Typisch ist ein BER von 10^{-3} (Freifeldausbreitung) – jedes tausendste Bit wird falsch übertragen [Wol07b].

^aBER = Bit Error Rate

^bPER = Packet Error Rate

Beispiele:

1. Der IEEE 802.15.4-Standard verlangt mind. -85 dBm Empfindlichkeit bei 1% PER.
2. Die für alle GSM-Handys vorgeschriebene Mindest-Empfindlichkeit ist gleich -102 dBm, was etwa $63 \cdot 10^{-15}$ Watt entspricht. Diese Empfangsleistung darf für eine störungsarme Verbindung nicht unterschritten werden.

Der Antennengewinn gibt an, in welcher Höhe die Antenne im bevorzugten Winkelsegment die Leistung abgibt beziehungsweise aufnimmt. Hierbei wird ein Vergleich zum isotropen Kugelstrahler hergestellt. Es stellt sich die Frage: „Wieviel Leistung benötigt der isotrope Kugelstrahler, um dieselbe Strahlungsleistung in Vorzugsrichtung abzugeben?“.

**Antennen-
gewinn**

**isotroper
Kugelstrahler**

Apertur- Antennen

Merksatz: Isotroper Kugelstrahler

Er stellt eine in der Realität nicht vorkommende verlustfreie Antenne dar. Sie strahlt die elektromagnetische Leistung in alle Richtungen gleichmäßig ab.

Wird der Antennengewinn auf den isotropen Kugelstrahler bezogen, erfolgt die Angabe in dBi¹¹⁸. In der Nachrichtentechnik wird hingegen vorzugsweise der Bezug zu Halbwellendipolen (siehe auch Abschnitt 2.3.3.3) hergestellt (z. B. Fernsehtechnik). Hierbei kommt die Einheit dBd¹¹⁹ zum Einsatz. Die Umrechnung erfolgt nach der Formel: dBi-Wert = dbd-Wert + 2,15 dB.

Der Antennengewinn ist unabhängig von der Antennenform (Apertur-Antennen¹²⁰): $G = G_{max} = 4 \cdot \pi \cdot A_a / \lambda^2$

Hierbei stellt A_a die geometrische Aperturfläche dar, bezogen auf einen isotropen Kugelstrahler.

Merksatz: Antennentechnik

Zur Abstrahlung elektromagnetischer Wellen muss ein Dipol der Länge der beiden Drahtelemente der halben Wellenlänge ($\lambda/2$) entsprechen (siehe auch Abschnitt 2.3.3.3).

Beispiel: $\lambda/2$ -Antennen haben somit bei einer Frequenz von 2,4 GHz eine Länge von 6,14 cm und bei einer Frequenz von 5 GHz eine Länge von $\approx 2,75$ cm (siehe auch Abbildung 2.68).

Merksatz: Stabantenne

Für eine kurze Stabantenne, auch Marconi-Antenne genannt, gilt für die Bauhöhe: $h = \lambda/4$ (siehe auch Abbildung 2.73). Somit ist h proportional zu $1/f$ – je größer die Frequenz, desto kleiner die Bauhöhe ([Sch78], S. 250 ff).

¹¹⁸dBi = dB isotrop

¹¹⁹dBd = dB dipolar

¹²⁰flächenhaft ausgedehnte Antennen

Beispiel: Die $\lambda/2$ -Antennen werden vorzugsweise als stationäre Antennen eingesetzt. In mobilen Systemen werden aufgrund der geringen Länge $\lambda/4$ -Antennen als Alternative verwendet. $\lambda/4$ -Antennen gehören zu den sogenannten „Groundplane“-Antennen. Sie benötigen eine leitende Fläche als elektrisches Gegengewicht, um die nötige Länge und Funktionsfähigkeit zu gewährleisten (siehe auch Abbildung 2.73).

Die maximale effektive Sendeleistung (EIRP¹²¹) ist für die einzelnen Regionen definiert. In Europa gilt für das 2,4 GHz-Frequenzband eine Obergrenze von 20 dBm und für 5 GHz ein Grenzwert von 30 dBm.

Die Leistungsflussdichte wird nach der Formel:

$$S = \underbrace{P_T \cdot G_T}_{\text{EIRP}} / (4 \cdot \pi \cdot d^2)$$

ermittelt.

Merksatz: Gesundheit

Die ZigBee-Technologie kommt aufgrund ihrer strahlungsarmen Funktechnologie auch in der Medizintechnik zum Einsatz. Beispielsweise wird ZigBee bei Armmanschetten der Firma BodyMedia verwendet. ZigBee gilt im Vergleich zu anderen Datenübertragungs-Technologien als äußerst sicher und strahlungsarm.

Ähnlich wie Bluetooth ist auch bei der ZigBee-Technologie die Strahlungsinintensität im Vergleich zu einem Handy (ein bis zwei Watt) 1000-fach schwächer. Verglichen mit anderen Umwelteinflüssen kann die Wirkung von ZigBee vernachlässigt werden (siehe auch [Bfs08]).

Die EIRP stellt die Strahlungsleistung dar, die ein fiktiver Kugelstrahler mit $G_T=1$ abgeben müsste, um dieselbe Leistungsflussdichte im Betrachtungspunkt zu erzeugen.

Beispiel: Gegeben ist eine 2,4 GHz WLAN-Einheit mit einer Sendeleistung von $P_T=14$ dBm und einem Antennengewinn G_T von 14dBi. Um die maximale Sendeleistung nicht zu überschreiten, ist der Einsatz eines Dämpfungsgliedes mit $C_T=14$ dBm + 14 dBi - 20 dBm = 8 dBm nötig. Als Dämpfungsglied können das Antennenkabel, der Blitzschutz o. ä. verwendet werden.

Weitere wichtige Dämpfungsfaktoren sind (siehe L_{Div} in Abbildung 2.37):

¹²¹EIRP = Equivalent Isotropic Radiated Power

Sendeleistung

EIRP

Dämpfungsfaktoren

- Fresnel-Zone: Berücksichtigung von Hindernissen, die nicht direkt zwischen der Sichtverbindung von Funkzellen liegen (siehe Abbildung 2.38)
- Schwund¹²²-Reserve: Kompensation der Wirkungen von Mehrwegeausbreitung (siehe Abschnitt 2.3.4.1)
- Gebäudedämpfung: In Abhängigkeit von Baustoffen, Größe und Bauart des Gebäudes ist es notwendig, in Gebäuden zusätzlich eine bestimmte Dämpfung einzuplanen (siehe Abschnitt 2.3.1.5).

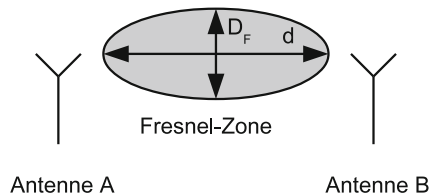


Abbildung 2.38: Fresnel-Zone

Bei der Realisierung von Richtfunkstrecken ist zu beachten, dass die erreichbare Reichweite der Funkzellen auch von Hindernissen beeinflusst wird, die nicht direkt zwischen der Sichtverbindung stehen.

Fresnel-Zone

Fresnel¹²³ hat deshalb gefordert, dass zwischen zwei Richtantennen eine direkte Sichtverbindung vorhanden und ein zusätzlicher Bereich (um diese Sichtverbindung) ebenfalls frei von Hindernissen sein muss. Diese Fresnel-Zone ist ellipsenförmig um die Sichtverbindung zweier Antennen angeordnet (siehe Abbildung 2.38). Der größtmögliche Durchmesser D_F bei $d/2$ kann mit der Formel

$$D_F = \sqrt{d \cdot \lambda}$$

ermittelt werden. Als Richtwert sollten mindestens 80% der Fresnel-Zone verfügbar sein.

Beispiel: Gegeben ist eine WLAN 802.11b-Einheit (2,4 GHz) mit:
Max. Sendeleistung (EIRP): 20 dBm
Sendeleistung: 14 dBm
Antennengewinn: 14 dBi
Empfangsempfindlichkeit: -82 dBm (bei 11 MBit/s)
Dämpfungsglied: 8 dBm
Hieraus ergibt sich für die Freiraumdämpfung (Leistungsbilanz):
 $L_{Fs} = 20 \text{ dBm} + 14 \text{ dBi} - 8 \text{ dB} - (-82 \text{ dBm}) = 108 \text{ dB}$.

¹²²engl.: Fading

¹²³Augustin Jean Fresnel: französischer Physiker

2.3.1.5 Reichweiten-Abschätzung

Grundsätzlich dämpft der Funkkanal (siehe Abbildung 2.37) das Ausgangssignal vom Sender zum Empfänger. Hinzu kommt, dass vom Sender zum Empfänger nicht immer eine direkte Sichtverbindung (LOS¹²⁴) besteht. Reflektions-, Beugungs-¹²⁵ und Abschattungseffekte sind die Folge.

Durch Reflektionen bilden sich Mehrwegeausbreitungen; hierdurch erreichen die Funkwellen den Empfänger zu unterschiedlichen Zeiten. Schwundeffekte sind schnelle Fluktuationen des Signals am Empfänger über sehr kurze Distanzen (in Bezug zur Wellenlänge). Abschnitt 2.3.4.1 liefert hierzu weitere Details.

Die Basis der meisten Ausbreitungsmodelle ist die Ausbreitung der Funkwelle im freien Raum. Hierzu dient die Freiraumformel nach Friis. Die Gleichung stellt die Empfänger-Signalleistung als Funktion der Entfernung d vom Empfänger dar: hierbei bedeutet:

$$P_R(d) = P_T \cdot G_T \cdot G_R \cdot \lambda^2 / ((4 \cdot \pi \cdot d)^2 \cdot L)$$

- d : Distanz zwischen Sender und Empfänger
- $\lambda = c/f$: Wellenlänge
Lichtgeschwindigkeit im Vakuum $c_0 = 2,997925 \cdot 10^8 \text{ m/s}$
- L : Systemverlust (unabhängig von der Wellenausbreitung)

Folgende Annahmen können zur Vereinfachung getroffen werden: $G_R = G_T = 1$ und $L = 1$ (kein Hardware-Verlust). Hieraus resultiert die vereinfachte Gleichung:

$$P_R(d) = P_T \cdot (\lambda / (4 \cdot \pi \cdot d))^2$$

oder als Pfadverlust (L_{Fs}):

$$L_{Fs} = 10 \cdot \log(P_T/P_R) = 10 \cdot \log((4 \cdot \pi \cdot d/\lambda)^2) \text{ [dB]} \quad (2.1)$$

$$\frac{L_{Fs}}{\text{dB}} = 32,44 + 20 \cdot \log\left(\frac{d}{\text{km}}\right) + 20 \cdot \log\left(\frac{f}{\text{MHz}}\right) \quad (2.2)$$

Abschnitt A.1 gibt eine Einführung in die dB-Rechnung.

Aufgabe: Leiten Sie Gleichung (2.2) aus Gleichung(2.1) her.

Der Pfadverlust stellt die Signalabschwächung als Verhältnis Ausgangsleistung (Sender) zu Eingangsleistung (Empfänger) in Dezibel dar. Hierbei ist zu beachten, dass die Freiraumformel nur im Fernfeld bzw. innerhalb der „Fraunhofer-

Funkkanal

Schwundeffekte

Freiraumformel

Fernfeld

¹²⁴engl.: Line Of Sight

¹²⁵Ablenkung von Wellen an einem Hindernis

Region“ ($r_f = (2 \cdot D_a^2)/\lambda$) gültig ist.

Für die Fernfeld-Bedingung muss gelten:

$$r_f \gg D_a \text{ (Antennendimension) und } r_f \gg \lambda \text{ (Wellenlänge).}$$

Aufgaben:

1. Bestimmen Sie für die Freiraumdämpfung $L_{FS} = 108$ dBm (siehe vorheriges Beispiel) die erreichbare Reichweite.
2. Ist die Reichweite realistisch? Begründen Sie Ihre Aussage.
3. Leiten Sie die Formel für den Pfadverlust aus der Freiraumformel nach Friis ab.

Merksatz: Freiraum-Reichweite

Die Reichweite ($n = 2$) mit Tabelle 2.15 beträgt bei $f = 900$ MHz 1050 m und bei $f = 2,45$ GHz 175 m (nur ein Sechstel).

Pfadverlust-exponent

Um den Umgebungseinflüssen (wie Reflexion, Beugung und Abschattung) bei der Ausbreitung gerecht zu werden, wird der Pfadverlustexponent n eingeführt. Hieraus ergibt sich (siehe auch Gleichung (2.1) und (2.2)) ein (allgemeiner) Pfadverlust (L) von:

$$L = 10 \cdot \log(P_T/P_R) = 10 \cdot \log\left(\frac{4 \cdot \pi \cdot d}{\lambda}\right)^n \quad (2.3)$$

$$L = 10,99 \cdot n + 10 \cdot n \cdot \log(d) + 10 \cdot n \cdot \log(1/\lambda) \quad (2.4)$$

Tabelle 2.14 zeigt die Pfadverlustexponenten für unterschiedliche Umgebungen.

Umgebung	Pfadverlust n
Freiraum (freie Sichtverbindung)	2,0
Städtische Umgebung	2,7 - 3,5
Abgeschattete städtische Umgebung	3,0 - 5,0
Innerhalb von Gebäuden, mit Sichtbehinderung	4,0 - 6,0
Innerhalb von Fabriken, mit Sichtbehinderung	2,0 - 3,0

Tabelle 2.14: Pfadverlustexponenten für verschiedene Umgebungen

Aufgaben:

1. Bestimmen Sie die Reichweiten für $f = 900 \text{ MHz}$ und $f = 2,45 \text{ GHz}$ bei $n = 2$.
2. Bestimmen Sie die Reichweiten für $f = 900 \text{ MHz}$ bei $n = 2,4$ und für $f = 2,45 \text{ GHz}$ bei $n = 2,7$.

Tabelle 2.15 fasst die Ergebnisse für die Frequenzbänder 900 MHz und 2,45 GHz zusammen.

Eigenschaft	900 MHz	2,45 GHz
Empfohlene Sendeleistung [dBm]	0	0
Minimale Empfindlichkeit [dBm] (IEEE 802.15.4)	-92	-85
LOS-Reichweite bei 0 dBm Sendeleistung [m] (Theorie)	1050	175
Innerhalb von Fabriken, mit Sichtbehinderung [m] (Theorie)	180	14
Applikation	proprietäre	WLAN, Bluetooth

Tabelle 2.15: Vergleich der Frequenzbänder 900 MHz und 2,45 GHz [Wal05]

Merksatz: Büro-Reichweite

Für bestimmte Büroumgebungen wurden durch Messreihen folgende Werte für n ermittelt: $n = 2,4$ ($f = 900 \text{ MHz}$), $n = 2,7$ ($f = 2,45 \text{ GHz}$). Somit ergeben sich mit Tabelle 2.15 Reichweiten von 180 m ($f = 900 \text{ MHz}$) und 14 m ($f = 2,45 \text{ GHz}$). Das Ergebnis zeigt, dass in Fabrik-Gebäuden $f = 900 \text{ MHz}$ bezüglich der Reichweite günstiger ist als $f = 2,45 \text{ GHz}$ [Wal05].

2.3.2 Modulation

Als Modulation bezeichnet man den Vorgang, bei dem ein Trägersignal mit einem niederfrequenten Informationssignal verknüpft wird. Eine modulierte Signalform besteht daher aus der Trägersignalform und aus der modulierenden Signalform des Informationssignals.

Trägersignale können hochfrequente (sinusförmige) Schwingungen oder Impulsfolgen sein. Im Falle der Funkkommunikation im Nahbereich werden hier ausschließlich hochfrequente sinusförmige Wechselspannungen betrachtet.

Abschnitt 2.3.3 zeigt die Realisierung in Sende- und Empfangsmodulen.

Nahbereich

2.3.2.1 Einführung und Überblick

sinusförmige Wechselspan- nung

Das Trägersignal liegt sowohl bei digitalen als auch analogen Modulationsverfahren der Funkkommunikation im Nahbereich stets als sinusförmiges Wechselspannungssignal vor. Die allgemeine Funktion einer sinusförmigen Wechselspannung u_T in Abhängigkeit von der Zeit (t) lautet:

$$u_T(t) = \hat{u}_T \cdot \cos(2 \cdot \pi \cdot f_T \cdot t \pm \varphi_T) \quad (2.5)$$

$$= \hat{u}_T \cdot \cos(\omega_T \cdot t \pm \varphi_T). \quad (2.6)$$

Die in dieser Funktion enthaltenen konstanten Parameter sind:

\hat{u}_T : Spitzenwert der Wechselspannung

f_T : Frequenz der Wechselspannung

φ_T : Phasenverschiebung der Wechselspannung

ω_T : Kreisfrequenz ($= 2 \cdot \pi \cdot f_T$)

Die Kreisfrequenz ω_T ist keine unabhängige Konstante. Sie wird bei Berechnungen gerne benutzt, um den etwas längeren Ausdruck ($2 \cdot \pi \cdot f_T$) zu vereinfachen.

Alle drei Konstanten \hat{u}_T , f_T und φ_T können zur Übertragung von Information herangezogen werden, indem man sie durch Variable ersetzt, die wiederum Funktionen des zu übertragenden Informationssignals $u_S(t)$ sind.

Die folgenden Abschnitte stellen analoge und digitale Modulationsverfahren vor.

2.3.2.2 Analoge Modulationsverfahren

Bei analogen Modulationsverfahren liegt das Informationssignal $u_S(t)$ in analoger Form vor. Je nach Modulationsart werden die oben (siehe Abschnitt 2.3.2.1) angeführten konstanten Parameter der sinusförmigen Trägerspannung mit dem Informationssignal $u_S(t)$ wie folgt verändert:

- Amplitudenmodulation: \hat{u}_T wird variiert
- Frequenzmodulation: f_T wird variiert
- Phasenmodulation: φ_T wird variiert.

Amplitudenmodulation

Das Informationssignal, also das Signal, mit dem der Träger moduliert wird, lautet (mit $\varphi_S = 0$):

$$u_S(t) = \hat{u}_S \cdot \cos(2 \cdot \pi \cdot f_S \cdot t \pm \varphi_S) \quad (2.7)$$

$$= \hat{u}_S \cdot \cos(\omega_S \cdot t). \quad (2.8)$$

Die Zeitfunktion des Trägersignals lautet nach Gleichung (2.6):

$$u_T(t) = \hat{u}_T \cdot \cos(\omega_T \cdot t), \quad (2.9)$$

wobei hier und bei Gleichung (2.8) auf die Angabe der Phasenverschiebung der beiden Cosinussignale verzichtet wird, denn sie spielen bei der Amplitudenmodulation keine Rolle.

Bei der Amplitudenmodulation (AM¹²⁶) schwankt die Amplitude \hat{u}_T des Trägersignals $u_T(t)$ in Abhängigkeit von der Amplitude des modulierenden Informationssignals $u_S(t)$. Damit erhält man die Zeitfunktion des amplitudenmodulierten Trägersignals $u_{AM}(t)$ ([Peh01], S. 79 ff.):

$$\begin{aligned} u_{AM}(t) &= \hat{u}_T(t) \cdot \cos(\omega_T \cdot t) \\ &= (\hat{u}_T + u_S(t)) \cdot \cos(\omega_T \cdot t) \end{aligned} \quad (2.10)$$

mit $u_S(t) = \hat{u}_S \cdot \cos(\omega_S \cdot t)$ (siehe auch Gleichung (2.8))

$$\begin{aligned} u_{AM}(t) &= (\hat{u}_T + \hat{u}_S \cdot \cos(\omega_S \cdot t)) \cdot \cos(\omega_T \cdot t) \\ &= \hat{u}_T \cdot (1 + m \cdot \cos(\omega_S \cdot t)) \cdot \cos(\omega_T \cdot t) \end{aligned} \quad (2.11)$$

Merksatz: Amplitudenmodulation

Das Signal $u_{AM}(t)$ (siehe Gleichung (2.11)) lässt sich in der Praxis einfach durch eine Addition eines Gleichanteils zum Informationssignal und der Multiplikation mit dem Trägersignal erzeugen.

Ein Maß für die relative Stärke der Trägersignaländerung durch das Informationssignal ist der Modulationsgrad. Er gibt das Verhältnis von Informationssignaländerung zur Trägersignaländerung an:

Modulationsgrad (m):

$$m = \hat{u}_S / \hat{u}_T, \quad (2.12)$$

bzw. als Prozentangabe:

$$m = \hat{u}_S / \hat{u}_T \cdot 100 [\%] \quad (2.13)$$

Abbildung 2.39 zeigt das sinusförmige Informationssignal $u_S(t)$, das Trägersignal $u_T(t)$ und das amplitudenmodulierte Ausgangssignal $u_{AM}(t)$ des Modulators. Alle Signale sind im Zeitbereich dargestellt.

Nach einer Umformung gemäß der Beziehung:

$$\cos \alpha \cdot \cos \beta = 1/2 \cdot \cos(\alpha + \beta) + 1/2 \cdot \cos(\alpha - \beta) \quad (2.14)$$

erhält man mit Gleichung (2.11)

$$\begin{aligned} u_{AM}(t) &= \hat{u}_T \cdot \cos(\omega_T \cdot t) + \\ &\quad m/2 \cdot \hat{u}_T \cdot \cos((\omega_T + \omega_S) \cdot t) + \\ &\quad m/2 \cdot \hat{u}_T \cdot \cos((\omega_T - \omega_S) \cdot t) \end{aligned} \quad (2.15)$$

Zeitbereich

Ansatz

Modulationsgrad

Frequenzbereich

¹²⁶AM = AmplitudenModulation

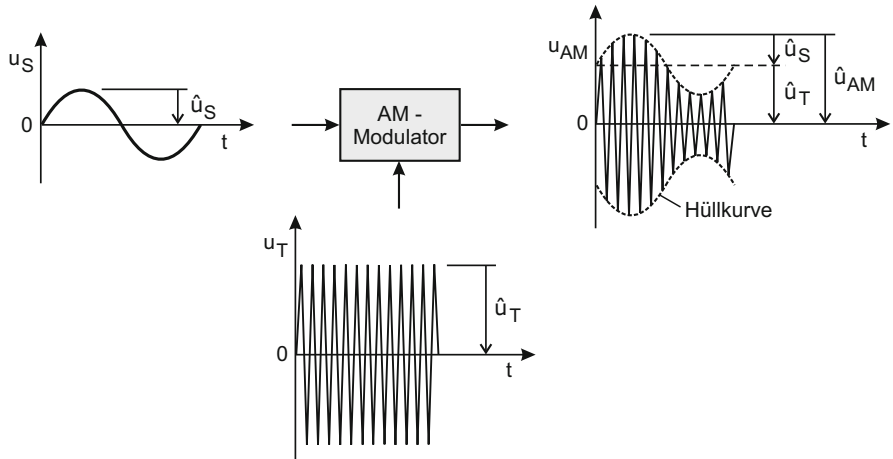


Abbildung 2.39: Amplitudenmodulation eines Trägers mit einem Sinussignal, Darstellung der Signale im Zeitbereich

**Doppel-
seitenband-
signal mit
Träger**

Dieses Signal $u_{AM}(t)$ bezeichnet man als Doppelseitenbandsignal mit Träger. Es enthält die Trägerfrequenz f_T , die Differenz zwischen Trägerfrequenz und Signalfrequenz ($f_T - f_S$) und die Summe von Trägerfrequenz und Signalfrequenz ($f_T + f_S$). Abbildung 2.40 ist die dazugehörige Darstellung im Frequenzbereich. Das modulierte Trägersignal $u_{AM}(t)$ weist rechts und links vom Träger f_T im Abstand von $|f_T - f_S|$ je einen Seitenträger auf.

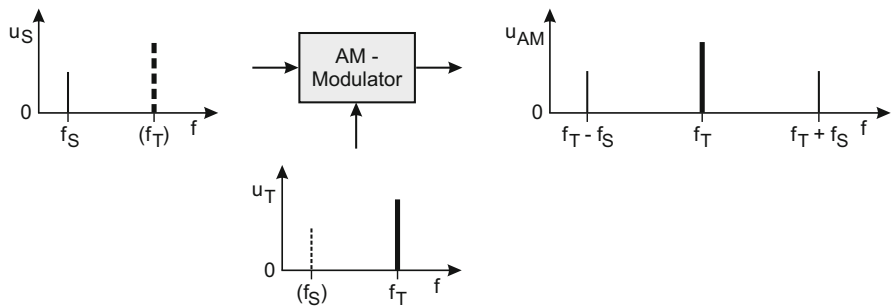


Abbildung 2.40: Amplitudenmodulation eines Trägers mit einem Sinussignal, Darstellung der Signale im Frequenzbereich

Das Informationssignal $u_S(t)$ ist selten sinusförmig, sondern es besteht entsprechend der zu übertragenden Information aus einem Signalamplitudenverlauf, der sich aus vielen Sinussignalen unterschiedlicher Amplituden zusammensetzt.

Dieses zusammengesetzte Informationssignal bezeichnet man als Basisband-signal. Ein Basisbandsignal umfasst einen Frequenzbereich (Bandbreite), der durch untere und obere Schranken (f_{Smin} und f_{Smax}) begrenzt wird. Beim Vorgang der Modulation wird dieses Basisfrequenzband in zwei Seitenbänder, rechts und links vom Träger, umgesetzt (siehe Gleichung (2.15) und Abbildung 2.41).

Basisband-signal

Das untere Seitenband (LSB¹²⁷) bezeichnet man auch als Kehrlage (oder Frequenzinversion), denn hier werden die hohen Frequenzanteile des Basisbandes auf den niedrigen Frequenzbereich des unteren Seitenbandes abgebildet (und entsprechend die niedrigen Frequenzanteile des Basisbandes auf den hohen Frequenzbereich des oberen Seitenbandes), wie auch in Abbildung 2.41 zu erkennen ist. Beim oberen Seitenband (USB¹²⁸) ist das nicht der Fall. Deshalb wird hier von einer Normallage gesprochen.

Seitenbänder

Den Aufbau beider Seitenbänder zeigt die Abbildung 2.41.

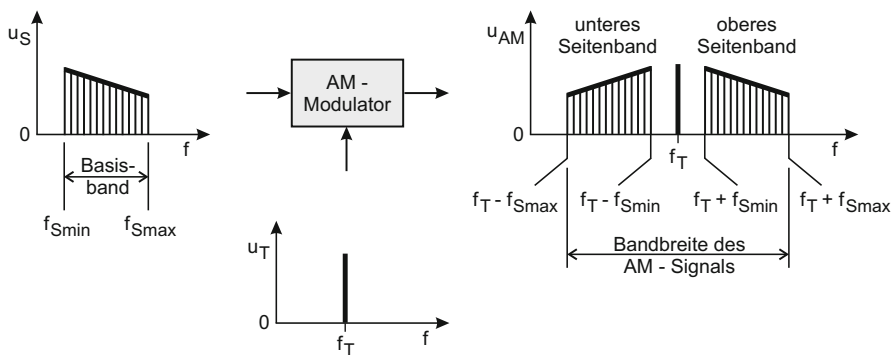


Abbildung 2.41: Amplitudenmodulation eines Trägers mit einem Basisbandspektrum von f_{Smin} bis f_{Smax} , Darstellung der Signale im Frequenzbereich

Die Bilanz der einfachen Amplitudenmodulation ist im Hinblick auf Leistungseffizienz und Bandbreiteneffizienz sehr schlecht.

Effizienz

Auch wenn kein Informationssignal vorliegt, wird z. B. trotzdem ein Trägersignal abgestrahlt (natürlich ohne Seitenbandsignale). Der Trägerleistungsanteil beim AM-Signal beträgt bei einem Modulationsgrad von 100% immerhin 50%.

Wie man der Abbildung 2.41 entnehmen kann, wird das Basisband je einmal oberhalb und unterhalb des Trägers als Seitenband abgebildet. Die von einem amplitudenmodulierten Signal belegte Bandbreite beträgt damit mehr als das Doppelte (maximale Basisbandfrequenz f_{Smax} multipliziert mit 2).

¹²⁷LSB = Lower Side Band (deutsch: unteres Seitenband)

¹²⁸engl.: Upper Side Band

**Doppel-
seitenband**

Beim Sonderfall der Doppelseitenbandmodulation wird der Träger unterdrückt (Gleichung (2.15) mit $\hat{u}_T = 0$). Die Leistungsbilanz ist dann etwas besser, die Bandbreiteneffizienz ist jedoch ebenso schlecht wie bei der einfachen Amplitudenmodulation.

ISB

Eine spezielle Modulationsart ist die unabhängige Seitenband-Modulation (ISB¹²⁹-Modulation). Hier werden das obere und das untere Seitenband mit getrennten und von einander unabhängigen Basisbandsignalen moduliert. Bei unterdrücktem Träger ist die Leistungsbilanz sehr gut. Die von einem ISB-Signal belegte Bandbreite ($2 \cdot f_{Smax}$) ist nur geringfügig größer als die zweier unabhängiger Basisbandsignale ($2 \cdot [f_{Smax} - f_{Smin}]$). Diese Modulationsart erfordert jedoch spezielle Demodulatoren im Empfänger. Deshalb bleibt der Einsatz auf Spezialfälle beschränkt.

SSB

Die effizienteste Modulationsart in Hinblick auf Bandbreitenbelegung und Leistungsverbrauch ist die Einseitenband-Modulation (SSB¹³⁰-Modulation). Hier werden z. B. durch Filterung oder auf anderem Wege der Trägeranteil und ein Seitenband des amplitudenmodulierten Signal unterdrückt (siehe auch Gleichung (2.15)):

$$u_{USB}(t) = \hat{u}_S/2 \cdot \cos((\omega_T + \omega_S) \cdot t) \quad (2.16)$$

Abbildung 2.42 zeigt den Aufbau eines SSB-Signals im Frequenzbereich (hier ein USB-Signal). Man erkennt, dass der modulierte HF-Träger eines SSB-Signals exakt die gleiche Bandbreite wie das Basisband benötigt. Liegt kein Basisbandsignal am Modulator an, entsteht auch kein HF-Signal am Ausgang des Modulators.

Dass Amplitudenmodulation heute noch im Einsatz ist, hat historische Gründe. Eine Umstellung auf effizientere Modulationsarten würde bedeuten, dass zeitgleich auch alle Empfänger ausgetauscht werden müssten. Das ist zurzeit (noch) nicht durchsetzbar.

Frequenzmodulation

Bei der Frequenzmodulation (FM¹³¹, oft auch als Winkelmodulation bezeichnet) schwankt die Frequenz f_T des Trägersignals in Abhängigkeit von der Amplitude des modulierenden Informationssignals (mit $\omega_T = 2 \cdot \pi \cdot f_T$):

$$\omega_T(t) = \omega_T + k_F \cdot u_S(t)$$

k_F : FM-Konstante

Für das modulierte Trägersignal gilt: $\hat{u}_T \cdot \cos(\varphi_T(t))$ und $\omega_T(t) = d\varphi_T(t)/dt$.

Ansatz

¹²⁹ISB = Independent Side Band (deutsch: unabhängiges Seitenband)

¹³⁰SSB = Single Side Band (deutsch: Einseitenband)

¹³¹FM = FrequenzModulation

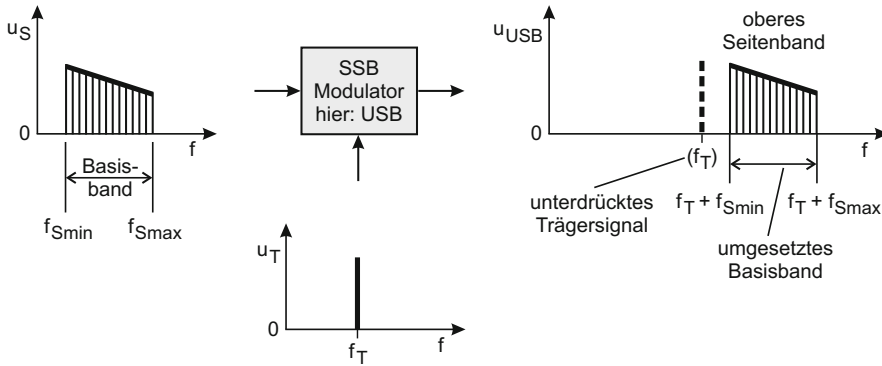


Abbildung 2.42: Einseitenbandmodulation eines Trägers (USB) mit einem Basisbandspektrum von f_{Smin} bis f_{Smax} , Darstellung der Signale im Frequenzbereich

Hieraus folgt allgemein für das frequenzmodulierte Signal $u_{FM}(t)$ ([Peh01], S. 113 ff.): **allgemein**

$$u_{FM}(t) = \hat{u}_T \cdot \cos\left(\int_{t_0=0}^t \omega_T(\tau) d\tau\right) \quad (2.17)$$

mit $u_S(t) = \hat{u}_S \cdot \cos(\omega_S \cdot t + \varphi_S)$ folgt für $\varphi_T(t)$

$$\begin{aligned} \varphi_T(t) &= \int_0^t \omega_T(\tau) d\tau \\ &= \omega_T \cdot t + k_F \cdot 1/\omega_S \cdot \hat{u}_S \cdot \sin(\omega_S \cdot t + \varphi_S) \end{aligned}$$

und für das frequenzmodulierte Signal $u_{FM}(t)$:

$$u_{FM}(t) = \hat{u}_T \cdot \cos(\omega_T \cdot t + k_F \cdot 1/\omega_S \cdot \hat{u}_S \cdot \sin(\omega_S \cdot t + \varphi_S)) \quad (2.18)$$

Die Amplitude des frequenzmodulierten Signals $u_{FM}(t)$ bleibt dabei konstant, denn deren Größe ist unabhängig vom Informationssignal $u_S(t)$ (siehe Abbildung 2.43).

Als Hub bezeichnet man die Auslenkung der Trägerfrequenz bei Ansteuerung des FM-Modulators durch $u_S(t)$. Da der Hub Δf_T nur die Auslenkung in eine Richtung bezeichnet, erfolgt die Angabe als **Hub**

$$Hub = \Delta f_T = 1/(2 \cdot \pi) \cdot k_F \cdot \hat{u}_S \quad (2.19)$$

Der Modulationsindex M ist der Quotient aus Hub und maximaler Signalfrequenz: **Modulations-index**

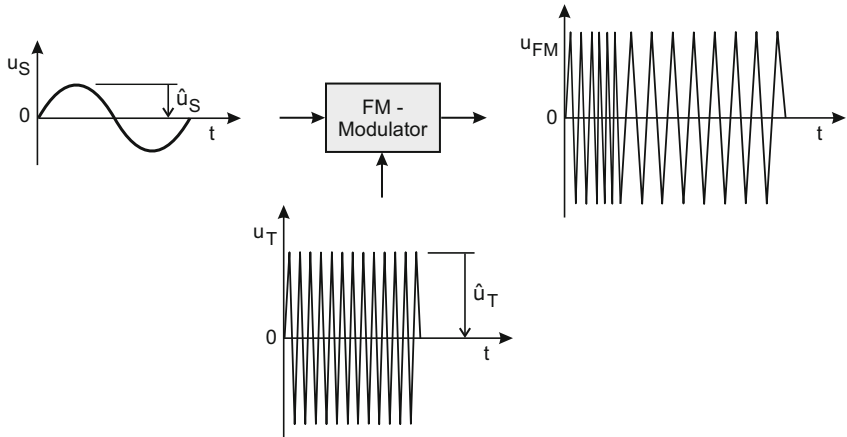


Abbildung 2.43: Frequenzmodulation eines Trägers mit einem Sinussignal, Darstellung der Signale im Zeitbereich

$$\begin{aligned} M &= \Delta f_T / f_{Smax} \\ &= \Delta \omega_T / \omega_{Smax} \end{aligned} \quad (2.20)$$

Bandbreiten-effizienz

Die benötigte Bandbreite B eines frequenzmodulierten Signals wird vom Hub und der maximalen Frequenz des Basisbandes bestimmt und ist theoretisch unendlich groß. Das genaue Spektrum eines FM-modulierten Signals lässt sich mit der Besselfunktion bestimmen. Als gute Näherungswerte gelten: Schmalband-FM:

$$B \approx 2 \cdot f_{Smax} \quad (2.21)$$

Breitband-FM:

$$B \approx 2 \cdot (\Delta f_T + f_{Smax}) \quad (2.22)$$

Leistungs-effizienz

Auch die Leistungseffizienz ist bei der Frequenzmodulation sehr ungünstig. Unabhängig von der Amplitude oder der Frequenz des Modulationssignals $u_S(t)$ wird stets die maximale Sendeleistung vom Sender abgestrahlt.

Dieser schlechten Leistungsbilanz steht ein gravierender Vorteil gegenüber: In einem FM-Empfänger kann leicht eine sehr effiziente Störunterdrückung durchgeführt werden. Störungen in einem Übertragungskanal sind meist Amplitudenstörungen. Wie oben gezeigt, sind jedoch die Informationsinhalte bei FM-Betrieb in der Frequenz des Trägers enthalten bzw. in dessen Nulldurchgängen. Damit können empfangene FM-Signale in ihren Amplituden begrenzt (limitiert) werden, denn diese haben, solange eine Mindestspannung nicht unterschritten wird, keinen Einfluss auf die Qualität des demodulierten Signals.

Durch eine Amplitudenbegrenzung im Empfänger werden Störungen sehr effizient ausgeblendet. Damit ist ein FM-Empfang wesentlich unempfindlicher gegen Amplitudenstörung als ein AM-Empfang, bei dem ja gerade das Informationssignal als Amplitudenänderung übertragen wird.

Phasenmodulation

Bei der Phasenmodulation (PM¹³², oft auch als Winkelmodulation bezeichnet) schwankt die Phase φ_T des Trägersignals $u_T(t)$ in Abhängigkeit von der Amplitude des modulierenden Informationssignals $u_S(t)$ ([Peh01], S. 116 ff.):

Ansatz

$$\varphi_T(t) = \omega_T \cdot t + k_P \cdot u_S(t)$$

k_P : PM-Konstante

Hieraus ergibt sich allgemein das folgende phasenmodulierte Signal $u_{PM}(t)$:

allgemein

$$\begin{aligned} u_{PM}(t) &= \hat{u}_T \cdot \cos(\varphi_T(t)) \\ &= \hat{u}_T \cdot \cos(\omega_T \cdot t + k_P \cdot u_S(t)) \end{aligned} \quad (2.23)$$

für $u_S(t) = \hat{u}_S \cdot \cos(\omega_S \cdot t + \varphi_S)$ gilt:

$$\varphi_T(t) = \omega_T \cdot t + k_P \cdot \hat{u}_S \cdot \cos(\omega_S \cdot t + \varphi_S)$$

hieraus ergibt sich für $u_{PM}(t)$:

$$u_{PM}(t) = \hat{u}_T \cdot \cos(\omega_T \cdot t + k_P \cdot \hat{u}_S \cdot \cos(\omega_S \cdot t + \varphi_S)) \quad (2.24)$$

Phasen- und Frequenzmodulation sind voneinander abhängig:

**Übergang
FM, PM**

$$\varphi = \int \omega dt = 2 \cdot \pi \cdot \int f dt \quad (2.25)$$

Jede Phasenveränderung hat ihre Ursache in einer Frequenzänderung, d. h. jede Frequenzveränderung hat eine Phasenänderung zur Folge. Der Zusammenhang ist in Abbildung 2.44 dargestellt.

Die Leistungseffizienz und die Bandbreiteneffizienz entsprechen weitgehend den Werten der Frequenzmodulation.

Effizienz

Aufgabe: Vergleichen sie die drei analogen Modulationsverfahren AM, FM und PM bezüglich Leistungs- und Bandbreiteneffizienz.

¹³²PM = PhasenModulation

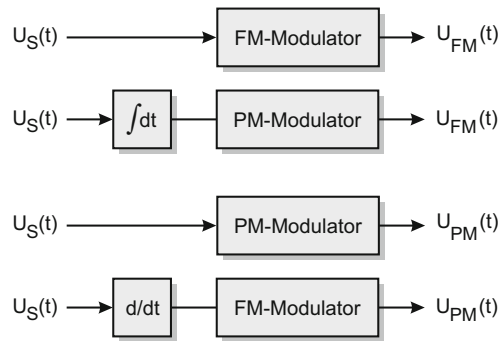


Abbildung 2.44: Zusammenhang zwischen Frequenzmodulation und Phasenmodulation

2.3.2.3 Digitale Modulationsverfahren

sinusförmiger
Träger

Bei den digitalen Modulationsverfahren werden wie bei den analogen Modulationsverfahren die Kennwerte einer sinusförmigen Trägerschwingung moduliert. Im Gegensatz zu den analogen Modulationsverfahren kann hier das Signal, welches einen Träger modulieren soll, nur diskrete Werte (Kennzustände) annehmen. Die Tabelle 2.16 zeigt die Beziehungen zwischen analogen und digitalen Modulationsverfahren. Tabelle 2.16 gibt hierzu einen Überblick:

Modulierendes Signal	analog	digital
Beeinflussung der Amplitude des Trägersignals	AM	ASK
Beeinflussung der Frequenz des Trägersignals	FM	FSK
Beeinflussung der Phase des Trägersignals	PM	PSK
Kombination von grundlegenden Modulationsarten		APSK (QAM)

Tabelle 2.16: Zusammenfassung der grundlegenden Modulationsarten mit sinusförmiger Trägerschwingung (auch Cosinusschwingung)

Im Folgenden die Bedeutung der Abkürzungen: AM¹³³, ASK¹³⁴, FM¹³⁵, FSK¹³⁶,

¹³³AM = AmplitudenModulation
¹³⁴ASK = Amplitude Shift Keying (deutsch: Amplitudenumtastung)
¹³⁵FM = FrequenzModulation
¹³⁶FSK = Frequency Shift Keying (deutsch: Frequenzumtastung)

PM¹³⁷, PSK¹³⁸, APSK¹³⁹ und QAM¹⁴⁰.

Bei digitalen Modulationsverfahren nimmt das Signal, welches einen Träger modulieren soll, nur diskrete Werte (Kennzustände) an. Interessant ist dabei, wie viel Information (Bits) pro Zeiteinheit in Abhängigkeit von der Schrittdauer (Schrittgeschwindigkeit) und der Anzahl der Kennzustände übertragen werden kann. Deshalb werden hier zunächst die Kennwerte digitaler Basisbandsignale vorgestellt.

Weiterführende Literatur zum Thema Modulationsverfahren findet man unter [MG02, Fre00, HL00, Rop06].

Kennwerte

Basisband

Schrittgeschwindigkeit

Die Schrittgeschwindigkeit V_S gibt die Anzahl der Einzelschritte mit der Schrittdauer T_S pro Sekunde an.

$$V_S = 1/T_S \text{ [Baud]} \quad (2.26)$$

Dieser Begriff kommt aus der Fernschreibtechnik und bezeichnet dort die Einheit für die Telefragiegeschwindigkeit Baud. Ein Baud bedeutet „ein Schritt pro Sekunde“.

Baud-Rate

Kennzustände

Im binären System gibt es die logischen Werte „logisch Null“ und „logisch Eins“. Diesen beiden Werten werden zwei physikalischen Kennzuständen zugeordnet, zum Beispiel der Spannungsbereich zwischen 0 und 0,8 Volt dem Wert „logisch Null“ und der Spannungsbereich vom 2,0 bis 5,0 Volt dem Wert „logisch Eins“. Die Dauer eines Symbols entspricht hier der Übertragungsdauer eines Bit.

Hat man nun vier Kennzustände, wie zum Beispiel in Tabelle 2.17 dargestellt, so kann man jedem Kennzustand einen Zwei-Bit-Wert zuordnen.

Das heißt, pro Schritt werden zwei Bits übertragen. Hat man „n“ Kennzustände,

Kennzustände [V]	Bits
-12 bis -9	00
-5 bis -2	01
+2 bis +5	10
+9 bis +12	11

Tabelle 2.17: Kennzustände I

¹³⁷PM = PhasenModulation

¹³⁸PSK = Phase Shift Keying (deutsch: Phasenumtastung)

¹³⁹APSK = Amplitude Phase Shift Keying (deutsch: Amplituden-Phasen-Umtastung)

¹⁴⁰QAM = Quadrature Amplitude Modulation

so werden pro Schritt ($\lg^{141} n$) Bits übertragen. Mit der Beziehung

$$\lg n = \lg n / \lg 2 \quad (2.27)$$

kann man den Informationsgehalt eines Schrittes in [Bit] in Abhängigkeit von den Kennzuständen bestimmen.

Die Tabelle 2.18 zeigt den Informationsgehalt eines Schrittes in Abhängigkeit von den Kennzuständen zwei bis zehn.

Kennzustände [V]	Bits
1	-
2	1,0
3	1,585
4	2,0
5	2,322
6	2,585
7	2,807
8	3,0
9	3,17
10	3,322
usw.	usw.

Tabelle 2.18: Kennzustände II

Übertragungsgeschwindigkeit

Bit-Rate

Die Übertragungsgeschwindigkeit V_{UE} (oder Bit-Rate) gibt an, wie viel Bits pro Sekunde übertragen werden. Sie ist das Produkt aus Schrittgeschwindigkeit und Anzahl der übertragenen Bits pro Schritt:

$$V_{UE} = V_S \cdot \lg n \text{ [Bit/s]} \quad (2.28)$$

Merksatz: Baud-Rate

Bei einem System mit nur zwei Kennzuständen ist die Baud-Rate gleich dem Wert der Bitübertragungsrate.

Symboldauer

In der Übertragungstechnik gibt die Symboldauer (oder Symbollänge) den Zeitraum an, in dem einem Trägersignal die gleichen Daten aufmoduliert werden. Sie entspricht damit der oben definierten Schrittdauer. Wenn dabei nur zwischen zwei Kennzuständen unterschieden wird, beträgt die übertragende Information

¹⁴¹Logarithmus zur Basis 2

während der Dauer eines Symbols, also pro Symboldauer (pro Symbollänge, pro Schritt) genau ein Bit.

Aufgabe: Bei einem ternären digitalen Signal unterscheidet man auf der Leitung zwischen den drei Kennzuständen, zum Beispiel zwischen (+U), (0) und (-U). Die Schrittdauer beträgt 0,1 ms. Wie hoch ist die Übertragungsgeschwindigkeit (Bit-Rate)?

Pro Schritt werden $\lg 3 = 1,585$ Bits übertragen. Es werden pro Sekunde $1\text{s} / (0,1\text{ ms}) = 10 \cdot 10^3$ Schritte durchgeführt. Die Bit-Rate beträgt somit $1,585 \cdot 10 \cdot 10^3 = 15,85\text{ kBit/s}$.

Lösung

Aufgabe: In einem QAM^a-16-System wird zwischen 16 Kennzuständen unterschieden. Die Symboldauer beträgt 0,1 μs . Wie viele Bits werden pro Sekunde übertragen?

^aQAM = Quadrature Amplitude Modulation

Pro Schritt werden $\lg 16 = 4$ Bits übertragen. Es werden pro Sekunde $1\text{s} / (0,1\text{ } \mu\text{s}) = 10 \cdot 10^6$ Schritte durchgeführt. Es werden somit $4 \cdot 10 \cdot 10^6 = 40\text{ MBit/s}$ übertragen.

Lösung

Frequenzbereich des Basisbandsignals (Bandbreite)

Eine binäre Folge hat ihre maximale Frequenz, wenn sie bei jedem Schritt ihren Kennzustand wechselt. Die Periodendauer ist in diesem Fall gleich der Dauer von zwei Schritten (mit den Kennzuständen für „logisch Eins“ und „logisch Null“). Die Grundfrequenz $f_0 = 1 / T_0$ mit $T_0 = 2 \cdot T_S$ beträgt dann

$$f_0 = 1 / (2 \cdot T_S) \quad (2.29)$$

Wenn hier von der „Grundfrequenz“ gesprochen wird, heißt das, dass auch noch andere Frequenzen auftreten. Eine unendlich lange „1-0-1-0-Folge“ ist ein Rechtecksignal, wie es in Abbildung 2.45 zu sehen ist.

Abbildung 2.45a) zeigt den zeitlichen Verlauf der „1-0-1-0-Folge“ in der logischen Ebene. Aus der Schrittdauer von $T_S = 5\text{ ms}$ erhält man die Periodendauer von $T_0 = 2 \cdot T_S$ mit der $T_0 = 10\text{ ms}$. Daraus ergibt sich die Frequenz $f_0 = 1/T_0$ des Rechtecksignals mit $f_0 = 100\text{ Hz}$.

Zu 2.45a)

Abbildung 2.45b) zeigt die dazugehörige Abbildung auf einer physikalischen Ebene

Zu 2.45b)

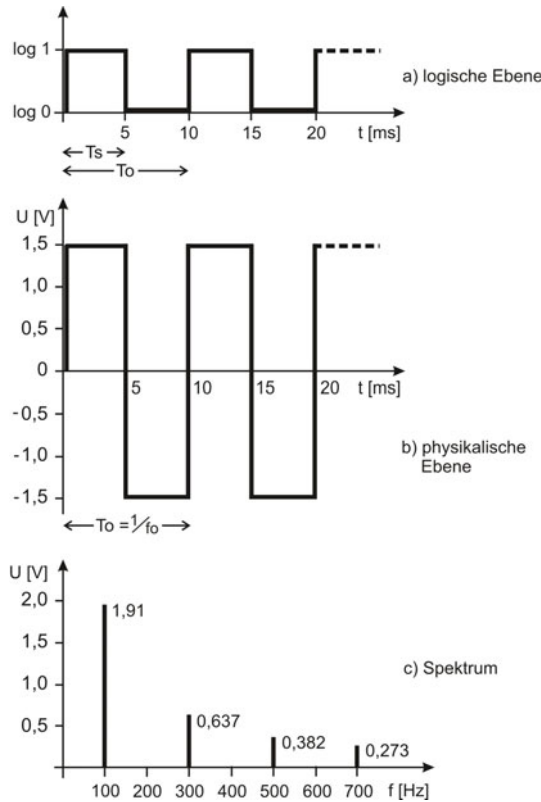


Abbildung 2.45: Darstellung einer „1-0-1-0-Folge“ in verschiedenen Ebenen

ne. Der logischen Null ist der Spannungswert $-1,5$ Volt zugeordnet und der logischen Eins der Spannungswert $+1,5$ Volt, das entspricht einem Amplitudenspitzenwert von $U_S=1,5V$ (oder einem Spitzen-Spitzenwert von $U_{SS}=3V$). Abbildung 2.45c) zeigt das Spektrum des Signals. Das Spektrum eines Rechtecksignals ist theoretisch unbegrenzt. Die Grundschiwingung des Rechtecksignals muss bei 100 Hz liegen. Deren Amplitude, die Frequenzen der Oberschwingungen und die Amplitudenwerte der Oberschwingungen kann man folgender Beziehung entnehmen:

$$u(\omega t) = U_S \cdot \frac{4}{\pi} \cdot (\sin \omega t + \frac{1}{3} \sin 3\omega t + \frac{1}{5} \sin 5\omega t + \frac{1}{7} \sin 7\omega t \dots)$$

Für das obige Beispiel ergeben sich damit die folgenden Amplitudenwerte (siehe Tabelle 2.19) für die ersten fünf Spektren eines Rechtecksignals (mit der Frequenz von $f_0 = 100$ Hz und einer Amplitude von $U_S = 1,5$ V).

Der Fall einer „1-0-1-0-Folge“, also eines Rechtecksignals mit dem Tastverhältnis $1:1$, wäre ein Sonderfall der Datenübertragung. Binäre Datensignale beste-

Zu 2.45c)

f [Hz]	U [V]
100	1,910
300	0,637
500	0,382
700	0,273
900	0,212
usw.	usw.

Tabelle 2.19: Amplitudenwerte für die ersten fünf Spektren nach Abbildung 2.45

hen in der Realität selten aus Schritten gleicher Dauer (T_S). Sie stellen dabei keinen periodischen Vorgang dar, denn die Datenbits ändern unregelmäßig ihre Werte. Eine weitgehend zufällige zusammengesetzte Bitfolge mit der Schrittdauer $T_S = 5\text{ms}$ (d. h. $T_0 = 10\text{ ms}$ bzw. $f_0 = 100\text{ Hz}$) erzeugt das in Abbildung 2.46 dargestellte Amplitudenspektrum.

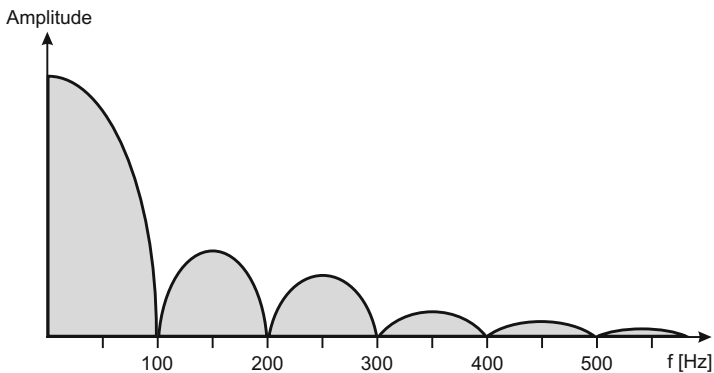


Abbildung 2.46: Amplitudenspektrum eines pseudozufälligen binären Datensignals

Amplitudenumtastung

Bei der Amplitudenumtastung (ASK¹⁴²) wird das digitale Informationssignal $u_S(t)$ auf die Amplitude eines Trägersignals $u_T(t)$ abgebildet. Das Ergebnis dieser Multiplikation zeigt Abbildung 2.47.

ASK

ASK-Modulation ist, historisch betrachtet, die erste Modulationsart, die in der Funktechnik zum Einsatz kam. In der Morsetelegraphie (CW¹⁴³) mussten die beiden Signalzustände „Ein“ und „Aus“ übertragen werden, was bei ASK den

¹⁴²ASK = Amplitude Shift Keying (deutsch: Amplitudenumtastung)

¹⁴³engl.: Continues Wave

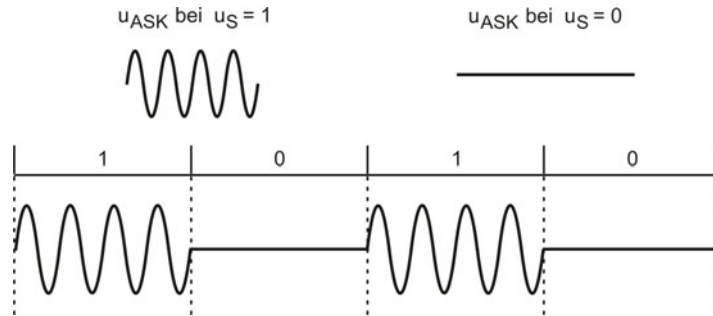


Abbildung 2.47: Abbildung der binären Werte „1“ und „0“ auf die Amplitude eines Trägers

Bit-Zuständen „1“ und „0“ entspricht. ASK bzw. CW kommen heute nur selten zum Einsatz. In Kombination mit der PSK-Modulation wird sie jedoch bei Modulationsverfahren für mehrwertige Codes eingesetzt (QAM).

Frequenzumtastung

FSK

Bei der Frequenzumtastung (FSK¹⁴⁴) wird das digitale Informationssignal $u_S(t)$ auf die Frequenz eines Trägersignals $u_T(t)$ abgebildet. Das Ergebnis zeigt Abbildung 2.48. Bei Frequenzumtastmodulation werden den beiden logischen Kenn-

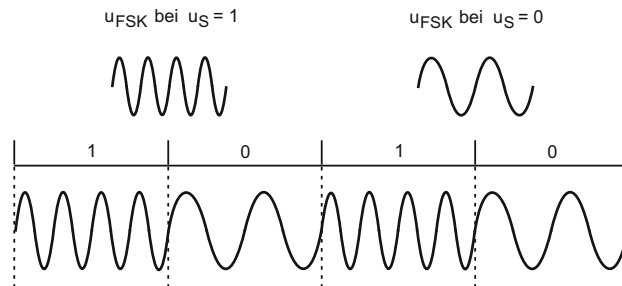


Abbildung 2.48: Abbildung der binären Werte „1“ und „0“ auf die Frequenz eines Trägers

zuständen „0“ und „1“ zwei verschiedene Frequenzen zugeordnet. Liegt dieses Frequenzpaar innerhalb des Sprachfrequenzbandes (0,3 bis 3,4 kHz), können vorhandene Sprachkommunikationseinrichtungen, wie z. B. das internationale Telefonnetz, für den digitalen Datenverkehr genutzt werden. Die ersten Telefonmodems arbeiteten nach diesem Prinzip, und heute wird dieses Verfahren

¹⁴⁴FSK = Frequency Shift Keying (deutsch: Frequenzumtastung)

noch bei einigen Diensten mit der Funkfernseh-technik eingesetzt.

Phasenumtastung

Bei der Phasenumtastung (PSK¹⁴⁵) wird das digitale Informationssignal $u_S(t)$ auf die Phasenlage eines Trägersignals $u_T(t)$ abgebildet. Das Ergebnis zeigt Abbildung 2.49 für eine 2-PSK-Modulation. Die „2“ steht für die Anzahl der Kennzustände. In diesem Fall wird zwischen den zwei Phasenlagen 0° und 180° unterschieden, denen die logischen Werte „0“ und „1“ des Informationssignals $u_S(t)$ zugeordnet sind. Andere Bezeichnungen dafür sind BPSK¹⁴⁶ oder PRK¹⁴⁷.

PSK

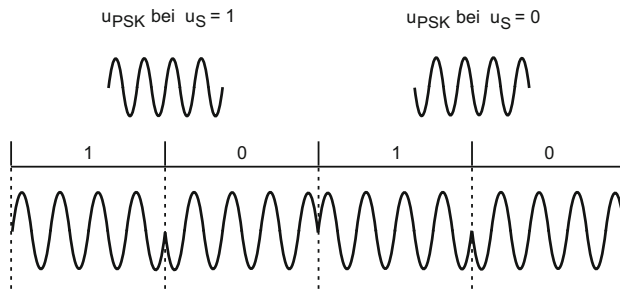


Abbildung 2.49: Abbildung der binären Werte „1“ und „0“ auf die Phase eines Trägers

Senderseitig wird eine 2-PSK-Modulation wie eine ASK-Modulation durch die Multiplikation des digitalen Informationssignals $u_S(t)$ mit dem Trägersignal $u_T(t)$ erzeugt.

Bei der ASK-Modulation werden jedoch die beiden logischen Bit-Werte „1“ und „0“ auf die diskreten Spannungswerte „1“ und „0“ des Informationssignals $u_S(t)$ abgebildet. Das hat zur Folge, dass bei der Multiplikation mit dem Bit-Wert „0“ eine Auslöschung des modulierten Trägersignals $u_{ASK}(t)$ stattfindet.

Bei der PSK-Modulation dagegen werden die logischen Bit-Werte „1“ und „0“ auf die diskreten Spannungswerte „+1“ und „-1“ des Informationssignals abgebildet. Damit bleibt bei der Multiplikation mit dem Bit-Wert „1“ der Träger unverändert, während er beim Bit-Wert „0“ invertiert wird, was einer Phasendrehung des Trägersignals $u_{PSK}(t)$ um 180° entspricht.

Die Demodulation eines PSK-Signals beim Empfänger ist kritisch, denn es wird ja keine Referenzphase mitgesendet, mit der entschieden werden kann, ob gerade eine 0° -Phase (= logisch „1“) oder eine 180° -Phase (= logisch „0“) empfangen wurde. Dieses Problem kann durch einen geeigneten Leitungscode gelöst wer-

¹⁴⁵PSK = Phase Shift Keying (deutsch: Phasenumtastung)

¹⁴⁶BPSK = Binary Phase Shift Keying

¹⁴⁷PRK = Phase Reversal Keying

den. Beim Differentialcode wird z. B. einer log. „1“ ein Phasenwechsel während eines Schrittes zugeordnet und einer log. „0“ kein Phasenwechsel während eines Schrittes (siehe Abbildung 2.50).

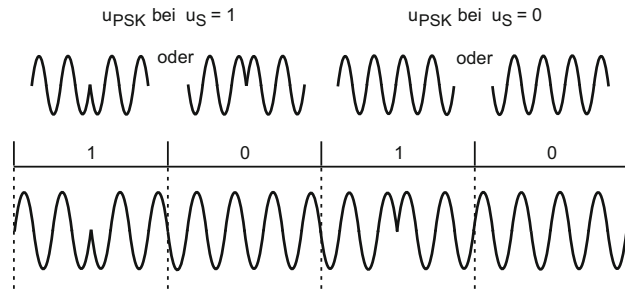


Abbildung 2.50: Differentielle PSK-Modulation

Ein anderes Problem bei der PSK-Modulation sind die steilen Flanken beim Wechsel der Phasenlagen (siehe Abbildung 2.51).

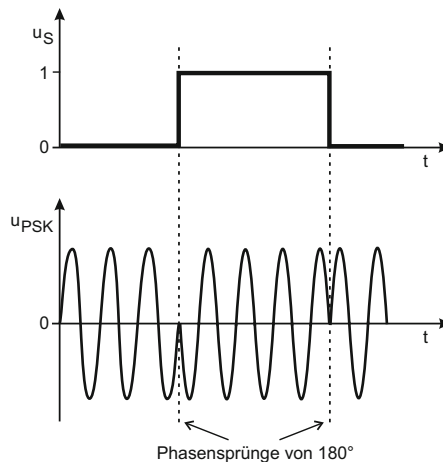


Abbildung 2.51: Phasensprünge bei der PSK-Modulation

Diese Flanken erzeugen Signale weit außerhalb des Übertragungsfrequenzbandes des Kanals und können dort (wie in den Abschnitten 2.2.2 und 2.2.3 bereits erwähnt) andere Kommunikationsverbindungen stören. Abhilfe schafft eine Filterung des Basisbandsignals und dessen Synchronisation mit dem Trägersignal. Dies bezeichnet man als MSK¹⁴⁸.

¹⁴⁸MSK = Minimum Shift Keying

Beim GMSK¹⁴⁹-Verfahren werden durch geeignete Tiefpassfilter die Flanken bei den ursprünglich rechteckförmigen „0-1“- und „1-0“-Übergängen des digitalen Informationssignals $u_S(t)$ verschliffen. Damit erfolgen auch die Übergänge beim PSK-Signal $u_{PSK}(t)$ nicht mehr abrupt, sondern kontinuierlich.

Eine andere Lösung ist die Absenkung der Trägeramplitude vor einem Phasenwechsel und die anschließende Wiederanhebung. In Abbildung 2.52 oben ist dies dargestellt. In der zeitgedehnten Darstellung in Abbildung 2.52 unten erkennt man den Phasenwechsel beim Nulldurchgang der Hüllkurve. Dieses Verfahren kommt vorwiegend bei der differentiellen PSK-Modulation zum Einsatz.

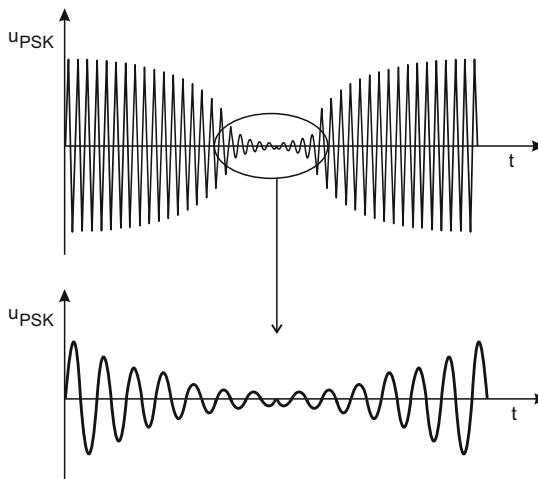


Abbildung 2.52: Dämpfung der Amplitude beim Phasensprung

In einem Schritt kann auch mehr als ein Bit übertragen werden. Abbildung 2.53 zeigt die möglichen Phasenlagen einer 2-PSK, einer 4-PSK und einer 8-PSK-Modulation. Pro Schritt werden hier ein, zwei oder drei Bits übertragen (siehe dazu Gleichung (2.27)).

Quadratur-Amplitudenmodulation

Eine andere Möglichkeit, mehr Bits pro Schritt zu übertragen, ist die Kombination von Phasen- und Amplitudenmodulation. Ein QAM-16-Signal ist z. B. aus 16 Kennzuständen aufgebaut. Damit werden pro Schritt vier Bits übertragen. Im oberen Teil der Abbildung 2.54 sieht man in der Phasenebene 16 diskrete Punkte, die jeweils einem 4-Bit-Wert zugeordnet sind. Jeder Zeiger auf einen der 4-Bit-Werte wird durch seinen Winkel zur 0° -Achse und durch seine Länge bestimmt.

¹⁴⁹GMSK = Gaussian Minimum Shift Keying

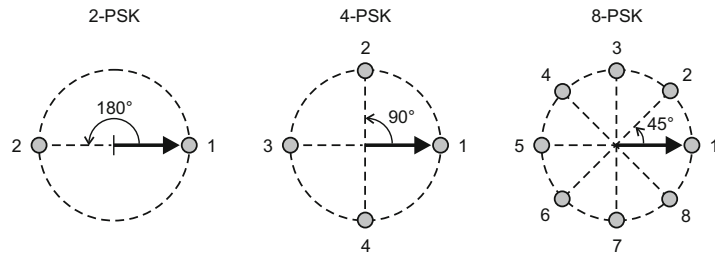


Abbildung 2.53: Mehrbit-PSK mit 1- 2- oder 3-Bit pro Schritt

Man erkennt ferner, dass hier drei unterschiedliche Amplitudenbeträge und 12 verschiedene Phasenlagen möglich sind. Von den damit möglichen 36 verschiedenen Kombinationen (= möglichen Zeigern) werden jedoch nur 16 genutzt. Dies dient der Störsicherheit bzw. der Fehlertoleranz. Um jeden gültigen Endpunkt eines Zeigers (also einer gültigen Kombination aus Phasenlage und Amplitudenbetrag) wird eine Umgebung definiert. Wenn durch einen Übertragungsfehler Phase und/oder Amplitudenbetrag vom Sollwert abweichen, sie aber dennoch innerhalb der definierten Umgebung um einen gültigen Endpunkt liegen, werden sie diesem zugeordnet. In Abbildung 2.54 ist dies am Zeiger „4“ für den Wert „1010“ aufgezeigt.

In Abbildung 2.55 sind noch einmal die grundlegenden Merkmale digitaler Modulationsarten als Übersicht zusammengefasst und in der Zeit-, Frequenz- und Phasenebene dargestellt.

2.3.3 Sender- und Empfängertechnik

Einige Komponenten (z. B. Frequenzaufbereitung, Antennenanpassung) werden sowohl bei der Sende-, als auch bei der Empfangsanlage benötigt. Sie sind deshalb in einer kombinierten Sende-/Empfangsanlage auch nur einmal vorhanden. Für solche kombinierten Anlagen hat sich die Bezeichnung Transceiver¹⁵⁰, eine Wortkombination aus den englischen Bezeichnungen „Transmitter“ und „Receiver“, eingebürgert.

2.3.3.1 Senderaufbau

Die Aufgabe eines Funksenders ist die drahtlose Übermittlung elektrischer Signale. Die ausgesendeten elektrischen Signale können in ihrer Amplitude, ihrer

¹⁵⁰Transceiver = **T**ransmitter**R**eciever

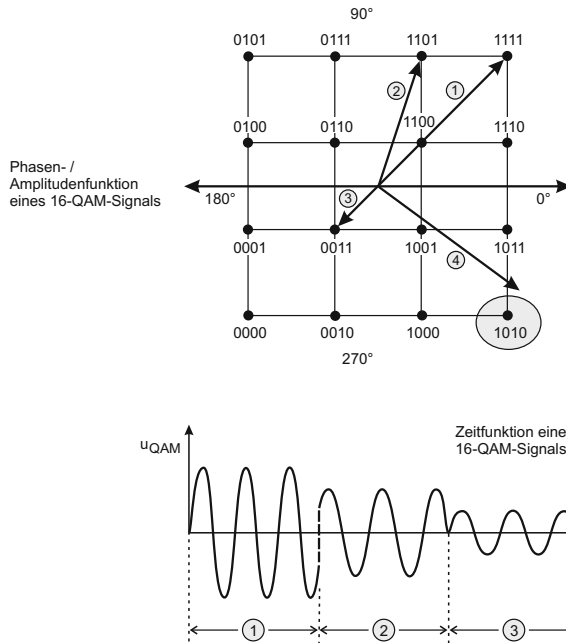


Abbildung 2.54: 16-QAM-Verfahren mit ASK- und PSK-Modulation

Frequenz und/oder ihrer Phase moduliert werden, um eine Information zu übertragen.

Zur Durchführung einer Kommunikation benötigt der Funksender einen Funkempfänger, der die ausgesendeten Funksignale empfängt und demoduliert.

Abbildung 2.56 zeigt das Blockschaltbild eines Hochfrequenzsenders.

Frequenzaufbereitung

Das „Herz“ eines Senders ist dessen Oszillator. Er erzeugt eine sehr präzise Wechselspannung konstanter Frequenz und konstanter Amplitude. Als Frequenzreferenz kommt ein Schwingquarz zum Einsatz, wobei häufig Oszillatorschaltung und Schwingquarz als ein Bauteil in einem Gehäuse angeordnet sind. Die moderne Form der Frequenzerzeugung erfolgt durch die direkte digitale Synthese (DDS¹⁵¹). Damit lassen sich beliebige Frequenzen sehr genau und rauscharm erzeugen. Moderne DDS-Schaltkreise besitzen ein Interface zur Kopplung mit Mikrocontrollern.

Bei einer direkten digitalen Synthese wird aus einem Takt ein digitaler Da-

Oszillator

DDS

**Mikro-
controller**

¹⁵¹DDS = Direct Digital Synthesis

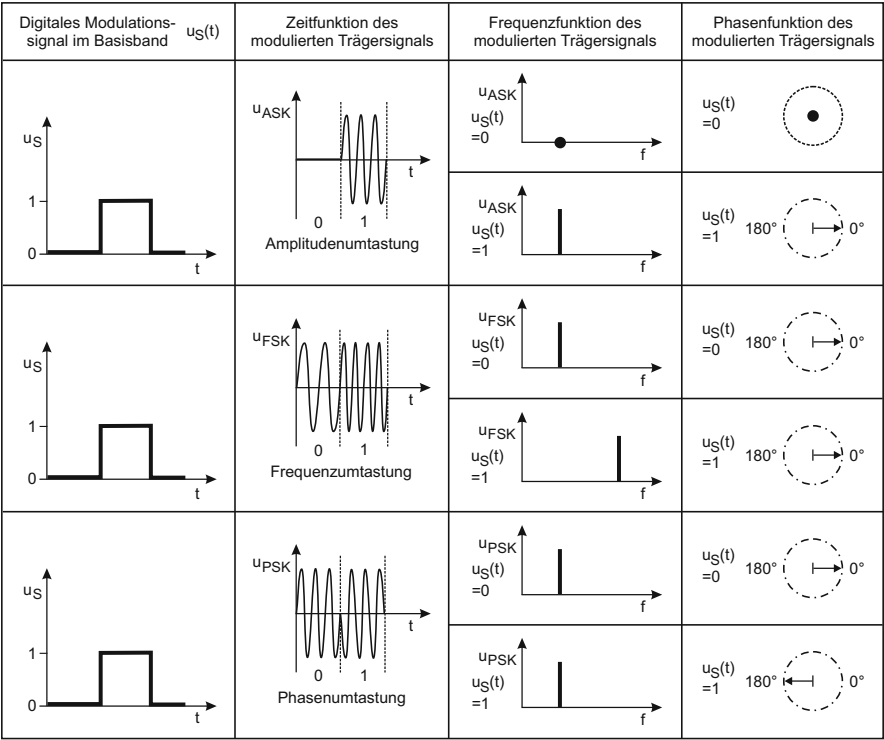


Abbildung 2.55: Modulation eines Trägers mit einem binären Basisbandsignal, Darstellung von AM (ASK), FM (FSK) und PM (PSK)

Mischung

tenwert erzeugt, der einen Speicher adressiert. In diesem Speicher¹⁵² sind die Amplitudenwerte eines Sinussignals abgelegt. Ein nachfolgender Digital-Analog-Wandler mit anschließender Filterung erzeugt daraus ein präzises Sinussignal. Der vom Mikrocontroller gelieferte Soll-Wert bestimmt die Frequenz des Ausgangssignals. Durch Mischung mit anderen Oszillatorfrequenzen wird die gewünschte Sendefrequenz erzeugt.

Die theoretische maximale Ausgangsfrequenz eines DDS-Generators beträgt 1/2 der Taktfrequenz. DDS-Schaltkreise mit integriertem D/A-Wandler sind heute mit Taktfrequenzen bis ca. 1000 MHz verfügbar.

Pufferstufe

Eine nachgeschaltete Pufferstufe dient dazu, Rückwirkungen von der dann folgenden Modulationsstufe auf die Frequenzaufbereitungsstufe auszuschließen. Der gesamte Vorgang der Frequenzaufbereitung kann rein digital in integrierten Schaltungen erfolgen. Dies schließt auch die Funktion des nachfolgenden Mo-

¹⁵²engl.: Look-Up-Table

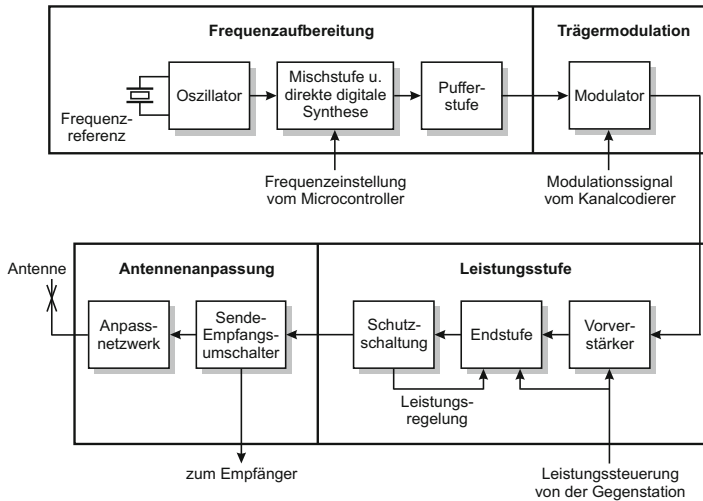


Abbildung 2.56: Vereinfachtes Blockschaltbild eines Hochfrequenzsenders

dulators mit ein. Da die Funktion der digitalen Bausteine programmierbar ist, konnen damit wesentliche Komponenten eines Hochfrequenzsenders per Software konfiguriert werden.

Tragermodulation

Im Modulator werden die Daten in geeigneter Form dem Tragersignal aufgepragt. Diesen Vorgang bezeichnet man als Modulation. Naheres dazu ist im Abschnitt 2.3.2 beschrieben.

Leistungsstufe

Aufgabe der Leistungsstufe ist es, das modulierte Tragersignal vom Oszillator zuverstarken. Das verstarkte Ausgangssignal wird in die Antennenanlage eingespeist und von dort als Hochfrequenzenergie in Form von elektromagnetischen Feldern abgestrahlt. Voraussetzung dafur ist eine funktionsfahige Antenne. Ist die Antenne nicht angeschlossen oder durch mechanische Veranderungen nicht korrekt auf die Sendefrequenz abgestimmt, kommt es zu einer Fehlanpassung. Fehlanpassung heit, dass die Sendenergie nicht mehr vollstandig abgestrahlt werden kann und daher in der Endstufe zur Erwarmung fuhrt. Das kann im Extremfall zur Zerstorung von Leistungskomponenten (Endstufentransistoren) fuhren. Deshalb ist eine Leistungsregelung vorgesehen. Diese basiert auf einer thermischen Schutzschaltung und/oder einer Stehwellenmesseinrichtung. Bei einer Stehwellenmessung wird die Leistung, die zur Antenne geschickt wird, und die Leistung, die von dort reflektiert wird, gemessen. Wird mehr als ein

**Fehl-
anpassung**

bestimmter prozentualer Teil von der Antenne reflektiert, so wird die Ausgangsleistung der Endstufe herunter geregelt.

Unabhängig von der Leistungsregelung wird in vielen drahtlosen Systemen auch eine Leistungssteuerung eingesetzt. Damit wird sichergestellt, dass nur die Sendeleistung erzeugt und abgestrahlt wird, die für eine sichere Kommunikation nötig ist. Die Steuerung erhält dabei ihre Sollwerte von der Gegenstation. Dort wird die Empfangsfeldstärke gemessen und dem Sender übermittelt. Das geschieht natürlich stets mit einer gewissen Zeitverzögerung, denn dieser Messwert kann erst dann übertragen werden, wenn die Datenflussrichtung wechselt. Drei wesentliche Gründe sind für die Implementierung einer Leistungssteuerung ausschlaggebend:

Gründe

- a) Es wird Energie eingespart. Da drahtlose Kommunikationssysteme oft mobil eingesetzt werden, ist die maximale Betriebsdauer bei einer vorgegebenen Batteriekapazität ein wichtiges Einsatzkriterium.
- b) Wenn man mit der geringst möglichen Sendeleistung arbeitet, wird die Ausstrahlung entsprechend räumlich begrenzt. Damit können in benachbarten Funkzellen die durch die eigene Kommunikation genutzten Frequenzen bereits neu belegt werden, ohne dass Störungen erfolgen.
- c) Es gibt Übertragungsverfahren, bei denen eine sehr exakte Leistungssteuerung Voraussetzung für einen störungsfreien Betrieb ist (siehe dazu CDMA, Abschnitt 2.3.4.3).

Antennenanpassung

In der Regel wird bei der drahtlosen Nahbereichs-Kommunikation über die gleiche Antenne gesendet und empfangen. Es muss daher eine Antennenweiche oder ein Antennenumschalter vorgesehen sein, der je nach momentaner Datenflussrichtung den Sende- oder den Empfangsteil des Transceivers mit der Antenne verbindet.

Aufgaben

Die eigentliche Antennenanpassung hat zwei Aufgaben:

- a) Im Resonanzfall stellt die Antenne einen realen (ohmschen) Widerstand dar. Wenn man die Verluste vernachlässigt, wird die in diesen Widerstand eingespeiste Leistung als Hochfrequenzenergie abgestrahlt. Deshalb wird dieser (virtuelle) Widerstand auch als Strahlungswiderstand bezeichnet. Um eine maximale Leistung in diesen Widerstand einzuspeisen, muss eine Leistungsanpassung vorliegen, d. h. der Innenwiderstand R_i des Generators (= der Endstufe) muss gleich dem Lastwiderstand R_L (= Strahlungswiderstand der Antenne) sein. Da das in der Regel nicht der Fall ist, muss hier eine Widerstandstransformation (= Widerstandsanpassung) durchgeführt werden.
- b) Eine Antenne wird selten genau mit ihrer Resonanzfrequenz betrieben. Oberhalb oder unterhalb der Resonanzfrequenz enthält der Ersatzwiderstand induktive oder kapazitive Blindkomponenten. Durch Reihen- oder Parallelschal-

tung von Spulen und Kondensatoren sind diese Blindkomponenten kompensierbar, und die von der Antenne reflektierte Leistung wird minimiert. Die Aufgaben a) und b) werden zusammengefasst und durch ein Anpassnetzwerk aus mehreren induktiven und kapazitiven Komponenten gelöst.

2.3.3.2 Empfangsprinzipien

Die Aufgabe eines Empfängers ist es, die von der Antenne gelieferten hochfrequenten Signale zu selektieren, zu verstärken und in das Basisband zu transformieren.

In Abbildung 2.57 sind diese Aufgaben anhand eines Geradeausempfängers verdeutlicht. Von der Antenne kommt ein breitbandiges Hochfrequenzsignal, welches sich aus den Signalen aller am Empfängerstandort empfangbaren Sender zusammensetzt. Durch einen Vorkreis wird nur der Frequenzbereich durchgelassen, der empfangen werden soll. Im nachfolgenden Hochfrequenzverstärker wird das Signal auf einen Pegel angehoben, der eine Demodulation (z. B. bei AM die Umsetzung eines Seitenbandes in das Basisband) ermöglicht. Das Basisbandsignal wird dann verstärkt und im Beispiel von Abbildung 2.57 auf einem Lautsprecher ausgegeben.

**Geradeaus-
empfänger**

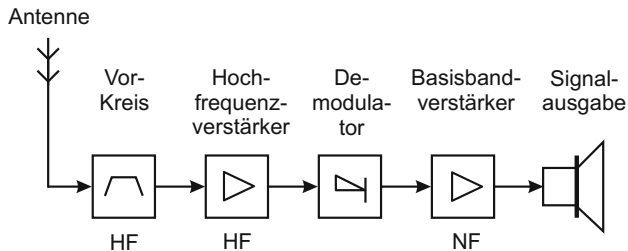


Abbildung 2.57: Prinzipielles Blockschaftbild eines Empfängers

Mit dem in Abbildung 2.57 dargestellten Prinzip eines Geradeausempfängers wurden die ersten Rundfunkempfänger in den 20er Jahren des vorigen Jahrhunderts gebaut. Heute kommt es nur noch bei Spezialempfängern zum Einsatz, denn es hat gravierende Nachteile.

**Rundfunk-
empfänger**

Um die erforderliche Selektivität zu erzielen, d. h. a) die Sperrung (Dämpfung) aller nicht gewünschten Frequenzbereiche und b) das Durchlassen des gewünschten Frequenzbereiches, müssen mehrere Filter hintereinander geschaltet werden. Bei einem Frequenzwechsel müssten dann alle diese Filter auf den neuen Frequenzbereich abgestimmt werden, was kaum praktikabel ist.

Ein weiterer Nachteil liegt darin, dass beim Geradesausempfänger der Empfangsfrequenzbereich direkt verstärkt werden muss. Je höher das Frequenzband ist, welches verstärkt werden soll, desto komplexer und teurer werden die Verstärkerkomponenten. Bei den Frequenzbereichen, die in der Nahbereichskommunikation per Funk benutzt werden, wäre das unökonomisch.

Beispiel: Funkuhrtechnik – Die Senderfrequenzen liegen unterhalb von 100 kHz, also in einem Frequenzband, welches leicht beherrschbar ist. Funkuhren empfangen in Mitteleuropa in der Regel nur den Zeitzeichensender DCF77 und damit ist kein Frequenzwechsel erforderlich (wobei es aber durchaus noch andere Zeitzeichenstationen gibt, z. B. MSF in Anthorn, Großbritannien, WWVB in Fort Collins, USA, und JJY40 in Tokio, Japan).

Die Nachteile eines Geradeausempfängers umgeht man mit einem Superheterodynempfänger (oder Superhet oder einfach nur Super genannt). Dabei wird das empfangene Sendersignal zunächst auf eine niedrige Zwischenfrequenz (ZF, IF¹⁵³) umgesetzt, auf der es dann leicht weiterverarbeitet werden kann.

Mischung

Diesen Vorgang der Umsetzung bezeichnet man als (Frequenz-) Mischung. Das von der Antenne kommende und gefilterte Frequenzband wird in der Mischstufe mit einer Oszillatorfrequenz multipliziert. Es entstehen dabei zwei Mischfrequenzbereiche, einmal die Summe ($f_o + f_e$) aus Empfangsfrequenz f_e und Oszillatorfrequenz f_o und einmal die Differenz ($f_o - f_e$). In Abbildung 2.58 ist dies dargestellt. Der Vorgang der Mischung entspricht dem einer Doppelseitenbandmodulation, wie sie im Abschnitt 2.3.2.2 beschrieben wurde.

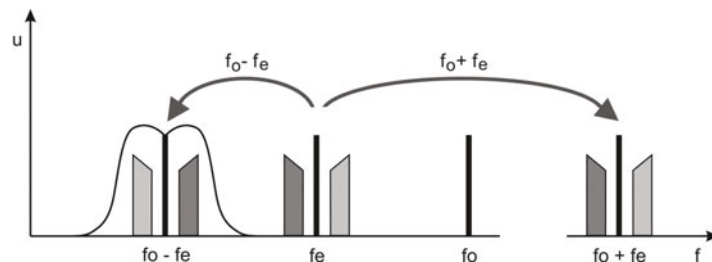


Abbildung 2.58: Prinzip einer Frequenzmischung

In Abbildung 2.59 wird das Prinzip eines Empfängers nach dem Superheterodynprinzip gezeigt. Die von der Antenne empfangenen HF-Signale werden nach

¹⁵³IF = Intermediate Frequency

**Super-
heterodyn-
prinzip**

einem Vorfilter verstärkt und zur Mischstufe¹⁵⁴ geleitet. Dort erfolgt die Multiplikation mit dem Ausgangssignal eines in der Frequenz einstellbaren Oszillators (f_O). Das entstehende Frequenzdifferenzsignal ($f_O - f_e$) fällt genau in den Durchlassbereich des Zwischenverstärkers. Das ebenfalls entstehende Frequenzsummensignal ($f_O + f_e$) wird nicht weiter verstärkt, da es hier nicht benötigt wird.

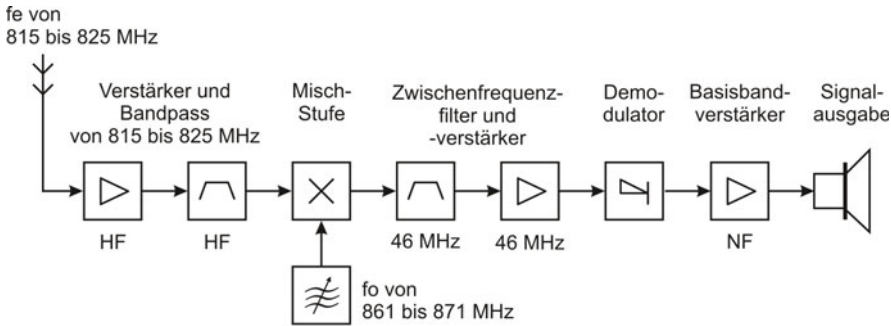


Abbildung 2.59: Blockschaltbild eines Superheterodynempfängers für das 800 MHz-Band

Die Eigenschaften eines Empfängers werden ganz wesentlich von dem Zwischenfrequenzverstärker bestimmt, denn die eigentliche Verstärkung des empfangenen Signals und die Bandbreitenbegrenzung (Selektion) erfolgen dort. Die Bandbreite des ZF-Verstärkers sollte dabei das Frequenzband des modulierten Trägersignals nicht oder nur unwesentlich überschreiten, denn sonst werden Störungssignale, die nicht zum Nutzsignal gehören (oder Rauschen) verstärkt. Die Bandbreite sollte aber auch keinesfalls kleiner als das Frequenzband des modulierten Trägersignals sein, denn sonst gingen Informationsanteile des Nutzsignals verloren.

Die Abstimmung auf die Empfangsfrequenz erfolgt bei diesem Empfangskonzept nur durch Veränderung der Oszillatorfrequenz.

Beispiel: Um ein Eingangssignal der Frequenz $f_e = 820$ MHz auf die Mittenfrequenz des ZF-Verstärkers $f_{ZF} = 46$ MHz umzusetzen, muss die Oszillatorfrequenz f_O auf den Wert $f_O = f_e + f_{ZF}$ ($= 866$ MHz) abgestimmt werden. Soll der Empfänger auf eine Empfangsfrequenz von $f_e = 821$ MHz abgestimmt werden, so muss der Oszillator auf den Wert $f_O = 867$ MHz eingestellt werden.

¹⁵⁴engl.: Mixer

Hinweis

Natürlich wäre auch eine Oszillatorfrequenz unterhalb der Empfangsfrequenz möglich. Um dann ein Eingangssignal der Frequenz $f_e = 820$ MHz auf die Mittenfrequenz des ZF-Verstärkers $f_{ZF} = 46$ MHz umzusetzen, muss die Oszillatorfrequenz f_O auf den Wert $f_O = f_e - f_{ZF}$ ($= 774$ MHz) abgestimmt werden. Soll der Empfänger auf eine Empfangsfrequenz von $f_e = 821$ MHz abgestimmt werden, so muss der Oszillator auf den Wert $f_O = 775$ MHz eingestellt werden. In beiden Fällen erfolgt eine Frequenzinversion, wie sie auch im Abschnitt 2.3.2.2, unteres Seitenband, erläutert wurde.

Geht man dagegen von einer eingestellten Oszillatorfrequenz von z. B. $f_O = 866$ MHz (siehe Beispiel oben) aus, so gibt es zwei mögliche Empfangsfrequenzen, die, gemischt mit dieser Oszillatorfrequenz, ein Zwischenfrequenzsignal von 46 MHz ergibt: a) $f_O - f_{ZF} = 820$ MHz und b) $f_O + f_{ZF} = 912$ MHz. Diese zweite mögliche (aber unerwünschte) Empfangsfrequenz wird als Spiegelfrequenz¹⁵⁵ bezeichnet (siehe Abbildung 2.60).

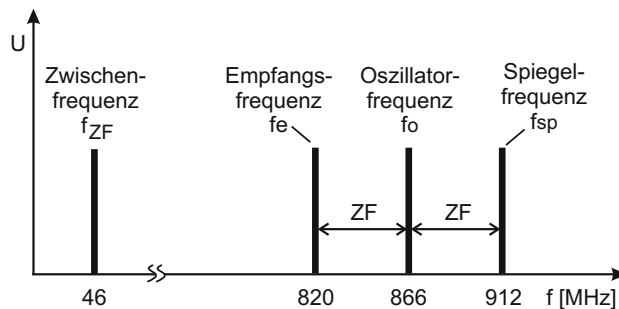


Abbildung 2.60: Frequenzschema beim Spiegelfrequenzempfang

Ohne weitere Vorkehrungen wird eine Sendestation auf der Spiegelfrequenz den Empfang des Trägersignals mit den Nutzdaten stören oder sogar verhindern können. Es muss deshalb mit einer Vorselektion verhindert werden, dass ein Signal, welches auf Spiegelfrequenz gesendet wird, auf den Eingang der Mischstufe gelangt. Im einfachsten Fall kann dies durch einen Tiefpass (oder einen Bandpass) geschehen, was aus Kostengründen besonders dann gerne gemacht wird, wenn im Spiegelempfangsbereich nur selten Sendestationen zu erwarten sind.

Ist jedoch der Empfangsbereich, den der Empfänger abdecken soll, größer als die doppelte Spiegelfrequenz ($f_{emax} - f_{emin} > 2 \cdot f_{ZF}$), muss eine mitlaufende Vorselektion implementiert werden. Mitlaufend heißt hier, dass außer der Oszillatorfrequenz f_O auch mindestens ein Vorkreis synchron abgestimmt werden muss. Der Abstand zwischen der Oszillatorfrequenz und der Mittenfrequenz der

¹⁵⁵engl.: Image Frequency

Vorkreise ist gleich der Zwischenfrequenz.

Der Demodulator ist auf die Modulationsart des Trägers zugeschnitten. Er trennt das Basisbandsignal vom (modulierten) Träger und leitet es einem Verstärker zu. In Abbildung 2.59 ist für die Ausgabe des analogen Signals ein Lautsprecher vorgesehen. Dieser elektroakustische Wandler hat hier die Funktion des Quelldecodierers.

Ein Empfängerkonzept, welches erst in den letzten Jahren mit vertretbarem Aufwand realisiert werden kann, ist das Prinzip der Direktmischung¹⁵⁶ (siehe Abbildung 2.61).

**Direkt-
mischung**

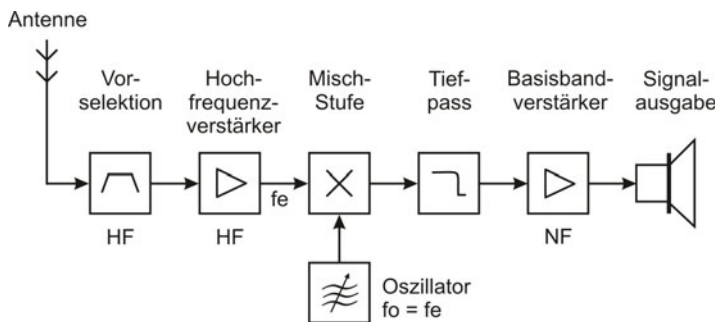


Abbildung 2.61: Empfängerschaltung mit Direktmischung

Dieses Empfängerprinzip erfordert sehr präzise und frequenzstabile Oszillatoren, die in der erforderlichen Qualität nicht zu einem akzeptablen Preis zur Verfügung standen. Erst durch die DDS¹⁵⁷-Technik (siehe Abschnitt 2.3.3.1, Frequenzaufbereitung) ist es praktikabel, Empfänger nach diesem Prinzip zu bauen.

DDS

Wie in Abschnitt 2.3.2.2, Einseitenbandmodulation, gezeigt wurde, wird das Basisbandsignal beim Vorgang der Modulation in einen anderen Frequenzbereich transferiert. Dieser Vorgang wird bei einem Direktmischer rückgängig gemacht, indem das Empfangssignal f_e mit einem Oszillatorsignal gleicher Frequenz gemischt wird. „Gleiche Frequenz“ heißt hier, dass der Oszillator genau auf die Frequenz des Trägers (oder bei SSB-Modulation des unterdrückten Trägers) abgestimmt wird. Die entstehenden Mischprodukte bestehen aus der Summe der Frequenzen ($f_e + f_o = f_e + f_e$), also doppelte Trägerfrequenz und der Differenz der Frequenzen, d. h. ($f_e - f_o = f_e - f_e$), 0 Hz bzw. eine Gleichspannung.

¹⁵⁶engl.: Direct Conversion

¹⁵⁷DDS = Direct Digital Synthesis

Aufgabe: Gegeben ist ein SSB-Signal im oberen Seitenband mit einem unterdrückten Träger bei $f_T = 10$ MHz (Basisband bei 0,3 bis 3,4 kHz).

a) Welchen Frequenzbereich belegt dieses Einseitenbandsignal im HF-Bereich?

b) Was passiert, wenn das HF-Signal aus a) mit einem Oszillator der Frequenz $f_O = 10$ MHz ($= f_T$) gemischt wird?

c) Was passiert, wenn das HF-Signal aus a) mit einem Oszillator gemischt wird, dessen Frequenz geringfügig von f_T abweicht? Welche prozentuale Frequenzabweichung ergibt sich für ein Basisbandsignal von $f_S = 1$ kHz, wenn es mit der Oszillatorfrequenz $f_O' = f_T + 0,001\%$ demoduliert wird?

Lösungen

Zu a) Es belegt den Frequenzbereich von 10,0003 bis 10,0034 MHz.

Zu b) Es wird einmal das Mischprodukt $f_e + f_O$ erzeugt, also eine Umsetzung von 10,0003 bis 10,0034 MHz nach 20,0003 bis 20,0034 MHz. Dieses Mischprodukt wird nicht weiter verarbeitet bzw. durch Filterung unterdrückt. Das andere Mischprodukt $f_e - f_O$ hat die gewünschte Umsetzung des HF-Bandes von 10,0003 bis 10,0034 MHz in das Basisband von 0,3 bis 3,4 kHz zum Ergebnis. Diese Umsetzung bezeichnet man als Demodulation.

Zu c) Der Oszillator mit $f_O' = f_T + 0,001\%$ weicht von der Frequenz des unterdrückten Trägers $f_T = 10$ MHz um 100 Hz nach oben ab. Damit erfolgt die Umsetzung des HF-Bandes von 10,0003 bis 10,0034 MHz in das Basisband von 0,2 bis 3,3 kHz. Ein senderseitiges Basisbandsignal von 1000 Hz wird im Empfänger als 900 Hz-Signal demoduliert. Diese Abweichung beträgt - 10%.

Aufgabe: Gegeben ist ein SSB-Signal im oberen Seitenband mit einem unterdrückten Träger bei $f_T = 10$ MHz (Basisband bei 0,3 bis 3,4 kHz).

a) Was passiert, wenn ein Signal im Spiegelfrequenzbereich des Empfängers gesendet wird? Zeigen Sie die Umsetzung eines Empfangssignals von 9,999 MHz in den Basisbandbereich, wenn $f_O = 10$ MHz ($= f_T$) beträgt.

b) Was passiert, wenn ein SSB-Signal im oberen Seitenband (mit einem unterdrückten Träger bei $f_T = 10$ MHz, Basisband bei 0,3 bis 3,4 kHz) mit einem Oszillator der Frequenz $f_O = 10,0037$ MHz gemischt wird?

Lösungen

Zu a) Bei einem Oszillatorsignal von $f_O = 10$ MHz wird das Empfangssignal $f_e = 9,999$ MHz im Basisband als 1 kHz demoduliert. Eine Ausblendung des Spiegelfrequenzbereiches von 9,9966 bis 9,9997 MHz ist wegen der erforderlichen Flankensteilheit der benötigten Filter nicht oder nur mit sehr hohem Aufwand möglich.

Es tritt eine Frequenzinversion bei der Demodulation auf, d. h. eine hohe Frequenz im ausgesendeten Basisbandsignal, z. B. von 3 kHz, wird als tiefe Frequenz von 700 Hz demoduliert und umgekehrt (siehe Tabelle 2.20). **Zu b)**

	Basisband (Quelle) [Hz]	HF- Bereich [MHz]	Oszillator- frequenz [MHz]	Basisband (Senke) [Hz]
$f_{S-unten}$	300	10,0003	10,0037	3400
f_{S1}	1000	10,001	10,0037	2700
f_{S2}	1850	10,00185	10,0037	1850
f_{S3}	3000	10,003	10,0037	700
f_{S-oben}	3400	10,0034	10,0037	300

Tabelle 2.20: Aufgabe SSB-Signal

Wie man beim Lösen der Aufgabe a) sieht, wird ein durchstimmbarer Empfänger jedes Signal zweimal empfangen: einmal direkt und einmal als Spiegelfrequenz. Die Unterdrückung des Spiegelfrequenzempfangs ist wegen des geringen Frequenzabstandes zwischen dem originalen Signal und dem Spiegelfrequenzsignal durch eine Vorfilterung des Antennensignals kaum möglich, denn diese müssten extrem steilflankig sein.

Neuere Techniken nutzen bei einem Empfänger mit Direktmischung ein anderes Demodulationsprinzip, wie es in Abbildung 2.62 dargestellt ist.

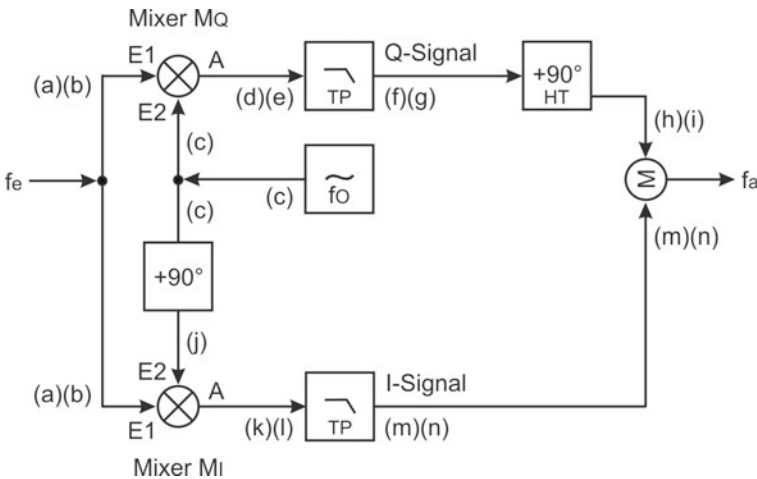


Abbildung 2.62: Quadraturdemodulatorschaltung

Um die unerwünschten Spiegelfrequenzen zu unterdrücken, wird das Empfangssignal in zwei getrennten Mischstufen mit der lokalen Oszillatorfrequenz f_O gemischt. Die eine Mischung erfolgt mit dem Signal $U_{OI}(t)$ und die andere mit dem Signal $U_{OQ}(t)$. $U_{OI}(t)$ und $U_{OQ}(t)$ sind zueinander um 90° phasenverschoben. Der Oszillator liefert das Signal

$$U_{OQ}(t) = \hat{u} \cdot \sin(\omega_O t) \text{ (Q-Signal}^{158}\text{),}$$

mit dem der obere Mischer angesteuert wird. Der untere Mischer wird mit dem um 90° phasenverschobenen Ausgangssignal des Oszillators angesteuert¹⁵⁹:

$$U_{OI}(t) = \hat{u} \cdot \cos(\omega_O t) \text{ (I-Signal}^{160}\text{).}$$

Aufgabe: Am Eingang der Schaltung aus Abbildung 2.62 liegen die Signale $U_{e1}(t) = \hat{u} \cdot \cos(\omega_0 t + \omega_1 t)$ im oberen Seitenband und $U_{e2}(t) = \hat{u} \cdot \cos(\omega_0 t + \omega_2 t)$ im unteren Seitenband an. Damit steht am Eingang der Schaltung ein Signal an, welches sich aus den zwei Frequenzen zusammensetzt: $f_{e1} = (\omega_0 - \omega_1)$ und $f_{e2} = (\omega_0 - \omega_2)$.

Die Generatorfrequenz sei gleich $f_O = \omega_0$.

Zeigen Sie, dass es mit dieser Schaltung möglich ist,

a) nur Signale mit $f_e < \omega_0$ bzw.

b) nur Signale mit $f_e > \omega_0$

zu demodulieren, d. h. ins Basisband umzusetzen.

Lösungen

Am Eingang E1 der Mischstufe M_Q liegen an:

(a): $\hat{u} \cdot \cos(\omega_0 t + \omega_1 t)$

und

(b): $\hat{u} \cdot \cos(\omega_0 t + \omega_2 t),$

sowie am Eingang E2 von M_Q das Oszillatorsignal

(c): $U_{OQ}(t) = \hat{u} \cdot \sin(\omega_O t).$

Zur Vereinfachung sei hier angenommen, dass der Amplitudenwert $\hat{u} = 1\text{V}$ betrage.

Am Ausgang A der Mischstufe M_Q erhält man

(d): (a) · (c)
 $\hat{u} \cdot \cos(\omega_0 t + \omega_1 t) \cdot \sin(\omega_O t)$

Gemäß $\cos\alpha \cdot \sin\beta = 1/2 \cdot [\sin(\alpha + \beta) - \sin(\alpha - \beta)]$ erhält man dann:

$$1/2 \cdot [\sin(\omega_0 t + \omega_1 t + \omega_O t) - \sin(\omega_0 t + \omega_1 t - \omega_O t)]$$

Da $f_O = \omega_0$ erhält man:

¹⁵⁸Q-Signal = **Quadrature-Signal**

¹⁵⁹In der Literatur werden mit dem Begriff „Quadratur-Signal“ häufig beide Komponenten (also das I- und das Q-Signal) bezeichnet.

¹⁶⁰I-Signal = **Inphase-Signal**

$$(d): \quad 1/2 \cdot [\sin(2 \cdot \omega_0 t + \omega_1 t) - \sin(\omega_1 t)].$$

Und entsprechend erhält man mit (b):

$$(e): \quad (b) \cdot (c)$$

$$\cos(\omega_0 t - \omega_2 t) \cdot \sin(\omega_0 t) = 1/2 \cdot [\sin(2 \cdot \omega_0 t - \omega_2 t) - \sin(-\omega_2 t)]$$

Gemäß der Beziehung $\sin(-\alpha) = -\sin(\alpha)$ erhält man für (e)

$$(e): \quad 1/2 \cdot [\sin(2 \cdot \omega_0 t - \omega_2 t) + \sin(\omega_2 t)]$$

Mit dem Tiefpass werden Signalfrequenzen $\gg \omega_0$ ausgeblendet. Damit erhält man aus (d):

$$(f): \quad -1/2 \cdot \sin(\omega_1 t).$$

Und entsprechend erhält man aus (e):

$$(g): \quad +1/2 \cdot \sin(\omega_2 t).$$

Gemäß der Beziehung $\sin(\alpha + 90^\circ) = \cos(\alpha)$ erhält man durch die nachfolgende Phasenverschiebung um $+90^\circ$ aus (f) und (g):

$$(h): \quad -1/2 \cdot \sin(\omega_1 t + 90^\circ) = -1/2 \cdot \cos(\omega_1 t)$$

und

$$(i): \quad +1/2 \cdot \sin(\omega_2 t + 90^\circ) = +1/2 \cdot \cos(\omega_2 t)$$

Für die untere Mischstufe aus Abbildung 2.62 ergibt sich entsprechend:

Am Eingang E1 der Mischstufe M_I liegen an

$$(a): \quad \hat{u} \cdot \cos(\omega_0 t + \omega_1 t)$$

und

$$(b): \quad \hat{u} \cdot \cos(\omega_0 t - \omega_2 t)$$

sowie am Eingang E2 von M_I das um 90° phasenverschobene Oszillatorsignal

$$(j): \quad U_{OI}(t) = \hat{u} \cdot \cos(\omega_0 t)$$

Zur Vereinfachung sei auch hier angenommen, dass der Amplitudenwert $\hat{u} = 1V$ betrage.

Am Ausgang der Mischstufe erhält man dann

$$(k): \quad (a) \cdot (j)$$

$$\hat{u} \cdot \cos(\omega_0 t + \omega_1 t) \cdot \cos(\omega_0 t)$$

Gemäß $\cos\alpha \cdot \cos\beta = 1/2 \cdot [\cos(\alpha - \beta) + \cos(\alpha + \beta)]$ erhält man dann:

$$1/2 \cdot [\cos(\omega_0 t + \omega_1 t - \omega_0 t) + \cos(\omega_0 t + \omega_1 t + \omega_0 t)]$$

Da $f_O = \omega_0$ erhält man:

$$(k): \quad 1/2 \cdot [\cos(\omega_1 t) + \cos(2 \cdot \omega_0 t + \omega_1 t)]$$

Und entsprechend erhält man mit (b):

$$(i): \quad (b) \cdot (j)$$

$$\cos(\omega_0 t - \omega_2 t) \cdot \cos(\omega_0 t) = 1/2 \cdot [\cos(-\omega_2 t) + \cos(2 \cdot \omega_0 t - \omega_2 t)]$$

Gemäß $\cos(-\alpha) = \cos(\alpha)$ erhält man

$$(i): \quad 1/2 \cdot [\cos(\omega_2 t) + \cos(2 \cdot \omega_0 t - \omega_2 t)].$$

Auch hier werden mit dem Tiefpass die Signalfrequenzen $\gg \omega_0$ ausgeblendet.

Damit erhält man aus (k):

$$(m): \quad +1/2 \cdot \cos(\omega_1 t)$$

Und entsprechend erhält man aus (i):

$$(n): \quad +1/2 \cdot \cos(\omega_2 t).$$

Werden nun die Signale (h), (i), (m) und (n) addiert, so erhält man:

$$-1/2 \cdot \cos(\omega_1 t) + 1/2 \cdot \cos(\omega_2 t) + 1/2 \cdot \cos(\omega_1 t) + 1/2 \cdot \cos(\omega_2 t) = \cos(\omega_2 t)$$

Man sieht, das Signal ω_1 im oberen Seitenband ist ausgeblendet worden, d. h. das Ausgangssignal $U_a(t)$ enthält nur die Frequenzanteile von $f_e < \omega_0$.

Werden die Signale (h) und (i) subtrahiert, erhält man:

$$-[-1/2 \cdot \cos(\omega_1 t) + 1/2 \cdot \cos(\omega_2 t)] + 1/2 \cdot \cos(\omega_1 t) + 1/2 \cdot \cos(\omega_2 t) = \cos(\omega_1 t)$$

In diesem Fall wird das Signal ω_2 im unteren Seitenband ausgeblendet. Das Ausgangssignal $U_a(t)$ enthält nur die Frequenzanteile von $f_e > \omega_0$.

Wenn man nach dem in Abbildung 2.62 dargestellten Prinzip einen durchstimmbaren Empfänger aufbaut, wird hier jedes Signal beim Durchstimmen des Oszillators nur einmal empfangen (im Gegensatz zur Schaltung nach Abbildung 2.61). Durch die Trennung in ein I- und Q-Signal wird der Spiegelfrequenzempfang unterdrückt und je nach Addition oder Subtraktion der I- und Q-Signale erhält man die Frequenzanteile unterhalb oder oberhalb von ω_0 .

Kritisch ist die Realisierung des Phasenschiebers, der im Q-Zweig von Abbildung 2.62 angeordnet ist. Das Berechnungsbeispiel wurde hier nur für die diskreten Frequenzen $f_{e1} = (\omega_0 + \omega_1)$ und $f_{e2} = (\omega_0 - \omega_2)$ durchgeführt. Tatsächlich tritt aber am Eingang der Schaltung ein ganzer Frequenzbereich auf: das im Modulator des Senders in den HF-Bereich umgesetzte Basisbandsignal.

Damit muss hier ein breitbandiger Phasenschieber zum Einsatz kommen, d. h. nicht ein Signal, welches aus einer einzelnen Frequenz besteht, muss phasenverschoben werden, sondern ein ganzer Frequenzbereich. So etwas ist in analoger Technik nicht einfach zu realisieren, denn bei phasenverschiebenden Bauteilen ist die Höhe der Phasenverschiebung eine Funktion der Frequenz.

Eine Lösung stellt die doppelte Umsetzung nach der Methode Weaver dar, die in Abbildung 2.63 dargestellt ist. Die hier zum Einsatz kommenden Phasenschieber müssen nur jeweils eine Frequenz (f_{O1} und f_{O2}) um 90° in der Phase verschieben.

Einfacher ist es, einen breitbandigen 90° -Phasenschieber arithmetisch mit der Hilbert-Transformation zu realisieren. Die Hilbert-Transformation, angewandt auf Sinus- bzw. Cosinus-signale lautet:

$$\begin{aligned} \text{HT}[\cos(\alpha)] &= \sin(\alpha) \\ \text{HT}[\sin(\alpha)] &= -\cos(\alpha) \\ &= +\cos(\alpha) \end{aligned}$$

Hilbert-Transformation

Aufgabe: Zeigen Sie auf, dass auch in der Schaltung nach Abbildung 2.63 die Spiegelfrequenzen ausgeblendet werden!

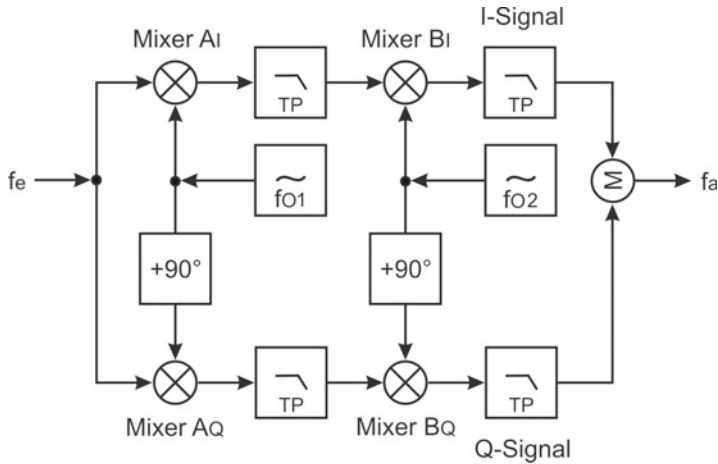


Abbildung 2.63: Demodulatorschaltung nach Weaver

Dazu müssen die I- und Q-Anteile des Empfangssignals digitalisiert werden, um anschließend in einem Mikrocontroller, DSP, FPGA oder sonstigen eingebetteten System auf arithmetischem Wege weiterverarbeitet zu werden. Abbildung 2.64 zeigt diese Variante einer Empfängerschaltung mit Direktmischung. Dieses Empfängerprinzip bezeichnet man als „Software Defined Radio“¹⁶¹, weil ein wesentlicher Teil der Demodulation und Filterung per Software erfolgt. Die hier maximal mögliche Eingangsfrequenz ist stark technologieabhängig. Deshalb wird derzeit überwiegend eine Kombination aus dem Superheterodynprinzip und der Schaltung nach Abbildung 2.64 eingesetzt. Die Zwischenfrequenz des Superheterodynempfängers ist dann die Eingangsfrequenz f_e der obigen Schaltung. Es ist natürlich auch denkbar, dass die Oszillatorfrequenz direkt auf die gewünschte Empfangsfrequenz abgestimmt wird. Das wäre ein wichtiger Schritt zu einem voll digitalen Empfänger. Zurzeit ist diese Technik zumindest für zivile Anwendungen noch unrentabel. Im Kapitel 6 sind die zu erwartenden Trends bei der Transceiverentwicklung auf diesem Gebiet beschrieben.

**Software
Defined
Radio**

Aufgabe: Nennen Sie jeweils drei Vor- und Nachteile der dargestellten Empfangsprinzipien.

¹⁶¹SDR = Software Defined Radio

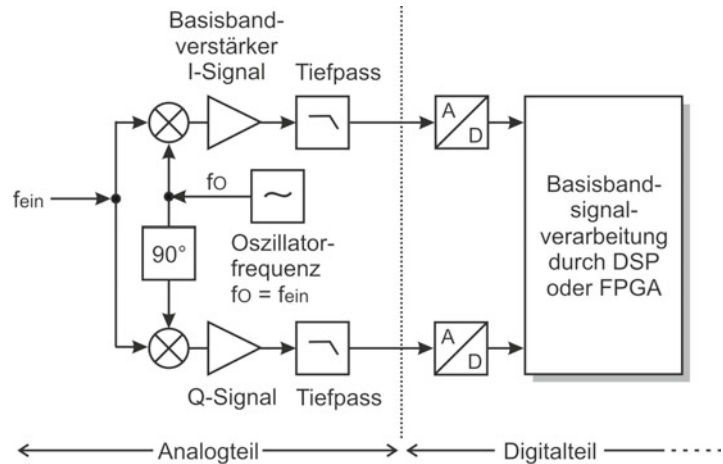


Abbildung 2.64: Demodulatorschaltung für ein Software Defined Radio

2.3.3.3 Antennentechnik

Aufgabe

H-, E-Feldkomponente

Die Aufgabe einer (Empfangs-) Antenne ist es, die Energie eines elektromagnetischen Feldes in eine Wechselspannung oder einen Wechselstrom zu wandeln. Dieser Vorgang ist reversibel. Mit einer (Sende-) Antenne wird elektrische Energie in ein elektromagnetisches Feld gewandelt. Jedes elektromagnetische Feld besteht aus einer magnetischen (H-) und einer elektrischen (E-) Feldkomponente (siehe Abbildung 2.65).

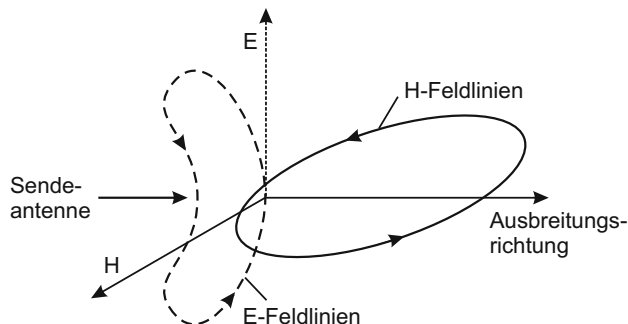


Abbildung 2.65: E- und H-Feld um eine Dipolantenne

Fernfeld

Im Fernfeld stehen diese Komponenten senkrecht aufeinander (siehe Abbildung 2.66). Je nachdem, welche Komponente des Feldes gewandelt werden soll, gibt es Antennen für die magnetische oder für die elektrische Feldkomponente des elektromagnetischen Feldes.

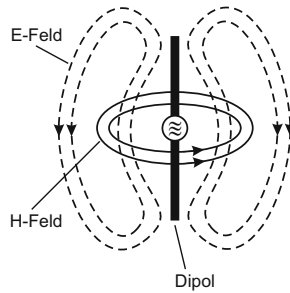


Abbildung 2.66: Elektromagnetisches Feld und Ausbreitungsrichtung

Für Antennen, welche die magnetischen Komponenten (H-Feld) des elektromagnetischen Feldes wandeln, hat sich der Begriff „magnetische“ Antenne eingebürgert. Diese Bezeichnung ist natürlich nicht ganz korrekt, denn diese Antennen sind ja keinesfalls „magnetisch“. Sie bestehen aus einer niederohmigen Schleife¹⁶², die Teil eines Serienresonanzkreises ist (Abbildung 2.67). Beim Empfang wird durch das magnetische Wechselfeld in der Schleife eine Spannung induziert. Oft erfolgt durch einen integrierten Transformator gleich die Anpassung an die Leitung bzw. den Senderausgang oder den Empfängereingang.

Magnetische Antennen kommen senderseitig nur in Spezialfällen zum Einsatz. Sie haben zwar im Vergleich zu elektrischen Antennen sehr geringe mechanische Abmessungen, jedoch sind die notwendigen hohen Schleifenströme nicht leicht beherrschbar.

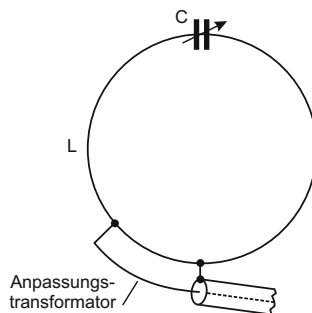


Abbildung 2.67: Magnetische Antennen mit Anpassungstransformator

Die in der drahtlosen Nahfeldkommunikation genutzten Frequenzbereiche sind so hoch, dass die mechanischen Abmessungen kein Entscheidungskriterium bei

¹⁶²engl.: Loop Antenna

Resonanz- betrieb

der Wahl zwischen magnetischer oder elektrischer Antenne sind (siehe Tabelle 2.13). Da die meisten Antennen sowohl für Sende- als auch für Empfangszwecke genutzt werden, kommen daher überwiegend Antennen zum Einsatz, welche die elektrische Feldkomponente wandeln.

Die Grundform einer Antenne, welche die elektrische Feldkomponente (E-Feld) wandelt, ist ein Halbwellendipol¹⁶³. Seine mechanischen Abmessungen richten sich nach der gewünschten Betriebsfrequenz, bei der er resonant sein soll.

Resonanzbetrieb heißt, dass die Antenne mit einer Frequenz betrieben wird, bei der sie sich wie ein ohmscher Widerstand (Fußpunktwiderstand) verhält. Das Verhältnis von Wellenlänge λ der Betriebsfrequenz und den mechanischen Abmessungen der Antenne bestimmen, ob die Antenne in Resonanz ist oder nicht. Die Resonanzfrequenz einer Antenne ist daher von deren mechanischer Abmessung abhängig.

Die Wellenlänge bestimmt sich aus Ausbreitungsgeschwindigkeit der Welle $c = 3000000000$ m/s multipliziert mit der Periodendauer T_P der Welle:

$$\begin{aligned}\lambda &= c \cdot T_P \\ &= 3000000000 / f\end{aligned}$$

Bei einer Frequenz von 1000 MHz ist dann $\lambda = 0,3$ m bzw. bei 2,4 GHz beträgt $\lambda = 12,4$ cm. Da im Antennenmaterial (Draht) die Ausbreitung geringer als mit Lichtgeschwindigkeit (c) erfolgt, ist auch die Wellenlänge entsprechend kleiner. Man muss deshalb bei der Dimensionierung einer realen Antenne einen Verkürzungsfaktor von etwa 0,7 bis 0,9 berücksichtigen.

In Abbildung 2.68 sieht man Spannungs- und Stromverteilung bei der Dipolantenne, wenn sie in Resonanz betrieben wird. An den Enden der beiden Dipoläste befinden sich jeweils Spannungsmaxima und Stromminima. Die Wellenausbreitung erfolgt senkrecht zu den beiden Dipolästen (Vergleiche Abbildung 2.66).

Resonanz- frequenz

In Nähe seiner Resonanzfrequenz verhält sich ein Dipol wie ein Serienschwingkreis. Wird er nicht resonant betrieben, enthält der Fußpunktwiderstand induktive oder kapazitive Blindkomponenten. Durch geeignete Anpassschaltungen (Kompensationsglieder) können Blindanteile kompensiert werden. Dies ist jedoch immer mit einer Verschlechterung der Abstrahlung verbunden (siehe auch Abschnitt 2.3.3.1, Antennenanpassung).

Dieser Fußpunktwiderstand darf nicht mit dem Strahlungswiderstand verwechselt werden.

¹⁶³engl.: Dipole Antenna

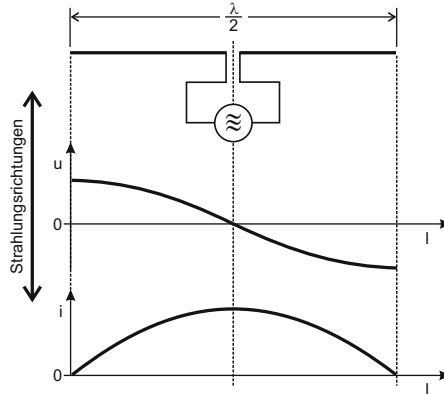


Abbildung 2.68: Spannungs- und Stromverteilung auf einem Halbwellendipol

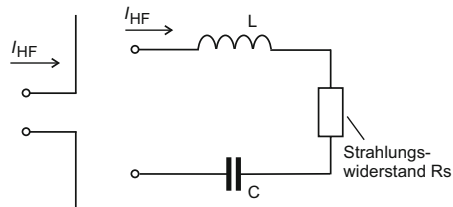


Abbildung 2.69: Ersatzbild einer Dipolantenne in Nähe ihrer Resonanzfrequenz

Definition: Strahlungswiderstand

Da in der Antennenkonstruktion noch Verluste auftreten (z. B. durch den ohmschen Widerstand des Antennenmaterials) wird nur ein Teil der in den Fußpunktwiderstand eingespeisten Hochfrequenzenergie in Strahlungsenergie umgesetzt. Man hat deshalb den Strahlungswiderstand eingeführt. Er stellt den (Ersatz-) Widerstand dar, dessen eingespeiste elektrische Leistung vollständig als elektromagnetische Wellen abgestrahlt wird.

In Abbildung 2.70 ist die Ortskurve des Fußpunktwiderstandes einer Antenne über den Frequenzbereich f_{min} bis f_{max} abgebildet. Unterhalb der Resonanzfrequenz $f < f_0$ zeigt die Antenne kapazitives Verhalten. Im Resonanzfall $f = f_0$ enthält der Fußpunktwiderstand keine Blindkomponenten, er ist ein ohmscher Widerstand. Die Ortskurve schneidet bei den Resonanzfrequenzen die Abszisse, d. h. der (imaginäre) Wert der Ordinate beträgt Null. Oberhalb von f_0 treten noch weitere Resonanzstellen f_1, f_2, f_3 usw. auf, an denen aber ein anderer reeller Fußpunktwiderstand als bei f_0 auftritt. In Abbildung 2.70 sind nur die drei Resonanzstellen f_0 bis f_2 dargestellt.

Ortskurve

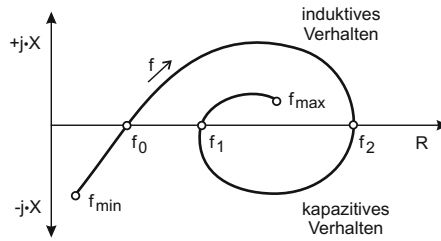


Abbildung 2.70: Verlauf des Fußpunktwiderstandes über die Betriebsfrequenz

Abbildung 2.71 zeigt die Strom- und Spannungsverteilung für resonante Dipole verschiedener Längen. Man erkennt, dass bei allen Resonanzstellen an den Enden der Dipoläste der Strom gleich Null ist. Das ist leicht einzusehen, denn dort endet der elektrische Leiter, also kann kein Strom mehr fließen. Wenn die erste Resonanzstelle bei f_0 ($l = \lambda_0 / 2$) liegt, spricht man von einem Halbwellendipol. Die zweite Resonanzstelle liegt dann bei f_1 ($l = \lambda_0$), die dritte bei f_2 ($l = 3/2 \cdot \lambda_0$) usw.

Resonanzstellen

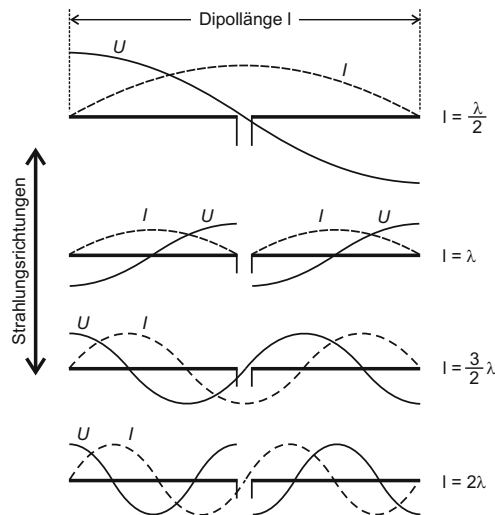


Abbildung 2.71: Strom- und Spannungsverteilung für resonante Dipole verschiedener Längen

An den verschiedenen Resonanzstellen zeigt die Dipol-Antenne unterschiedliche Strahlungsdiagramme. In Abbildung 2.72 sind die Strahlungsdiagramme für einen Halbwellendipol (links) und einen Vollwellendipol (rechts) dargestellt. Man kann leicht erkennen, dass die der Antenne zugeführte Energie nicht in alle

Strahlungsdiagramme

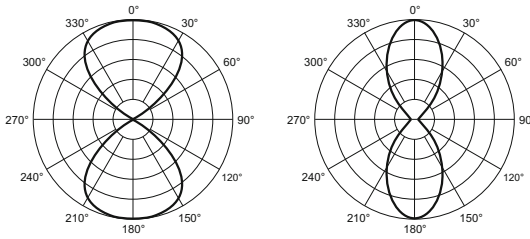


Abbildung 2.72: Links: Strahlungsdiagramm eines Halbwellendipols ($l = \lambda/2$); rechts: Strahlungsdiagramm eines Vollwellendipols ($l = \lambda$)

Richtungen gleichstark abgestrahlt wird. Die Hauptstrahlung erfolgt senkrecht zu den Dipolästen, parallel zu den Dipolästen erfolgt keine Abstrahlung. Die Ausprägung der Richtwirkung ist stark von der Antennenbauform abhängig. Im Beispiel Abbildung 2.72 erkennt man, dass die Richtwirkung eines Vollwellendipols deutlich stärker ausgeprägt ist als die eines Halbwellendipols.

Definition: Richtantennen

Wenn das von einer Antenne ausgestrahlte elektromagnetische Feld in eine Richtung (oder mehrere Richtungen) stärker ausgestrahlt wird als in andere, spricht man von einer Richtwirkung der Antenne. Die Ausprägung der Richtwirkung kann man dem Strahlungsdiagramm einer Antenne entnehmen, welches genau genommen für alle Raumwinkel erstellt werden müsste. Von einer Richtantenne spricht man, wenn die Ausstrahlung überwiegend in eine Hauptstrahlrichtung erfolgt. Im Vergleich zu einer Antenne, die in alle Richtungen gleichstark abstrahlt, haben Richtantennen in ihrer Hauptstrahlrichtung eine Antennenverstärkung.

Dieses Verhalten einer (passiven) Antenne ist reziprok, d. h. die Antennenverstärkung ist sowohl bei Sende- als auch bei Empfangsantennen wirksam.

Richtantennen sind bei Punkt-zu-Punkt-Verbindungen vorteilhaft. Beim Empfänger können nicht erwünschte Empfangsrichtungen ausgeblendet und damit eventuelle Störungen aus diesen Richtungen unterdrückt werden. Sendeseitig wird durch eine Richtantenne die maximale Sendeenergie zielgerichtet zur Empfangsstelle ausgesandt. Damit wird nicht nur eine bessere Energiebilanz erzielt, sondern auch unerwünschtes Mithören an anderen Positionen als der Empfangsstelle erschwert (bzw. der belegte Übertragungsfrequenzbereich kann an diesen anderen Positionen bereits wieder für andere Zwecke genutzt werden).

Bei der Kommunikation von Mobilstationen müsste eine Richtantenne ständig nachgeführt werden. Techniken, eine Antenne rein elektronisch ohne mechanische Bewegung in die gewünschte Sende- oder Empfangsrichtung auszurichten, sind in Entwicklung (Smart Antennas).

isotrope Antenne

Zur Angabe der Antennenverstärkung (auch als Antennengewinn bezeichnet) wird die Strahlungsleistung P_S der Richtantenne in Hauptstrahlrichtung mit der Strahlungsleistung P_i einer isotropen Antenne (siehe unten) verglichen, wenn in beide Antennen die gleiche Leistung eingespeist wird. Der Antennengewinn G_i gibt an, um welchen Faktor die Strahlungsleistung der Richtantenne in Hauptstrahlrichtung höher ist, als die einer isotropen Antenne.

$$G_i = P_S / P_i$$

Entsprechend dem Reziprozitätsprinzip (siehe unten) gilt diese Definition sowohl für den Empfangs- als auch für den Sendefall. Die Angabe ist dimensionslos und erfolgt oft in dB.

$$g_i = 10 \log P_S / P_i \text{ [dB]}$$

Da als Vergleichsantenne ein verlustloser isotroper Kugelstrahler (siehe unten) nicht so ohne weiteres realisierbar ist, kommt in der Praxis oft auch ein Halbwellendipol als Messantenne zum Einsatz. Entweder wird dann der Gewinn bezüglich des Halbwellendipols (G_d bzw. g_d) angegeben oder er wird auf den Gewinn eines isotropen Strahlers umgerechnet. Der Gewinn eines Halbwellendipols bezüglich eines verlustlosen isotropen Kugelstrahlers beträgt $g_i = 2,15$ dB.

Beispiel: Die Vergleichsmessung einer Antenne mit einem Halbwellendipol ergibt einen Antennengewinn von $g_d = 5$ dBd. Der Antennengewinn g_i bezüglich eines isotropen Strahlers beträgt dann $g_i = 7,15$ dBi.

Wenn die in eine Antenne eingespeiste Leistung und ihr Antennengewinnfaktor G_i bekannt sind, kann die in Hauptstrahlungsrichtung abgegebene Wirkleistung P_S bestimmt werden. Wenn die Weglänge r zwischen Sender und Empfänger bekannt ist (und $r < 4 \cdot \lambda$), können am Empfangsort (im Fernfeld die Strahlungsdichte S (oder Leistungsflussdichte (LDF)), und die Stärke des E- und H-Feldes) bestimmt werden (siehe unten).

Definition: Isotroper Kugelstrahler

Der isotrope Kugelstrahler ist eine theoretische Modellantenne, die nicht realisiert werden kann. Sie hat eine punktförmige mechanische Ausdehnung und strahlt in alle Richtungen gleichstark (omnidirektional). Wenn der isotrope Kugelstrahler im Mittelpunkt einer Kugel angeordnet wäre, könnte man an jeder Stelle der Kugeloberfläche die gleiche Strahlungsdichte messen. Der isotrope Kugelstrahler wird als Vergleichsantenne herangezogen, wenn der Antennengewinn angegeben werden soll.

Während der isotrope Kugelstrahler senderseitig nicht realisierbar ist, kann für Empfangszwecke eine solche omnidirektionale Antenne in gewissen Grenzen

durch drei aktive Dipolantennen (siehe unten), die in den Raumkoordinaten x , y , und z orthogonal zueinander angeordnet sind, nachgebildet werden. Um die Ersatzfeldstärke E_r ¹⁶⁴ richtungsunabhängig zu messen, müssen die von den drei Antennen gemessenen Feldstärkewerte vektoriell addiert werden (vergleiche Abschnitt 5.5.1, Feldstärkemessgerät). Der Betrag des Feldstärkevektors ergibt sich dann aus

$$E_r = \sqrt{E_x^2 + E_y^2 + E_z^2}$$

Eine omnidirektionale Empfangsantenne wird als Messantenne eingesetzt, denn sie muss zur Feldstärkebestimmung nicht auf die Position des Senders ausgerichtet werden.

Definition: Aktivantenne

Als aktive Antenne bezeichnet man die mechanische Einheit aus Empfangsantenne (passiv) und einem integrierten Verstärker (aktiv). Der Verstärker ist optimiert auf die elektrischen Eigenschaften der Antenne und kompensiert deren Frequenzgang. Man erhält so ein kleines, hochempfindliches, breitbandiges, nichtreziprokes Antennensystem, welches für den Verstärker eine separate Stromversorgung benötigt. Mit Hilfe geeigneter Weichen erfolgt die Stromversorgung oft über die Antennenzuleitung.

Feldstärkenbestimmung im Fernfeld

Im Nahfeld einer Antenne stehen, bedingt durch die mechanische Antennenkonstruktion, die elektrischen und magnetischen Feldlinien nicht senkrecht aufeinander. Die Feldstärke ist deshalb in diesem Bereich schwierig berechenbar. Im Fernfeld (siehe unten) geht man davon aus, dass diese Bedingung erfüllt ist, bzw. der Fehler vernachlässigbar klein ist. H- und E-Feld befinden sich in Phase, und es wird mit dem strahlenden Feld eine Wirkleistung abgestrahlt.

Die Strahlungsdichte S (auch als Leistungsflussdichte¹⁶⁵ bezeichnet) ist eine gerichtete Größe. Sie ist proportional zur der von der Antenne in den Raum abgegebenen Wirkleistung P_S und nimmt mit dem Faktor $1/r^2$ ab (r : Abstand der Antenne zum Messpunkt):

$$S = P_S / (4 \cdot \pi \cdot r^2)$$

Auch die elektrischen und magnetischen Komponenten des elektromagnetischen Feldes beschreiben die Intensität des elektromagnetischen Feldes:

$$S = E \cdot H$$

Da sich H- und E-Feld in Phase befinden, gilt die Beziehung:

$$E = H \cdot Z_0$$

Z_0 ist der Freiraum-Feldwellenwiderstand, oft auch nur als Feldwellenwiderstand bezeichnet:

$$Z_0 = \sqrt{\mu_0 / \epsilon_0} = 376,68 \, \Omega$$

**Wellen-
widerstand**

¹⁶⁴engl.: Resultant Field Strength

¹⁶⁵engl.: Power Flux Density (PFD)

Damit erhält man für die elektrische Feldstärke E:

$$E = \sqrt{S \cdot Z_0}, \\ = (1 / 2 \cdot r) \cdot \sqrt{P_S \cdot Z_0 / \pi},$$

und für die magnetische Feldstärke H entsprechend:

$$H = \sqrt{S / Z_0} \\ = (1 / 2 \cdot r) \cdot \sqrt{P_S / (Z_0 \cdot \pi)},$$

Die elektrische und die magnetische Feldstärke nehmen mit dem Faktor $1/r$ ab. Die oben angeführten Berechnungen gelten für den verlustfreien Raum. Für eine reale hindernisfreie Verbindung muss das Freiraum-Dämpfungsmaß α_0 berücksichtigt werden, dessen Wert von der Wellenlänge λ und der zu überbrückenden Distanz d abhängig ist: $\alpha_0 = 20 \cdot \log[(4 \cdot \pi \cdot d) / \lambda]$ dB

Aufgabe: Eine Antenne strahlt eine Leistung von $P_S = 1$ W auf einer Frequenz von $f = 800$ MHz ab. Bestimmen Sie die Strahlungsdichte, die elektrische und die magnetische Feldstärke in $r = 1000$ m Entfernung!

Lösungen

Überprüfung der Fernfeldbedingung:

$$\lambda = c / (800 \cdot 10^6) = 37,5 \text{ cm.}$$

Da $(r \gg 4 \cdot \lambda)$ ist, gilt hier die Fernfeldbedingung.

Die Leistungsflussdichte S ergibt sich aus:

$$S = P_S / (4 \cdot \pi \cdot r^2) \\ = 1 / (4 \cdot \pi \cdot 10^6) \\ = 79,6 \text{ nW/m}^2$$

In 1000 m Entfernung stehen somit pro Quadratmeter Empfangsfläche 79,6 nW zur Verfügung. Wenn diese verfügbare Leistung nicht für einen Empfang ausreicht, muss entweder die wirksame Antennenfläche am Empfangsort vergrößert oder die Sendeleistung erhöht werden.

Entsprechend zu oben errechnet sich die elektrische Feldstärke E (ohne Berücksichtigung des Freiraum-Dämpfungsmaßes):

$$E = (1 / 2 \cdot r) \cdot \sqrt{P_S \cdot Z_0 / \pi} \\ = 5,476 \cdot \sqrt{P_S} / r \\ = 5,475 \text{ mV/m.}$$

Die magnetischen Feldstärke H bestimmt sich aus:

$$H = (1 / 2 \cdot r) \cdot \sqrt{P_S / (Z_0 \cdot \pi)} \\ = 14,53 \cdot 10^{-3} \cdot \sqrt{P_S} / r \\ = 14,53 \mu \text{ A/m.}$$

Es gibt verschiedene Definitionen für den Beginn des Fernfeldbereiches. Sie sind abhängig von den mechanischen Abmessungen der Antenne. Die deutsche Bundesnetzagentur (ehemals Regulierungsbehörde für Post und Telekommunikation

(RegTP)) geht bei ihren Berechnungen der Grenzwerte von gesundheitsschädlichen elektromagnetischen Feldstärken vom Beginn des Fernfeldes bei Entfernungen von mehr als $(4 \cdot \lambda)$ zwischen strahlender Antenne und dem Messpunkt aus. Danach beginnt das Fernfeld für

$f = 1000 \text{ MHz}$ bei 1,2 m

$f = 2,4 \text{ GHz}$ bei 0,5 m

(vergleiche hierzu Abschnitt 2.3.1.5).

Wird eine Antenne als Sendeantenne betrieben und ist in deren Fernfeld eine zweite Antennen als Empfangsantenne installiert, so stellt diese Anordnung, von ihren Anschlussklemmen her betrachtet, einen Vierpol dar. Der Vertauschungssatz der Vierpoltheorie besagt hier, dass es keine Rolle spielt, welche Antenne als Sende- und welche als Empfangsantenne betrieben wird, das Verhalten des Gesamtsystems ist davon unabhängig.

Definition: Reziprozitätsprinzip

Das Reziprozitätsprinzip bedeutet, dass die Verstärkung durch eine Antenne (Richtantenne) sowohl senderseitig als auch empfangsseitig erfolgen kann. Das gilt nicht für Aktivantennen, also Antennen mit einem eingebauten Hochfrequenzverstärker. Auch die auftretenden Feldstärken können natürlich unterschiedlich sein.

$\lambda/4$ -Antenne

Wenn ein Ast eines Dipols senkrecht über einer elektrisch leitenden Fläche angeordnet ist (Abbildung 2.73, Links), verhält er sich elektrisch so, als wäre ein zweiter Ast spiegelbildlich vorhanden. Diese Antennenform bezeichnet man als Monopol, als Vertikalstrahler oder als $\lambda/4$ -Strahler.

Sie wird gern für Sende- und Empfangszwecke eingesetzt, denn wenn eine elektrisch gut leitende Fläche vorhanden ist (z. B. eine Kfz-Karosserie), beträgt ihre mechanische Abmessung nur $\lambda/4$. Wenn eine gut leitende Fläche nicht zur Verfügung steht, z. B. wenn die Antenne auf dem Erdboden platziert wird, kann ein „künstliches Gegengewicht“ geschaffen werden. Dazu wird z. B. im Erdboden eine elektrisch leitende Fläche durch dort verlegte Drähte geschaffen (siehe Abbildung 2.73, Mitte).

Eine Weiterentwicklung dieser Antennenform ist die Groundplane-Antenne (Abbildung 2.73, rechts). Hier wird das Gegengewicht durch wenige $\lambda/4$ -Leiterstücke, so genannte Radials, geschaffen. Bei der Dimensionierung des Vertikalstrahlers und der Radials ist der Verkürzungsfaktor V (siehe oben) zu berücksichtigen.

Weiterführende Literatur zum Thema Antennentechnik findet man unter [RK02].

Groundplane-Antenne

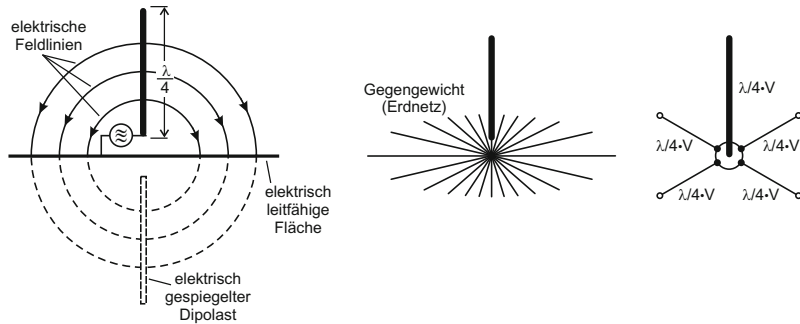


Abbildung 2.73: Links: Monopol über leitender Fläche; Mitte: $\lambda/4$ -Monopolstrahler mit Gegengewicht; Rechts: Groundplane-Antenne

2.3.4 Betriebsarten

Im nachfolgenden Teil werden verschiedene Betriebsarten behandelt. Ziel der hier vorgestellten Techniken ist es stets, die Übertragungssicherheit zu erhöhen. Diese wird bei Funkübertragungsstrecken durch den Mehrfachempfang und die daraus resultierenden Interferenzzonen mit Schwunderscheinungen¹⁶⁶ der Signalstärke beeinträchtigt.

2.3.4.1 Mehrwege-Empfang und Schwund

Die Verhältnisse sind selten so, dass den Empfänger nur ein Signal direkt vom Sender erreicht. Die Regel ist der Mehrfachempfang, bedingt durch Reflexionen des Sendesignals. Im Frequenzbereich unter 30 MHz ist dies häufig erwünscht. Abbildung 2.74 zeigt, wie erst durch Reflexionen in der Ionosphäre ein weltweiter Rundfunkempfang möglich wird.

Unmittelbar um den Sender, jedoch bereits außerhalb der Sichtweite, existiert eine „Tote“ Zone, in der, bedingt durch die Erdkrümmung, kein Empfang möglich ist. Weitere „Tote“ Zonen, wie man nach der Abbildung 2.74 vermuten könnte, sind durch die instabile Reflexionsschicht in der Ionosphäre selten (siehe Abbildung 2.75). Die Reflexionen erfolgen in verschiedenen Schichten der Ionosphäre etwa zwischen 150 und 400 km Höhe. Die verschiedenen Höhen bewirken auf der Erdoberfläche Interferenzzonen, in denen ein Mehrfachempfang erfolgt. Da sich die reflektierenden Schichten auch noch zeitlich ändern, führt das zu zeitlich stark schwankenden Feldstärkewerten im Bereich der Interferenzzone. Dies bezeichnet man als „Schwund“.

Mehrwege-empfang

Fading

¹⁶⁶engl.: Fading

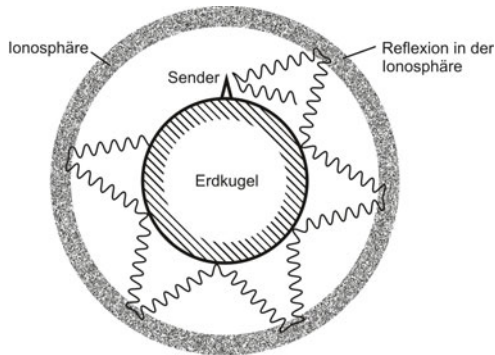


Abbildung 2.74: Fernempfang bei Frequenzen unter 30 MHz Interferenzzonen

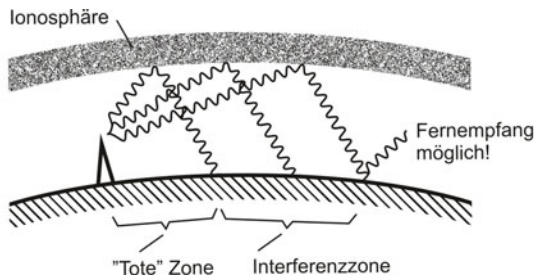


Abbildung 2.75: Interferenzzonen

Diese Effekte treten prinzipiell auch bei Nahbereichs-Kommunikation auf. Hier erfolgt eine Reflexion jedoch nicht in der Ionosphäre, sondern an elektrisch leitenden Objekten (in der Regel an Metallflächen) in der Umgebung der Sende- und Empfangseinrichtungen. Ein Mehrfachempfang ist hier die Regel (siehe Abbildung 2.74). In Extremfällen kann es zu Signalverdoppelungen und Signalauslöschungen kommen, die stabil bleiben, wenn sich die Weglängen und Sendefrequenzen nicht ändern. Eine „Tote“ Zone hingegen gibt es selten. Wenn keine direkte Sichtverbindung zwischen Sender und Empfänger besteht, ist häufig dennoch eine Kommunikationsverbindung möglich. Diese erfolgt dann zum Beispiel über Reflexionen an metallischen Objekten.

Am Empfänger treffen zwei Signale des Senders ein (siehe Abbildung 2.76), eines auf direktem Weg (S_D) und eines über eine Reflexion (S_R). Die Phasendifferenz zwischen diesen beiden Signalen kann je nach Verzögerung durch den Umweg zwischen 0° und 180° betragen. Vernachlässigt man die Dämpfung entlang der Übertragungsstrecke, kann es so bei Gleichphasigkeit zur einer Amplitudenverdopplung kommen (siehe Abbildung 2.77, links) oder im ande-

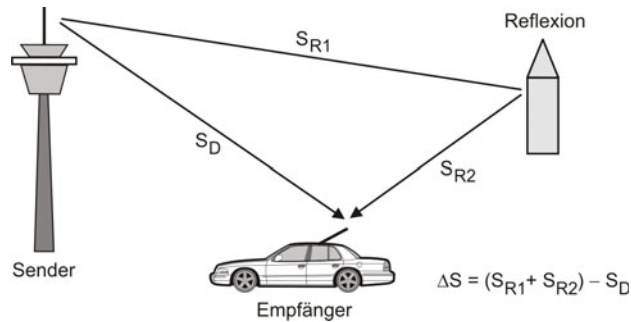


Abbildung 2.76: Doppelempfang durch Reflexion

ren Extremfall bei Gegenphasigkeit (180° Phasendifferenz) zu einer kompletten Auslöschung des Signals (siehe Abbildung 2.77, rechts).

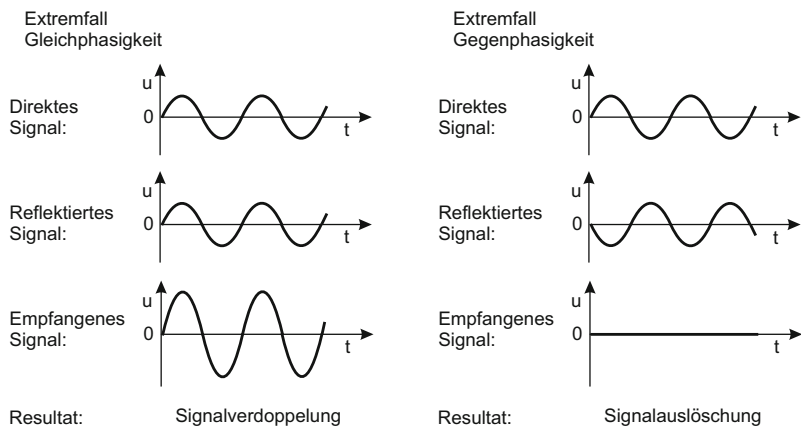


Abbildung 2.77: Auswirkungen des Doppelempfangs

Ob einer der Extremfälle eintritt oder ein Wert dazwischen, hängt von den Parametern der Nachrichtenstrecke ab. Die Phasendifferenz berechnet sich wie folgt:

$$\varphi_{Rad} = 2 \cdot \pi \cdot [\text{mod}^{167}(\Delta s / \lambda_s)] / \lambda_s, \text{ bzw.}$$

$$\varphi_{Grad} = 360 \cdot [\text{mod}(\Delta s / \lambda_s)] / \lambda_s [^\circ]$$

mit

Δs : Streckendifferenz zwischen der direkten Verbindung zwischen Sender (S_D) und Empfänger und Weglänge des reflektierten Signals ($S_R = S_{R1} + S_{R2}$),
 $\Delta s = S_R - S_D$

¹⁶⁷Mod = **Modulo**

f_s : Sendefrequenz [Hz]

λ_s : Wellenlänge des Sendesignals,

$\lambda_s = c/f_s$ [m] mit $c \approx 300 \text{ km/s} = 300 \cdot 10^6 / f_s$ [m]

T_s : Periodendauer des Sendesignals, $T_s = 1/f_s$ [s]

Beispiel: Signalverdoppelung durch Interferenz:

Ein Sender arbeitet auf der Frequenz $f_s=1000$ MHz, das entspricht einer Wellenlänge von $\lambda_s=0,3$ m. Die Weglänge des direkten Signals betrage $S_D=10$ m, die Weglänge des reflektierten Signals betrage $S_R=13$ m. Somit ist die Streckendifferenz $\Delta s=3$ m. Daraus berechnet sich die Phasendifferenz zwischen direktem Signal und reflektiertem Signal zu

$$\varphi_{Grad} = 360 \cdot [\text{mod}(3/0,3)]m/0,3m [^\circ] \\ = 0^\circ$$

Erläuterung

Die beiden Signale am Empfängereingang treffen dort gleichphasig ein, d. h. ihre Amplitudenwerte addieren sich. Es kommt gegenüber der Amplitude des direkt empfangenen Signals zu einer Amplitudenverdoppelung, wenn man eine gleiche Streckendämpfung für beide Signale annimmt.

Beispiel: Signalauslöschung durch Interferenz:

Ein Sender arbeitet auf der Frequenz $f_s = 1000$ MHz, das entspricht einer Wellenlänge von $\lambda_s = 0,3$ m. Die Weglänge des direkten Signals betrage $S_D = 10$ m, die Weglänge des reflektierten Signals betrage $S_R = 13,45$ m. Somit ist die Streckendifferenz $\Delta s = 3,45$ m. Daraus berechnet sich die Phasendifferenz zwischen direktem Signal und reflektiertem Signal zu

$$\varphi_{Grad} = 360 \cdot [\text{mod}(3,45/0,3)]m/0,3m [^\circ] \\ = 360 \cdot 0,15m/0,3m [^\circ] = 180^\circ$$

Erläuterung

Die beiden Signale am Empfängereingang treffen dort mit einer Phasendifferenz von 180° , also gegenphasig, ein, d. h. ihre Amplitudenwerte subtrahieren sich. Es kommt gegenüber der Amplitude des direkt empfangenen Signals zu einer Signalauslöschung, wenn man eine gleiche Streckendämpfung für beide Signale annimmt.

Beispiel: Signalauslöschung durch Interferenz:

Ein Sender arbeitet auf der Frequenz $f_s = 1050$ MHz, das entspricht einer Wellenlänge von $\lambda_s = 0,2857$ m. Die Weglänge des direkten Signals betrage $S_D = 10$ m, die Weglänge des reflektierten Signals betrage $S_R = 13$ m. Somit ist die Streckendifferenz $\Delta s = 3$ m.

Daraus berechnet sich die Phasendifferenz zwischen direktem Signal und reflektiertem Signal zu $\varphi_{Grad} = 360 \cdot [\text{mod}(3/0,2857)]m/0,2857m [^\circ]$
 $= 360 \cdot 0,143m/0,2857m [^\circ] = 180^\circ$

Erläuterung

Die beiden Signale am Empfängereingang treffen dort mit einer Phasendifferenz von 180° , also gegenphasig, ein, d. h. ihre Amplitudenwerte subtrahieren sich. Es kommt gegenüber der Amplitude des direkt empfangenen Signals zu einer Signalauslöschung, wenn man eine gleiche Streckendämpfung für beide Signale annimmt.

Wenn sich, wie in Abbildung 2.76 dargestellt, der Empfänger bewegt, kommt es auf Grund der sich ständig ändernden Weglängendifferenz zwischen dem direkten und dem reflektierten Signal zu Schwankungen der Empfangsfeldstärke entlang der Fahrtstrecke. Diese würden auch entstehen, wenn sich der Sender bewegt oder der Reflektor. Letzteres ist bei Schwankungen der Reflexionsschicht in der Ionosphäre der Fall (Abbildung 2.75).

Aus den Beispielen (1) bis (3) ist leicht zu entnehmen, dass es in Abhängigkeit von der Weglängendifferenz und der Sendefrequenz bei sonst gleichen Bedingungen in den Extremfällen zu einer Amplitudenverdoppelung oder zu einer Signalauslöschung kommen kann. Damit ist auch bereits gezeigt, mit welchen Maßnahmen eine unerwünschte Signalauslöschung verhindert werden kann: Man diversifiziert die Signalübertragung auf unterschiedliche Weglängen, unterschiedliche Betriebsfrequenzen oder bei bewegtem Sender, Empfänger oder Reflektor auf unterschiedliche Zeiten. Man spricht in diesem Fällen vom Diversity-Betrieb.

2.3.4.2 Diversity-Verfahren

Im Folgenden wird auf die drei Diversity-Betriebsarten Raum, Frequenz und Zeit einzeln eingegangen.

Raum-Diversity

- Raum-Diversity¹⁶⁸:

Beim Raumdiversity-Betrieb werden zwei Sende-Empfangs-Wege mit unterschiedlichen Weglängen parallel betrieben. Eine Signalauslöschung ist, wie oben gezeigt, von der Weglängendifferenz abhängig. Damit ist zu

¹⁶⁸engl.: Space Diversity

erwarten, dass bei der Auslöschung des Signals eines Weges das andere keine Auslöschung zeigt. Auf der Empfangsseite wird dazu ständig geprüft, welcher der beiden Signalwege am Empfänger eine höhere Feldstärke bewirkt. In der Praxis wird ein Raum-Diversity durch zwei räumlich getrennte Empfangsantennen realisiert. Der Empfänger schaltet zwischen diesen beiden Antennen (oder weiteren Antennen) hin und her, misst deren Empfangssignalstärke und wählt das stärkere Signal aus. Der Abstand zwischen den beiden Empfangsantennen sollte idealerweise 2 bis 5 λ betragen. Aus Platzgründen ist dies nicht immer realisierbar. In diesem Fall kann auf den Frequenz-Diversity-Betrieb zurückgegriffen werden

- Frequenz-Diversity¹⁶⁹:

Hierbei senden zwei Sender gleichzeitig, jedoch auf zwei unterschiedlichen Frequenzen. Eine Signalauslöschung ist, wie oben gezeigt, von der Wellenlänge und damit von der benutzten Sendefrequenz abhängig. Damit ist zu erwarten, dass bei der Auslöschung einer Sendefrequenz die andere keine Auslöschung zeigt. Auf der Empfangsseite wird dazu ständig geprüft, welcher der beiden Sendefrequenzen am Empfänger eine höhere Feldstärke bewirkt. Entsprechend hoch ist der senderseitige Aufwand, denn man muss einen zweiten Sender parallel betreiben und einen zweiten Übertragungskanal (Sendefrequenz) belegen. In der früheren Kurzwellen-Fernübertragungstechnik hat man das in Kauf genommen. In der heutigen digitalen Nahbereichs-Kommunikation geht man jedoch einen anderen Weg. Man realisiert einen Frequenz-Diversity-Betrieb durch das „Frequency-Hopping“-Verfahren. Im Abschnitt 2.3.5.1 wird es näher beschrieben.

Frequenz-Diversity

- Zeit-Diversity¹⁷⁰:

Beim Empfang eines direkten und eines reflektierten Signals und gleichzeitiger Bewegung des Senders, des Empfängers oder des Reflektors verändern sich die Phasenlagen der beiden am Empfänger eintreffenden Signale in Abhängigkeit von der Änderung der Weglängen. Am Empfangsort nimmt man das als zeitabhängige Schwankungen der Empfangsfeldstärke wahr. Beim Zeit-Diversity-Betrieb wird deshalb eine Aussendung nach einer gewissen Zeit einmal oder mehrfach wiederholt. Es ist zu erwarten, dass bei einer der Wiederholungen keine Auslöschung auftritt und so die zu übertragende Information übermittelt werden kann. Der Aufwand beim Sender und beim Empfänger ist so am niedrigsten. Die Informationen müssen jedoch für die Wiederholungen auf der Senderseite und für die Auswertungen auf der Empfangsseite zwischengespeichert werden. Je

Zeit-Diversity

¹⁶⁹engl.: Frequency Diversity

¹⁷⁰engl.: Time Diversity

nach Wahl der Zeit zwischen den Wiederholzyklen ist dieses Verfahren nur bedingt echtzeitfähig. Dies ist bei einer Datenübertragung eher unkritisch, kann jedoch zum Beispiel eine Sprachübertragung stark behindern.

In der Praxis werden in der Regel Mischformen der verschiedenen Diversity-Betriebsarten eingesetzt. Sie bieten nicht nur Schutz gegen das Fading, sondern auch gegen sonstige kurzzeitige Störungen in einem Übertragungskanal.

Mobilität und Portabilität

Es ist aus den oben angeführten Gründen sinnvoll, zwischen den Begriffen „Mobilität“ und „Portabilität“ in der Nahbereichs-kommunikation zu unterscheiden.

Merksatz: **Mobiles Kommunikationssystem**

Ein mobiles Kommunikationssystem muss während einer Ortsveränderung betriebsbereit sein. Das auf Grund der Bewegung übertragungstechnisch bedingte Fading kann durch eine optimierte Kanalcodierung abgefangen werden.

Eine Signalauslöschung durch Fading ist zeitlich begrenzt. Die Informationsmenge, die während einer Signalauslöschung verloren geht, ist abhängig von der Datenübertragungsrate im Kanal. Die Fehlerkorrektureigenschaften des Kanaldecodierers müssen so mächtig sein, dass die während der Signalauslöschung nicht übertragenen Informationen regeneriert werden können.

Der zeitliche Zyklus zwischen den Auslöschungen ist wiederum abhängig von der Bewegungsgeschwindigkeit. Unterschreitet dieser ein unteres Limit, treten mehr Fehler auf, als der Kanaldecodierer korrigieren kann, und die Kommunikation bricht ab.

Das heißt, ein mobiles Kommunikationssystem kann in Abhängigkeit von der Datenübertragungsrate und seinen Fehlerkorrektureigenschaften nur bis zu einer maximalen Bewegungsgeschwindigkeit betrieben werden. Beim Mobilfunkstandard GSM beträgt diese etwa 220 km/h.

Merksatz: **Portables Kommunikationssystem**

Ein portables Kommunikationssystem ist ortsungebunden, solange sich der Empfänger in Reichweite des Senders befindet. Auch die Energieversorgung ist oft für einen stromnetzunabhängigen Betrieb ausgelegt (siehe Abschnitt 2.4).

Während der Aussendung oder des Empfanges von Daten darf jedoch der Standort nicht gewechselt werden, oder wenn doch, dann nur mit einer so kleinen Geschwindigkeit, dass die durch Fading bedingte Signalauslöschung vernachlässigbar ist.

2.3.4.3 Multiplex-Verfahren

Eine häufige Anforderung an einen Übertragungskanal ist dessen Möglichkeit zur Mehrfachnutzung. In der klassischen Fernmeldetechnik war es aus Effizienzgründen erforderlich, mehrere Telefongespräche über ein Kabel (Kanal) zu übertragen. Bei der drahtlosen Übertragungstechnik bedeutet dies, dass ein Funkkanal (Frequenzband) von mehreren Quellen unabhängig voneinander genutzt wird. Auf der Empfangsseite müssen die verschiedenen Signale wieder voneinander getrennt werden. Man unterscheidet dabei drei Grundprinzipien:

**Grund-
prinzipien**

- SDMA¹⁷¹
- TDMA¹⁷²
- FDMA¹⁷³

Darüber hinaus sind auch Kombinationen dieser Verfahren möglich. Weiterentwicklungen mit besonderen Eigenschaften sind:

**Weiterent-
wicklungen**

- CDMA¹⁷⁴
- OFDM¹⁷⁵

Raummultiplex

In der Funktechnik wird beim Raummultiplexverfahren um jeden Sender eine Zone eingerichtet, in der kein anderer Sender auf der gleichen Frequenz sendet. Damit können die benutzten Frequenzen (Kanäle) mehrfach vergeben werden, denn gegenseitige Störungen sind ausgeschlossen (siehe Abbildung 2.78).

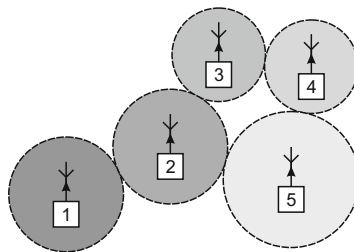


Abbildung 2.78: Raummultiplex

¹⁷¹SDMA = Space Division Multiplex Access (deutsch: Raummultiplex)

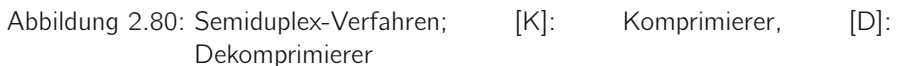
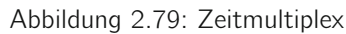
¹⁷²TDMA = Time Division Multiplex Access (deutsch: Zeitmultiplex)

¹⁷³FDMA = Frequency Division Multiple Access (deutsch: Frequenzmultiplex)

¹⁷⁴CDMA = Code Division Multiple Access (deutsch: Codemultiplex)

¹⁷⁵OFDM = Orthogonal Frequency Division Multicarrier

Beim Zeitmultiplexverfahren wird jeder Quelle ein „Zeitschlitz“ zugewiesen. Zeitlich nacheinander bekommt jede Quelle für eine vorgegebene Zeit das gesamte Übertragungsfrequenzband (Kanal) zugeteilt. Diese Zuweisungen müssen natürlich zwischen Sender und Empfänger zeitlich synchronisiert sein (siehe Abbildung 2.79). Eine besondere Form des Zeitmultiplexbetriebs ist das TDD¹⁷⁶-Verfahren. Hier wird zwischen den zwei Stationen A und B eine „Quasi“-Vollduplexverbindung geschaltet, die jedoch durch zwei Zeitscheiben, einmal in die Richtung A ' B und dann in Richtung B ' A realisiert wird. Dieser Semiduplexbetrieb erspart die aufwendigen Filter (Duplexweiche) zwischen Senderausgang / Empfängereingang und der gemeinsamen Sende- / Empfangsantenne. Abbildung 2.80 zeigt den zeitlichen Ablauf. Bevor ein Datenpaket von der Station A auf dem Kanal ausgesendet wird, muss es so komprimiert werden, dass es den Kanal nur während der halben seiner ursprünglichen Übertragungszeit belegt.



¹⁷⁶TDD = Time Division Duplex

Die so gewonnene freie Übertragungszeit des Kanals wird für die ebenfalls (von der hier in Abbildung nicht dargestellten Station B) komprimierten Datenpakete von der Station B nach A genutzt. Station A muss die komprimierten Datenpakete dekomprimieren. Die Komprimierungs- und Dekomprimierungsvorgänge kosten Zeit und müssen zwischen den beteiligten Stationen synchronisiert werden.

Frequenzmultiplex

Beim Frequenzmultiplex-Verfahren beansprucht jeder Sender einen Teil der Gesamtübertragungsbandbreite (in Abbildung 2.81 dargestellt als Bandbreiten B_1 bis B_5 der Unterkanäle). Der Abstand der Trägerfrequenzen muss so groß sein, dass sich auf der Empfängerseite die durch die Modulation entstehenden Seitenbänder nicht überlagern und somit stören.

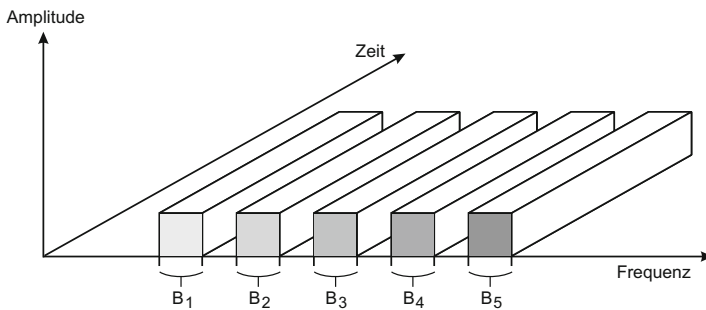


Abbildung 2.81: Frequenzmultiplex

Beispiel: Anwendungen

- Raummultiplex: Kabelinfrastruktur (mehrere Adern pro Kabel)
- Zeitmultiplex: GSM (D- und E-Netz, acht Sprachkanäle)
- Frequenzmultiplex: Hörfunk
- Codemultiplex: Spreizspektrumtechnik

Codemultiplex

Nach einem ganz anderen Prinzip arbeitet das Codemultiplex-Verfahren. Es basiert auf dem Prinzip der Spreizspektrumtechnik (siehe Abschnitt 2.3.5.2). Alle Sender teilen sich gleichzeitig das gleiche Übertragungsfrequenzband. Die Unterscheidung erfolgt durch unterschiedliche Spreizcodes (Chip-Sequenzen).

Beispiel: Erzeugung eines Codemultiplexsignals. Abbildung 2.82 zeigt in den Zeilen a) bis d) die Generierung eines gespreizten Signals A_S aus der Nachrichtenbitfolge

$$A_N = [1, 0, 1, 1, 0]$$

und der Pseudozufallsbitfolge

$$A_C = [1, 0, 1, 0, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 1, 1, 0, 1, 0].$$

Dieser Vorgang wurde bereits im Abschnitt 2.3.5.1 beschrieben. In den Zeilen e) bis h) wird die Generierung eines gespreizten Signals B_S aus der Nachrichtenbitfolge $B_N = [1, 0, 0, 0, 1]$

und der Pseudozufallsbitfolge

$$B_C = [1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1]$$

gezeigt.

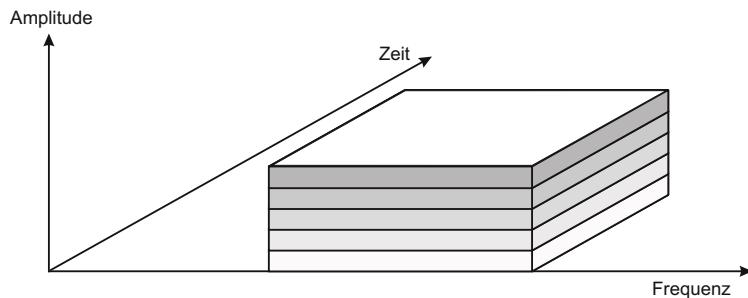


Abbildung 2.82: Codemultiplex

Erläuterung

Beide Spreizfolgen werden einem Träger aufmoduliert und im gleichen Frequenzbereich (zur gleichen Zeit) abgestrahlt. Die besonderen Anforderungen an die Spreizfolgen werden unten unter „Anforderung an die Modulation beim CDMA-Verfahren“ beschrieben.

An einem Empfänger im Empfangsbereich der beiden Sender wird daher das Summensignal aus A_S und B_S eintreffen (siehe Abbildung 2.83, Zeile i)).

- Zeile a): Darstellung der zu spreizenden Nachrichtenbitfolge $A_N = [1,0,1,1,0]$, hier kombiniert als logische und binäre Signalfolge.
- Zeile b): Darstellung der Spreizsequenz A_C , hier kombiniert dargestellt als logische und binäre Signalfolge.
- Zeile c): Gespreiztes Nachrichtensignal A_S im Basisband, hier dargestellt als logische Signalfolge und erzeugt durch bitweise Verknüpfung der Zeilen a) und b): $A_S = A_N \otimes A_C$ (siehe dazu Tabelle 2.25).
- Zeile d): Diese Zeile entspricht der Zeile c), jedoch ist das gespreizte Nachrichtensignal A_S als binäre Signalfolge dargestellt.
- Zeile e): Darstellung der zu spreizenden Nachrichtenbitfolge $B_N = [1,0,0,0,1]$, hier kombiniert als logische und binäre Signalfolge.
- Zeile f): Darstellung der Spreizsequenz B_C , hier kombiniert dargestellt als logische und binäre Signalfolge.
- Zeile g): Gespreiztes Nachrichtensignal B_N im Basisband, hier dargestellt als logische Signalfolge und erzeugt durch bitweise Verknüpfung der Zeilen e) und f): $B_S = B_N \otimes B_C$ (siehe dazu Tabelle 2.25).

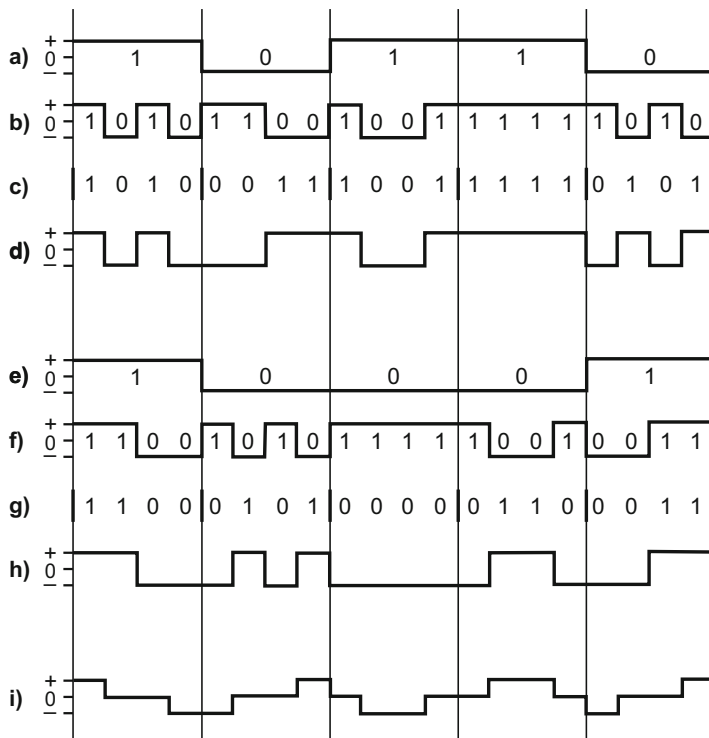


Abbildung 2.83: Beispiel für Codemultiplex

- Zeile h): Diese Zeile entspricht der Zeile g), jedoch ist das gespreizte Nachrichtensignal B_S als binäre Signalfolge dargestellt.
- Zeile i): Summensignal (siehe Tabelle 2.21) aus den Zeilen d) und h), welches bei einem Doppelempfang der Aussendungen der gespreizten Signalfolgen A_S und B_S am Empfängereingang anliegt ($[A_N \otimes A_C] \oplus [B_N \otimes B_C]$).

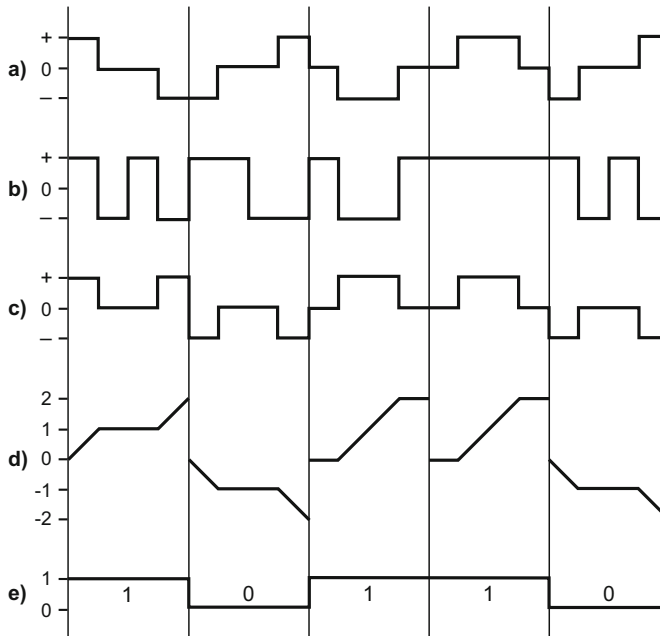
\oplus	+	-
+	+	0
-	0	-

Tabelle 2.21: \oplus -Funktion für die Addition binärer Pegel

Beispiel: Decodierung der Nachrichtensequenz aus einem Codemultiplexsignal. Beim Empfänger muss die Pseudozufallsfolge bekannt sein, damit er eine erfolgreiche De-Spreizung durchführen kann. In den Abbildungen 2.84 und 2.85 wird gezeigt, wie aus dem Summensignal $[A_N \otimes A_C] \oplus [B_N \otimes B_C]$ (Abbildung 2.83, Zeile i) durch Verknüpfung mit der entsprechenden Spreizsequenzen die Nachrichtenbitfolgen A_N (siehe Abbildung 2.83) und B_N (Abbildung 2.85) decodiert werden.

Erläuterung

- Zeile a) Summensignal, welches bei einem Doppelempfang der Aussendungen der gespreizten Signalfolgen A_S und B_S am Empfängereingang anliegt: $([A_N \otimes A_C] \oplus [B_N \otimes B_C])$.
- Zeile b): Darstellung der Spreizsequenz A_C als binäre Signalfolge.
- Zeile c): Verknüpfung des Summensignals aus Zeile a) mit der Spreizsequenz A_C aus Zeile b). Hier kommt die \otimes -Funktion (Tabelle 2.22) zur Anwendung, was einer Multiplikation entspricht (Vorzeichen beachten!).
- Zeile d): Das Signal aus Zeile c) wird bitweise integriert (kohärente Demodulation). Am Ende des Integrationsintervalls erfolgt die Bewertung, d. h. die Zuordnung zu einem logischen Pegel: $<1 \rightarrow$ „logisch Eins“ und $>1 \rightarrow$ „logisch Null“.
- Zeile e): De-spreiztes Nachrichtensignal A_N im Basisband, hier kombiniert dargestellt als logische und binäre Signalfolge. Die Folge $A_N = [1, 0, 1, 1, 0]$ ist identisch mit der Folge in Zeile a) in Abbildung 2.83.

Abbildung 2.84: De-Spreizung der Nachricht A_N **Erläuterung**

- Zeile a): Summensignal, welches bei einem Doppelempfang der Aus-sendungen der gespreizten Signalfolgen A_S und B_S am Emp-fängereingang anliegt: $([A_N \otimes A_C] \oplus [B_N \otimes B_C])$.
- Zeile b): Darstellung der Spreizsequenz B_C als binäre Signalfolge.
- Zeile c): Verknüpfung des Summensignals aus Zeile a) mit der Spreiz-sequenz B_C aus Zeile b). Hier kommt die \otimes -Funktion (Tabelle 2.22) zur Anwendung, was einer Multiplikation entspricht (Vor-zeichen beachten!).
- Zeile d): Das Signal aus Zeile c) wird bitweise integriert (kohärente De-modulation). Am Ende des Integrationsintervalls erfolgt die Be-wertung, d. h. die Zuordnung zu einem logischen Pegel: $(< 1) \rightarrow$ „logisch Eins“ und $(> 1) \rightarrow$ „logisch Null“.
- Zeile e): De-spreiztes Nachrichtensignal B_N im Basisband, hier kombi-niert dargestellt als logische und binäre Signalfolge. Die Folge $B_N = [1, 0, 0, 0, 1]$ ist identisch mit der Folge in Zeile e) in Abbildung 2.83.

\otimes	+	0	-
+	+	0	-
0	0	0	0
-	-	0	+

Tabelle 2.22: \otimes -Funktion für analoge Pegel (Vorzeichenregel der Multiplikation, entspricht der Tabelle 2.21, jedoch ergänzt um den Eingangswert "0")

Anforderung an die Modulation beim CDMA-Verfahren

Während es bei der direkten Spreizspektrumtechnik (Abschnitt 2.3.5.2, DSSS) nur darum geht, das Basisbandsignal mit einem Pseudozufallscode zu spreizen und auf der Empfangsseite die De-Spreizung durchzuführen, sind beim Code-Division-Multiple-Access-Verfahren die Anforderungen höher. Eine erfolgreiche De-Spreizung muss auch für mehrere gleichzeitig im gleichen Frequenzband empfangene Nachrichten durchführbar sein. In den Beispielen der Abbildungen 2.84 und 2.85 wurde dies für die Nachrichtensignale A_N und B_N beispielhaft gezeigt. Die Rückgewinnung der Basisbandsignale A_N und B_N durch den Vorgang der De-Spreizung ist nur möglich, wenn einige Voraussetzungen an das eingesetzte Modulationsverfahren, an die Empfangspegel der empfangenen Signale und an die verwendeten Spreizsequenzen erfüllt sind.

Additive Überlagerung

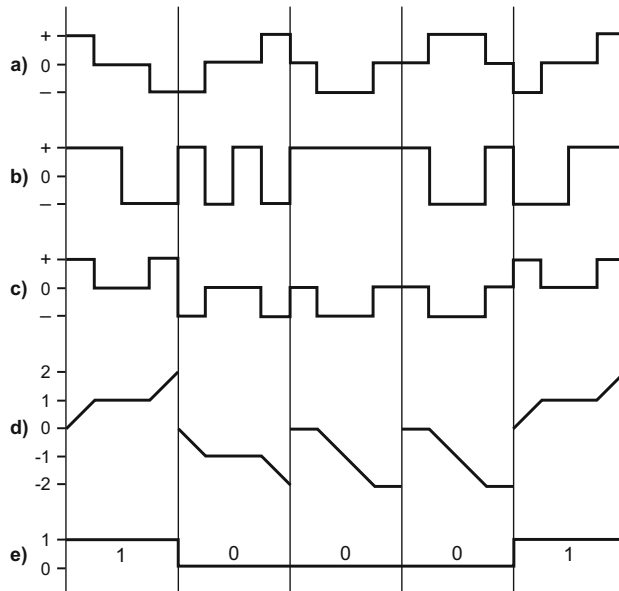
Eine Voraussetzung für die Dekodierbarkeit eines CDMA-Signals bei Mehrfachempfang ist, dass sich verschiedene Spreizspektrumsignale additiv überlagern und in dieser Form auch im Basisbandbereich zur Verfügung stehen.

Die tatsächliche Überlagerung der Signalpegel erfolgt auf dem Übertragungsweg, also im Hochfrequenzbereich. Es dürfen deshalb nur Modulationsverfahren zum Einsatz kommen, bei denen eine Addition zweier modulierter Träger im HF-Bereich auch eine Addition der Nutzsignale zur Folge hat. Nach der Demodulation müssen die Nutzsignale im Basisband additiv überlagert zur Verfügung stehen. Dies ist zum Beispiel bei der BPSK-Modulation der Fall.

Leistungsregelung

In CDMA-Systemen erfolgt nach der Integration die Bewertung des analogen Signals und damit die Zuordnung zu einem logischen Pegel (Zeile d) in den Abbildungen 2.84 und 2.85.

Für eine korrekte Bewertung ist es erforderlich, dass alle zeitgleich im gleichen Frequenzband ausgestrahlten Spreizspektrumsignale am Antenneneingang des Empfängers mit dem gleichen Pegel zu empfangen sind. Nur dann kann die Ab-

Abbildung 2.85: De-Spreizung der Nachricht B_N

bildung vom analogen Signal auf ein logisches Signal fehlerfrei erfolgen. Ist das nicht der Fall, kann dies die Teilnehmerkapazität, also die Anzahl der gleichzeitig in einem Frequenzbereich arbeitenden und dekodierbaren Stationen, erheblich mindern.

Erreicht wird dies durch eine genaue Leistungsregelung, d. h. die Empfangsstation teilt den Sendern mit, wie stark sie senden dürfen. Bei einem Mobilfunknetz gibt daher die Feststation vor, wie stark eine Mobilstation senden darf.

Anforderung an die Spreizsequenzen

Damit aus gleichzeitig empfangenen CDMA-Signalen die ursprünglichen Informationen zurückgewonnen werden können, müssen die zum Spreizen eingesetzten Pseudozufallssequenzen zueinander orthogonal sein.

Definition: **Orthogonale Vektoren**

Zwei Vektoren $[A]$ und $[B]$ mit $[A] = (a_1, a_2, a_3, \dots, a_n)$ und $[B] = (b_1, b_2, b_3, \dots, b_n)$ sind zueinander orthogonal, wenn ihr Skalarprodukt gleich Null ist, wenn $[A] \neq [B]$. Geometrisch betrachtet stehen die Vektoren, deren Skalarprodukt gleich Null ist, senkrecht aufeinander.

Definition: Skalarprodukt

Das Skalarprodukt ist die Summe aus den Produkten der korrespondierenden Komponenten jedes Vektors.

Beispiel: Wenn $A = (a_1, a_2, a_3, \dots, a_n)$ und $B = (b_1, b_2, b_3, \dots, b_n)$, ist das Skalarprodukt S aus $[A]$ und $[B]$: $S = (a_1 \cdot b_1 + a_2 \cdot b_2 + a_3 \cdot b_3 + \dots a_n \cdot b_n)$.

Walsh-Hadamard

Erzeugung von orthogonalen Spreizsequenzen

Es ist relativ einfach, mit Hilfe der diskreten Walsh-Hadamard-Transformation zueinander orthogonale Spreizsequenzen zu erzeugen.

Ausgehend von der Matrix H_2 für die Transformation zweiter Ordnung

$$H_2 = 2^{-0,5} \cdot \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix} \equiv \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

kann durch rekursives Ersetzen eine neue Matrix vierter Ordnung erzeugt werden (Der Normierungsfaktor $2^{(-1/N)}$ mit N = Ordnung der Matrix wurde hier weggelassen.).

$$H_4 = \begin{bmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & +1 \end{bmatrix} \equiv \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

Die Spalten- oder Zeilenvektoren einer Hadamard-Matrix sind zueinander orthogonal. Die Zuordnung der Matrixwerte zu den logischen Werten erfolgt durch die Beziehungen $(-1) \equiv „0“$ und $(+1) \equiv „1“$.

Beispiel: Die Sequenz in der dritten Zeile $[A] = (+1 +1 -1 -1)$ und die Sequenz in der vierten Zeile $[B] = (+1 -1 -1 +1)$ sind zueinander orthogonal, denn das Skalarprodukt ist Null.

$$[A] \cdot [B] = ((+1) \cdot (+1) + (+1) \cdot (-1) + (-1) \cdot (-1) + (-1) \cdot (+1)) \\ = ((+1) + (-1) + (+1) + (-1)) = 0$$

Gold-Codes

In der Erzeugung vorteilhafter Spreizsequenzen, die orthogonal sind und auch den Bedingungen eines Zufallscodes genügen, liegt eine Menge Wissen, vor allem, wenn es darum geht, ein Signal für Dritte „unentschlüsselbar“ zu machen. Oft kommt hier ein Code aus der Familie der „Gold“-Codes zum Einsatz, der aus der Modulo-2-Addition mehrerer Pseudozufallscodes erzeugt wird.

Beispiel für die De-Spreizung beim Mehrfachempfang

Für das folgende Beispiel wird zur Vereinfachung nur von einem Spreizfaktor $m = 4$ ausgegangen, das heißt, ein Bit des Nutzsignals wird um den Faktor vier gespreizt. Für eine Spreizung sollen die zueinander orthogonalen Chip-Sequenzen $[R]$ und $[S]$ verwendet werden mit

$$[R] = (+1 +1 -1 -1) \text{ und}$$

$$[S] = (+1 -1 -1 +1).$$

Wegen der Orthogonalität von $[R]$ und $[S]$ gilt (wenn $[R] \neq [S]$):

$$[S] \cdot [S] = 0; [S] \cdot [S] = „1“$$

$$[R] \cdot [S] = 0; [S] \cdot [S] = „1“$$

Zur Übertragung einer logischen „1“ wird vom Sender A die Chipsequenz $[R]$ und vom Sender B die Chipsequenz $[S]$ ausgesandt. Zur Übertragung einer logischen „0“ wird vom Sender A die negierte Chipsequenz $[R]$ und vom Sender B die negierte Chipsequenz $[S]$ ausgesandt. Damit erhält man Tabelle 2.23.

	Ausgesandte Chipsequenz
Sender A sendet eine „1“	(+1 +1 -1 -1)
Sender A sendet eine „0“	(-1 -1 +1 +1)
Sender B sendet eine „1“	(+1 -1 -1 +1)
Sender B sendet eine „0“	(-1 +1 +1 -1)

Tabelle 2.23: Beispiel De-Spreizung I

Zur De-Spreizung auf der Empfängerseite muss das Skalarprodukt aus der empfangenen Chip-Sequenz $[E]$ und der entsprechenden Spreizsequenz $[R]$ oder $[S]$ gebildet werden. Das Skalarprodukt wird anschließend durch den Spreizfaktor m dividiert, um den ursprünglich ausgesandten logischen Wert zu erhalten. (-1) entspricht dann einer logischen „0“ und $(+1)$ einer logischen „1“ bzw.

$$([E] \cdot [S]) / m > 1 \equiv \text{log. Wert} = 1$$

$$([E] \cdot [S]) / m < 1 \equiv \text{log. Wert} = 0$$

Wird zur De-Spreizung nicht die passende Spreizsequenz verwendet, hat das Skalarprodukt wegen der Orthogonalität den Wert 0.

Die Empfangssequenz E ist abhängig von den verschiedenen ausgesandten Bits der Sender A und B. Wegen der additiven Überlagerung addieren sich die gleichzeitig am Empfängereingang anliegenden Signale linear (siehe Tabelle 2.24).

Wie oben bereits gezeigt, muss zur De-Spreizung der Aussendung von A der

A sendet	B sendet	Ausgesandte Chipsequenz
0	0	$[E_{00}] = (-2 \ 0 \ +2 \ 0)$
0	1	$[E_{01}] = (0 \ -2 \ 0 \ +2)$
1	0	$[E_{10}] = (0 \ +2 \ 0 \ -2)$
1	1	$[E_{11}] = (+2 \ 0 \ -2 \ 0)$

Tabelle 2.24: Beispiel De-Spreizung II

Empfänger die empfangene Sequenz $[E_{XX}]$ mit der Chip-Sequenz $[R]$ multiplizieren. (Entsprechend muss zur De-Spreizung von B die Multiplikation mit der Spreiz-Sequenz $[S]$ erfolgen.)

Zur De-Spreizung von A aus $[E_{00}]$ erhält man:

$$\begin{aligned}[E_{00}] \cdot [R] &= (\overline{A} + \overline{B}) \cdot [R] \\ &= [\overline{A}] \cdot [R] + [\overline{B}] \cdot [R]\end{aligned}$$

Das Skalarprodukt $[\overline{B}] \cdot [R]$ ergibt wegen der Orthogonalität den Wert Null. Daher ist

$$\begin{aligned}[E_{00}] \cdot [R] &= (-2 \ 0 \ +2 \ 0) \cdot (+1 \ +1 \ -1 \ -1) \\ &= (-2 \ 0 \ -2 \ 0) = (-4)\end{aligned}$$

Den logischen Wert „0“ erhält man durch die Division des Ergebnisses durch den Spreizfaktor $m = 4$ und die Zuordnung $(-1) \equiv „0“$ und $(+1) \equiv „1“$.

Entsprechend gilt für die De-Spreizung von A aus $[E_{01}]$:

$$\begin{aligned}[E_{01}] \cdot [R] &= (\overline{A} + B) \cdot [R] \\ &= [\overline{A}] \cdot [R] + [B] \cdot [R]\end{aligned}$$

Das Skalarprodukt $[B] \cdot [R]$ ergibt wegen der Orthogonalität den Wert Null. Daher ist

$$\begin{aligned}[E_{01}] \cdot [R] &= (0 \ -2 \ 0 \ +2) \cdot (+1 \ +1 \ -1 \ -1) \\ &= (0 \ -2 \ 0 \ -2) = (-4)\end{aligned}$$

Den logischen Wert „0“ erhält man wieder durch die Division des Ergebnisses durch den Spreizfaktor $m = 4$ und die Zuordnung $(-1) \equiv „0“$ und $(+1) \equiv „1“$.

Abschließend soll die De-Spreizung von A aus $[E_{11}]$ gezeigt werden:

$$\begin{aligned}[E_{11}] \cdot [R] &= (A + B) \cdot [R] \\ &= [A] \cdot [R] + [B] \cdot [R]\end{aligned}$$

Das Skalarprodukt $[B] \cdot [R]$ ergibt wegen der Orthogonalität den Wert Null. Daher ist

$$\begin{aligned}[E_{11}] \cdot [R] &= (+2 \ 0 \ -2 \ 0) \cdot (+1 \ +1 \ -1 \ -1) \\ &= (+2 \ 0 \ +2 \ 0) = (+4)\end{aligned}$$

Den logischen Wert „1“ erhält man auch hier durch die Division des Ergebnisses durch den Spreizfaktor $m = 4$ und die Zuordnung $(-1) \equiv „0“$ und $(+1) \equiv „1“$.

Orthogonal Frequency Division Multiplexing

In Abbildung 2.86 ist die Situation beim Mehrwegeempfang dargestellt. Er kann Eigenstörungen verursachen. Die zum Empfang nutzbare Symboldauer wird unter Umständen drastisch verkürzt.

In Abbildung 2.86 ist in der unteren Zeile die zeitliche Abfolge der (ungestört) empfangenen Symbole 1 bis 3 dargestellt. In der mittleren Zeile sieht man den Verlauf des reflektierten, um die Zeit T_b verzögerten Signals (b).

Das erste Drittel (T_b) des direkt empfangenen Symbols 2 wird durch das zeitgleich empfangene letzte Drittel des (verzögerten) Symbols 1 gestört. Ungestört

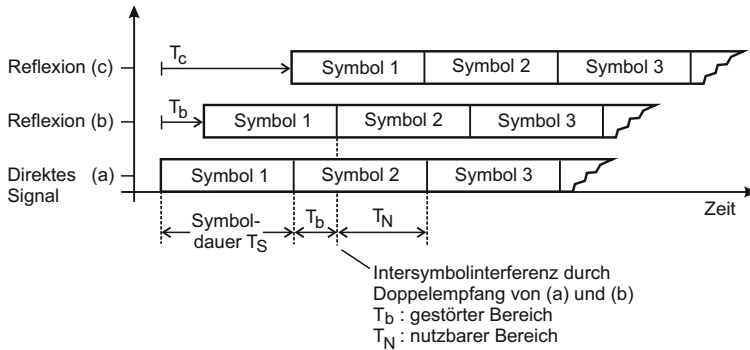


Abbildung 2.86: Intersymbolinterferenzen durch Mehrwegeempfang

ist das Symbol 2 daher nur während der Zeit T_N empfangbar. Die tatsächlich nutzbare Symboldauer ist vom Umweg des reflektierten Signals (b) abhängig. Die nutzbare Symboldauer T_N ist um die Dauer der Verzögerungszeit T_b gekürzt ($T_N = T_S - T_b$).

Im Falle des gleichzeitigen Empfangs des direkten (a) und des um T_c verzögerten Signals (c), ist kein ungestörter Empfang mehr möglich, denn die Verzögerungszeit T_c ist gerade so groß wie die Symboldauer T_S .

Eine Lösung dieses Problems ergibt sich durch die Einfügung von Schutzintervallen¹⁷⁷ zwischen den Symbolen (siehe Abbildung 2.87, untere Zeile: GI-Bereiche). Schutzintervalle sind Zeiten, in denen keine Aussendungen erfolgen. Ein Schutzintervall erhöht die effektive Symboldauer um die Dauer des Schutzintervalls T_G . Seine Dauer wird durch die maximal zulässige (Um-)Weglänge eines reflektierten Signals und damit durch dessen maximale Verzögerungszeit bestimmt. Da dies ist ein Teil der Systemspezifikationen des Kommunikationssystems ist, kann der Wert variiert werden.

In Abbildung 2.87 ist leicht erkennbar, dass gleichzeitiger Empfang des direkten Signals (a) und des reflektierten Signals (b) nicht zu Störungen führt. Am Beispiel des Symbols 2 kann man sehen, dass der Beginn des Symbols 2 sich mit dem Schutzintervall am Ende des Symbols 1 im reflektierten Signalpfad (b) überlappt. Die effektiv nutzbare Symboldauer erhöht sich in diesem Fall sogar, da sich in den Pfaden (a) und (b) entweder ein Schutzintervall und das Symbol 2 überlappen oder über beide Pfade das Symbol 2 empfangen wird. Der kritische, zu Störungen führende Fall, die Überlappung unterschiedlicher Symbole, ist durch die Schutzintervalle ausgeschlossen worden. Ob die Eigenstörungen, also der doppelte Empfang des Symbols 2 zu Störungen führt, hängt von der eingesetzten Modulationsart ab.

¹⁷⁷engl.: Guard Interval, GI

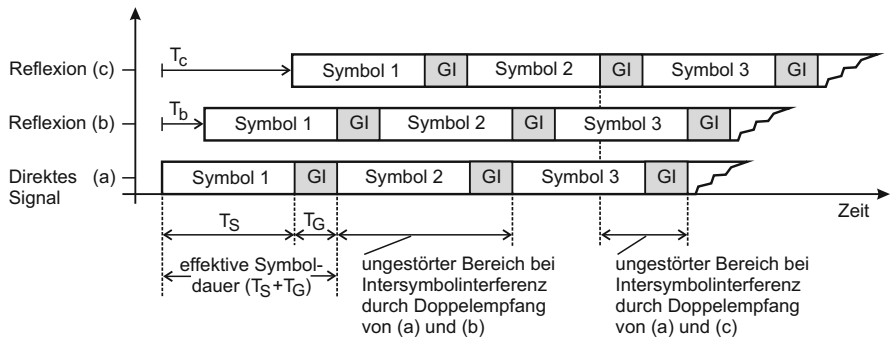


Abbildung 2.87: Intersymbolinterferenzen durch Mehrwegeempfang bei Signalen mit Schutzintervall

Selbst im Falle des gleichzeitigen Empfangs des direkten Signals (a) und des um eine komplette Symboldauer ($T_C = T_S$) verzögerten reflektierten Signals (c) ist noch ein Bereich mit ungestörtem Empfang möglich, wie dem obigen Bild am Beispiel des Symbols 3 zu entnehmen ist.

Da die Dauer eines Schutzintervalls ungenutzte Übertragungszeit und damit ungenutzte Kanalkapazität bedeutet, sollte sie möglichst klein gehalten werden. Andererseits soll die Schutzintervallsdauer möglichst groß sein, um auch die Auswirkungen größerer Verzögerungszeiten bzw. Umweglängen eines reflektierten Signals kompensieren zu können.

Die maximal zulässige Umweglänge ist, wie oben bereits erwähnt, durch die Systemspezifikationen vorgegeben und kann nicht verändert werden. Um dennoch den Verlust an Kanalkapazität klein zu halten, wird der Nutzinformationsfluss in mehrere Subkanäle mit deutlich höherer Symboldauer unterteilt.

Beispiel: Es soll die Informationsmenge von 10.000 Bits übertragen werden.

Es steht eine binäre Datenstrecke (Bitdauer = Symboldauer) mit einer Übertragungsrate von 1000 Baud zur Verfügung. Die Dauer eines Datenbits beträgt dann $T_{Bit} = 1/1000$ s. Die Übertragungsdauer von 10.000 Bits beträgt somit $10.000/1000 = 10$ Sekunden. Die maximal auftretende Frequenz des Rechtecksignals (also einer 0-1-Datenfolge) bestimmt die notwendige Kanalbandbreite. Sie ermittelt sich aus der kürzest möglichen Periodendauer $T_P = 2 T_{Bit}$ mit $B = 1/(2 T_{Bit})$ als $B = 500$ Hz.

Für die Übertragung von 10.000 Bits benötigt man also 10 Sekunden bei einer Bandbreite von 500 Hz.

Beispiel: Es soll die Informationsmenge von 10.000 Bits übertragen werden. Steht eine binäre Datenstrecke mit einer Übertragungsrate von 100 Baud zur Verfügung, ergibt sich nach obiger Rechnung für die Übertragung von 10.000 Bits eine Übertragungsdauer von 100 Sekunden bei einer Bandbreite von 50 Hz.

Das Produkt aus Übertragungsdauer und benötigter Übertragungsbandbreite ist bei einer vorgegebenen Informationsmenge konstant. In Abbildung 2.88 ist das in den Skizzen

- a): kurze Symboldauer T_S bei großer Bandbreite B_S und
 b): lange Symboldauer T_S bei kleiner Bandbreite B_S dargestellt.

Zu a) und b)

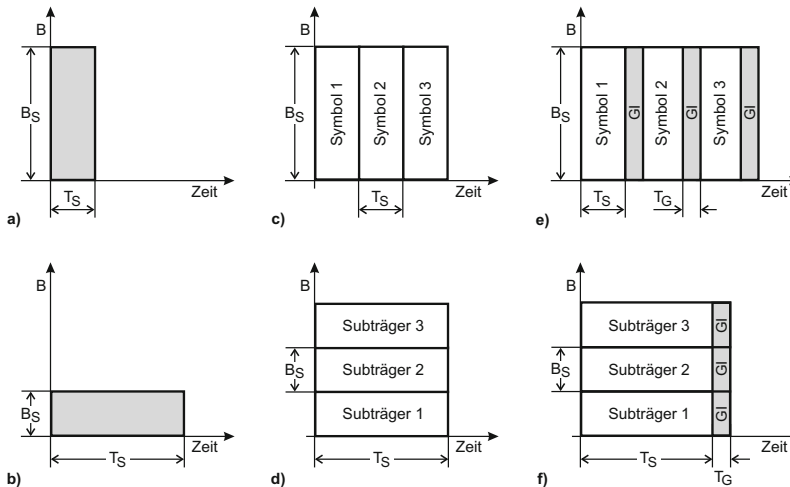


Abbildung 2.88: Auswirkungen der Schutzintervalle: T_S : Symboldauer; B_S : Signalbandbreite; G_I : Schutzintervall; T_G : Schutzintervalldauer

Im Abbildungsteil c) werden die Symbole 1 bis 3 mit jeweils kurzer Symboldauer T_S nacheinander übertragen. Die gesamte Übertragungsdauer beträgt $3 \cdot T_S$. Die kurze Symboldauer T_S bedeutet eine hohe Baud-Rate und zugleich eine große benötigte Bandbreite B_S .

Zu c)

Im Abbildungsteil d) ist der Informationsgehalt der drei Symbole aus c) auf die drei Subträger 1 bis 3 aufgeteilt. Jeder Subträger benötigt für sich eine kleinere Bandbreite als die Symbole im Fall c). Da die Baud-Rate jedes Subträgers geringer ist als bei c), muss die Symboldauer T_S entsprechend größer sein. Diese Verteilung auf drei Subträger entspricht dem Frequenzmultiplexverfahren

Zu d)

(FDMA¹⁷⁸), wie es in Abschnitt 2.3.4.3 beschrieben wurde. Jeder Sender (Subträger) beansprucht einen Teil der Gesamtübertragungsbandbreite (in Abbildung 2.81 dargestellt als Bandbreiten B_1 bis B_5 der Unterkanäle). Der Abstand der Trägerfrequenzen muss so groß sein, dass sich auf der Empfängerseite die durch die Modulation entstehenden Seitenbänder nicht überlagern und somit stören. Die gesamte Übertragungsdauer und die gesamte benötigte Bandbreite ist in den Fällen c) und d) gleich groß, denn es wird ja in beiden Fällen die gleiche Informationsmenge übertragen.

Um die Übertragung robust gegen Eigenstörungen durch Reflexionen zu machen, wird, wie oben bereits gezeigt, nach jedem Symbol ein Schutzintervall (GI) der Dauer T_G eingefügt. Diese Verlängerung der Symboldauer T_S um die Schutzintervalldauer T_G verringert die nutzbare Datenrate.

Zu e) Im Fall e) erhöht sich daher die gesamte Übertragungsdauer um die Dauer der drei eingefügten Schutzintervalle.

Zu f) Im Fall f) werden dagegen drei Symbole zeitgleich auf mehreren Übertragungsstrecken (Subträgern) übertragen. Damit muss nur ein Schutzintervall am Ende der Symbole in jedem Subkanal eingefügt werden. Die daraus resultierende Gesamtübertragungsdauer ist daher im Falle f) deutlich geringer als im Falle e).

Bandbreite Die benötigte Bandbreite für einen modulierten Subträger ist abhängig von der Bandbreite des Nutzsignals im Basisband. Bei einem nichtorthogonalen Frequenzmultiplexverfahren müssen die Frequenzen f_0 , f_1 und f_3 der Subträger (siehe Abbildung 2.89) so gewählt werden, dass sich die Seitenbänder nicht gegenseitig stören bzw., dass sie im Empfänger noch separiert und demoduliert werden können. Das bedingt Mindestschutzabstände zwischen den Subträgerfrequenzen, was eine Erhöhung der belegten Bandbreite zur Folge hat. Durch eine besondere Anordnung der Subträgerfrequenzen kann auf die Mindestschutzabstände zwischen den Subträgerfrequenzen verzichtet werden. Durch das Einfügen des Schutzintervalls haben die Subträger das Frequenzspektrum der „si“-Funktion $[\sin(x)/x]$, die in Abbildung 2.90 dargestellt ist. Werden jetzt Subträgerfrequenzen so gewählt, dass ihr Abstand voneinander gerade gleich dem Kehrwert der Symboldauer T_S ist (also gleich der reinen Nutzsignaldauer ohne Berücksichtigung der Schutzintervalldauer), so löschen sich die Ränder der „si“-Funktion gegenseitig aus. Eine solche Wahl der Subträgerfrequenzen bezeichnet man als orthogonale Anordnung, und deshalb spricht man hier auch von OFDM¹⁷⁹. In Abbildung 2.91 ist das Frequenzspektrum eines OFDM-Systems mit drei Subträgerfrequenzen dargestellt. Man erkennt, dass beim Amplitudenmaximum eines Subträgers alle anderen Subträger gerade einen Nulldurchgang haben und daher nicht stören können.

Das Spektrum eines OFDM-Signals erstreckt sich von f_0 bis $f_0 + N/T_S$, wobei

¹⁷⁸FDMA = Frequency Division Multiple Access (deutsch: Frequenzmultiplex)

¹⁷⁹OFDM = Orthogonal Frequency Division Multicarrier

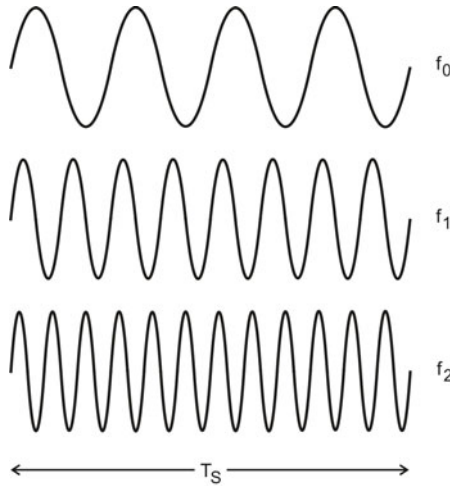


Abbildung 2.89: Subträger f_0 , f_1 und f_2 mit der Symboldauer T_S

T_S die Symboldauer ist und N die Anzahl der Subträger. Die benötigte Bandbreite B_{OFDM} eines OFDM-Signals ergibt sich daher als $B_{OFDM} = N/T_S$. Das typische Leistungsspektrum eines OFDM-Signals ist in Abbildung 2.92 dargestellt. Die einzelnen Subträger sind nicht mehr unterscheidbar. An den Rändern sind noch spektrale Anteile der „sinc“-Funktion erkennbar, die von den Subträgern an den Bandgrenzen herrühren.

2.3.5 Frequenzspreiz-Verfahren

Der folgende Abschnitt stellt verschiedene Frequenzspreiz-Verfahren wie Frequenzsprung, Spreizspektrum und Codemultiplex vor.

2.3.5.1 Frequenzsprung

Dieses Verfahren ist eine Weiterentwicklung des Frequenz-Diversity-Verfahrens, angepasst an den Digitalbetrieb. Hierbei wechselt der Sender in kurzen Abständen seine Sendefrequenz. Synchron dazu wird auf Empfangsseite die Empfangsfrequenz abgestimmt.

Die Frequenzwechselfolgen und die Umschaltzeitpunkte müssen Sender und Empfänger bekannt sein. Die Frequenzwechselfolgen basieren auf einem vorher vereinbarten (pseudo-) Zufallscode.

**Frequenz-
Diversity**

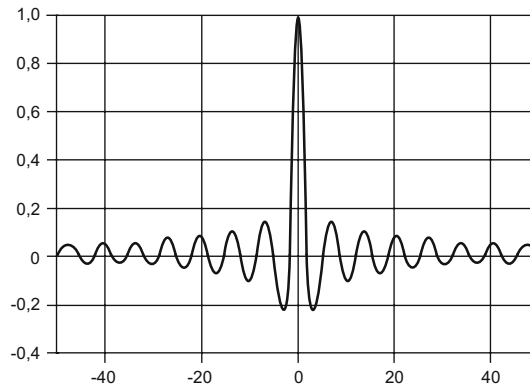


Abbildung 2.90: $\sin(x)/x$ -Funktion

Voraussetzung für den sicheren Betrieb des Frequenzsprungverfahrens (FHSS¹⁸⁰) ist eine einwandfreie Synchronisation zwischen Sender und Empfänger.

SFH

Jedes Kästchen in Abbildung 2.93 stellt die Dauer der Aussendung auf einer Frequenz (in einem Übertragungskanal) dar. Wenn während der Aussendung auf einer Frequenz (während eines Kästchens) mehrere Bits übertragen werden, spricht man vom SFH¹⁸¹ (siehe Abbildung 2.94).

FFH

Wird dagegen nur ein Bit übertragen oder erfolgt sogar während der Übertragungsdauer eines Bits bereits ein Frequenzsprung, spricht man vom FFH¹⁸². Beim Mobilfunkstandard GSM wird „Slow Frequency Hopping“ eingesetzt.

Das Frequenzsprung-Verfahren geht auf ein Patent der österreichischen Hollywood-Schauspielerinnen Hedwig Kiesler (Künstlername: Hedy Lamarr) und ihres Lebensgefährten George Antheil zurück. Das System sollte bei der Funkfernsteuerung von Torpedos zum Einsatz kommen. Die ständigen Wechsel der Übertragungsfrequenz erschweren eine gezielte Störung, wenn die Folge der Frequenzwechsel einem potentiellen Störer vorher nicht bekannt ist.

Der Schutz gegen gezielte Störungen ist nur ein Aspekt. Durch die schnellen Frequenzwechsel ist es möglich, die Auswirkungen des Fading-Effektes aufzuheben oder zumindest zu verringern. Eine Auslöschung tritt zu einer Zeit immer nur auf einer Frequenz (einem Kanal) auf. Da sich die benutzten Frequenzen (Kanäle) ständig ändern, ist der Datenverlust bei einer Signalauslöschung auf die Informationen beschränkt, die während der Benutzung des gestörten Kanals ausgesandt wurden. Ein in der Kanalcodierung implementiertes Fehlerkorrekturverfahren kann diesen Informationsverlust kompensieren.

¹⁸⁰FHSS = Frequency Hopping Spread Spectrum

¹⁸¹SFH = Slow Frequency Hopping

¹⁸²FFH = Fast Frequency Hopping

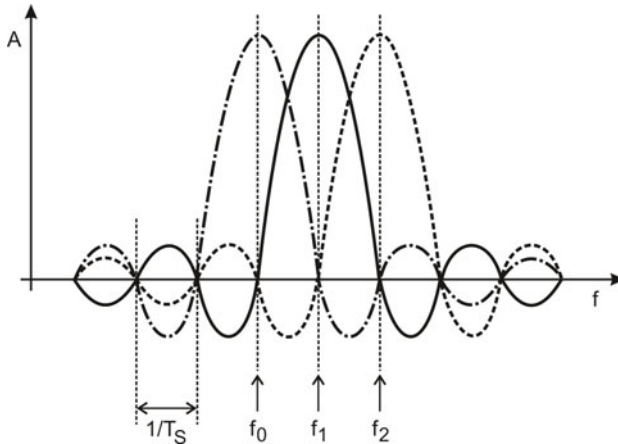


Abbildung 2.91: Anordnung der OFDM-Subträger f_0 , f_1 und f_3 im Frequenzbereich

Die Weiterentwicklung ist ein adaptives Frequenzsprung-Verfahren. Bei häufigen Störungen auf einer Frequenz (einem Kanal) kann diese bei weiteren Sprung-Sequenzen ausgeblendet werden, d. h. sie wird nicht mehr angesprungen und der Fehlerkorrekturmechanismus wird damit entlastet. Dazu bedarf es jedoch eines Rückmeldeverfahrens im Protokoll zwischen Sender und Empfänger.

Eine gezielte Störung im Sinne der Sabotage einer Kommunikationsverbindung ist beim Frequenzsprung nicht ausgeschlossen. Es muss dazu zum Beispiel ein breitbandiger Störträger (Störsender) erzeugt werden, der gleichzeitig mehrere Frequenzen (Kanäle) so stört, dass das Fehlerkorrekturverfahren an seine Grenzen stößt. Nach den geltenden fernmeldetechnischen Gesetzen in Industriestaaten ist das natürlich illegal und strafbar.

Ein weiterer Nebeneffekt ist Abhörsicherheit. Nur wenn der Mithörer die Sprungfolge kennt, ist ein Mithören ohne weiteres möglich. Dieser Aspekt tritt jedoch bei dem hier betrachteten Nahbereichs-Kommunikationssystem in den Hintergrund, denn eine Abhörsicherheit kann auch durch geeignete Kryptoverfahren sichergestellt werden.

Durch die ständigen Frequenzwechsel wird für die Übertragung ein größeres Frequenzband benötigt. Der Frequenzsprung führt zu einer Quasi-Spreizung des genutzten Übertragungsfrequenzbandes. Man bezeichnet deshalb dieses Verfahren auch als „Spread-Spectrum“-Technik, da im Vergleich zu einem Einkanalverfahren die ursprünglich benötigte Bandbreite aufgespreizt wurde (Abbildung 2.93). Die Spreizung eines modulierten Trägers kann auch anders realisiert werden; wie die Spreizspektrumtechnik dabei vorgeht, zeigt der folgende Abschnitt.

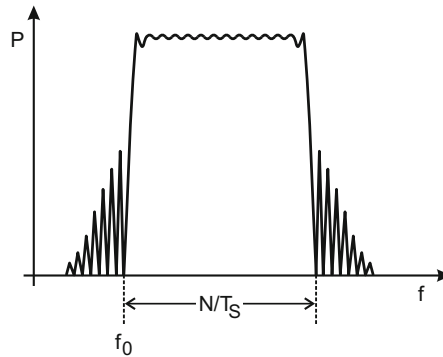


Abbildung 2.92: Leistungsspektrum der OFDM-Subträger über den Frequenzbereich f_0 bis $N \cdot 1/T_s$

2.3.5.2 Direkte Spreizspektrum

Auch bei der Direkten Spreizspektrumtechnik oder DSSS¹⁸³ wird die zu übertragende Information auf ein breites Frequenzband (Übertragungskanal) gespreizt. Dies geschieht jedoch nicht durch einen Wechsel der Sendefrequenz, sondern durch Spreizung des Basisbandsignals vor der Modulation des Trägers. Ein (frequenz-) verbreitertes Basisbandsignal hat auch größere Bandbreiten des modulierten (HF-) Signals zur Folge.

Spreizung des Spektrums eines Basisbandsignals

Abbildung 2.95 zeigt den prinzipiellen Ablauf bei der Erzeugung eines Spreizspektrumsignals. Das Spektrum SN einer binären (nichtperiodischen) Datenbitfolge ist hier mit den negativen Frequenzanteilen dargestellt (siehe auch Abbildung 2.46 im Abschnitt 2.3.2.3). Diese Nachrichtenbitfolge wird mit einer Spreizsequenz gemischt (Mischstufe 1). Für eine sichere Decodierung müssen beim Empfänger nur die Amplitudenanteile innerhalb des Frequenzbereichs der Bandbreite B_N zur Verfügung stehen.

Die Spreizsequenz S_C besteht aus einem Zufallscode, der auch der Empfangsstation bekannt sein muss, um dort die Decodierung zu ermöglichen. Es können deshalb in der Praxis nur Zufallscodes zum Einsatz kommen, bei denen sich die Zufallswerte berechnen lassen. Damit sind das keine echten zufälligen Bitfolgen. Man spricht hier deshalb von einem Pseudozufallscode (PN-Sequenz¹⁸⁴). Eine

Mischstufe 1

**Pseudo-
zufallscode**

¹⁸³DSSS = Direct Sequence Spread Spectrum

¹⁸⁴PN = Pseudo Random Noise

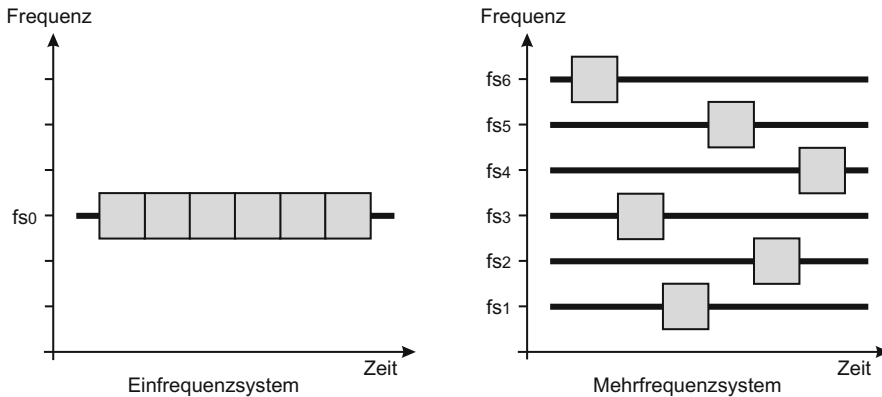


Abbildung 2.93: Gegenüberstellung von Einfrequenzsystem (links) und Mehrfrequenzsystem (rechts)

PN-Sequenz hat alle Eigenschaften eines echten Zufallscodes, jedoch lässt sich aus einem bekannten Wert (N) der nächste Wert ($N+1$) berechnen.

Beim Spreizvorgang werden Teile der Pseudozufallsfolge in Abhängigkeit von den zu übertragenden Datenbits invertiert. Die Pseudozufallsfolge bleibt deshalb unverändert eine Pseudozufallsfolge. Wenn daher ein konventioneller Empfänger die Pseudozufallsfolge empfängt, wird vor und nach der Modulation mit der Datenfolge nur ein Rauschen empfangen. Auch die Bandbreite des Spektrums wird durch diese Modulation kaum verändert ($B_C = B_S$). Erst wenn dem Empfänger die Pseudozufallsfolge bekannt ist und er damit phasengenau demoduliert (De-Spreizung), kann er die ursprüngliche Datenbitfolge zurückgewinnen.

Die Schrittdauer dieser pseudozufälligen binären Datenfolge bezeichnet man auch als Chipdauer. Für eine wirksame Spreizung des Basisbandsignals werden Spreizfaktoren von 10 bis 1000 realisiert. Damit darf die Chipdauer maximal $1/10$ bis max. $1/1000$ der Schrittdauer der Nachrichtenbitfolge betragen (d. h. die Bitfolgefrequenz des Zufallscodes soll mindestens 10 bis 1000 Mal so hoch, wie die Bitfolgefrequenz der Nutzdaten (Nachricht) sein). Damit benötigt das gespreizte Signal etwa die 10- bis 1000fache Bandbreite des Basisbandsignals B_N .

Das gespreizte Basisbandsignal SS wird in einer Modulatorstufe (Mischstufe 2) auf den Träger f_T moduliert. Das so entstandene breitbandige Hochfrequenzsignal wird dann von der Sendeantenne abgestrahlt.

Im Folgenden die Legende zu den Abbildungen 2.95 und 2.96:

S_N : Amplitudenspektrum des Nachrichtensignals im Basisband

S_C : Amplitudenspektrum der Chipsequenz (Pseudozufallscode, PN-Sequenz)

Schrittdauer

Mischstufe 2

Legende

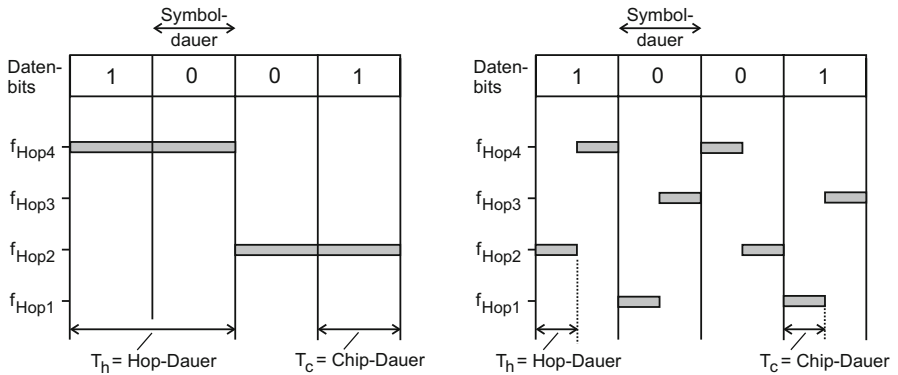


Abbildung 2.94: Slow (links) und Fast (rechts) Frequency Hopping

- S_S : Amplitudenspektrum des gespreizten Nachrichtensignals
- S_T : Spektrum des Trägers (Sinussignal mit der Frequenz f_T)
- S_A : Spektrum des von der Antenne abgestrahlten Hochfrequenzsignals
- B_N : Bandbreite des Nachrichtensignals
- B_C : Bandbreite des Spreizsignals (Chip)
- B_S : Bandbreite des gespreizten Nachrichtensignals
- B_A : Bandbreite des Hochfrequenzsignals
- f_T : Frequenz des Hochfrequenzträgers (Mittenfrequenz HF-Signals)
- f_{St} : Störträger

Komprimierung eines Spreizspektrumsignals

Im Abbildung 2.96 ist der Komprimierungsablauf auf der Empfangsseite dargestellt. Das von der Antenne empfangene Spreizspektrumsignal wird in der ersten Mischstufe (1) (Demodulator) in den Basisbandbereich (herunter) gemischt. Das so entstandene Signal S_S ist aber noch gespreizt und belegt die Bandbreite B_A .

De-Spreizung

Erst in der zweiten Mischstufe findet die Komprimierung (De-Spreizung) statt, indem es mit der Spreizsequenz S_C , die beim Sender zur Spreizung verwendet wurde, phasensynchron gemischt wird. Am Ausgang der Mischstufe (2) erhält man das übertragene Nachrichtensignal.

Beispiel: Spreizung einer Nachrichtenbitfolge. In Abbildung 2.97 ist die Spreizung der zu übertragenden Datenbitfolge 1-0-1-1 dargestellt.

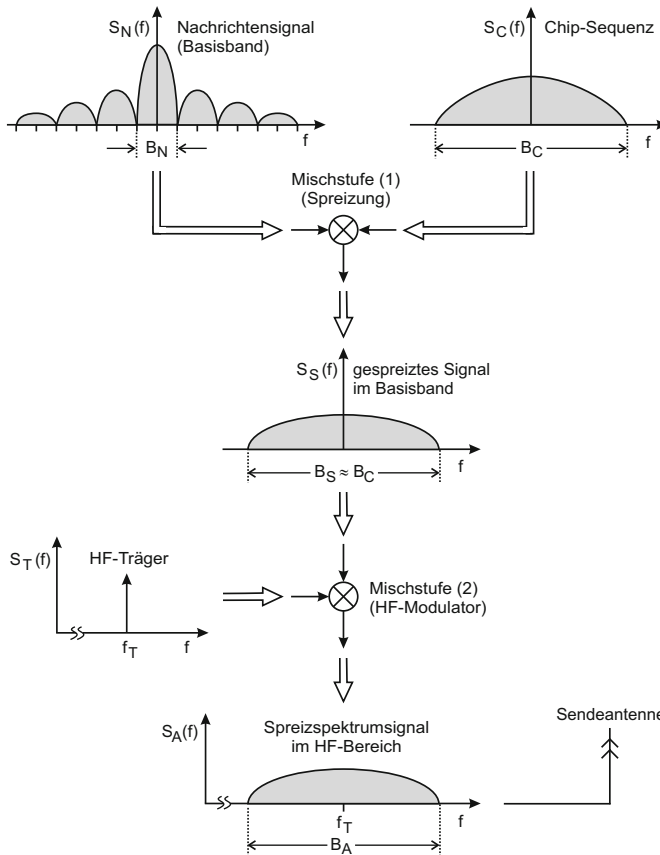


Abbildung 2.95: Übersicht über die Bandbreiten bei der Erzeugung eines Spreizspektrumsignals

Erläuterung

- Zeile a): Zu spreizende Nachrichtenbitfolge 1-0-1-1 als logische Signalfolge.
- Zeile b): Darstellung der Nachrichtenbitfolge als binäre Pegel (physikalische Pegel) „logisch Eins“ \rightarrow $[+]$ und „logisch Null“ \rightarrow $[-]$.
- Zeile c): Zufallsbitfolge als logische Signalfolge.
- Zeile d): Darstellung der Zufallsbitfolge als binäre Pegel (physikalische Pegel) „logisch Eins“ \rightarrow $[+]$ und „logisch Null“ \rightarrow $[-]$.
- Zeile e): Gespreiztes Nachrichtensignal im Basisband, hier dargestellt als logische Signalfolge und erzeugt durch bitweise Verknüpfung der Zeilen a) und c) durch die \otimes -Funktion (Äquivalenzfunktion oder „Nicht“-EX-OR-Funktion, siehe dazu Tabelle 2.25 für die logischen Signale).

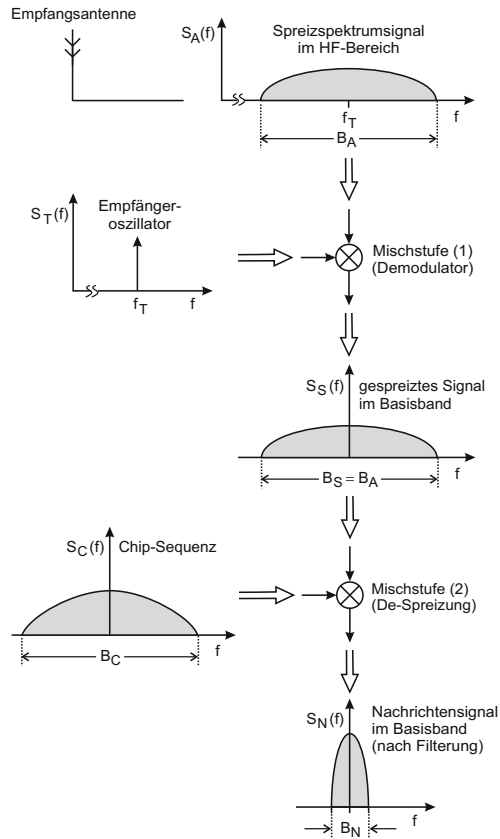


Abbildung 2.96: Übersicht über die Bandbreiten beim Empfang und der De-Spreizung eines Spreizspektrumsignals

Zeile f): Gespreiztes Nachrichtensignal im Basisband, hier dargestellt mit physikalischen Pegeln. Diese Darstellung entspricht der Darstellung von Zeile e). Erzeugt wird das Signal durch die bitweise Verknüpfung der Zeilen b) und d) mittels der \otimes -Funktion (Äquivalenz-Funktion oder „Nicht“-EX-OR-Funktion, siehe dazu Tabelle 2.26 für die physikalischen Pegel).

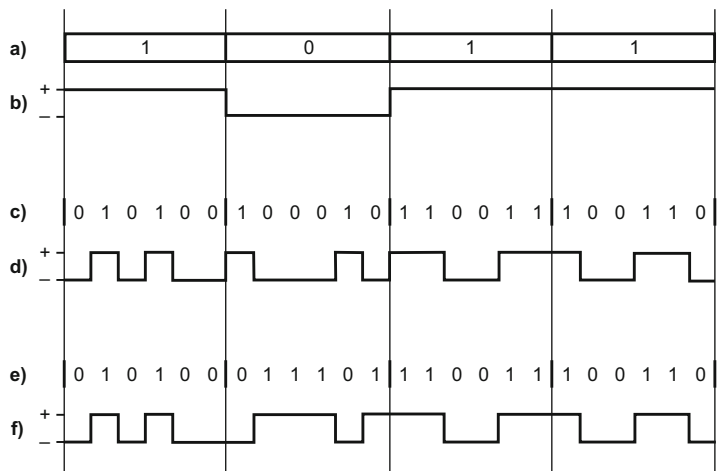


Abbildung 2.97: Erzeugung eines gespreizten Nachrichtensignals im Basisband

\otimes	0	1
0	1	0
1	0	1

Tabelle 2.25: \otimes -Funktion für logische Signale (Äquivalenz-Funktion oder „Nicht“-EX-OR-Funktion)

Beispiel: De-Spreizung (Komprimierung) eines Spreizspektrumsignals: In Abbildung 2.98 ist die De-Spreizung eines Spreizspektrumsignals dargestellt. Der Vorgang der De-Spreizung ist identisch zum Vorgang der Spreizung. Beide Abläufe setzen die Kenntnis der Pseudozufallsfolge und deren phasenrichtige Anwendung voraus. Man spricht hier auch von einer kohärenten Demodulation.

\otimes	-	+
-	+	-
+	-	+

Tabelle 2.26: \otimes -Funktion für binäre Pegel (physikalische Pegel) (entspricht den Vorzeichenregeln bei der Multiplikation)

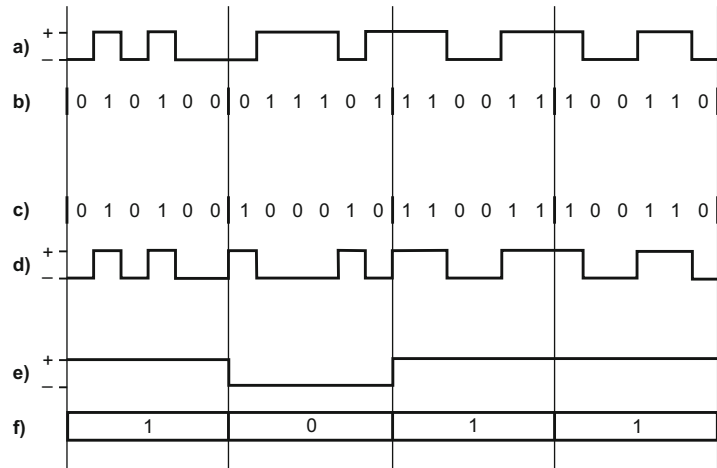


Abbildung 2.98: Rückgewinnung der Nachrichtenbitfolge aus einem Spreizspektrumsignal (De-Spreizung)

Erläuterung

- Zeile a): Zu despreizende Spreizspektrumfolge als binäre Signalfolge.
- Zeile b): Darstellung des Spreizspektrumsignal als logische Pegel [+]
→ „logisch Eins“ und [-] → „logisch Null“.
- Zeile c): Zufallsbitfolge als logische Signalfolge.
- Zeile d): Darstellung der Zufallsbitfolge als binäre Pegel (physikalische Pegel) „logisch Eins“ → [+]
und „logisch Null“ → [-].
- Zeile e): Komprimierte binäre Spreizspektrumfolge, entstanden durch die bitweise Verknüpfung der Zeilen a) und d) durch die \otimes -Funktion (siehe Tabelle 2.26 für binäre Signale). Dies ist die ursprüngliche Signalfolge aus Abbildung 2.98, Zeile b).
- Zeile f): Hier ist die Darstellung der Zeile e) als logische komprimierte Spreizspektrumfolge, entstanden durch die bitweise Verknüpfung der Zeilen b) und c) mittels der \otimes -Funktion (siehe Tabelle 2.25 für die logischen Signale). Dies ist die ursprüngliche Signalfolge aus Abbildung 2.97, Zeile a).

Vorteile

Die Spreizspektrumtechnik bietet eine Reihe von Vorteilen:

- Robustheit gegen frequenzselektives Fading
Durch die große Bandbreite kann eine Signalauslöschung (Fading) immer nur einen Teil des Spektrums betreffen. Damit weist dieses System im Vergleich zu einem Schmalbandsystem eine geringere Anfälligkeit gegen Schwunderscheinungen durch Mehrwegeempfang auf.

- **Systembedingte Chiffrierung**
Ein Spreizspektrumsignal kann nur dann entspreizt werden, wenn die Zufallsfolge (der Spreizcode) dem Empfänger bekannt ist. Damit ist eine relative hohe Hürde gegen ein Mithören geschaffen, was diese Technik z. B. für militärische Anwendungen sehr interessant macht.
- **Erschwerte Erkennung einer Nachrichtenaussendung**
Durch die Spreizung wird die Sendeenergie auf einen weiten Frequenzbereich verteilt. Durch diese geringe spektrale Leistungsdichte ist eine Aussendung ohne besondere Maßnahmen im Empfänger nicht mehr erkennbar. Man registriert nur eine (geringe) Erhöhung des Rauschpegels. Ein nicht erkennbares Signal ist auch schwerer gezielt stöbar.
- **Robustheit gegen Störträger im Übertragungsband**
Durch den Vorgang des De-Spreizens beim Empfang wird ein möglicher Störträger selbst gespreizt. Seine Sendeleistung wird auf ein breites Spektrum verteilt, während das Nutzsignal auf die kleine Bandbreite der Nachrichtbitfolge komprimiert wird. Dieses Signal hat einen deutlich höheren Signalpegel als der gespreizte Störträger, so dass dieser keine Störungen verursachen kann (siehe auch die Abbildungen 2.99 und 2.100).
- **Robust gegen Gleichkanalinterferenzen**
Durch besondere Spreizfolgen können in einem Übertragungsband zeitgleich mehrere Spreizspektrumsignale ausgestrahlt und in einem Empfänger wieder komprimiert werden (CDMA¹⁸⁵). Dieser Betrieb ist im Abschnitt 2.3.4.3 in den Abbildungen 2.83 bis 2.85 beschrieben.

Nachteilig bei der Spreizspektrumtechnik ist der hohe Aufwand für die Synchronisierung der Zufallsfolgen beim Sender und Empfänger. Bei gleichzeitigem Empfang mehrerer Spreizspektrumsignale gibt es besondere Anforderung an die Spreizfolge und an eine genaue Sendeleistungsregelung (siehe dazu Abschnitt 2.3.4.3, CDMA).

Nachteile

Abbildung 2.99 zeigt das Auftreten eines Störträgers f_{St} im Bereich des gespreizten Nachrichtenbandes. Bei der De-Spreizung (Abbildung 2.100) wird dasselbe Verfahren wie bei der Spreizung durchgeführt. Nach der Mischstufe (2) ist der Störträger f_{St} gespreizt und das Nutzsignal entspreizt. Das Nutzsignal hat durch die De-Spreizung ein höheres Leistungsspektrum als das gespreizte Störsignal. Der Signal-Rauschabstand wird durch einen Filter (Integrator bzw. Tiefpassfilter) weiter verbessert.

¹⁸⁵CDMA = Code Division Multiple Access (deutsch: Codemultiplex)

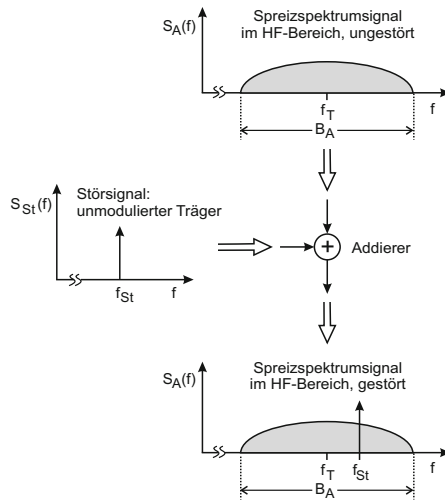


Abbildung 2.99: Störung des Spreizspektrumsignals durch einen Träger innerhalb der Übertragungsbandbreite

Beispiele:

1. FHSS setzt zum Beispiel Bluetooth ein (siehe Abschnitt 3.2.1).
2. DSSS kommt bei WLAN (siehe Abschnitt 3.2.3), UMTS, UWB, Zig-Bee (siehe Abschnitt 3.2.2) und WirelessUSB zum Einsatz.

Beispiel: Der Spreizcode (Chip-Sequenz) lautet: „1 1 0 0 0 1 1 1“. Acht Chips, das heißt ein Bit muss durch 8 Chips codiert werden. Dies geschieht mittels einer XOR-Verknüpfung. Das zu sendende Nutzsignal sei die Bitfolge „1 0“.

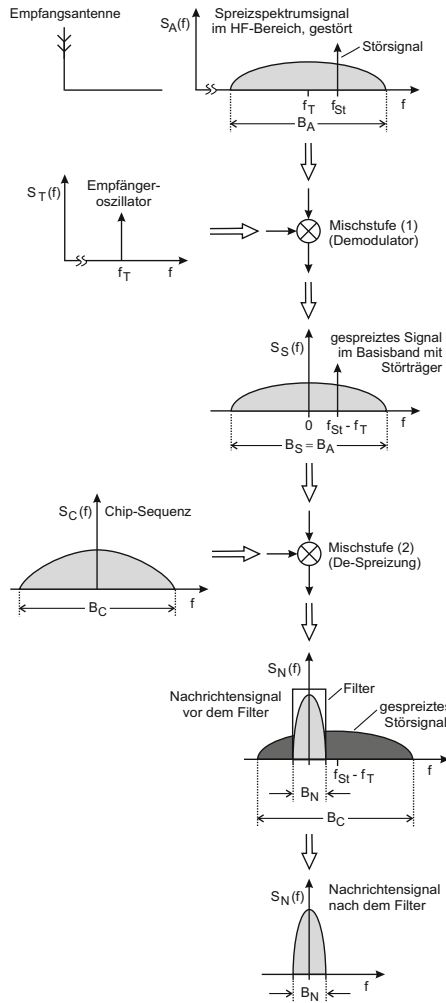


Abbildung 2.100: Ablauf bei Empfang eines gestörten Spreizspektrumsignals

Signal:	1	0
Chip-Seq.:	11000111	11000111
XOR:	00111000	11000111

Lösung

Es würde das Ergebnis der XOR-Operation mit einer nun um den Faktor acht erhöhten zu übertragenden Datenmenge übermittelt. Es ergibt sich also eine niedrigere nutzbare Datenrate auf dem Übertragungsweg.

Der Empfänger kennt die Chipfolge und ist synchron zum Sender, wodurch er die ursprünglichen Daten wieder decodieren kann [Wik08]:

Signal:	00111000	11000111
Chip-Seq.:	11000111	11000111
XOR:	11111111	00000000

Chip-Rate

Die Bezeichnung Chip-Rate wurde

in den technischen Sprachgebrauch eingeführt zur Unterscheidung zwischen den Bitraten an Ein- und Ausgang des Kanalcodierers. Die Chip-Rate ist die Bitrate am Ausgang und somit gleich der Bitrate auf dem Übertragungskanal. Für die Bitrate am Eingang wird unverändert der Begriff Bitrate benutzt. Die Chip-Rate wird in CPS¹⁸⁶ angegeben. Die Nutzbitrate ist immer kleiner als die Chip-Rate (Bitrate des Spreizcodes).

2.4 Eingebettete Systeme

Die Eingebetteten Systeme bilden die Plattform zur Implementierung eines Wireless-Netzwerkes für den Nahbereich.

2.4.1 Definition

Eingebettete Systeme

Wenn man von Computern spricht, denkt man zunächst an Geräte wie PCs, Laptops, Workstations¹⁸⁷ oder Großrechner. Aber es gibt auch noch andere, weiter verbreitete Rechnersysteme: die eingebetteten Systeme¹⁸⁸. Das sind Rechenmaschinen, die in elektrischen Geräten „eingebettet“ sind, z. B. in Kaffeemaschinen, CD-, DVD-Spielern oder Mobiltelefonen.

Unter eingebetteten Systemen verstehen wir alle Rechensysteme außer den Desktop-Computern. Verglichen mit Millionen produzierter Desktop-Systeme werden Milliarden eingebetteter Systeme pro Jahr hergestellt. Vielfach findet man bis zu 50 Geräte pro Haushalt und Automobil [VG00]. Im Folgenden gilt diese Definition:

¹⁸⁶engl.: Chips Per Second

¹⁸⁷Arbeitsplatzrechner mit hoher Rechenleistung

¹⁸⁸engl.: Embedded Systems

Definition: Eingebettete Systeme

Rechenmaschinen, die für den Anwender weitgehend unsichtbar in einem elektrischen Gerät „eingebettet“ sind.

Eingebettete Systeme weisen folgende Merkmale auf:

Merkmale

1. Ein eingebettetes System führt eine Funktion (wiederholt) aus.
2. Es gibt strenge Randbedingungen bezüglich Kosten, Energieverbrauch, Abmessungen usw.
3. Sie reagieren auf ihre Umwelt in Echtzeit¹⁸⁹.

Beispiel: Eine Digitalkamera

1. führt die Funktion „Fotografieren“ aus,
2. soll wenig Strom verbrauchen, kompakt und leicht sein und kostengünstig herzustellen sein,
3. soll das Foto innerhalb einer definierten Zeitschranke erstellen und abspeichern.

Beispiel: Die „Ulmer Zuckeruhr“ ist ein portables System zur Messung und Regelung des „Zuckers“ (Glukose) im Unterhautfettgewebe (subkutan). Die Einstellung des Blutzuckers ist essentiell bei der im Volksmund als „Zucker“ bekannten Krankheit Diabetes mellitus [Ges00].

Aufgaben:

1. Warum ist ein portabler MP3-Player ein eingebettetes System?
2. Warum ist die Elektronik einer Kaffeemaschine ein eingebettetes System?
3. Nennen Sie drei eingebettete Systeme aus Ihrem Alltagsleben.

¹⁸⁹innerhalb einer definierten Zeitschranke

2.4.2 Entwicklung

Problem

Für die Entwicklung eingebetteter Systeme gibt es – wie für jede Entwicklung – einen Anlass: Kunden haben ein Bedürfnis, z. B. nach kleinen, leistungsfähigen MP3-Spielern, das wir durch unser eingebettetes System befriedigen. Am Ausgangspunkt der Entwicklung steht die Marktnachfrage, das Problem des Kunden. Unsere Aufgabe ist es nun, das Problem des Kunden zu lösen. Um diesen Auftrag ausführen zu können, wollen wir die verschiedenen, in Abbildung 2.101 dargestellten Facetten der Entwicklung eingebetteter Systeme betrachten.

Die Abkürzungen KDS¹⁹⁰ und VDS¹⁹¹ stehen für (fest) verdrahtete und konfigurierbare digitale Schaltungen. VDS kommen vorzugsweise bei ASICs¹⁹² zum Einsatz. Die Abkürzung DS*¹⁹³ umfasst VDS und KDS.

Die Herausforderung in der Entwicklung besteht darin:

1. die relevanten Randbedingungen zu identifizieren
2. die Aufgabe oder Problemstellung zu verstehen
3. die „beste“ Lösung für die gegebene Aufgabe unter den gegebenen Randbedingungen zu erarbeiten

Randbedingungen

Das Kapitel 4 im Buch [GM07] geht ausführlich auf die Anforderungsanalyse ein. In dieser Analyse werden die funktionalen Anforderungen herausgearbeitet und die Randbedingungen an die Entwicklung des eingebetteten Systems deutlich. Solche Randbedingungen sind beispielsweise:

- technische Randbedingungen an
 - Verzögerungszeit (Latenz): Zeit zwischen dem Starten und dem Beenden einer Aufgabe¹⁹⁴
 - Datendurchsatz¹⁹⁵: verarbeitete Datenmenge pro Zeit
 - Ressourcenverbrauch: Speicher, Logikgatter, Anschlüsse¹⁹⁶ usw.
 - Energieverbrauch
 - Abmessungen und Gewicht

¹⁹⁰KDS = Konfigurierbare Digitale Schaltung

¹⁹¹VDS = Verdrahtete Digitale Schaltung

¹⁹²ASIC = Application Specific Integrated Circuit (*deutsch: Anwendungsspezifische Integrierte Schaltung*)

¹⁹³DS* = Digitale Schaltungen aus VDS und KDS

¹⁹⁴engl.: Task

¹⁹⁵Aus einer Forderung an den Datendurchsatz können sich weitere Forderungen ableiten lassen, z. B. an die Taktfrequenz oder an die Anzahl der verarbeiteten Befehle pro Sekunde.

¹⁹⁶engl.: Pins

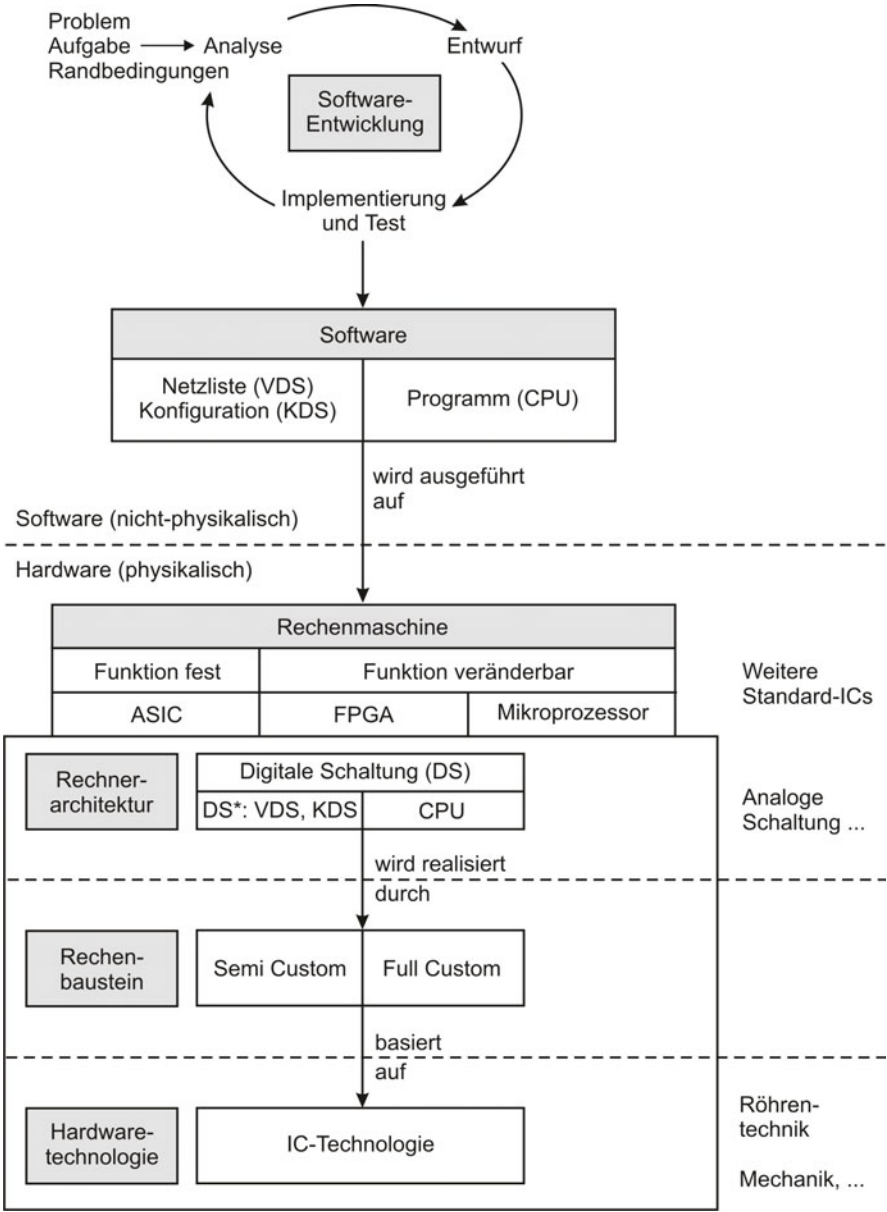


Abbildung 2.101: Entwicklung eingebetteter Systeme

- ökonomische Randbedingungen
 - einmalige Kosten der Fertigungseinrichtung oder der Entwicklung (NRE¹⁹⁷)
 - Stückkosten: laufende Fertigungskosten zur Duplizierung eines Systems
 - Entwicklungsdauer eines Prototypen
 - Entwicklungsdauer bis zur Markteinführung
- weitere Randbedingungen
 - Flexibilität: schnelle und kostengünstige Änderung der Funktionalität
 - Pflbarkeit: kostengünstige Wartung eines Gerätes
 - Zuverlässigkeit: korrekte Funktion und sicherer Betrieb
 - Verfügbarkeit: Ausführbeschränkungen
 - Markt- und Konkurrenzsituation
 - juristische Randbedingungen, Patente

Kompromiss

Die Lösung wird in der Regel ein Kompromiss sein, denn häufig gibt es technische Anforderungen, die einen Spielraum erlauben und sich gegenseitig beeinflussen. Es gilt, die priorisierten Randbedingungen gegeneinander abzuwägen. Da die unterschiedlichen Einflussfaktoren untereinander gekoppelt sein können, kann die Verbesserung einer Produkteigenschaft zur Verschlechterung einer anderen führen.

Beispiel: MP3-Spieler

1. Der MP3-Spieler soll möglichst lange spielen (mindestens 20 h).
2. Der MP3-Spieler soll möglichst leicht sein (höchstens 100 g).

Die erste Forderung könnte man erfüllen, indem man eine zusätzliche Batterie einsetzt. Dies würde jedoch das Gewicht des MP3-Spielers erhöhen und somit der zweiten Forderung entgegenwirken.

Ein ähnliches Abwägen gilt ebenfalls für die Auswahl der Subsysteme des eingebetteten Systems, z. B. für Prozessoren, Speicher, Platinen, die häufig als fertige Komponenten (COTS¹⁹⁸) gekauft und dem System hinzugefügt werden.

¹⁹⁷NRE = Non Recurring Engineering (deutsch: Einmalige Entwicklungskosten)

¹⁹⁸COTS = Commercial Off-The-Shelf (deutsch: Kommerzielle Produkte aus dem Regal)

Im Folgenden unterscheiden wir die drei in Abbildung 2.101 gezeigten Technologien, die für die Entwicklung eingebetteter Systeme eine besondere Rolle spielen:

1. Software-Entwicklung (siehe [GM07], Kapitel 4 und 5) : Analyse der Anforderungen, Entwurf der Lösung unter Berücksichtigung verschiedener Randbedingungen, Implementierung und Test der Lösung.
2. Rechnerarchitektur (siehe [GM07], Kapitel 3) : Die Software (inklusive digitaler Schaltung) wird auf einer Hardware, dem Rechenbaustein, ausgeführt. Die Rechenbausteine unterscheiden sich in ihren Rechnerarchitekturen.
3. Hardware-Technologie: Die Rechenbausteine werden auf Grundlage einer Hardware-Technologie (hier IC-Technologie) hergestellt.

Diese Technologien werden in den folgenden drei Abschnitten näher beschrieben.

2.4.2.1 Software-Entwicklung

In der Vergangenheit haben sich die Methoden in der Software-Entwicklung für Mikroprozessoren einerseits und des Schaltungsentwurfs für programmierbare Logikbausteine andererseits weitgehend unabhängig voneinander entwickelt. Dabei verbindet doch beide Disziplinen eine zentrale Aufgabe: die Lösung für ein technisches Problem zu liefern. Dazu müssen das Problem analysiert und die Lösung entworfen, implementiert und getestet werden. Anforderungsanalyse, Software-Entwurf, Testverfahren und die gesamte Vorgehensweise während der Entwicklung sind zum Großteil unabhängig davon, auf welcher Rechenmaschine die entstehende Software schließlich ausgeführt wird.

Die einzelnen Phasen der Software-Entwicklung sollten in den allermeisten Fällen iterativ-inkrementell abgearbeitet werden. Der Kreiszyklus in der Abbildung 2.101 hebt diese wichtige Forderung hervor: komplexe System entstehen nicht wasserfallartig, sondern evolutionär.

Kapitel 4 aus [GM07] behandelt den in Abbildung 2.101 unter der Rubrik „Software-Entwicklung“ dargestellten Themenkomplex anhand der Software-Entwicklung für Mikroprozessoren. Die Software-Entwicklung für digitale Schaltungen wird in Kapitel 5 aus [GM07] vorgestellt.

**Analyse,
Entwurf,
Implementierung**

**iterativ-
inkrementelle
Entwicklung**

2.4.2.2 Rechnerarchitekturen

Die Rechnerarchitektur legt den inneren Aufbau einer Rechenmaschine¹⁹⁹ fest. Eine Rechenmaschine kann eine Rechenaufgabe entweder sequentiell oder – falls die Rechenaufgabe es erlaubt – auch parallel verarbeiten (siehe Abbildung 2.102). Ein Mikroprozessor arbeitet sequentiell (CIT²⁰⁰). Er kann verschiedene Teile eines parallelisierbaren Algorithmus nicht parallel ausführen, ein programmierbarer Logikbaustein dagegen schon. Verschiedene Teile des Algorithmus werden gleichzeitig an verschiedenen Stellen des Chips ausgeführt (CIS²⁰¹).

CIT

CIS

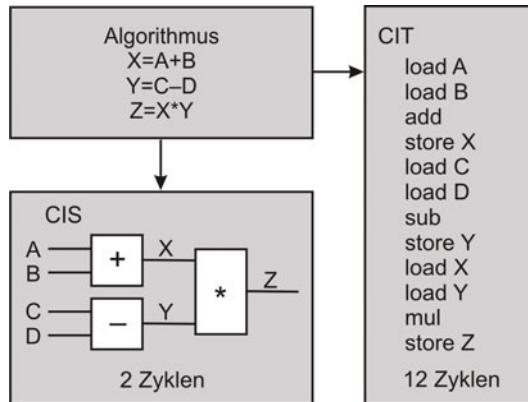


Abbildung 2.102: Ein parallelisierbarer Algorithmus kann parallel (CIS) oder sequentiell (CIT) ausgeführt werden [Rom01].

Abbildung 2.103 zeigt die drei prinzipiell möglichen Rechnerarchitekturen in Bezug auf ihre Universalität:

- CPU (CPU²⁰²): siehe CIT-Architektur in Abbildung 2.102
- Digitale Schaltung: die Architektur ist für eine Aufgabe „maßgeschneidert“ (single purpose), siehe CIS-Architektur in Abbildung 2.102.

Die in der Abbildung 2.103 durch den oberen Kreis symbolisierten Anforderungen lassen sich durch die drei Rechnerarchitekturen (CPU, KDS, VDS), die ebenfalls durch geometrische Objekte symbolisiert werden, mehr oder weniger präzise überdecken. Bei der vollständigen Abdeckung durch das Quadrat (CPU) bleibt jedoch ein Teil der Fläche des Quadrats ungenutzt – und damit ein Teil

¹⁹⁹kurz Rechner oder Prozessor

²⁰⁰CIT = Computing In Time

²⁰¹CIS = Computing In Space

²⁰²CPU = Central Processing Unit (deutsch: Zentrale Verarbeitungseinheit)

der Funktionalität der CPU. Dafür lässt sich mit dem großen Quadrat leichter eine beliebige Anforderungsfläche überdecken.

Applikationsspezifische Architekturen (ASIP²⁰³) sind eine Mischform zwischen CPU und Digitaler Schaltung. **ASIP**

Aufgabe: Welche Vor- und Nachteile hätte die CPU- und DS-Rechnerarchitektur für die Entwicklung eines MP3-Spielers?

Die verschiedenen Rechnerarchitekturen werden ausführlicher in Kapitel 3 in [GM07] miteinander verglichen.

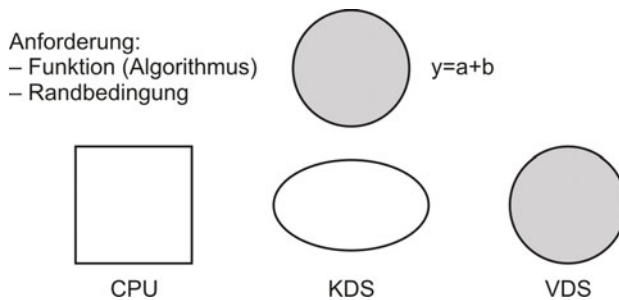


Abbildung 2.103: Geometrische Metapher für die Abdeckung von Anforderungen durch verschiedene Rechnerarchitekturen

2.4.2.3 Rechenbaustein und Hardware-Technologie

Die Transistoranzahl von integrierten Schaltungen (IC) hat sich in den letzten Jahrzehnten im Mittel alle 18 Monate verdoppelt. Diese als „Moore’s Gesetz“ bekannte Gesetzmäßigkeit hat bereits Intel-Gründer Gordon Moore im Jahr 1965 vorhergesagt: „Die Transistoranzahl von integrierten Schaltungen verdoppelt sich alle 18 Monate“. Diese Vorhersage traf in den letzten Dekaden stets zu. Abbildung 2.104 zeigt, welch kleinen Teil ein Chip aus dem Jahr 1981 auf einem Chip des Jahres 2008 einnehmen würde. Die Chip-Kapazität aus dem Jahr 2008 würde für ≈ 260.000 Chips des Jahres 1981 ausreichen.

Je nachdem, *wer* eine Schaltung *wann* auf einem Chip integriert hat, unterscheiden wir zwischen zwei IC-Technologien:

**Moore’s
Gesetz**

²⁰³ASIP = Application Specific Instruction Set Processor

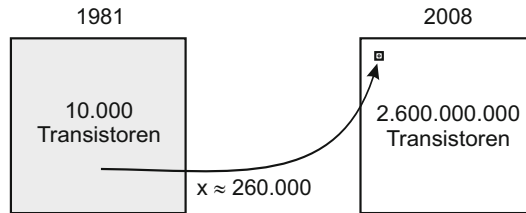


Abbildung 2.104: Moore's Gesetz: „Verdoppelung der Transistoren je Chip alle 18 Monate“ [GS98]

Full Custom

1. Voll nach Kundenwunsch (Full Custom): hierbei sind alle Verdrahtungen der Schaltung und die Konfiguration der Transistoren den spezifischen Anforderungen optimal angepasst. Diese Flexibilität bietet eine sehr gute Leistungsfähigkeit bei geringer Chipfläche und niedrigem Energieverbrauch. Die einmaligen Einrichtungskosten (Größenordnung: mehrere 100 kEuro) sind jedoch hoch, und der Markteintritt dauert lange. Anwendungsgebiete sind Applikationen mit hohen Stückzahlen (Größenordnung mehrere Hunderttausend). Beispiel: Rechenbausteine für MP3-Spieler. Die hohen Einrichtungskosten amortisieren sich durch hohe Stückzahlen und senken so die Stückkosten.

Semi Custom

2. Teilweise nach Kundenwunsch (Semi Custom): hier sind Schaltungsebenen ganz oder teilweise vorgefertigt. Dem Entwickler bleibt die Verdrahtung und teilweise die Platzierung der Schaltung auf dem Chip. Diese Bausteine bieten eine gute Leistungsfähigkeit bei kleinen Chipflächen und geringeren Entwicklungskosten als bei einem voll nach Kundenwunsch gefertigten Rechenbaustein (Größenordnung: mehrere 10 kEuro). Die Entwicklungszeit liegt jedoch immer noch im Bereich von Wochen bis Monaten.

Die IC-Technologien werden in Abschnitt 3.2 in [GM07] detaillierter dargestellt.

2.4.2.4 Rechenmaschine

Abbildung 2.105 vergrößert einen Ausschnitt aus Abbildung 2.101: die Realisierung einer Rechnerarchitektur auf Basis einer bestimmten Hardware-Technologie (hier IC-Technologie). Diese Realisierung nennen wir Rechenbaustein!

Definition: Rechenmaschine

Die Rechenmaschine besteht aus einer Rechnerarchitektur, die auf einem Rechenbaustein abgebildet wird. Der Rechenbaustein basiert auf einer Hardware-Technologie (hier IC-Technologie).

Die Darstellung konzentriert sich auf die beiden Rechenmaschinen Mikroprozessor und FPGA.

Rechnerarchitektur und Rechenmaschinen sind in der Zuordnung unabhängig voneinander. Die verschiedenen Kombinationsmöglichkeiten sind in Abbildung 2.105 gezeigt.

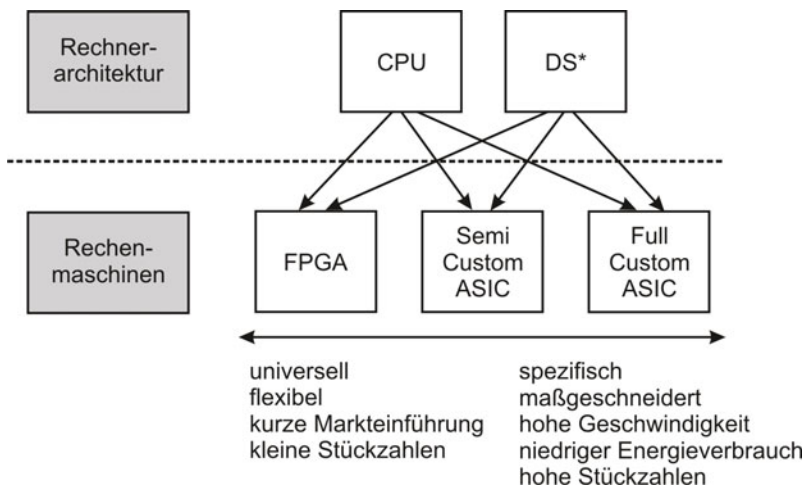


Abbildung 2.105: Die Rechnerarchitektur lässt sich auf verschiedene Rechenmaschinen abbilden.

Beispiele:

1. Eine CPU kann sowohl auf einem FPGA als eingebettetes System als auch als separater Baustein in Form eines Mikroprozessors realisiert werden (siehe „Ein-Bit-Rechner“ [Stu06], S. 15 ff).
2. Hersteller von Prozessoren entwerfen CPUs kundenspezifisch (Full Custom oder Semi Custom ASIC).

FPGA

Bei den FPGAs sind alle Schaltungsebenen vorgefertigt. Der Entwickler eines eingebetteten Systems kauft ein fertiges IC, bestehend aus Logikgattern und Kanälen zur Verdrahtung, bildet die digitale Schaltung auf die vorhandenen Logikgatter ab (Platzierung) und verbindet die Gatter untereinander. Aus Sicht des Entwicklers eines eingebetteten Systems sind die Entwicklungskosten einer FPGA-basierten Lösung gegenüber einem full custom ASIC gering, da der Baustein für den Entwickler unmittelbar verfügbar ist und er nur noch die gewünschte Schaltung aufprägen muss – eine reine Software-Tätigkeit. Diese Flexibilität und Anpassbarkeit eines FPGAs auf spezielle Anforderungen führt jedoch dazu, dass für eine konkrete Applikation in der Regel nicht alle Funktionalitäten des FPGAs genutzt werden können, ein Teil der Hardware also „verschwendet“ wird. Gegenüber einer Schaltung mit einem maßgeschneiderten Full Custom ASIC bietet ein FPGA eine geringere Verarbeitungsgeschwindigkeit bei einer höheren elektrischen Leistung. Außerdem sind die Stückkosten eines FPGAs höher (Größenordnung: mehrere 10 Euro).

Aufgabe: Wie unterscheiden sich Full und Semi Custom ASICs bezüglich Leistungsfähigkeit, einmaliger Kosten der Fertigungseinrichtung und der Entwicklungsdauer bis zur Markteinführung?

Der Trend von eingebetteten Systemen geht in Richtung „Ubiquitous Computing“ – der Allverfügbarkeit von Rechenmaschinen [Wei91]. Diese Systeme stehen in enger Verbindung zu drahtlosen Technologien. Weiterführende Literatur zu eingebetteten Systemen findet der Leser bei [GM07, VG02].

2.4.3 Mikroprozessoren

Abbildung 5.3, Kapitel 5 zeigt eine Zwei-Chip-Lösung mit Schnittstellen, bestehend aus Transceiver und Mikroprozessor. Der Mikroprozessor beinhaltet hierbei den Protokollstapel.

2.4.3.1 Grundlegende Funktionsweise

Mikroprozessoren sind ein zentrales Bauelement jedes Mikrocomputer-Systems. Der Kern eines Computers ist die zentrale Verarbeitungseinheit (siehe Abbildung 2.106). Sie besteht aus den Komponenten

- Steuerwerk,
- Rechenwerk,

- Register und
- Verbindungssystem zur Ankopplung von Speicher und Peripherie.

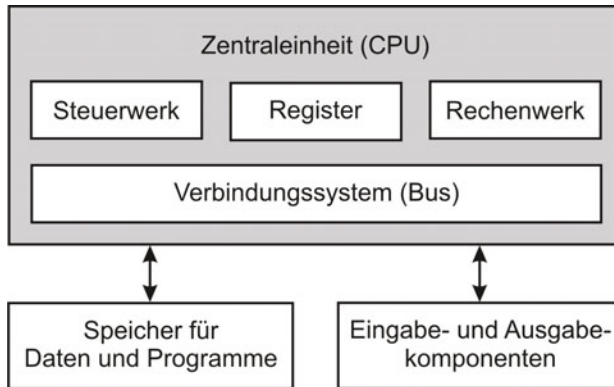


Abbildung 2.106: Prinzipieller Aufbau eines Mikroprozessors [BH01]

Der Programmspeicher (ROM²⁰⁴) sichert das Programm und der Datenspeicher (RAM²⁰⁵) die Daten. Der Zugriff erfolgt über den Adress- und Datenbus.

Zur Ankopplung der CPU an die Außenwelt dient ein Verbindungssystem. Typisch ist die Einteilung der Leitungen (ausgenommen Betriebsspannungszuführung) in ein Drei-Bus-System²⁰⁶ aus:

Bus

- Datenbus: Wortbreite des Mikroprozessors
- Adressbus: Ein Bündel von meistens 16 Leitungen bei den 8-Bit-Prozessoren oder 24 Leitungen bei 16-Bit-Prozessoren. Der Mikroprozessor gibt über diesen Bus die Adresse eines Speicherplatzes oder Ein-/Ausgaberegisters aus. Der Prozessor liest oder schreibt dann die adressierten Daten (hierzu gehört auch das Programm).
- Steuerbus: Hierunter sind alle übrigen Leitungen zusammengefasst, die der Steuerung der Peripherie dienen. Die Anzahl dieser Leitungen ist variabel. Die Leitungen sind nicht so streng parallel geordnet wie die der anderen beiden Sammelschienen. Steuerleitungen sind beispielsweise \overline{CS} („logisch 0 aktiv“)²⁰⁷, \overline{RD} ²⁰⁸ und \overline{WR} ²⁰⁹.

²⁰⁴ROM = Read Only Memory

²⁰⁵RAM = Random Access Memory

²⁰⁶Bus = Sammelschiene

²⁰⁷CS = Chip Select

²⁰⁸RD = Read (engl.:) = lesen

²⁰⁹WR = Write (engl.:) = schreiben

Mikro- prozessor- technik

Die Mikroprozessortechnik befasst sich mit der Architektur, der Entwicklung, der Implementierung, dem Bau, der Programmierung und dem Einsatz von Mikroprozessoren. Rechner oder Computer, bei denen Mikroprozessoren zum Einsatz kommen, werden als Mikrorechner oder Mikrocomputer bezeichnet.

Definition: Mikroprozessor^a

auf einem integrierten Schaltkreis (IC) realisierte Zentraleinheit

^aMP = MikroProzessor (μ P)

Definition: Mikrocontroller

Mitte der 70er Jahre gelang es, die peripheren Komponenten zusätzlich auf einem Chip zu integrieren. Es entstand der Mikrocontroller (MC^a). Mikrocontroller beinhalten ein vollständiges Computersystem (siehe 2.4.3.1 Minimalsystem) auf einem Chip. Neben den Daten- und Programmspeichern sind Peripheriekomponenten wie Ein- und Ausgabe, Zeitgeber, Analog-/Digitalwandler (ADC^b) integriert. Mikrocontroller werden z. B. bei eingebetteten Systemen zur Steuerung von Waschmaschinen oder des Motormanagements in Kraftfahrzeugen genutzt.

^aMC = MikroController

^bADC = Analog Digital Converter (deutsch: Analog-Digital-Wandler)

Im Folgenden werden elementare Begriffe definiert:

- Mikrocomputer, auch Mikrorechner genannt, sind Systeme, bestehend aus Mikroprozessoren und Peripherie.
- Die Mikroprozessortechnik befasst sich mit der Architektur, der Entwicklung, der Implementierung, dem Bau und der Programmierung von Mikroprozessoren

Minimal- system

Zu einem funktionsfähigen System gehören außer dem eigentlichen Mikroprozessor noch weitere Schaltkreise. Eine arbeitsfähige Minimalkonfiguration besteht gewöhnlich aus:

- Mikroprozessor mit Taktversorgung
- Nur-Lese-Speicher (ROM) für das Programm
- Schreib-/Lesespeicher (RAM) für variable Daten
- Ein-/Ausgabebaustein (Interface) von und zur Peripherie
- Stromversorgung

Mikro- computer

Ein derartig komplettes System nennt man einen Mikrocomputer.

Aufgaben:

1. Was ist der Unterschied zwischen Mikrocomputer und Mikroprozessor?
2. Nennen und beschreiben Sie die Funktion der Komponenten einer CPU.
3. Skizzieren Sie den prinzipiellen Aufbau eines Mikroprozessors.
4. Welche Elemente gehören zu einer Minimalkonfiguration?

Das Computerprogramm bestimmt durch eine Folge von Anweisungen oder Befehlen die Arbeitsweise der CPU und der Ein- und Ausgabekomponenten.

Programm

Aufgaben:

1. Skizzieren Sie ein Computersystem aus CPU, Speicher und Bussen.
2. Der Adressbus der CPU ist 16 Bit breit. Wie viele Adressen können damit angesprochen werden?
3. Im System werden Speicherbausteine mit jeweils 8K Speicherzellen verwendet. Wie viele Adressleitungen werden zur Auswahl der Speicherzellen benötigt?
4. Skizzieren Sie mit NAND- Bausteinen eine Dekodierschaltung für zwei Speicherbausteine (alle Steuereingänge der Speicher sind „logisch 0 aktiv“).

Das Steuerwerk koordiniert die Operationsausführung. Die Befehlsverarbeitung erfolgt in diesen Schritten:

Steuerwerk

1. Laden des Befehls in das Befehlsregister²¹⁰
2. Dekodierung des Befehls durch das Steuerwerk
3. Erzeugung von Steuersignalen für die ALU, Multiplexer, Speicher und Register

Befehls- verarbeitung

²¹⁰IR = Instruction Register (*engl.:*) = Befehlsregister

4. ALU²¹¹ verknüpft Operanden
5. Ergebnisse in Register schreiben
6. bei Lade- und Speicheroperationen Adressen erzeugen
7. Statusregister aktualisieren, die Flags für bedingte Sprünge setzen
8. Befehlszähler²¹² neu schreiben

Beispiel: Der Assemblerbefehl ADD #10, R5 steuert das Rechenwerk des MSP430, so dass zum Register R5 der Wert 10 hinzuaddiert wird.

2.4.3.2 Arten

Eingebettete Systeme

In der Vergangenheit wurden Mikroprozessoren hauptsächlich für numerische Berechnungen eingesetzt. Heute gibt es viele neue Applikationen, zum Beispiel im Bereich der Kommunikations- und Automatisierungstechnik, in Kraftfahrzeugen und bei Chipkarten. Diese eingebetteten Systeme wurden erst durch den Einsatz von hochintegrierten elektrischen Bauelementen möglich. Für bestimmte Anwendungsgebiete ist eine Spezialisierung von Mikroprozessoren erkennbar, was sich in unterschiedlichen Architekturen bemerkbar macht.

Architekturen

Die zunehmende Dezentralisierung (verteilte Systeme) der Mikrocomputeranwendungen und die Fortschritte bei der Hochintegration führten zu einer Spezialisierung der Mikroprozessoren für verschiedene Anwendungsbereiche:

Standard-prozessor

- Standardprozessoren (auch Universalprozessoren genannt, GPP²¹³) werden beispielsweise in PCs eingesetzt.

Mikro-controller

- Mikrocontroller (MC²¹⁴): Mitte der 70er Jahre gelang es, die peripheren Komponenten zusätzlich auf einem Chip zu integrieren. Es entstand der Mikrocontroller. Mikrocontroller beinhalten ein vollständiges Computersystem (siehe 2.4.3.1 Minimalsystem) auf einem Chip. Neben den Daten- und Programmspeichern sind Peripheriekomponenten wie Ein- und Ausgabe, Zeitgeber, Analog-/Digitalwandler (ADC²¹⁵) integriert. Mikrocontroller werden z.B. bei eingebetteten Systemen zur Steuerung von Waschmaschinen oder des Motormanagements in Kraftfahrzeugen genutzt.

²¹¹ALU = **A**rithmetical **L**ogical **U**nit (deutsch: *Arithmetische Logische Einheit*)

²¹²PC = Program Counter (engl.:) = Befehlszähler

²¹³GPP = General Purpose Processor (deutsch: *Universalprozessor*)

²¹⁴MC = MikroController

²¹⁵ADC = Analog Digital Converter (deutsch: *Analog-Digital-Wandler*)

- Digitale Signalprozessoren (DSP²¹⁶) sind Spezialprozessoren zur sehr schnellen Verarbeitung von mathematischen Algorithmen und kommen z.B. in der Sprach- und Bildverarbeitung zum Einsatz.
- Hochleistungsprozessoren finden beispielsweise in Großrechnern Verwendung.

Signalprozessor

Hochleistungsprozessor

Während bei Mikrocontrollern die funktionale Integration auf einem Chip im Vordergrund steht, ist bei Hochleistungsprozessoren und digitalen Signalprozessoren deren Verarbeitungsgeschwindigkeit von entscheidender Bedeutung ([BH01], S. 18 ff).

Beispiel: Mikrocontroller-Familie MSP430 von TI^a: Anwendungsgebiete sind unter anderem portable Messgeräte für Wasser, Gas, Heizung, Energie und Sensorik. Die Familie hat folgende Eigenschaften: sehr geringe Leistungsaufnahme, 0,1 μ A Power down („Schlafmodus“), 0,8 μ A Bereitschaftsbetrieb, 250 μ A/1 MIPS^b bei 3V, 4 Low power modi (Stromspar-modi), < 50 nA Port Leckstrom (Verluste), Spannungsbereich 1,8 bis 3,6 V, Temperaturbereich -40 bis +85 °C, 16-bit RISC^c, Peripheriemodule: E/A-Leitungen, Zeitgeber, LCD-Controller^d, ADC, DAC^e, USART^f usw. [Stu06]. Der Stückpreis liegt für den kleinsten Baustein <1 Euro (Abnahmemenge von 1000).

^aTI = Texas Instruments

^bMIPS = Million Instructions Per Second

^cRISC = Reduced Instruction Set Computer

^dLCD = Liquid Cristal Display

^eDAC = Digital Analog Converter (*deutsch: Digital-Analog-Wandler*)

^fUSART = Universal Synchronous Asynchronous Receiver Transmitter

Weiterführende Literatur findet man unter [Sik04b], [Ste04].

Einstieg: Starter Kit

Das eZ430 ist ein Entwicklungswerkzeug, bestehend aus USB-Stick mit IAR-Kickstart IDE (limitierte C-Entwicklungsumgebung) für circa 20 Euro [TI06].

²¹⁶DSP = Digital Signal Processor

Aufgaben:

1. Was versteht man unter Mikrocontrollern?
2. Wo werden DSPs eingesetzt?

Tabelle 2.27 zeigt die „Artenvielfalt“ eingebetteter Systeme.

	Wasch- ma- schine	Maus	Druk- ker	Handy	Key- board	Tele- fon- anlage	Auto	Werk- zeug- ma- schine
Prozes- sor	μC	ASIC	μP , ASIP	DSPs	μP , DSPs	μP , DSP	≈ 100 μC , μP , DSP	μC , ASIP
Bus [Bit]	8	k. A.	16.. 32	32	32	32	8.. 64	.. 32
Spei- cher [Bit]	<8k	<1k	1.. 64M	1.. 64M	<512M	8.. 64M	1k.. 10M	<64M
Netz- werk	k. A.	RS232	diverse Schnitt- stellen	GSM	MIDI	V.90	CAN,...	I ² C,...
Echtzeit	keine	weich	weich	hart	weich	hart	hart	hart
Zuver- lässig- keit	mittel	keine	gering	gering	gering	gering	hoch	hoch

Tabelle 2.27: Artenvielfalt eingebetteter Systeme [TAM03]. Die Abkürzung „k. A.“ steht für „keine Angabe“.

2.4.4 Energieverbrauch

Der Energieverbrauch stellt eine der wichtigen Randbedingungen bei der Entwicklung von Eingebetteten Systemen (siehe Abschnitt 2.4.2) dar. Der Abschnitt dient als Grundlage für die Kapitel 3 und 4.

Für batteriebetriebene portable Systeme ist ein niedriger Energieverbrauch besonders wichtig. Er setzt sich zum einen aus der abgestrahlten Energie der

Sendeeinheit und zum anderen aus dem Verbrauch des Eingebetteten Systems, bestehend aus Mikroprozessoren, weiteren ICs etc., zusammen. Der Energieverbrauch der Sendeeinheit geht unmittelbar in die Reichweite ein und unterliegt gegensätzlichen Anforderungen und Abhängigkeiten wie [Neu07]:

Sendeeinheit

- Zyklische/azyklische Kommunikation versus Energieverbrauch
- Echtzeitverhalten in unterschiedlichen Klassen versus Energieverbrauch
- Notwendigkeiten von redundanten Übertragungswegen versus Echtzeitverhalten

Um den Energieverbrauch Eingebetteter Systeme erklären zu können, wird im Folgenden auf die CMOS²¹⁷-Technologie näher eingegangen. Die CMOS-Technologie ist auf dem Gebiet der modernen digitalen Hochleistungselektronik von zentraler Bedeutung. Aufgrund der Eigenschaft, dass Energie nur für die Zustandsübergänge²¹⁸ benötigt wird, ist die CMOS-Technologie die erste Wahl bei ICs²¹⁹ mit hoher Packungsdichte.

Eingebettete Systeme

CMOS

Ein CMOS-Logikgatter (siehe Abbildung 2.107) arbeitet derart, dass ein „Pull-up-Netz“ aus p-Typ-Transistoren mit U_B^{220} oder ein „Pulldown-Netzwerk“ aus n-Typ-Transistoren mit Masse (Gnd^{221}) verbunden wird. Liegt der Pegel am Eingang nahe bei der Schwellspannung²²² $U_{tn,p}$, so leitet eines der beiden Netze, während das andere sperrt – es gibt keine Verbindung von U_B zur Masse. Der gesamte Energieverbrauch von CMOS-Schaltungen setzt sich zusammen aus den drei Anteilen

Funktion

Anteile Energieverbrauch

- Schaltleistung: verbrauchte Energie bei Laden oder Entladen der Ausgangskapazitäten. Der Energieverbrauch pro Zustandsübergang wird wie folgt bestimmt: $E = 1/2 \cdot C_L \cdot U_B^2 \approx 1$ Picojoule
- Kurzschlussleistung: Wenn sich die Gattereingänge in einer Zwischenstufe befinden, können p- und n-Typ-Netzwerke leiten. Dies führt zu einem vorübergehenden Kurzschluss zwischen U_B und Masse. Bei gut entworfenen Schaltungen (schnelle Übergänge) beträgt dieser Anteil nur einen Bruchteil der Schaltungsleistung.
- Leckstrom: Die Transistornetzwerke leiten auch im ausgeschalteten Zustand minimal (Bruchteil eines Nanoamperes pro Gatter). In der Regel kann der Leckstrom vernachlässigt werden.

²¹⁷CMOS = Complementary Metal-Oxid-Semiconductor

²¹⁸Transitionen

²¹⁹IC = Integrated Circuit (deutsch: Integrierter Schaltkreis)

²²⁰positive Betriebs- oder Versorgungsspannung

²²¹engl.: Ground; negative Versorgungsspannung

²²²engl.: Threshold

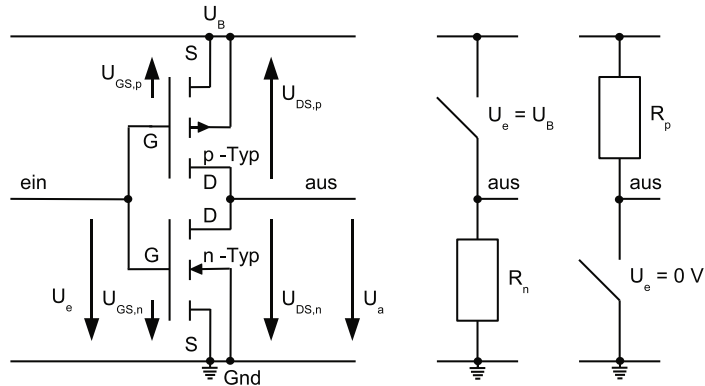


Abbildung 2.107: CMOS-Inverter

Beispiel: Funktion CMOS-Inverter (siehe Abbildung 2.107):

- 1. Fall: $U_e = 0 \text{ V}$ – logisch '0'
 $U_{GS,n} = 0 \text{ V}$ (n-Typ sperrt)
 $U_{GS,p} = -U_B < U_{tp}$ (p-Typ leitet)
- 2. Fall: $U_e = U_B$ – logisch '1'
 $U_{GS,n} = U_B > U_{tn}$ (n-Typ leitet)
 $U_{GS,p} = 0 \text{ V}$ (p-Typ sperrt)

Gesamt- leistung

Die Gesamtleistung unter Vernachlässigung der Anteile Kurzschlussleistung und Leckstrom kann wie folgt ermittelt werden:

$$P_S = 1/2 \cdot f \cdot U_B^2 \cdot \sum A_g \cdot C_L^g.$$

Die Abkürzung „S“ steht für Schaltung und „g“ für Gate. Hierbei sind f die Taktfrequenz, A der Aktivitätsfaktor (nicht alle Gatter schalten in jedem Taktzyklus) und C_L die Lastkapazität des Gatters.

Anhand der Gleichung können die Maßnahmen für eine energiesparende Schaltungen erläutert werden:

- Reduktion der Versorgungsspannung U_B : Aufgrund der quadratischen Abhängigkeit ist dies offensichtlich.
- Minimierung der Schaltungsaktivitäten A : Wenn Schaltungsaktivitäten nicht benötigt werden, sollten diese vermindert werden, zum Beispiel über „Clock Gating“.

- Reduktion der Gatteranzahl: Einfache Schaltungen benötigen weniger Energie als komplexe (bei ansonsten gleichen Randbedingungen). Somit tragen weniger Gatter zur Summenbildung bei.
- Reduktion der Taktfrequenz f : Die Vermeidung unnötig hoher Taktfrequenzen ist sicher wichtig. Allerdings reduziert sich hierdurch auch die Rechenleistung (MIPS²²³).

Die Verringerung der Geometrie von CMOS-Prozessen hat zwangsläufig auch eine Verringerung von U_B zur Folge. Werden Transistoren kleiner, so steigt die Feldstärke bei konstanter U_B . Dies führt zu einer Zerstörung der Halbleiter. Wird U_B verringert, so verringert sich ebenfalls die Leistung der Schaltung. Der maximale Transistorstrom I_{sat} in Sättigung²²⁴ lässt sich wie folgt bestimmen:

$$I_{sat} \propto (U_B - U_t)^2$$

Die maximale Taktfrequenz beträgt hierbei:

$$f_{max} \propto (U_B - U_t)^2 / U_B.$$

Verringert man U_B , so verringert sich ebenfalls die Frequenz f_{max} . Es ist offensichtlich, dass eine Reduktion von U_t zu einer Reduktion der Verlustleistung führt. Allerdings ist der Leckstrom²²⁵ I_{leak} stark von U_t abhängig:

$$I_{leak} \propto \exp(-U_t/35mV).$$

Bei kleiner Verringerung von U_t hat dies eine deutliche Erhöhung des Leckstroms zur Folge. Der Leckstrom verursacht ein schnelleres Entladen der Batterien von inaktiven Schaltungsteilen. Daher ist auf Wechselwirkungen zwischen der Maximierung der Leistung und des Energieverbrauchs im Ruhezustand²²⁶ zu achten.

**Reduktion
Versorgungs-
spannung**

Beispiel: Aktuelle Technologien verwenden Betriebsspannungen von 1 bis 2 Volt.

Abschließend noch einige Wege zur Einsparung von Energie im Allgemeinen:

- Reduktion von U_B : niedrigste geforderte Taktfrequenz wählen. Die Versorgungsspannung ist bei gegebener Taktfrequenz soweit wie möglich zu reduzieren. Hierbei darf U_B nicht so weit vermindert werden, dass die Verluste den Verbrauch im Ruhezustand stark beeinflussen.
- Reduktion der On-Chip-Aktivität: Vermeidung überflüssiger Taktung von Schaltfunktionen. Dies kann zum Beispiel mittels Schlafmodi („Gated Clocks“) erfolgen.

**Energie-
einsparung**

Schlafmodi

²²³MIPS = Million Instructions Per Second

²²⁴engl.: Saturation

²²⁵engl.: Leakage

²²⁶engl.: Standby

- Reduktion der „Off-Chip“-Aktivität: Minimierung von Aktivitäten außerhalb des Chips, zum Beispiel von Caches, um den Zugriff auf externe Speicher zu senken.
- Nutzung von Nebenläufigkeit (CIS): Bei Verdopplung einer Schaltung kann dieselbe Rechenleistung bei halber Taktfrequenz erreicht werden ([Fur02], S. 50 ff).

Aufgaben:

1. Erläutern Sie den Zusammenhang zwischen abgestrahlter Energie und Reichweite.
2. Nennen Sie Maßnahmen zur Reduktion des Energieverbrauchs.
3. Wie werden prinzipiell „Schlafmodi“ bei Mikroprozessoren implementiert?

Beispiel: Handies schalten schrittweise die Anzeige ab, um Energie zu sparen. Ein Tasten-Interrupt „weckt“ dann das Gerät wieder zur Kommunikation auf.

2.4.5 Echtzeit-Datenverarbeitung

Die Echtzeit stellt eine der wichtigen Randbedingungen bei der Entwicklung von Eingebetteten Systemen (siehe Abschnitt 2.4.2) dar. Diese Grundlagen werden für die Kapitel 3 und 4 benötigt.

„Echt-Zeit“ setzt sich zusammen aus der „Zeit“ als physikalische Größe der Datenverarbeitung und „Echt“ im Sinne von „für den Menschen als real empfundene Zeit“.

Abbildung 2.108 und Tabelle 2.28 (E/A^{227}) zeigt einen Vergleich zwischen Echtzeit- und konventioneller Verarbeitung. Echtzeitsystem²²⁸ setzt sich aus den beiden Komponenten Prozess und Datenverarbeitungsgerät zusammen. Unter Prozess versteht man einen beliebigen Ablauf, der aufeinander folgende

²²⁷ $E/A = \text{Ein-/Ausgänge}$

²²⁸ engl.: Real-Time System

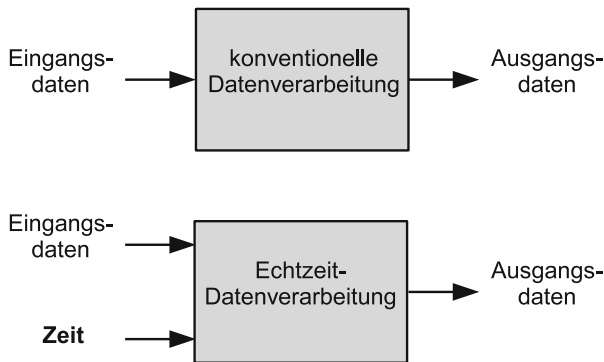


Abbildung 2.108: Echtzeit-Datenverarbeitung im Vergleich zur konventionellen Datenverarbeitung

Kategorie	Informationssysteme	Echtzeitsysteme
Steuerung	datengesteuert	ereignisgesteuert
Datenstrukturen	komplex	einfach
Eingangsdaten	große Menge	kleine Menge
Fokussierung	E/A-intensiv	rechenintensiv
Portierbarkeit	maschinenunabhängig	auf Hardware zugeschnitten

Tabelle 2.28: Vergleich Informations- und Echtzeitsysteme

zeitliche Resultate erzeugt. Heutzutage sind Datenverarbeitungsgeräte digitale Rechner mit Kommunikations-Schnittstellen. Die Geräte bestehen aus Hard- und Software-Einheiten. Die Hardware ist kompakt und verfügt über einen hohen Integrationsgrad – man nennt diese Art von Geräten auch eingebettete Systeme.

**eingebettete
Systeme**

Beispiele:

1. Prozesse: Regelung des Wohnraumklimas; Ablaufsteuerung einer Waschmaschine
2. Datenverarbeitungsgeräte: Controller für Motormanagement

Unter harten Echtzeitsystemen versteht man Systeme, die harte zeitliche Schranken vorgeben. Die Datenverarbeitung muss gewährleisten, dass die Zeitvorgaben eingehalten werden, da es sonst zu schwerwiegenden Ausfällen kommt. Es wer-

**harte
Echtzeit**

weiche Echtzeit

den keine Verstöße gegen die zeitlichen Vorgaben erlaubt. Derartige Systeme müssen im Normalfall hundertprozentig deterministisch sein.

Mit weichen Echtzeitsystemen sind Systeme gemeint, die prinzipiell in Echtzeit arbeiten, aber über sehr dehnbare Zeitgrenzen verfügen. Nichtsdestotrotz muss das Echtzeitsystem deterministisch sein, wobei die vorhersagbaren zeitlichen Grenzen sehr große Toleranzen aufweisen.

Beispiele:

1. harte Echtzeitsysteme: Flugzeuge, Kraftwerke
2. weiches Echtzeitsystem: Bankomat

synchrone Programmierung

Bei der Einteilung der Art, wie eine Applikation verwaltet wird, unterscheidet man zwischen synchroner und asynchroner Programmierung.

Die synchrone Programmierung, auch „Zeitscheibenprogrammierung“ genannt, basiert auf einem festen Takt eines Zeitgebers – man spricht deshalb von einer Zeitsteuerung. Die Festlegung der Reihenfolge der verschiedenen Teilprogramme einer Applikation wird beispielsweise in einer Tabelle hinterlegt. Hierbei erhält jedes Teilprogramm eine bestimmte „Zeitscheibe“²²⁹, die nicht verlängert werden kann. Die synchrone Programmierung ist somit hundertprozentig deterministisch.

*Merksatz: **Synchrone Programmierung***

Die Organisation des Ablaufs lässt sich mit einer Zustandsmaschine vergleichen. Im jeweiligen Zustand wird ein Programmabschnitt mit bestimmter Länge abgearbeitet.

asynchrone Programmierung

Die asynchrone Programmierung reagiert flexibel auf asynchrone Ereignisse; man spricht von einer Ereignissteuerung. Aufgrund des schwer voraussagbaren zeitlichen Ablaufs ist der Nachweis der Determiniertheit schwerer nachzuweisen als bei der synchronen Programmierung. Systemunterbrechungen²³⁰ werden von externen Ereignissen ausgelöst. Aufgrund der hohen Priorität wird das laufende Hauptprogramm unterbrochen. Es folgt die Abarbeitung der mit dem Interrupt fest verbundenen „Interrupt Service Routine“.

*Merksatz: **Polling, Interrupt***

Ereignisse können durch eine spezielle Hardware direkt Einfluss auf den Ablauf des Programms nehmen (Interrupt-Steuerung) oder durch Software-Abfrage (Polling) erkannt werden.

²²⁹engl.: Time Slice

²³⁰engl.: Interrupts

In der Praxis trifft man oft Mischsysteme aus synchroner und asynchroner Programmierung an ([Wit00], S. 26 ff).

Aufgaben:

1. Was ist ein Echtzeitsystem?
2. Was versteht man unter asynchroner und synchroner Programmierung?
3. Was ist der Unterschied zwischen Interrupt und Polling?

2.4.6 Eingebettete Funksysteme

Abbildung 2.109 zeigt die Beziehung zwischen dem Entwurf von Eingebetteten Systemen und der Netzwerk-Technologie (siehe Nachrichten- und Kommunikationstechnik Abschnitt 2.3, 2.2).

Eine weitere Schnittstelle bilden die Telemetrie und Telesurveillance.

Hieraus resultiert der prinzipielle Aufbau eines Eingebetteten Funksystems. Abbildung 5.3 zeigt eine Zwei-Chip-Lösung mit Schnittstellen, bestehend aus Transceiver und Mikroprozessor. Der Mikroprozessor beinhaltet hierbei den Protokollstapel.

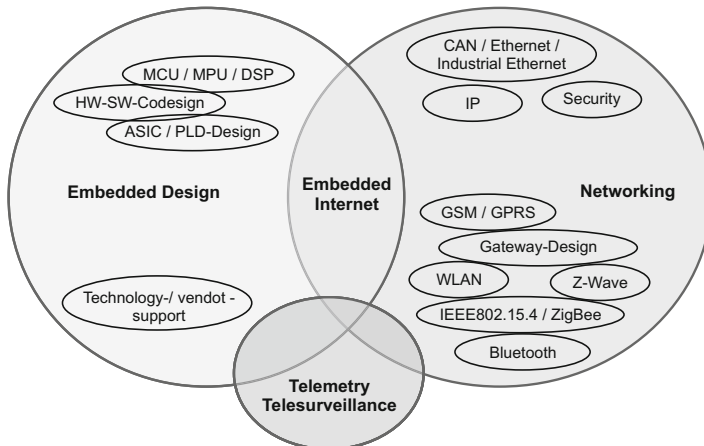


Abbildung 2.109: Einordnung der Eingebetteten Funksysteme [Sik05]

Definition: **Telemetrie**

Telemetrie^a wird als Messwert- und Datenübermittlung sowie Fernsteuerung, Fernregelung und Fernüberwachung definiert. Telemetrie ist Messen, Steuern und Regeln über große Distanzen. Telemetrie dient zur Überwachung von Lebewesen, Geräten, Innen- und Außenanlagen und zur Kontrolle von Anlagen. Typische Anwendungen sind Raumfahrt, Medizin (Telesurveillance), Umwelttechnik etc. [Zog02]

^aengl.: Telemetry

Kapitel 5 beschreibt die Implementierung der Eingebetteten Funksysteme.

Merksatz: **Hochintegration**

Eine weitere Hochintegration kann mittels der unter den Abschnitten 2.4.2.2 und 2.4.2.4 vorgestellten Methoden erreicht werden.

Rand- bedingungen

Abbildung 2.110 zeigt die Herausforderungen beim Entwurf von Wireless-Systemen. Die einzelnen Entwurfs-Parameter sind miteinander gekoppelt. Allen Randbedingungen gleichzeitig gerecht zu werden, ist in den seltensten Fällen möglich und stellt die „Ingenieurs-Kunst“ dar.

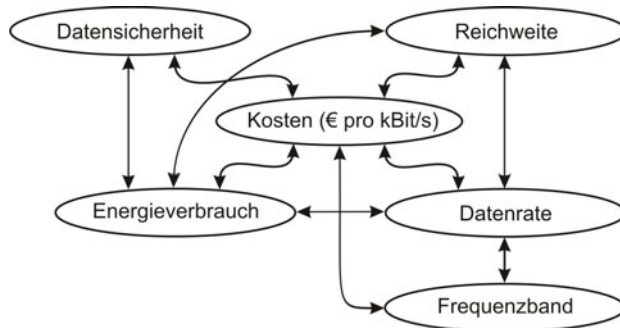


Abbildung 2.110: Herausforderungen beim Entwurf von Wireless-Systemen [Sik05].

„Ubiquitous Computing“

Die Entwicklung der Funknetzwerke im Nahbereich geht immer in Richtung allverfügbare, verteilte Netze im intelligenten Einzelknoten. Diese Entwicklungen werden unter dem Begriff der Allverfügbarkeit „Ubiquitous Computing“ zusammengefasst: 6 A's – anyone, anytime, always on, anywhere, any application, any device. Weitere Begriffe in diesem Kontext sind:

- Intelligenz in der Umgebung²³¹: Einsatz von Sensor-Aktor-Netzen mit

²³¹engl.: Ambient Intelligence

dem Ziel einer massiven Vernetzung der Sensoren, Funkmodule und Computer, um den Alltag zu verbessern. Dieser Ansatz steht im engen Zusammenhang mit dem EU-Forschungsprogramm „Information Society Technologies“.

- Rechnerdurchdringung²³²: ist die von der Industrie geprägte Variante der Abdeckung von Geschäftsprozessen, aber auch von Lebensbereichen [KS07].

Zusammenfassung^a:

1. Die Leser können die Kommunikations- und Nachrichtentechnik einordnen.
2. Die Anwender kennen Huffman-Codierung und Hamming-Code.
3. Sie sind geübt in der Bestimmung der Hamming-Distanz und der Kanalkapazität.
4. Die Leser kennt die verschiedenen Modulationsarten und sind geübt in der Ermittlung von Leistungsbilanz und Reichweiten-Abschätzung.
5. Sie sind vertraut mit den Eingebetteten (Funk-)Systemen bezüglich der Definition, Randbedingungen, Arten, Energieverbrauch und Echtzeit-Datenverarbeitung.

^amit der Möglichkeit zur Lernziele-Kontrolle

²³²engl.: Pervasive Computing

Wireless-Netzwerke für den Nahbereich
Eingebettete Funksysteme: Vergleich von
standardisierten und proprietären Verfahren
Gessler, R.; Krause, Th.
2015, XXI, 354 S. 151 Abb., Softcover
ISBN: 978-3-8348-1239-1