Grundlagen der Übertragungstechnik

Vorlesung an der Dualen Hochschule Baden-Württemberg, Karlsruhe

Kurs: Tel17NT

Prof. Dr.-Ing. Gerald Oberschmidt [1]
Duale Hochschule Baden-Württemberg Karlsruhe
Fakultät für Technik
Erzbergerstraße 121, 76133 Karlsruhe, Germany
Raum 520, Tel.: +49 721 9735 886, Mail: oberschmidt@dhbw-karlsruhe.de

25. Januar 2020

[1]Urheberrechtlich geschütztes Material, Kopie, Weitergabe, Veröffentlichung auch auszugsweise nur nach Rücksprache mit dem Rechteinhaber.

Inhaltsverzeichnis

1 Zum Start ein Überblick **1**

1.1 Zielsetzung und Kompetenzen 1

1.2 Struktur der Vorlesung 2

1.2.1 Unterstützende Lehrmaterialien 3

2 Einführung in die Übertragungstechnik **5**

2.1 Das einfachste Nachrichtensystem 5

2.2 Vereinbarungen, Protokolle, Standards und Modelle 7

2.2.1 Standards 7

2.2.2 Referenzmodelle 11

2.3 Ethernet 12

2.3.1 Referenzmodell einer Ethernetvariante: 100 GBase-… 16

2.3.2 Weitere “Referenzmodelle” 19

2.4 Lernkontrolle 19

3 Leitungen und Pegel **20**

3.1 Pegelmaße 20

3.1.1 Lernkontrolle 22

3.2 Leitungen allgemein 23

3.2.1 Ersatzschaltbild der Leitung 24

3.2.2 Leitungsgleichungen 24

3.2.3 Harmonische Lösung der Leitungsgleichung 26

3.2.4 Wellenlänge und Wellenimpedanz 28

3.2.5 Lösung der Feldgleichungen im statischen Fall: Parameter von TEM-Leitungen 30

3.2.6 Dispersion und das Drumherum 31

3.2.6.1 Gruppen- und Phasengeschwindigkeit 32

3.2.6.2 Definition der Dispersion und Konsequenzen: Die Breiwerdung des Impulses . . . 35
3.2.6.3 Kompensation der Dispersion: Zwei Methoden . . . 36
3.3 Optische Wellenleiter . . . 36
3.3.1 Grundsätzliche Funktion . . . 38
3.3.2 Eigenschaften des Glases . . . 46
3.3.3 Multimoden und Einmoden-Fasern . . . 47
3.3.3.1 Patching und Verbinden von Glasfasern . . . 49
3.3.3.2 Dispersion . . . 52
3.3.4 Messungen an Fasern: OTDR . . . 55
3.4 Übungsaufgaben . . . 56
3.4.1 Dimensionierung einer optischen Übertragungsstrecke . . . 56
3.4.2 Übertragung mit koaxialen Leitungen . . . 56
3.4.3 Dispersion . . . 57

4 Übertragung im Basisband 58
4.1 Das Übertragungssystem im Basisband . . . 58
4.2 Repräsentation der Daten auf der Leitung . . . 61
4.2.1 Spektrum und Spektrale Eigenschaften . . . 66
4.2.1.1 Eigenschaften der Fouriertransformation . . . 67
4.2.1.2 Sätze zur Fouriertransformation . . . 67
4.2.1.3 Einige Beispielfunktionen zur Fouriertransformation . . . 75
4.2.1.4 Diskrete Fourier-Transformation . . . 76
4.2.1.5 Leistung und Energie . . . 77
4.3 Komplexe Signale im Basisband . . . 77
4.4 Bitfehler und Störungen durch Rauschen . . . 81
4.4.1 Gauss'sches Rauschen . . . 81
4.4.2 Bitfehler . . . 83
4.4.3 Kanalkapazität . . . 85
4.4.4 Bitfehlerraten für komplexe Modulationsformen . . . 85
4.5 Spektrum und Impuls . . . 87
4.5.1 Cos-Roll-Off-Filter . . . 91

5 Datensicherung und Kodierung 93
5.1 Sicherung der Datenübertragung gegen Bitfehler . . . 93
5.2 Lineare Codes in Matrizendarstellung . . . 95

5.2.1 Hammingdistanz . . . 95
5.2.2 Linearer Code . . . 96
5.2.3 Kodierung . . . 98
5.2.4 Dekodierung und Prüfmatrix . . . 99
5.2.5 Leistungsfähigkeit eines Codes . . . 102
5.2.5.1 Kodiergewinn und Dekodierung . . . 102
5.2.6 Konkret verwendete lineare Codes . . . 106
5.2.6.1 Hamming-Codes . . . 106
5.2.6.2 Hadamard-Code . . . 107
5.2.6.3 Low-Density Parity-Check Codes . . . 109
5.2.6.4 Tanner Grafen und Dekodierung von LDPC . . . 112
5.3 Zyklische Codes . . . 116
5.3.1 Generatorpolynom . . . 116
5.3.1.1 Systematischer zyklischer Code . . . 117
5.3.2 Realisierung mit rückgekoppeltem Schieberegister . . . 119
5.3.3 Verwendete Generatorpolynome und Frameaufbau . . . 120
5.4 Reed-Solomon Codes . . . 125
5.4.1 Galois-Felder, rechnen auf endlichen Körpern . . . 125
5.4.1.1 Erweiterungskörper . . . 126
5.4.2 Polynomkörper und Kodierung . . . 128
5.4.3 Dekodierung und Fehlerkorrektur . . . 131
5.4.3.1 Ablauf der Dekodierung und Fehlerkorrektur . . . 133
5.4.3.2 Bestimmung der Fehlerstellen (Berlekamp-Massey Algorithmus) . . . 133
5.4.3.3 Korrektur der Lösung . . . 135

6 Analoge Modulation 136
6.1 Amplitudenmodulation . . . 136
6.1.1 Zweiseitenband Amplitudenmodulation (ZSB-AM) . . . 136
6.1.2 Zweiseitenband-Amplitudenmodulation mit unterdrücktem Träger (ZSB-AM-uT) . . . 138
6.1.3 Einseitenband-Amplitudenmodulation (ESB-AM) . . . 140
6.2 Phasenmodulation (PM) . . . 142
6.3 Frequenzmodulation (FM) . . . 143

7 Mobilkommunikation - Einführung **146**
7.1 Technische Basis Trivia . . . 150
7.2 Der Mobilfunkkanal . . . 151
7.2.1 Mehrwegeausbreitung . . . 151
7.2.2 Rayleigh-Kanal . . . 154
7.2.3 Rice-Kanal . . . 159
7.2.4 Mobilfunkkanäle mit Symbol-Interferenzen . . . 160

8 Anpassung an die Kanäle: Spreizung des Spektrums **164**
8.1 Frequenz-Sprung-Verfahren (Frequency Hopping Spread Spectrum) . . . 165
8.2 Direct Sequence Spread Spectrum und Code Division Multiple Access . . . 167
8.2.1 PN-Sequenzen . . . 169
8.2.1.1 Kreuzkorrelation . . . 171
8.2.1.2 Gold-Sequenzen . . . 171
8.2.2 Spread-Spectrum-Modulation . . . 174
8.2.3 Mehrfachausnutzung des gespreizten Kanals . . . 176
8.2.4 Das Beispiel UMTS . . . 178
8.3 Othogonal Frequency Division Multiplexing in 4G, 5G usw. . . . 180

9 Modulation, Demodulation und Trägerrückgewinnung **186**
9.1 Ideale Modulation und De-Modulation . . . 186
9.2 Trägerrückgewinnung . . . 189
9.2.1 Phase Locked Loop (Phasenregelkreis) (PLL) und Trägerrückgewinnung aus dem unmodulierten Träger . . . 189
9.2.2 Costas-Loop und Trägerrückgewinnung aus dem modulierten Signal 195
9.2.3 Trägerrückgewinnung für allgemeine Modulationen . . . 198
9.3 Taktrückgewinnung . . . 199

10 Kanalkodierung: Erweiterung für Funkkanäle **200**
10.1 Faltungscodes . . . 201
10.1.1 Weitere Beispiele für Faltungscodes . . . 205
10.1.2 Leistungsfähigkeit von Codes . . . 206
10.1.3 Turbocodes und IIR-Faltungscodes . . . 207
10.2 Interleaver . . . 210

11 Kanal-Zugriffstechniken und Spektrumsausnutzung **213**
11.1 Einführung und Überblick . . . 213

11.2 Frequency Division Multiple Access . . . 215
11.3 Time Division Multiple Access . . . 216
11.4 Access gegenüber Duplex . . . 217
11.5 Beispielhafte Untersuchung des GSM-Systems . . . 218
11.5.1 Datenrate und Spektrale Ausnutzung . . . 218
11.5.2 Netzwerkbetrachtungen . . . 221
11.6 Code Division Multiple Access . . . 222
11.7 Orthogonal Frequency Division Multiple Access . . . 223
11.8 Ungesteuerter Zugriff: ALOHA und CSMA . . . 223

12 Sender und Empfänger **224**
12.1 Empfängerstrukturen . . . 224
12.1.1 Quadratischer Detektor und Geradeausempfänger . . . 225
12.1.2 Heterodyner Empfang (Superhet) . . . 226
12.1.3 Homodyner Empfang . . . 227
12.2 Optische Strukturen . . . 229
12.2.0.1 Wellenlängenmultiplex . . . 231
12.2.1 Optische Senderstruktur . . . 233
12.2.1.1 Halbleiter . . . 233
12.2.1.2 PN-Übergang und Light Emitting Diode [Leuchtdiode] (LED) . . . 234
12.2.1.3 Laserdiode und Resonator . . . 235
12.2.2 Optische Empfängerstruktur . . . 241
12.2.2.1 Photodioden . . . 241
12.2.3 Small Formfactor Plugable (SFP)-Module . . . 245
12.2.4 Optische Verstärker (EDFA) . . . 246
12.2.5 Optische Filter (Multiplexer (MUX), De-Multiplexer (DeMUX)) . . . 248
12.3 Optische Übertragungsverfahren und 100 GBit/s . . . 251
12.3.1 Optische Filter im DWDM (ROADM) . . . 251

A Simulationstechniken **252**
A.1 Simulationsumgebung . . . 252
A.1.1 Matlab und Clone Gnu-Octave . . . 252
A.1.2 C/C++ und Java . . . 253
A.1.3 Python . . . 255
A.1.4 Ptolemy . . . 256

A.1.5 LabView 257
A.2 Simulation 257
A.2.1 Datenerzeugung, die PN-Sequenz 257
A.2.1.1 Eigenschaften von PN-Sequenzen 259
A.2.1.2 Aufgaben 260
A.2.2 Übertragungskanal und Bitfehlerrate 261
A.2.2.1 Kanal mit Arbitrary Whit Gaussian Noise (AWGN) . . . 261
A.2.2.2 Aufgaben 261
A.2.2.3 Bitfehlerrate im Schwundkanal 262
A.2.2.4 Aufgaben 262

Literaturverzeichnis **263**

Glossar **268**

Index **274**

Kapitel 1

Zum Start ein Überblick

1.1 Zielsetzung und Kompetenzen

- Sie lernen die grundsätzlichen Aufbau eines Kommunikationssystems kennen.
- Sie verstehen die physikalischen Grundlagen der Informationsübertragung und der daraus resultierenden Konsequenzen für Übertragung von Information bei unterschiedlichen Herausforderungen (Strecke, Geschwindigkeit, Sicherheit, Datenrate etc.).
- Sie können verschiedenen Arten der Informationsübertragung benennen und Lösungen sowie Bausteine üblicher Systeme benennen und charakterisieren.
- Sie verstehen die Erzeugung und Übertragung von Daten im Digitalen.
- Sie verstehen theoretische und praktische Grenzen der Informationsübetragung und sind in der Lage, Lösungen zum Erreichen und/ oder Herausschieben dieser Grenzen anzugeben.
- Sie verstehen, wie verschiedene Nachrichtenquellen und -senken auf ein Übetragungsmedium zugreifen können und dieses optimal ausnutzen können.
- Sie sind in der Lage, Information auch bei Übertragung über nichtideale Nachrichtenkanäle zu sichern.
- Sie kennen technisch aktuell bedeutsame Übertragungssysteme und deren Parameter.

- Sie wissen, wo man sich über die Details der Kommunikationssysteme informieren kann.

1.2 Struktur der Vorlesung

Die Vorlesung ist in zwei große Teile geteilt:

1. Ethernet (and friends)

   beschreibt die physikalische Schicht (physikal Layer) der heute in nahezu allen Bereichen üblichen Übertragung im Kabel nach dem Ethernet (IEEE 802.3) Standard. Im Mittelpunkt steht die leitungsgebundenen Übetragung von elektrischen und optischen Signalen. Es werden die angrenzenden Gebiete des Ethernet im Wide Area Network (WAN) wie Synchronous Digital Hierarchy (SDH) und Synchronous Optical Network (SONET), wie auch im Zugangsnetz bspw. Asynchronous Digital Subscriber Line (ADSL) erläutert.

   Wichtige Themen sind

   - Eigenschaften von Kabeln und Wellenleitern (Glasfasern)
   - Aussendung und Detektion von elektrischen und optischen Signalen
   - Erzeugung von Signalen (eigentlich: Wandlung analoger akustischer Signale in digitale)
   - Digitale Übertragung der Signale in einfachster Form (Kodierung und Basisband)
   - Digitale Übertragung in etwas erweiterter Form (Modulation, Spektrum und Abtastung)
   - Sicherung der Daten gegen Bitfehler in einfacher Form
   - Zugriff mehrerer Stellen auf das gleiche Medium
   - Immer wieder: Umsetzung der Techniken für die Datenübertragung nach dem Ethernet-Standard
   - Technik von Zugangsnetzen

2. Universal Mobile Communication System (UMTS)/ Long Term Evolution (LTE) (and friends)

beschreibt im zweiten Teil die drahtlose Übertragung im Mobilfunknetz. Hier stehen Erweiterungen der im ersten Teil angesprochenen Techniken im Vordergrund, wie z.B.

- Die Luft als Kommunikationskanal
- Senden und Empfangen (im Überblick)
- Mehrwegeausbreitungen und (der extreme) Mehrfachzugriff auf das Medium
- Erweiterte Modulationsverfahren
- Erweiterte Datensicherungsverfahren (Faltungs-Codes/ Block-Codes) und deren Dekodierung (Trellis und Viterbi)
- Breitbandübertragung mit Code Division Multiple Access (CDMA) und Orthogonal Frequency Division Multiplexing (OFDM).
- Ausblick auf Übertragungsstandards für kurze Reichweite (Wireless Local Area Network (WLAN), Bluetooth, ZigBee)

Die Vorlesung steht im engen Zusammenhang mit der Lehrveranstaltung zur Hochfrequenztechnik, in der verschiedenen Themen z.B. der Leitungen und Antennen deutlich vertieft behandelt werden. Weiterhin wird die Vorlesung durch Laborübungen unterstützt.

1.2.1 Unterstützende Lehrmaterialien

Im Laufe der Vorlesung wird auf die jeweils frei erhältliche Software

- *Gnu-Octave* [1] zur Simulation und Darstellung physikalischer und mathematischer Zusammenhänge,
- *WinIQSim* [2] und als neuer Version *WinIQSim2* [3] der Firma Rohde und Schwarz zur Darstellung komplexer digitaler Signale
- *Ptolemy II* [4] der University of califormia at Bekeley wird als Programmier- und Simulationsumgebung verwendet. Dieses Programm hat eine grafische Oberfläche und unterstützt optimal das Datenflussprinzip.

zurückgegriffen.
Natürlich werden wir Lehrbücher verwenden, von denen einige schon im Vorfeld genannt werden sollen:

- **Nachrichtenübertragung von D. Kammeyer** [5] Als ein umfassendes Standardlehrbuch zur digitalen Nachrichtenübertragung
- **Grundlagen der digitalen Nachrichtenübertragung von C. Roppel** [6] gibt einen sehr schönen Überblick über die Kommunikationstechnik und lässt fast kein Thema aus. Allerdings wird die Physik der Übertragung auf Leitungen oder anderswo nicht behandelt, weshalb
- **Elemente Optischer Netze von V. Brückner** [7] herangezogen werden muss. Hierin werden die optischen Übertragungsverfahren kurz und zum Punkt erläutert.
- **Ethernet von Jörg Rech** [8] erläutert das Ethernet und liefert damit die Überschriften und den roten Faden durch diese Vorlesung. Geht als technisches Buch für das, was wir hier tun aber nicht tief genug.
- **Digital Communication von J. Proakis** [9] gibt einen tiefen und theoretischen sowie umfassenden Einblick in die digitale Kommunikationstechnik und ist als Standardwerk im angelsächsichen Raum zu sehen.

Kapitel 2

Einführung in die Übertragungstechnik

Zunächst wollen wir uns einen Überblick verschaffen, was ein Übertragungssystem überhaupt ist, und über was wir hier überhaupt reden. Grundsätzlich werden in einem solchen System Daten (Nachrichten) von einem Sender (bzw. der Nachrichtenquelle) an einen Empfänger (die Nachrichtensenke) über einen Kanal übertragen. In Abb. 2.1 ist so ein generisches System einmal aufgeführt. Ziel ist es in der eins zu eins Kommunikation, dass die Nachrichten fehlerfrei übertragen werden und natürlich auch auf der anderen Seite verstanden werden.

2.1 Das einfachste Nachrichtensystem

Hierzu sind nun einige Überlegungen notwendig:

- Sender und Empfänger müssen die gleiche Sprache sprechen, den gleichen Code benutzen.

  Das heißt, dass sämtliche Parameter der Nachrichtenübertragung auf beiden Seiten bekannt sein müssen, was beinhaltet, dass klar sein muss, ob die Übertragung optisch (etwa durch Taschenlampe oder Laser), elektrisch (Strom, welche Amplitude),

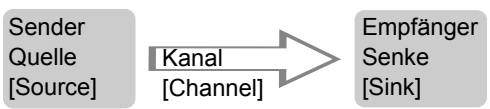


Abbildung 2.1: Einfachste Darstellung eines Übertragungssystems

akustisch (Sprache oder Ultraschall) oder mechanisch (Seilzug) erfolgen soll. Welche Pegel zugrunde zu legen sind, in welcher Geschwindigkeit (Takt) zu übertragen ist, aber auch, was ein Zeichen bedeutet. Heißt etwa ein lauter Schrei: “Komm’ mir zur Hilfe” oder (wie etwa das Gebrüll des Löwen): “Bleib mir fern!”. Weiterhin muss der Code der Übertragung verstanden werden. Auch das ist klar, lernt man doch sogar als Mensch verschiedene Sprache, um genau dieses zu erreichen. Warum sollte es also bei Maschinen anders sein?

- Sender und Empfänger müssen den Kanal erreichen können.

  Haben Sie sich schon einmal unter Wasser unterhalten? Das geht nicht oder nur sehr schwer, Ihre Organe (Mund und Ohren) sind nicht gut auf das Übertragungsmedium (Wasser) abgestimmt. Zeichensprache, also optisch, geht, so lange das Wasser nicht zu trübe ist. Im luftleeren Raum - mal abgesehen von Ihren Überlebenschancen - können Sie sich auch nicht unterhalten. Auch ein optisches Unterseekabel hilft Ihnen nichts, wenn Sie keinen passenden Laser haben, um die zu übertragenden Signale dort anzuregen.

- Die Übertragung soll fehlerfrei, bzw. mindestens fehlerarm sein

  Beim Spiel “Stille Post” dienen viele Kinder, die sich Worte ins Ohr flüstern als Übertragungskanal. Was herauskommt ist jedesmal lustig. So ist es aber nicht, wenn Sie ernsthaft Daten übertragen wollen. Hier soll die Fehlerrate gering sein, Sie müssen bei der Übertragung also immer dafür sorgen, dass die Information niemals zu schwach wird, dass sie nicht mehr verstanden werden kann, dass vielleicht Datenworte ineinander laufen und so unverständliches Kuddelmuddel dabei herauskommt. Und natürlich müssen Sie sich vor bewussten Störungen des Kanals (Kinder finden so etwas lustig) schützen.

Und natürlich sind ein Sender und ein Empfänger niemals alleine auf der Welt, sondern agieren im Verein mit vielen. Viele wollen kommunizieren und dafür den gleichen Kommunikationskanal nutzen. Sie brauchen Regeln. In der Schule ist es “Wenn alles schläft und einer spricht, das nennt der Mensch den Unterricht” [Wilhelm Busch?] und verschiedener Unterricht findet gleichzeitig in verschiedenen Räumen statt (Space Division Multiple Access (SDMA)), in Besprechungen lässt man den anderen ausreden (Time Division Multiple Access (TDMA)). Im Funk (Radio) bekommt man Frequenzen zugewiesen (Frequency Division Multiple Access (FDMA)).

2.2 Vereinbarungen, Protokolle, Standards und Modelle

2.2.1 Standards

Nach der oben geführten Vorrede ist klar geworden, dass Vereinbarungen, auf verschiedenen Ebenen notwendig sind. Diese Vereinbarungen sind in der Regel in Standards niedergelegt. Und so wie Sender und Empfänger über die verwendete Kommunikationsart einig werden müssen, müssen auch Menschen, Firmen, Konsortien, Interessengemeinschaften und Länder bzw. deren Regierungsvertreter bei der Definition eines erfolgreichen Kommunikationssystems Einigkeit zeigen.
In diesem Sinne ist es nicht weiter verwunderlich, dass schon 1865 in Paris die International Telecommunication Union (ITU) (damals als International Telegraph Union) als älteste internationale Organisation gegründet wurde [10]. Der Anlass war, dass eine rein einzelstaatliche Organisation des Telegraphiebetriebs schon nur ca. 20 Jahre nach der ersten Morsetelegraphie (1844 von Samuel Morse) den tatsächlichen Bedürfnissen nur ungenügend Rechnung trug. Die Staaten schlossen also Abkommen mit ihren Nachbarn, so dass allein Preussen fünfzehn verschiedene Vereinbarungen mit seinen (deutschen) Nachbar(kleinstaaten) für einen effizienten Nachrichtenverkehr brauchte. Denn natürlich hatte der Nachbarstaat im Westen andere Vorstellungen als der im Osten usw. Und so vereinbarten zwanzig europäische Staaten einheitliche Regeln für Ausrüstung, Betrieb und Abrechnung internationaler Verbindungen. Die ITU teilt die Welt sehr grob in drei Bereiche auf, nämlich Afrika, Europa, Nord Asien (frühere UdSSR) und Vorderasien in Region 1, beide amerikanische Teilkontinente in Region 2 und Südasien, Australien, Ozeanien in Region 3.
Heute besteht die ITU aus drei Hauptbereichen, der Telecommunication Standardization (ITU-T), Radiocommunication (ITU-R) und Telecommunication Development (ITU-D).

Standardization (ITU-T) sorgt für eine weitgehende Standardisierung von Techniken von Funkübertragung über Telefonie hin zu Internet und dort verbreiteten Verfahren bspw. zur Video- und Audiokompression. Die Standards heißen übrigen "recommendations".

Radiocommunication (ITU-R) regelt die Funk-Kommunikationsdienste, und zwar sowohl die terrestrischen wie auch die Satelliten gestützten. Hier wird international das gesamte Frequenzspektrum verwaltet und werden Bereiche und Dienste koordiniert. Diese Aktivität ist nicht nur in Grenzgebieten sinnvoll, oder bei Diensten,

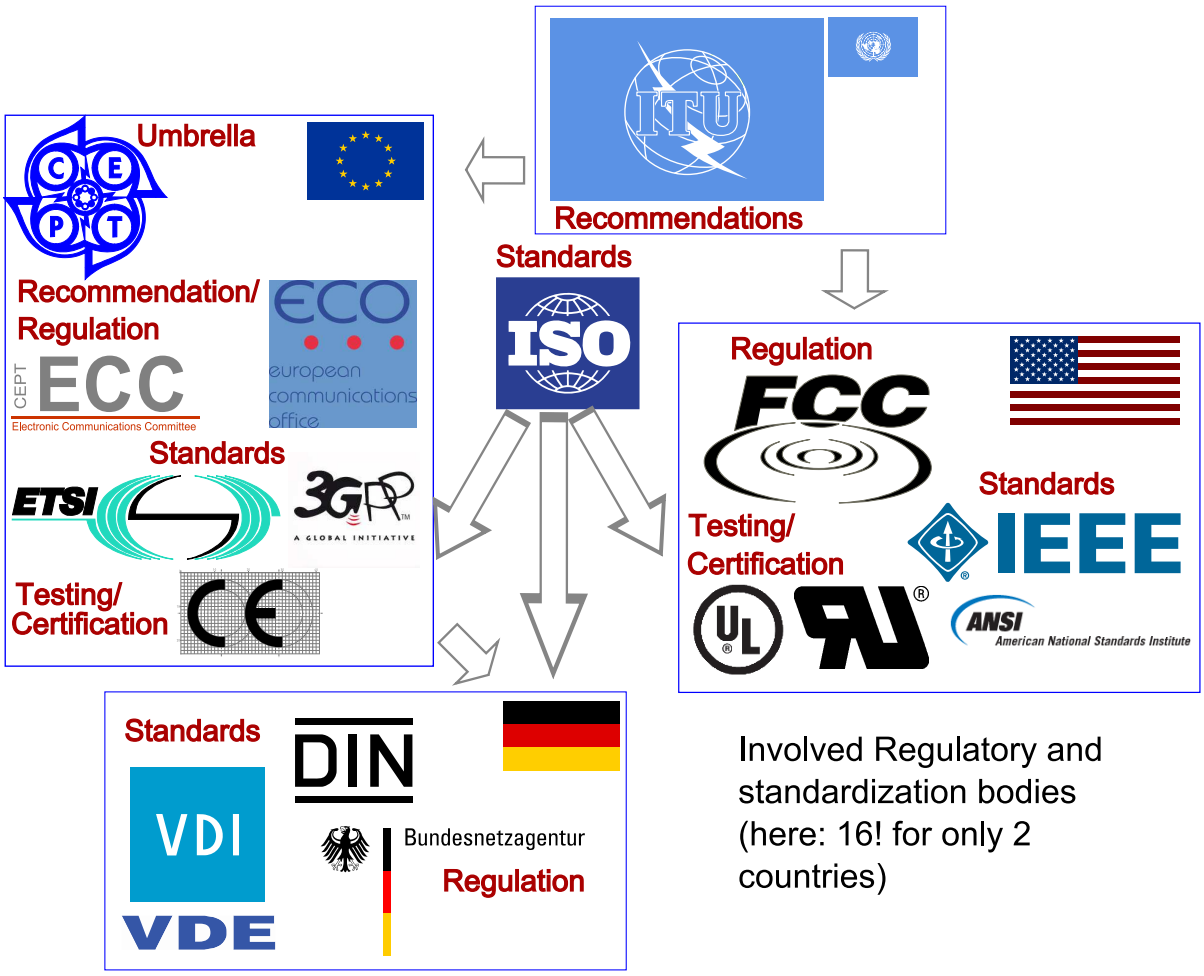


Abbildung 2.2: Übersicht über einige der wichtigsten Standardsierungsgremien.

in denen der Träger bspw. ein Schiff auch schon mal den regionalen Bereich wechseln kann, sondern eine international einheitliche Vergabe von Frequenzbereichen ermöglicht auch die Herstellung von Geräten für einen wesentlich größeren Markt.

Development (ITU-D) hat die Aufgabe, Ländern und Regionen bei der Entwicklung von Telekommunikationstechniken zu helfen. Damit sollen die Lebensbedingungen in den Entwicklungsländern verbessert werden.

Vor allem der erste Punkt ist für die Entwicklung von Kommunikationssystemen bedeutend. Im Rahmen von ITU gesteuerten Meetings werden Standards (Recommendations) erstellt und publiziert, nach denen sehr viele der heute bekannten Systeme funktionieren. Als weltweites Gremium hat die ITU natürlich keine nationale Gesetzesgewalt, und kann dementsprechend auch keine Industrieinteressensverbände lenken. Diese Aufgabe muss

von nationalen Instiutionen übernommen werden. In Deutschland ist dieses für die Bereich der Frequenzregulierung und auch der Zugangsgewährung zu (vorhandenen) Kommunikationsnetzen die Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen. Hier werden auf nationaler und damit gesetzlich relevanter Ebene bspw. die Frequenzen zugewiesen und Regularien verabschiedet, wie verschiedene Interessenten im Bereich der Kommunikationsnetze miteinander interagieren können.
Dazwischen geschaltet ist auf europäischer Ebene das European Communications Office (ECO), das als beratendes Gremium des European Conference of Postal and Telecommunications Administration (CEPT) Post und Telekommunikation in Europa unter einen Hut bringen will.
Auf europäischer und nationaler Ebene praktisch bedeutsam für die Fortentwicklung der Kommunkationstechnik sind die technisch orientierten und von Industrieverbänden unterstützten Organisationen

ETSI European Telecommunications Standards Institute (ETSI) ist durch den durchdringenden Erfolg des Global System for Mobile Telecommunication (GSM)-Standards berühmt geworden, als die bedeutsamste Standardisierungsinstitution weltweit. Seitdem werden von der ETSI Normen (mit dem Kürzel (European Norm (EN)) als Nachfolger bzw. übergeordnet zu den Deutsches Institut für Normung (DIN)-Normen) über die Funktion und Anforderungen von Funk- und Telekom-Geräten veröffentlicht. Die ETSI ist dementsprechend auch ein wichtiger Grundstein für das

3GPP , das Third Generation Partnership Project (3GPP) hat sich die Aktualisierung der Mobilkommunikation auf die Fahnen geschrieben und in diesem Sinne in den 1990er-Jahren den UMTS-Standard entwickelt, mit dem Ziel einen weltweit einheitlichen Standard für die Mobilkommunikation zu entwickeln. Formal ist dieses sicherlich gelungen, auch wenn natürlich die verschiedenen Varianten und regional nutzbaren Frequenzbänder die weltweite Diversität widerspiegeln und nicht ein einzelnes, einfaches Gerät genügt, um weltweit mobil verbunden zu sein. In den letzten Jahren wurde mit LTE eine Weiterentwicklung voran getrieben.

Entscheidend für uns ist, dass sämtliche Standards frei zugänglich über das Internet sind und somit jeder die Funktionsweise der genannten Systeme detailliert nachlesen kann. Seit einigen Jahren sind hier auch die Beschreibungen für GSM/EDGE Radio Access Network (GERAN), also letztlich GSM von der ETSI auf diese Organisation umgezogen worden.

VDE , der Verband der Elektrotechnik Elektronik Informationstechnik e.V. (VDE) als

der wichtigste Berufsverband der Elektroingenieure in Deutschland spielt in der Nachrichtentechnik - im Gegensatz zu seinem amerikanischen Pendant - kaum eine Rolle. Hier werden eher 230 V-Stecker genormt.

Ein Blick über den Atlantik lohnt. Nicht nur deshalb, weil dort ein großer Absatzmarkt liegt, sondern vor allem, weil dort unter dem Dach des Institute of Electrical and Electronics Engineers (IEEE) sehr sehr viel technisches Know-How entwickelt und verbreitet wird.

FCC Die Federal Communications Commission (FCC) ist die amerikanische Behörde, die die Frequenznutzung und die Zulassung von Geräten regelt und ist damit vielleicht vergleichbar mit der deutschen Bundesnetzagentur und auch den Normungsaufgaben der ETSI. Hier werden bspw. auch die Elektromagnetische Verträglichkeit (EMV)-Anforderungen der Funkgeräte für den US-Markt festgelegt.

IEEE der IEEE ist formal wohl die Schwestergesellschaft des VDE aber sehr viel größer und schlagkräftiger. Neben vielen Journalen und Konferenzen hat der IEEE eine starke Normungsgruppe und unter den 802.er-Reihe eine ganze Menge von wichtigen Telekommunikationssystemen beschrieben, die am besten unter `http://standards.ieee.org/getieee802/index.html` abgerufen werden können:

- .1 High Level Interface (HILI)
- .2 Logical Link Control
- .3 CSMA/CD (Ethernet)
- .5 Token Ring Access Method
- .11 Wireless LAN (WLAN)
- .15 Wireless Personal Area Network (WPAN)
- .16 Broadband Wireless Access (BWA)
- .17 Resilient Packet Ring (RPR)
- .18 Radio Regulatory Technical Advisory Group
- .19 Coexistence Technical Advisory Group
- .20 Mobile Broadband Wireless Access
- .21 Media Independent Handover
- .22 Wireless Regional Area Networks

Einige davon, nämlich der IEEE 802.3, der IEEE 802.11, der IEEE 802.15 und der IEEE 802.16 werden in dieser Vorlesung, die letzten drei eher in der zweiten Hälfte, eine Rolle spielen.

ANSI das American National Standards Institute (ANSI)

Wie schon eingangs erwähnt, ist der Schlüssel in der Kommunikation(stechnik), dass sich alle an die vereinbarten Regeln halten. Das ist wie im richtigen Leben, nur dann funktioniert es. Einige Regularien wie z.B. die von vom Gesetzgeber beauftragten Normungsgremien, hier sind das am ehesten die Bundesnetzagentur, die FCC, das ETSI und das ANSI, bekommen Gesetzeskraft und man begeht eine Ordnungswidrigkeit oder sogar Straftat, wenn man dagegen verstößt. Insbesondere wird es heikel, wenn Sachen oder gar Menschen zu Schaden kommen und Verstöße nachgewiesen werden können.
Andere Standards, hier sind Ergebnisse von 3GPP, ETSI, IEEE zu nennen, sind theoretisch nur Empfehlungen. Aber wird natürlich ein Kunde, und das kann einer der großen Netzbetreiber sein, nur Geräte kaufen oder als bspw. Handys in sein Netz lassen, wenn zertifiziert ist, dass diese Geräte den Standards entsprechen und keinen Ärger verursachen. Das Thema Zertifizierung wird aber hier nicht weiter behandelt.

2.2.2 Referenzmodelle

Damit man weiß, wovon man spricht, hat man sich in der Industrie auf Modelle, sogenannten Referenzmodelle geeinigt. Die Mutter der Referenzmodelle ist das bekannte Open System Interconnection (OSI)-Referenzmodell, das auch von der International Standardization Organization (ISO) adaptiert wurde. Grundidee des Modells ist es, dass Geräte, Funktionen und Applikationen immer zwischen einzelnen Schichten ausgeführt werden. Eine Endapplikation arbeitet also bei beiden (Sender und Empfänger) auf der höchsten Schicht und übergibt die Daten “nach unten” an die niedrigere. Die eingehende Diskussion des Modells obliegt anderen Vorlesungen, und wird hier als bekannt vorausgesetzt. Erwähnung findet das OSI-Referenzmodell hier, um Orientierung zu geben. In dieser Vorlesung werden - bis auf wenige Ausnahmen - ausschließlich über die Schichten 1 (Bitübertragungsschich =Physical Layer) und 2 (Sicherungsschicht = Logical Link Control (LLC) und Medium Access Control (MAC)) sprechen. Die Aufgabe werden hier kurz umrissen [11]:

Bitübertragungsschicht hier werden die physikalischen Parameter und Verfahrensweisen eines Kommunikationssystems festgelegt. Hier geht es um die Repräsentation

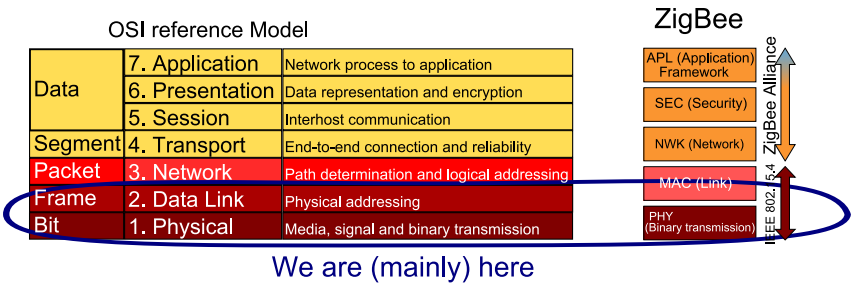


Abbildung 2.3: OSI-Referenzmodell und Markierung, wo wir uns befinden (noch zu überarbeiten)

des Bit als physikalische Größe. Dieses sind in der Regel Spannungs- bzw. elektromagnetische Leistungspegel, optische Signale. Es werden die Trägermedien, also die (Glasfaser-)Kabel, Antennen usw. definiert und dargelegt welchen zeitlichen und spektralen Verlauf ein gültiges Signal haben darf.

Die Details dieser Schicht dominieren diese Lehrveranstaltung

Sicherungsschicht Beschreibt die Art und Weise, wie mit logischen/ mathematischen Methoden die Daten so vorbreitet werden, dass sie optimal über den Physical Layer geschickt werden können. Hier geht es darum, wie evtl. Fehler in der Übertragung über einen Träger ausgeglichen und korrigiert werden können. Außerdem geschieht in dieser Schicht als MAC die Paketierung der Daten (auf der untersten Ebene) und die die Adressierung der beiden hier beteiligten Partner Sender und Empfänger.

Ein paar Aspekte, insbesondere die Fehlerkorrektur wie auch der Kanalzugriff werden in dieser Veranstaltung am Beispiel des Ethernet und dessen Umgebung besprochen.

2.3 Ethernet

Die Geschichte des Ethernet geht zurück in die 70er-Jahre und ist bspw. unter [8, 12] nachzulesen. Die Geschichte ist nicht Thema hier. Allerdings hat sich aus dem anfänglichen einfachen Standard, der nur wenige Computer miteinander vernetzen sollte, eine Technologie entwickelt, die heute - allerdings in deutlich unterschiedlicher Ausprägung - zur Peer-to-Peer-Vernetzung zweier Computer wie auch im Weitverkehrsnetz als Carrier-Ethernet eingesetzt wird. Datenraten von wenigen Megabit bis zu 100 GigaBit pro Sekunde werden unterstützt, vollkommen unterschiedliche Technologien zur Datenübertragung

werden eingesetzt. Bevor wir tief in die Technologie einsteigen, soll eine kurze Auflistung gegeben werden (Tabelle 2.1).

Tabelle 2.1: Ausgewählte Physical Layer Parameter des Ethernet bis 100 Gb/s [8, 13, 14, 15, 16, 17, 18] mit den im Text eingeführten Bezeichungen, l_{max} ist die max. Länge einer Strecke ohne Repeater.

Bezeichnung	Bit/s	Übertragungsmedium	Topologie	l_{max}/m	Jahr/ Annex
10BASE5	10 M	RG-8 Koax	Bus	500	1983/
10BASE2	10 M	RG-58 Koax	Bus	185	1985/a
10BROAD36[1]	10 M	Koax 75 Ω	P2P	2800	1985/b
10BASE-T	10 M	TP	P2P	100	198X/XXX
10BASE-FP	10 M	LWL	Stern	500	1987/d
10BASE-FB	10 M	LWL	Stern	2000	1987/d
10BASE-FL	10 M	LWL	Stern	2000	1987/d
100BASE-TX	100 M	UTP	P2P	200-400[2]	1995/u
100BASE-T2	100 M	TP[3]	Stern	100	1995/x,y
100BASE-T4	100 M	TP[4]	Stern	100	1995/x,y
100BASE-FX	100 M	LWL (SC)	Stern	200-400[2]	1995/x,y
100BASE-LX10	100 M	LWL 1310 nm	P2P	0,5-10k	sec-5
100BASE-BX10	100 M	LWL[5]	P2P	0,5-10k	sec-5
1000BASE-T	1 G	TP[6]	Stern	>100	1998/z
1000BASE-SX	1 G	LWL 850 nm	Stern	550	1998/z
1000BASE-LX	1 G	LWL 1310 nm	Stern	550/5000[7]	1998/z
1000BASE-CX	1 G	STP[8]	Stern	550/5000[7]	1998/z
1000BASE-PX10	1 G	LWL[9]	P2MP	0,5-20k	sec-5
1000BASE-PX20	1 G	LWL[9]	P2MP	0,5-20k	sec-5
10GBASE-T	10 G	TP	Stern	100	2006/an
10GBASE-SR/W[10]	10 G	LWL 850 nm	Stern	300	2002/ae
10GBASE-LR/W[10]	10 G	LWL 1310 nm	Stern	10 000	2002/ae
10GBASE-ER/W[10]	10 G	LWL 1550 nm	Stern	40 000	2002/ae
10GBASE-CX4	10 G	Twinax	P2P	????	20??/an
10GBASE-KX4	10 G	BP(4 Lanes)	P2P	1	2004/ap
10GBASE-KR	10 G	BP (1 Lane)	P2P	1	2004/ap

Weiter auf der nächsten Seite

Bezeichnung	Bit/s	Übertragungsmedium	Topologie	l_{max}/m	Jahr/ Annex
40GBASE-KR4[11]	40 G	BP	P2P	1	2010/ba
40GBASE-CR4	40 G	4×Twinax[12]	P2P	7	2010/ba
100GBASE-CR10	100 G	10×Twinax[12]	P2P	7	2010/ba
40GBASE-SR4	40 G	4×LWL 850 nm (MMF)	Stern	150	2010/ba
100GBASE-SR10	100 G	10×LWL 850 nm (MMF)	Stern	100	2010/ba
40GBASE-LR4	40 G	4×LWL 1310 nm (SMF)	Stern	10 000	2010/ba
100GBASE-LR4	100 G	4×LWL 1310 nm (SMF)	Stern	10 000	2010/ba
100GBASE-ER4	100 G	4×LWL 1550 nm (SMF)	Stern	40 000	2010/ba

Als Abkürzungen wurden hier verwendet

TP Twisted Pair (Cable) (TP) meint ein Kabel mit einer üblicherweise Wellenimpedanz von 100, 120, oder 150 Ω (je nach Spezifikation), in der die für jede einzelne Übertragungsrichtung notwendigen zwei Adern miteinander verdrillt sind. Dieses verbessert die Abstrahleigenschaften und immunisiert ein wenig gegen Einstrahlungen.

UTP Unshielded Twisted Pair (Cable) (UTP) spezifiziert, dass die beiden Adern eines TP nicht geschirmt sein müssen. Die bekannten und meistverwendeten Cat 5 Kabel sind häufig solche ungeschirmten Kabel.

STP Shielded Twisted Pair (Cable) (STP) verwendet einzeln gegeneinander abgeschirmte Adernpaare.

BP Backplane (BP) ist die Rückseitenverdrahtung in einem Gerät, das mehrere Baugruppen trägt.

LWL Lichtwellenleiter (LWL) ist eben der optische Wellenleiter.

P2MP Point to Multipoint (P2MP) bedeutet eine Verbindung von einem Sternpunkt auf viele andere Endpunkte. Hier ist der Sternpunkt allerdings nicht aktiv und verteilt tatsächlich die Daten, sondern er stellt lediglich eine einfache Verzweigung dar. Es handelt sich also um einen Splitter. Der perfekte Punkt zu Mehrpunktverteiler wäre eine Antenne, die in viele Richtungen aussendet.

P2P Point to Point (P2P) bzw. Stern bezeichnet die Topologie des Netzwerkes und meint, dass immer nur ein Sender/ Empfänger mit einer Gegenstelle verbunden ist. Einer der Partner ist idR. ein Sternpunkt, der mehrere Endpunkte oder weitere Sternpunkte miteinander verbindet. Dieses steht im Gegensatz zum

Bus in dem mehrere Stationen auf einen solchen, in der Regel ist das ein koaxiales Kabel, zugreifen. Damit teilen sich mehrere Stationen ein Medium und müssen sich irgendwie über dessen Nutzung einig werden. Weiterhin ist jeder Abgriff von dem Bus eine Störung, mit der das System klar kommen muss.

Und zudem sind in der Tabelle folgende Bemerkungen zu beachten:

1. Dieser ist ein Exot unter den Standards, denn es wird eine modulierte Übertragung verwendet und das Medium kann (auf anderen Frequenzbereichen) noch für weitere Dienste genutzt werden.
2. Nominelle Länge des Netzwerkes inkl. Repeater etc., sonst auch nur 100 m.
3. Voll Duplex, selten implementiert, lebt als Technologie bei 1 Gb-Übertragung fort.
4. verwendet alle vier Adernpaare und ist so gut wie nicht implementiert.
5. Bidirektional mit 1550 nm im Downlink und 1310 nm im Uplink, also bidirektional über eine Faser
6. Die vier Adern des "üblichen" Ethernet Cat-5 Kabels werden für jeweils 250 Mb/s genutzt, und zwar gleichzeitig im Voll-Duplex Modus.
7. Mit Einmodenfasern (Single Mode Fiber [Einmodenfaser] (SMF))
8. Belegung ein Adernpaar je Richtung, daher inkompatibel zu 1000BASE-T
9. Bidrektional auf einer Faser mit 1310 nm Uplink und 1490 nm Downlink für Passive Optical Networks (PON)
10. R=Local Area Network (LAN) und W=WAN
11. Siehe auch 10GBASE-KR
12. Twinax-Kable (Cat7-Kabel) haben einzeln geschirmte Adernpaare.

Noch einen Blick auf die Ethernet-Nomenklatur werfend, erkennen wir, dass sich der Name zusammensetzt aus der Übertragungsrate, der Art der Übertragung (bis auf eine Ausnahme ist das Basisband-Übertragung) und der physikalischen Verbindung (also "T" für Twisted Pair, "S,L,E" für kurze (betrieben mit Licht der Wellenlänge von 850 nm), lange (Wellenlänge 1310 nm) und sehr lange (Wellenlänge 1550 nm) optische Verbindung, "K" für die Backplane "R" die LAN und "W" die WAN-Variante).

Anwendung	Ethernet	Alternative
Backplane	10GBASE-K	PCI-Express 3.0 (3 Gb/s) [19]
P2P Geräte	10-1000Base-T	USB (3.0: 5 Gb/s) [20]
WAN	10-100GBASE-...	SDH/ SONET (ITU G.783/803 [21, 22])
Access (optisch)	10GBASE-P...	PON (ITU G.983.1 [23])

Tabelle 2.2: Alternativen zum Ethernet für verschiedene Anwendungsfälle (Auswahl)

Wie zu erkennen ist, deckt der Ethernet-Standard eine Vielzahl unterschiedlicher, drahtgebundener Kommunikationsvarianten ab. Von der Kurzstreckenübertragung zwischen einzelnen Chips oder Baugruppen innerhalb eines Netzelementes über Verbindung einzelner Geräte, Kommunikation innerhalb eines Büros oder auf einem Campus bis hin zur Weitverkehrstechnik, in der große Datenmengen von einer Stadt oder Region in eine andere gebracht werden müssen ist alles vertreten. Für unterschiedliche Zwecke werden auch unterschiedliche Träger verwendet. Ist es im Nahbereich und für geringe Datenraten die Kupferleitung, so sind für hohe Datenmengen und große Entfernungen nur noch spezielle Glasfaserkabel, geeignet. Alle diese Träger und die richtige Anwendung derer wollen wir uns im Weiteren detailliert anschauen.
Die verschiedenen Varianten des Ethernet stehen hier exemplarisch für eine Vielzahl von leitungsgebundenen Kommunikationslösungen. Das Ethernet ist - wenn auch weit verbreitet - nicht einmal in jedem Fall die beste Lösung. Alternativen für verschiedene Anwendungsfälle sind bspw. in Tab. 2.2 angegeben.
Allen gemeinsam ist aber, dass ähnliche oder gleiche Kanäle (Kabel) verwendet werden und daher oft vergleichbare Lösungen verfolgt werden, so dass es sich lohnt einige dieser teilweise sehr umfangreichen und komplizierten Techniken auch von einem theoretischen Punkte her besser kennen zu lernen.

2.3.1 Referenzmodell einer Ethernetvariante: 100 GBase-...

Die schon oben unter Abschnitt 2.2.2 Schichten im Referenzmodell werden für die meisten komplexeren Protokolle weiter aufgeteilt. Dieses ist im Bild 2.4 dargestellt. So teilt sich der uns vornehmlich interessierende “physical layer” auf in

Reconciliation Reconciliation Sublayer (RS) ist eine der Unterschichten. Sie definiert den (logischen) Zusammenhang zwischen der MAC und der Physikalischen Schicht, bzw. genauer dem Medium Independent Interface (MII). Die RS ist im Prinzip

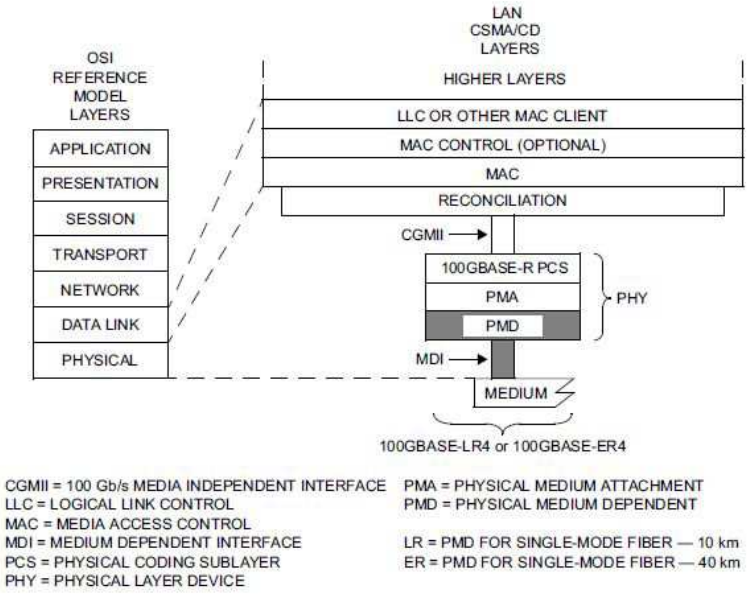


Abbildung 2.4: Referenzmodell des 100GBASE-LR/ER Substandards nach [18]

Foto von MAU einfügen.

Abbildung 2.5: Media Access Unit (MAU) aus “alten Zeiten” für 10BASE-…

logisch eine Summe von Funktionsaufrufen, die die bedienenden Chips dazu veranlassen Daten und Signale auf den Bus zu schicken oder von dort abzunehmen.

MII ist heute in diesem Standard nur eine logische Schicht. In früheren (und langsameren) Versionen des Ethernetstandards war diese Schnittstelle tatsächlich als Leitung existent. Es handelte sich um eine 15-polige Leitung, angeschlossen mit D-Sub15 Steckern und Buchsen, die in einer kleines Box mündeten, die die letzte Umsetzung auf das Medium machten. Bei 10BASE-…, wo diese Technik eingesetzt wurde, ging das MII dann auf das Koax- oder das Twisted-Pair Kabel, je nachdem, was man wollte (Bild 2.5).

Bei neueren “Netzwerkkarten” (bzw. Ethernetkarten) und auch im Standard ist diese Schicht nur noch eine logische Einheit, die man tatsächlich realisieren kann, aber nicht muss. Für die hier betrachteten Datenraten von 40 bzw. 100 Gb/s wird die Schnittstelle XLGMII oder CLGMII genannt, wobei “G” für Gigabit steht und der Rest (inklusive der römischen Ziffern am Anfang) erklärt sein sollte. Für beide Richtungen werden 64 Bit parallel in 8×8 “Lanes” übertragen dazu kommen noch 8 Steuerungssignale/Leitungen und eine Taktleitung. Wir sprechen also über 146

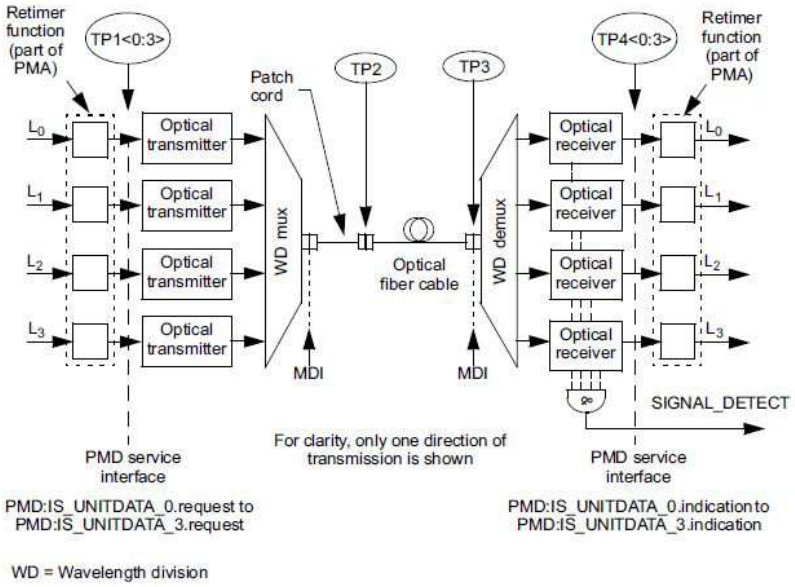


Abbildung 2.6: Blockdiagramm der PMD (mit Verbindung zum PMA) für 100GBASE-LR4

Leitungen insgesamt.

PCS im Physical Coding Sublayer (PCS) findet die Kodierung, also die Abbildung der oberhalb noch logisch als Daten vorhandenen Signale auf die Symbole, die tatsächlich übertragen werden sollen, statt. Namentlich wird hier die Umsetzung von 64Bit auf 66Bit (64B/66B) vorgenommen. Dieses und die in anderen Varianten des Standards (optional) hier ansässigen Forward Error Correction (FEC) Mechanismen werden in dieser Vorlesung einigen Raum einnehmen.

PMA und PMD , also Physical Medium Attachment (PMA) und Physical Medium Dependent (PMD) stellen nun die Verbindung zur physikalischen Medium, der Glasfaser dar. Ganz praktisch ausgedrückt sind dieses vier Laserdioden zum Aussenden und vier Fotodioden zum Empfang der Signale. In der Verbindung zur Faser ist noch ein Multiplexer, bzw. De-Multiplexer anzubringen (Abb. 2.6), der die vier verschiedenen „Farben“ voneinander trennt. Im Endeffekt werden die Daten in vier Strömen auf vier verschiedene „Farben“im Infraroten parallel in die Glasfaser eingespeist.

2.3.2 Weitere “Referenzmodelle”

…Hier evtl. noch ein noch detaillierteres Referenzmodell einbauen, in dem schon die einzelnen Bausteine des Transceivers erkennbar sind.

2.4 Lernkontrolle

1. Nennen Sie die Deutsche Behörde deren Regularien bzgl. Frequenznutzung und Telekommunikationsdienste bindend sind!

2. In welchen Gremien werden wesentliche Normen zur Telekommunikation auf internationaler Ebene erarbeitet? In welcher Institution/ Organisation einigt man sich auf internationale Normen?

3. Welche Berufsverband bringt maßgeblich Impulse in der Informatinostechnologie und in welcher Nomenklatur werden diese veröffentlicht?

4. Nennen Sie die Schichten des OSI-Referenzmodells und wie sind die unteren Schichten im Ethernet detailliert aufgeteilt?

5. Wie wird bei 40 GigE und 100 GigE die Datenrate im Vergleich zum 10 GigE übertragen?

6. Welche Wellenlängenbereiche werden in der optischen Nachrichtentechnik im Ethernet verwendet und für welche Übertragungsstrecken werden sie typischerweise eingesetzt?

Kapitel 3

Leitungen und Pegel

3.1 Pegelmaße

In der Nachrichtentechnik sind oft Signale über einen weiten Wertebereich darzustellen. Dieses rührt insbesondere daher, dass am Ausgang eines Senders (im Funk ganz extrem) üblicherweise recht hohe Leistungen vorliegen. Hier sind einige Watt üblich, aber sogar hunderte Kilowatt auch möglich. Am Empfänger kommen diese Signale allerdings nur stark gedämpft an. Hier liegen Pegel von Mikrowatt oder sogar nur Picowatt vor. Diesen Wertebereich von nun siebzehn Zehnerpotenzen zu erfassen, ist schwierig, daher hat man als griffigeres Darstellungsinstrument die Pseudoeinheit Bel erfunden, um noch schönere Werte zu bekommen, wird meist sogar nur das Dezi-Bel (dB) verwendet. Es handelt sich dabei um ein Verhältnismaß, das immer Leistungsverhältnisse darstellt, auch wenn manchmal Spannungen gemeint sind.

$$\left.\frac{P_2}{P_1}\right|_{\mathrm{dB}} = 10 \times \log_{10}\left(\frac{P_2}{P_1}\right). \tag{3.1}$$

Hier ist natürlich der Logarithmus zur Basis Zehn gemeint. Den Index Zehn werden wir in Zukunft weglassen. Bekanntlich stehen Spannung und Leistung, die diese Spannung an einen Wirkwiderstand abgibt im Zusammenhang

$$P = \frac{U^2}{R} \tag{3.2}$$

zueinander. Setzt man dieses in (3.1) ein, so erhält man

$$\left.\frac{U_2}{U_1}\right|_{\mathrm{dB}} = 10 \times \log_{10}\left(\frac{U_2^2}{R} \times \frac{R}{U_1^2}\right) = 20 \times \log_{10}\left(\frac{U_2}{U_1}\right), \tag{3.3}$$

was hier zwar als Spannungsverhältnis dargestellt ist, effektiv aber das Verhältnis der Leistungen angibt, die diese beiden Spannungen bei ihrem Abfall über dem Widerstand R abgeben.

Wie zu sehen ist, werden hier nur Verhältnisse angegeben, denn nur dieses macht Sinn. Würde man absolute Werte in dB angeben wollen, so käme man an die Frage, was denn der Logarithmus einer Einheit wäre. Dieses ist unsinnig. Um dennoch absolute Spannungen und Leistungen in Dezi-Bel darstellen zu können bedient man sich des Tricks, dass man die Leistung, bzw. Spannung einfach auf eine andere, frei gewählte bezieht. Diese Bezugsleistungen sind in der Regel 1 W, 1 mW, 1 V, 1 μW, so dass die Definitionsgleichungen

$$P|_{\mathrm{dBm}} = 10 \times \log\left(\frac{P}{1\,\mathrm{mW}}\right), \tag{3.4}$$

$$P|_{\mathrm{dBW}} = 10 \times \log\left(\frac{P}{1\,\mathrm{W}}\right), \tag{3.5}$$

$$U|_{\mathrm{dBV}} = 20 \times \log\left(\frac{U}{1\,\mathrm{V}}\right), \tag{3.6}$$

$$U|_{\mathrm{dB}\mu\mathrm{V}} = 20 \times \log\left(\frac{U}{1\,\mu\mathrm{V}}\right) \tag{3.7}$$

$$\tag{3.8}$$

folgen. Einige Beispielwerte und einfache Berechnungsregeln sind in Tab. 3.1 dargestellt. Um ein Spannungsverhältnis in ein Leistungsverhältnis umzurechnen, ist es grundsätzlich erforderlich, den Widerstand zu kennen, über dem die Spannung abfällt, bzw. an dem die Leistung umgesetzt wird. Ist dieser bekannt, so kann man mit (3.2) und

$$U = 1\,\mu\mathrm{V} \times 10^{U|_{\mathrm{dB}\mu\mathrm{V}}/20} \tag{3.9}$$

Tabelle 3.1: Einige Charakteristische Werte der dB Darstellung für Leistungen (P) und Spannungen (U)

dB	Summe/Differenz von 10,5,3	Mult., Div.	Linear(P)	Linear(U)
0	Memorize		1	1
1	$10-3-3-3$	$10/2/2/2$	1.25	
2	$5-3$	$3/2$	1.5	1.25
3	Memorize		2	
4	$10-3-3$	$10/2/2$	2.5	1.5
5	Memorize		3	
6	$3+3$	$2*2$	4	2
7	$10-3$	$10/2$	5	
8	$5+3$	$3*2$	6	2.5
9	$3+3+3$	$2*2*2$	8	
10	Memorize		10	3

berechnen:

$$P|_{dBm} = 10 \times \log\left[\frac{(1\,\mu\mathrm{V})^2}{1\,\mathrm{mW}} \times \frac{1}{R} \times \left(10^{U|_{\mathrm{dB}\mu\mathrm{V}}/20}\right)^2\right] \tag{3.10}$$

$$= 10\left[\log\left(\frac{(1\,\mu\mathrm{V})^2}{1\,\mathrm{mW}} \times \frac{1}{R}\right) + 2 \times \log 10^{U|_{\mathrm{dB}\mu\mathrm{V}}/20}\right] \tag{3.11}$$

$$= 10\log\left(\frac{(1\,\mu\mathrm{V})^2}{1\,\mathrm{mW}} \times \frac{1}{R}\right) + \frac{20}{20} \times \log 10^{U|_{\mathrm{dB}\mu\mathrm{V}}} \tag{3.12}$$

$$= 10\log\left(\frac{(1\,\mu\mathrm{V})^2}{1\,\mathrm{mW}} \times \frac{1}{R}\right) + U|_{\mathrm{dB}\mu\mathrm{V}} \tag{3.13}$$

$$= C + U|_{\mathrm{dB}\mu\mathrm{V}} \tag{3.14}$$

$$\tag{3.15}$$

Wobei der Teil links in der Formel als ein konstanter Faktor betrachtet werden kann, der nur von dem Wirkwiderstand und den Bezugsgrößen der Bel für Spannung und Ströme abhängt, also davon, ob wir mit dBm, dBW, dBμV oder dBV rechnen. In Tab. 3.2 sind einige Umrechungswerte für typische Widerstände angegeben.

3.1.1 Lernkontrolle

1. Welche Leistung in dBm und in W wird in einem 1 Ω-Widerstand bei einer Spannung von 100 dBV umgesetzt? (Hinweis: Das ist sehr viel.)

Tabelle 3.2: Umrechung von dBW und dBV und Ähnliches für verschiedenen Bezugswiderstände

R/Ω	$C(\mathrm{dB}\mu\mathrm{V} \to \mathrm{dBm})$	$C(\mathrm{dBV} \to \mathrm{dBW})$	$C(\mathrm{dBV} \to \mathrm{dBm})$
1	−90	0	30
50	−107	−17	13
100	−110	−20	10
120	−110,8	−20,8	9,2
150	−111,8	−21,8	8,2
600	−117,8	−27,8	2,2

2. Eine Leistung von 10 dBm wird um den Faktor 10 gedämpft. Wie groß ist sie dann?

3. Am Eingang einer Glasfaser werden 17 dBm eingespeist. Am Ende der langen Leitung haben wir noch 100 μW, wie groß ist die Dämpfung der Leitung?

4. Eine Leistung von 100 nW wird in einem Eingangswiderstand von 120 Ω umgesetzt und dann mit einer Spannungsverstärkung von 30 dB verstärkt. Wie groß sind Leistung und Spannung am Ausgang an einer Impedanz von 50 Ω?

3.2 Leitungen allgemein

Leitungen haben in der Übertragungstechnik in der Regel Längen, die deutlich über eine Wellenlänge hinaus gehen, es gilt also $l \gg \lambda$. Damit grenzt sich die hier durchgeführte Betrachtung der Leitungen deutlich von anderen Themengebieten ab. In der Gleichstrom- und auch herkömmlichen Wechselstromtechnik geht man nämlich davon aus, dass die Leitungen kurz sind und Welleneigenschaften nicht betrachtet werden müssen. In der Hochfrequenztechnik, insbesondere hier bei Schaltungen, sind Leitungen vorhanden und drehen die Phase eines Signals, allerdings sind die Leitungslängen in der Größenordnung der Wellenlänge, je nach Frequenz aber meist sogar darunter.
In der Übetragungstechnik geht es um sehr lange Leitungen, deren hier bedeutsamste Eigenschaften die

- Phasendrehung: (ggf. ist es undefiniert in welcher Phasenlage ein Signal am Ende einer Leitung beobachtet werden kann)

- Dämpfung: Das Signal am Ende ist schwächer als am Eingang

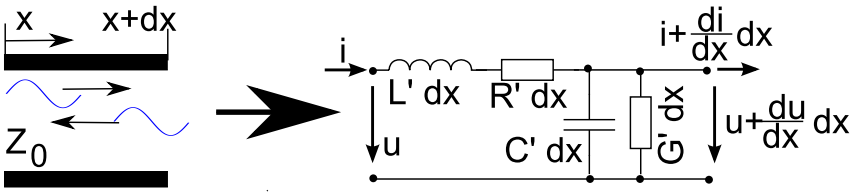


Abbildung 3.1: Ersatzschaltbild eines infinitesimal kurzen Leitungsstückes

- Laufzeit und Laufzeitverzerrungen (Dispersion): Aus einem schönen Impuls wird hierdurch ein verschwommenes Signal.

sind.

3.2.1 Ersatzschaltbild der Leitung

Trotz der Vorrede wird zunächst ein infinitesimal kurzes Leitungsstück in den Fokus gerückt (Abb. 3.1). Es besteht aus den

Kapazitiven Belägen C', die dadurch entstehen, dass zwei Leiter parallel geführt werden und dadurch natürlich Flächen mit unterschiedlicher Ladung gegenüber liegen, wodurch eine Ladungsspeicherung, wie im Kondensator, zustande kommt,

Induktiven Belägen L', weil natürlich durch einen fließenden Strom ein Magnetfeld aufgebaut wird, in dem Energie gespeichert ist,

Leitwert Belägen G', weil es immer der ein oder andere Ladungsträger schafft, durch ein Dielektrikum (und sei es auch nur Luft) von einem Leiter zum nächsten zu wandern, und den

Widerstands Belägen R', weil die metallischen Leiter ja auch ohmsche Verluste mit sich bringen.

3.2.2 Leitungsgleichungen

Auch mit den o.g. Spannungen und Strömen, bzw. den Spannungsabfällen und Stromflüssen können folgende Gleichungen aufgestellt werden:

- Für den Spannungsabfall an $L'dx$ und $R'dx$

$$-\frac{\partial u}{\partial x}dx = iR'dx + \frac{\partial i}{\partial t}L'dx \tag{3.16}$$

$$\Leftrightarrow \frac{\partial u}{\partial x} = -R'i - L'\frac{\partial i}{\partial t} \tag{3.17}$$

als erste Differentialgleichung.

- Für die Änderung des Stromes ergibt sich analog $G'dx$ und $C'dx$

$$-\frac{\partial i}{\partial x}dx = uG'dx + \frac{\partial u}{\partial t}C'dx \tag{3.18}$$

$$\frac{\partial i}{\partial x} = -G'u - C'\frac{\partial u}{\partial t} \tag{3.19}$$

als zweite Differentialgleichung.

- Zusammen werden diese die Telegraphengleichungen genannt.

Dieses System kann in eine einzige Gleichung überführt werden, indem man die erste Gleichung nach dem Ort, und die zweite nach der Zeit ableitet:

$$\frac{\partial^2 u}{\partial x^2} = -R'\frac{\partial i}{\partial x} - L'\frac{\partial}{\partial x}\frac{\partial i}{\partial t} \tag{3.20}$$

$$\frac{\partial}{\partial t}\frac{\partial i}{\partial x} = -G'\frac{\partial u}{\partial t} - C'\frac{\partial^2 u}{\partial t^2} \tag{3.21}$$

Und zusammen erhält man dann die Wellengleichung

$$\frac{\partial^2 u}{\partial x^2} = R'G'u + (R'C' + L'G')\frac{\partial u}{\partial t} + L'C'\frac{\partial^2 u}{\partial t^2} \tag{3.22}$$

und bei Vernachlässigung der Verluste durch G' und R' bleibt noch die Laplace-Gleichung

$$\frac{\partial^2 u}{\partial x^2} = L'C'\frac{\partial^2 u}{\partial t^2} \tag{3.23}$$

stehen. Letztere ist äußerst angenehm, da jede Funktion mit

$$u(t,x) = f\left(t - \frac{x}{v}\right) \tag{3.24}$$

die Funktion löst. Es ist dann nämlich

$$\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 f\left(t - \frac{x}{v}\right)}{\partial x^2} = \frac{1}{v^2}\frac{\partial^2 f(b)}{\partial b^2} \tag{3.25}$$

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial^2 f\left(t - \frac{x}{v}\right)}{\partial t^2} = \frac{\partial^2 f(b)}{\partial b^2} \tag{3.26}$$

wobei b einfach nur ein Platzhalter für das Argument ist (Stichwort: Äußere Ableitung). In (Gl. 3.23) eingesetzt hat man dann

$$\frac{1}{v^2}\frac{\partial^2 f(b)}{\partial b^2} = L'C' \times \frac{\partial^2 f(b)}{\partial b^2} \tag{3.27}$$

$$\Leftrightarrow \frac{1}{v^2} = L'C' \Leftrightarrow \qquad v = \pm\frac{1}{\sqrt{L'C'}}. \tag{3.28}$$

Hieraus lernt man zweierlei, nämlich

1. bleibt eine völlig beliebige Wellenform bei der Übertragung über die Leitung unverändert (L' und C' sind Konstanten). Die Leitung ist vollkommen dispersionsfrei, alle Frequenzanteile werden gleich übertragen und

2. ist die Größe v die Geschwindigkeit dieser Übertragung gemäß (Gl. 3.28) unabhängig von der Frequenz und nur abhängig von den o.g. konstanten Leitungsbelägen.

Das Verhältnis der Geschwindigkeit v zur Lichtgeschwindigkeit c im freien Raum bezeichnet man auch als Verkürzungsfaktor, er ist immer größer als eins.

3.2.3 Harmonische Lösung der Leitungsgleichung

Die Lösung, dass jedes Signal störungsfrei und nur - für alle Frequenzen gleich stark - verzögert durch die Leitung wandert, gilt in seiner absoluten Schönheit nur für die Leitung nach (Gl. 3.23). Schon bei einer verlustbehafteten Leitung nach (Gl. 3.21) sind die Verhältnisse verwickelter.
Ein Lösungsansatz ist hier eine spezielle Form der vorherigen Lösung nach (Gl. 3.24), die harmonische Schwingung mit

$$u(x,t) = u_0 \times e^{\pm(j\omega t+\gamma x)}. \tag{3.29}$$

Mit den partiellen Ableitungen (wir betrachten nur das negative Vorzeichen)

$$\frac{\partial^2 u(x,t)}{\partial x^2} = \gamma^2 u(x,t) \tag{3.30}$$

$$\frac{\partial u(x,t)}{\partial t} = -j\omega u(x,t) \tag{3.31}$$

$$\frac{\partial^2 u(x,t)}{\partial t^2} = -\omega^2 u(x,t) \tag{3.32}$$

ergibt sich dann aus (Gl. 3.21) unter Kürzung der $u(x,t)$—-Terme auf beiden Seiten

$$\gamma^2 = R'G' - j(R'C' + L'G')\omega - \omega^2 L'C' \tag{3.33}$$

$$\Leftrightarrow \gamma = \pm\sqrt{R'G' - j(R'C' + L'G')\omega - \omega^2 L'C'} \tag{3.34}$$

$$= \pm\sqrt{(R' + j\omega L')(G' + j\omega C')} \tag{3.35}$$

$$= \pm j\omega\sqrt{L'C'} \times \sqrt{1 + j\frac{R'C' + L'G'}{\omega L'C'} - \frac{R'G'}{\omega^2 L'C'}} \tag{3.36}$$

für den Ausbreitungskoeffizienten. Dieser wird oft auch gerne in

$$\gamma = \alpha + j\beta \tag{3.37}$$

in einen reellen Phasenkoeffizienten β und einen ebenfalls reellen Dämpfungskoeffizienten α aufgeteilt. Beide Teile hängen von der Frequenz ab. Somit beschreibt der Term $e^{-\alpha x}$ eine Abnahme einer (Spannungs-)Amplitude mit der Strecke. Der imaginäre Anteil ist verantwortlich für die Phasendrehung und auch hiermit wird dann die Ausbreitungsgeschwindigkeit der Wellenfront bestimmt. Leider sind die Koeffizienten nicht so leicht zu trennen. Betrachtet man die Dämpfung genauer, so erhält man mit

$$\begin{aligned} a|_{dB} &= 20 \times \log_{10}\left(e^{-\alpha x}\right) && (3.38)\\ &= 20 \times \log_{10}\left(10^{-\alpha x/\ln 10}\right) && (3.39)\\ &= -\frac{20}{\ln 10} \times \alpha x \approx 8{,}686\,\mathrm{dB} \times \alpha x && (3.40) \end{aligned}$$

die Dämpfung einer Leistung der Länge x in dB.
Für geringe Verluste, also $R' \ll \omega L'$ und $G' \ll \omega C'$ kann (Gl. 3.36) zu

$$\begin{aligned} \gamma &\approx \pm j\omega\sqrt{L'C'} \times \left(1 + j\frac{R'C' + L'G'}{2\omega L'C'} - \frac{R'G'}{2\omega^2 L'C'}\right) && (3.41)\\ &\approx \pm j\omega\sqrt{L'C'} \times \left(1 + j\frac{R'}{2\omega L'}\right) && (3.42)\\ &= \pm j\omega\sqrt{L'C'} \mp \frac{R'}{2} \times \sqrt{\frac{C'}{L'}} && (3.43) \end{aligned}$$

genähert werden. Letztere Zeilen gelten für vernachlässigbaren Leitwert durch das Dielektrikum $G' = 0$. Dieses gilt nur für geringe Verluste und dann - und nur dann - sind Geschwindigkeit der Wellenausbreitung

$$v = \omega/\mathrm{Im}\{\gamma\} = 1/\sqrt{L'C'} \tag{3.44}$$

und Dämpfung

$$\alpha = R'/2 \times \sqrt{C'/L'} \tag{3.45}$$

unabhängig von der Frequenz. Bei größeren Dämpfungen ist das nicht mehr so.
Diese Frequenzabhängigkeit ist schädlich, weil sie Impulse verfälscht. Der Effekt wird Dispersion genannt und wird uns im Weiteren noch verfolgen. Zu bemerken ist weiterhin, dass der Induktivitätsbelag im Nenner der Dämpfung steht. Eine Leitung mit höherer Induktivität hat also eine geringere Dämpfung und auch - leider - eine geringere Geschwindigkeit. Hinzufügen von Induktivitätsbelag ggf. auch als lokalisierte Spule würde die Dämpfung verringern.

Bei stärker verlustbehafteten Leitungen kann man durch geschicktes Einstellen der Parameter die Frequenzabhängigkeit vermindern oder sogar eliminieren. Man wählt

$$\frac{R'}{L'} = \frac{G'}{C'} \tag{3.46}$$

und erhält aus (Gl. 3.36)

$$\gamma = \pm j\omega\sqrt{L'C'} \times \sqrt{1 + \frac{2jR'}{\omega L'} - \left(\frac{R'}{\omega L'}\right)} \tag{3.47}$$

$$= \pm j\omega\sqrt{L'C'} \times \sqrt{\left(1 + \frac{jR'}{\omega L'}\right)^2} \tag{3.48}$$

$$= \pm j\omega\sqrt{L'C'} \times \left(1 + \frac{jR'}{\omega L'}\right) \tag{3.49}$$

$$= \pm j\omega\sqrt{L'C'} \mp R'\sqrt{\frac{C'}{L'}}. \tag{3.50}$$

auch ohne, dass der Leitwertsbelag verschwinden muss. Hiermit sind Dämpfung und Geschwindigkeit schon wieder unabhängig von der Frequenz, ein Impuls würde nicht verformt werden. Im Vergleich zu (Gl. 3.45) sind die Verluste verdoppelt, da nach Maßgabe von (Gl. 3.46) die dielektrischen Verluste und die Leistungsverluste gleich eingestellt wurden. In der Regel ist aber $R'/L' \gg G'/C'$. Eine Gleichheit erreicht man hier durch Hinzufügen von Induktivitätsbelag, ggf. auch lokalisiert als Pupinspule. Dieses hat man früher in Seekabeln, aber auch in Telefonleitungen gemacht [24]. Mit dem Aufkommen von ISDN und ADSL, die einen deutlichen höheren Frequenzbereich als herkömmliche Sprachtelefonie verwenden, hat sich die Methode als schädlich herausgestellt.

3.2.4 Wellenlänge und Wellenimpedanz

Neben den Ausbreitungseigenschaften einer Welle ist noch zu definieren, in welchem Verhältnis Strom und Spannung, bzw. allgemeiner elektrisches und magnetisches Feld zueinander stehen. Verhältnisse von Strom und Spannung werden üblicherweise in Form von Widerständen (bzw. allgemeiner) Impedanzen angegeben. Wir haben oben unter (gl. 3.36) zwei - eine positive und eine negative - Lösung für den Ausbreitungskoeffizienten γ erhalten, zu dessen Interpretation wir nun einmal die Verluste weglassen und damit

$$\gamma = \alpha + j\beta = 0 \pm j\omega\sqrt{L'C'} \tag{3.51}$$

erhalten. Eingesetzt in den harmonischen Ansatz nach (Gl. 3.29) und nur Betrachtung des Realteils erhält man

$$u(x,t) = u_0 \cos\left(\omega(t \pm \sqrt{L'C'}x)\right). \tag{3.52}$$

Das Verhalten einer Phasenfront sieht man, wenn man hier das Argument (es reicht der Teil $(t \pm \sqrt{L'C'}x)$ konstant, also bspw. $= 0$) hält. Die Zeit läuft nun und damit folgt für die Strecke, die diese Wellenfront zurück legt,

$$x = \mp \frac{1}{\sqrt{L'C'}} t = vt, \tag{3.53}$$

durch das Auflösen nach t ist das Vorzeichen gedreht worden und man sieht, dass für die negative Lösung von γ die Phasenfront in positive Richtung wandert (und umgekehrt). Mit anderen Worten: Oben wurden zwei Lösungen erhalten, eine für eine Welle mit Ausbreitung in positive, eine mit Ausbreitungsrichtung in negative Richtung. Die Gesamtlösung ist irgendeine Mischung aus beiden:

$$U(x) = U_0^+ e^{-\gamma x} + U_0^- e^{\gamma x} \tag{3.54}$$

$$I(x) = I_0^+ e^{-\gamma x} + I_0^- e^{\gamma x}. \tag{3.55}$$

Der Term γ meint nun nur noch die positive Lösung, die Vorzeichen werden extern ausgedrückt. Die Länge der Welle kann man „messen“, wenn man sie genau eine Periode $T = 1/f$ laufen lässt und man erhält

$$\lambda_l = \frac{1}{f\sqrt{L'C'}} = \frac{v}{f} \tag{3.56}$$

für die Wellenlänge auf der Leitung. Ein Vergleich mit den Bedingungen im freien Raum, wo ja - mit Lichtgeschwindigkeit c -

$$\lambda_0 = \frac{c}{f} \tag{3.57}$$

gilt, ergibt

$$\frac{\lambda_l}{\lambda_0} = \frac{v}{c} \, (< 1) \,, \tag{3.58}$$

den Verkürzungsfaktor, der in den hier meist betrachteten Fällen der (quasi) TEM-Welle kleiner als eins ist, und angibt, um wieviel die Welle auf der Leitung kürzer und langsamer als im freien Raum ist.
Den Zusammenhang für U aus (Gl. 3.54) eingesetzt in (Gl. 3.17) ergibt (nach Auflösen)

$$I(x) = \frac{\gamma}{R' + j\omega L'} \left(U_0^+ e^{-\gamma x} + U_0^- e^{\gamma x} \right) \tag{3.59}$$

und ein Vergleich mit dem Ausdruck für den Strom bringt

$$Z_0 = \frac{R' + j\omega L'}{\gamma} = \sqrt{\frac{R' + j\omega L'}{G' + j\omega C'}} \approx \sqrt{\frac{L'}{C'}}, \tag{3.60}$$

was man als die Wellenimpedanz bezeichnet und mit

$$\frac{U_0^+}{I_0^+} = Z_0 = -\frac{U_0^-}{I_0^-} \tag{3.61}$$

den Zusammenhang zwischen den einzelnen Wellenamplituden herstellt.
Dieser Wert ist reell, was zunächst überraschen mag, da doch die Leitung frei von Verlusten ist, und daher keine Energie verloren geht. Aus Sicht der Quelle geht bei einer unendlich langen Leitung - und nur die wird hier betrachtet - sehr wohl Energie verloren, denn die Quelle liefert elektrische Energie in die Leitung und - da diese unendlich lang ist - sieht die Quelle die Energie niemals wieder, es wird also tatsächlich Arbeit verrichtet.

3.2.5 Lösung der Feldgleichungen im statischen Fall: Parameter von TEM-Leitungen

Die oben gemachten Ausführungen beruhen auf einer idealisierten Leitung, von der alle vier Parameter C',G',L',R' bekannt sind. Während man ggf. G' und R' noch mittels Multimeter messen könnte, sind die anderen Parameter recht schwer zugänglich. Für eine Berechnung gilt Ähnliches. Das übliche Berechnungsverfahren ist für die hier interessierenden Leitungen die Methode der konformen Abbildungen, die z.B. in [25] beschrieben ist. Es wird nur das Ergebnis in Tab. 3.3 angegeben. Die charakteristische Impedanz eines homogenen Dielektrikum ist

$$Z_0 = \sqrt{\frac{\mu_0}{\epsilon_r \epsilon_0}} \approx \frac{1}{\sqrt{\epsilon_r}} 120\pi\,\Omega \approx \frac{1}{\sqrt{\epsilon_r}} 377\,\Omega \tag{3.62}$$

Außerdem ist die Oberflächenimpedanz des Materials

$$Z_m = R_m + jX_m = \frac{1+j}{\sigma\delta_s} \tag{3.63}$$

$$\delta_s = \sqrt{\frac{2}{\omega\mu_0\sigma}} \tag{3.64}$$

$$\Rightarrow R_m = \sqrt{\frac{\sigma}{2\omega\mu_0}} \tag{3.65}$$

verwendet worden, wobei δ_s die Skin-Eindringtiefe ist und σ die Leitfähigkeit des Metalls beschreibt. Da R_m schon frequenzabhängig ist, wird auch jeder daraus resultierende Dämpfungsfaktor frequenzabhängig sein.
In Tab. 3.4 sind Beispiele für tatsächlich verfügbare Kabel aufgeführt

Tabelle 3.3: Leitungsparameter

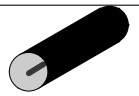
Wert	Koaxleitung	Paralleldrahtleitung
		
Konf. Abb.	$W = \ln(Z)$	$W = \ln\left(\frac{Z+a}{Z-a}\right)$
Z_c	$\frac{Z_0}{2\pi}\ln\left(\frac{b}{a}\right)$	$\frac{Z_0}{\pi}\operatorname{arcosh}\left(\frac{D}{2R}\right)$
β	$\omega\sqrt{\epsilon_r\epsilon_0\mu_0}$	$\omega\sqrt{\epsilon_r\epsilon_0\mu_0}$
α	$\frac{R_m}{4\pi Z_c}\left(\frac{1}{a}+\frac{1}{b}\right)$	$\frac{R_m}{2\pi Z_c R}\frac{\frac{D}{2R}}{\sqrt{\left(\frac{D}{2R}\right)^2-1}}$

Tabelle 3.4: Beispiel für RG58 und Cat5-Kabel (Datenblatt aus Internet)

RG-58-Koax		Cat5-e TP	
Z_0	50 Ω	Z_0 1…100 MHz	$100 \pm 15\,\Omega$
DC-Innenleiter R	48,56 Ω/km	DC-loop R	188 Ω/km
Kapazitätsb.	101 nF/km	Kapazitätsb.	52 nF/km
Verkürzungsf.	0,66	Verkürzungsf.	0,64
		Laufzeit	530 ns/100 m
Dielektrikum	PE	Dielektrikum	PE (d=1 mm)
Leiterdurchmesser	0,9 mm	Leiter	AWG24 (0,5 mm)
Aussenleiter d	2,95 mm		

3.2.6 Dispersion und das Drumherum

Schon oben wurde wiederholt ausgeführt, dass einige Parameter der Wellenausbreitung auf Leitungen abhängig von der Frequenz sind. Dieses Phänomen erlebt man am eindrucksvollsten am Regenbogen (Abb. 3.2), an dem die einzelnen Farben aus unterschiedlicher Richtung am Auge ankommen. Und das deshalb, weil die einzelnen Wellenlängen (Frequenzen) der parallelen Strahlen von der Sonne in den Regentröpfchen unterschiedlich stark abgelenkt (gebrochen) werden. Die blauen etwas mehr, die roten etwas weniger. Das macht den Unterschied. Dieser Effekt wird Dispersion genannt. Allgemein sind einfach unterschiedliche Frequenzkomponenten eines Signals unterschiedlich schnell. Geschieht dieses beispielsweise auf einer Leitung zur Nachrichtenübertragung, auf der Pulse vorliegen, so werden diese verwaschen. Denn Pulse bestehen aus vielen unterschiedlichen Frequenzkomponenten, die die steile Flanke ausmachen. Werden diese einzelnen Komponenten unterschiedlich gedämpft und mit unterschiedlicher Geschwindigkeit geführt, so verschwimmen

Abbildung 3.2: Regenbogen über einem Dorf nördlich Karlsruhe

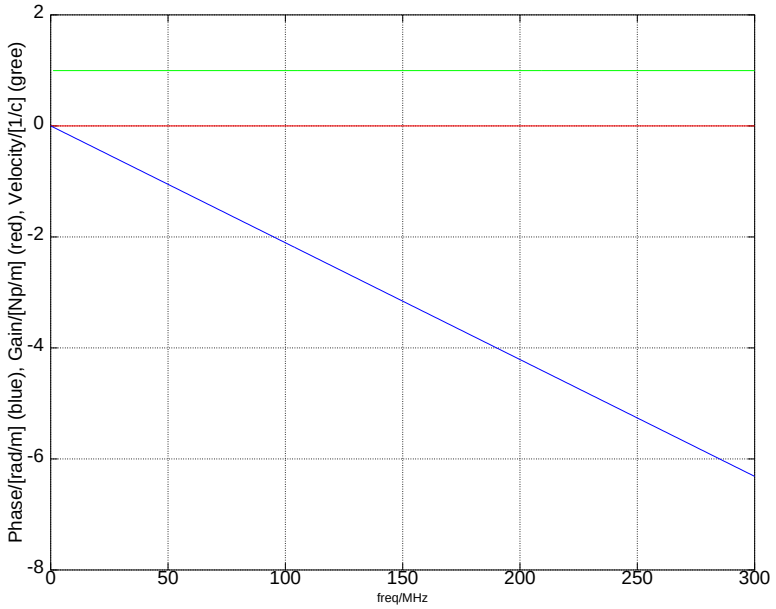
die schönen steilen Flanken irgendwann zu Brei. Dieser Effekt soll nun näher untersucht werden.

3.2.6.1 Gruppen- und Phasengeschwindigkeit

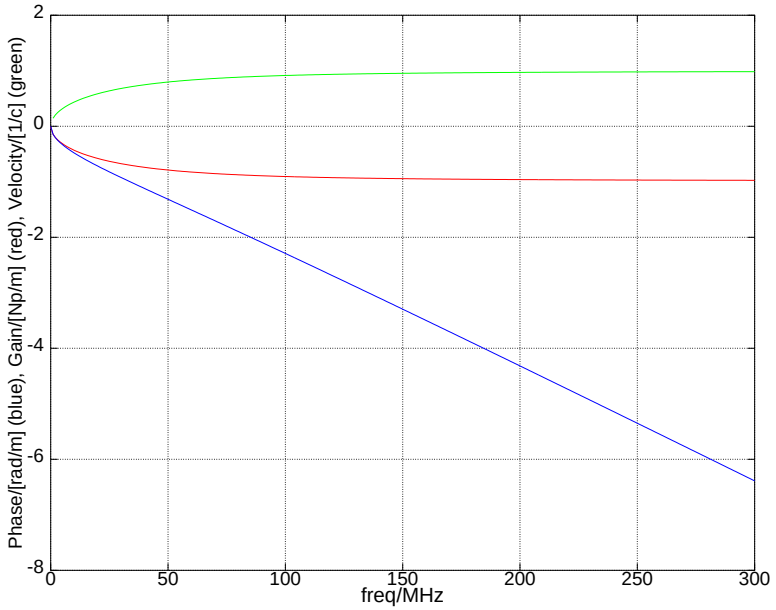
Phasengeschwindigkeit Um Laufzeitverzerrungen besser zu verstehen, wird nun die Ausbreitung von Wellen auf Leitungen in einfachen Beispielen näher betrachtet. Man geht hier (noch) davon aus, dass es sich um den eingeschwungenen Zustand handelt, dass also die Welle auf der Leitung schon immer da war und immer da sein wird. Von wirklicher Signalausbreitung ist demnach noch keine Rede.
Der einfachste Fall ist die Ausbreitung einer wandernden Welle und hier die Betrachtung der Bewegung einer Phasenfront, was im Prinzip schon in (Gl. 3.28) und (Gl. 3.44) vorgenommen wurde. Für den gesamten Phasenterm nach (G. 3.36) kann man im verlustfreien Fall

$$\mathrm{Im}\{\gamma\} = \omega\sqrt{L'C'} \tag{3.66}$$

schreiben und ist in Abb. 3.3(a) nochmals dargestellt. Es ist ersichtlich, dass die Phase linear mit der Frequenz abfällt und bei 300 MHz ca. -2π aufweist. Hierin steckt die Aussage, dass sich die Phase der Welle auf einen Meter um genau 360° bzw. 2π ändert, und damit die Welle wohl genau einen Meter lang sein muss. Diese Leitung verhält sich also gegenüber der Welle wie der freie Raum. Nach der bekannten Vorschrift (Gl. 3.44) $v_p = \omega/\mathrm{Im}\{\gamma\}$



(a)



(b)

Abbildung 3.3: Phasen-(blau), Dämpfungsverläufe (rot) und Phasengeschwindigkeit (grün) in Abhängigkeit von der Frequenz für (a) verlustfreie und (b) verlustbehaftete generische Leitungen

ergibt sich ein konstanter Wert v_p, der von nun an als Phasengeschwindigkeit bezeichnet wird und beschreibt, mit welcher Geschwindigkeit ein Nulldurchgang einer Welle (im eingeschwungenen Zustand) über die Leitung wandert. Dieser Wert kann kleiner, gleich oder sogar größer als die Lichtgeschwindigkeit $c = 299\,792\,458\,\mathrm{m/s} \approx 3 \times 10^8\,\mathrm{m/s}$ werden. Letztere Tatsache, also dass die Welle schneller als Licht werden kann, ist nicht etwa nobelpreisverdächtig, sondern rührt daher, dass die Welle schon eingeschwungen ist. Es wird hier keine Energie oder Information übertragen. In Abb. 3.3(b) ist der Fall einer verlustbehafteten Leitung gezeigt. Man erkennt sowohl die Dämpfung, als auch die frequenzabhängigen Phasenverläufe und die zugehörige Geschwindigkeit. In diesem Fall ist die Phasengeschwindigkeit immer kleiner als die Lichtgeschwindigkeit. Die Variation ist bei geringen Frequenzen größer, daher mag es sich anbieten, das Signal in der Frequenz zu versetzen.

Gruppengeschwindigkeit Was passiert aber nun, wenn man mit einem Wellenpaket, im einfachsten Fall also zwei sinusförmigen Signalen unterschiedlicher Frequenz an die Leitung herantritt?

$$u(t) = \mathrm{Re}\left\{u_0\left(e^{-j(\omega_1 t-\beta_1 x)} + e^{-j(\omega_2 t-\beta_2 x)}\right)\right\} \tag{3.67}$$

$$= \mathrm{Re}\left\{u_0 e^{-j\left(\frac{\omega_1+\omega_2}{2}t-\frac{\beta_1+\beta_2}{2}x\right)}\left(e^{-j\left(\frac{\omega_2-\omega_1}{2}t-\frac{\beta_2-\beta_1}{2}x\right)} + e^{j\left(\frac{\omega_2-\omega_1}{2}t-\frac{\beta_2-\beta_1}{2}x\right)}\right)\right\} \tag{3.68}$$

$$= 2u_0\mathrm{Re}\left\{e^{-j\left(\frac{\omega_1+\omega_2}{2}t-\frac{\beta_1+\beta_2}{2}x\right)}\right\}\cos\left(\frac{\omega_2-\omega_1}{2}t-\frac{\beta_2-\beta_1}{2}x\right) \tag{3.69}$$

Der letzte, im Kosinus versteckte Term bestimmt eine periodische Amplitudenschwankung mit der Kreisfrequenz $(\omega_2 - \omega_1)/2$, und nur die Ausbreitung dieser Schwankung wird betrachtet, da sie nämlich das Wellenpaket sehr gut beschreibt. Die Schwankung pflanzt sich nach

$$\frac{\omega_2-\omega_1}{2}t - \frac{\beta_2-\beta_1}{2}x = const \tag{3.70}$$

$$\Leftrightarrow x = \frac{\omega_2-\omega_1}{\beta_2-\beta_1}t + \frac{2\times const}{\beta_2-\beta_1} \tag{3.71}$$

fort. Ohne Verlust der Allgemeinheit kann man die Konstante zu Null wählen und erhält dann

$$x = \frac{\omega_2-\omega_1}{\beta_2-\beta_1}t \tag{3.72}$$

$$\Rightarrow v_g = \frac{dx}{dt} = \frac{d\omega}{d\beta} \tag{3.73}$$

für die Gruppengeschwindigkeit. Diese ist ein Maß dafür, wie sich ein Wellenpaket, bestehend aus zwei Sinussignalen mit kaum unterschiedlicher Frequenz auf der Leitung fortplanzt. Auch diese Geschwindigkeit kann kleiner, gleich oder größer als die Phasengeschwindigkeit sein. Die Gruppengeschwindigkeit kann sogar kleiner als Null werden. Das Wellenpaket wandert dann scheinbar den eigentlichen Wellen entgegen.
Mit dem bekannten Zusammenhang (Gl. 3.44) folgt schnell

$$v_g = \frac{d(v_p\beta)}{d\beta} = v_p + \beta\frac{dv_p}{d\beta} \tag{3.74}$$

und gibt einen Zusammenhang zwischen der Phasengeschwindigkeit und der Gruppengeschwindigkeit an.
Man mag sich nun fragen, mit welcher Geschwindigkeit Energie, Information oder eben echte Impulse übertragen werden. Das oben Geschriebene gilt ja nur im eingeschwungenen Zustand, wo alle Signale schon immer da waren und immer da sein werden. Wie man sich denken kann, drücken sich alle um diese zentrale Frage herum. Die Herausforderung besteht darin, zu definieren, was man als den Start des Pulses ansieht. Ist es das erste Eintreffen einer minimalen Leistung am Ende der Leitung? In dem Fall ist die Geschwindigkeit immer die Lichtgeschwindigkeit [26, S.333], denn ein (unendlich kleiner) Teil der Energie wird immer den Weg durch den freien Raum finden und sich somit mit Lichtgeschwindigkeit ausbreiten können. So richtig wie diese Antwort ist, so wenig praktikabel ist sie auch, denn niemand kann mit infinitesimal wenig Energie etwas anfangen. Definiert man eine Mindestschwelle, so kann man eigentlich nur noch sagen, dass die Übertragungsgeschwindigkeit geringer als die Lichtgeschwindigkeit sein wird, eine genauere Angabe ist mit erheblichem Denk- und Rechenaufwand verbunden, den wir hier nicht ausführen.

3.2.6.2 Definition der Dispersion und Konsequenzen: Die Breiwerdung des Impulses

Die Abhängigkeit der Phasengeschwindigkeit von der Frequenz nennt man Dispersion. Dispersion ist eigentlich die Regel, entsteht sie doch alleine schon durch Verluste auf der Leitung, die niemals zu vermeiden sind. Man unterscheidet drei Fälle:

Dispersionsfrei Hier ist die Phasengeschwindigkeit unabhängig von der Frequenz und damit sind Gruppengeschwindigkeit und Phasengeschwindigkeit gleich ($v_g = v_p$).

Normale Dispersion heißt es, wenn hohe Frequenzen langsamer als niedrige Frequenzen ein Medium durchqueren (blaues Licht ist langsamer als rotes). Dann ist die

Gruppengeschwindigkeit kleiner als die größte Phasengeschwindigkeit und kann sogar negativ werden. Dabei scheint das Wellenpaket dann rückwärts zu laufen. Allgemein ist hier $v_g < v_p$.

Anomale Dispersion meint den umgekehrten Fall, in dem die niedrigere Frequenz langsamer als die hohe ist, und damit $v_g > v_p$ gilt.

Da man es in der Kommunikationstechnik immer mit mehrfrequenten Signalen zu tun hat, in der Regel Pulse oder Änderungen von Signalen betrachtet werden müssen, ist Dispersion eine Herausforderung. Nach einer, durch die Leitungsparameter, gegebenen Strecke kann es nämlich zu einem Zerlaufen der Pulse kommen (Abb. 3.4).
Die einzelnen Parameter der Dispersion sind in Abb. 3.5 dargestellt. Hier ist noch der Parameter der Dispersion als

$$d^{-1} = \frac{d\omega}{d\tau_g} = \frac{d^2\omega}{d\beta^2} \tag{3.75}$$

dargestellt.

3.2.6.3 Kompensation der Dispersion: Zwei Methoden

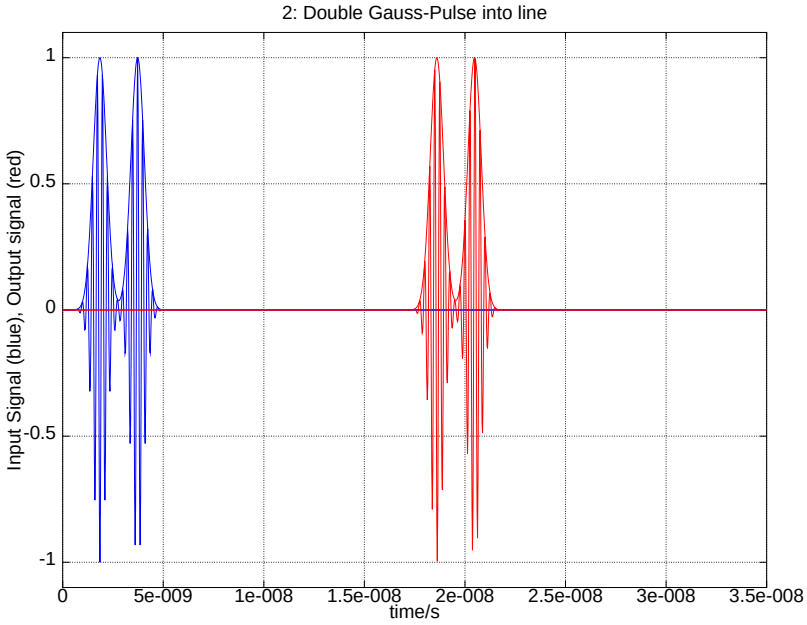
Das Angenehme ist, dass man eine bekannte Dispersion wieder zurückdrehen, also kompensieren kann.

Kompensation durch Rückdrehung, die optische Alternative Hier fügt man der eigentlichen Leitung eine Leitung mit genau umgekehrtem Dispersionsverhalten an. Das geht natürlich nur dann, wenn man das Ursprungsverhalten genau kennt. Mit anderen Worten: Man kann einen verbreiterten Impuls wieder schmaler machen, wenn man genau weiß, was diesen Puls breit gemacht hat. Man kann allerdings nicht allgemein einen breiten Puls in einen schmalen verwandeln, das muss scheitern. Eine Darstellung ist in Abb. 3.6 gezeigt. Diese Methode kann man bei Glasfasern recht gut anwenden.

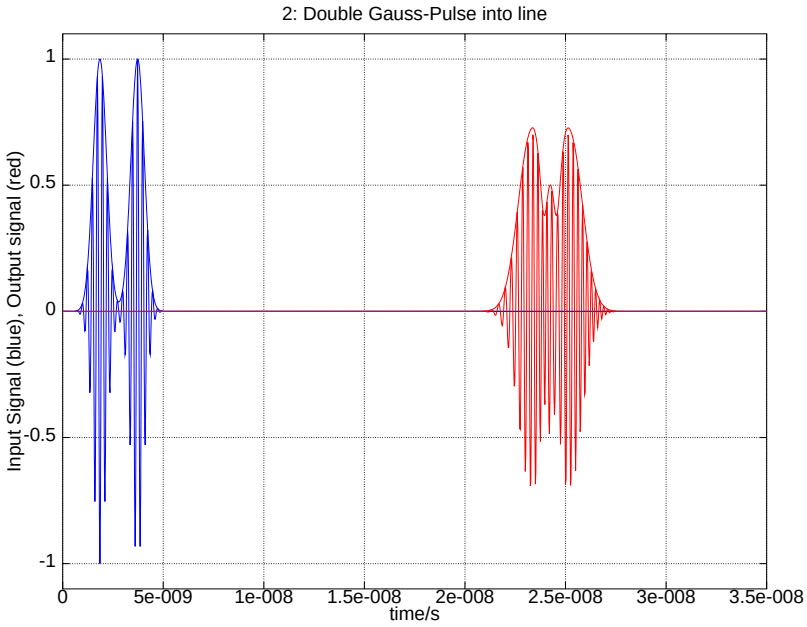
Kompensation durch Induktive Belastung Bei elektrischen Leitern ist es möglich, wie schon in (Gl. 3.50) gezeigt, durch gezieltes Einstellen insbesondere der Induktivität der Leitungskonfiguration die Dispersion zu minimieren.

3.3 Optische Wellenleiter

Die wohl bei weitem meisten Datenkilometer werden heutzutage in Form von Lichtsignalen in LWL zurückgelegt. Wahrscheinlich kann man behaupten, dass jedes übertragene Bit



(a)



(b)

Abbildung 3.4: Beispielhafte Darstellung der Dispersion bei zwei aufeinanderfolgenden Gauss'schen Pulsen (a) ohne Dispersion und (b) mit Dispersion. Annahme: Verlustfrei

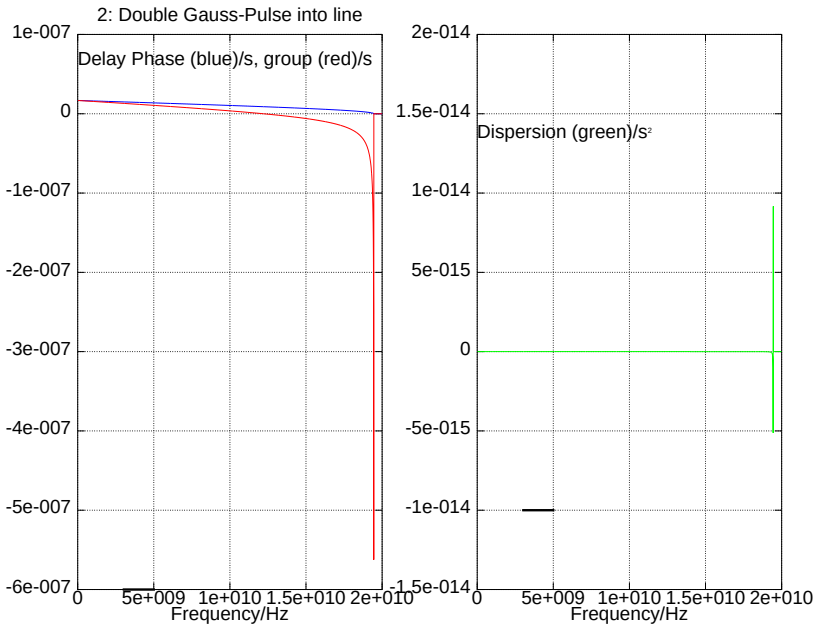


Abbildung 3.5: Darstellung der Gruppen und Phasengeschwindigkeiten (a) und der Laufzeiten, sowie der Dispersion in (b) für das Beispiel nach Abb. 3.4(b).

auf seiner Strecke vom ersten Erzeuger der Daten zur letzten Senke irgendwann einmal über einen LWL läuft. Daher wollen wir diesen etwas genauer betrachten.

3.3.1 Grundsätzliche Funktion

Die grundsätzliche Funktion beruht auf dem Effekt der Totalreflexion (Abb. 3.8) an Grenzschichten. Dieses ist z.B. bei einem Lichstrahl aus Wasser heraus an der Grenzschicht zur Luft gut bekannt. Nach dem Snellius'schen Brechungsgesetz gilt doch

$$\frac{\sin \alpha_1}{\sin \alpha_2} = \frac{n_2}{n_1}, \tag{3.76}$$

wobei hier bspw. $n_1 > n_2$ gilt und dann ist

$$\sin \alpha_2 = \frac{n_1}{n_2} \sin \alpha_1 \tag{3.77}$$

und nur für

$$\frac{n_1}{n_2} \sin \alpha_1 < 1 \tag{3.78}$$

$$\Leftrightarrow \sin \alpha_1 < \frac{n_2}{n_1} \tag{3.79}$$

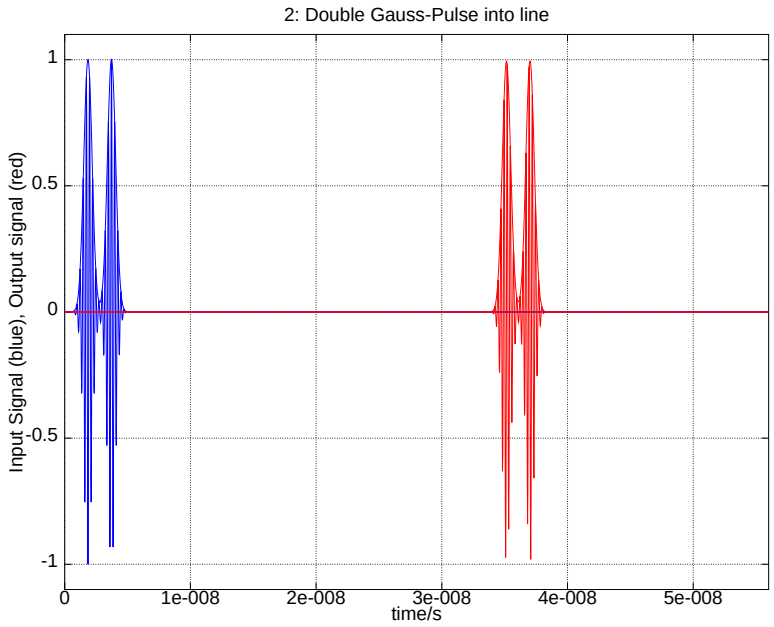


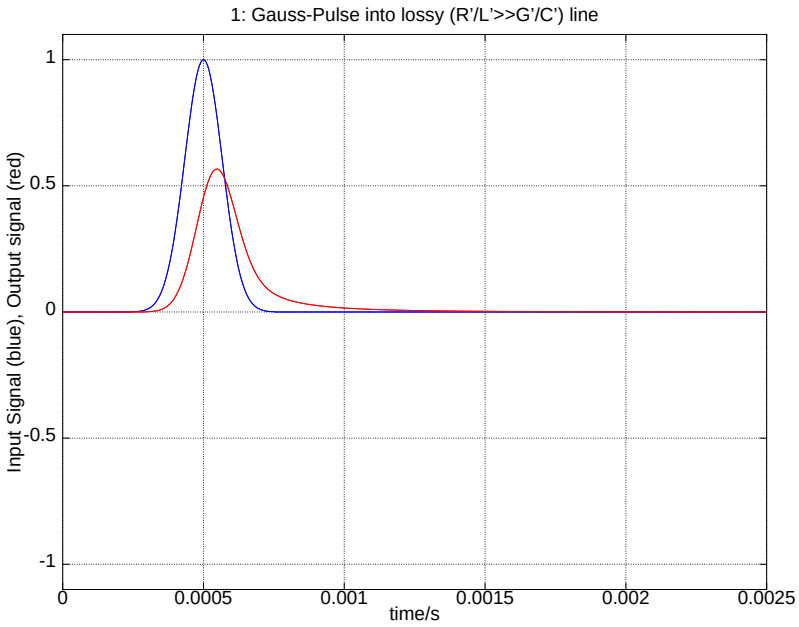
Abbildung 3.6: Kompensierte Dispersion, Ursprung hat die gleichen Parameter wie Abb. 3.4(b).

wäre ein reeller Winkel α_2 bestimmbar, eine Brechung in das Medium 2 findet darüber nicht mehr statt, alle Energie des Strahls wird reflektiert. Umgibt man nun ein Medium 1 komplett mit einem Medium 2 mit $n_2 < n_1$ komplett, so ist ein Lichtstrahl dort eingesperrt, da er an beiden Grenzschichten total reflektiert wird. Ist der Wellenleiter geometrisch als Film, also planar ausgeführt, so spricht man von einem Filmwellenleiter, ist er zylindrisch, also als "Draht" mit innerer Struktur ausgeführt, so entsteht die Glasfaser (Abb. 3.9).
Die Ausbreitung und Einkopplung in dem Wellenleiter soll nun noch näher untersucht werden. Dazu sind folgende Beobachtungen festzuhalten:

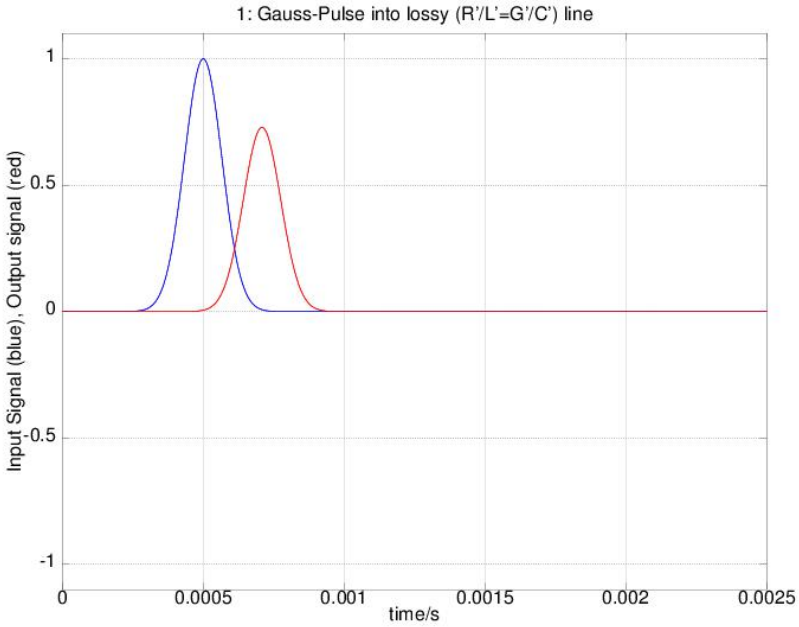
- Es können nur Anteile des Lichts in dem Wellenleiter geführt werden, die unterhalb des maximalen Winkels der Totalreflexion eingestrahlt werden. Für diesen Winkel gilt dann in dem Einkoppelmedium Luft mit einem Brechungsindex von $n = 1$

$$\frac{\sin\alpha_A}{\sin(90^\circ - \alpha_1)} = \frac{n_K}{1} \tag{3.80}$$

$$\sin\alpha_A = n_K \cos\alpha_1 = n_k\sqrt{1-\sin^2\alpha_1} = n_K\sqrt{1-\left(\frac{n_M}{n_k}\right)^2} = \sqrt{n_k^2 - n_M^2} \tag{3.81}$$



(a)



(b)

Abbildung 3.7: Dispersion (a) und Kompensation durch Induktivität (b) bei einer (elektrischen) Leitung bei Betrieb im Basisband

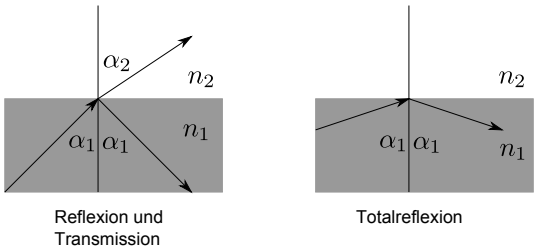


Abbildung 3.8: Zum Snellius'schen Brechungsgesetz und zur Totalreflexion

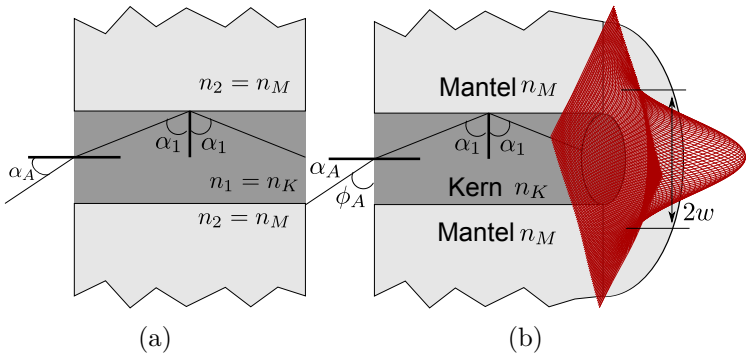


Abbildung 3.9: Aufbau eines Filmwellenleiters (a) und einer Glasfaser (b)

Dieser Wert wird auch Numerische Apertur genannt und ist ein Maß dafür, unter welchem Winkel noch Leistung in die Faser eingekoppelt werden kann.

$$\mathrm{NA} = \sin \alpha_A = \sqrt{n_k^2 - n_M^2}. \tag{3.82}$$

- Die Totalreflexion des Lichtes an der Grenzschicht ist nur eine Vorstellung aus der Strahlenoptik. In Wirklichkeit ragt das Feld der elektromagnetischen (Licht)Welle in den Mantel mit n_M hinein und die Feldstärke fällt dort exponentiell ab. Ein Teil der Welle wird also auch im Mantel geführt, so dass sich für die komplette Ausbreitung der Welle eine Mischung als Kern- und Mantelausbreitung ergibt.

- Die Reflexion an der Grenzschicht geht einher mit einer Phasenverschiebung an dieser Schicht. Man kann sich vorstellen, dass die Welle nicht direkt am Übergang, sondern etwas dahinter reflektiert wird. Dieses wird der Goos-Haenchen-Effekt ge-

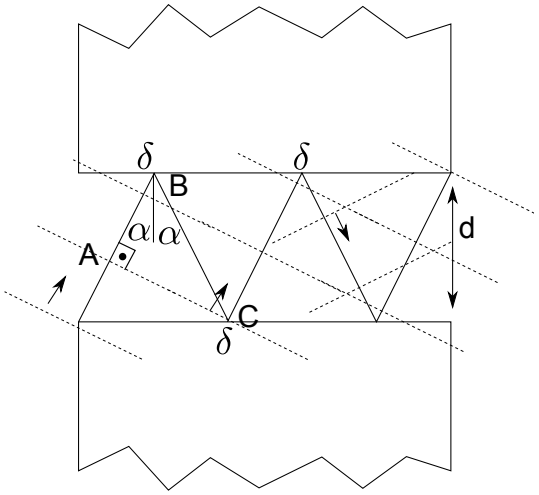


Abbildung 3.10: Strahlenoptik des LWL.

nannt und ist für die beiden möglichen Polarisationen der Welle unterschiedlich:

$$\tan\frac{\delta_\perp}{2} = -\frac{1}{\cos\alpha} \times \sqrt{\sin^2\alpha - \left(\frac{n_K}{n_M}\right)^2} \tag{3.83}$$

$$\tan\frac{\delta_\parallel}{2} = -\frac{1}{\left(\frac{n_K}{n_M}\right)^2\cos\alpha} \times \sqrt{\sin^2\alpha - \left(\frac{n_K}{n_M}\right)} \tag{3.84}$$

Für die Ausbreitung der Moden, das sind diskrete erlaubte Wellenformen in dem Wellenleiter, betrachtet man das Strahlenbild in Abb. 3.10. Hier wird die Betrachtung für den Filmwellenleiter nach Abb. 3.9(a) und Abb. 3.10 durchgeführt und erst dann auf den LWL nach Abb. 3.9(b) angepasst. Grund ist die einfachere Mathematik:
Die gesamte Phasenverschiebung der Welle auf dem Weg von A nach C muss so sein, dass die (ggf. gedachten) Phasenfronten an A mit denen an C so überlappen, dass eine konstruktive Interferenz entsteht. Die Phasenverschiebung muss also ein Vielfaches von 2π sein. Ist dieses nicht gegeben, so werden sich die Wellenanteile gegenseitig auslöschen, am Ende der Leitung kommt nichts an. Die Phase an C ist inklusive der Phasenverschiebung an den Reflexionsflächen

$$\phi_C = \phi_A + 2\delta + \beta\left(\overline{AB} + \overline{BC}\right) \tag{3.85}$$

Die Strecke $\overline{BC}$ ist einfach der um $\sin(90^\circ - \alpha)$ verkippte Durchmesser d des Films:

$$\overline{BC} = \frac{d}{\sin(90^\circ - \alpha)}. \tag{3.86}$$

Die Strecke $\overline{AB}$ ist die Ankathete des Dreiecks ABC mit der Hypothenuse BC, die schon bekannt ist:

$$\overline{AB} = \overline{BC}\cos(2\alpha) = \frac{d}{\sin(90^\circ - \alpha)}\cos(2\alpha). \tag{3.87}$$

Also ist die Phasendifferenz des gesamten Weges

$$\phi_C - \phi_A = 2\delta + \beta\frac{d}{\cos\alpha}\left(1 + \cos(2\alpha)\right) \tag{3.88}$$

$$= 2\delta + 2\beta\frac{d}{\cos\alpha}\cos^2\alpha \tag{3.89}$$

$$= 2\delta + 2\beta d\cos\alpha = 2m\pi \tag{3.90}$$

wobei das Additionstheorem $\cos^2 x = 1/2 \times (1 + \cos 2x)$ verwendet wurde und natürlich $m \in \mathbb{N}$ ist.
Somit folgt, dass nur Wellen in den Ausbreitungsrichtungen mit

$$\delta + \beta d\cos\alpha = m\pi \tag{3.91}$$

laufen können. In der Überlagerung der an der oberen Fläche reflektierten mit den an der Unterseite reflektierten Anteilen ergibt sich natürlich eine Ausbreitung in horizontaler Richtung, bzw. in Richtung des Films. Der Phasenkoeffizient in dieser, horizontalen, Richtung ist allerdings nur noch

$$\beta_{eff} = \beta \times \sin\alpha \tag{3.92}$$

oder mit Einsetzen des Kernbrechungsindexes und des Phasenkoeffizienten des freien Raumes $\beta_0 = 2\pi/\lambda_0$ folgt

$$\beta_{eff} = n_k \sin\alpha\beta_0 = n_{eff}\beta_0 \tag{3.93}$$

$$\Leftrightarrow \lambda_{eff} = \frac{1}{n_{eff}}\lambda_0 \tag{3.94}$$

womit der effektive Brechungsindex der Ausbreitung in axialer Richtung eingeführt wurde. Aus (Gl. 3.91) folgt für den Grenzfall $m = 0$, dass $\beta d\cos\alpha = -\delta$ sein muss, also der Goos-Haenchen-Effekt gerade die Laufstrecke kompensieren muss. Dieses ist möglich. Dieses ist die Grundmode des Wellenleiters, die schon in Abb. 3.9(a) angedeutet wurde. Es existieren je nach Durchmesser der Faser weitere Lösungen, die aber auf eine transzendente Gleichung führen, die nur numerisch oder grafisch gelöst werden kann [7, 27]. Eine Vereinfachung, nämlich die Annahme keiner Phasenverschiebung an den Grenzflächen mit

$\delta = 0$, macht das Leben leichter und ergibt

$$\beta d \cos\alpha = m\pi \tag{3.95}$$

$$\cos\alpha = \frac{m\pi}{\beta d}. \tag{3.96}$$

Der mögliche Winkel α muss größer als der maximal mögliche Winkel der Totalreflexion sein, denn ansonsten entschwindet die Welle in den Mantel und dann in die Armierung. Es muss also

$$\sin\alpha > \frac{n_M}{n_K} \tag{3.97}$$

$$\Leftrightarrow \sin^2\alpha - 1 > \frac{n_M^2}{n_K^2} - 1 \tag{3.98}$$

$$\Leftrightarrow \cos\alpha < \sqrt{1 - \frac{n_M^2}{n_K^2}} \tag{3.99}$$

gelten und damit dann

$$\sqrt{1 - \frac{n_M^2}{n_K^2}} > \frac{m\pi}{\beta d} \tag{3.100}$$

$$\Leftrightarrow m < \sqrt{1 - \frac{n_M^2}{n_K^2}} \frac{2n_K d}{\lambda_0} \tag{3.101}$$

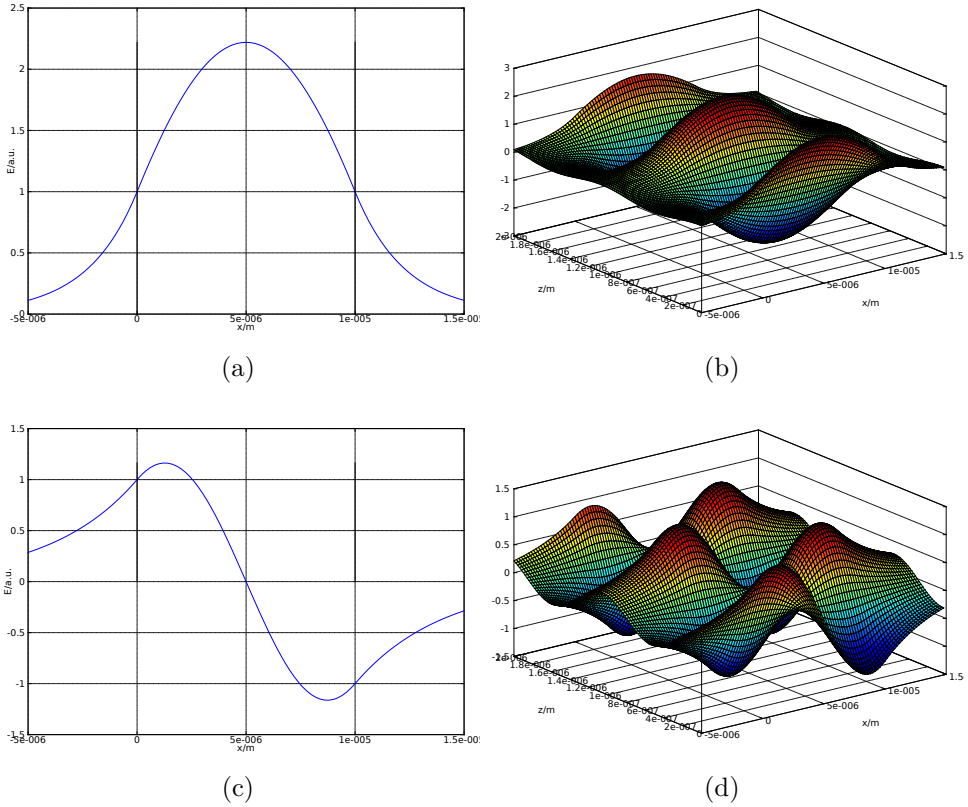
$$< \sqrt{n_K^2 - n_M^2} \frac{2d}{\lambda_0} \tag{3.102}$$

$$< \mathrm{NA}\, \frac{2d}{\lambda_0} \tag{3.103}$$

als die größte mögliche Ordnung m der Mode. Diese bestimmt damit dann natürlich auch die Anzahl der Moden, die in einem Filmwellenleiter ausbreitungsfähig sind, was auch ein sehr wichtiges Unterscheidungskriterium für LWL ist. Diese Zahl hängt von

- Wellenlänge, die meist vorgegeben ist,
- dem Brechzahlsprung, ausgedrückt durch die Numerische Apertur und
- dem Durchmesser des Kerns ab.

Beispiele für Modenverteilungen in einem Filmwellenleiter mit $n_K = 1{,}463, n_M = 1{,}458$ und einer Dicke von $10\,\mu$m bei $\lambda = 1550\,$nm sind in Abb. 3.11 gezeigt. Ersichtlich sind die unterschiedliche Anzahl der Maxima bzw. Nulldurchgänge und die aus dem eigentlichen



(a) (b)

(c) (d)

Abbildung 3.11: Modenverteilungen in planarem Filmwellenleiter für Ordnung 0 (a,b) und Ordnung 2 (c,d) mit Querschnittsverteilung (a,c) und der Wellenausbreitung (b,d).

Film deutlich herausragenden Felder. Insgesamt unterstützt dieser Film bei dieser Wellenlänge 4 Moden mit Ausbreitungskoeffizienten von $\beta_{eff} = 1{,}459304$ bis $\beta_{eff} = 1{,}461987$. Ähnliche, aber mathematisch komplexere Überlegungen führen auf die Ergebnisse der runden Glasfaser. Hier führt man den Faserparameter

$$V = \mathrm{NA}\,\frac{\pi d}{\lambda_0} \tag{3.104}$$

ein und kann die Anzahl der Moden mit

$$M = \frac{1}{2}V^2 \tag{3.105}$$

für eine Faser mit einem harten Brechzahlsprung (Stufenfaser) abschätzen. Die Anzahl der ausbreitungsfähigen Moden (unter Berücksichtigung aller Polarisationen) macht den Unterschied zwischen der SMF und der Multi Mode Fiber [Mehrmodenfaser] (MMF) gemäß ihres Namens aus. Der kritische Faserparameter $V_C = 2{,}405$ trennt die SMF von

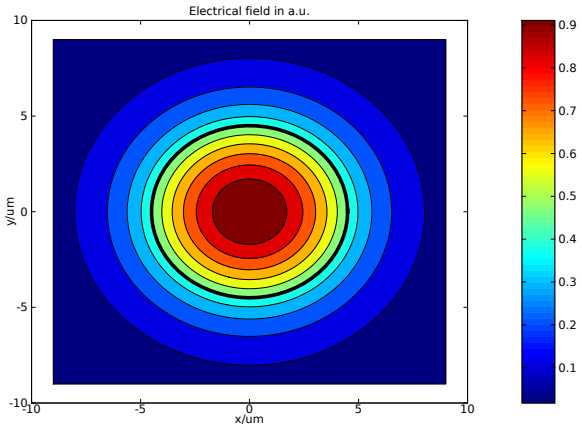


Abbildung 3.12: Feldbild der Mode einer SMF.

der MMF, denn für $V < V_C$ ist $M < 3$ und damit ist nur eine Mode in zwei Polarisationen vorhanden.
In Abb. 3.12 ist das Feldbild einer SMF gezeigt. Dieses ist die einzige ausbreitungsfähige Mode mit $\beta_{eff} = 1{,}4604$ und den Parametern Faserdurchmesser $9\,\mu$m, $n_K = 1{,}463, n_M = 1{,}458$ bei $\lambda = 1550\,$nm. Es ist ersichtlich, dass die Felder ein wenig in den Mantel hinein ragen.
In Abb. 3.13 sind dagegen verschiedenen Feldverteilungen in einer MMF gezeigt. Die Brechungsindizes sind wie oben, jedoch ist der Radius des Kernes $50\,\mu$m, was doppelt so viel wie bei einer Standard-Faser ist. Die Faser unterstützt Moden mit unterschiedlichen Ausbreitungskoeffizienten $\beta_{eff} = 1{,}4581 \ldots 1{,}4627$, einer Spanne von $0{,}3\,\%$.

3.3.2 Eigenschaften des Glases

Entscheidend für optische Wellenleiter ist zum einen die Transparenz des Stoffes, als auch die Möglichkeit der Herstellung der Totalreflexion an Grenzschichten, wofür ein Brechzahlsprung, bzw. die Modellierung eines Brechzahlverlaufs notwendig ist (Abb. 3.14). Quarzglas bietet beides. Zum einen kann eine hohe Transparenz erreicht werden, zum anderen kann durch gezielte Verunreinigung der Gläser die Brechzahl nach Abb. 3.14 eingestellt werden. Der resultierende Dämpfungsverlauf einer Faser ist in Abb. 3.15 dargestellt.
Die Absorptionsmaxima rühren von Resonanzen unvermeidlicher OH^--Ionen in dem Glasmaterial her, diese Stellen sind zu meiden, wenn man über weite Strecken kommunizieren möchte. Daher sind von der ITU die in Tab. 3.5 dargestellen optischen Bänder definiert worden, dem noch das Multimedia-Band zur Kurzstreckenübertragung bei 850 nm Wellenlänge hinzuzufügen ist.

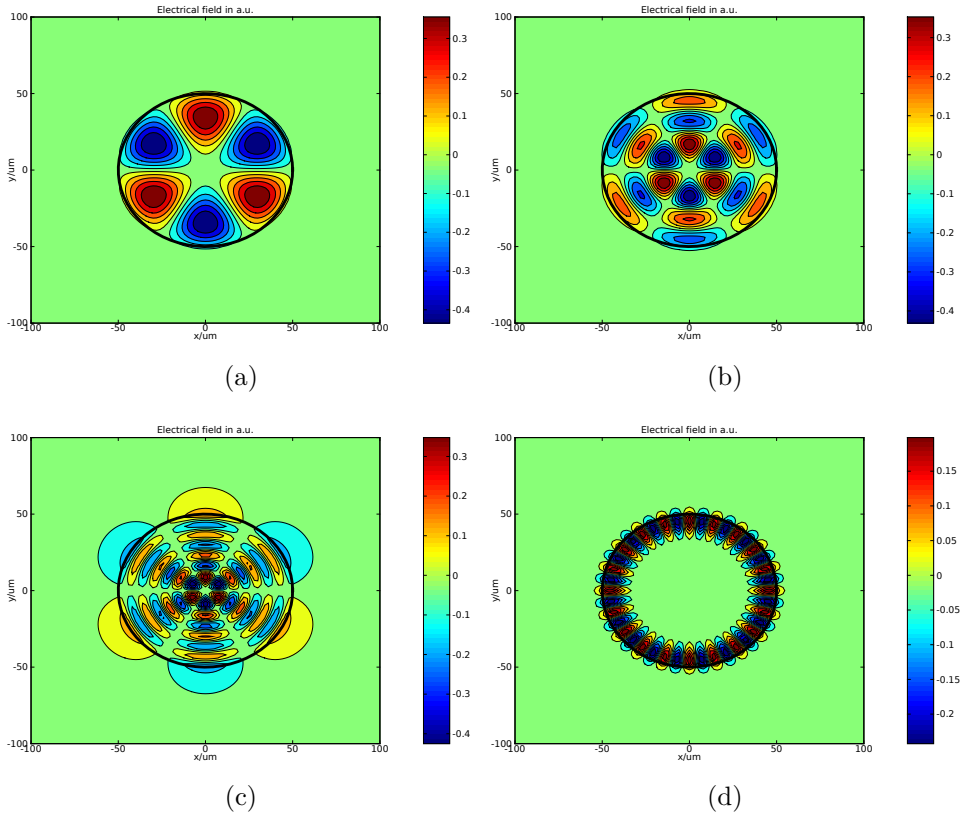


Abbildung 3.13: Feldbild der Mode in einer MMF mit Umfangsordnung 4 (a,b,c) und 21 (d) und radialer Ordnung 0,4,12.

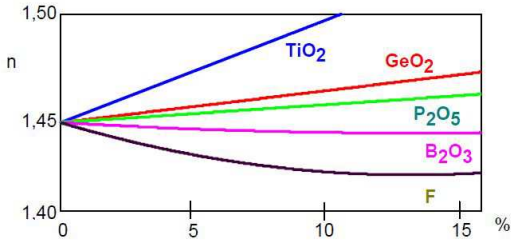


Abbildung 3.14: Brechzahlindex von Glas bei verschiedener Verunreinigung/ Dotierung (kopiert von [7, S.18].

3.3.3 Multimoden und Einmoden-Fasern

Handelsübliche Fasern selber sind in [28, 29] definiert worden. Man unterscheidet zwischen solchen, in denen nur eine fundamentale Mode (eigentlich aufgrund der Polarisationsmehrdeutigkeit zwei) oder von vornherein mehrere Moden ausbreitungsfähig sind. Wesentlicher Unterschied in den Fasern ist der Kerndurchmesser. Dieser und weitere Parameter sind in

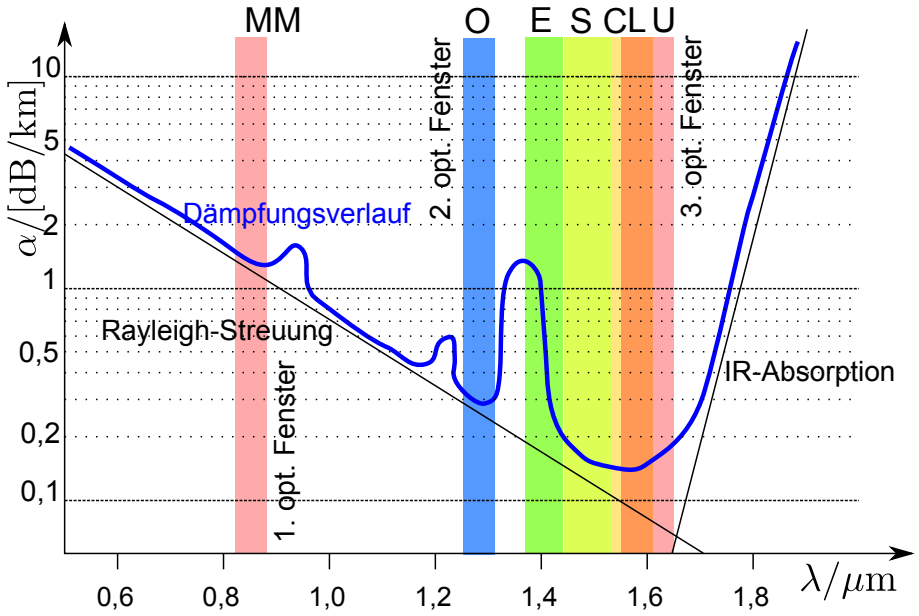


Abbildung 3.15: Typischer Dämpfungsverlauf einer Glasfaser mit Bezeichnung der Bänder.

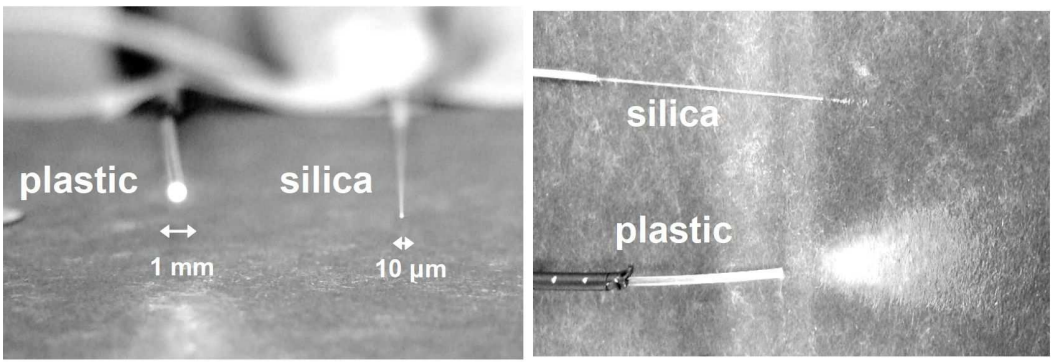
Tabelle 3.5: Definition optischer Bänder nach ITU

Band	Description	Wavelength range
O band	original	1260-1360 nm
E band	extended	1360-1460 nm
S band	short wavelengths	1460-1530 nm
C band	conventional (“erbium window”)	1530-1565 nm
L band	long wavelengths	1565-1625 nm
U band	ultralong wavelengths	1625-1675 nm

Tab. 3.6 dargestellt. Zu erkennen ist die leicht unterschiedliche Definition der Parameter, so ist z.B. bei der SMF nur der Modenfelddurchmesser vorgegeben, bei der MMF aber der Durchmesser. Dargestellt ist auch der unterschiedliche Betriebsbereich.
Eine Sonderstellung nimmt die Polymer Optical Fiber (POF) ein, die in Abb. 3.16 gezeigt ist. Diese ist nicht aus mineralischem Glas, sondern aus Kunststoff, vor allem gern aus Polymeren, hergestellt. Sie zeichnet sich durch Leichtigkeit und vor allem den Preis aus. Technisch weist sie bei Weitem höhere Verluste (Abb. 3.17) und höhere Dispersion auf. Letzteres liegt natürlich auch an der hohen Zahl ausbreitungsfähiger Moden. Die POF wird daher nur für Übertragung mit vergleichsweise geringen Bandbreite über kurze Distanzen eingesetzt. Gebiete sind die Intra-Office, In-House Verkabelung im LAN oder der Einsatz im MOST-Bus im Auto.

Tabelle 3.6: Faserparameter nach ITU-T [28, 29]

Parameter	Wert MMF	Wert SMF
Cladding diameter	125 μm	125 μm
Core diameter	50 μm	(9 μm)
Modenfelddurchm. bei 1310 nm		8,6-9,5 μm
Numerical aperture	0.20	n.d.
Cut-off Wavelevelength		1260 nm
Att. Max. at 850 nm	3.5 dB/km	n.d.
Att. Max. around 1300 nm	1.0 dB/km	0,4 dB/km
Att. Max. at 1550 nm	n.d.	0,3 dB/km



(a) (b)

Abbildung 3.16: Vergleich zwischen POF und mineralischem LWL, kopiert von [30].

3.3.3.1 Patching und Verbinden von Glasfasern

Die Verbindungstechnik von Glasfasern verdient besondere Aufmerksamkeit. So ist es das Ziel, das optische Signal möglichst Dämpfungsarm und Reflexionsarm von einer Faser in die anschließende zu bekommen. Die Standardverbindung ist das Glasfaserspleißen, wobei die einzelnen Enden mit einem Lichtbogen miteinander verschweißt werden. dazu ist folgendes zu beachten:

- Alles Coating muss von den Spleißenden entfernt werden, was heute mit einem speziellen Werkzeug gemacht wird
- Die Faserenden müssen glatt abgebrochen werden. Dazu werden die Fasern zunächst mit einem Diamanten geritzt und dann gebrochen.
- Die Faserenden müssen sehr sauber sein.

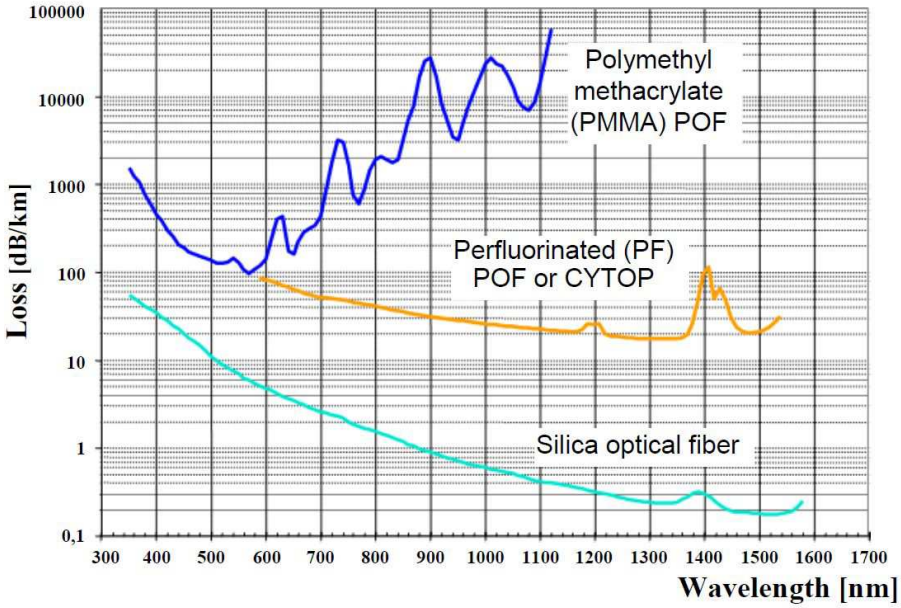
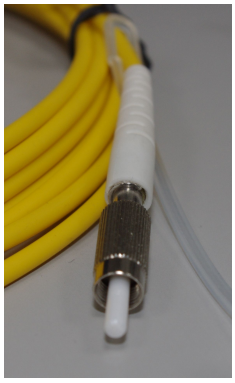
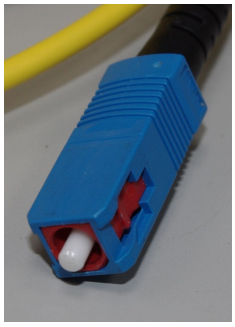
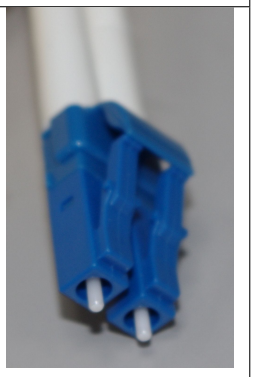
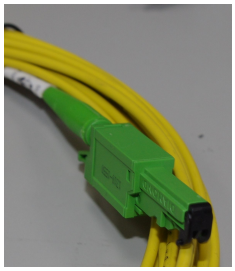
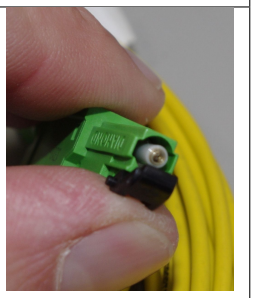
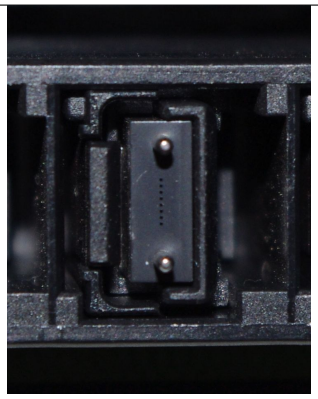


Abbildung 3.17: Dämpfungsverläufe verschiedener LWL, kopiert von [30].

- Die Faserenden werden genau gegenüber positioniert
- Nach dem eigentlichen Spleiß wird die Durchgangsdämpfung gemessen, diese soll sehr niedrig (typisch $< 0{,}1\,\mathrm{dB}$, sicher aber besser als $0{,}3\,\mathrm{dB}$) sein.
- Die Spleißstelle muss mechanisch und dicht fest verschlossen werden.

Verbindungen zwischen der eigentlichen Faserstrecke und den aktiven optischen Netzelementen werden typischerweise mit Steckverbindern vorgenommen, von denen einige in Tab. 3.7 gezeigt sind (siehe dazu auch [31]. Die Unterschiede der Stecker liegen in der mechanischen Verbindung und dem Ferrulendurchmesser. Dieser ist mit Ausnahme des LC-Steckers 2,5 mm. Der LC-Stecker hat 1,25 mm. Bspw. weisen der FC und der DIN(LSA)-Stecker Schraubverbindungen auf, die sehr fest und zuverlässig sind, und mechanisch für gute Verbindung sorgen. Daher hat dieser Verbindngstyp sehr geringe Dämpfung und eignet sich recht gut für Messgeräte. Gleichzeitig ist die Schraubverbindung aufwändig. Der E2000-Stecker weist einen integrierten Staubschutz auf, er schließt sich automatisch, wenn keine Verbindung vorliegt. Eine Sonderstellung nimmt der MTRJ-Stecker (oben nur die Buchse gezeigt) ein. Hier sind acht optische Anschlüsse auf eine Verbindung geführt.

Tabelle 3.7: Optische Steckverbinder (Auswahl)

Name	Bild		
DIN/ LSA		FCPC	
SC		LC (Duplex)	
E2000			
MTRJ			

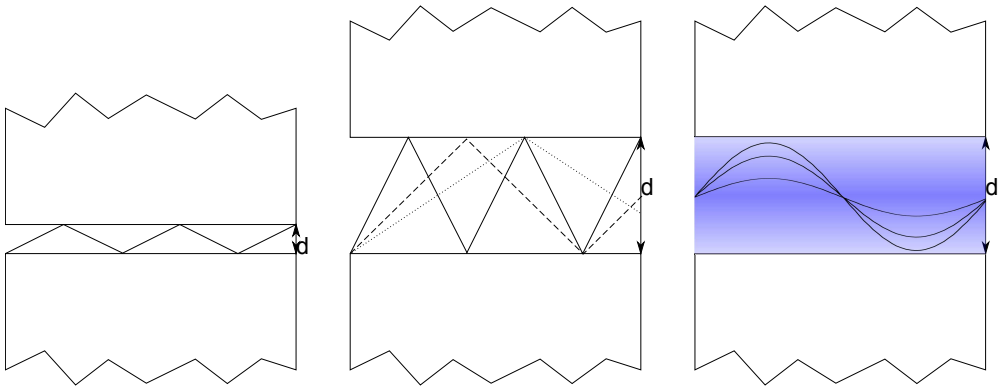


Abbildung 3.18: Wege der Moden in (a) Einmodenfaser, (b) Stufenindex Mehrmodenfaser und (c) einer Gradientenfaser.

3.3.3.2 Dispersion

Warum gibt es unterschiedliche Typen von Fasern, warum genügt nicht eine Sorte, um alle Aufgaben zu lösen? Der wichtigste Grund sind wohl die unterschiedlichen Kosten. So ist eine Faser aus Kunststoff sehr leicht und günstig herzustellen, mit ihrem großen Durchmesser ist sie gut an bspw. eine einfache LED angepasst, für kurze Übertragungsstrecken ist die ausreichend. Ebenso relativ günstig herstellbar ist eine Quarzglasfaser mit großem Innendurchmesser, die MMF. Auch diese genügt für moderate Übertragungsstrecken. Teuer, aber auch sehr leistungsfähig ist dagegen die SMF.
Neben der Dämpfung ist die Dispersion das wichtigste Unterscheidungskriterium der Glasfasern. Diese Effekte sollen nun behandelt werden.

Modendispersion In LWL, in denen mehrere Moden ausbreitungsfähig sind, kommt es zur Modendispersion. Im Strahlenbild nach Abb. 3.10 kann man sich leicht vorstellen, dass unterschiedliche Moden unterschiedliche Wege gehen und daher auch unterschiedliche Zeit zum Ziel benötigen. Ein Teil der Energie verläuft also im spitzen Winkel auf einem relativ weiten Weg, während ein anderer Teil den kurzen Weg im flachen Winkel durch die Faser sucht. Es kommt unweigerlich zum Zerlaufen eines Pulses.
Mit einer geschickten Formung des Verlaufes des Brechungsindexes kann man dieses etwas verbessern. Der Effekt ist in Abb. 3.18 dargestellt. Durch die weichere Modellierung des Brechungsindex erleben die Moden, die den weiten Weg gehen müssen, einen effektiv kleineren Brechungsindex und haben dadurch effektiv eine höhere Ausbreitungsgeschwindigkeit, wodurch sie gegenüber den Moden, die die Abkürzung nehmen, wieder aufholen

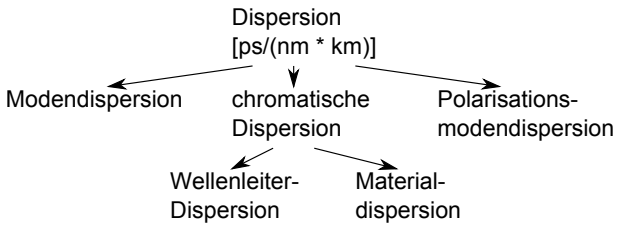


Abbildung 3.19: Hierarchie der Dispersionen [32].

können.
Diese Modendispersion ist die dominierende Größe bei MMF.

Materialdispersion Fällt die Modendispersion weg, wie in einer SMF, so ist zwar schon viel, aber noch nicht alles gewonnen. Es bleibt als wesentlicher Beitrag zur Pulsverbreiterung die Materialdispersion. Die Ausbreitungsgeschwindigkeit von Licht in Glas ist schlicht - wie schon vom Prisma bekannt - abhängig von der Farbe, also der Wellenlänge des Lichts.

Wellenleiterdispersion Wellen unterschiedlicher Frequenz zeigen eine unterschiedliche Feldverteilung der geführten Moden. Dadurch lappen die Feldverteilungen unterschiedlich in den Mantel hinein (Abb. 3.9(b)) und erfahren natürlich unterschiedliche Ausbreitungsbedingungen und -geschwindigkeiten.

Polarisationsmodendispersion Zusätzlich breiten sich in der Faser aufgrund des zylindrischen Querschnitts immer zwei Moden (in unterschiedlicher Polarisation) aus, die demnach ebenso durch Anisotropien im Glasmedium mit verschiedenen Geschwindigkeiten übertragen werden.
Zusammengefasst in einer Hierarchie sind die Dispersionseffekte in Abb. 3.19.
Die Dispersion wird in verschiedenen Koeffizienten zusammengefasst, wobei

$$D = \frac{\Delta\tau_g}{\ell \times \Delta\lambda} \tag{3.106}$$

sehr gebräuchlich ist und mit der Einheit ps/(km×nm) angegeben wird. $\Delta\tau_g$ ist damit die Laufzeitverzerrung der Gruppe über eine Länge ℓ, wenn am Eingang ein Signal mit der Linienbreite $\Delta\lambda$ eingespeist wurde. Ebenso häufig wird die Steigung der Dispersion mit der Wellenlänge, also

$$S_0 = \frac{dD}{d\lambda} \qquad \text{mit} \qquad [S_0] = \frac{\text{ps}}{\text{nm}^2\text{km}} \tag{3.107}$$

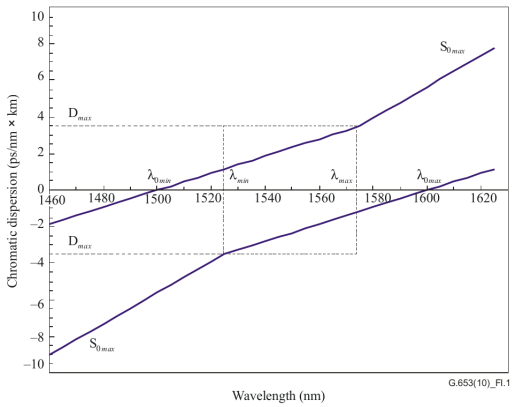


Figure I.1 – Table 2 fibre dispersion boundary

Abbildung 3.20: Dispersionsgrenzen einer dispersionverschobenen Glasfaser nach [33].

angegeben. Typische Werte für die chromatische Dispersion liegen bei [32]

$D_{Chrom} < 3{,}5\,\mathrm{ps/(km \times nm)}$ für $\lambda = 1310\,\mathrm{nm}$

$D_{Chrom} < 18\,\mathrm{ps/(km \times nm)}$ für $\lambda = 1550\,\mathrm{nm}$,

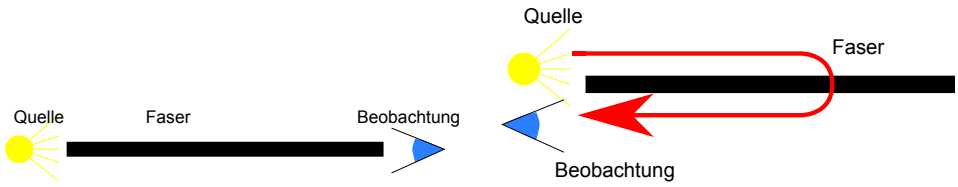
und sind für eine spezielle, dispersionsverschobene Glasfaser nach ITU-T G.653 [33] in Abb. 3.20 gezeigt.

In Kombination mit Tab. 3.6 ist klar, dass bei 1550 nm zwar das Dämpfungsminimum, aber nicht das Dispersionminimum vorliegt. Es muss also - bei Übertragung von hohen Datenraten bei dieser Wellenlänge - mit Dispersionskompensation gearbeitet werden.

Bandbreite-Länge-Produkt Das Bandbreite-Länge-Produkt [7] gibt an, über welche Strecke welche Datenrate ohne Dispersionsverluste übertragen werden kann.

$$BR \times L = \frac{1}{D_{Chrom} \times \Delta\lambda} \tag{3.108}$$

Die Bandbreite BR hängt über $BR = \frac{1}{\tau_B}$ mit der Bitdauer τ_B und damit der Übertragungsrate zusammen. Mit der o.g. Dispersion von $D_{Chrom} = 18\,\mathrm{ps/(km \times nm)}$ im 1550 nm Bereich und einem Laser mit einer Linienbreite von $\Delta\lambda = 1{,}4\,\mathrm{nm}$ ergibt sich also $BR \times L = 40\,\mathrm{GBit/s \times km}$, so dass ein 40 GBit/s-Signal über einen Kilometer ohne Dispersionsverluste übertragen werden kann. Eine Übertragung über 100 km würde also umgekehrt eine Linienbreite von 14 pm erfordern. Mit extrem schmalbandigen Lasern kann man zwar dieses als CW-Signal hin bekommen, doch führt allein die Modulation des Signals zu einer endlichen Bandbreite und damit selbst zur Linienverbreiterung. Mit



(a) (b)

Abbildung 3.21: Faser Durchgangsmessungen (a) und Reflektometrie (b).

$\lambda_0 = 1550\,\text{nm}$ und einer Bandbreite von $2 \times f_{mod} = 40\,\text{GHz}$ folgt

$$f_0 = \frac{c}{\lambda_0} = 193{,}55\,\text{THz} \tag{3.109}$$

$$f_{min} = f_0 - f_{mod} = 193{,}53\,\text{THz} \qquad \lambda_{max} = 1550{,}1\,\text{nm} \tag{3.110}$$

$$f_{max} = f_0 + f_{mod} = 193{,}57\,\text{THz} \qquad \lambda_{min} = 1549{,}8\,\text{nm} \tag{3.111}$$

$$\Rightarrow \Delta\lambda = 0{,}3\,\text{nm} \tag{3.112}$$

als Linienverbreiterung allein der Modulation. Allgemeiner kann man

$$\Delta\lambda = \Delta\lambda_{Laser} + BR \times \frac{\lambda^2}{c} \tag{3.113}$$

$$\Leftrightarrow L = \frac{c}{BR^2 \times \lambda^2 \times D_{Chrom}} \quad \Delta\lambda_L \approx 0 \tag{3.114}$$

als die maximale Länge für eine Übertragung mit der Bandbreite BR angeben.
In unserem Beispiel ergibt sich somit $L \approx 4{,}5\,\text{km}$, was auch in den obigen Überlegungen in etwa bestätigt wird.

3.3.4 Messungen an Fasern: OTDR

Glasfaserverbindungen müssen getestet werden. Ein einfachster Test auf Verbindung ist einfach die Beleuchtung der Faser, natürlich mit einer adäquaten Lichtquelle - also einem Laser -, nebst Messung des Durchgangssignals. Von einfachem Durchgang über Dämpfungsmessungen bis hin zu Messung von modulierten Signalen und damit sogar Messung der Dispersion ist hiermit alles möglich. Nachteilig ist, dass die Faser an zwei Stellen angegriffen werden muss und man natürlich zwischen beiden Punkte eine minimale Kommunikation aufbauen muss. Alternativ kann man auch ein und dasselbe Ende der Faser beleuchten und dort die Reflexionen beobachten. Man betreibt Reflektometrie.
Weiteres in “Reference Guide to Fiber Optic Testing”

3.4 Übungsaufgaben

3.4.1 Dimensionierung einer optischen Übertragungsstrecke

Bestimmen Sie die Empfangsleistung am Ende einer optischen Übertragungsstrecke mit einer Ausgangsleistung des Sendelasers mit $P_{aus} = 50\,\text{mW}$ bei 1550 nm Wellenlänge. Die Einmodenfaser hat 0,3 dB/km Dämpfung und ist 120 km lang. Zusätzlich liegt eine Dispersionskompensationsfaser mit $d = 1\,\text{dB}$ vor und es sind drei Spleiße bzw. Übergänge mit je 0,5 dB Dämpfung vorhanden. Der Empfänger braucht eine Eingangsleistung von $50\,\mu\text{W}$, welcher Verstärker ist vorzuschalten?

Lösung Die Ausgangsleistung entspricht mit 50 mW 17 dBm. Und so ergibt sich

$$P_{empf} = 17\,\text{dBm} - 120\,\text{km} \times 0{,}3\,\text{dB/km} - 1\,\text{dB} - 3 \times 0{,}5\,\text{dB} = -21{,}5\,\text{dBm} \tag{3.115}$$

Die erforderliche Eingangsleistung entspricht −13 dBm, man braucht also einen Verstärker, der das Signal um 8,5 dB bzw. den Faktor 7,1 verstärkt. Dieses kann mit einem Faserverstärker erreicht werden.

3.4.2 Übertragung mit koaxialen Leitungen

An den Eingang einer Koax-Leitung mit $Z_c = 50\,\Omega$ wird eine Spannung von $V_{pp} = 10\,\text{V}$ (Spitze-Spitze) angelegt. Die Spannung ist sinusförmig und die Frequenz beträgt 2 GHz. Nach welcher Strecke ist die verfügbare Leistung am Empfänger auf 1 mW abgefallen, wenn der Dämpfungsfaktor der Leitung $\alpha = 0{,}1\,/\text{m}$ beträgt?

Lösung Die Spitzen-Spitzenspannung entspricht $U_{eff} = U_{pp}/2/\sqrt{2} = 3{,}536\,\text{V}$, also einer Ausgangsleistung von

$$P_{out} = \frac{U_{eff}^2}{Z_C} = 250\,\text{mW} \tag{3.116}$$

mit

$$\frac{P_{in}}{P_{out}} = e^{-\alpha l} \qquad \Leftrightarrow \qquad l = -\frac{1}{\alpha} \times \ln\left(\frac{P_{in}}{P_{out}}\right) \tag{3.117}$$

$$l = 55{,}2\,\text{m}. \tag{3.118}$$

Nach nur 55,2 m ist also die Grenze der Strecke erreicht. Ein Vergleich mit einem handelsüblichen RG 58 Kabel, das bei 2 GHz ein $\alpha = 0{,}32\,/\text{m}$ hat, zeigt, dass wir hier zu

optimistisch gerechnet haben. Bei 50 MHz hat auch RG 58 ein $\alpha = 0{,}11\,/\mathrm{m}$ und liegt damit in unserem Bereich.

3.4.3 Dispersion

Bestimmen Sie das Bandbreite-Länge-Produkt eines LWL mit $D_{chrom} = 4\,\mathrm{ps/(nm \times km)}$ bei $\lambda = 1310\,\mathrm{nm}$ und $D_{chrom} = 20\,\mathrm{ps/(nm \times km)}$ bei $\lambda = 1550\,\mathrm{nm}$. Die Faser wird mit einer Quelle der Linienbreite $\Delta\lambda = 1\,\mathrm{nm}$ bestrahlt.
Wenn die Linienbreite vernachlässigt werden kann, wie groß ist dann bei einer Bitrate von $BR = 10\,\mathrm{Gbit/s}$ die überbrückbare Strecke ohne Dispersionskompensationsmaßnahmen? Welche Maßnahmen zur Dispersionskompensation bzw. Vermeidung kennen Sie allgemein und speziell in der optischen Übertragungstechnik?

Lösung Mit

$$BR \times L = \frac{1}{D_{chrom} \times \Delta\lambda} \tag{3.119}$$

folgt

$$250\,\mathrm{Gbit \times km/s} \quad \text{für} \quad \lambda = 1310\,\mathrm{nm} \tag{3.120}$$

$$50\,\mathrm{Gbit \times km/s} \quad \text{für} \quad \lambda = 1550\,\mathrm{nm}, \tag{3.121}$$

wenn die Linienbreite des Lasers eine Rolle spielt. Im ersten Fall kann man also 25 km, im zweiten 5 km überbrücken. Spielt die Linienbreite des Lasers keine Rolle, so kann man mit der o.g. Bitrate eine Strecke wie folgt überbrücken:

$$L = \frac{c}{BR^2 \lambda^2 D_{chrom}} \tag{3.122}$$

und damit

$$437\,\mathrm{km} \quad \text{für} \quad \lambda = 1310\,\mathrm{nm} \tag{3.123}$$

$$62\,\mathrm{km} \quad \text{für} \quad \lambda = 1550\,\mathrm{nm}. \tag{3.124}$$

Die Linienbreite des Lasers kann nicht vernachlässigt werden, denn allein durch seine Breite ergeben sich deutlich schlechtere Werte als allein durch die Modulationsverbreiterung möglich wären. Die bei $BR = 10\,\mathrm{GBit/s}$ überbrückten Strecken sind oben gegeben.
Auswirkungen der Dispersion kann man verringern durch besonders schmalbandige Quellen. Weiterhin kann man durch induktive Kompensation in Kabeln die Dispersion kompensieren, oder im Optischen kann man die Wirkung der Dispersion durch Dispersionskompensationsfasers zurückdrehen.

Kapitel 4

Übertragung im Basisband

Als Basisband bezeichnet man den Frequenzbereich, in dem die Signale ursprünglich vorliegen. Bei Sprachsignalen ist das natürlich der Bereich von ca. 300 Hz bis 3 400 Hz. Videosignale im Deutschen PAL-Standard beanspruchen eine Bandbreite von 5 MHz, man würde dieses Band auch das Videoband nennen. Der Begriff ist verallgemeinert synonym zum Basisband zu sehen. Bei Datensignalen wird als Basisband die Grundfrequenz (Symbolrate) der Übertragung verstanden.
Übertragung im Basisband ist als Gegensatz zur Übertragung der Daten aufmoduliert auf einen Träger zu verstehen. Bei der herkömmlichen (terrestrischen) Radio- oder Fernsehübertragung wird ja nicht das eigentliche Tonsignal ausgesendet, sondern eines, das in der Frequenz versetzt wurde. Hier findet eine Übertragung im Bandpassbereich statt, da das Zielsignal nur in dem Frequenzbereich um den Träger vorliegt und nur durch Filterung und zurück-Umsetzung regeneriert werden kann.
Betrachtet man allerdings - auch im Sinne von Protokollschichten - die Ende zu Ende Beziehung zwischen Sender und Empfänger, so kann man zunächst einmal eine einfache Übertragung im Basisband als Modell für die Gesamtübertragung verwenden. Die ggf. vorhandene Umsetzung auf verschiedenen Frequenzen wird dann einfach hinterher als spezieller Übertragungskanal eingefügt.

4.1 Das Übertragungssystem im Basisband

Ein einfaches - generisches - Übertragungssystem ist in Abb. 4.1 dargestellt. Die großen Blöcke sind natürlich Sender (kurz: Transceiver oder Transmit (Tx)) und Empfänger (kurz: Receiver oder Receive (Rx)) und Übertragunskanal. Eine weitere Aufschlüsselung bringt die folgenden Bausteine zum Vorschein:

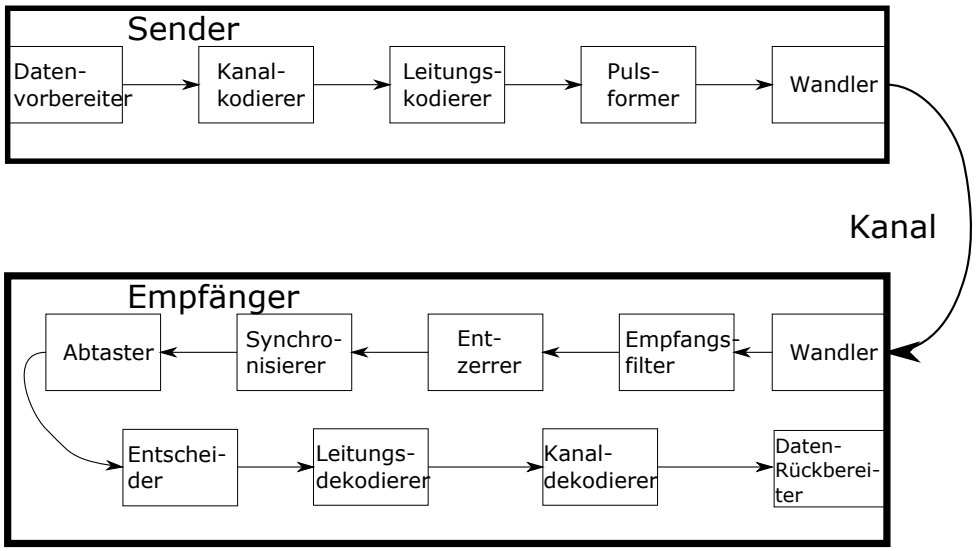


Abbildung 4.1: Modell eines Übertragungssystems für Übertragung im Basisband

Sender besteht aus

Datenvorbereiter steht zusammenfassend für Scrambler , der die Daten "würfelt", so dass keine langen Sequenzen von Einsen und Nullen entstehen, oder sogar chiffriert, wenn Geheimhaltung erforderlich ist. Hier können auch Elemente wie Parallel-Serien-Wandler vorliegen, die aus einem parallelen Datenstrom (Bytes bzw. Octetts) einen seriellen machen. Ggf. werden hier die Daten auch paketiert, wenn sie nicht in einem kontinuierlichen Strom auf die Leitung geschickt werden können.

Kanalkodierer Der Kanalkodierer bereitet die Signale durch Zufügen gezielter Redundanz auf die Übertragung vor. Bspw. wird hier die Checksumme zugefügt, oder werden andere, ggf. kompliziertere Mechanismen zur Datensicherung vorgenommen.

Leitungskodierer Der Leitungskodierer bringt nun die immer noch digital vorliegenden Signale in eine Form, die sich zur Übertragung anbietet. Hier werden aus Null und Eins Signale, die sich einer Spannung oder Phase zuordnen lassen. So werden hier bspw. bipolare Signale (also solche mit positivem und negativem Spannungspegel) gebildet. Es liegen hier aber noch *nicht* die Spannungs-/Leistungswerte und -verläufe vor, die wir auf der Leitung finden.

Pulsformung In diesem Schritt der Pulsformung wird aus dem als hart umgetastetes (Rechteck-)Signal ein an den Flanken Verschliffeneres gebildet. Es wird

Tiefpass gefiltert, wobei allerdings die Form des Filters so gewählt ist, dass die einzelnen Symbole nur so ineinander laufen, dass sie sich gegenseitig nicht stören. Hier muss mit einer Überabtastung (engl. Oversampling) gearbeitet werden, so dass die weicheren Verläufe auch wirklich dargestellt werden können.

Wandler Der Wandler ist der letzte Schritt vor dem eigentlichen Kanal, der Leitung. Hier wird das physikalische Signal tatsächlich erzeugt und über bspw. einen Verstärker oder - im Optischen - einen Laser in die Leitung eingespeist.

Kanal besteht aus dem Übertragungsmedium. Bei Übertragung im Basisband ist das einfach die Leitung, die allerdings als Kupferleitung für elektrische Signale oder als Lichwellenleiter für optische Signale ausgeführt sein kann. Dient die Vorstellung der Übertragung im Basisband nur als Modell, so ist hier jeder beliebige Kanal anzusetzen, also beispielsweise auch ein Funkkanal. Aber um auf diesen zugreifen zu können, muss natürlich das Signal anders aufbereitet werden und insbesondere in der Frequenz umgesetzt werden.

Empfänger Im Prinzip ist der Empfänger der Sender rückwärts, wobei einige zusätzliche Elemente zwingend sind:

Wandler wandelt das auf der Leitung vorliegende Signal zunächst zurück in eines, das weiter verarbeitet werden kann. In der Regel ist dieses Element ein Empfangsverstärker/ Empfangswandler oder - im Optischen - eine Fotodiode. Je nach Ausprägung des Kanals kann das erste Element auch ein Tiefpassfilter sein, um Störsignale daran zu hindern, überhaupt in den Empfänger einzudringen.

Empfangsfilter Dieses Filter ist angepasst zur Pulsformung und sorgt dafür, dass die einzelnen Symbole in einer möglichst interferenzfreien Form vorliegen.

Entzerrer Der Entzerrer soll die Effekte des Kanals kompensieren. Hierzu ist natürlich Kenntnis über den Kanal nötig, so dass dieser herausgerechnet werden kann.

Synchronisation Die Synchronisation ist notwendig, da ja in der Regel Sender und Empfänger keine, oder jedenfalls keine genau genuge Information über den Takt des jeweils anderen haben. Während der Synchronisation muss also aus dem ankommenden Signal neben der Nutzinformation auch noch die Taktinformation zurückgewonnen werden, so dass am

Abtaster immer zum richtigen Zeitpunkt der Signalwert abgenommen wird, um dann im

Entscheider festlegen zu können, um welchen übertragenen Wert es sich handelt. Wird einfach ein Bit pro Symbol übertragen, so liegen hier wieder die Einsen und Nullen, bipolare Signale oder fest entschiedenen Symbole vor. Handelt es sich um eine höherwertige Übertragung, in der pro Symbol mehrere Bit übertragen wurden, so haben wir hier nun diese Werte (also mehrere Bits) vorliegen.

Leitungsdecodierer bilden nun aus den vorher entschiedenen Werten die digitalen Daten zurück, die dann im

Kanaldekodierer von der Redundanz befreit werden, ggf. werden auch die Fehler, die auf der Übertragungsstrecke vorlagen korrigiert, die Checksummen werden überprüft.

Datenrückbereiter übernimmt nun noch die Aufgabe, die Daten wieder so vorzulegen, bspw. zu descramblen, dass die nachfolgenden Protkollschichten etwas damit anfangen können.

4.2 Repräsentation der Daten auf der Leitung

Als Erstes wollen wir uns um die Darstellung von Daten auf dem physikalischen Übertragungsmedium kümmern. Betrachtet werden in diesem Abschnitt ausschließlich digitale Signale, die - entsprechend dem binären System - als Einsen und Nullen dargestellt werden. Physikalisch wird aber natürlich nicht eine Eins oder Null, sondern ein Spannungsverlauf oder ein Verlauf eines optischen Signals (Leistung) übertragen. Wesentlich ist, dass neben dem eigentlichen Wert zudem noch die Taktinformation im Empfänger reproduzierbar sein muss. Daher gibt es neben einfacher Darstellung des digitalen Signals als ein Hochpegel (Eins) und ein niedriger Pegel (Null) einige weitere Repräsentationen, die teilweise sehr viel besser an den verwendeten Übertragungskanal angepasst sind. Wir werden einige davon kennenlernen (Abb. 4.2):

NRZ Die Non Return (to) Zero (NRZ) Darstellung eines digitalen Signals ist wohl die am leichtesten vorstellbare. Hier wird einfach eine Eins als ein Hochpegel (allgemein: Spannung oder Pegel U_1 oder P_1) und eine Null als ein Nullpegel (allgemein: Spannung U_0 oder Pegel P_0) dargestellt. In den meisten Anwendungen ist $U_1 = U_{max}$ und $U_0 = 0$ oder $U_0 = -U_{max}$, so dass entweder ein- und ausgeschaltet (On Off

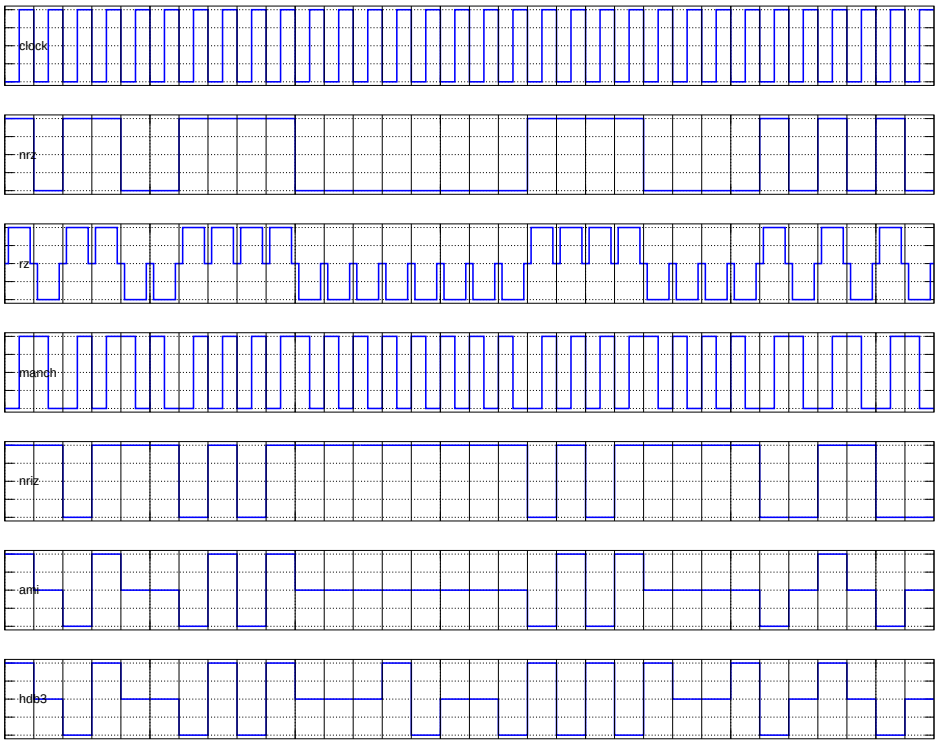


Abbildung 4.2: Zeitverläufe verschiedener Kodierungsarten mit NRZ, RZ, Manchester, NRZI, AMI, HDB3

Keying (OOK)) wird, oder zwei bipolare Pegel verwendet werden (Binary Phase Shift Keying (BPSK)). Im Empfänger wird das Signal dann irgendwo im mittleren Bereich des Bit-Intervalls abgenommen, jedenfalls dort, wo vermeintlich die Flanken das Signal nicht verfälschen können. Der Vorteil ist, dass mit dieser Darstellung - einmal abgesehen von spektralen Transienten an den Flanken - die zur Verfügung stehende Bandbreite des Übertragungskanals gut ausgenutzt werden kann. Nachteil ist vor allem bei langen Sequenzen eines Pegels (also viele Einsen oder viele Nullen hintereinander) zu sehen, dann nämlich kann der Takt über eine recht lange Zeit nicht mehr zurückgewonnen werden, es entsteht eine Unsicherheit in der Abtastung, die im schlimmsten Fall zum Verlust des Taktes und damit des Signals führen kann. Außerdem ist bei schlechter Wahl $U_1 \neq -U_0$ die Übertragung nicht mittelwertfrei, es wird also auch eine Gleichspannungskomponente übertragen, die eine Entkopplung bzw. eine Abnahme des Signals über induktive Koppler unmöglich macht.

Die höchste auftretende Frequenz kann hier der periodische Wechsel von “Eins” auf “Null” sein. Diese Situation kann durch einen Sinus mit einer Periodendauer, die dem doppelten der Bitdauer entspricht, dargestellt werden.

NRZI Die Non Return to Zero, Inversion on one (NRZI) heißt zwar fast wie die vorgenannte, ist aber schon wesentlich etwas anderes. In der originären Version wird nun eine “Eins” dargestellt als eine Inversion des Pegels, also einen Pegelübergang, während eine “Null” den (vorhergehenden) Pegel einfach hält.

Hier wird die höchste notwendige Frequenz durch eine permanente Folge von “Einsen” erreicht, und das heißt wie auch schon beim NRZ ein Pegelwechsel pro Bitdauer und die Spektralen Eigenschaften, insb. die notwendige Bandbreite, entsprechen der der NRZ.

Manchester Die Manchester Kodierung legt die binären Signale als Übergang von einem unteren Pegel auf einen höheren (Eins) oder von einem höheren auf einen unteren Pegel (Null) fest. Damit ist gewährleistet, dass in jedem Intervall mindestens einmal der Pegel gewechselt wird und damit eine Flanke entsteht. Aus der Flanke ist natürlich immer der Takt abzuleiten. Der Preis ist der im Extremfall im Vergleich zur NRZ-Kodierung doppelt so schnelle Signalwechsel, was natürlich eine doppelt so hohe Bandbreite erfordert.

AMI Die Alternate Mark Inversion (AMI) ist eine sehr einfache - und in der reinen Form ungenügende - Erweiterung der NRZ-Darstellung. Hier wird die Null als physikalischer Nullpegel übertragen und eine Eins wird abwechselnd als U_{max} und als $-U_{max}$ dargestellt. Damit ist das Problem des Gleichspannungsanteils gelöst, denn ein AMI-kodiertes Signal ist ein reines Wechselsignal, Transformatoren/ induktive Koppler stellen kein Hindernis mehr dar. Weiterhin ist auch bei einer langen Folge von Einsen ein genügender Pegelwechsel gegeben, so dass der Takt zurückgewonnen werden kann. Allerdings gilt dieses nicht für lange Nullfolgen. Hier ist einfach die Spannung aus, ein Takt kann noch weiterhin verloren gehen. Faktisch ist hier nun ein ternäres Signal mit drei Pegeln, nämlich $-U_{max}, 0, +U_{max}$, eingeführt worden.

HDB3 Der High Densitiy Bipolar (3 consecutive zeros) (HDB3) ist eine Variante des AMI, bei dem gezielt die einfache Regel: “Eine Null wird durch den Nullpegel dargestellt.” durch einen komplizierteren Zusatz erweitert wird: ``Each block of 3 (or 4) successive zeros is replaced by 00V (or 000V respectively) or B0V (B00V). The choice of 00V (000V) or B0V (B00V) is made so that the number of B

`pulses between consecutive V pulses is odd. In other words, successive V pulses are of alternate polarity so that no DC component is introduced.''` [34]

Wobei sich hier die 3 auf HDB2 und 4 auf HDB3 bezieht und dies damit auch die maximale Anzahl von Nullpegeln hintereinander (nämlich zwei oder drei) festlegt. In der o.g. Definition meint ein "B", dass der Puls der AMI-Regel gehorcht, also gegenüber dem letzten "Eins"-Puls invertiert ist, währen ein "V" für "Violation" also Verletzung (der Regel) steht, also der Puls gegenüber dem letzten Puls **nicht** invertiert ist.

In der Implementation benötigen wir also zwei Zähler, einer, n_0, zählt die Anzahl der Nullen (seit letztem Puls) und der zweite, n_B, zählt die Anzahl der gültigen Pulse "B" zwischen zwei Verletzungen. Ist also $n_B = 1,3,5,...$ ungerade, und kommt nun eine Folge von vier Nullen, dann wird $000V$ eingesetzt, der Zähler wird $n_B = 0$ gesetzt. Ist $n_B = 0,2,4,6,...$ also gerade und es kommt eine Folge von vier Nullen, dann wird $B00V$ gewählt, wobei das erste B die Anzahl von binären Pulsen n_B ungerade macht und dann natürlich obige Regel angewendet wird und ein V eingeschoben werden darf. Mit dem V wird natürlich wieder der Zähler $n_B = 0$ gesetzt.

Mit dem HDB3 haben wir nun erreicht, dass der Symbolstrom mittelwertfrei ist und gleichzeitig auch hinreichend oft ein Pegelwechsel vorkommt, damit der Takt regeneriert werden kann.

RZ der Return (to) Zero (RZ)-Code ist letztlich nur eine leichte Abwandlung des schon oben beschriebenen NRZ-Codes. Beim RZ fällt man in jedem Intervall auf "Null" zurück. Damit ergibt sich in jedem Intervall ein Pegelwechsel, womit der Takt zurück zu gewinnen ist. Voraussetzung ist natürlich, dass die Eins als positive Spannung und die Null als negative Spannung dargestellt wird. Dieses ist der Bipolare-RZ-Code, während beim unipolaren (also nur die Eins hat eine Spannung, bei der Null ist keine Spannung) wieder nur die Hälfte der Probleme (also lange Eins-folgen) gelöst ist.

Aufgrund der notwendigerweise kürzeren Pulse - es wird ja noch Platz für den Rückgang zur Null benötigt - ist natürlich der Bandbreitebedarf dieser Kodierung größer als beim NRZ-Verfahren.

Im Ethernet wird im 10BASE-T die Manchester Kodierung verwendet, im 100BASE-X wird eine 4B/5B-Kodierung eingesetzt, in der ein Nibble, also jeweils vier Bits, auf fünf

Tabelle 4.1: 4B/5B Abbildung im 100BASE-X

PCS code-group [4 : 0] 4 3 2 1 0	Name	MII (TXD/RXD) < 3 : 0 > 3 2 1 0	Interpretation
Data			
1 1 1 1 0	0	0 0 0 0	Data 0
0 1 0 0 1	1	0 0 0 1	Data 1
1 0 1 0 0	2	0 0 1 0	Data 2
1 0 1 0 1	3	0 0 1 1	Data 3
0 1 0 1 0	4	0 1 0 0	Data 4
0 1 0 1 1	5	0 1 0 1	Data 5
0 1 1 1 0	6	0 1 1 0	Data 6
0 1 1 1 1	7	0 1 1 1	Data 7
1 0 0 1 0	8	1 0 0 0	Data 8
1 0 0 1 1	9	1 0 0 1	Data 9
1 0 1 1 0	A	1 0 1 0	Data A
1 0 1 1 1	B	1 0 1 1	Data B
1 1 0 1 0	C	1 1 0 0	Data C
1 1 0 1 1	D	1 1 0 1	Data D
1 1 1 0 0	E	1 1 1 0	Data E
1 1 1 0 1	F	1 1 1 1	Data F
1 1 1 1 1	I	undefined	IDLE;
Control			
1 1 0 0 0	J	0 1 0 1	Start-of-Stream Delimiter, 1/2; always used in pairs with K
1 0 0 0 1	K	0 1 0 1	Start-of-Stream Delimiter, 2/2; always used in pairs with J
0 1 1 0 1	T	undefined	End-of-Stream Delimiter, 1/2; always used in pairs with R
0 0 1 1 1	R	undefined	End-of-Stream Delimiter, 2/2; always used in pairs with T

Bit abgebildet werden. Dieses ist eigentlich nur eine konsequente Fortführung der Idee des HDB3-Codes, so dass nicht nur einzelne “pathologische” Fälle wie viele folgende Nullen betrachtet werden, sondern eben durchgängig alle Daten angeschaut werden und gezielt

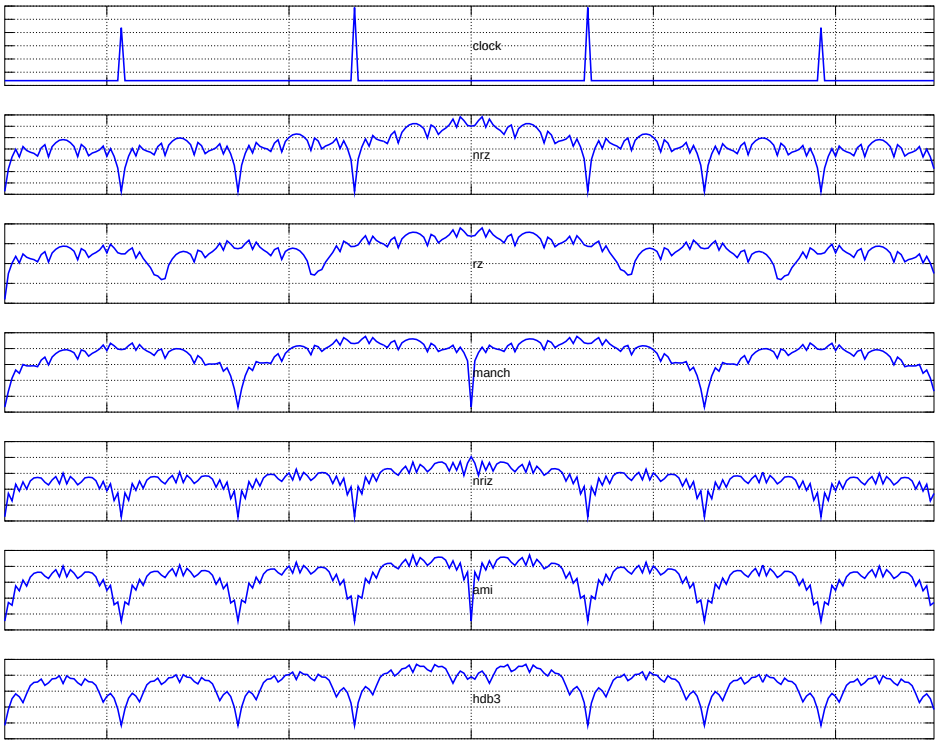


Abbildung 4.3: Spektrale Eigenschaften verschiedener Kodierungsarten mit NRZ, RZ, Manchester, NRZI, AMI, HDB3

umgesetzt werden (Tab. 4.1). Es werden von den mit fünf Bit möglichen Bitfolgen nur 16 für Daten und vier für Control, also zwanzig, verwendet, die anderen sind ungültig und werden als Fehler erkannt und zurückgewiesen. Auf diese Weise ist schon eine erhöhte Sicherheit in der Datenübertragung durch geschickte Kodierung entstanden. Diese so kodierten Daten werden mit der oben beschriebenen NRZI weiter kodiert und auf die Leitung geschickt. Für höherwertige Übertragungsverfahren wird dieses fortgesetzt, so dass bspw. für das 10 GigE 64 Datenbits auf 66 Leitungskodes umgesetzt werden. Der Overhead ist natürlich mit $2/64 = 1/32 = 3{,}125\%$. Der Kode heißt dann 64B/66B [16]

4.2.1 Spektrum und Spektrale Eigenschaften

Bisher haben wir uns explizit nur um eine Darstellung des Signals im Zeitbereich gekümmert. Zur Analyse des Signals bzw. auch zur Auslegung des Übertragungssystems ist eine

Betrachtung im Spektralbereich oft sehr hilfreich.
Mit der bekannten Transformation

$$S(f) = \mathcal{F}\{x(t)\} = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}dt \tag{4.1}$$

und Rücktransformation

$$x(t) = \mathcal{F}^{-1}\{S(f)\} = \int_{-\infty}^{\infty} S(f)e^{j2\pi ft}df \tag{4.2}$$

wird der Übergang zwischen beiden Welten bewerkstelligt. Während das Zeitsignal die bspw. Dimension einer Spannung und damit der Einheit "V" hat, bekommt das Signal im Spektralbereich die Einheit Vs, oder V/Hz. Die hier gewählte Darstellung hat den Nachteil, dass man einige 2π mitschleppen muss, aber den Vorteil, dass die Darstellung symmetrisch ist, und nach Transformation (4.1) auch der Gleichanteil der Transformierten dem Mittelwert der Zeitbereichsfunktion entspricht.
Im Prinzip ist die Fouriertranformation nach (Gl. 4.1) nichts weiter als ein Spektrumanalysator, der ein Zeitsignal an einem (komplexen) Sinus-Signal - repräsentiert durch $e^{-j2\pi ft}$ - heruntermischt und dann mit einem Tiefpass filtert, um wirklich nur Komponenten bei dieser Frequenz zu messen - das Filtern macht die Mittelung durch das Integral.
Auch ein rein reelles Zeitsignal hat in der Regel ein komplexes Spektrum. Für spezielle Fälle kann Tab. 4.2 zu Rate gezogen werden.

4.2.1.1 Eigenschaften der Fouriertransformation

Die wichtigsten Eigenschaften zur Fouriertransformation wie Komplexität und Symmetrie sind in Tab. 4.2 zusammen gefasst.

4.2.1.2 Sätze zur Fouriertransformation

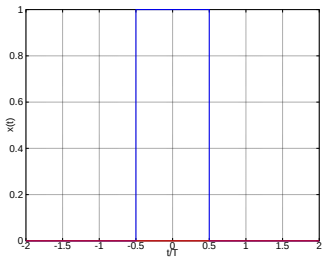
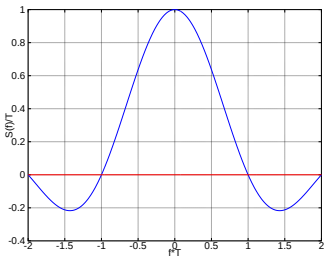
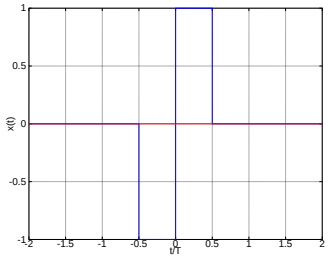
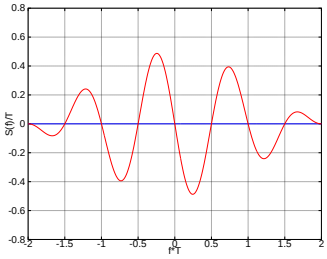
Diese Zusammenstellung findet sich für zwei Dimensionen z.B. in [35]. Hier wurde sie (die Darstellung) auf eine Dimension angepasst.

Linearitatstheorem

$$\mathcal{F}\{af(t)\} = a\mathcal{F}\{f(t)\} \tag{4.3}$$

was direkt aus der Definition des Fourierintegrals folgt.
Man merke sich, dass die Verstärkung einer Funktion im Zeitbereich genauso auf das Spektrum wirkt.

Tabelle 4.2: Spezielle Fourier-Beziehungen

Zeitbereich $x(t)$	Frequenzbereich $S(f)$
gerade	gerade
ungerade	ungerade
reell	reell, gerade + imaginär, ungerade
imaginär	reell, ungerade + imaginär, gerade
reell, gerade 	reell, gerade 
reell, ungerade 	imaginär, ungerade 
imaginär, gerade	imaginär, gerade
imaginär, ungerade	reell, ungerade

Ähnlichkeitstheorem Mit

$$\mathcal{F}\{x(t)\} = S(f) \tag{4.4}$$

folgt

$$\mathcal{F}\{x(at)\} = \frac{1}{|a|} S\left(\frac{f}{a}\right), \tag{4.5}$$

wobei a reell ist.

Beweis:

$$\mathcal{F}\{x(at)\} = \int_{-\infty}^{\infty} x(at)e^{-j2\pi ft}dt \tag{4.6}$$

$$= \int_{-\infty}^{\infty} x(t')e^{-j2\pi ft'/a}\frac{1}{a}dt' \tag{4.7}$$

$$= \frac{1}{|a|}S\left(\frac{f}{a}\right). \tag{4.8}$$

Der Betrag rührt daher, dass die Integrationsgrenzen immer von minus nach plus gehalten werden. Eine Skalierung der Zeit führt also zu einer umgekehrten Skalierung der Frequenz, wird also durch ein $|a| > 1$ der zeitliche Verlauf der Funktion beschleunigt, so wird die Änderung der Frequenz verlangsamt, was zu einer Aufweitung des Spektrums führt.
Man merke sich: Je schneller sich eine Funktion im Zeitbereich ändert, desto weiter wird ihr Spektrum, und zwar im genau umgekehrten Maße.

Verschiebungstheorem Mit

$$\mathcal{F}\{x(t)\} = S(f) \tag{4.9}$$

folgt

$$\mathcal{F}\{x(t-a)\} = e^{-j2\pi af} \times S(f) \tag{4.10}$$

wobei a reell ist.
Beweis:

$$\mathcal{F}\{x(t-a)\} = \int_{-\infty}^{\infty} x(t-a)e^{-j2\pi ft}dt \tag{4.11}$$

$$= \int_{-\infty}^{\infty} x(t')e^{-j2\pi f(t'+a)}dt \tag{4.12}$$

$$= e^{-j2\pi af} \times S(f). \tag{4.13}$$

Das heißt also, dass eine Verschiebung der Funktion im Zeitbereich zu einer Modulation im Frequenzbereich führt. Der Satz ist in seiner umgekehrten Form praktisch von hoher Bedeutung:

Modulationstheorem

$$\mathcal{F}\{x(t)e^{+j2\pi f_0 t}\} = S(f - f_0) \tag{4.14}$$

Beweis:

$$\mathcal{F}\{x(t)e^{+j2\pi f_0 t}\} = \int_{-\infty}^{\infty} x(t)e^{-j2\pi(f-f_0)t}dt \tag{4.15}$$

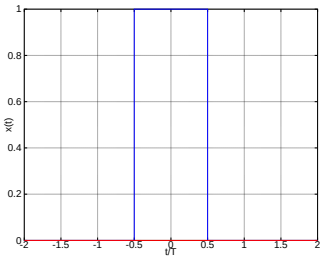
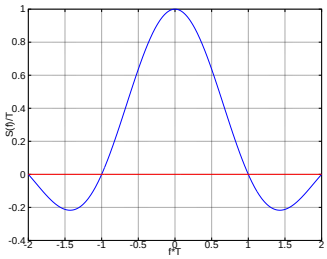
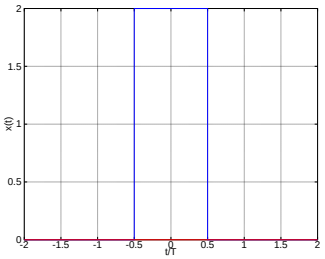
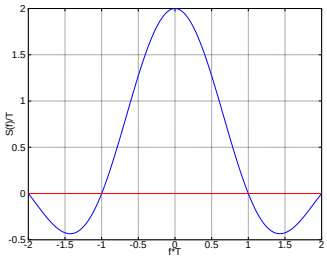
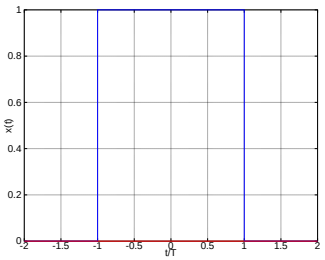
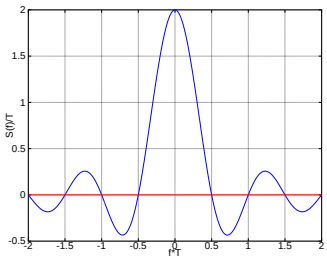
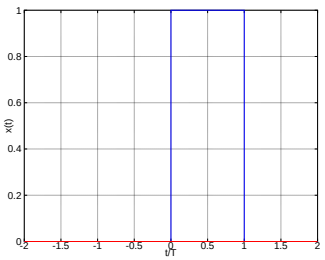
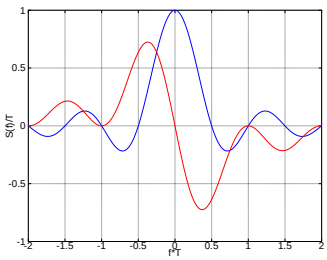
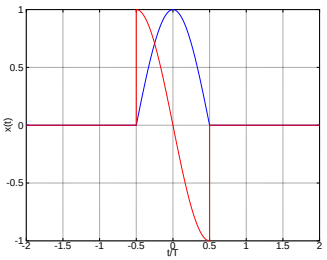
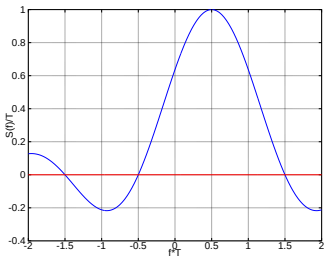
$$= \int_{-\infty}^{\infty} x(t)e^{-j2\pi f' t}dt, \qquad f' = f - f_0 \tag{4.16}$$

$$= S(f') \tag{4.17}$$

$$= S(f - f_0)\,. \tag{4.18}$$

Eine Modulation eines Signals schlägt sich also als Frequenzverschiebung nieder.
Diese ersten drei Sätze sind in Tab. 4.3 bildlich dargestellt.

Tabelle 4.3: Wichtige Sätze zur Fouriertransformation in bildlicher zusammenfassender Darstellung

Satz	Zeitbereich	Frequenzbereich
Original		
Linearität		
Ähnlichkeit		
Verschiebung		
Modulation		

Satz zur Ableitung Mit

$$\mathcal{F}\{x(t)\} = S(f) \tag{4.19}$$

folgt

$$\mathcal{F}\left\{\frac{dx(t)}{dt}\right\} = 2\pi i f S(f) \tag{4.20}$$

Beweis

$$\mathcal{F}\left\{\frac{dx(t)}{dt}\right\} = \int_{-\infty}^{\infty} \frac{dx(t)}{dt} e^{-j2\pi ft} dt \tag{4.21}$$

$$= \int_{-\infty}^{\infty} \frac{d\{x(t)e^{-j2\pi ft}\}}{dt} - x(t)\frac{d\{e^{-j2\pi ft}\}}{dt} dt \tag{4.22}$$

$$= \int_{-\infty}^{\infty} 2\pi f x(t) e^{-j2\pi ft} dt \tag{4.23}$$

$$= 2\pi i f S(f) \tag{4.24}$$

Es wurde hier die Ableitung von Produkten von Funktionen verwendet und die Tatsache, dass das Integral über die Ableitung beider Funktionen (erster Term) verschwinden muss. Dieses ist so, weil die Funktionen als solche an den Enden im Unendlichen verschwinden müssen, denn sonst wäre die Fouriertransformation gar nicht ausführbar.

Parsevals Theorem Mit

$$\mathcal{F}\{x(t)\} = S(f) \tag{4.25}$$

folgt

$$\int_{-\infty}^{\infty} |x(t)|^2\} dt = \int_{-\infty}^{\infty} |S(f)|^2 \, df \tag{4.26}$$

Beweis, den wir anders herum starten:

$$\int_{-\infty}^{\infty} |S(f)|^2 \, df = \int_{-\infty}^{\infty} S(f) \, S^*(f) \, df \tag{4.27}$$

$$= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt \int_{-\infty}^{\infty} x^*(t') e^{j2\pi ft'} dt' \right\} df \tag{4.28}$$

$$= \iiint_{-\infty}^{\infty} x(t) x^*(t') e^{-j2\pi f(t-t')} dt dt' df \tag{4.29}$$

$$= \iint_{-\infty}^{\infty} x(t) x^*(t') \delta(t-t') dt dt' \tag{4.30}$$

$$= \int_{-\infty}^{\infty} x(t) x^*(t) dt \tag{4.31}$$

$$= \int_{-\infty}^{\infty} |x(t)|^2 \} dt \tag{4.32}$$

wobei die Integrationsreihenfolge getauscht wurde und die Distribution δ (Dirac Stoß) verwendet wurde. Für die gilt:

$$\delta(t) = \begin{cases} \infty, \, t = 0 \\ 0, \, t \neq 0 \end{cases} \tag{4.33}$$

$$\text{und} \int_{-\infty}^{\infty} \delta(t) dt = 1. \tag{4.34}$$

Merken muss man sich hier, dass *die gesamte Energie einer Funktion im Zeit- und Spektralbereich die gleiche ist.*

Faltungstheorem Mit

$$\mathcal{F}\{x(t)\} = S(f), \qquad \text{und} \qquad \mathcal{F}\{y(t)\} = T(f) \tag{4.35}$$

folgt

$$\mathcal{F}\left\{x(t) * y(t-\tau) d\tau\right\} = \mathcal{F}\left\{ \int_{-\infty}^{\infty} x(\tau) y(t-\tau) d\tau \right\} \qquad = S(f)T(f) \tag{4.36}$$

Beweis:

$$\mathcal{F}\left\{\int_{\infty}^{\infty} x(\tau)y(t-\tau)d\tau\right\} = \int_{-\infty}^{\infty} x(\tau)\mathcal{F}\left\{y(t-\tau)d\tau\right\} \tag{4.37}$$

$$= \int_{-\infty}^{\infty} x(\tau)e^{-j2\pi f\tau}T(f)d\tau \tag{4.38}$$

$$= S(f)T(f) \tag{4.39}$$

Eine Faltung im Zeitbereich kann also durch einfache Multiplikation der Frequenzbereichsfunktionen dargestellt werden. Dieser Satz ist die Basis für viele, eigentlich alle, Filteroperationen in Zeit- und Frequenzbereich.

Autokorrelation Mit

$$\mathcal{F}\{x(t)\} = S(f) \tag{4.40}$$

folgt

$$\mathcal{F}\left\{\int_{\infty}^{\infty} x(\tau)x^*(\tau-t)d\tau\right\} = |S(f)|^2 \tag{4.41}$$

Beweis:

$$\mathcal{F}\left\{\int_{\infty}^{\infty} x(\tau)x^*(\tau-t)d\tau\right\} = \mathcal{F}\left\{\int_{\infty}^{\infty} x(\tau)x^*\left(-(t-\tau)\right)d\tau\right\} \tag{4.42}$$

$$= S(f)S^*(f) \tag{4.43}$$

wobei noch verwendet wurde, dass

$$\mathcal{F}\{x^*(t)\} = \int_{-\infty}^{\infty} x^*(t)e^{-j2\pi ft}dt \tag{4.44}$$

$$= \left\{\int_{\infty}^{\infty} x(t)e^{j2\pi ft}dt\right\}^* \tag{4.45}$$

$$= \left\{\int_{\infty}^{\infty} x(-t')e^{-j2\pi ft'}dt'\right\}^* \tag{4.46}$$

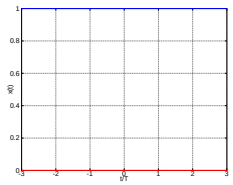
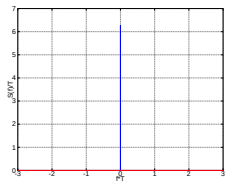
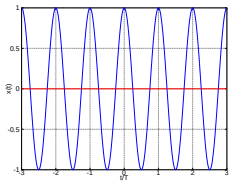
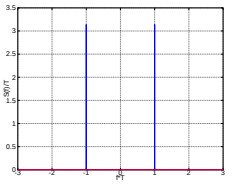
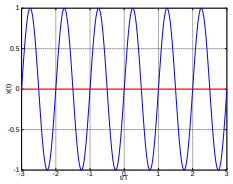
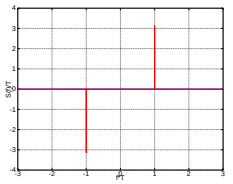
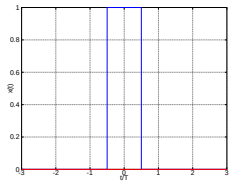
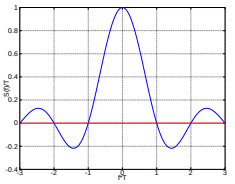
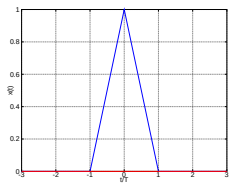
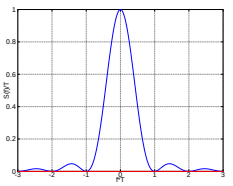
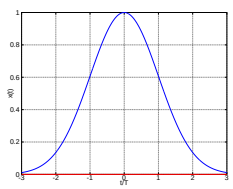
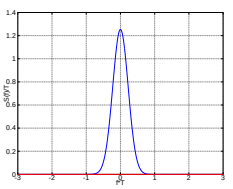
$$= S^*(-f) \tag{4.47}$$

ist.

4.2.1.3 Einige Beispielfunktionen zur Fouriertransformation

Tabellarisch sollen nun einige Beispielfunktionen in Tab. 4.4 angegeben werden

Tabelle 4.4: Beispielfunktionen zur Fouriertransformation (blau: Real-, rot: Imaginärteil)

Zeitbereich		Frequenzbereich	
$x(t) =$		$S(f)$	
1		$\delta(f)$	
$\cos(2\pi f_0 t)$		$\frac{1}{2}\left(\delta(f-f_0)+\delta(f+f_0)\right)$	
$\sin(2\pi f_0 t)$		$\frac{j}{2}\left(\delta(f-f_0)-\delta(f+f_0)\right)$	
$\mathrm{rect}\left(\frac{t}{T}\right)$		$\frac{\sin(\pi f T)}{\pi f}$	
$\mathrm{tri}\left(\frac{t}{T}\right)$		$\frac{\sin^2(\pi f T)}{(2\pi f)^2 T}$	
$e^{-t^2/(2T^2)}$		$\sqrt{\frac{\pi}{2}}Te^{-\pi^2 f^2 T^2}$	

4.2.1.4 Diskrete Fourier-Transformation

Das Arbeitspferd der digitalen Signalverarbeitung und damit der Übertragungstechnik ist die Diskrete Fourier-Transformation [36]. Der Ursprung ist die Begrenzung und Diskretisierung der beiden Gleichungen (4.1) und (4.2) zu

$$S(f) = \mathcal{F}\{x(t)\} \approx \sum_{n=-N}^{N} x(n \cdot \Delta t) e^{-j2\pi f n \Delta t} \Delta t \tag{4.48}$$

und Rücktransformation

$$x(t) = \mathcal{F}^{-1}\{S(f)\} \approx \sum_{k=-N}^{N} S(k \cdot \Delta f) e^{j2\pi k \Delta f t} \Delta f. \tag{4.49}$$

Für die abgetasteten Werte kann man auch abkürzend

$$x_n = x(n\Delta t) \tag{4.50}$$

$$S_k = S(k\Delta f) \tag{4.51}$$

schreiben und die diskreten Abtastpunkte sind in Zeit und Frequenz dann

$$t_n = n\Delta t, \qquad f_k = k\Delta f = \frac{k}{2N\Delta t} \tag{4.52}$$

wobei der letzte Teil den Zusammenhang zwischen der Abtastung im Zeitbereich und im Frequenzbereich darstellt. Damit folgt dann

$$S(f_k) = \Delta t \sum_{n=-N}^{N} x_n e^{-j2\pi f_k n \Delta t} \tag{4.53}$$

$$= \Delta t \sum_{n=-N}^{N} x_n e^{-j2\pi k n/N} \tag{4.54}$$

und schließlich in einem Übergang auf eine Zahlenfolge

$$S_k = \sum_{n=-N}^{N} x_n e^{-j2\pi k n/N} \tag{4.55}$$

$$S(f_k) = \Delta t S_k, \tag{4.56}$$

hier verbindet die letzte Zeile die diskrete Fouriertransformierte mit der Fouriertransformation an der Stelle f_k. Rücktransformation erfolgt mit

$$x_n = \frac{1}{N} \sum_{k=-N}^{N} S_n e^{j2\pi k n/N} \tag{4.57}$$

und der Skalierung mit $1/N$, die hier notwendig wird.
Es sei wiederholt, dass die Abtastung einer Funktion hoch genug erfolgen muss, um die gewollten Frequenzen mitzunehmen. Wenn eine Frequenzkomponente f_{max} der Ursprungsfunktion mitzunehmen ist, dann muss diese nach dem Nyquist-Kriterium mit mindestens der doppelten Frequenz (also $2f_{max}$) abgetastet worden sein. Mit einem Zahlenbereich von $-N...N$ liegen die Frequenzen nach (Gl. 4.52) im Bereich $\pm\frac{1}{2\Delta t}$ und die feinste Frequenzauflösung ist $\frac{1}{2N\Delta t}$. Das heißt also, je schneller abgetastet wird (Δt klein), desto höher werden die Frequenzkomponenten, die in der abgetasteten Funktion noch vorhanden sind, desto gröber wird allerdings auch die Frequenzauflösung. Die Anteile werden nur noch in "Eimern" (Bins) der Größe $\frac{1}{2N\Delta t}$ zusammengefasst. Diese Bins können durch Erhöhung der Anzahl der Abtastpunkte, also letztlich durch Erhöhung der gesamten Abtastdauer (Sammeldauer) $N\Delta t$, verfeinert werden.

4.2.1.5 Leistung und Energie

Die reine Fouriertransformierte ist in vielen Fällen für Ingenieure das, was sie ist: eine mathematische Transformation ohne große Anschauungsmöglichkeiten und damit von begrenzter Aussagekraft. Man spricht auch selten von der "Fouriertransformierten" sondern meist von dem "Spektrum" eines Signals und meint damit die Leistung (oder Energie) des Signals in einem gewissen Frequenzbereich. Es steht die Frage im Raum: "Wieviel Watt (oder Milliwatt) sind im Frequenzbereich von f_1 bis f_2?" Streng genommen ist dieses dann die Frage nach einer "spektralen Dichte" , denn die Frage impliziert eine Antwort mit der Dimension Leistung/Frequenz(bereich) $[P]/[f]$.
Mathematisch liefert die Fouriertransformation schon eine Spannungsdichte mit der Dimension $[U]/[f]$ erhält man mit dem oben schon Gesagten diese Antwort aus der Fouriertransformierten mit

$$P' = \int_{f_1}^{f_2} |S(f)|^2 df. \tag{4.58}$$

4.3 Komplexe Signale im Basisband

Bislang war die mindestens still schweigende Annahme, dass nur die drei Zustände -1, 0 und 1 vorkommen dürfen, wobei immer versucht wurde, die eigentlich digitalen Daten Null und Eins als -1 und +1 darzustellen und zwar so, dass über lange Zeiten kein Mittelwert der Sequenz existiert. Der Null-Zustand ist also als Übergangszustand verwendet worden. Soll die Übertragungsrate erhöht werden, so ist einfach die Taktrate zu erhöhen.

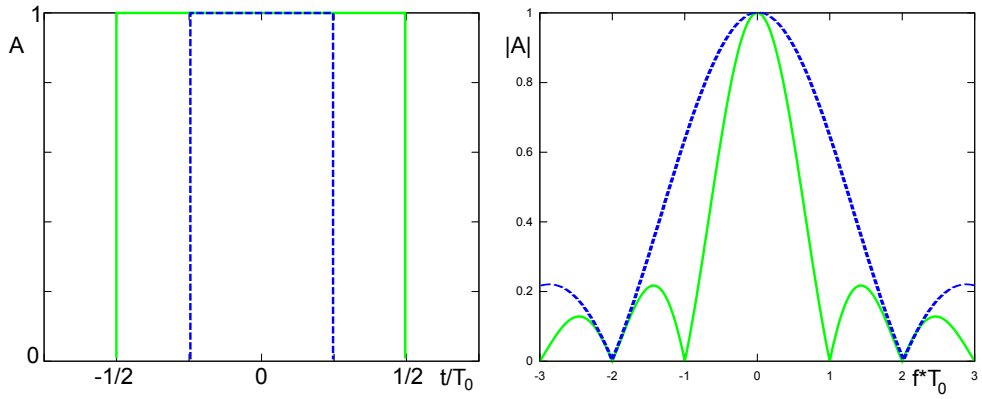


Abbildung 4.4: Aufweitung des Spektrums (links) durch Verkürzung der Signalperiode (rechts).

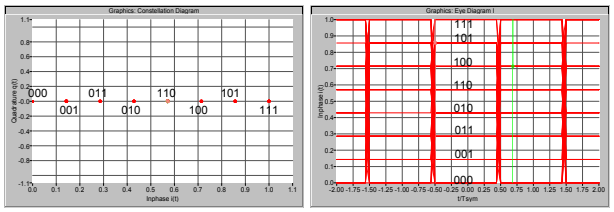


Abbildung 4.5: Mehrere Zustände in einer 8-ASK mit Augendiagramm in Gray-Kodierung.

Damit einher geht zwingend eine Aufweitung des Spektrums (Abb. 4.4). Dieses kann natürlich problematisch werden, wenn für die Übertragung nur eine bestimmte Bandbreite vorgesehen ist, weil z.B. nur ein gewisser Kanal in der Funkübertragung gesetzlich erlaubt wurde, oder ggf. ein Kabel bei noch höherer Übertragungsraten an seine Grenzen kommen würde.
Eine Alternative zur Erhöhung der Datenrate ist die Erhöhung der Anzahl der in einer Taktperiode übertragenden Information. Hierzu werden in einer Periode z.B. nicht nur die Zustände -1 und +1 erlaubt, sonder Zwischenzustände, wie in Abb. 4.5 für eine Amplitude Shift Keying (ASK) mit acht Zuständen gezeigt. In einer Taktperiode können hier also $N = \mathrm{ld} N_{zus} = 3$ Bit übertragen werden. Der Preis ist, dass die Zustände enger beieinander liegen, was aber hier - ohne Rauschen - noch nicht interessiert.
Zu einer weiteren oder auch alternativen Art der Erhöhung der Datenrate muss man auf die Eigenschaften des Sinus und Kosinus zurückgreifen, die selbst bei der gleichen Frequenz orthogonal zueinander stehen (Abb. 4.6).

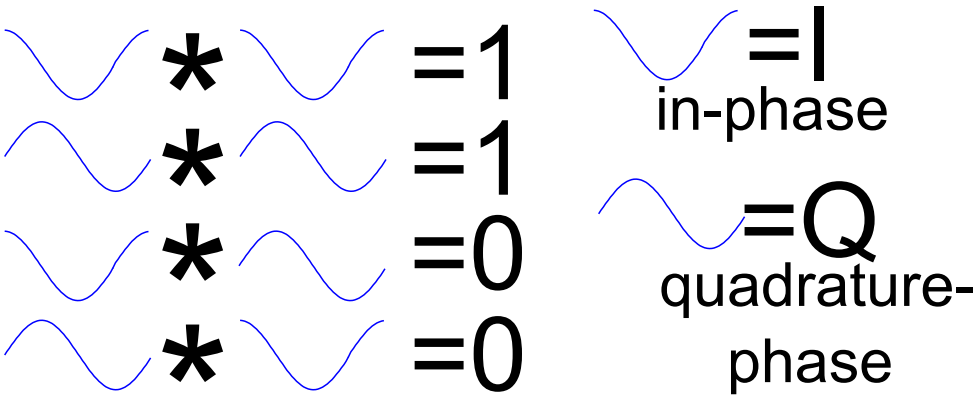


Abbildung 4.6: Grafische Darstellung der Orthogonalität von Sinus und Kosinus.

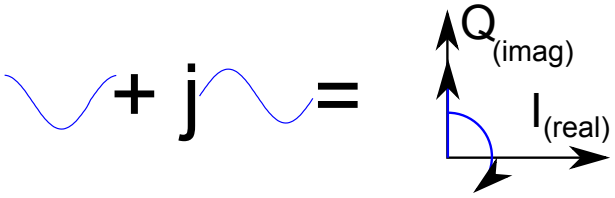


Abbildung 4.7: Darstellung von Signalen auf Sinus und Kosinus in der komplexen Ebene.

$$\int_0^T \sin\left(\frac{2\pi}{T}t\right) \times \sin\left(\frac{2\pi}{T}t\right) dt = 1 \tag{4.59}$$

$$\int_0^T \cos\left(\frac{2\pi}{T}t\right) \times \cos\left(\frac{2\pi}{T}t\right) dt = 1 \tag{4.60}$$

$$\int_0^T \sin\left(\frac{2\pi}{T}t\right) \times \cos\left(\frac{2\pi}{T}t\right) dt = 0 \tag{4.61}$$

$$\int_0^T \cos\left(\frac{2\pi}{T}t\right) \times \sin\left(\frac{2\pi}{T}t\right) dt = 0 \tag{4.62}$$

Dieses kann man ausnutzen, indem man den Sinus mit Informationen moduliert und ebenso den Kosinus und diese beide gleichzeitig auf den Kanal schickt (Abb. 4.7). Dazu wird die De-Moivre'sche Formel

$$e^{j\omega t} = \cos(\omega t) + j\sin(\omega t) \tag{4.63}$$

ausgenutzt, so dass man den Signalen auf dem Sinus die imaginäre Achse (Q=Quadratur-Phase) und dem Kosinus die reelle Achse (I=In-Phase) zuweist. Eine einfache $2 \times 2 = 4$-

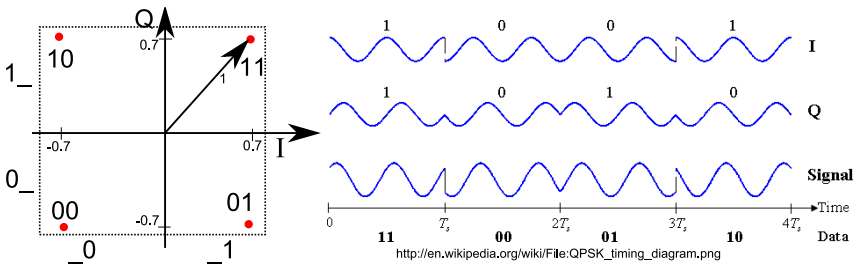


Abbildung 4.8: QPSK mit ihrem Zeitsignal (links:Konstellationsdiagramm).

wertige Modulation ist in Abb. 4.8 darstellt. Diese wird als Quadrature Phase Shift Keying (QPSK) bezeichnet und kann pro Symboltakt vier Zustände also zwei Bit übertragen. Die Darstellung ist das bekannte Konstellationsdiagramm. Hervorzuheben ist, dass die imaginäre Achse nur in der menschlichen Vorstellung, nicht aber in der Realität, existiert. Dort gibt es nur reelle Signale, so dass bei $N+1$ Symbolen Folgendes übertragen wird:

$$S(t) = \sum_{n=0}^{N} s_I(t-nT)\cos\left(\omega(t-nT)\right) + s_Q(t-nT)\sin\left(\omega(t-nT)\right) \tag{4.64}$$

$$= \mathrm{Re}\left\{\sum_{n=0}^{N}\left[s_I(t-nT) - js_Q(t-nT)\right]e^{j\omega(t-nT)}\right\}. \tag{4.65}$$

Mit dem Minuszeichen in $-js_Q$ werden wir in Zukunft flexibel umgehen. Die einzelnen Symbolformen s_I und s_Q sind hier definiert als

$$s_{(I,Q)}(t) = d_{(I,Q)} \times \begin{cases} 1, & 0 < t < T \\ \frac{1}{2}, & t = 0, t = T \\ 0, & \text{sonst.} \end{cases} \tag{4.66}$$

und hier sind $d_{(I,Q)}$ die Bitdarstellungen (hier -1 und 1).
Konsequent fortgesetzt werden die Darstellungen nach Abb. 4.5 und 4.8 kombiniert, so ergeben sich Modulationsmöglichkeiten in zwei Dimensionen, in denen Phase und Amplitude moduliert werden. Es entstehen die Quadrature Amplitude Modulation (QAM), die in Abb. 4.9 gezeigt werden.
Technisch werden QAM mit bis zu 256 bzw. 1024 Zuständen eingesetzt. Hiermit können dann 8 (10) Bit in einem Symbol übertragen werden. Für eine rechteckförmige Darstellung der Impulse nach Gl. 4.66 ist in Abb. 4.10 nochmals das Spektrum gezeigt. Auch hier ist natürlich die Ausdehnung bis zur ersten Nullstelle im Spektrum $-1/T...1/T$ und damit liegt gegenüber der einfachen ASK bzw. BPSK kein Spektrumsgewinn vor, jedoch werden im Unterschied zu den vorgenannten Modulationsarten 2 Bit pro Symbol übertragen, womit die Übertragungsrate verdoppelt ist.

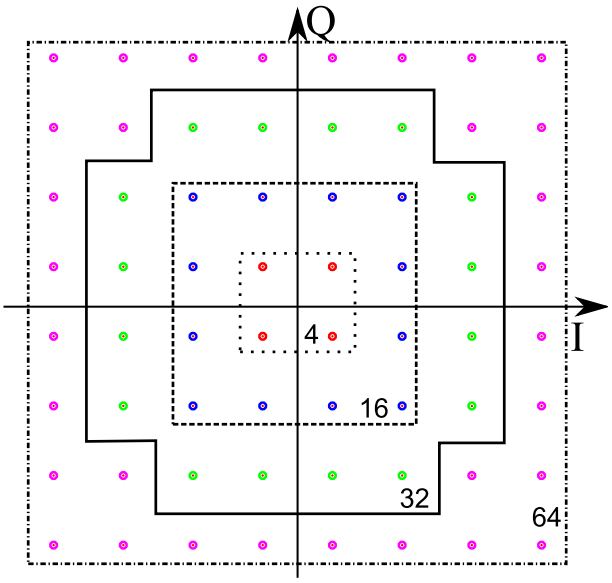


Abbildung 4.9: QAM in mehreren Varianten von der QPSK (rot) über 16-QAM (blau) 32-QAM (grün) zur 64-QAM (violett) (Konstellationsdiagramm).

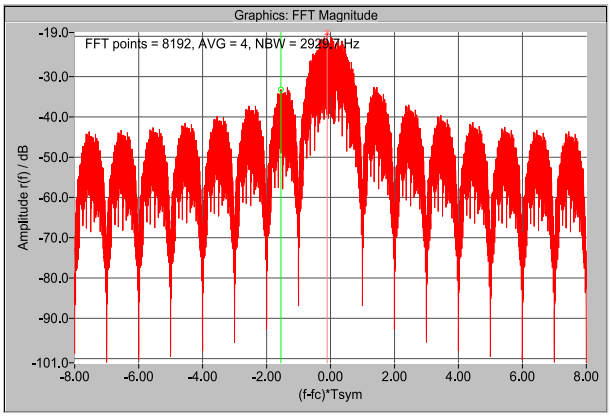


Abbildung 4.10: Spektrum einer QPSK-Modulation.

Ein prinzipieller Modulator zur Erzeugung der komplexen Modulationsarten ist in Abb. 4.11 gezeigt.

4.4 Bitfehler und Störungen durch Rauschen

4.4.1 Gauss'sches Rauschen

Bislang wurde der Übertragungskanal nur aus mikroskopischer Sicht beschrieben. Es wurde dargestellt, dass bei Übertragung die Signale gedämpft und durch unterschiedliche

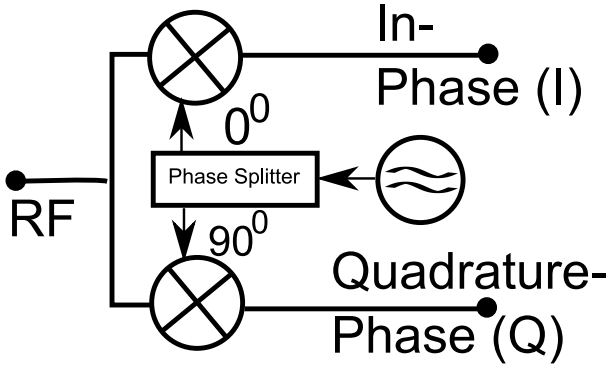


Abbildung 4.11: Einfacher QPSK/ QAM modulator.

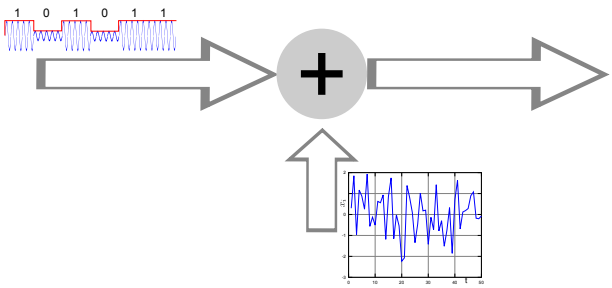


Abbildung 4.12: Darstellung des einfachen AWGN-Kanals.

Ausbreitungsgeschwindigkeiten (Dispersion) verzerrt wurden. Nun soll der erstere Effekt in seiner Auswirkung auf die digitale Übertragung betrachtet werden.
In der Modellbildung geht man davon aus, dass dem Nutzsignal Rauschen zugefügt wird. Das ist natürlich real nicht der Fall, sondern das Rauschen bleibt gleich, nur das Signal wird schwächer. Im Endeffekt führen aber beide Betrachtungen auf gleiche Ergebnisse. Die Betrachtung führt auf den Additive White Gaussian Noise (AWGN)-Kanal nach Abb. 4.12.
Dem Signal wird hier ein stochastisches, also wahllos in der Amplitude u fluktuierendes Signal zugeführt. Nimmt man an, dass dieses Signal im Wesentlichen aus thermischen Prozessen herrührt (thermisches Rauschen), dann kann man davon ausgehen, dass die Wahrscheinlichkeitsdichtefunktion einer Gauss'schen Glockenfunktion ($\approx e^{u^2/\sigma^2}$) nach folgt. Es sei betont, dass es sich hier nicht um das zeitliche Verhalten der Funktion, sondern nur um die Darstellung der Wahrscheinlichkeit handelt, mit der eine bestimmte Amplitude u anzutreffen ist (Abb. 4.13).
Analysen zeigen, dass die Standardabweichung der mittleren Leistung entspricht (wenn

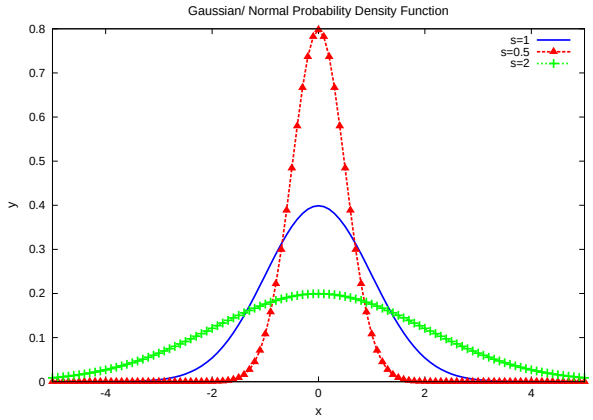


Abbildung 4.13: Repräsentation einer Gaus'schen Wahrscheinlichkeitsdichtefunktion mit unterschiedlichen Standardabweichungen $\sigma = 0.5; 1; 2$.

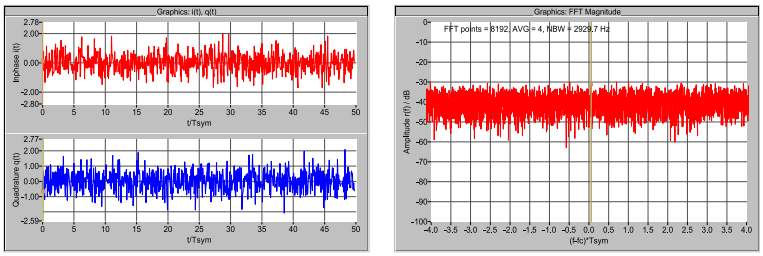


Abbildung 4.14: Zeitlicher und spektraler Verlauf eines weißen Rauschsignals.

die Amplitude die Dimension $[u] = \sqrt{\mathrm{W}}$ hat) und damit ist

$$N = \sigma^2 = kTB \tag{4.67}$$

mit T der Temperatur, $k = 1{,}380\,6488(13) \times 10^{-23}\,\mathrm{Ws/K}$ der Boltzmannkonstante und B der Bandbreite. Ein solcher Gauss'scher Prozess ist in Abb. 4.14 dargestellt. Man sieht, dass der zeitliche Verlauf tatsächlich regellos ist und das Spektrum ideal unendlich ausgedehnt ist. Letzteres macht natürlich physikalisch keinen Sinn, da es ja unendlicher Energie entspricht, kann aber in der Modellbildung für unsere Zwecke gut genutzt werden.

4.4.2 Bitfehler

Wird nun einem digitalen Signal aus Einsen und Nullen Rauschen überlagert, so werden die ehemals scharfen Signalwerte verschwimmen (Abb. 4.15). Es wird Fälle geben, bei denen das Rauschen aus einer Eins eine Null macht und umgekehrt, also die Entscheidungsschwelle in der Mitte zwischen beiden Punkten durchschritten wird. In diesem Fall

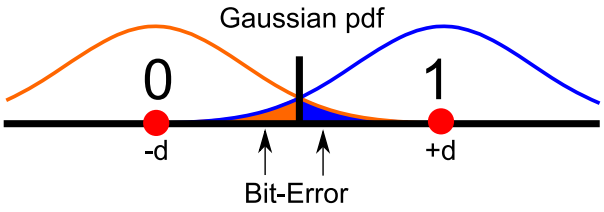


Abbildung 4.15: BPSK mit überlagertem Rauschen

kann der Empfänger nicht anders als sich für den falschen Wert zu entscheiden und damit einen Bitfehler zu produzieren.
Je größer das Rauschen ist, also je breiter die Verteilung der Rauschamplituden, desto größer ist die Wahrscheinlichkeit auf einer Fehlentscheidung, die Bitfehlerrate steigt. Diese Wahrscheinlichkeit wird durch die Flächen unter der Kurve angegeben und lässt sich mit einem Rauschsignal von

$$n(\xi) = \frac{1}{\sigma_n\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\xi}{\sigma_n}\right)^2} \tag{4.68}$$

bei einem Signal mit der Amplitude $\pm d$ berechnen zu

$$P_b = P\{n < -d\} = P\{n > +d\} \tag{4.69}$$

$$P_b = \frac{1}{\sqrt{2\pi}\sigma_n} \int_d^\infty e^{-\frac{1}{2}\frac{\xi^2}{\sigma_n^2}} d\xi = \frac{1}{2}\mathrm{erfc}\left(\frac{d}{\sqrt{2}\sigma_n}\right). \tag{4.70}$$

Hiermit wurde die komplementäre Fehlerfunktion erfc eingeführt. Mit Rauschleistungen $N \propto \sigma^2$ und Signalleistungen $S \propto d^2$ ergibt sich dann

$$P_b = \frac{1}{2}\mathrm{erfc}\sqrt{\frac{1}{2}\frac{S}{N}} \tag{4.71}$$

als Tiefpass gefiltertes Signal und im Basisband mit symmetrischer (positiver und negativer Frequenzaufteilung) $N = N_0/2$

$$P_b = \frac{1}{2}\mathrm{erfc}\sqrt{\frac{E_S}{N_0}} \tag{4.72}$$

im Basisband mit der Signalenergie E_S. Als Bezugsgröße, also Rauschleistung, verwendet man nicht das einfache Integral über das Quadrat der Gaussfunktion, sondern es ist nur die einseitige Rauschleistung(sdichte) in dem betrachteten Frequenzbereich gefragt. Diese ergibt $2\delta^2 = N_0 = 2N$ [9, S.160].

4.4.3 Kanalkapazität

Allein aus der Überlegung folgt, dass ein Kanal umso besser ist, je größer das Signal-Rauschverhältnis E_S/N_0 ist. Ist das Verhältnis sehr klein, so kommt es häufig zu Bitfehlern, die Information wird am Empfänger nicht verstanden und muss ggf. wiederholt werden. Die Zeit, in der die fehlerhaften Bits wiederholt werden, steht nicht zur Übertragung neuer Informationen zur Verfügung. Das Rauschen senkt die verfügbare Kapazität des Kanals.
Ist die Signal Noise Ratio (SNR) sehr gut, so kann man sich u.U. gemäß Abb. 4.5 sogar trauen, zu einer Symbolzeit differenziertere Informationen zu übertragen, die im Empfänger sauber getrennt werden können. Die Kanalkapazität wird besser ausgenutzt und steigt damit.
Genauere Untersuchungen zeigen, dass die Kapazität eines Kanals $C \propto \ln(E_S/N)$ ist und nach Claude Shannon als

$$C = B \log_2 \left[1 + \frac{E_S}{N}\right] \tag{4.73}$$

mit der Bandbreite B des Kanals geschrieben werden kann. Eine Näherung für gute Kanäle (mit wenig Rauschen) ist

$$C \approx \frac{1}{3} B \left.\frac{E_S}{N}\right|_{dB} . \tag{4.74}$$

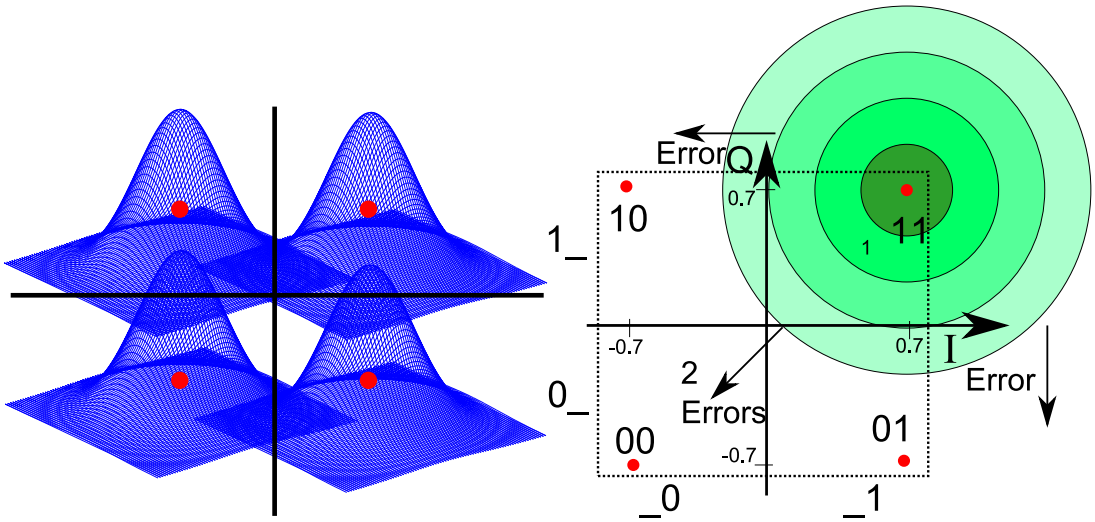
Der Ausdruck (Gl. 4.73) stellt allerdings lediglich die obere Grenze dar, die erst vor wenigen Jahren wirklich mit den Turbo-Codes erreicht wurde.
Weitere Konsequenzen der Aussage sind

- Bandbreite und SNR können ausgetauscht werden, also kann man bei großer Bandbreite auch über einen recht schlechten Kanal noch eine beachtliche Übertragungsrate erzielen.

- Einen sehr guten Kanal kann man gut ausnutzen, indem man auch bei niedriger Bandbreite Information dicht gepackt in z.B. Modulationsformen hoher Ordnung überträgt.

4.4.4 Bitfehlerraten für komplexe Modulationsformen

Nach den oben geführten prinzipiellen Überlegungen zur Bitfehlerrate soll nun eine quantitative Untersuchung zu den verschiedenen Modulationsarten folgen. Als Abschätzung



(a) (b)

Abbildung 4.16: Rauschen in QPSK-Modulation (a) bildlich, (b) zur Analyse geeignet.

der Wahrscheinlichkeit für einen Bitfehler wird nur der wahrscheinlichste Fehlerfall angenommen (Abb. 4.16(a)), nämlich der, in dem durch Rauschen ein Übergang in einen benachbarten Quadranten stattfindet, nicht aber über die Diagonale hinweg. In der QPSK Modulation ist die Energie pro Symbol die doppelte der BPSK, wenn die horizontalen bzw. vertikalen Amplitudenwerte der Punkte jeweils $\pm d$ betragen. Es liegen ja zwei orthogonale Entscheidungen vor:

$$E_S|_{QPSK} = 2\, E_S|_{BPSK} = 2E_b. \tag{4.75}$$

Damit ist die Bitfehlerwahrscheinlichkeit

$$P_b = \frac{1}{2}\mathrm{erfc}\sqrt{\frac{E_b}{N_0}} = \frac{1}{2}\mathrm{erfc}\sqrt{\frac{E_S}{2N_0}} \tag{4.76}$$

und für höherwertige Modulationsformen kann die Bitfehlerwahrscheinlichkeit mit

$$P_S \approx \mathrm{erfc}\left(\sqrt{\frac{E_S}{N_0}}\sin\frac{\pi}{M}\right) \tag{4.77}$$

für eine $M-$wertige Phase Shift Keying (PSK) angegeben werden. Für die Werte $M = 2$ und $M = 4$ ergeben sich zwei Auffälligkeiten, nämlich, dass zum einen für $M = 2$ (Gl. 4.77) nicht der genauen Berechnung nach (Gl. 4.76) entspricht. Hier fehlt ein Faktor zwei. Dieser ist der Näherung in (Gl. 4.77) geschuldet.

Mit dem Zusammenhang zwischen Symbolenergie und Energie pro Bit

$$E_S = \mathrm{ld}(M) E_b \qquad \Leftrightarrow \qquad E_b = \frac{E_S}{\mathrm{ld}(M)} \tag{4.78}$$

ergibt sich nach (Gl. 4.77) für BPSK und QPSK die gleiche Abhängigkeit der Bitfehlerrate in Abhängigkeit vom Signal-Rausch-Abstand pro Bit zu

$$P_b \approx \mathrm{erfc}\sqrt{\frac{E_b}{N_0}}. \tag{4.79}$$

Allgemein erhält man

$$P_S \approx \mathrm{erfc}\left(\sqrt{\frac{\log_2(M) E_b}{N_0}} \sin\frac{\pi}{M}\right) \tag{4.80}$$

Für die QAM folgt nach [9, S. 278]

$$P_S \approx 1 - \left(\left(1 - \frac{1}{\sqrt{M}}\right) \mathrm{erfc}\left(\sqrt{\frac{3E_S}{(M-1)2N_0}}\right)\right)^2, \tag{4.81}$$

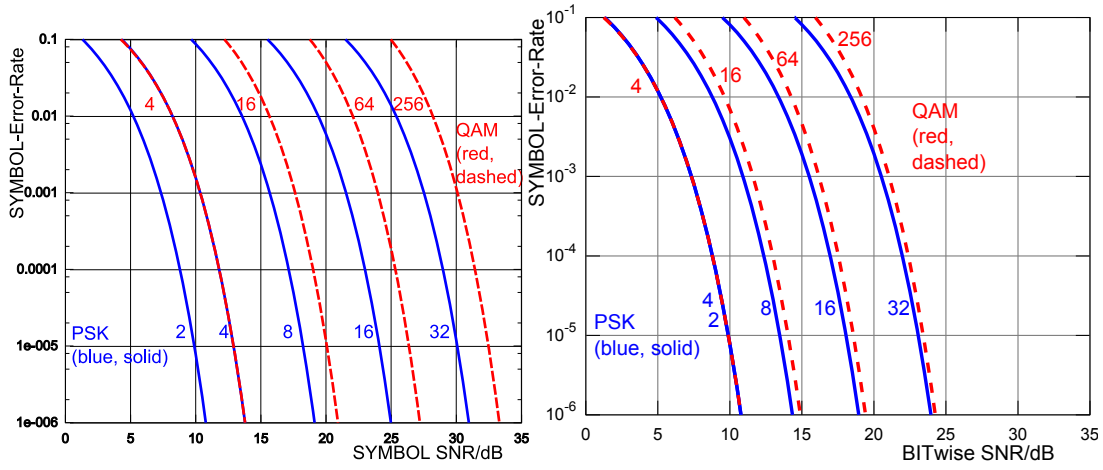
wobei auch hier die Fehlerrate über der Symbolenergie aufgetragen ist. Bezogen auf die Energie pro Bit ergibt sich dann

$$P_S \approx 1 - \left(\left(1 - \frac{1}{\sqrt{M}}\right) \mathrm{erfc}\left(\sqrt{\frac{3\log_2(M) E_b}{(M-1)2N_0}}\right)\right)^2. \tag{4.82}$$

Der Übergang auf Symbolfehlerrate ist für unsere Näherungen nicht relevant, da hier genähert wird, dass in Gray-Kodierung ein Symbolfehler auch genau einem Bitfehler entspricht und der Übergang zum übernächsten Nachbarzustand nicht stattfindet. Die Fehlerraten für die verschiedenen Modulationsarten sind in Abb. 4.17 aufgetragen.

4.5 Spektrum und Impuls

Dieses Kapitel ist stark angelehnt an [5].
Wie schon in Abschnitt 4.2.1 erläutert und in Abb. 4.3 dargestellt, führt die Übertragung "roher" digitaler Datensignale nach Abb. 4.2 zu sehr weit ausgedehnten Spektren. In Kanälen, in denen eine nahezu unendliche Bandbreite allein für einen Übertragung zur Verfügung steht, wie z.B. in einem guten Kabel, ist dieses kein Problem und es sind keine weiteren Maßnahmen nötig. In der Praxis ist der Kanal jedoch bandbegrenzt, was wir nun untersuchen wollen.



(a) (b)

Abbildung 4.17: Leistungsfähigkeit verschiedener Kodierungsarten in rauschenden Kanälen (a) dargestellt über dem Signal-Rausch-Abstand pro Symbol und (b) Signal-Rausch-Abstand pro Bit.

Idealerweise wünschen wir uns einen verzerrungsfreien, unendlich schnellen Kanal, bei dem am Ausgang das Eingangssignal unverfälscht erscheint:

$$y(t) = x(t), \qquad \text{wodurch} \tag{4.83}$$

$$y(t) = h_0(t) * x(t), \text{mit } h_0(t) = \delta(t) \tag{4.84}$$

ist. $\delta(t)$ ist der Diracstoß, der hier die Impulsantwort des Kanals mit h_0 beschreibt (Abb. 4.18:

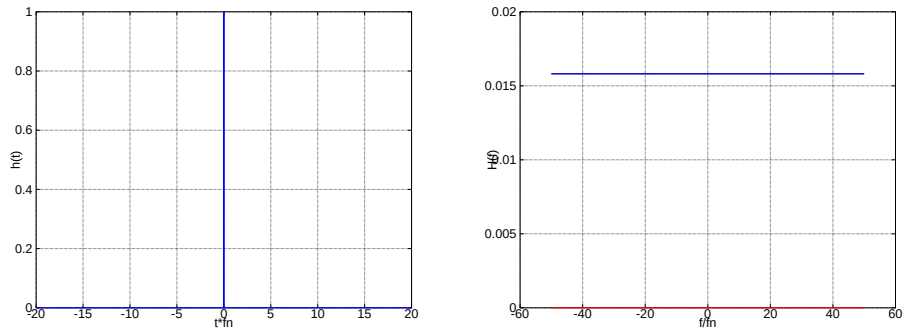
$$h_0(t) = \delta(t) \qquad \circ\!\!-\!\!\bullet \qquad H_0(f) = 1. \tag{4.85}$$

Dass dieses nicht so sein kann, wissen wir schon lange aus der Betrachtung der Leitungen. Mindestens liegt im Kanal (Kabel) eine endliche Ausbreitungsgeschwindigkeit vor, wodurch das Signal am Ausgang eine verzögerte und skalierte Variante des Signals am Eingang ist:

$$y(t) = a \times x(t - t_0) \tag{4.86}$$

und damit

$$h_1(t) = a \times \delta(t - t_0) \qquad \circ\!\!-\!\!\bullet \qquad H_1(f) = a \times e^{-j2\pi f t_0}. \tag{4.87}$$



(a) (b)

Abbildung 4.18: Idealer Kanal (a) Impulsantwort und (b) Spektrum für unendliche Bandbreite und keine Verzögerung.

Immer noch hat der Kanal eine unendliche Bandbreite, die Amplitude von H_1 fällt nicht mit steigender Frequenz ab. Auch dieses entspricht nicht der Erfahrung. Stattdessen hat man es mit einem bandbegrenzten Kanal zu tun:

$$H_2(f) = \begin{cases} a \times e^{-j2\pi f t_0}, & \text{für } |f| < |f_N| \\ \frac{a}{2} \times e^{-j2\pi f t_0}, & \text{für } |f| = |f_N| \\ 0 & \text{sonst} \end{cases} \tag{4.88}$$

$$\bullet\!\!-\!\!\circ \tag{4.89}$$

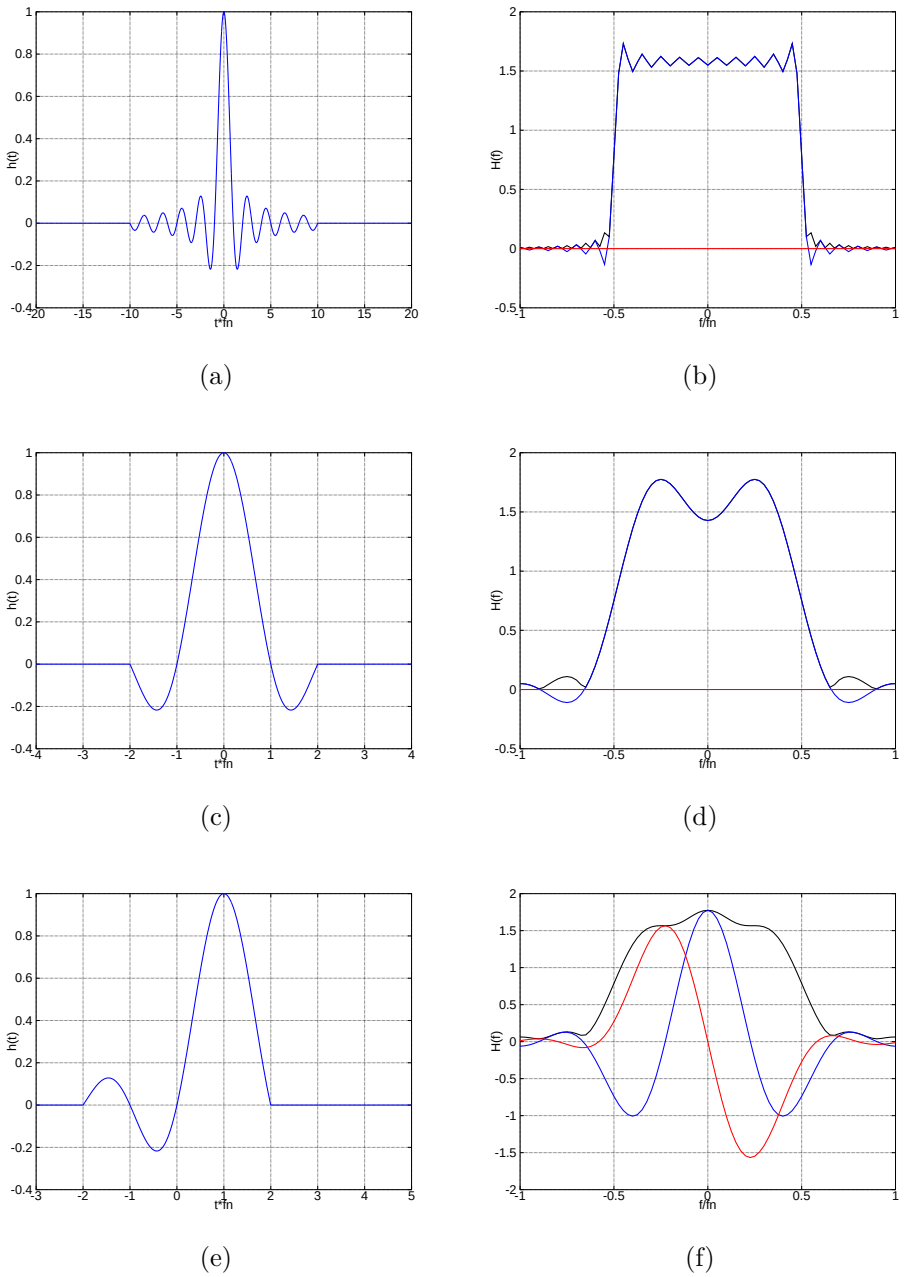
$$h_2(t) = a \times \frac{\sin\left(2\pi f_N (t - t_0)\right)}{2\pi(t - t_0)}, \tag{4.90}$$

wobei $f_N = 1/(2T)$ ist. Er hat letztlich zwei Eigenschaften, die ihn unsinnig machen: Seine Impulsantwort ist unendlich ausgedehnt und sie ist vor allem symmetrisch um den Nullpunkt der Zeit, wodurch das System nicht kausal wird, denn es müsste ja schon reagieren, bevor überhaupt ein Eingangssignal vorliegt. Man muss, um ein kausales System zu konstruieren, die Impulsantwort auch im Zeitbereich noch mit einer Fensterfunktion w belegen, die die Antwort an den unsinnigen Stellen zu Null zwingt (Abb. 4.19):

$$h_3(t) = h_2(t) \times w(t - t_0) \tag{4.91}$$

$$= \begin{cases} a \times \frac{\sin(2\pi f_N (t - t_0))}{2\pi(t - t_0)} \times w(t - t_0) & \text{für } 0 \le t \le 2t_0 \\ 0 & \text{sonst.} \end{cases} \tag{4.92}$$

Natürlich weitet diese Operation im Zeitbereich die Darstellung im Frequenzbereich wieder auf und das Verhalten ist kein idealer Bandpass, sondern einer, dessen Spektrum nun



(a) (b)

(c) (d)

(e) (f)

Abbildung 4.19: Bandbegrenzter (f_N) und mit dem Rechteck gewichteter verzögerungsfreier (a,b,c,d) und $2/f_N$ verzögerter (e,f) Kanal mit (a,c,e) Impulsantwort und (b,d,f) dem Spektrum. Der Filterparameter ist einmal $t_0 = 2/f_n$ und $t_0 = 10/f_n$ (a,b).

wieder unendlich ausgedehnt ist, aber doch recht schnell abfällt. Dieses ist unter praktischen Gesichtspunkten akzeptabel.
Ist der Kanal jedoch bandbegrenzt, entweder durch ein Bandpassverhalten (Tiefpassver-

halten), wie in einem (schlechten) Kabel, oder dadurch, dass Sie auch anderen Nutzern einen Teil des Frequenzbandes zur Verfügung stellen, müssen Sie weitere Maßnahmen ergreifen. Einfache analoge Filter zur Bandbegrenzung führen zur Verzerrung des Signals und würden somit eine Übertragung erschweren. Besser ist es schon im digitalen Bereich eine Bandbegrenzung vorzunehmen. Da man hier im Zeitbereich arbeitet, spricht man eher von Impulsformung, indem man das digitale Signal durch ein digitales Filter schiebt. Die Impulsantworten in Abb. 4.19 zeigen verschwindende Werte an den Punkten $t \times f_N = n/2$ mit ganzen Zahlen n. Es gilt

$$t_n = t_0 \pm \frac{n}{2f_N} \tag{4.93}$$

mit $f_N = 1/(2T)$. Tastet man einen der Impulse an nur diesen Punkten ab, so ergibt sich die Folge

$$h_3(kT) = \begin{cases} 1 & \text{für } k = k_0 \\ 0 & \text{sonst} \end{cases} \tag{4.94}$$

der Impuls erscheint also zwar verzögert aber ansonsten immer verzerrungsfrei, es ist nichts hinzugekommen oder weggenommen worden. Diese Beobachtung steht in engem Zusamenhang mit dem Abtasttheorem, denn eine Bandbegrenzung des Übertragungskanals auf $|f_g| > 1/(2T)$, wie in allen Abbildungen zu beobachten, verzerrt den diskret abgetasteten Impuls in keiner Weise.

4.5.1 Cos-Roll-Off-Filter

Praktisch häufig verwendet und gut zu realisieren ist ein Filter mit einer abfallenden kosinus-förmigen Flanke, im Spektralbereich auch Cos-Roll-Off-Filter genannt. So eine Impulsformung wird z.B. im UMTS (W-CDMA) und im ZigBee eingesetzt (Abb. 4.20). Die Beschreibung des Fensters im Zeitbereich ergibt

$$w(t) = \frac{\cos(\alpha 2\pi t/T)}{1 - (4\alpha t/T)^2} \tag{4.95}$$

und führt zu einer Impulsantwort von

$$h_c = a \times \frac{\sin\left(2\pi f_N (t - t_0)\right)}{2\pi(t - t_0)} \times \frac{\cos(\alpha 2\pi t/T)}{1 - (4\alpha t/T)^2}, \tag{4.96}$$

was dann im Spektralbereich (normiert auf eins) das Verhalten

$$H_c(f) = \begin{cases} 1, & |f| \le \frac{1-\alpha}{2T} \\ \cos^2\left(\frac{\pi T}{2\alpha}\left(|f| - \frac{1-\alpha}{2T}\right)\right), & \frac{1-\alpha}{2T} < |f| \le \frac{1+\alpha}{2T} \\ 0, & \text{sonst} \end{cases} \tag{4.97}$$

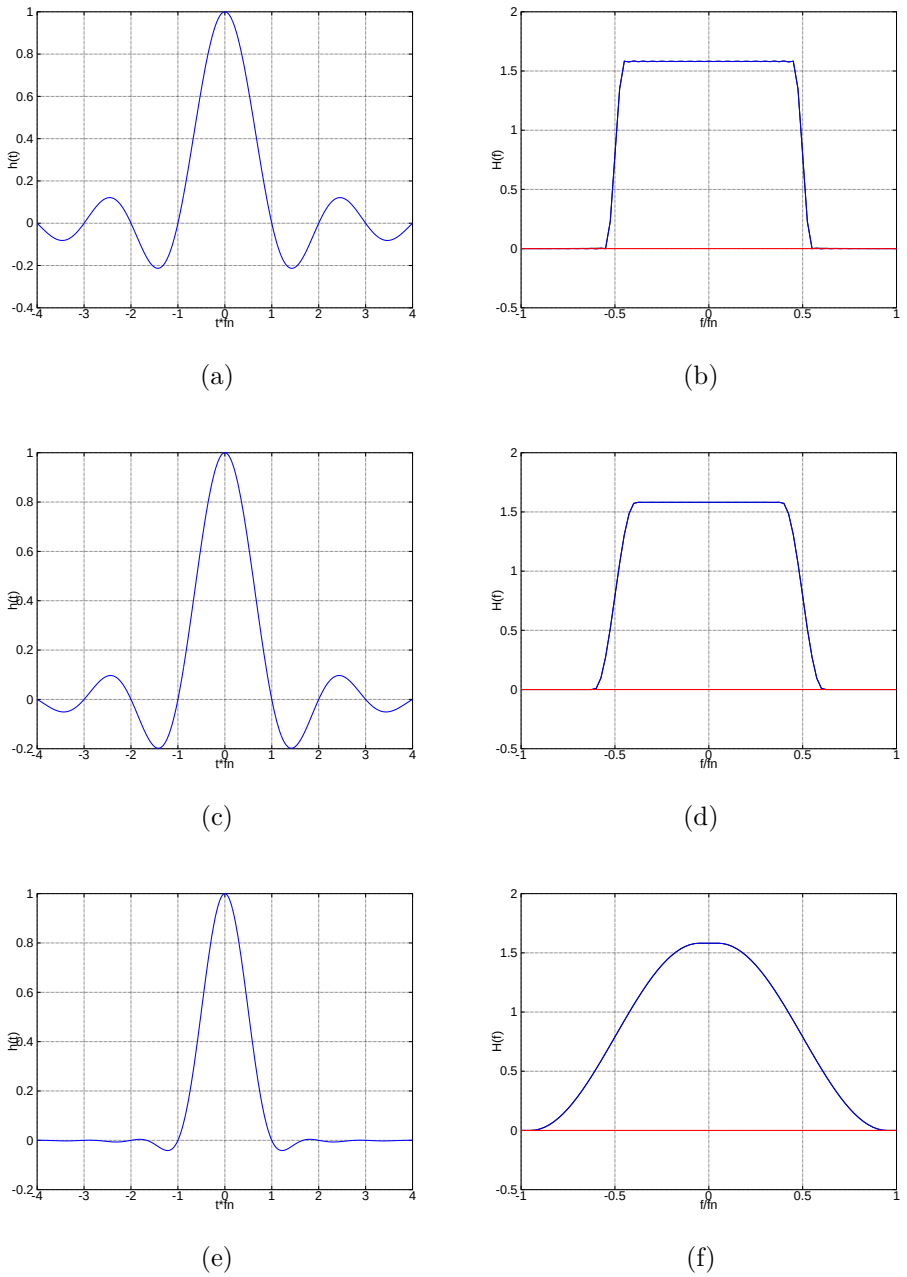


Abbildung 4.20: Cos-Roll-Off Filter im Zeitbereich (a,c,e) und Spektralbereich (b,d,f) mit Parameter $\alpha = 0{,}1$ (a,b), $\alpha = 0{,}22$ (c,d) und $\alpha = 0{,}9$ (e,f).

ergibt. Die benötigte Bandbreite zur Übertragung ist nominal die durch $1/(2T)$ gegeben. Praktisch können Kanäle aber nur im Abstand von $1/(2T) \times (1 + \alpha)$ gesetzt werden. Für das UMTS-System heisst das, dass die nominale Bandbreite von 3,84 MHz im Raster von minimal $3{,}84 \times (1 + 0{,}22)\,\text{MHz} = 4{,}68\,\text{MHz}$, also ca. 5 MHz übertragen wird.

Kapitel 5

Datensicherung und Kodierung

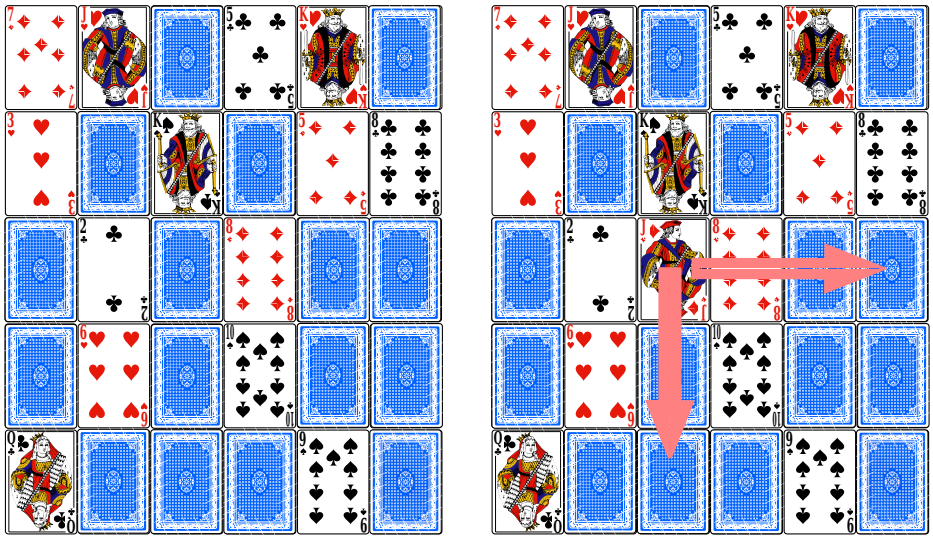
Im Kapitel 4.4.2 wurde dargestellt, wie leicht es zu Übertragungsfehlern kommen kann. Um dennoch eine sichere Datenübertragung zu ermöglichen, bzw. Fehler wenigstens zu erkennen, wurden verschiedene Methoden zur Datensicherung und Kodierung entwickelt. Sie sind ein wichtiger Schlüssel in heutigen mobilen und drahtgebundenen Kommunkationssystemen. Weite Teile dieses Kapitels lehnen sich an [9, 6] an.

5.1 Sicherung der Datenübertragung gegen Bitfehler

Die Basis der Fehlererkennung, in binären Systemen, ist eine begrenzte Liste zulässiger Codewörter, also ein definiertes Code-Buch. Als Beispiel werden die Zahlen von Null bis Drei im binären System als

Zahl	Binärdarstellung
0	00
1	01
2	10
3	11

dargestellt. Wird bei der Übertragung nun ein Bit verfälscht, dann fällt es immer wieder mit einem anderen gültigen Codewort zusammen. Z.B. $00 \rightarrow 01$ macht aus der gesendeten Null eine Eins. Es gibt keine Chance den Fehler zu erkennen. Eine Fehlererkennung wird durch hinzufügen von Redundanz möglich, so dass hier beispielsweise einfach die Codes doppelt übertragen werden:



(a) (b)

Abbildung 5.1: Kartentrick ohne Fehler (a) und mit einem Bitfehler (b). Umgedrehte Karte ist die Eins, offene die Null.

Zahl	Binärdarstellung
0	00 00
1	01 01
2	10 10
3	11 11

Wird nun ein Bit verfälscht 00 00 $\rightarrow$ 00 01, so ergibt sich ein ungültiges Codewort, das verworfen werden kann - und muss. Es ist auch noch nicht möglich, das Original zu rekonstruieren, denn beim Empfang von 00 01 könnte - bei einem Bitfehler - der Ursprung 00 00 oder auch 01 01 gewesen sein.

Die Idee fehlerkorrigierender Codes soll anhand eines Kartentricks nach Abb. 5.1 gezeigt werden. Hier stellt eine verdeckte (man sieht die Rückseite) Karte eine Eins dar, die offene Karte eine Null. Die 4×5 Karten links oben sind die Nutzbits, die rechte Spalte und die untere Zeile ergänzen die kompletten Zeilen oder Spalten zu einer gerade Parität, also immer einer gerade Anzahl verdeckter Karten. Die Karte unten rechts ergibt sich immer aus der entsprechenden Spalte und Zeile gleichermaßen. Ein Bitfehler ist in Abb. 5.1(b) gezeigt. Eine verdrehte Karte zerstört Zeilen- und Spaltenparität gleichzeitig und kann daher gefunden und behoben werden. Insgesamt wurden bei 20 Nutzkarten 10 Redundanzkarten zugefügt.

5.2 Lineare Codes in Matrizendarstellung

5.2.1 Hammingdistanz

Es werden die oben gegebenen Codebücher betrachtet und es wird der Unterschied zwischen einzelnen Codewörtern gemessen, und zwar so, dass jeweils ein Unterschied an einer Stelle der binären Codes den gemessenen Abstand im Sinne der Metrik erhöht, also

$$d_{ij} = \sum_{k=1}^{N} b_{i,k} \oplus b_{j,k} \tag{5.1}$$

wobei $b_{i,k}$ die einzelnen Bits des Codes sind. Die Operation $\oplus$ ist die Modulo-2-Addition mit

$$0 \oplus 0 = 0, \qquad 1 \oplus 0 = 0 \oplus 1 = 1 \qquad 1 \oplus 1 = 0. \tag{5.2}$$

Den Abstand d_{ij} bezeichnet man als die Hammingdistanz. Im einfachen Code ist das für alle gültigen Wörter des Codes

d_{ij}	00	01	10	11
00	0	1	1	2
01	1	0	2	1
10	1	2	0	1
11	2	1	1	0

Der minimale Abstand zwischen zwei Codewörtern

$$d_{min} = \min\left\{d_{ij} \text{ für alle } i \neq j\right\} = 1\ , \tag{5.3}$$

was dazu führt, dass ein Bitfehler schon die Distanz zum nächsten gültigen Code überwindet und ein gültiges Codewort erzeugt. Anders im verdoppelten Code:

d_{ij}	00 00	01 01	10 10	11 11
00 00	0	2	2	4
01 01	2	0	4	2
10 10	2	4	0	2
11 11	4	2	2	0

Hier ist $d_{min} = 2$, die Distanz kann nicht überwunden werden, ein durch einen einzigen Bitfehler verändertes Codewort fällt nicht mit einem anderen gültigen zusammen, der Fehler kann erkannt, aber nicht korrigiert werden.

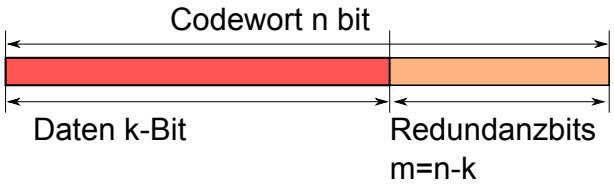


Abbildung 5.2: Codewort eines separierbaren/ systematischen (n,k)-Blockcodes

Nun ist die einfache Verdopplung der Übertragung nicht notwendigerweise die effizienteste Methode, die Datenübertragung zu sichern, denn schon durch Hinzufügen eines Paritätsbits, also immer ein Bit anhängen, so dass die Anzahl der Einsen immer gerade ist, führt mit

Zahl	Binärdarstellung
0	00 0
1	01 1
2	10 1
3	11 0

d_{ij}	00 0	01 1	10 1	11 0
00 0	0	2	2	2
01 1	2	0	2	2
10 1	2	2	0	2
11 0	2	2	2	0

zur gleichen minimalen Hammingdistanz des Codes $d_{min} = 2$, nur dass nun keine Kombination mit der Distanz vier mehr vorliegt. Im letzten Fall ist das gleiche Ergebnis durch Hinzufügen nur eines Bits erreicht worden.
Verallgemeinert ist der Code nun aufgebaut (Abb. 5.2) aus insgesamt n-Bits, wobei k Datenbits und $m = n - k$ Redundanzbits vorliegen. Die Rate des Codes, die auch seine Effizienz beschreibt, ist

$$R = \frac{k}{n}, \qquad \text{mit} \qquad 0 < R < 1. \tag{5.4}$$

5.2.2 Linearer Code

Wenn - wie im obigen Beispiel gezeigt - das Codewort durch Multiplikation mit einer Generatormatrix aus den Datenbits hervorgeht, spricht man von linearen Codes. Es ist dann

$$\vec{b}^T = \vec{x}^T \cdot \overline{\overline{G}}, \tag{5.5}$$

wobei nun $\vec{x}^T = (x_1,x_2,\ldots x_k)$ die Datenbits - geschrieben als übliche Zeilenvektoren - und $\vec{b}^T = (b_1,b_2,\ldots b_n)$ die ebenso geschriebenen zu übertragenden Bits meint. Mit $\oplus$ ist die exklusive-Oder-Funktion auszuführen. Zur Berechnung der Parität ist

$$\overline{\overline{G}} = \left(\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 1 \end{array}\right). \tag{5.6}$$

Die Matrix ist natürlich rechteckig mit k Zeilen und n Spalten und allgemein

$$\overline{\overline{G}} = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ g_{21} & g_{22} & \cdots & g_{2n} \\ \vdots & \vdots & & \vdots \\ g_{k1} & g_{k2} & \cdots & g_{kn} \end{pmatrix}, \tag{5.7}$$

so dass die Multiplikation für das $j-$te Element in $\vec{b}$ als

$$b_j = x_1 g_{1j} \oplus x_2 g_{2j} \oplus \ldots x_k g_{kj} \tag{5.8}$$

auszuführen ist. Es ist klar, dass bei Eingabe des Nullvektors auch immer der Nullvektor wieder herauskommen muss. Dieses ist eine Eigenschaft linearer Codes.
Auch wenn oben ein konkretes Beispiel gewählt wurde, so sind doch zunächst die Elemente von $\overline{\overline{G}}$ beliebig, aber sinnvoll zu wählen. Wichtig ist nur, dass die Zeilen linear unabhängig von einander sind, so dass der Rang k der Matrix erhalten bleibt. Man kann die Matrix immer so umformen, dass sie in einen Teil, der die Einheitsmatrix enthält, und einen, der eine Permutationsmatrix erhält, aufgeteilt werden kann:

$$\overline{\overline{G}} = \left(\overline{\overline{I_k}} | \overline{\overline{P}}\right) = \left(\begin{array}{ccccc|cccc} 1 & 0 & 0 & \ldots & 0 & p_{11} & p_{12} & \cdots & p_{1,n-k} \\ 0 & 1 & 0 & \ldots & 0 & p_{21} & p_{22} & \cdots & p_{2,n-k} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & p_{k1} & p_{k2} & \cdots & p_{k,n-k} \end{array}\right). \tag{5.9}$$

Mit dem ersten Teil der Matrix wird einfach nur das Datenwort durchgereicht, während im zweiten Teil die Parität berechnet wird. Diese so generierten Codes nennt man systematische Codes.
Die Fähigkeit eines Codes, Fehler zu erkennen bzw. zu beheben, wird - wie oben besprochen - durch den minimalen Abstand zweier gültiger Codewörter d_{min} beschrieben. Nach (Gl. 5.3) musste jede mögliche Kombination untersucht werden. Bei linearen Codes, bei denen ja der Nullvektor ein gültiges Wort ist, genügt es den minimalen Abstand zu diesem Wort zu finden, also

$$d_{min} = \min \left\{ \sum_{k=1}^{N} b_{i,k} \right\} \text{ für alle } i \neq 0\,. \tag{5.10}$$

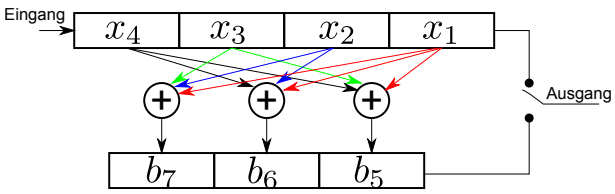


Abbildung 5.3: Schieberegisterdarstellung eines separierbaren (7,4)-Blockcodes

5.2.3 Kodierung

Für einen separierbaren (7,4)-Hamming Code mit vier Datenbits und drei Kontrollbits ergibt sich eine Generatormatrix

$$\overline{\overline{G}} = \left(\begin{array}{cccc|ccc} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{array}\right). \tag{5.11}$$

Die einzelnen Paritätsbits gehen somit durch

$$b_5 = x_1 \oplus x_3 \oplus x_4 \tag{5.12}$$

$$b_6 = x_1 \oplus x_2 \oplus x_4 \tag{5.13}$$

$$b_7 = x_1 \oplus x_2 \oplus x_3 \tag{5.14}$$

aus den Datenbits hervor. Dieses ließe sich rechentechnisch durch ein Schieberegister wie folgt realisieren (Abb. 5.3):

1. Die vier (k) Datenbits werden in ein Schieberegister der Länge, k geladen
2. durch exklusive Oder-Verknüpfung werden die $n - k$ Redundanzbits bestimmt,
3. die Datenbits werden übertragen,
4. die Redundanzbits werden übertragen.

Die Generatormatrix des Kartentricks nach Abb. 5.1 für 4×4 Karten und entsprechenden Paritäten ist

$$\overline{\overline{G}}_{karten} = \begin{pmatrix}
1 & & & & & & & & & & & & 1 & & & 1 & & & 1 \\
 & 1 & & & & & & & & & & & 1 & & & & 1 & & 1 \\
 & & 1 & & & & & & & & & & 1 & & & & & 1 & 1 \\
 & & & 1 & & & & & & & & & 1 & & & & & & 1 & 1 \\
 & & & & 1 & & & & & & & & & 1 & & 1 & & & & 1 \\
 & & & & & 1 & & & & & & & & 1 & & & 1 & & & 1 \\
 & & & & & & 1 & & & & & & & 1 & & & & 1 & & 1 \\
 & & & & & & & 1 & & & & & & 1 & & & & & 1 & 1 \\
 & & & & & & & & 1 & & & & & & 1 & 1 & & & & 1 \\
 & & & & & & & & & 1 & & & & & 1 & & 1 & & & 1 \\
 & & & & & & & & & & 1 & & & & 1 & & & 1 & & 1 \\
 & & & & & & & & & & & 1 & & & 1 & & & & 1 & 1 \\
 & & & & & & & & & & & & 1 & & & 1 & 1 & & & 1 \\
 & & & & & & & & & & & & & 1 & & 1 & & 1 & & 1 \\
 & & & & & & & & & & & & & & 1 & 1 & & & 1 & 1 \\
 & & & & & & & & & & & & & & & 1 & 1 & & & 1
\end{pmatrix}. \tag{5.15}$$

Hier wird davon ausgegangen, dass erst die Daten, dann die Parätsspalte und erst am Ende die Zeile, die die Paritäten der Spalten enthält, gesendet werden.

5.2.4 Dekodierung und Prüfmatrix

Die Kodierung ist nur die Hälfte der Arbeit, nämlich die, die im Sender der Information gemacht wird. Im Empfänger muss die Nachricht auf Fehlerfreiheit geprüft werden und ggf. (wenn möglich) korrigiert werden. Zu diesem Zweck wird der duale Code generiert. Dieser ist ebenfalls ein linearer Code, aber mit den Parametern (n,$n-k$), also der Länge n mit nun nur $n-k$ Datenbits. Der duale Code ist nun so zu konstruieren, dass für jedes Codewort des originalen Codes

$$\vec{b}^T \cdot \overline{\overline{H}}^T = 0 \tag{5.16}$$

mit $\overline{\overline{H}}^T$ der Transponierten der Generatormatrix des dualen Codes folgt. $\overline{\overline{H}}$ ist auch die Prüfmatrix, denn die Prüfung besteht darin, dass die Multiplikation eines gültigen Codeworts mit ihr Null ergeben muss. Da (Gl. 5.16) für alle $\vec{b}$ gelten muss, muss auch die

Matrixgleichung

$$\overline{\overline{G}} \cdot \overline{\overline{H}}^T = 0 \tag{5.17}$$

gelten. Hiermit ist eine Bildungsvorschrift für $\overline{\overline{H}}$ gefunden, die sich zu

$$\overline{\overline{H}} = \left(\overline{\overline{P}}^T | \overline{\overline{I}}_{n-k} \right) \tag{5.18}$$

ergibt. Die Multiplikation mit dem Generator nach (Gl. 5.9) ergibt

$$\overline{\overline{G}} \cdot \overline{\overline{H}}^T = \left(\overline{\overline{I}}_k | \overline{\overline{P}} \right) \cdot \left(\frac{\overline{\overline{P}}}{\overline{\overline{I}}_{n-k}} \right) \tag{5.19}$$

$$= \overline{\overline{I}}_k \cdot \overline{\overline{P}} \oplus \overline{\overline{P}} \cdot \overline{\overline{I}}_{n-k} = 0, \tag{5.20}$$

wobei auffällt, dass bei herkömmlicher Addition hier ein Minuszeichen zwischen den Teilen des Ausdrucks stehen müsste, um die Null rechts erzwingen zu können, und damit dieses auch in der Prüfmatrix auftauchen müsste. Es kann entfallen, da die Modulo-2 Addition dieses unnötig macht.
Für die bislang betrachteten Codes ergeben sich die Prüfmatrizen

$$\overline{\overline{H}}_{Par} = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \tag{5.21}$$

für die einfache gerade Parität. Jedes hierfür gültige Codewort besteht aus drei Nullen oder zwei Einsen und einer Null. Nachrechnen zeigt, dass die Multiplikation mit $\overline{\overline{H}}_{Par}$ immer Null ergibt. Für den (7,4)-Code mit (Gl. 5.11) ergibt sich die Prüfmatrix

$$\overline{\overline{H}}_{7,4} = \left(\begin{array}{cccc|ccc} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{array} \right). \tag{5.22}$$

In einem Dekodierer wird nun ein empfangenes Wort $\vec{b}^*$ mit

$$\vec{s}^T = \vec{b}^{*T} \cdot \overline{\overline{H}}^T = \vec{b}^{*T} \cdot \left(\frac{\overline{\overline{P}}}{\overline{\overline{I}}_{n-k}} \right) \tag{5.23}$$

$$= \left[(b_1^*, b_2^*, \ldots, b_k^*) \cdot \overline{\overline{P}} \oplus (b_{k+1}^*, \ldots, b_n^*) \cdot \overline{\overline{I}}_{n-k} \right] \tag{5.24}$$

einer Prüfung unterzogen. Der Vektor $\vec{s}$ wird das Syndrom genannt und ist Null, wenn ein gültiges Codewort empfangen wurde. Ein einzelnes Element des Syndroms ist

$$s_j = b_1^* p_{1j} \oplus b_2^* p_{2j} \oplus \ldots \oplus b_k^* p_{kj} \oplus b_{k+j}^* p_{k+j\,j}. \tag{5.25}$$

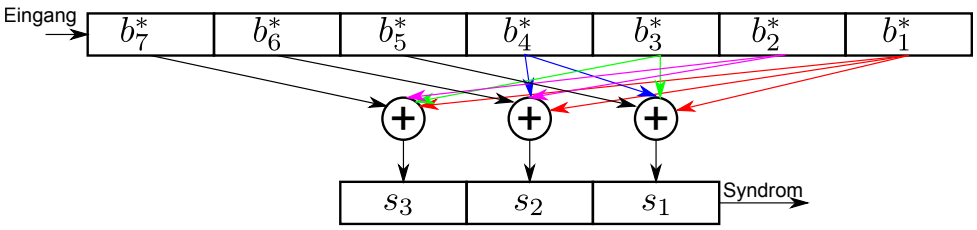


Abbildung 5.4: Schieberegisterdarstellung zur Berechnung des Syndroms eines separierbaren (7,4)-Blockcodes

Tabelle 5.1: Zuordnung von Syndrom und Fehler in einem (7,4)-Blockcode (LSB steht links: $(e_1, \ldots, e_7)$.

Fehlervektor $\vec{e}^T$	Syndrom $\vec{s}^T$	Fehlervektor $\vec{e}^T$	Syndrom $\vec{s}^T$
0000 000	000	0001 000	110
1000 000	111	0000 100	100
0100 000	011	0000 010	010
0010 000	101	0000 001	001

Die hierzu gehörige Schieberegisterrealisierung ist in Abb. 5.4 gezeigt. Es ist ersichtlich, dass die Datenbits in mehrere Verknüpfungen eingehen, während die Redundanzbits nur jeweils einfach verwendet werden.
Stellt man den empfangenen Datenstrom nun als

$$\vec{b}^* = \vec{b} + \vec{e} \tag{5.26}$$

dar, wobei $\vec{b}$ das gesendete, also gültige Codewort ist und $\vec{e}$ einen Fehlerterm beschreibt, so ergibt sich für das Syndrom

$$\vec{s}^T = \left(\vec{b}^T + \vec{e}^T\right) \cdot \overline{\overline{H}}^T = \vec{b}^T \cdot \overline{\overline{H}}^T + \vec{e}^T \cdot \overline{\overline{H}}^T = \vec{e}^T \cdot \overline{\overline{H}}^T. \tag{5.27}$$

Wenn nur ein Fehler an der Stelle j auftritt, so kann man jedem beliebigen Fehlervektor ein Syndrom zuordnen (Tab. 5.1) und somit die Fehlerstelle finden und korrigieren.
Für den einfachen Paritätscode besteht das Syndrom (es ist ja nur eine Stelle) bei einem Bitfehler immer aus einer Eins. Man kann also zwar einen Fehler feststellen, da alle Syndrome gleich sind, aber nicht korrigieren. Wird Fehlerkorrektur verwendet, so ist Fehlererkennung darüber hinaus nicht mehr möglich, man muss sich also entscheiden.

5.2.5 Leistungsfähigkeit eines Codes

Die Beschreibung der Leistungsfähigkeit eines Codes beschränkte sich bislang auf die Angabe der minimalen Hammingdistanz (Gl. 5.10) und der Coderate (Gl. 5.4). Damit ist noch keine Aussage getroffen, was der Code leisten kann.
Mit einem Code, der die minimale Hammingdistanz von d_{min} aufweist können

- $d_{min} - 1$ Fehler erkannt werden,
- $< d_{min}/2$ Fehler korrigiert werden,

so dass bei dem Paritätscode mit $d_{min} = 2$ nur ein Fehler erkannt und keiner korrigiert werden kann, während für den (7,4)-Blockcode mit $d_{min} = 3$ entweder zwei Fehler erkannt oder einer korrigiert werden können.

Übungsaufgabe Weisen Sie die o.g. Aussage bzgl. des (7,4)-Blockcodes nach!

5.2.5.1 Kodiergewinn und Dekodierung

Die Leistung eines Codes misst sich natürlich wesentlich darin, um wieviel er die Situation verbessert. Dieses kann gemessen werden als der Kodiergewinn, also um welchen Betrag kann die Übertragungsleistung bei gleicher Bitfehlerrate reduziert werden (Siehe dazu auch Bild 4.17). Diese Reduktion bezeichnet man als Kodiergewinn. Da dieser massiv von der Art der Dekodierung abhängt, soll diese nochmals beleuchtet werden.
In einem verrauschten Kanal nach Abschnitt 4.4.1 stehen im Empfänger analog verschobene Signalwerte zur Verfügung, die in der Regel im Empfänger mittels Analog Digital Converter (ADC) abgetastet werden. Nehmen wir an, dass die Folge $\vec{b} = (1; 0; 0; 0; 1; 1; 1)$ bipolar als $\vec{b}_b = (1; -1; -1; -1; 1; 1; 1)$ gesendet wurde und nach dem ADC nun die Werte $\vec{b}_a = (1{,}2; -1{,}2; -0{,}5; 0{,}1; 1{,}1; 0{,}3; 0{,}9)$ zur Verfügung stehen.

Hard-Decision Dekodierung Nun kann in einem Entscheider sofort die Entscheidung für die Bit-Ebene gefällt werden und an dessen Ende liegt die Folge $\vec{b}^* = (1001111)$ vor, man erkennt sofort den einen Bitfehler an der vierten Stelle. Es erfolgt die Berechnung des Syndroms mittels Multiplikation an die Prüfmatrix und es ergibt sich dieser zu $\vec{s} = (110)$ und aus der Syndromtabelle Tab. 5.1 kann der Fehlervektor zu $\vec{e} = (0001000)$ abgelesen werden. Eine Addition mit der empfangenen und entschiedenen Bitfolge korrigiert den Bitfehler und ergibt das richtige Ergebnis. Die Bitfehlerrate und damit der Gewinn dieses Verfahrens lässt sich näherungsweise berechnen.

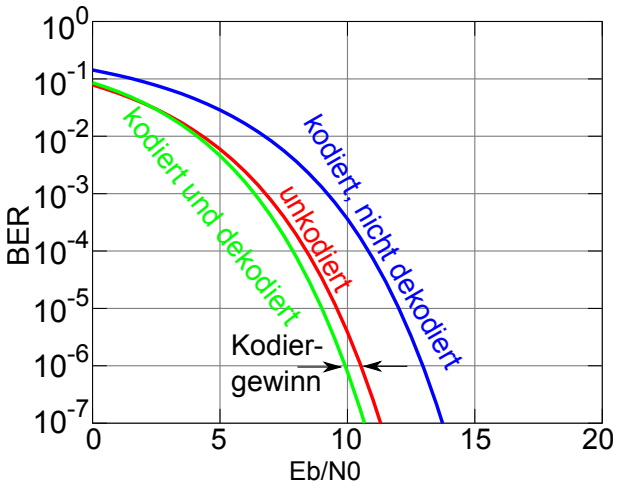


Abbildung 5.5: Bitfehlerrate in Abhängigkeit der Bitenergie bei einer Hamming (7,4) Kodierung nach Hard-Decision.

Nach Abschnitt. 4.4.2 und insbesondere (Gl. 4.79) kann der Bitfehler einer BPSK bzw. äquivalent einer Übertragung mit QPSK angegeben werden. Diese Kurve ist in Abb. 5.5 dargestellt. Nach Kodierung werden die vier Nutzbits mit den drei Redundanzbits in Blöcken zu siebenen verschickt, die für ein einzelnes Bit zur Verfügung stehende Energie reduziert sich um die Kodierungsrate R und ergibt sich so zu

$$P_c = \frac{1}{2}\mathrm{erfc}\sqrt{R \times \frac{E_b}{N_0}}. \tag{5.28}$$

Dieses bedeutet natürlich eine Verschlechterung der Übertragung und die Bitfehlerrate steigt zunächst. Jedoch können einige der Bitfehler, nämlich y, korrigiert werden. Die Wahrscheinlichkeit, dass ein einziger Bitfehler in einem dieser Blöcke vorliegt, ist hier

$$P_1 = P_c \times (1 - P_c)^6 \times 7, \tag{5.29}$$

was aus der Kombinatorik folgt. Die Wahrscheinlichkeit für keinen Bitfehler in diesem Block ist

$$P_0 = (1 - P_c)^7 \tag{5.30}$$

und die Wahrscheinlichkeit für zwei Bitfehler ausgeschrieben

$$P_2 = P_c^2 \times (1 - P_c)^5 \times (6 + 5 + 4 + 3 + 2 + 1), \tag{5.31}$$

Tabelle 5.2: Codebuch des (7,4)-Blockcode (LSB steht links)

Datenvektor $\vec{d}^T$	Code $\vec{b}^T$	Datenvektor $\vec{d}^T$	Code $\vec{b}^T$
0000	0000 000	1000	1000 111
0001	0001 110	1001	1001 011
0010	0010 101	1010	1010 010
0011	0011 011	1011	1011 100
0100	0100 011	1100	1100 100
0101	0101 101	1101	1101 010
0110	0110 110	1110	1110 001
0111	0111 000	1111	1111 111

oder allgemein

$$P_i = \binom{n}{i} \times P_c^i \times (1 - P_c)^{n-i}. \tag{5.32}$$

Da Einbit-Fehler korrigiert werden können, ist die Summenwahrscheinlichkeit für alle anderen Mehrbit-Fehler interessant und die ist

$$P_{s,y+1} = \sum_{i=y+1}^{n} \binom{n}{i} \times P_c^i \times (1 - P_c)^{n-i}. \tag{5.33}$$

Die Rate für Fehler ergibt sich nun aus der Anzahl der Fehler bezogen auf die Blockgröße zu

$$P_{cd} \approx \frac{1}{n} \sum_{i=y+1}^{n} i \binom{n}{i} \times P_c^i \times (1 - P_c)^{n-i} \tag{5.34}$$

und ist ebenso in Abb. 5.5 dargestellt. Nicht berücksichtigt worden ist, dass der Dekodierer noch mehr Fehler hinzufügen könnte, als im Bitstrom eigentlich vorhanden sind.
Für eine Zielfehlerrate von 10^{-6} ergibt sich in dieser Situation also ein Kodierungsgewinn von ca. 0,6 dB.

Soft-Decision Dekodierung Eine deutliche Verbesserung lässt sich erzielen, wenn die analog vorliegenden Daten während der Dekodierung tatsächlich verwendet werden. Hierzu stellt man die Liste der erlaubten Codewörter zusammen (Tab. 5.2). Im nächsten Schritt wird der Abstand des empfangenen Wortes zu jedem der gültigen Codewörter bestimmt. Die dazu verwendete Metrik kann gewählt werden und bestimmt natürlich auch

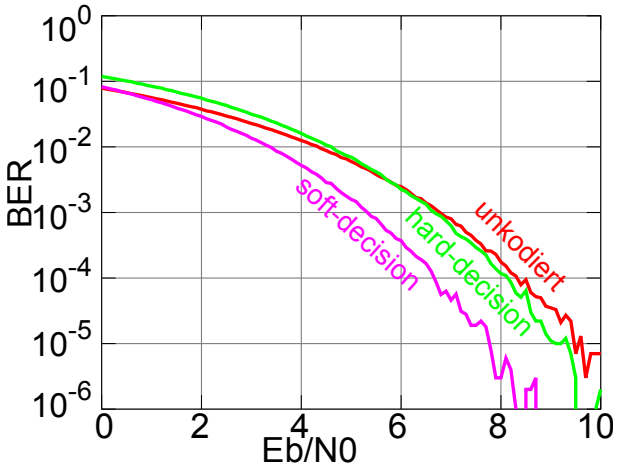


Abbildung 5.6: Bitfehlerrate in Abhängigkeit der Bitenergie im Vergleich zwischen keiner Kodierung und Hard-Decision sowie Soft-Decision Dekodierung mit einem Hamming (7,4) Code.

die Leistungsfähigkeit des Codes. Hier kann der einfache Abstand nach

$$a_j = \left(\sum_{i=1}^{n} (b_{bj,i} - b_{a,i})^p \right)^{1/p} \tag{5.35}$$

von der analogen Folge $\vec{b}_a$ verwendet werden. Es ist $\vec{b}_{bj}$ das j-te Codewort nach Tab. 5.2 in der bipolaren Form. Weiterhin werden für $p = 1$ einfach die Differenzen addiert, und für $p = 2$ wird der euklidische Abstand angenommen. Alternativ kann auch die Korrelation nach

$$a_j = \sum_{i=1}^{n} (b_{bj,i} \times b_{a,i}) \tag{5.36}$$

verwendet werden. In jedem Fall wird das Codewort als das wahrscheinlichste ausgewählt, das den geringsten Abstand zu bzw. die höchste Korrelation mit der empfangenen Folge aufweist. Simulationen in Abb. 5.6 zeigen hier eine deutliche Verbesserung und einen deutlich größeren Kodierungsgewinn als mit der Hard-Decision Dekodierung.
Beispielhaft ist in Tab. 5.3 empfange Amplitudenfolge mit Hamming (7,4)-Kodierung soft dekodiert worden.

Tabelle 5.3: Soft-Decision Dekodierung: Beipiel

$\vec{b}_a$	1,2	-1,2	-0,5	0,1	1,1	0,3	0,9	
0	-1	-1	-1	-1	-1	-1	-1	-1,9
1	-1	-1	-1	1	1	1	-1	1,1
2	-1	-1	1	-1	1	-1	1	1,1
3	-1	-1	1	1	-1	1	1	-0,3
4	-1	1	-1	-1	-1	1	1	-1,9
5	-1	1	-1	1	1	-1	1	-0,1
6	-1	1	1	-1	1	1	-1	-2,5
7	-1	1	1	1	-1	-1	-1	-5,1
8	1	-1	-1	-1	1	1	1	5,1
9	1	-1	-1	1	-1	1	1	3,1
A	1	-1	1	-1	-1	1	-1	0,1
B	1	-1	1	1	1	-1	-1	1,9
C	1	1	-1	-1	1	-1	-1	0,3
D	1	1	-1	1	-1	1	-1	-1,1
E	1	1	1	-1	-1	-1	1	-1,1
F	1	1	1	1	1	1	1	1,9

5.2.6 Konkret verwendete lineare Codes

5.2.6.1 Hamming-Codes

Die binären Hamming-Codes sind $(2^m - 1, 2^m - 1 - m)$ Block-Codes mit der Eigenschaft, dass die minimale Hamming-Distanz $d_{min} = 3$ beträgt, also zwei Bitfehler erkannt, bzw. einer korrigiert werden kann. Hamming-Codes sind perfekte (optimale) Codes.
Weiterhin haben die Hamming-Codes die in Tab. 5.4 dargestellten Eigenschaften. Die Coderate wird natürlich immer höher, der Schutzmechanismus immer schlechter, bei $m = 11$ können ja nur noch zwei Bitfehler in 2047 Bit (also eine Fehlerrate von ca. 0,1 %) erkannt werden, während mit dem (7,4) 28 % oder mit (3,1) sogar 66 % Bitfehlerrate erkannt werden könnten.
Durch das Anhängen eines einfachen Paritätsbits erhöht sich die minimale Distanz der Codes um eins auf $d^T_{min} = 4$ und somit können nun drei Bitfehler erkannt werden. Dieses kann man allerdings nur einmal machen, weiteres Anhängen eines Paritätsbits ist unsinnig.

Tabelle 5.4: Eigenschaften von Hamming-Codes

m	(n,k)	R_c	m	(n,k)	R_c
2	(3,1)	0,3333	7	(127,120)	0,9449
3	(7,4)	0,5714	8	(255,247)	0,9686
4	(15,11)	0,7333	9	(511,502)	0,9824
5	(31,26)	0,8387	10	(1023,1013)	0,9902
6	(63,57)	0,9048	11	(2047,2036)	0,9946

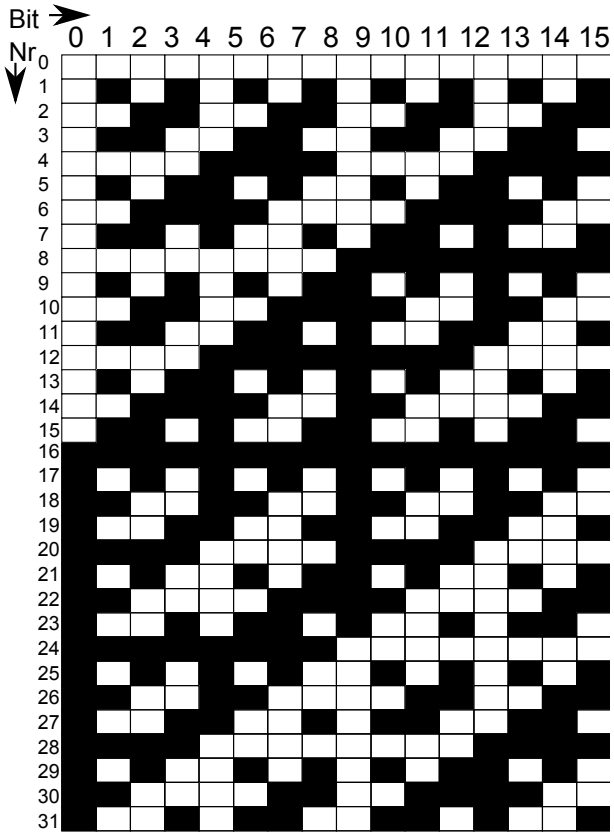


Abbildung 5.7: Grafische Darstellung des [16,5,8]-Hadamard-Codes.

5.2.6.2 Hadamard-Code

Hier werden die Codewörter aus den Zeilen der Hadamard-Matrizen $\overline{\overline{M}}_n$ generiert. Diese entstehen (in bipolarer Darstellung) aus der Bildungsvorschrift

$$\overline{\overline{M}}_{2n} = \begin{pmatrix} \overline{\overline{M}}_n & \overline{\overline{M}}_n \\ \overline{\overline{M}}_n & -\overline{\overline{M}}_n \end{pmatrix} \tag{5.37}$$

und dem Beginn $\overline{\overline{M}}_1 = 1$. Somit ist

$$M_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \qquad M_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix} \tag{5.38}$$

$$M_8 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix} \tag{5.39}$$

für die Matrizen bis zur Ordnung $n = 8$. In dieser Darstellung sind die einzelnen Zeilen orthogonal zueinander, das heißt, dass eine Zeile, multipliziert mit einer anderen und dann über die Einzelsummen aufsummiert, Null ergibt:

$$0 = \sum_{i=1}^{n} M_{n,ji} \cdot M_{n,ki} \qquad \text{für } j \neq k. \tag{5.40}$$

Durch die Konstruktion der Codewörter

$$\overline{\overline{C}}_{2n} = \begin{pmatrix} \overline{\overline{M}}_n \\ -\overline{\overline{M}}_n \end{pmatrix} \tag{5.41}$$

ergeben sich nun

$$C_8 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ -1 & -1 & 1 & 1 \\ -1 & 1 & 1 & -1 \end{pmatrix} \tag{5.42}$$

$$\tag{5.43}$$

$2n = 8$ Codewörter mit einem maximalen Abstand, nämlich $n/2 = 4$ zueinander. Die Angabe des Codes erfolgt dann als $[2^{k-1},k,2^{k-2}]$-Code und hat dementsprechend die

Tabelle 5.5: Eigenschaften von Hadamard-Codes. Die korrigierbare Bitfehlerrate ist immer 25 %.

k	$[2^{k-1},k,2^{k-2}]$	R_c	k	$[2^{k-1},k,2^{k-2}]$	R_c
2	[2,2,1]	1	7	[64,7,32]	0,1094
3	[4,3,2]	0,75	8	[128,8,64]	0,0625
4	[8,4,4]	0,5	9	[256,9,128]	0,0352
5	[16,5,8]	0,3125	10	[512,10,256]	0,0195
6	[32,6,16]	0,1875	11	[1024,11,512]	0,0107

- Blocklänge 2^{k-1} (hier: 4)
- Anzahl übertragener Informationsbits k (hier: 3)
- Minimaler Abstand 2^{k-2} und damit die Fähigkeit $2^{k-2}-1$ Bitfehler zu erkennen und 2^{k-3} zu korrigieren (hier: minimaler Abstand 2 und damit kann kein Fehler korrigiert werden, da die Anzahl korrigierbarer Fehler echt kleiner als die Hälfte des Abstandes sein muss)
- Coderate $R_c = k/2^{k-1}$, (hier 0,75).
- Erkennbare Bitfehlerrate ist $(2^{k-2}-1)/2^{k-1} \approx 0{,}5$
- Korrigierbare Bitfehlerrate ist $2^{k-3}/2^{k-1} = 0{,}25$ allgemein, im speziellen Fall mit $k = 3$ ist der Abstand zu gering für eine Korrektur.

Man sieht (Tab. 5.5), dass die Coderate sehr schnell rapide sinkt, aber natürlich die Korrektureigenschaften des Codes erhalten bleiben. Zur Illustration ist in Abb. 5.7 das Code-Buch für [16,5,8]-dargestellt. Den schwarzen und weißen Feldern können Null, Eins, oder in bipolarer Darstellung auch -1 und 1 zugeordnet werden.

5.2.6.3 Low-Density Parity-Check Codes

Eine vierzig Jahre alte, danach mehrfach wiederentdeckte und in den 1990er dann umgesetzte Variante sind die Low Density Parity Check Codes (LDPC) oder auch Gallager-Codes [37]. Wichtiges Charakteristikum ist, dass die Prüfmatrizen dünn besetzt sind, was recht einfache und vor allem schnelle Dekodierer ermöglicht. Die Konstruktion der Codes ist leicht und erfolgt mit Pseudo-Zufalls-Prozessen. Die Codes eignen sich besonders für große Blöcke. Wie auch andere Codes, werden diese beschrieben durch die Blocklänge n

der Übertragung und die Anzahl k der Quellenbits, es liegen damit $m = (n - k)$ Redundanzbits vor. Zur Konstruktion der Prüfmatrix, und mit der fängt man an, wird zunächst das Gewicht t (das sind die Anzahl der Einsen) einer Spalte festgelegt. Es kann bspw. $w_c = 3$ gewählt werden. Die Konstruktion nach Gallager erfolgt dann so, dass die Matrix in w_c Untermatrizen der Zeilenzahl n/w_c aufgeteilt wird. Bei $n = 16$, $k = 12, w_c = 3$ sind das also drei dreizeilige Matrizen. In der oberen Reihe werden dann immer w_r (hier wird $w_r = 4$ gewählt) folgende Einsen eingetragen. In den folgenden Reihen der Untermatrizen tauchen dann nur noch wahllos gewählte Permutationen der oberen Reihen auf. So eine Kostruktion fürhrt leicht auf Matrizen, bei denen zwei Zeilen in mehr als einer Spalte gleichzeitig eine Eins haben. das ist zu vermeiden, da es in einem Tanner-Graph zu kurzen Zyklen kommt. Eine verfeinerte Konstruktion der weiteren Bereiche ist, zunächst als Blöcke die Einheitsmatrizen ein zu tragen und in dem letzten Bereich verschobene Einheitsmatrizen zu verwenden, also solche bei denen die Einsen zyklisch verschoben auf einer Nebendiagonalen liegen. Ein Beispiel ist in Gl. 5.44 gezeigt.

$$\overline{\overline{H}} = \left(\begin{array}{cccccccccccccccc}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
\hline
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
\hline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0
\end{array}\right). \tag{5.44}$$

Die Konstruktion mit zyklisch verschobenen Einheitsmatrizen ist der Konstruktion der Codes in 5G NR angelehnt [38], aber dort natürlich für wesentlich größere Blocklängen und damit Matrizen beschrieben. Der durch Gl. 5.44 beschriebene Code ist natürlich unsystematisch, das heißt in der hieraus zu errechnenden Generatormatrix taucht das eigentliche Datenwort nicht mehr auf. Hinzu kommt noch, dass zumindest für dieses Beispiel der Rang der Matrix kleiner als die erforderliche zwölf ist, er ist nur zehn. Damit ist keine einfache Berechnung der Generatormatrix, die ja eine Invertierung eines Matrixteils benötigt, möglich. Ohne es mathematisch bewiesen zu haben, scheint der Fall der nicht

invertierbaren Matrizen bei der Konstruktion nach Gallager für geringe Blocklängen und geringe Coderaten (die oben hat die Rate 3/4) recht wahrscheinlich zu sein.
Dennoch hat die Matrix die Eigenschaft, dass

$$n \times w_c = m \times w_r \tag{5.45}$$

ist. Diese Codes bezeichnet man als reguläre LDPC. Wenn diese Bedingung nicht erfüllt ist, dann spricht man von irregulären LDPC. Diese Klasse, auch wenn sie hier nicht vertieft behandelt wird, bringt beste Kodiereigenschaften mit sich.
Eine andere, genauso einfache, Konstruktionsvorschrift nach MacKay-Neal [39] verfolgt einen etwas offeneren Weg:

1. Starte mit einer (m,n) Matrix $\overline{\overline{H}}$, gefüllt mit nur Nullen.
2. Fülle nacheinander jede Spalte mit w_c Einsen in beliebigen Reihen, die noch nicht w_r Einsen enthalten.
3. (Optional) überprüfe den Überlapp zwischen allen Spaltenpaaren und verändere die Matrix so, dass nur noch Überlapp Eins übrig bleibt.

Für einen nach dieser Vorschrift generierten Code mit $n = 20$, einer Rate von 3/4 und 3 bzw. 4 Einsen in Spalten und Zeilen folgt mit Rang 15

$$\overline{\overline{H}} = \begin{pmatrix}
0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}. \tag{5.46}$$

Diese Matrix ist regulär, aber es gibt einige Zeilen, die mehr als eine überlappende eins haben. Dieses kann man optimieren. Auch diese Methode führt nicht immer zu angenehmen und leicht invertierbaren Matrizen. Erst nach einigen Versuchen lassen sich hier gute Prüfmatrizen finden. Mathematische Beschreibungen, warum das so ist, muss ich selber noch suchen, für diese Vorlesung soll es erstmal reichen.
Da der Rang der Matrix genügend hoch ist, nämlich der Zeilenzahl entspricht, kann man einen systematische Code daraus generieren, wozu die Gauss'sche Eliminiationstechnik verwendet wurde:

$$\overline{\overline{H}}_s = \begin{pmatrix}
1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}. \tag{5.47}$$

Der linke Teil der Matrix ist der Paritätsanteil, mit dem nach Gl. 5.9 die Generatormatrix aufgebaut werden kann. Vorteile der dünn besetzten Matrizen sind vor allem die schnelle Kodierung und Dekodierung, die Matrix-Inversionen sind ja nur einmal zu machen und können dann in den Systemen fest hinterlegt werden. Zur Dekodierung können graphenbasierte iterative Verfahren verwendet werden, die in dieser Vorlesung in diesem Jahr aber nicht behandelt werden. Das ist etwas für die Zukunft!

5.2.6.4 Tanner Grafen und Dekodierung von LDPC

Die Dekodierung kann über einen iterativen Algorithmus mit Hilfe der Umwandlung der Matrizen in einen Grafen (Factor-Graf oder spezieller Tanner-Graph) erfolgen. In dem

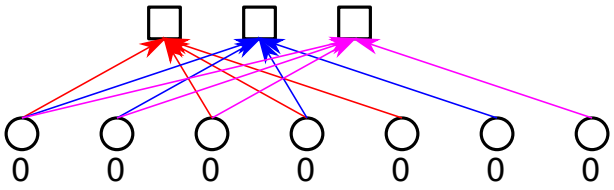


Abbildung 5.8: Tannergraph des Hamming (7,4)-Codes

Grafen sind runde Knoten Variablen-Knoten und Quadratische Check-Knoten. Als Beschreibung fallen wir hier zurück auf den Hamming (7,4)-Code. das ist natürlich kein LDPC, aber ein Blockcode mit einer Länge, die man zeichnerisch und von Hand beherrschen kann. Die Matrix ist in Gl. 5.48 wiederholt

$$\overline{\overline{H}} = \begin{pmatrix} 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{5.48}$$

Hier kommen vom Kanal sieben Variablen an, und es gibt drei Paritätschecks. Jede Eins in einer Zeile zeigt, wie jeder Paritätscheck mit einer Variablen verbinden ist (Abb. 5.8). In den Checkknoten werden im einfachsten Fall die Xor-Operation aller gelieferten Bit-werte ausgeführt.
So ein Graf schafft nun Übersicht, gibt aber auch die Möglichkeit sich vor zu stellen, wie Nachrichten von einem Knoten zum anderen Knoten wandern würden. Offensichtlich erfüllt der Nullvektor alle Checks, da alle Xor-Operationen die erforderliche Null liefern. Wenn nun ein Fehler auftritt, also bspw. das Wort {0,0,0,0,0,1,0} empfangen wurde, wird der mittlere Check eine Eins ergeben und damit einen Fehler signalisieren. Nun könnte man mithife der Syndromtabelle (Tab. 5.1) den Fehler finden und sogar korrigieren. Hat man noch mehr Informationen über die Bits, wie z.B. einen Soft-wert über die Wahrscheinlichkeit, dass man hier auch das richtige Bit entschieden hat, so kann man weiter gehen. Es wird davon ausgegangen, dass der Vektor wie in Gl. 5.49 dargestellt wird, der Prozess ist visualisiert in Abb. 5.9.

$$\begin{array}{lccccccc} c = & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ p = & 0{,}9 & 0{,}5 & 0{,}8 & 0{,}9 & 0{,}7 & -0{,}1 & 0{,}9 \end{array} \tag{5.49}$$

Die Wahrscheinlichkeiten p geben an, wie sicher man sich ist, die jeweils richtige Entscheidung (0 oder 1) getroffen zu haben. Man kann sie als Logarithmic Likelihood Ratio (LLR) berechnen, indem man die Wahrscheinlichkeit, dass eine Null gesendet wurde, durch die

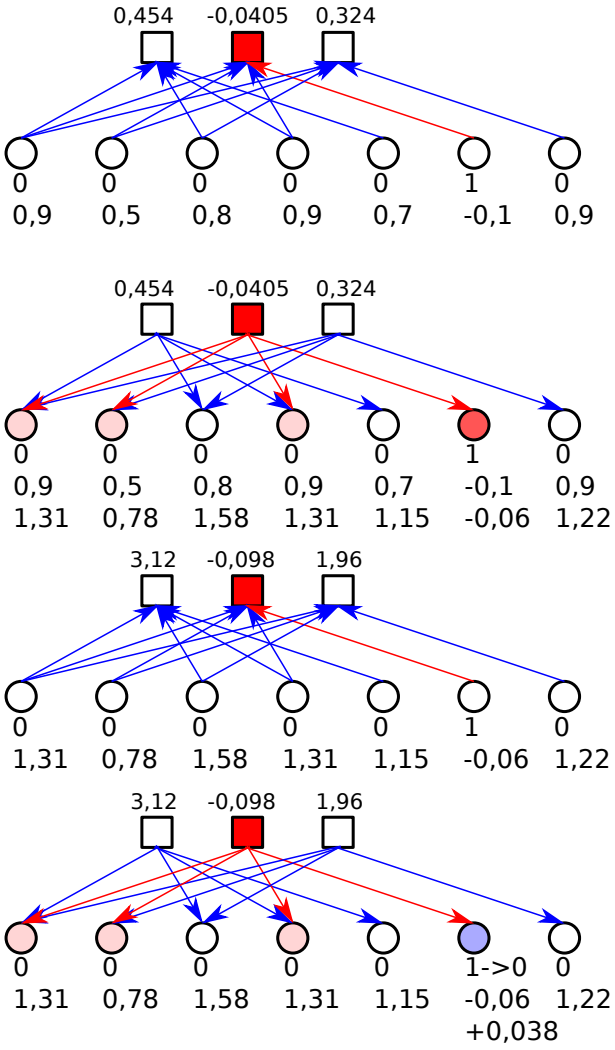


Abbildung 5.9: Vereinfachte Dekodierung mit Fehlerkorrektur

Wahrscheinlichkeit, dass eine Eins gesendet wurde dividiert, man erhält das Wahrscheinlichkeitsverhältnis bzw. die Likelihood Ratio (LR).

$$\mathrm{LR}(y) = \frac{P(y|c=0)}{P(y|x=1)}\,. \tag{5.50}$$

y ist das empfangene Symbol und x ist das gesendete Symbol Der Wert $P(y|x=0)$ ist die bedingte Wahrscheinlichkeit, dass man das Symbol y erhält, wenn ein $x=0$ gesendet wurde. Entsprechend ist $P(y|x=1)$ die Chance mit der nach Aussenden einer $x=1$ am Empfänger das beachtete y auftritt. $\mathrm{LR}(\cdot)$ ist nun ein positiver Wert mit

$$\mathrm{LR}(y) < 1 \qquad P(y|x=1) > P(y|x=0) \tag{5.51}$$

$$\mathrm{LR}(y) > 1 \qquad P(y|x=0) > P(y|x=1) \tag{5.52}$$

also größer Eins für eine wahrscheinlich gesendete Null und sonst kleiner als Eins.
Um das oben schon angedeutete Vorzeichen (Null ist positiv, Eins ist negativ) zu bekommen, nimmt man den Logarithmus und erhält eben die LLR:

$$\mathrm{LLR}(y) = \log\left(\mathrm{LR}(y)\right) = \log\left(\frac{P(y|c=0)}{P(y|x=1)}\right) \tag{5.53}$$

und damit ist

$$\mathrm{LLR}(y) < 0 \qquad P(y|x=1) > P(y|x=0) \tag{5.54}$$

$$\mathrm{LLR}(y) > 0 \qquad P(y|x=0) > P(y|x=1) \tag{5.55}$$

In dem Beispiel nach Gl. 5.49 und Abb. 5.9 ist die einzige Eins der Fehler, daher kann man realistisch davon ausgehen, dass dieses evtl. nur ein kleiner Pegel war, und man sich hier besonders unsicher ist. Die Wahrscheinlichkeit einer Eins wird negativ bewertet. In jedem Check-Knoten wird nun das Produkt der Wahrscheinlichkeiten gebildet. Man sieht, dass ein Knoten einen schwach negativen Wert hat. Hier liegt also ein Fehler vor, eigentlich müssten bei einem richtigen Codewort alle Werte positiv sein.
Im nächsten Schritt übermitteln die Check-Knoten ihre Werte zurück an die Variablen-Knoten. An dieser Stelle muss man nun mit den neuen Informationen evtl. die Variablen überarbeiten. Hier macht man das so: Es wird die Summe aller Eingangswahrscheinlichkeiten gebildet. Da gehen eben alle Werte von den Check-Knoten ein, wie auch die Werte von ursprünglich empfangenen Variablenknoten. Es fällt auf, dass es nun Werte größer als Eins gibt, hier müsste man skalieren, was in diesem Beispiel nicht gemacht wird, und das Vertrauen in die (falsche) Eins ist weiter geschwunden, während alle anderen Wahrscheinlichkeiten gewachsen sind.
Der Prozess wird nun mit neuen Werten wiederholt. Nach der zweiten Iteration kippt der Wert an der Fehlerstelle in die richtige Richtung. Im nächsten Schritt wären alle Check-Knoten erfüllt, und der Dekodierungsprozess könnte abgebrochen werden.
Der hier beschriebene Prozess ist stark vereinfacht dargestellt, damit man ihn leicht von Hand nachvollziehen kann. Es gibt verschiedenen Varianten hier, die bspw. in [**sas19**, 40] dargestellt sind.
LDPC haben sich als äußerst leistungsfähig erwiesen. Sie kommen - wie sonst nur noch die später zu behandelnden Turbocodes der Shannongrenze nach (Gl. 4.73) sehr nahe. Auch solche Leistungsuntersuchungen stellen wir hier (noch) nicht an.

5.3 Zyklische Codes

Gerade für lange Blöcke ist die erforderliche Matrix-Multiplikation sehr aufwändig und auch die Schieberegisterdarstellung erfordert viele Register, deren Inhalt dann schließlich miteinander verrechnet werden muss. Darum sucht man nach einfacheren Codes, die durch ihre interne hohe Ordnung eine einfachere Codierung der Eingabewerte ermöglichen. Dieses sind Codes, bei denen die zyklische Verschiebung eines Codewortes wiederum ein gültiges Codewort ergeben. Ist also

$$\vec{c}_i^T = (c_0,c_1,\ldots,c_{n-2},c_{n-1}) \tag{5.56}$$

gültig, so ist auch

$$\vec{c}_j^T = (c_{n-1},c_0,\ldots,c_{n-3},c_{n-2}) \tag{5.57}$$

gültig. Zur Betrachtung dieser Codes ist es bequem, die Bitströme, bzw. Vektoren der Länge n mit Indizes von 0 bis $n-1$ als Polynome vom Grad $n-1$ zu schreiben:

$$c(x) = c_0 + c_1 x + \ldots + c_{n-2}x^{n-2} + c_{n-1}x^{n-1}. \tag{5.58}$$

Die einzelne Stelle des Bits wird also einfach durch die entsprechende Potenz von x markiert. Für die Beschreibung mittels Polynomen ist es zweckmäßig, den konstanten Term mit der Potenz Null und damit auch mit dem Index (also c_0) zu versehen. Daher sind ggü. der Betrachtung mit Matrizen die Indizes um eine Stelle verschoben.

5.3.1 Generatorpolynom

Liegt ein Generatorpolynom vom Grad m

$$g(x) = g_0 + g_1 x + \ldots + g_m x^m \tag{5.59}$$

mit vorzugsweise mindestens $g_0 = g_m = 1$ vor, dann ist klar, dass die (Modulo-2) Multiplikation mit irgendeinem Datenwort mit nur einer Stelle ungleich 0, auch dargestellt als Polynom, eine jeweils verschobene Version des Generators ergibt:

$$d(x) = x^o \tag{5.60}$$

$$d(x) \times g(x) = g_0 x^o + g_1 x^{o+1} + \ldots + g_m x^{o+m}. \tag{5.61}$$

Die Multiplikation des Generators mit irgendeinem Datenwort

$$d(x) = d_0 + d_1 x + \ldots + d_{k-1}x^{k-1} \tag{5.62}$$

ergibt dann ein zu übertragendes Polynom mit

$$b(x) = d(x) \times g(x) = b_0 + b_1 x + ... + b_{n-1} x^{n-1}\,, \tag{5.63}$$

wobei $n - 1 = m + k - 1$ gilt. Damit liegt - wie auch in der Betrachtung mit Matrizen - eine Blocklänge von $n = m + k$ vor. Ein nach (Gl. 5.63) gebildetes Wort ist ohne Rest durch das Generatorpolynom teilbar, es ergibt sich folgerichtig

$$\frac{b(x)}{g(x)} = d(x). \tag{5.64}$$

In dem so gebildeten Code ist das Datenwort nicht mehr sichtbar enthalten, der Code ist also unsystematisch.

5.3.1.1 Systematischer zyklischer Code

Ein systematischer Code lässt auch hier leicht konstruieren. Dazu wird das Datenwort zunächst mit x^{n-k} multipliziert und dann durch das Generatorpolynom dividiert. Es ergibt sich also

$$\frac{x^{n-k} \times d(x)}{g(x)} = z(x) + \frac{k(x)}{g(x)}. \tag{5.65}$$

Hier ist $z(x)$ uninteressant und wird weggeworfen, während $k(x)$ den Rest der Division darstellt. Zur Übertragung wird dann das Wort

$$b(x) = x^{n-k} \times d(x) + k(x) \tag{5.66}$$

vorgesehen. Die Datenbits d werden durch die einfache Multiplikation einfach in die höherwertigen Bitstellen verschoben, der Divisionsrest als Checksumme wird einfach angehängt. Im Decoder wird eine Division mit dem Generatorpolynom durchgeführt:

$$\frac{b(x)}{g(x)} = \frac{x^{n-k} \times d(x)}{g(x)} \oplus \frac{k(x)}{g(x)} \tag{5.67}$$

$$= z(x) \oplus \frac{k(x)}{g(x)} \oplus \frac{k(x)}{g(x)} \qquad = z(x). \tag{5.68}$$

Der letzte Schritt gilt, weil in der Modulo-2 Arithmetik die Addition ($\oplus$) dasselbe bewirkt wie eine Subtraktion. Das nach Glgn. (5.65,5.66) konstruierte Codewort ist also auch wieder ohne Rest durch den Generator teilbar. Taucht ein Rest ungleich Null auf, so stellt dieser wiederum das Syndrom dar und kann zur Fehlerkorrektur verwendet werden.

Beispiel zum zyklischen Code Das Generatorpolynom für die (7,4)-Hammingcode ist mit

$$g(x) = x^3 + x + 1 \tag{5.69}$$

gegeben und es soll das Datenwort (1111), also

$$d(x) = x^3 + x^2 + x + 1 \tag{5.70}$$

gesichert werden. Dazu wird $d(x) \times x^3$ durch den Generator dividiert:

$$\begin{array}{lll}
 & (x^6 + x^5 + x^4 + x^3 &) : (x^3 + x + 1) = x^3 + x^2 + 1 \\
\oplus & \underline{(x^6 \quad + x^4 + x^3} & \underline{)} \\
 & x^5 & \\
\oplus & \underline{(x^5 \quad + x^3 + x^2} & \underline{)} \\
 & x^3 \quad + x^2 & \\
\oplus & \underline{(x^3 \quad + x + 1} & \underline{)} \\
 & x^2 \quad + x + 1 = k(x) &
\end{array} \tag{5.71}$$

Übertragen werden dann die Daten gefolgt von dem gerade berechneten Rest, also

$$b(x) = x^6 + x^5 + x^4 + x^3 + x^2 + x + 1. \tag{5.72}$$

Die Probe zeigt, dass dieses Polynom ohne Rest durch den Generator teilbar ist:

$$\begin{array}{lll}
 & (x^6 + x^5 + x^4 + x^3 + x^2 + x + 1 &) : (x^3 + x + 1) = x^3 + x^2 + 1 \\
\oplus & \underline{(x^6 \quad + x^4 + x^3} & \underline{)} \\
 & x^5 & \\
\oplus & \underline{(x^5 \quad + x^3 + x^2} & \underline{)} \\
 & x^3 \quad + x + 1 & \\
\oplus & \underline{(x^3 \quad + x + 1} & \underline{)} \\
 & 0 \quad = k(x) &
\end{array} \tag{5.73}$$

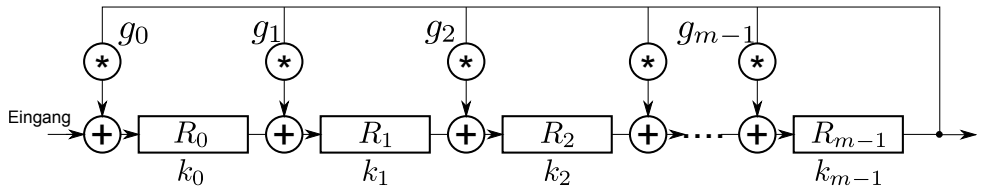


Abbildung 5.10: Prinzipielle Darstellung eines Schieberegisters zur Berechnung des Divisionsrests.

Ein Bitfehler an der letzten Stelle ergibt aber

$$\begin{array}{rl}
 & (x^6 + x^5 + x^4 + x^3 + x^2 + x + 0) : (x^3 + x + 1) = x^3 + x^2 + 1 \\
\oplus & \underline{(x^6 \qquad + x^4 + x^3 \qquad\qquad\quad)} \\
 & \quad x^5 \\
\oplus & \quad \underline{(x^5 \qquad + x^3 + x^2 \qquad\quad)} \\
 & \qquad\quad x^3 \qquad + x \\
\oplus & \qquad\quad \underline{(x^3 \qquad + x + 1)} \\
 & \qquad\qquad\quad 1 \quad = k(x)
\end{array} \tag{5.74}$$

einen Rest, der dem Syndrom entspricht und entsprechend der Syndromtabelle 5.1 einem Fehlervektor (hier: letztes Bit) zugeordnet werden.

5.3.2 Realisierung mit rückgekoppeltem Schieberegister

Eine besonders effiziente Realisierung ist die durch rückgekoppelte Schieberegister. Betrachtet man sich nämlich die Ausführung der Polynomdivision, so wird die in den folgenden Schritten vorgenommen:

1. Subtrahiere die Generatorfolge von den höherwertigsten Bits $(n-1), \ldots, (n-k-1)$. Das höherwertigste Bit ist hier Teil des Ergebnisses (wenn es überhaupt interessiert).

2. Schiebe das Datenpolynom eine Stelle weiter und wiederhole den obigen Schritt

3. Wenn das letzte Datenbits ins Schieberegister geschoben wurde, und die letzte Subtraktion ausgeführt wurde, dann steht der Divisionsrest im Speicher.

Die Realisierung ist in Abb. 5.10 dargestellt. Die Berechnung des Syndroms erfolgt genauso. Ein Beispiel für das schon oben verwendete fehlerhafte Übertragungswort (1111110)

Nr.	Daten							Schiebe		
	b_0	b_1	b_2	b_3	b_4	b_5	b_6	k_0	k_1	k_2
0	0	1	1	1	1	1	1	0	0	0
1		0	1	1	1	1	1	1	0	0
2			0	1	1	1	1	1	1	0
3				0	1	1	1	1	1	1
4					0	1	1	1⊕ 1	1⊕ 1	1
								0	0	1
5						0	1	1⊕1	0⊕1	0
								0	1	0
6							0	1⊕0	0⊕0	1
								1	0	1
7								0⊕1	1⊕1	0
								1	0	0

Tabelle 5.6: Schieberegister bei Dekodierung von (1111110) = ($d_6 \ldots d_0$) mit Generator (1011) = ($g_3 \ldots g_0$). Es bleibt Rest (001) = (k_2,k_1,k_0) (grün) und der Quotient (1101) = ($q_3 \ldots q_0$) grau hinterlegt.

ist in Tab. 5.6 gezeigt. Entscheidend ist, auf die richtige Bit-Ordnung zu achten, dass nämlich das höherwertigste Bit (MSB) in dieser Darstellung rechts steht und damit zuerst in das Register geschoben wird.

5.3.3 Verwendete Generatorpolynome und Frameaufbau

Praktische Verwendung finden nur wenige Generatorpolynome, die dann auch immer wieder auftauchen. Die Koeffizienten werden oft - je nach Geschmack - in polynomialer Darstellung wie oben, also Liste der Rückkoppelstellen, als Hex- oder auch als Octaldarstellung angegeben.
In Tab. 5.7 sind einige Beispiele und deren Verwendung angegeben.
Damit können wir hier den Aufbau eines kompletten Übertragungsrahmens im Ethernet besprechen. Der Aufbau für 10/100Base-T ist in Abb. 5.11 gezeigt.
Für die Übertragung von 10 Gbit/s über das gleiche physikalische Medium, nämlich das Twisted Pair Cable, ist erheblich mehr Aufwand zu treiben. Hier werden verschiedene Kodierungen ineinander verschachtelt, eben um näher an die Shannongrenze heran zu kommen. Der Aufbau eines solchen Rahmens ist in Abb. 5.12 und 5.13 beschrieben und

Tabelle 5.7: Einige Generatorpolynome

Name	Polynom	Hex	Bemerkung
Hamming (7,4)	$x^3 + x + 1$	$0x0B$	
Hamming (15,11)			
CRC-4	$x^4 + x + 1$	$0x13$	
CRC-8	$x^8 + x^6 + x^5 + x + 1$	$0x163$	10GBase-T
CRC-16	$x^{16} + x^{15} + x^2 + 1$	$0x18005$	
CRC-16	$x^{16} + x^{12} + x^5 + 1$	$0x11021$	ZigBee
CRC-24	$x^{24} + x^{10} + x^9 + x^6$		
	$+x^4 + x^3 + x + 1$	$0x100065B$	Bluetooth LE
CRC-32	$x^{32} + x^{26} + x^{23} + x^{22}$		
	$+x^{16} + x^{12} + x^{11} + x^{10} + x^8$		
	$+x^7 + x^5 + x^4 + x^2 + x + 1$	$0x104C11DB7$	Ethernet (multi)

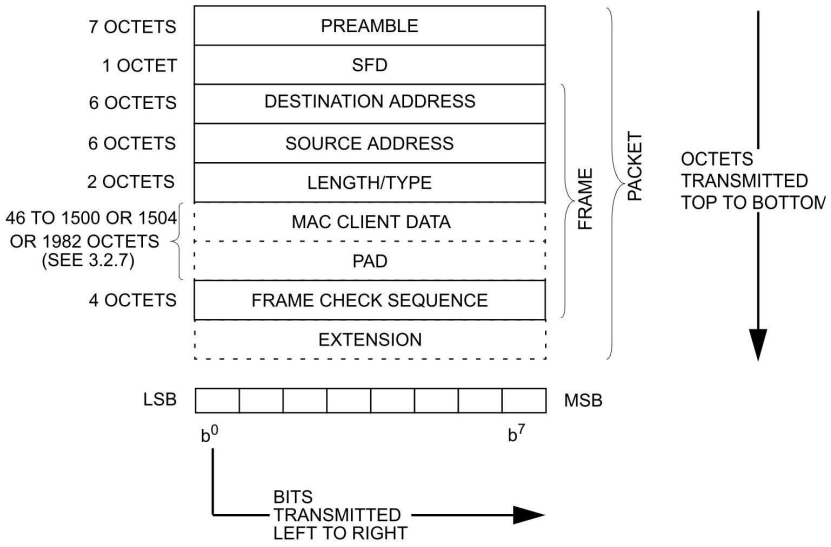


Abbildung 5.11: Rahmenaufbau im 10/100Base-T nach [13, Figure 3-1].

wird im Folgenden nach dem Original [16] ausgeführt:

55.3.2.2.5 Block structure

Blocks consist of 65 bits. The first bit of a block is the data/ctrl header. Blocks are either data blocks or control blocks. The data/ctrl header is 0 for data blocks and 1 for control blocks. The remainder of the block contains the payload.

Data blocks contain eight data characters. Control blocks begin with an 8-bit block type

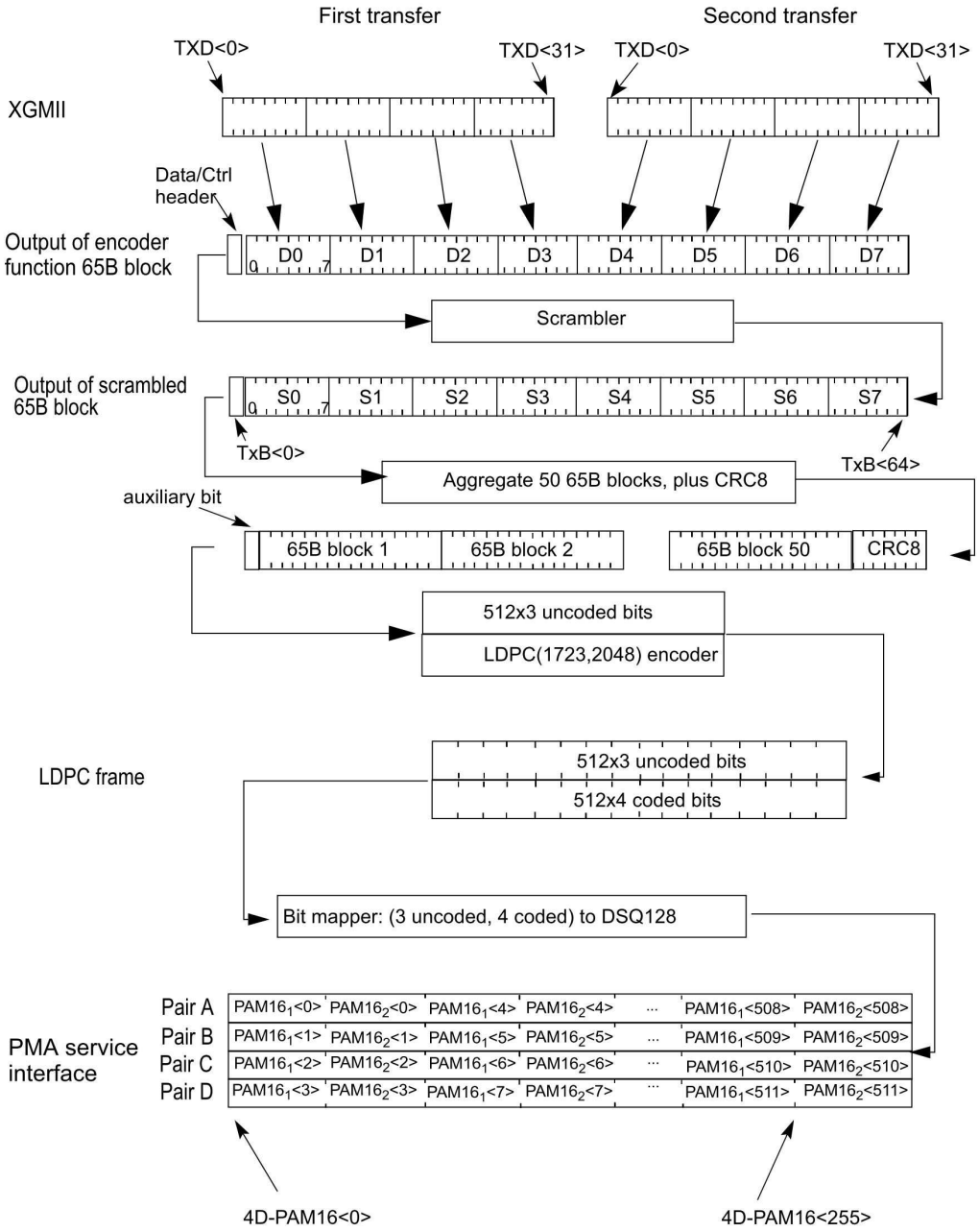


Abbildung 5.12: Bitordnung im 10GBase-T [16, Figure 55-6].

field that indicates the format of the remainder of the block. For control blocks containing a Start or Terminate character, that character is implied by the block type field. Other control characters are encoded in a 7-bit control code or a 4-bit O Code. Each control block contains

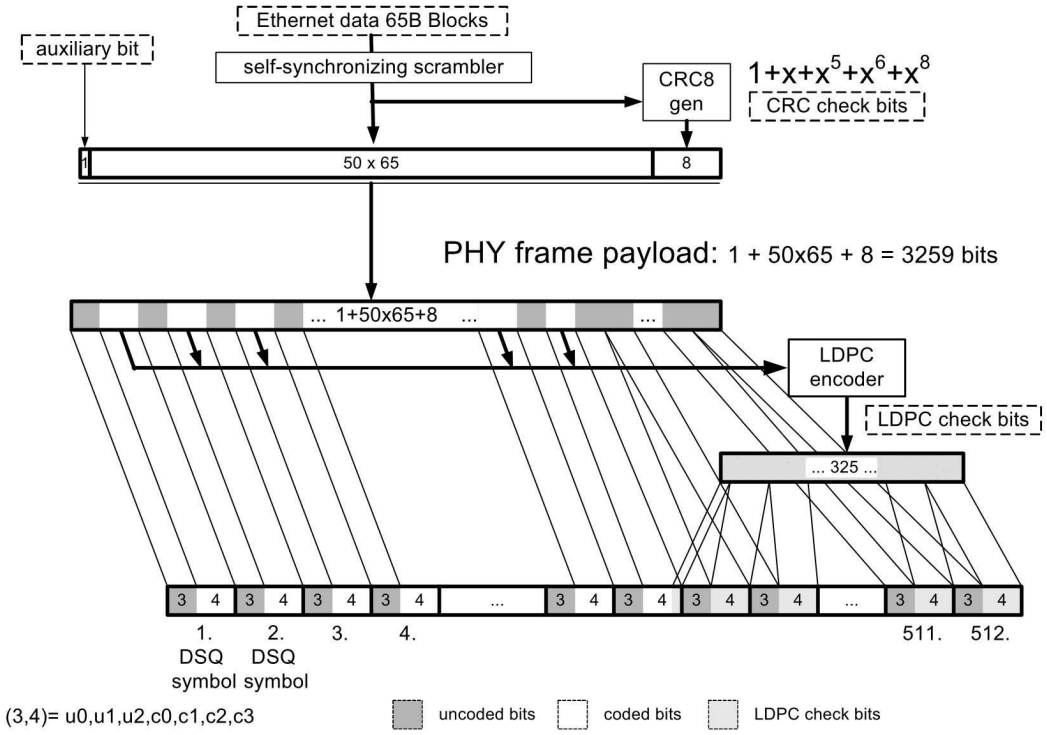


Abbildung 5.13: Details der Bitordnung im 10GBase-T [16, Figure 55-8].

eight characters.

The format of the blocks is as shown in Figure 55-9. In the figure, the column labeled Input Data shows, in abbreviated form, the eight characters used to create the 65-bit block. These characters are either data characters or control characters and, when transferred across the XGMII interface, the corresponding TXC or RXC bit is set accordingly. Within the Input Data column, D0 through D7 are data octets and are transferred with the corresponding TXC or RXC bit set to zero. All other characters are control octets and are transferred with the corresponding TXC or RXC bit set to one. The single bit fields (thin rectangles with no label in the figure) are sent as zero and ignored upon receipt.

Bits and field positions are shown with the least significant bit on the left. Hexadecimal numbers are shown prepended with '0x', and with the least significant digit on the right. For example the block type field 0x1E is sent as 01111000 representing bits 1 through 8 of the 65-bit block. The least significant bit for each field is placed in the lowest numbered position of the field.

All unused values of the block type field are reserved.

55.3.2.2.16 CRC8

The aggregated 50 65B blocks shall be used to calculate the cyclic redundancy check parity

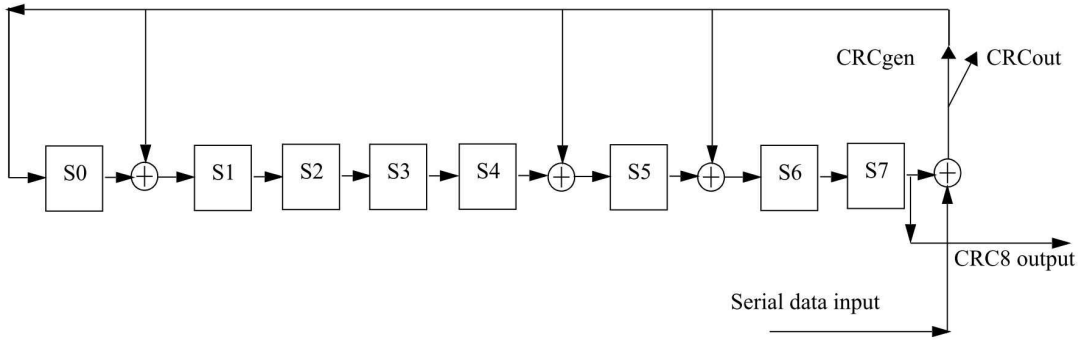


Abbildung 5.14: Schieberegisterdarstellung der Checksummenberechnung nach CRC-8 [16, Figure 55-11].

bits for each PHY frame. The parity bits are generated by the following CRC8 cyclic generator polynomial:

The CRC8 shall produce the same result as the implementation shown in Figure 55-11. In Figure 55-11 the 8 delay elements S0,..., S7, shall be initialized to zero. Afterwards the 50 65B blocks of serial data input are used to compute the CRC8 with the switch connected, which is setting CRCgen in Figure 55-11. After all the 50 65B blocks have been processed, the switch is disconnected (setting CRCout) and the 8 values stored in the delay elements are transmitted in the order illustrated, first S7, followed by S6, and so on until the final value S0.

55.3.2.2.17 LDPC encoder

The resulting payload of scrambled 50 65B blocks, followed by the corresponding 8 bits from the CRC8 and preceded by 1 auxiliary bit results in a total payload of $50 \times 65 + 8 + 1 = 3259$ bits. The use of this bit for vendor-specific communication is outside the scope of this document. For the purposes of this standard it is ignored by the link partner. From the total payload of 3259 bits, 1536 (3 bits for each of the 512 DSQ128 symbols) are uncoded bits and the remaining 1723 shall be encoded by the LDPC(1723, 2048) generator matrix G. G is described in Annex 55A.

The LPDC encoding takes the 1723 bit input code vector x = [x0 x1 x2 ... x1722], and shall generate the 2048 bit codeword c represented by the matrix multiplication $c = x \times G$. For both x and c the encoder shall follow the notation described in 55.3.2.2.3 where the LSB (leftmost element of the vectors x and c) is the first bit into the LDPC encoder and the first transmitted bit.

5.4 Reed-Solomon Codes

Eine praktisch sehr wichtige Klasse von Kodierungsverfahren arbeitet nicht nur bitweise, sondern auf Gruppen von Bits, üblicherweise auf acht Bits, also einem Byte. Diese Verfahren basieren auf den Arbeiten von Reed und Solomon [] und sollen hier nun dargestellt werden. Einsatzgebiete sind die Compact Disk (CD) und auch die Fehlerkorrektur im DVB.

5.4.1 Galois-Felder, rechnen auf endlichen Körpern

Zur effektiven Mathematik auf Gruppen von Bits (Bytes) ist eine weitgehende Umdefinition der bisher bekannten Mathematischen Operationen notwendig. Hier streubt sich der Geist, da das in Jahrzehnten Schulzeit erlernte teilweise über Bord geworfen werden muss. Es soll eine kurze, praxisorientierte Einführung folgen.
Bei der Anwendung der bisher bekannten Rechenoperationen Addition und Multiplikation ist der Zahlenraum unbegrenzt, d.h. je mehr addiert oder multipliziert wird, desto weiter wachsen die Ergebnisse und desto weiter muss man den Zahlenraum fassen. Dieses ist unpraktisch, wenn man eigentlich nur Bytes (typischerweise) behandeln möchte, allerdings bleibt man bei den genannten Operationen immer noch im Bereich der natürlichen Zahlen, wenn man nur natürliche Zahlen verwendet. Die Inverse Operation zur Multiplikation, die Division, liefert allerdings schon bei natürlichen Zahlen Ergebnisse aus einer vollkommen anderen Zahlenmenge, aus den rationalen Zahlen.
Somit definiert man Operationen in endlichen Körpern, auch endliche Felder oder Galois Felder genannt, die immer Ergebnisse liefern, die streng im Körper bleiben. Die Grundmenge dazu sind Galois-Felder $GF(p)$ wobei p hier zunächst eine Primzahl ist, als Beispiel wird $p = 5$ verwendet.
Der Addition und der Multiplikation wird eine Modulo-Operation nachgeschaltet, so dass alle Ergebnisse mod p zu sehen sind. So wird aus $4+4 = 8 = 5+3 = 3 \mod 5$ und $4 \cdot 4 = 16 = 5 \cdot 3+1 = 1 \mod 5$. Die gesamten Operationen sind in Tabelle 5.8 zusammengefasst. Die nachgeschaltete Modulo-Operation stellt sicher, dass die Ergebnisse immer im Körper bleiben. Assoziativgesetz und Distributivgesetz können leicht nachgeprüft werden. Auffällig ist auch, dass es bzgl. der Multiplikation ein inverses Element aus dem Zahlenkörper gibt, also ein Element, das mit der gewählten Zahl multipliziert die Eins ergibt. Für 2 ist das 3, denn $2 \cdot 3 = 6 = 1 \mod 5$. Damit ist dann auch die Division definiert und auch bezüglich der Division wird der Körper nicht verlassen. Bleibt noch die Subtraktion, die wird einfach genau wie die Addition ausgeführt, kein Unterschied.

Tabelle 5.8: Addition (links) und Multiplikation (rechts) in $GF(5)$

$+$	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

$\cdot$	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Weiterhin existiert ein primitives Element α, das ist letztlich eine Zahl deren $p-1 = 4$ Potenzen in irgendeiner Reihenfolge alle Elemente (ohne die Null) des Körpers überstreichen, also hier mit 2

$$2^0 = 1 \qquad 2^1 = 2 \qquad 2^2 = 4 \tag{5.75}$$

$$2^3 = 3 \qquad 2^4 = 1 \qquad (2^5 = 2)\,. \tag{5.76}$$

Ab α^{p-1} beginnt alles wieder von vorn. So ein Element existiert immer.

5.4.1.1 Erweiterungskörper

Man fragt sich nun, was diese Primkörper mit Bytes zu tun haben, nun, sie sind die Basis der erweiterten Galois-Felder $GF(p^m)$. Zwar kann man nun allgemein p als Primzahl annehmen und m als irgendeine andere Zahl, aber das ist praktisch nicht von Bedeutung. In den bedeutsamen Fällen verwendet man $p = 2$ und (in Rechnerimplementationen idR) $m = 8$, also $2^8 = 256$, ein Byte. Für übersichtliche Beispiele wird in dieser Abhandlung oft auch $m = 3$, also $2^3 = 8$ verwendet. Nun wird man schnell feststellen, dass zwar die Modulorechnung oben noch funktioniert, aber kein primitives Element mehr vorhanden ist. Das ist aber wichtig.
Um diesen Mangel nun auch noch zu beheben, wird zunächst die Zahl (das Element im $GF(p^m)$ als Polynom mit Koeffizienten $0 \leq a_i < p$ geschrieben. Für die binären Zahlen $p = 2$ ist dieses schon aus vorangegangenen Kapiteln bekannt, und andere braucht man praktisch nicht. Damit ist eine Zahl in $GF(2^3)$ bzw. $GF(2^8)$

$$e(x) = a_2x^2 + a_1x^1 + a_0 \tag{5.77}$$

$$e(x) = a_7x^7 + a_6x^6 + a_5x^5 + a_4x^4 + a_3x^3 + a_2x^2 + a_1x^1 + a_0 \tag{5.78}$$

Tabelle 5.9: Primitive Polynome

m	Polynome	m	Polynome
1	0x03	2	0x07
3	0x0B	4	0x13
5	0x25	6	0x43
7	0x83	8	0x11d, 0x171

mit Koeffizienten $a_i \in \{0,1\}$. Die Addition und damit auch die Subtraktion erfolgen nun komponentenweise im Grundkörper, bei $p = 2$ handelt es sich also einfach um die exklusiv-oder Verknüpfung. Somit wird aus $7 + 3 = 0b111 \oplus 0b011 = 0b100 = 4$. Da nur komponentenweise gearbeitet wird ist klar, dass der Körper endlich ist, also die Elemente niemals über $p^m - 1$ wachsen.
Die Multiplikation erfolgt nun wie man es für Polynome macht, bei $m = 3$ ist das

$$e(x) = e_2x^2 + e_1x^1 + e_0 \tag{5.79}$$

$$f(x) = f_2x^2 + f_1x^1 + f_0 \tag{5.80}$$

$$e(x) \cdot f(x) = e_2f_2x^4 + (e_1f_2 + e_2f_1)x^3 + (e_0f_2 + e_2f_0 + e_1f_1)x^2 + (e_0f_1 + e_1f_0)x^1 + e_0f_0 \tag{5.81}$$

und sprengt mit Exponenten über $m-1 = 2$ den Körper. Es gibt aber spezielle, so genannte primitive Polynome, die eine Modulo-Rechnung ermöglichen, mit der die Eigenschaften des endlichen Körpers erhalten bleiben. Diese gehen letztlich wieder auf Primzahlen zurück und sind in Tabelle. 5.9 angegeben. Für manche Werte von m gibt es mehrere solche Polynome, für $m = 8$ sind zwei aufgeführt, das erste wird bspw. im AES zur Chiffrierung von Daten verwendet.
Zur Erläuterung werden nun noch die Eigenschaften dieser Polynome aufgeführt. So ist

Irreduzibles Polynom ein Polynom vom Grad m, das nicht durch ein Produkt von zwei anderen Polynomen mit Grad $< m$ dargestellt werden kann, das irreduzible Polynom ist damit also nicht ohne Rest durch andere Polynome teilbar.

Primitives Polynom ist nun ein Irreduzibles, wenn alle Potenzen mit $0 \dots 2^m - 1$ seiner Nullstellen Modulo dieses Polynoms alle Elemente des Körpers $1 \dots 2^m - 1$ in irgendeiner Reihenfolge erzeugen.

So ist bei $m = 3$ und primitivem Polynom $0x0b = 0b1011$ eine Nullstelle die $0b10 = 2$, denn

Tabelle 5.10: Logarithmiertabelle

i	Komponente			Binär	Dezimal	Berechnung
$-\infty$			0	000	0	
0			1	001	1	
1		α		010	2	
2	α^2			100	4	
3		α	$+1$	011	3	$\alpha^3 = \alpha + 1$
4	α^2	$+\alpha$		110	6	$\alpha^4 = (\alpha + 1)\alpha$
5	α^2	$+\alpha$	$+1$	111	7	$(\alpha + 1)\alpha^2 = \alpha^3 + \alpha^2$
6	α^2		$+1$	101	5	$\alpha^3\alpha^3 = (\alpha + 1)^2 = \alpha^2 + 1$
7			1	001	1	$(\alpha^2 + 1)\alpha = \alpha^3 + \alpha = 1$

$$(0b10)^3 + 0b10 + 1 = 0b1000 + 0b10 + 1 = 0b1011 = 0 \bmod 0b1011\,. \tag{5.82}$$

Und was nützt das nun? Damit können Logarithmiertabellen aufgestellt werden, die sehr effektive Rechenalgorithmen unter Umgehung der oben explizit ausgeführten Polynom-Multiplikation erlauben. Für $GF(2^3)$ ist dieses in Tabelle 5.10 gezeigt. Praktisch macht man sich zu Nutze, dass

$$a \cdot b = e^{\log a + \log b} \tag{5.83}$$

ist, diese Rechnung also auf zwei Logarithmierungen (über Tabellen), eine Addition und eine Exponentiation (einfach die umgekehrte Tabelle verwenden) zurückgeführt werden kann. Für die Division ist entsprechend die Differenz der Logarithmen zu bilden.
Damit sind die wesentlichen Vorarbeiten für die Reed-Solomon Codes und auch die Kryptographie gemacht. Natürlich kann man nach der Darstellung hier alles nur glauben, es wurde nichts bewiesen. Beweise und weitere Verfeinerungen finden sich in der weiterführenden Literatur.
Aus der Tabelle 5.10 lässt sich nun die Multiplikationstabelle 5.11 ableiten, die hier der Vollständigkeit halber und zur Referenz gegeben ist.

5.4.2 Polynomkörper und Kodierung

Viele Kodierungsverfahren fassen Symbole zu Blöcken zusammen. In den vorherigen Kapiteln waren diese Symbole immer einzelne Bits, nun werden zunächst einige wenige (prak-

Tabelle 5.11: Multiplikationstabelle

	0	1	2	3	4	5	6	7
0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7
2	0	2	4	6	3	1	7	5
3	0	3	6	5	7	4	1	2
4	0	4	3	7	6	2	5	1
5	0	5	1	4	2	7	3	6
6	0	6	7	1	5	3	2	4
7	0	7	5	2	1	6	4	3

tisch bis zu einem Byte, also acht) Bits zu einem Symbol zusammengefasst. Eine Kette dieser Symbole dient dann als Eingang zum Kodieralgorithmus.
Besonders verwirrend ist nun, dass diese Kette von Symbolen wieder als ein Polynom beschrieben wird. Dieses Polynom ist unabhängig von dem oben bei der Behandlung der Galois-Felder eingeführten.
Ein Datenpaket wird also als

$$c(x) = \sum_{i=0}^{k-1} c_i x^i = c_{k-1}x^{k-1} + ... + c_1 x^1 + c_0 \tag{5.84}$$

dargestellt.
Die Reed-Solomon Codes werden wie folgt bezeichnet:

$$\text{RS}(p^m,k,n) \tag{5.85}$$

mit bekanntem p^m, also dem Werteraum, n der Länge und k der Anzahl der Nachrichtensymbole. Es liegt damit eine Minimaldistanz von $d_{min} = n-k+1$ vor, und es können somit $(n-k)/2$ fehlerhafte Symbole erkannt und behoben werden. Da ein Symbol mehrere (nämlich m) Bits enthält, werden entsprechend mehr Bitfehler behoben. Diese symbolhafte Struktur ist vorteilhaft, wenn man davon ausgeht, dass Fehler oft in Gruppen, also dicht benachbart auftreten.
Ganz analog zu den Hamming-Codes und Zyklischen Checksummen wird auch hier wieder ein Generatorpolynom konstruiert. Hierzu verwendet man ein recht einfaches, nämlich

$$g(x) = \prod_{i=0}^{n-k-1} (x + \alpha^i) \tag{5.86}$$

Dieses ist nun ein Polynom, das die ersten Potenzen der o.g. Nullstelle (also der Zwei) des Primitiven Polynoms selbst als Nullstellen hat. In $GF(2^3)$ kann so ein Generatorpolynom

$$g(x) = x^2 + 3x + 2 \tag{5.87}$$

sein, die Nullstelle $x_0 = 1$ führt auf $1^2 + 3 \cdot 1 + 2 = 0b001 + 0b011 + 0b010 = 0$ und $x_1 = 0b10$ führt auf $0b10^2 + 0b11 \cdot 0b10 + 0b10 = 0b100 + 0b110 + 0b010 = 0$.
Es ergeben sich hiermit verschiedene Darstellungen des Codes, so kann man gemäß Gl. 5.84 einfach n verschiedene Auswertungen des Polynoms an den Stellen α^i übertragen. es werden also nicht die ursprünglichen Symbole c_i verwendet, sondern die

$$C_i = c(\alpha^i) = \sum_{j=0}^{k-1} c_j \left(\alpha^i\right)^j = c_{k-1}\alpha^{k-1} + ... + c_1\alpha^1 + c_0\,, \tag{5.88}$$

wobei α wieder die Nullstelle des primitiven Polynoms aus der GF-Rechnung ist, also in der Regel $\alpha = 2$ gilt. Damit werden nun nicht nur k Werte übertragen, sondern $n > k$. Da man weiß, dass diese n Werte die Auswertung eines Polynoms mit k Koeffizienten sind, ist klar, dass Redundanz vorliegt, und dass man, wenn ein paar dieser Auswertungen auf dem Übertragungsweg zerstört werden, sie aus den verbleibenden mindestens k fehlerfreien Werten rekonstruieren kann. Der Nachteil dieser sehr einfachen Darstellung ist aber, dass die Ursprungsnachricht nicht mehr in den übertragenen Symbolen enthalten ist. Der Code ist unsystematisch.
Besser ist ein systematischer Code, denn dann muss nach Empfang der Symbole nur noch das Syndrom (die Checksumme) berechnet werden, und wenn dieses Null ist, dann liegt kein Fehler vor. Vor allem bei geringen Fehlerraten bereitet dieses Verfahren einen deutlich geringeren Aufwand. Einen systematischen Code generiert man durch

1. Verschieben der k Informationsstellen so dass man das Informationspolynom

$$c'(x) = c'_{n-1}x^{n-1} + ... + c'_{n-k}x^{n-k} \tag{5.89}$$

   erhält.

2. Polynomdivision mit dem Generatorpolynom

$$c'(x) : g(x) = i(x) \text{ mit Rest } r(x) \tag{5.90}$$

3. Zusammenbau zu

$$c''(x) = c'(x) + r(x) = c'_{n-1}x^{n-1} + ... + c'_{n-k}x^{n-k} + r(x)\,. \tag{5.91}$$

Da Subtraktion und Addition hier die gleiche Operation sind, ist nun $c''(x)$ ohne Rest durch $g(x)$ zu dividieren, bzw. es ergibt sich das Syndrom

$$s(x) = c''(x) \bmod g(x)\,. \tag{5.92}$$

Ist $s(x) = 0$, dann trat kein Fehler auf.
Ein Beispiel soll hier noch betrachtet werden. Dazu werden im $GF(2^3)$ die Daten $k = 5$

$$c'(x) = (5x^4 + 4x^3 + 3x^2 + 2x + 1)x^2 \tag{5.93}$$

kodiert mit obigem Generatorpolynom, so dass mit dem Rest

$$c'(x) : g(x) = i(x) \text{ mit Rest } 5x + 4 \tag{5.94}$$

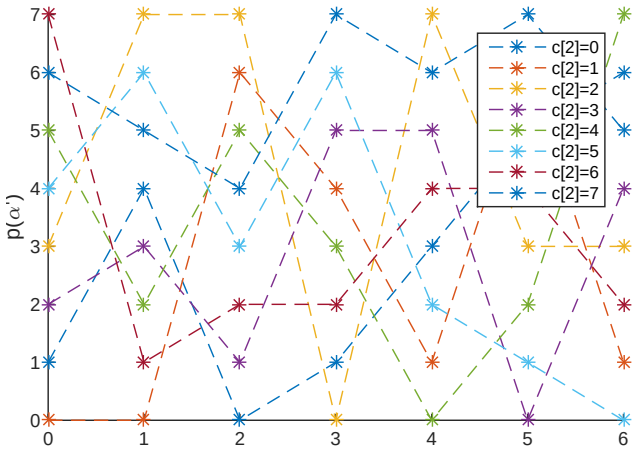
der kodierte Datenstrom 4,5,1,2,3,4,5 übertragen wird. Unter der Annahme, dass an dritter Stelle (also an $c_0 = c''_2$, was vorher 1 war) ein Fehler entsteht, ergeben sich die $2^3 - 1$ Auswertungen des Polynoms an den in Tabelle 5.10 gegebenen Stellen wie in Abb. 5.12 und Tabelle 5.12.
Nur das eine, nämlich unverfälscht übertragene, Wort hat an den ersten beiden Stellen den Wert Null, alle anderen weisen von Null verschiedene Werte auf, sind also ungültig. Diese beiden Stellen genügen aber, um (a) heraus zu finden, welche Stelle fehlerhaft ist, und (b) wie diese Stelle korrigiert werden muss. Dieses sind zwei Freiheitsgrade, weswegen auch zwei Stellen für die vollständige Korrektur benötigt werden. Aus der Konstruktion der Codes heraus ist auch einsichtig, wenn auch hier nicht bewiesen, dass die Länge der kodierten Wörter maximal $p^m - 1$, also hier 7, werden darf.

5.4.3 Dekodierung und Fehlerkorrektur

Wie so häufig ist die Fehlerkorrektur aufwändiger als die Kodierung, so auch hier. Im ersten Schritt wird davon ausgegangen, dass die Position der Fehler bekannt ist. Im obigen Beispiel ist das die dritte Position im empfangenen Wort.

Tabelle 5.12: Auswertungen des übertragenen Polynoms bei Fehler an dritter Stelle

$c_0 = c_2''$	s_0	s_1	Polynomwerte				
0	1	4	0	1	3	5	6
1	0	0	6	4	1	6	1
2	3	7	7	0	7	3	3
3	2	3	1	5	5	0	4
4	5	2	5	3	0	2	7
5	4	6	3	6	2	1	0
6	7	1	2	2	4	4	2
7	6	5	4	7	6	7	5



Man weiß über das (notwendigerweise gültige) Codewort, dass die Syndrompositionen, das sind hier die ersten beiden Stellen, Null sein müssen, und dass die dritte Stelle falsch ist. Letzteres ist in der Regel erst nach Anwendung eines weiteren, später beschrieben Algorithmus, bekannt. Abstrakter dargestellt, muss für ein gültiges Codewort

$$c(\alpha^i) = 0 \,\text{für}\, 0 \leq i < (n-k) \tag{5.95}$$

gelten.
Aus diesen Informationen kann man nun die fehlerhafte Stelle rekonstruieren, indem man nämlich den passenden Koeffizienten findet, der das Syndrom zu Null zwingt. Man schreibt das empfangene, fehlerhafte Wort als

$$u(x) = c''(x) + e(x) = c'_{n-1}x^{n-1} + ... + c'_0x^0 + e_{i_1}x^{i_1} + ... + e_{i_\tau}x^{i_\tau} + ... \tag{5.96}$$

mit den bekannten $1 \leq j \leq \tau(=1)$ Fehlerstellen $i_j(=2)$. Für obiges einfaches Beispiel hat man also

$$u(x) = 5x^6 + 4x^5 + 3x^4 + 2x^3 + (1+e)x^2 + 5x + 4 \tag{5.97}$$

$$u(\alpha^0 = 1) = 5 + 4 + 3 + 2 + (1+e) + 5 + 4 = s_0 \tag{5.98}$$

$$u(\alpha^1 = 2) = 5 \cdot 2^6 + 4 \cdot 2^5 + 3 \cdot 2^4 + 2 \cdot 2^3 + (1+e) \cdot 2^2 + 5 \cdot 2^1 + 4 = s_1\,. \tag{5.99}$$

Hier wurde, natürlich ohne dass man es kennt, das richtige Codewort plus den Fehlerterm eingesetzt. Das komplette richtige Codewort muss jeweils zu Null führen, so dass der Teil

ohne den Fehlerterm e in der Modulo-Rechnung Null ergibt. Somit ist

$$e = s_0 = 1 \tag{5.100}$$

$$e \cdot 0b100 = s_1\,. \tag{5.101}$$

Für die fehlerhafte Übertragung mit $c_0 = c_2'' = 0$ (richtig wäre 1) findet man also aus Tab. 5.12 $s_0 = 1 = e$ und $s_1 = 4 = 4e$ und somit auch $e = 1$. Die Korrektur ergibt sich, wenn man nun die empfangene und fehlerhaft Stelle (hier 0) additiv, also xor, verknüpft mir e und so erhält man hier den korrigierten Wert $c_0 = 1$.
Ein weiteres Beispiel verdeutlicht das: Es wird dieselbe Stelle als Fehlerstelle verwendet, weil ansonsten eine komplett neue Tabelle 5.12 aufwändig erstellt werden müsste. Nun sei $c_0 = c_2'' = 7$ empfangen worden, und $s_0 = e = 6$ und $s_1 = 4e$ und hier auch $e = 6$, weil $4 \times 6 = 5 = s_1$ gemäß Tab. 5.11. Die additive xor-Verknüpfung mit der empfangenen 7 ergibt den korrigierten Wert 1.
Aus diesen Überlegungen erhält man also grundsätzlich zwei Gleichungen, die hier gleiche Ergebnis für den Fehler (die Korrektur) e liefern. Eine Gleichung hätte gereicht, aber beide müssen stimmen. In der Realität verwendet man eine Gleichung, um den Fehler zu lokalisieren, was hier über die Voraussetzung geschehen ist, und den anderen, um die Korrektur zu berechnen.

5.4.3.1 Ablauf der Dekodierung und Fehlerkorrektur

Der Ablauf der Fehlerkorrektur ist wie folgt:

1. Bestimmung des Syndroms, bspw. durch Auswertung des übertragenen Polynoms an den vermutlichen Nullstellen. Wenn diese Null sind, ist alles getan und die Nutzdaten können einfach übernommen werden.

2. Ist das Syndrom nicht Null sind mit dem Berlekamp-Massey-Algorithmus die Positionen der Fehler zu bestimmen

3. Die Fehler an den nun bekannten Stellen sind zu korrigieren.

5.4.3.2 Bestimmung der Fehlerstellen (Berlekamp-Massey Algorithmus)

Die Bestimmung der Nullstellen erfolgt nach dem Berlekamp-Massey Algorithmus (BMA), der auf dem Syndrom-Polynom arbeitet. Wenn $s(x)$ das Syndrom ist, dann ist der Algorithmus im folgenden C-Programm dargelegt.

```
/**
 * \brief Uses the Berlekamp–Massey Algorithm to find the error locations
 */
int8_t rs_bma(uint16_t ls,uint8_t *s,uint16_t *ll,uint8_t *lambda,uint8_t *eloc)
{
        uint16_t ii,ll_old;
        uint8_t l,j,lb;
        uint8_t b[RS_MAX_LOCEN];
        uint8_t lambda_old[RS_MAX_LOCEN];
        uint8_t delta;

        lambda[0]=1;
        *ll=1;
        b[0]=1;
        lb=1;

        l=0;
    delta=0;
    for (j=0;j<ls;j++)
    {
        for (ii=lb;ii>0;ii--)
        {
                b[ii+1-1]=b[ii-1];
        }
        b[0]=0;
        lb++;
        delta=s[j];
        for (ii=1;ii<*ll;ii++)
        {
                delta=gf_add(delta,gf_mul(s[j-ii],lambda[ii]));
        }
        printf("Delta %2x\n",delta);
        if (delta!=0)
        {
                if (lb>*ll)
                {
                        ll_old=lb;
                        gf_poly_cpy(lb,b,lb,lambda_old,0);
                        gf_poly_sca(ll_old,lambda_old,delta);
                        lb=*ll;
                        gf_poly_cpy(*ll,lambda,*ll,b,0);
                        gf_poly_sca(lb,b,gf_div(1,delta));
                        *ll=ll_old;
                        gf_poly_cpy(ll_old,lambda_old,*ll,lambda,0);
                }
                printf("Delta %2x\n",delta);
                printf("lambda:\t");for (ii=0;ii<*ll;ii++){printf("%2x, ",lambda[ii]);} printf("\n");
                printf("b:\t");for (ii=0;ii<lb;ii++){printf("%2x, ",b[ii]);} printf("\n");
                gf_poly_add_sca(*ll,lambda,1,lb,b,delta,0);
                *ll=(*ll>lb?*ll:lb);
        }
        printf("lambda:\t");for (ii=0;ii<*ll;ii++){printf("%2x, ",lambda[ii]);} printf("\n");
    }
        l=*ll-1;
        if (2*l>ls)
        {
                return -1;
        }


        // Search for zero
        j=0;
        for (ii=0;ii<gf.max;ii++)
        {
                if (gf_poly_eva(*ll,lambda,(uint8_t)gf.exp[gf.max-ii])==0)
                {
                        eloc[j++]=ii;
                        printf("alfa=%2x\n",(uint8_t)gf.exp[gf.max-ii]);
                }
        }
        if (j!=l)
        {
                return -1;
        }
```

```
        else
        {
                return l;
        }
}
```

5.4.3.3 Korrektur der Lösung

Die Korrektur kann nun erfolgen, wie oben angedeutet, indem man die Interpolationseigentschaften des Polynoms verwendet. Ein erweiterter Lösungsalgorithmus wird erst im nächsten Semester vorgestellt...

```
void rs_correct(uint16_t l, uint8_t *c, uint8_t *s, uint16_t lpos, uint8_t *pos)
{
        uint16_t ii,lq,lp;
        uint8_t x[RS_MAX_LOCEN];
        uint8_t q[RS_MAX_LOCEN];
        uint8_t p[RS_MAX_LOCEN];
        uint8_t u,v,w;

        x[1]=1;
        lq=1;
        q[0]=1;
        lp=1;
        for (ii=0;ii<lpos;ii++)
        {
                x[0]=gf.exp[pos[ii]];
                lp++;
                gf_poly_mul(lp,p,lq,q,2,x);
                lq=lp;
                gf_poly_cpy(lp,p,lq,q,0);
        }
        lp=lpos;
        gf_poly_cpy(lpos,s,lp,p,0);
        for(ii=0;ii<(lp/2);ii++){u=p[ii];p[ii]=p[lp-1-ii];p[lp-1-ii]=u;}
        gf_poly_mul(lp+lq,x,lp,p,lq,q);
        lp+=lq-1;
        gf_poly_cpy(lp,x,lp,p,0);
        for(ii=lp-lpos;ii<lp;ii++){p[ii-lp+lpos]=p[ii];}
        lp=lpos;
        for(ii=0;ii<(lp/2);ii++){u=p[ii];p[ii]=p[lp-1-ii];p[lp-1-ii]=u;}
//      lq--;v
        for(ii=(lq&1);ii<lq;ii+=2)
        {
                q[(ii-(lq&1))/2]=q[ii];
        }
        lq=lq/2;
        for(ii=0;ii<(lq/2);ii++){u=q[ii];q[ii]=q[lq-1-ii];q[lq-1-ii]=u;}
        for(ii=0;ii<lpos;ii++)
        {
                u=gf.exp[gf.max-(uint16_t)pos[ii]];
                v=gf_poly_eva(lp,p,u);
                w=gf_poly_eva(lq,q,gf_mul(u,u));
                c[pos[ii]]=gf_add(c[pos[ii]],gf_div(v,gf_mul(u,w)));
        }
}
```

Kapitel 6

Analoge Modulation

Zum besseren Verständnis und des besseren Überblicks wegen, wird hier ein Ausflug in die analoge Modulation unternommen. In diesem Fall orientiert sich das Kapitel an [41, Kapitel O].

6.1 Amplitudenmodulation

6.1.1 Zweiseitenband Amplitudenmodulation (ZSB-AM)

Die wohl einfachste Modulationsform ist die Veränderung der Amplitude eines Trägersignals $u_T(t)$ mit einem niederfrequenten Signal (NF):

$$u_0(t) = u_T(t)\left(1 + m \times s(t)\right) , \tag{6.1}$$

wobei $s(t)$ den NF-Signalverlauf darstellt und als Betragsmaximum $\max\{|s(t)|\} = 1$ normiert ist. Dann ist m der Modulationsindex. Der Träger ist natürlich zeitharmonisch mit der Kreisfrequenz Ω und für ein ebenso sinusförmiges NF-Signal folgt

$$u_0(t) = \hat{u}_T \cos(\Omega t + \Phi)\left[1 + m\cos(\omega t + \phi)\right] . \tag{6.2}$$

Dieses kann man leicht mittels Additionstheorem in

$$u_0(t) = \hat{u}_T \cos(\Omega t + \Phi) + \hat{u}_T \frac{m}{2} \cos((\Omega - \omega)t + \Phi - \phi) + \hat{u}_T \frac{m}{2} \cos((\Omega + \omega)t + \Phi + \phi) \tag{6.3}$$

umwandeln. In komplexer Zeigerdarstellung folgt

$$u_0(t) = \mathrm{Re}\left\{\hat{u}_T e^{j[\Omega t + \Phi]} + \hat{u}_T \frac{m}{2} e^{j[(\Omega-\omega)t + \Phi - \phi]} + \hat{u}_T \frac{m}{2} e^{j[(\Omega+\omega)t + \Phi + \phi]}\right\} . \tag{6.4}$$

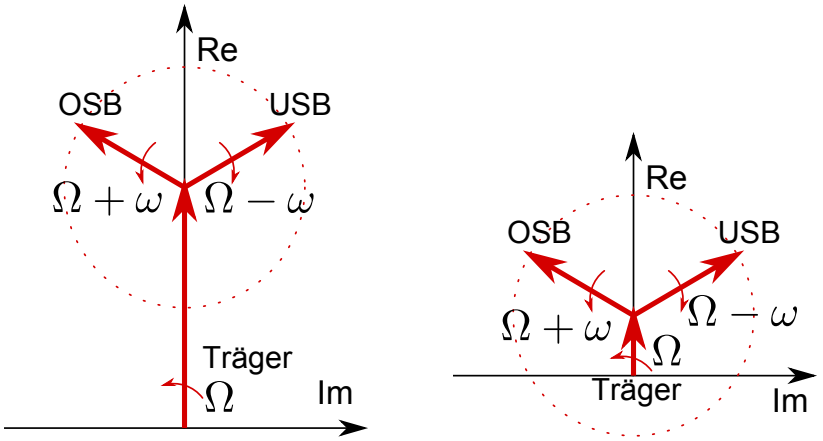


Abbildung 6.1: Zeigerdiagramm der Amplitudenmodulation mit $m = 1$ (a) und mit unterdrücktem Träger, also $m > 1$ (b)

Diese Zeigerdarstellung ist in Abb. 6.1a ausgeführt. Darstellung des Zeitverlaufs der Amplitudenmodulation und das resultierende Spektrum sind in Abb 6.2 dargestellt. Mit Modulationsindex $m = 1$ wird die Nulllinie erreicht, aber nicht durchbrochen. Wegen der zwei symmetrisch angeordneten Bänder zu jeder Seite des Trägers wird die Modulationsform Zweiseitenband-Amplitudenmodulation (ZSB-AM) genannt.
Die Frequenzbetrachtung offenbart, dass mit $B_{HF} = 2B_{NF}$ in der HF das doppelte Spektrum wie in der Niederfrequenz belegt wird. Eine Leistungsbetrachtung zeigt, dass Träger und Unteres Seitenband (USB) und Oberes Seitenband (OSB) die Energien

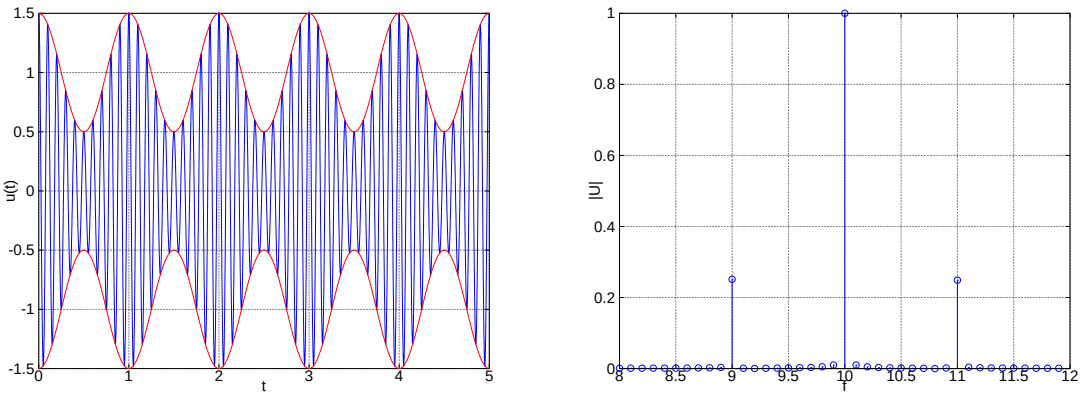
$$P_T \times R = \frac{\hat{u}_T^2}{2} \tag{6.5}$$

$$P_{USB} \times R = P_{OSB} \times R = \frac{m^2 \hat{u}_T^2}{8} \tag{6.6}$$

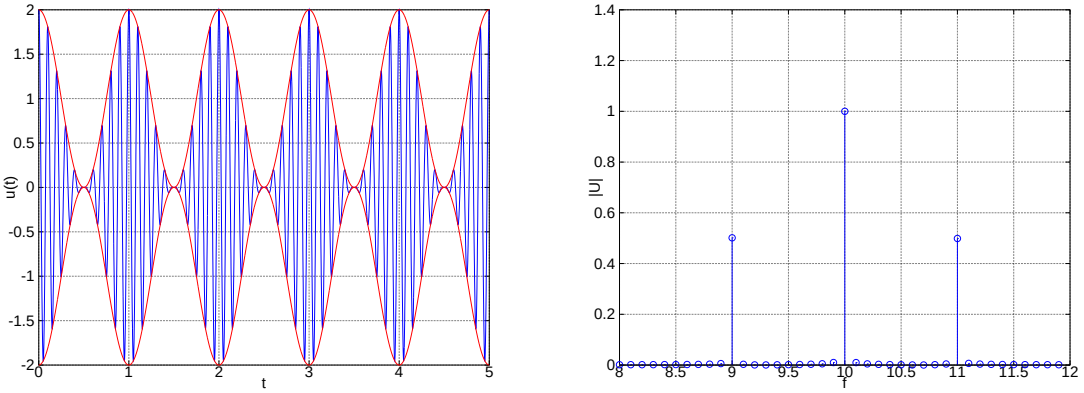
übertragen werden. Das Signal wird nur in den Seitenbändern übertragen, für eines davon steht der Anteil

$$\frac{P_{OSB}}{P_{ges}} = \frac{\frac{m^2}{8}}{\frac{1}{2} + 2\frac{m^2}{8}} = \frac{m^2}{2(2+m^2)} \tag{6.7}$$

zur Verfügung. Dieses ist im besten Fall mit $m = 1$ 1/6 der Leistung. In beiden signaltragenden Seitenbändern stehen damit 33 % der Leistung zur Verfügung, der Rest geht in den Träger. Dieser beinhaltet außer der Tatsache, dass er da ist und seiner Frequenz und Phase keine Information. In diesem Sinne ist das verschwendete Leistung.
Eine einfache AM kann mit dem Modulator nach Abb. 6.3 erzeugt werden. Die Demodulation kann einfachst durch einen Spitzenwertgleichrichter nach Abb. 6.4(a) erfolgen.



(a) (b)



(c) (d)

Abbildung 6.2: Amplitudenmodulation mit $m = 0.5$ (a,b) und $m = 1$ (c,d) inlusive Spektrum

6.1.2 Zweiseitenband-Amplitudenmodulation mit unterdrücktem Träger (ZSB-AM-uT)

Energetisch günstiger ist die Übertragung mit unterdrücktem Träger als Zweiseitenband-Amplitudenmodulation mit unterdrücktem Träger (ZSB-AM-uT) (siehe auch Abb. 6.1(b) und 6.5). Dieses kann (im Grenzübergang $m \to \infty$) durch eine einfache Modulation des NF-Signals mit dem Träger dargestellt werden:

$$u_0(t) = \hat{u}_T \cos(\Omega t + \Phi) \cos(\omega t + \phi) \tag{6.8}$$

$$= \frac{\hat{u}_T}{2} \cos((\Omega - \omega)t + \Phi - \phi) + \frac{\hat{u}_T}{2} \cos((\Omega + \omega)t + \Phi + \phi). \tag{6.9}$$

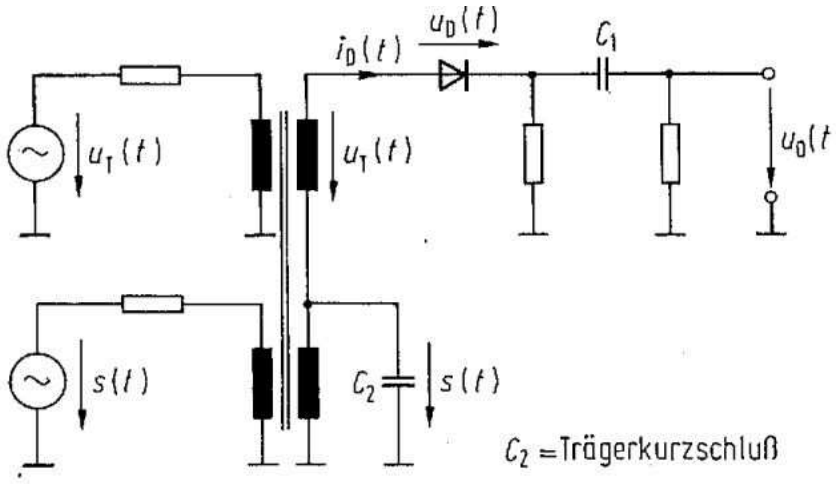
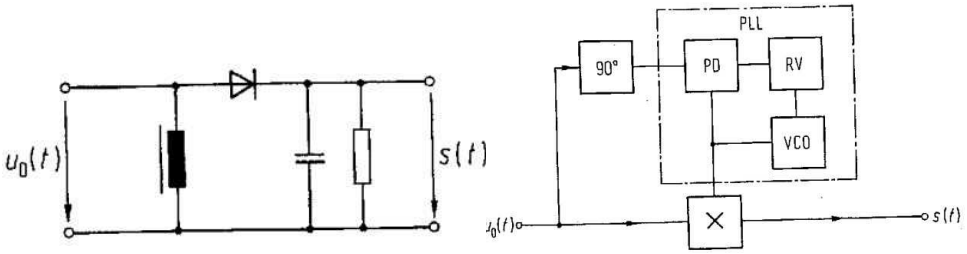
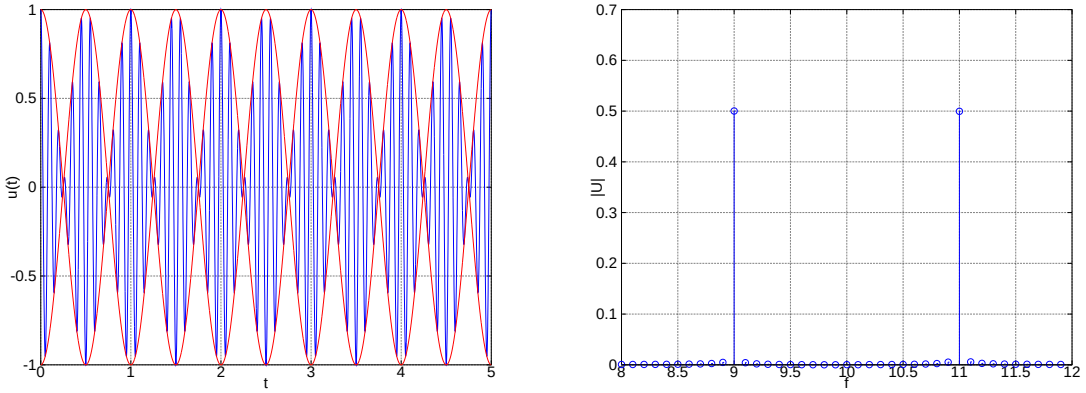


Abbildung 6.3: Amplitudenmodulator nach [41, S.O3]



(a) (b)

Abbildung 6.4: Amplitudendemodulator mit Spitzenwertgleichrichter (a) und als Synchrondemodulator (b) nach [41, S.O6]



(a) (b)

Abbildung 6.5: Amplitudenmodulation mit unterdrücktem Träger inklusive Spektrum

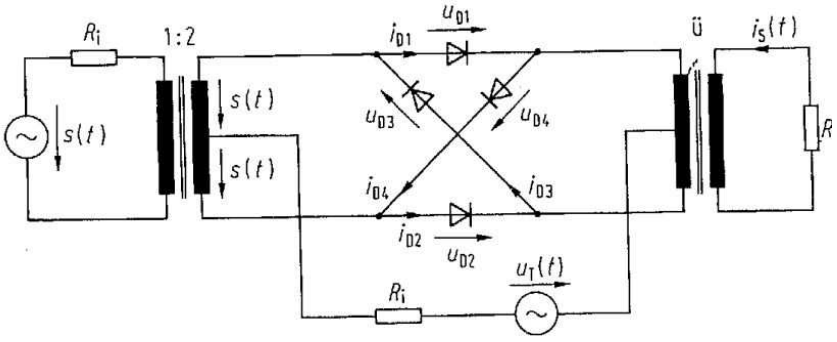


Abbildung 6.6: Amplitudenmodulator mit Trägerunterdrückung [41, S.O4]

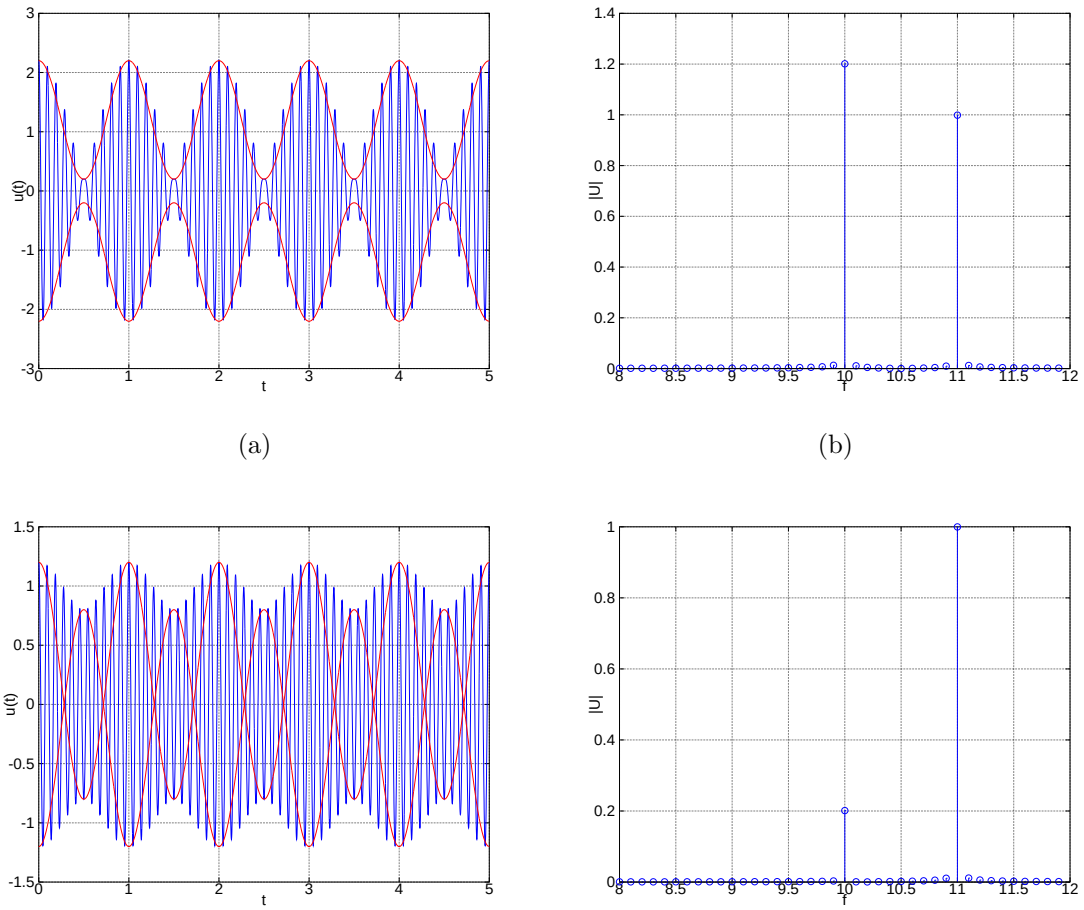
Man beachte, dass hier die Einhüllende eigentlich den Nullpunkt schneidet. Eine Demodulation mit Spitzenwertgleichrichter ist hier nicht mehr möglich, da ansonsten die phasengedrehten Anteile bei negativem NF-Signal nach "oben geklappt werden" und dort zu Verzerrungen führen. Eine solche Modulation wird mit dem Modulator nach Abb. 6.6 erzeugt. Aufgrund der Symmetrie wird der Träger zum Ausgang hin effektiv unterdrückt. Der Träger schaltet lediglich die entsprechenden Diodenpaare gerade oder verkreuzt durch und führt dadurch die Multiplikation nach Gl. 6.9 effektiv aus. Noch immer wird aber die doppelte Bandbreite des Niederfrequenzsignals benötigt.

6.1.3 Einseitenband-Amplitudenmodulation (ESB-AM)

Durch Filter oder Phasenmethode wird nur eines der Seitenbänder ausgewählt und dann als Einseitenband-Amplitudenmodulation (ESB-AM) übertragen. In der Regel wird auch der Träger unterdrückt, so dass die Formulierung (mit K der Trägerunterdrückung)

$$u_0(t) = \hat{u}_T \cos((\Omega \pm \omega)t + \Phi \pm \phi) + K\hat{u}_T \cos(\Omega t + \Phi) \tag{6.10}$$

bleibt. Abb. 6.7 zeigt Spektren und Zeitverläufe der AM-ESB. Im Vergleich mit dem rot eingezeichneten NF-Signalverlauf ist eine Phasenverschiebung zu erahnen, die in der Zeigerdarstellung 6.8 am Winkel θ abzulesen ist. In der ESB wird also durch den Wegfall des einen Seitenbandes nicht nur die Amplitude, sondern auch die Phase moduliert. Aus diesem Grunde ist die AM-ESB nur mit dem Synchrondemodulator nach Abb. 6.4(b) zu demodulieren.



(a) (b)

(c) (d)

Abbildung 6.7: Amplitudenmodulation mit $K = 1.2$ (a,b), also Trägeramplitude über Signal, und $K = 0.2$ (c,d) inlusive Spektrum (Träger bei 10, OSB bei 11). In rot ist das ursprüngliche Signal eingetragen.

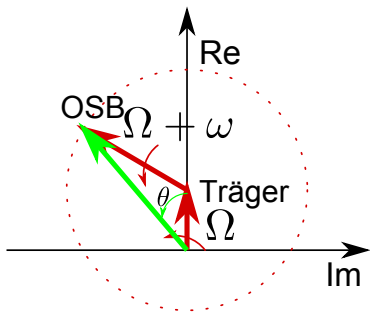
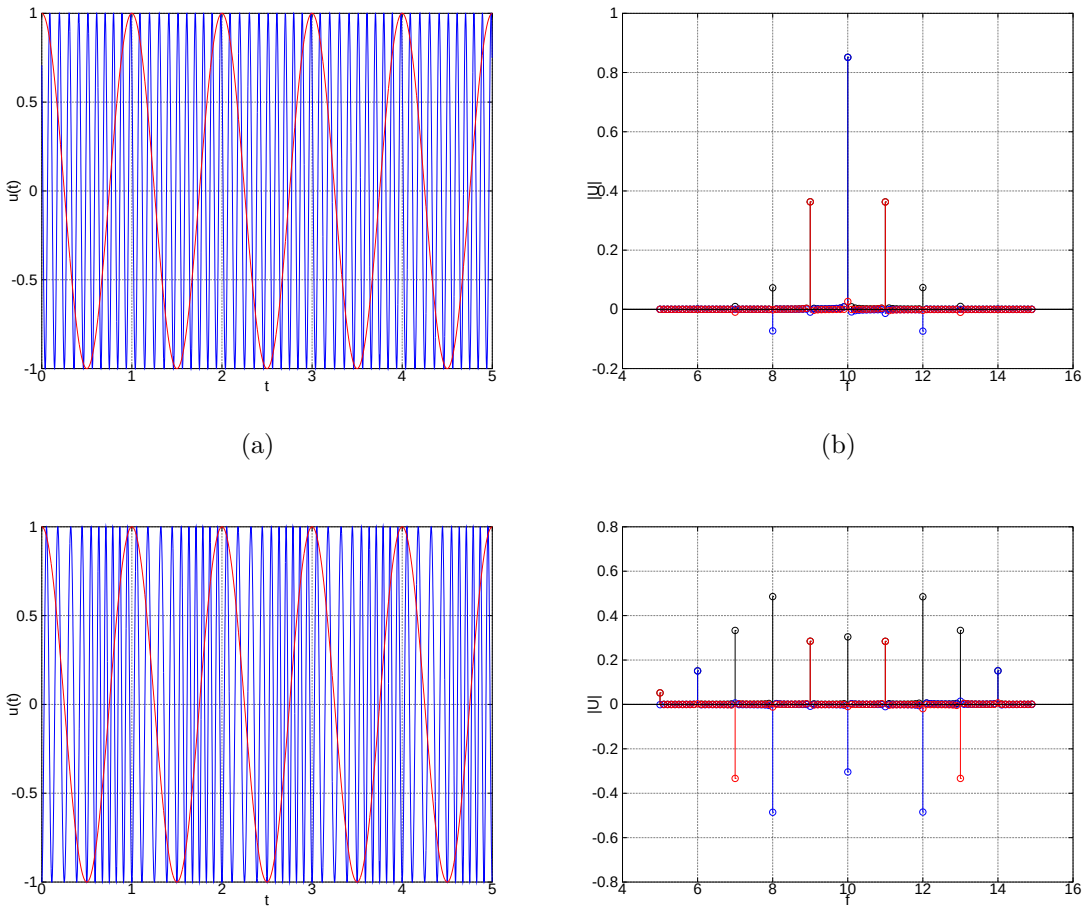


Abbildung 6.8: Zeigerdiagramm der Einseitenband-Amplitudenmodulation.



(a) (b)

(c) (d)

Abbildung 6.9: Phasenmodulation mit Spektrum für $\Delta\phi = \pi/4$ (a,b) und $\Delta\phi = \pi$ (c,d). Im Zeitverlauf ist das Signal in rot angedeutet, im Spektrum sind Real-, Imaginärteil und Betrag in rot, blau und schwarz (tlw. verdeckt) zu unterscheiden. Es sind wieder nur die positiven Frequenzen in einem Ausschnitt gezeigt.

6.2 Phasenmodulation (PM)

Neben der Amplitude eines Trägers kann natürlich auch die Phase nach

$$u(t) = \cos\left(\Omega t + \Delta\phi s(t)\right) \tag{6.11}$$

$$= \mathrm{Re}\left\{e^{j(\Omega t + \Delta\phi s(t))}\right\} \tag{6.12}$$

moduliert werden, so dass die Phasenmodulation (PM) entsteht. Die vollständige Beschreibung für eine kosinusförmige Anregung ist damit

$$u(t) = \cos\left(\Omega t + \Delta\phi\cos(\omega t)\right), \tag{6.13}$$

und es sind Zeitverlauf und Spektrum für $\Delta\phi = \pi/4$ und π in Abb. 6.9 dargestellt. Mit erhöhter Phasenverschiebung steigt auch die Ausdehnung des Spektrums. Diese ist prinzipiell unendlich ausgedehnt, fällt aber für geringe Phasenhübe schnell ab, so dass sich die relevante Information eng um den Träger konzentriert.

6.3 Frequenzmodulation (FM)

Im Allgemeinen gilt die Beziehung

$$\phi(t) = \int_0^t \omega(\tau)d\tau \tag{6.14}$$

zwischen momentaner Phase und der momentanen Frequenz. Die Phase ist demnach das Resultat aller bis zum Betrachtungszeitpunkt t da gewesenen Frequenzen $\omega(t)$. Wird nun die Frequenz gemäß einem Signal im Sinne einer Frequenzmodulation (FM) geändert, so ergibt sich schließlich

$$\phi(t) = \int_0^t \Omega + \Delta\Omega \times s(\tau)d\tau \tag{6.15}$$

$$= \Omega t + \Delta\Omega \int_0^t s(\tau)d\tau, \tag{6.16}$$

wobei eine Anfangsphase zu Null angenommen wurde. Bei wiederum kosinusförmiger Anregung ergibt sich dann

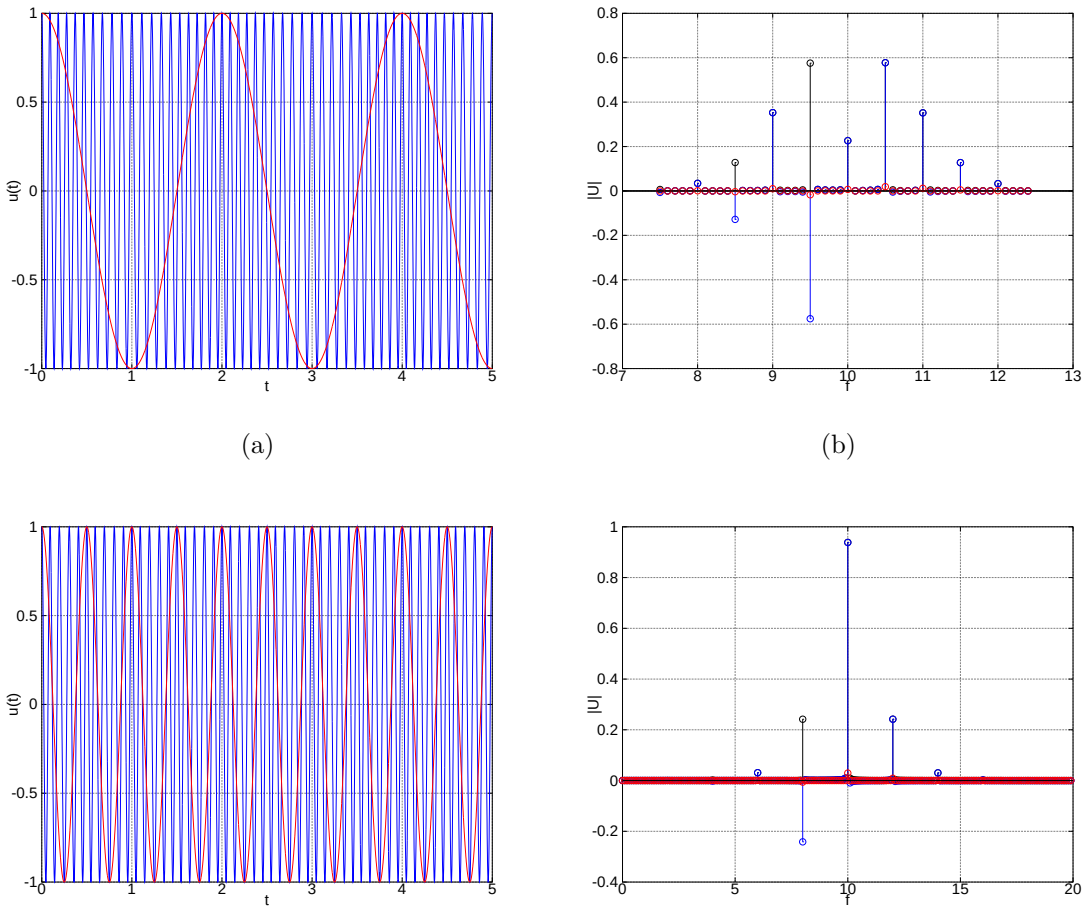
$$\phi(t) = \Omega t + \Delta\Omega \int_0^t \cos(\omega\tau)d\tau \tag{6.17}$$

$$= \Omega t + \frac{\Delta\Omega}{\omega}\sin(\omega t) \tag{6.18}$$

$$= \Omega t + M\sin(\omega t) \tag{6.19}$$

mit einem Modulationsindex $M = \Delta\Omega/\omega$, der von der Signalfrequenz abhängt und zwar derart, dass er groß für geringe und gering für große Frequenzen ist. Somit stellt sich das komplette Signal als

$$u(t) = \cos\left(\Omega t + M\sin(\omega t)\right) \tag{6.20}$$



(a) (b)

(c) (d)

Abbildung 6.10: Frequenzmodulation mit Spektrum für $\Delta\Omega/\Omega = 0{,}1$ und $\omega = \Omega/20$ ($M = 2$) (a,b) und $\omega = \Omega/5$ ($M = 0{,}5$) (c,d). Im Zeitverlauf ist das Signal in rot angedeutet, im Spektrum sind Real-, Imaginärteil und Betrag in rot,blau und schwarz (tlw. verdeckt) zu unterscheiden. Es sind wieder nur die positiven Frequenzen in einem Ausschnitt gezeigt.

dar.

Ansonsten sieht die FM für ein harmonisches NF-Signal der PM sehr ähnlich. Die Zeitverläufe für $\Delta\Omega = 0{,}1\Omega$ und unterschiedliche Signalfrequenzen sind in Abb. 6.10 dargestellt. Die geringe Signalfrequenz ($\omega = \Omega/20$) führt bei einheitlicher Signalamplitude zu einem sehr weit ausgedehnten Spektrum, hier ist $M = 2$. Die hohe NF-Frequenz von $\omega = \Omega/5$ führt wegen $M = 0{,}5$ zu einem deutlich weiter eingeengten Modulationsspektrum. Es ist zu beachten, dass die Realteile des Spektrums bei ungeraden Ordnungen links und rechts

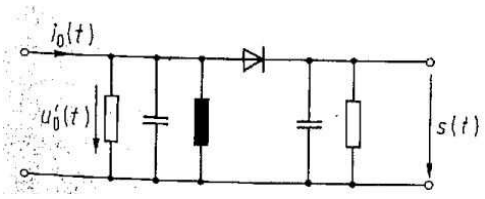


Abbildung 6.11: Flankendiskriminator [41, S.O9]

des Trägers unterschiedliches Vorzeichen aufweisen. Ansonsten entsprechen die Spektren denen einer Phasenmodulation mit entsprechendem Hub.
Prinzipiell ist die FM sehr leicht mit einem Flankendiskriminator nach Abb. 6.11 in eine Amplitudenmodulation umzuwandeln und danach an der Diode zu demodulieren.

Kapitel 7

Mobilkommunikation - Einführung

Im zweiten Teil der Vorlesung steht der Mobilfunk im Mittelpunkt, der Untertitel dieses Abschnittes könnte auch **UMTS and Friends** lauten. Seitdem Heinrich Hertz in Karlsruhe am 11. November 1886 die erste Übertragung von elektromagnetischen Wellen von einem Sender zu einem Empfänger nachgewiesen hat, entwickelte sich die Funktechnik durch Protagonisten wie Guglielmi Marconi, Nicola Tesla und viele andere sehr rasant.
In Deutschland wurde mit dem C-Netz 1985 erstmalig ein noch analoges Mobilfunksystem einer etwas breiteren Masse zugänglich. Fast gleichzeitig wurde 1982 von der CEPT (Europäische Konferenz der Verwaltungen für Post und Telekommunikation) die Groupe Spécial Mobile (kurz: GSM) ins Leben gerufen. Die Aufgabe war es, einen europaweiten oder sogar Europa übergreifenden Mobilfunkstandard ins Leben zu rufen. 1992 startete der Massen-Roll-Out der „Handys“ und damit der unglaubliche Siegeszug der Mobilfunktechnik. In Deutschland startete das System als D1-Netz der Deutschen Telekom und als D2-Netz von Mannesmann-Arcor auf 900 MHz. Später kam als weitere Frequenz im 1800 MHz-Band das E-Netz hinzu. Mittlerweile ist auch GSM weltweit etabliert, ja sogar eigentlich schon lange wieder überholt.
GSM war natürlich für Sprachkommunikation konzipiert, was bedeutet, dass Daten (auch Sprachdaten) verbindungsorientiert übertragen werden, und relativ geringe Datenraten von ca. 9,6 kbps pro Nutzer ausreichend waren, aber eben keine Pakete zwischendurch verloren gehen sollten und ein quasi-kontinuierlicher Strom von Daten (Sprache) entstehen sollte. Mit reinem GSM Daten zu übertragen ist auch heute noch eine Qual, es steht in der ursprünglichen Form ein Datenkanal von 9,6 kbps bzw. 14,4 kbps zur Verfügung. Das genügt heute natürlich nicht mehr. Auch der verbindungsorientiere Kanalaufbau, also der Aufbau des Kanals und die Reservierung desselben für eine Zeit, egal, ob gerade Daten gesandt werden oder nicht, ist der Übertragung von digitalen Daten, wie Mails oder

kleinen Paketen nicht zuträglich. So wurde das Netzwerk aufgerüstet und mit General Packet Radio System (GPRS) ein paketorientiertes Radio-Interface aufgebaut. Dieses allein erhöht aber nicht die Datenrate. Letzteres wird parallel mit High Speed Circuit Switched Data (HSCSD) erreicht, indem man einem Nutzer nicht nur einen von acht Zeitschlitzen, sondern einfach bis zu acht (praktisch vier) der acht Zeitschlitze zuordnet. Eine Datenrate von - theoretisch - $8 \times 14{,}4\,\text{kbps} = 115{,}2\,\text{kbps}$ bzw. 57,6 kbps bei vier Zeitschlitzen ist möglich. Auch das ist noch kein wirkliches Vergnügen, weshalb die Modulationsform geändert wurde und mit Enhanced Data Rates for GSM Evolution (EDGE) eine Verdreifachung der Rohdatenrate erreicht werden konnte, was dann - auch mit Veränderungen anderer Parameter der Übertragung - in Enhanced General Packet Radio System (EGPRS) zu Datenraten von 59,2 kbps und theoretisch insgesamt 473 kbps bzw. praktisch ca. 220 kbps führt.
Wem dieses noch nicht genügt, der muss auf UMTS umsteigen. Dieser Standard wurde seit 1992 federführend von der ETSI innerhalb des 3GPP entwickelt und sehr viel besser, wenn auch nicht perfekt, auf Datenverkehr ausgerichtet. Datenraten von 384 kbps bzw. in der Spitze 2 Mbps sind praktisch und theoretisch in dem erstmalig als Release 99 bekannten Version breit realisiert. Das Radio-Interface ist vollkommen anders als das des GSM. In der ursprünglichen Version ist auch UMTS noch sehr verbindungsorientiert ausgelegt. Natürlich sind Datenraten von 384 kbps bzw. 2 Mbps zu gering, weshalb auch UMTS mittels High Speed Packet Access (HSPA) erst als High Speed Downlink Packet Access (HSDPA) für den Downlink und ab Release 5 mit 14,4 Mbps und als High Speed Uplink Packet Access (HSUPA) im Uplink mit 5,7 Mbps (Rel. 6) aufgerüstet worden. Als HSPA+ werden verschiedene Verfahren zur Erhöhung der Datenrate angewandt, die dann in Downlink-Datenraten von bis zu 168 Mbps münden. Hier werden ggf. mehrere Antennen(systeme) verwendet und die Datenraten verschiedenen Träger bzw. sogar Zellen miteinander kombiniert. Unstrittig ist, dass diese Systeme mindestens unter dem Begriff 3. Generation Mobilfunk firmieren können.
In 2003 hat die International Mobile Telecommunications (IMT) als IMT-Advanced die Anforderungen an Systeme der 4. Generation (4G) herausgegeben [42], die unter (`http://www.itu.int/ITU-R/index.asp?category=information&rlink=imt-advanced&lang=en`) zusammengefasst werden:

- a high degree of commonality of functionality worldwide while retaining the flexibility to support a wide range of services and applications in a cost efficient manner;
- compatibility of services within IMT and with fixed networks;
- capability of interworking with other radio access systems;

- high quality mobile services;
- user equipment suitable for worldwide use;
- user-friendly applications, services and equipment;
- worldwide roaming capability; and,
- enhanced peak data rates to support advanced services and applications (100 Mbit/s for high and 1 Gbit/s for low mobility were established as targets for research).

Demnach ist klar HSPA noch kein Kandidat für die vierte Generation. Auch LTE in der Version nach Release 8 und 9 erreicht diese Anforderungen nicht, erst Release 10 mit über 1 Gbps kann als „4G“ bezeichnet werden. Auch LTE, welches unter dem Dach des 3GPP entwickelt und spezifiziert wird, hat gegenüber UMTS ein vollkommen überarbeitetes Radio-Interface. Methoden zur Erhöhung der Datenrate, wie

- Modulationen hoher Ordnung,
- Kanalbündelung (Carrier-Aggregation)
- meherere Antenne Multiple Inputs Multiple Outputs (MIMO)

werden natürlich in allen Systemen gleichermaßen verwendet. Einen kleinen Eindruck gibt Abb. 7.2.
Mit der IMT-2020, die in 2015 verabschiedet wurde [43], wird die fünfte Generation im digitalen Mobilfunk definiert, das sogenannte 5G oder im 3GPP als NR (New Radio), oder insgesamt als 5.th Generation Mobile Communication New Radio (5G NR)bezeichnet. Die wesentlichen Parameter sind

- Datenrate 20 Gbit/s und 10 Gbit/s im Down- und Up-link, wobei die Nutzer zu 95 % der Zeit jeweils immerhin 0,5 % davon auch wirklich haben sollen.
- Sehr kurze Latenzzeiten von 4 ms oder sogar nur 1 ms bei den sehr hoch verfügbaren und mit wenig Latenz behafteten Diensten.
- Mobilität bis 500 km/h wird unterstützt
- Es werden bis zu 1.000.000 Endgeräte pro Quadratkilometer unterstützt bei einer Kapazität von 10 Mbit/s/m^2
- Die Spektrale Effizient soll 30 bit/Hz/Sektor betragen.

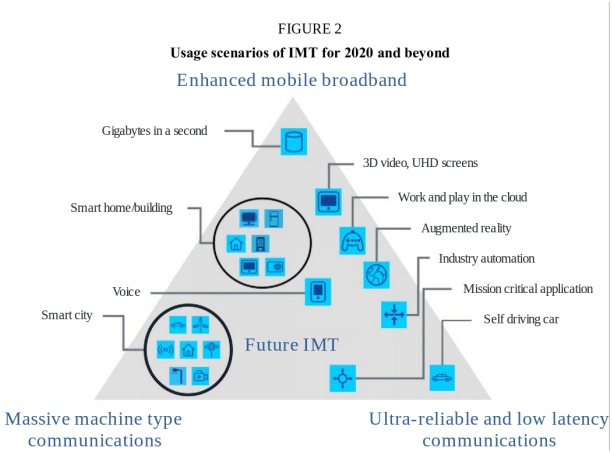


Abbildung 7.1: Nutzungs-Szenarios nach [43]

Die 5G Parameter wurden in Release 15 in 2019 erstmalig festgeschrieben und ab da kann der Roll-out beginnen, so dass in 2020 mit 5G Umsetzungen zu rechnen ist. Entscheidend ist zudem, dass Nutzungs-Szenarien noch weiter in den Vordergrund geschoben wurden, so soll es vor allem die in Abb. 7.1 beschriebenen drei Anwendungsgebiete:

eMBB enhanced Mobile Broadband für hohe Datenraten und Verfügbarkeiten auch bei hoher Geschwindigkeit und Gerätedichte,

mMTC massive Machine Type Communication, wo Maschinen, bspw. auch Autos miteinander kommunizieren, und

URLLC Ultra-Reliable and Low Latency Communications, bei dem auch wieder kritisch und mit hohen Sicherheitsanforderungen kommuniziert wird. Hier kann es ebenso bspw. um autonome Autos oder andere Systeme und auch chirurgische Eingriffe auf Distanz gehen.

In dieser Vorlesung werden einige Ideen, die den heutigen Mobilfunksystemen zugrunde liegen erläutert. Es wird aber kaum möglich sein, im Rahmen dieser Vorlesung alles dar zu legen, und gleichzeitig auch noch zu vertiefen und zu üben. Es wird eine Mischung aus Ideen, Theorien und praktischen Umsetzungen mit Übungen gezeigt.
Völlig unbeachtet geblieben sind in dieser Einführung die Varianten der Mobilkommunikation auf anderen Kontinenten, die natürlich vergleichbare Dienste, tlw. aber mit vollkommen unterschiedlichen technischen Arbeitspferden, zur Verfügung stellen.

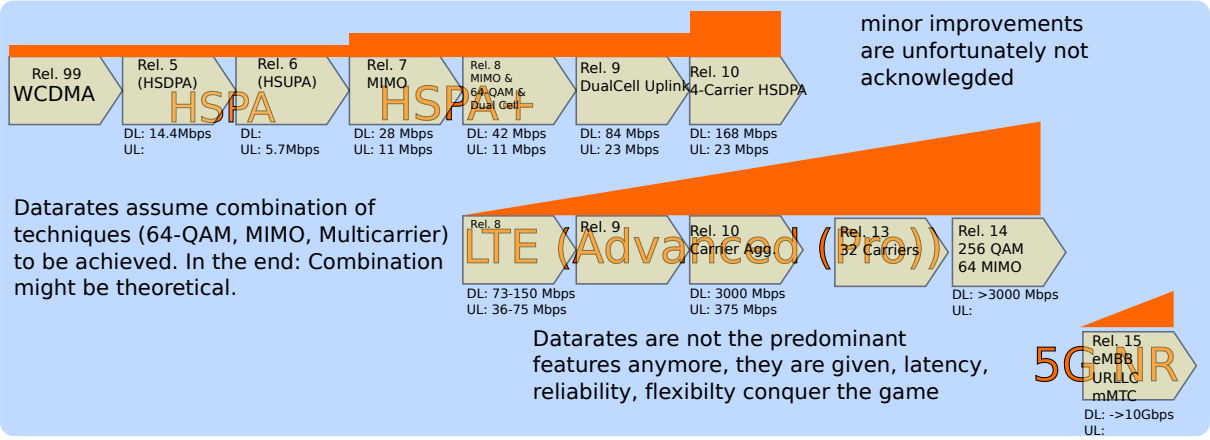


Abbildung 7.2: Übersicht über die UMTS, LTE und 5G Versionen.

7.1 Technische Basis Trivia

Folgende grundsätzliche Bemerkungen sind bei der Betrachtung von drahtlosen Kommunikationssystemen anzustellen:

1. Durch die Funkfelddämpfung ist die empfangene Leistung ggf. sehr sehr viel geringer, als die ausgesendete.

⇒ Diese großen Leistungsunterschiede von tlw. 13-14 Zehnerpotenzen sind in einem tragbaren Gerät auf engstem Raum getrennt zu verarbeiten, eine gute und sorgfältige Auslegung und Isolation von Sender und Empfänger ist notwendig.

2. Auf das Übertragungsmedium greifen tlw. unkontrollierbar viele verschiedene Nutzer zu, die koordiniert werden wollen, diese müssen sich kooperativ verhalten und sich auch gegenseitig voreinander schützen. Sender müssen optimiert werden, um nur minimale Störungen zu erzeugen, Empfänger müssen in der Lage sein auch mit Störungen durch den Nachbarn klar zu kommen.

3. Übertragungskanäle sind zeitvariant, können sich ändern, es kann zu Abschattungen und auch zu Mehrwegeausbreitungen kommen, tlw. ist gar keine direkte Sichtverbindung zwischen Sender und Empfänger mehr möglich, oder gar nötig. Ausgefeilte Konzepte zur Datensicherung und auch zur stör-immunen Übertragung sind notwendig.

4. Tragbare, batteriebetriebene Geräte müssen sehr energieeffizient arbeiten. Vor allem effiziente Endverstärker und eine effiziente Datenverarbeitung sind zu erstellen.

5. Ein Datenkanal steht in der Regel nur in einem eng definierten Frequenzbereich zur Verfügung, der gesetzlich vorgeschrieben und einzuhalten ist. Übertragung im Basisband ist daher unmöglich, Modulation unerlässlich.

7.2 Der Mobilfunkkanal

Wie schon bei der Betrachtung der drahtgebundenen Übertragung in früheren Kapiteln, steht am Beginn eine detaillierte Analyse des Übertragungskanals. Hierbei handelt es sich bei den drahtlosen Übertragungsverfahren um elektromagnetische Wellen als Träger der Information, die sich - idealerweise - im freien Raum ausbreiten. Die Ausbreitung im wirklich freien Raum gehorcht dem Frijschen Gesetz und ist mit

$$\frac{P_2}{P_1} = G_1 G_2 \left(\frac{\lambda}{4\pi r}\right)^2 \tag{7.1}$$

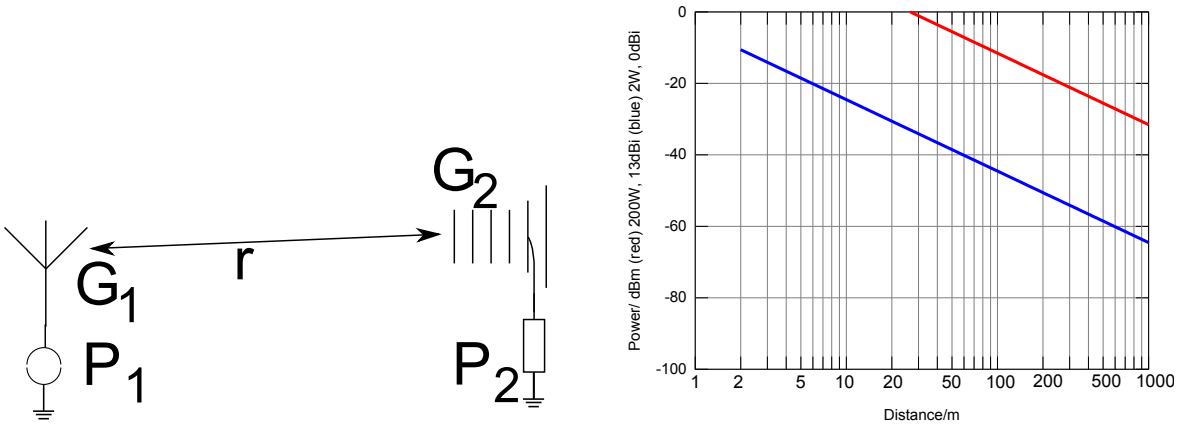
bzw. in rein logarithmischer Darstellung

$$\left.\frac{P_2}{P_1}\right|_{dB} = G_1|_{dB} + G_2|_{dB} - 20\log_{10}\left(\frac{r}{1\,\mathrm{km}}\right) - 20\log_{10}\left(\frac{f}{1\,\mathrm{GHz}}\right) - 92.44\,\mathrm{dB} \tag{7.2}$$

gegeben. Die Parameter sind in Abb. 7.3(a) gegeben. Es kann im freien Raum von einer Verringerung der Leistung mit dem reziproken Quadrat der Entfernung (relativ bezogen auf die Wellenlänge) ausgegangen werden. Dieses ist nochmal in Abb. 7.3(b) dargestellt. Praktisch gilt dieser Verlauf natürlich nur im wirklich freien Raum, also letztlich zwischen Satelliten, oder mindestens unter der Bedingung, dass die Fresnelzone von Störungen frei bleibt (Abb. 7.4). Wird diese Zone gestört, indem die eine Hälfte der Zone abgedeckt wird, so sinkt die zu empfangene Feldstärke auf die Hälfte, und damit die Leistung auf ein Viertel, um 6 dB. In der Realität findet man auf der Erde sehr stark abweichende Ausbreitungsbedingungen vor. So kann z.B. eine Verbindung wenigstens teilweise gestört werden, oder es gibt gar keine Sichtverbindung zwischen Sender und Empfänger, etwa, weil sich einer der Partner in einem Gebäude befindet, oder bspw. in einer Häuserschlucht nicht erreichbar ist. Dieses schlägt sich in sehr unangenehmen Effekten nieder, die in Abb. 7.5 dargestellt sind.

7.2.1 Mehrwegeausbreitung

Die Mehrwegeausbreitung wird nun detaillierter betrachtet (Abb. 7.6). Hier besteht neben der direkten Verbindung zwischen Sender und Empfänger ein zweiter Pfad durch die Reflexion an einem Gebäude. Natürlich legt die Welle auf dem Pfad eine längere Strecke



(a) (b)

Abbildung 7.3: Ein einfaches Kommunikationssystem in (a) und zwei beispielhafte Verläufe für bspw. eine GSM-Basisstation (rot) und ein Handy (blau) (b) im freien Raum und bei 900 MHz.

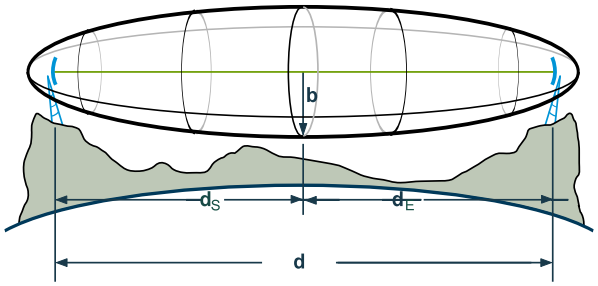


Abbildung 7.4: Darstellung der Fresnelzone (nach http://de.wikipedia.org/wiki/Fresnelzone)

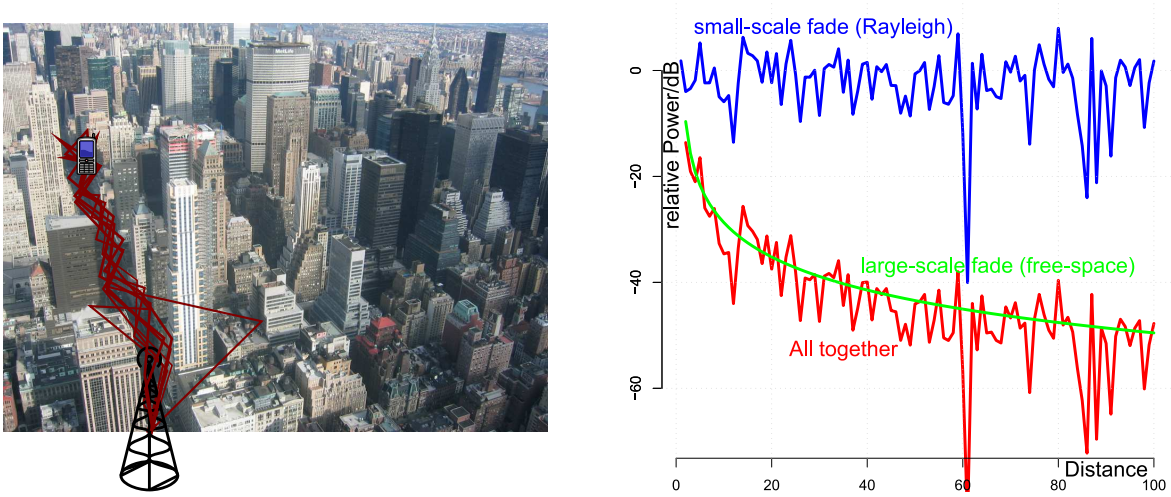
zurück und kommt deswegen gegenüber dem direkten Signal verzögert am Empfänger an. Mit mehreren betrachteten Wegen, ergibt sich eine Kanalimpulsantwort mit

$$s(t) = \sum_{m=0}^{M} \rho_m \delta_0(t - t_m), \tag{7.3}$$

die zu einer Frequenzantwort des Kanals mit

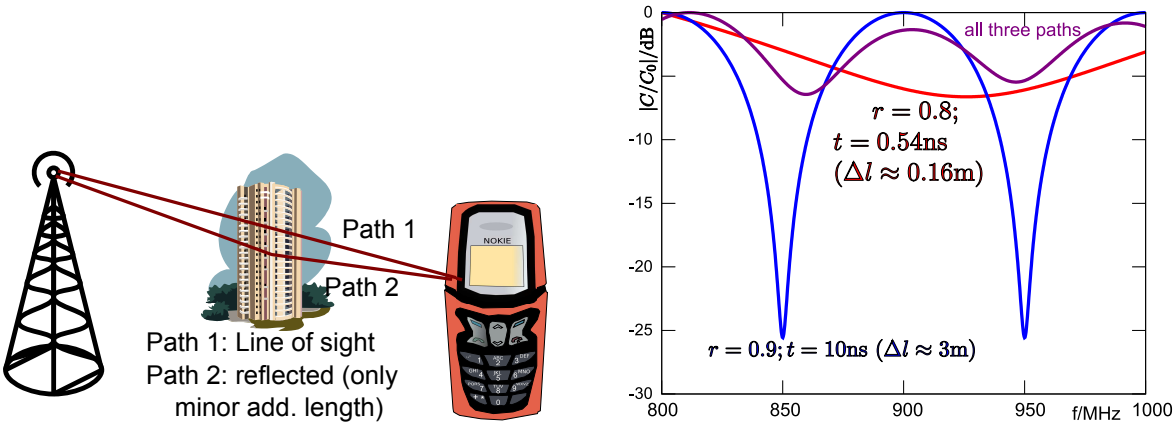
$$S(\mathrm{j}\omega) = \sum_{m=0}^{M} \rho_m e^{-\mathrm{j}\omega t_m} \tag{7.4}$$

führt. In beiden Fällen ist M die Gesamtzahl der betrachteten Pfade und m der Laufindex eines Pfades mit der Verzögerungszeit t_m und der Dämpfung ρ_m. Abb. 7.6 stellt



(a) (b)

Abbildung 7.5: Bildliche Darstellung (in Manhattan) (a) und Wirkung (b) der Mehrwegeausbreitung.



(a) (b)

Abbildung 7.6: Zwei Ausbreitungswege (a) bildlich dargestellt und (b) für ein Beispiel mit drei Wegen ausgerechnet.

die Frequenzabhängigkeit gemäß Gl. 7.4 dar, und es zeigt sich so ein frequenzselektiver Kanal. Im Klartext heißt es, dass dieser Kanal einige Frequenzanteile deutlich stärker bedämpft als andere, weil es durch die Gangunterschiede zu konstruktiven und destruktiven Interferenzen kommen kann. Hierdurch fehlen einfach einige Signalanteile oder im

schlimmsten Fall sogar ganze Frequenzbänder. Zudem ist in vielen mobilen Anwendungen der Kanal auch noch zeitvariant, das heißt, die Kanaleigenschaften ändern sich mit der Zeit. Dieses ist nicht weiter verwunderlich, denn ein Nutzer bewegt sich in der Regel; oder allein durch Änderungen von Sonnenaktivität oder auch Luftfeuchte, bzw. andere bewegte Störer (Autos, LKW) im oder am Ausbreitungsweg ändern sich die Bedingungen leicht, so dass es am Empfänger immer wieder zu (ggf. leichten) Verschiebungen der Charakteristika kommt. Eine feste, und für alle Zeiten gleiche Beschreibung des Kanals ist daher unter praktischen Gesichtspunkten wenig sinnvoll. Aus diesem Grunde hat man sich für statistische Beschreibungen entschieden.

7.2.2 Rayleigh-Kanal

Unter den Voraussetzungen, dass

- ein Kanal mit geringen Verzögerungsunterschieden vorliegt, also nur die Phase des Träger verändert wird, aber - aufgrund der deutlich geringeren Frequenz - die Phase und zeitliche Abfolge des modulierten Signals gleich bleibt,
- keine Sichtverbindung vorliegt, statt dessen aber sehr viele sehr dicht beieinander liegende Streukörper (bspw. Ionosphärenausbreitung oder städtische Umgebung) betrachtet werden müssen,

spricht man von dem Rayleigh-Kanal. Ein einfacher Kanal, wie er hier beschrieben wird, ist nicht frequenzselektiv. Die Amplitude und Phase des empfangenen Signals schwankt hier regellos und unterliegt nur einer statistischen Beschreibung. Einen Rayleigh-Prozess stellt man dar, wie in Abb. 7.7 gezeigt: Ähnlich einer IQ-Modulation werden zwei Stränge aufgebaut und mit Gauß'schem Rauschen gespeist. Dieses Rauschen wird mit dem Träger moduliert und als I- und Q-Komponente addiert. Dadurch, dass I und Q jeweils einem (weißen) Rauschprozess entstammen sind auch Betrag und Phase regellos und ebenfalls dargestellt. Gauß'sches Rauschen ist weiß, enthält auch beliebig hohe Frequenzkomponenten, die erst durch ein Tiefpassfilter am Eingang der Modulatoren effektiv und sinnvoll auf das interessierende Band begrenzt werden. Ein solcher Rayleighprozess, schon mit Tiefpassfiltern, die hier noch keine Rolle spielen, ist in Abb. 7.7 dargestellt. Betrachtet man die Fluktuationen und auch die Wahrscheinlichkeitsdichteverteilungen des Betrages und der Phase (ohne Tiefpass) dieses Prozesses, so gelangt man zu den Darstellungen nach Abb. 7.8. Im Endeffekt gehorcht die statistische Verteilung der (empfangbaren) Leistungen in einem Rayleigh-Kanal der Rayleighverteilung (eigentlich eine Weibullverteilung mit

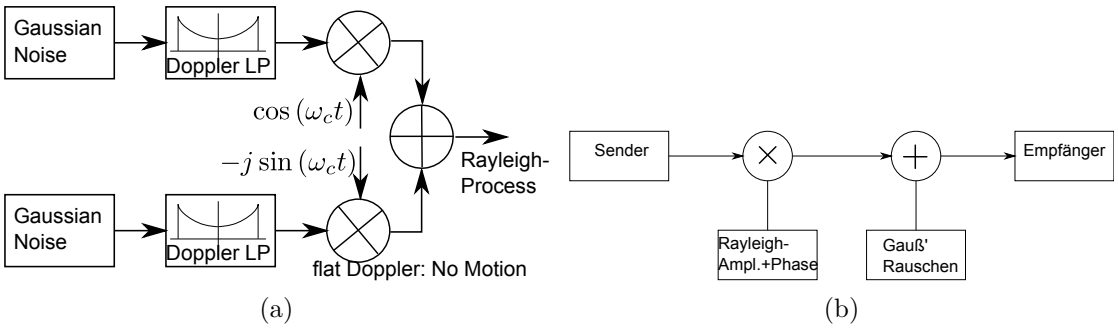


Abbildung 7.7: Darstellung eines Rayleigh-Prozesses zur Erzeugung der Fluktuationen (a) und Verwendung in der Simulation eines Kanals zusätzlich zum weißen Rauschen (b).

Form zwei und Exponent $\alpha = 2$) gemäß

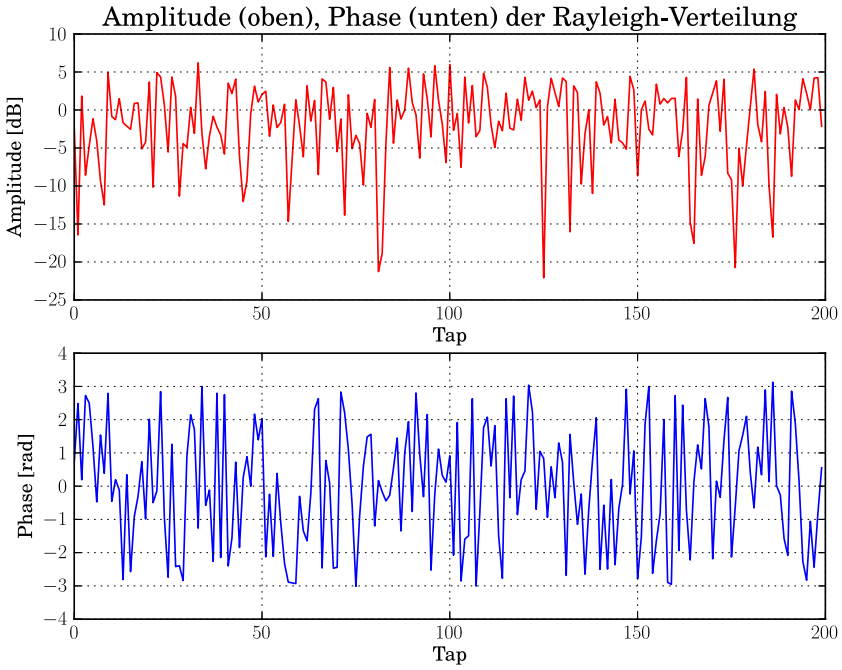
$$f(x|\sigma) = \frac{2xe^{-\frac{x^2}{\sigma^2}}}{\sigma^2} \tag{7.5}$$

und einer gleichverteilten Phase zwischen $-\pi ... \pi$. σ^2 ist die Leistung der kombinierten Gauß-Prozesse in der Erzeugung der Rayleigh-Verteilung. Ein Gauß-Prozess kommt also mit einer Standardabweichung $\sigma/\sqrt{2}$ daher.
Bestimmt ist hiermit aber nun nur die Wahrscheinlichkeit, mit der in einem Funkkanal eine bestimmt Dämpfung und Phasendrehung auftritt. Bleibt der Kanal konstant, die Nutzer bewegen sich nicht und alles andere ist auch immer gleich, dann hat man eine konstante Phasendrehung und konstante Dämpfung und das war es. Die Fluktuationen kommen erst dazu, wenn es sich um bewegte Ziele handelt. Bei ortsfestem Sender kann man sich vorstellen, dass sich der Empfänger entlang der x-Achse in Abb. 7.8(a) bewegt. Je nach seiner Geschwindigkeit ändern sich Amplitude und Phase des empfangenen Signals unterschiedlich schnell. Zusätzlich bedeuten Bewegung und Geschwindigkeit immer Frequenzverschiebung durch den Doppler-Effekt.
Bewegt sich eine Signalquelle (oder auch Senke) mit einer Geschwindigkeit v direkt auf die Gegenstelle zu, so sieht diese das ausgesandte Signal mit der Wellenlänge

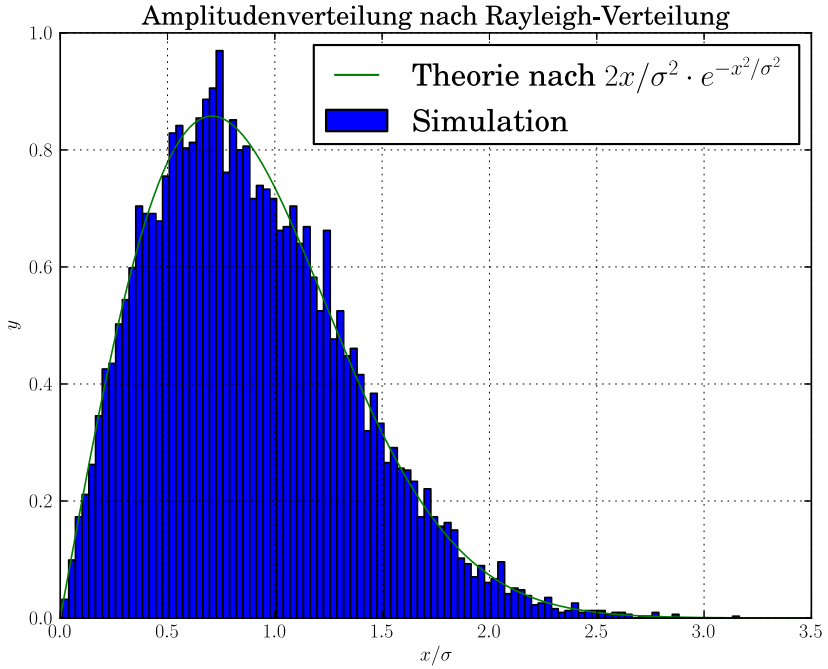
$$\lambda = \lambda_0 - vT \tag{7.6}$$

wobei $\lambda_0 = c_0/f_0 = c_0 T$ die Wellenlänge des Trägers mit Frequenz und Periodendauer $f_0 = 1/T_0$ im unbewegten Zustand ist und c_0 die Lichtgeschwindigkeit meint. Es folgt dann weiter

$$\frac{c_0}{f} = \frac{c_0}{f_0} - v\frac{1}{f_0} \Leftrightarrow \qquad f = \frac{f_0}{1 - \frac{v}{c_0}} \tag{7.7}$$



(a)



(b)

Abbildung 7.8: Ergebnis nach Betrag und Phase (a) und Wahrscheinlichkeitsdichte (b) eines rohen Rayleighprozesses.

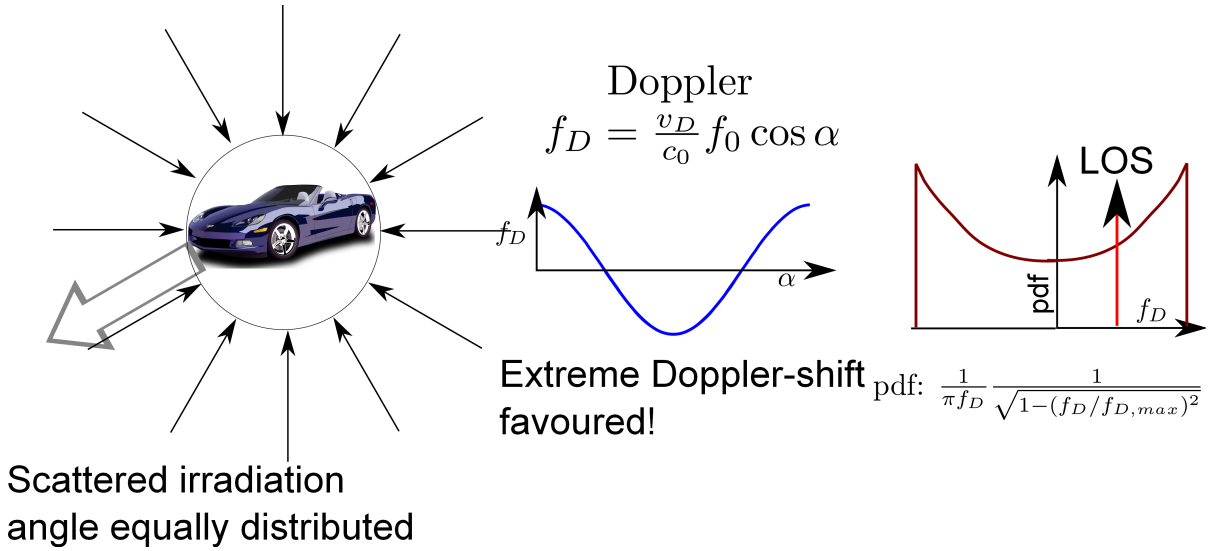


Abbildung 7.9: Zustandekommen der Spektralverteilung nach einer Dopplerverschiebung.

für die verschobene Frequenz und

$$f_D = f - f_0 = \frac{f_0}{1 - \frac{v}{c_0}} - f_0 \approx \frac{f_0 v}{c_0} \tag{7.8}$$

für die Dopplerverschiebung. Bei Quellen, die sich nicht direkt auf den Empfänger zu bewegen, interessiert nur die Geschwindigkeitskomponente in Richtung des Empfängers. Wenn sich die Bewegung im Winkel α relativ zur direkten Verbindung stattfindet, dann ist die Dopplerverschiebung

$$f_D = \frac{f_0 v}{c_0} \cos\alpha \tag{7.9}$$

wie auch in Abb. 7.9 gezeigt. Für einen GSM-Kanal bei $f_0 = 900\,\text{MHz}$ und einer Geschwindigkeit von $v = 200\,\text{km/h}$ ergibt sich eine Dopplerverschiebung von $f_D = 166{,}7\,\text{Hz}$. Physikalisch ist der Prozess für genau einen Fall nun klar, es fehlt noch die Berechnung der Auswirkung auf die Wahrscheinlichkeitsverteilungen. Hierzu sind einige stochastische Betrachtungen notwendig. Man kann davon ausgehen, dass die Richtungen α der Relativbewegungen gleichverteilt im Wertebereich $0 \ldots 2\pi$ sind. Damit ist die pdf von α

$$\text{pdf}(\alpha) = \frac{1}{2\pi} \tag{7.10}$$

und die kumulative Verteilungsfunktion ist

$$F(\alpha) = \begin{cases} 0 & \alpha < 0 \\ \frac{\alpha}{2\pi} & 0 \leq \alpha \leq 2\pi \\ 1 & 2\pi < \alpha \end{cases}. \tag{7.11}$$

Damit ist dann die kumulative Verteilungsfunktion für die Dopplerverschiebungen f_D

$$G(f_D) = \text{Prob}\,(f \leq f_D) = \text{Prob}\left(\frac{f_0 v}{c_0}\cos\alpha \leq f_D\right) \tag{7.12}$$

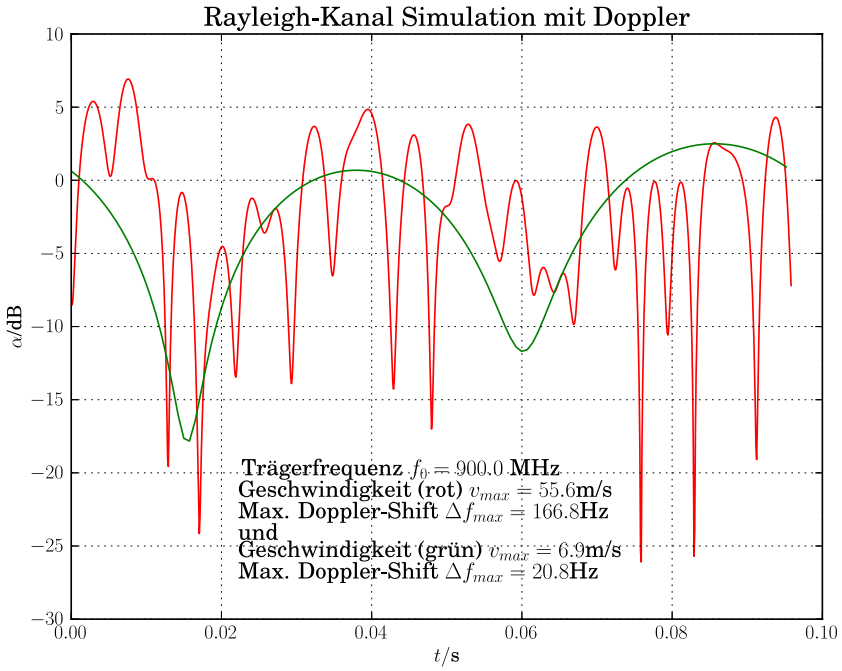
$$= \text{Prob}\left(\alpha < \arccos\left[\frac{f_D c_0}{f_0 v}\right]\right) \tag{7.13}$$

$$= \frac{1}{2\pi}\arccos\left(\frac{f_D c_0}{f_0 v}\right) = \frac{1}{2\pi}\arccos\left(\frac{f_D}{f_{D,max}}\right), \tag{7.14}$$

wobei als Abkürzung $f_{D,max} = \frac{v}{c_0} f_0$ als maximal auftretende Dopplerverschiebung verwendet wurde. Die pdf ist die Ableitung der kumulativen Funktion

$$g(f_D) = \frac{dG(f_D)}{df_D} = \frac{1}{2\pi f_{D,max}} \times \frac{1}{\sqrt{1-(f_D/f_{d,max})^2}}. \tag{7.15}$$

Diese Funktion ist in Abb. 7.9 ebenfalls dargestellt.
Ein Durchfahren des Rayleigh-Kanals ohne Doppler entspricht einem Tiefpassverhalten, es werden einfach nach Vorgabe der Änderung des Ortes Betrag und Phase des Kanals geändert. Mit dem o.g. Doppler-Effekt kommt es zu einer Überbetonung der schnellen Änderungen, weil ja die Frequenz leicht geändert wird. Für einen schnell bzw. einen langsam fahrenden Empfänger gebärdet sich der Rayleigh-Kanal wie in Abb. 7.10 dargestellt. Für eine schnelle Bewegung ändern sich die Amplitudenverhältnisse sehr schnell, es kann zu Einbrüchen im Empfangssignal kommen, die nicht mehr ausgeregelt werden können. Auch bei einem langsamen Beobachter gibt es tiefe Einschnitte, die jedoch sehr viel langsamer von statten gehen.
Natürlich ist klar, dass bei den Zeiten starker Dämpfung das Signal-Rauschverhältnis durch den nachgeschalteten AWGN-Kanal (Abb. 7.7(b)) sehr schlecht werden kann. Die Bitfehlerrate steigt hier sicherlich an. Die Länge eines Symbols (bits) im GSM ist bspw. 3,69 μs lang, ein ganzer Zeitschlitz mit 577 μs. Dieses ist sehr viel kürzer als die Dauer eines Einbruchs. Es ist davon auszugehen, dass, wenn Bitfehler auftreten, diese nicht isoliert sind, sondern in Büscheln auftreten. Kodierungsverfahren, die nur vereinzelte Fehler in einem Block korrigieren oder erkennen können, sind dementsprechend ungeeignet.



(a)

Abbildung 7.10: Fluktuationen im Rayleigh-Kanal mit Doppler für zwei verschiedenen Geschwindigkeiten.

7.2.3 Rice-Kanal

Mit dem Rayleigh-Kanal und der Freiraumausbreitung sind zwei extreme Ausbreitungsbedingungen gezeigt worden, beim einen gibt es gar keine direkte (Sicht)Verbindung, beim anderen gibt es gar keine äußeren Einflüsse und Störungen. In der Regel liegen Mischungen zwischen beiden vor. Wenn zu einem (nicht frequenzselektiven) Rayleigprozess noch eine direkte Komponente, die natürlich auch einer Dopplerverschiebung unterliegen kann, hinzukommt, dann liegt ein Rice-Kanal vor. Dieser konstruiert sich wie in Abb. 7.11 gezeigt und liefert die dort auch gezeigten entsprechenden Amplitudenschwankungen. Diese sind umso stärker, je geringer die direkte Komponente ausfällt. Die Wahrscheinlichkeitsdichteverteilug ist ebenfalls gegenüber dem ursprünglichen Rayleigh-Prozess eingeengt. Vor allem ist die Phase nicht mehr gleichförmig verteilt, sondern konzentriert sich auch um die Nullphase herum, bzw. um die durch die Dopplerverschiebung und deren Frequenz vorgegebene Phasendrehung. Auf eine analytische, bzw. statistisch berechnende Darstellung wird hier verzichtet.

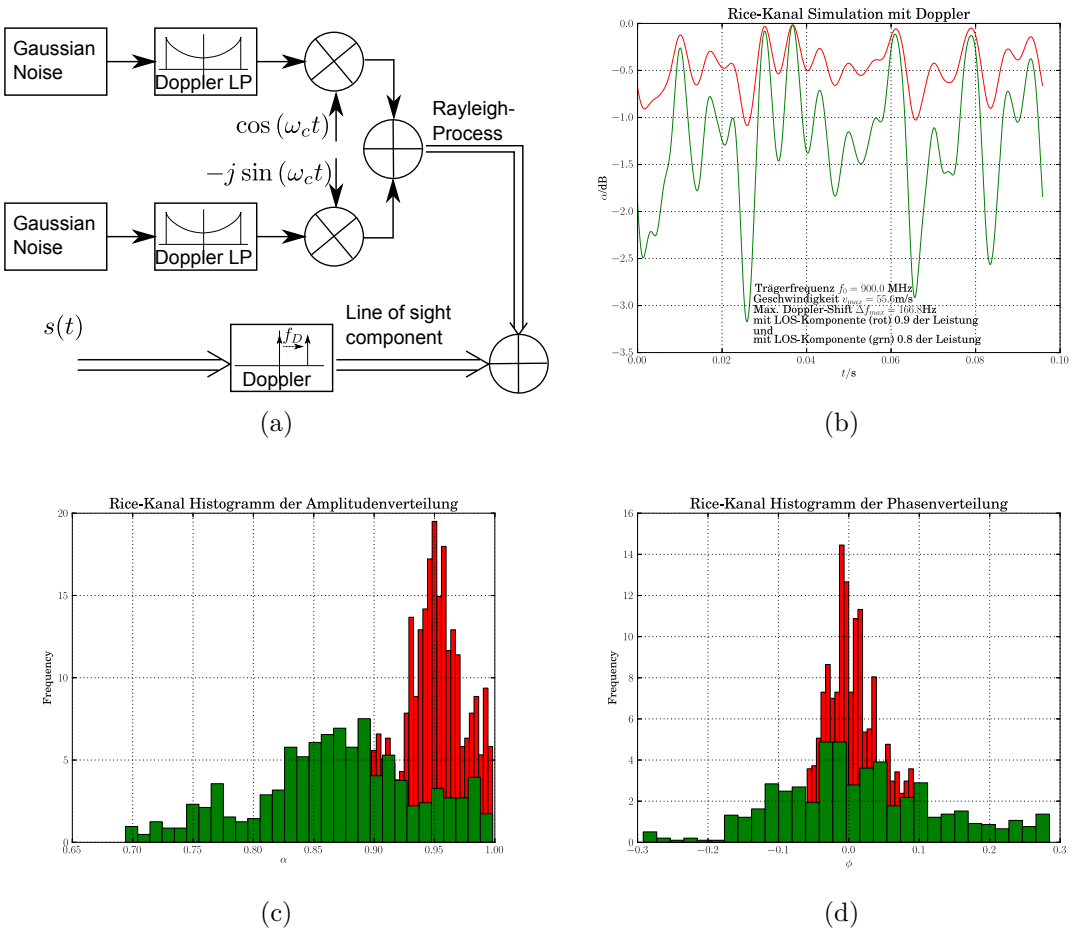


Abbildung 7.11: Schema eines Rice-Kanal (a) und die resultierende Amplitudenfluktuation (b). In (c) und (d) sind die Verteilungen der Amplitude (d) und der Phase (c) für die nach (b) berechneten zwei Fälle für einen stationären direkten Kanal gezeigt.

7.2.4 Mobilfunkkanäle mit Symbol-Interferenzen

Die bislang vorgestellten Rayleigh- und Rice-Kanäle, wie auch der ungestörte Kanal (AWGN) sowieso, sind nicht frequenzselektiv und die einzelnen Streukomponenten liegen zeitlich bzw. örtlich so dicht beieinander, dass sie sich auf wenige Perioden der Trägerkomponente erstrecken. Die natürlich sehr viel längeren (weil langsamer verlaufenden) Signalkomponenten bleiben ungestört. Bei einer (GSM)-Bitdauer bedeutet dieses, dass die maximalen Umwege sehr viel kürzer als 3,69 μs laufen müssen, dieses entspricht einer Strecke von ca. 1 km. Es sei betont, dass es sich nicht um die Entfernung zwischen Sender und Empfänger, sondern um den Umweg handelt. In dichten städtischen Gebieten kann

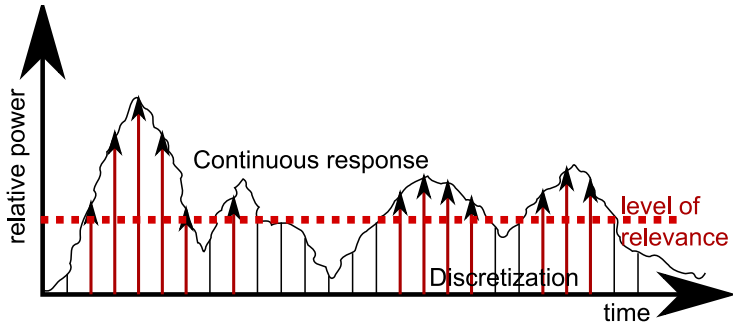


Abbildung 7.12: Darstellung der Impulsantwort in einem Kanal mit Mehrwegeausbreitung. Zur Simulation werden nur relevante Anteile verwendet.

diese Abschätzung sicherlich als gültig angenommen werden. In hügeligem oder gar bergigem Gelände kann es allerdings zu Umwegen kommen, deren Verzögerungen gegenüber dem direkten Pfad in die Größenordnung oder sicher auch deutlich über den o.g. Wert liegen können. In dem Fall werden verzögerte Signalanteile eines Symbols mit den schnellen Signalanteile des folgenden Symbols interferieren. Die Impulsantwort eines solchen Kanals besteht daher aus vielen verschiedenen verzögerten Komponenten

$$\tilde{h}(t) = \left\{ \sum_{n=0}^{N} r_n e^{\mathrm{j}\phi_n} \delta(t - \tau_n) \right\} \tag{7.16}$$

mit r,ϕ,τ den Amplituden, Phasen und Verzögerungsparametern einzelner direkter, Rayleigh- oder Rice-Prozesse. Eine graphische Darstellung liegt in Abb. 7.12 vor.
Die Beschreibung/ Simulation dieses Phänomens ist kritisch und erfolgt in der Regel durch die Annahme mehrerer direkter, Rayleigh oder Rice-Prozesse. In Abb. 7.13 ist das Schema eines solchen Prozesses dargestellt. Diese Kanäle sind nun auch frequenzselektiv, da in der Tat viele verschiedenen Verzögerungselemente zusammenwirken.
Beispielhaft sind in Tab. 7.1 einige im UMTS verwendete Performance-Kanäle und deren Parameter aufgeführt. Verschiedene Szenarien werden untersucht, vom langsamen Fußgänger mit wenig Verzögerungsunterschieden bis hin zum schnell bewegten Empfänger in einer Umgebung mit großer Verzögerungsunterschieden.
Die Empfänger sind auf die vorzufindenden Kanäle zu optimieren, dieses ist eine der Kernaufgaben im Mobilfunk.

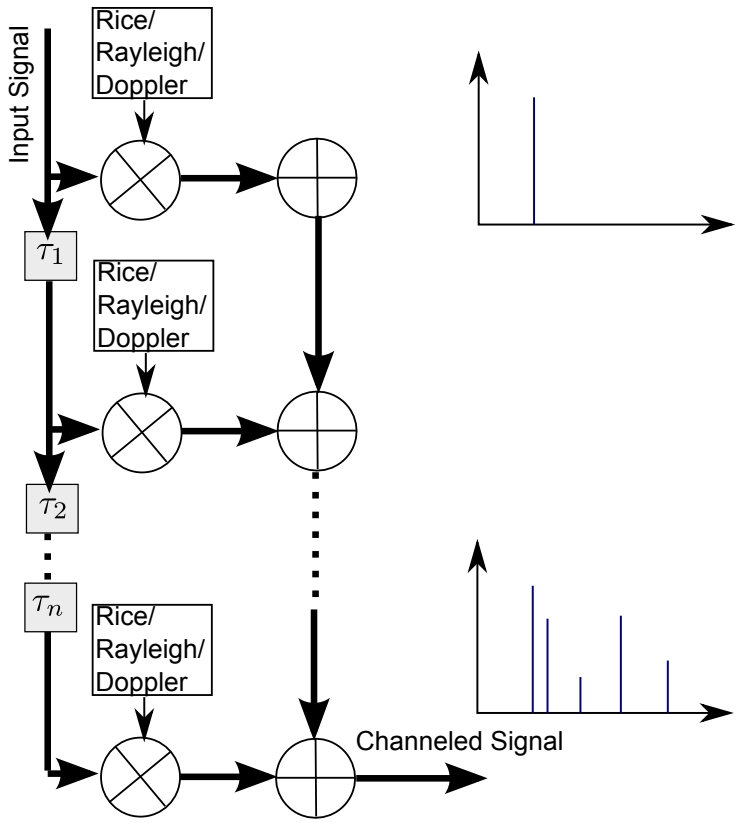


Abbildung 7.13: Schematische Darstellung einer Kanalsimulation mit vielen Verzögerungselementen.

Tabelle 7.1: Ausbreitungsbedingungen für Mehrwegeausbreitungen für die E-DPDCH und E-DPCCH Performance Anforderungen mit Delay= Relative Delay, Power = relative Mean Power

ITU Pedestrian A Speed 3km/h (PA3)		ITU Pedestrian B Speed 3km/h (PB3)		ITU vehicular A Speed 30km/h (VA30)		ITU vehicular A Speed 120km/h (VA120)	
Delay ns	Power dB	Delay ns	Power dB	Delay ns	Power dB	Delay ns	Power dB
0	0	0	0	0	0	0	0
110	-9.7	200	-0.9	310	-1.0	310	-1.0
190	-19.2	800	-4.9	710	-9.0	710	-9.0
410	-22.8	1200	-8.0	1090	-10.0	1090	-10.0
		2300	-7.8	1730	-15.0	1730	-15.0
		3700	-23.9	2510	-20.0	2510	-20.0

Kapitel 8

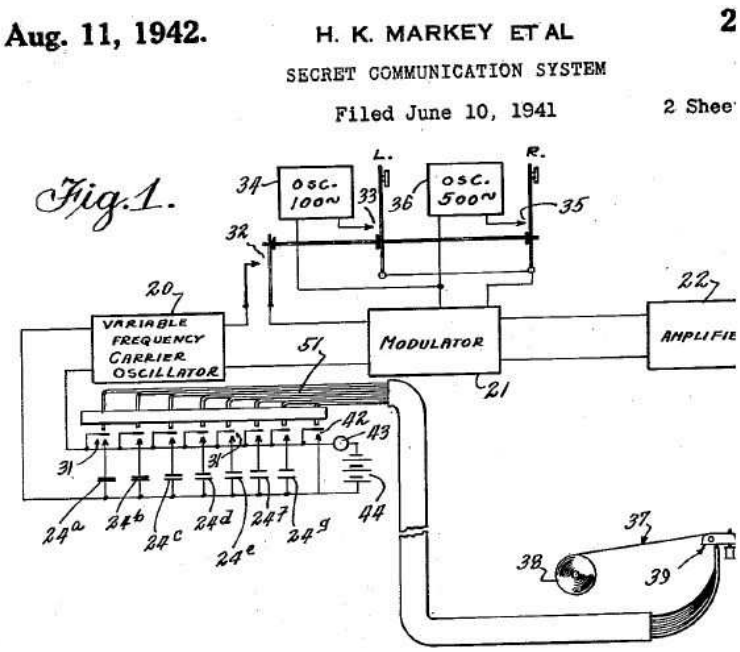
Anpassung an die Kanäle: Spreizung des Spektrums

In den vorangegangenen Kapiteln wurden die speziellen Eigenschaften der Mobilfunkkanäle aufgezeigt:

- Große Signalunterschiede zwischen Nutzern nahe an einer Basisstation und solchen weit entfernt davon,
- Frequenz- und Phasenverwerfungen und Änderungen durch bewegte Nutzer und ändernde Kanaleigenschaften (Doppler)
- Frequenzselektiv: Einzelne, schmale Frequenzbereiche unterliegen einer starken Dämpfung (Zwei-Wege-Kanal),
- Zeitvariant mit "Deep Fades": Zu bestimmten, ggf. kurzen Zeiten kann es zu starken Dämpfungen im Ausbreitungspfad kommen (Rayleigh-Fading),
- Inter-Symbol-Interferenz: Durch starke Umwege und Reflexionen im Übertragungskanal können Reste eines Symbols in das folgende hineinlaufen,
- Zugriff vieler Nutzer auf das gleiche Medium

In diesem Kapitel wird es um die Anpassung der Signale an diese oben genannten Eigenschaften gehen.

(a)



(b)

Abbildung 8.1: Hedy Lamarr (a) und die Zeichung eines FHSS Senders aus ihrem Patent (b) (Quelle: Wikipedia, Patentamt).

8.1 Frequenz-Sprung-Verfahren (Frequency Hopping Spread Spectrum)

Das einfachste Verfahren, einem frequenzselektiven Kanal wenigstens die meiste Zeit aus dem Weg zu gehen, ist es, die Frequenz schnell und oft zu ändern. Schon beim herkömmlichen Analog-Funk vereinbart man einen Frequenzwechsel, wenn man auf einem Kanal schlechte Ausbreitungseigenschaften vorfindet. Die Digitaltechnik, in der Daten ohnehin nur in Bursts versandt werden und die schnelle und automatische Steuerung von Sendern und Empfängern ermöglichen es, dieses Verfahren zu kultivieren. Es ist bekannt als Frequency Hopping Spread Spectrum (FHSS) und wurde erfunden/ patentiert unter Nummer und Titel “US2292387 (A) - Secret communication system” von der Schauspielerin Hedy Lamarr (Abb. 8.1). Nach einer gewissen Zeit wechseln Sender und Empfänger einfach den Kanal und machen mit der Übertragung dort weiter. Das bekannteste Funksystem, das dieses Verfahren nutzt ist das berühmte Bluetooth, bei dem im Frequenzbereich von 2402 MHz bis 2480 MHz insgesamt 79 1 MHz breite Kanäle zur Verfügung stehen, die in einer Pseudo-zufälligen Abfolge durchlaufen werden (Abb. 8.2) .

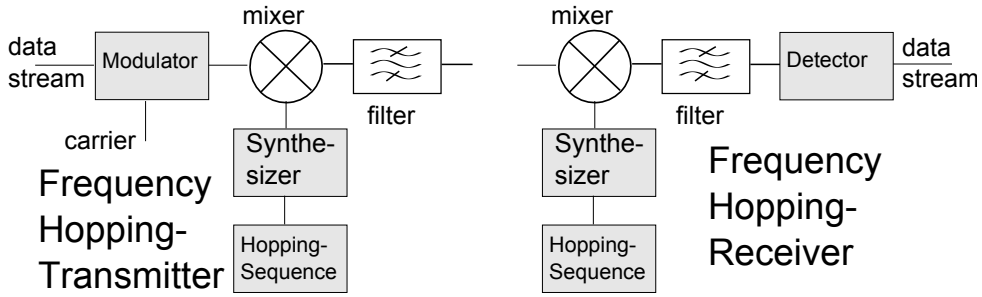


Abbildung 8.2: Sender und Empfänger für FHSS.

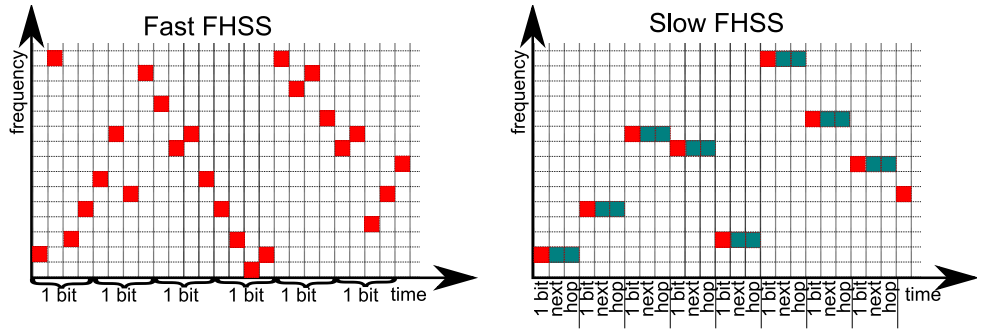


Abbildung 8.3: Schnelles und langsames FHSS

Grundsätzlich unterscheidet man zwischen einer Änderung des Kanals schneller als ein Bit und langsamer als eine Bitdauer in schnelles (fast) und langsames (slow) FHSS (Abb. 8.3). Genau genommen ist nur das schnelle FHSS wirklich eine Aufspreizung des Spektrums. Diese Variante wird in der kommerziellen Kommunikationstechnik nicht angewandt. Alle Systeme arbeiten nach einem Verfahren, in dem komplette Bursts (Bluetooth 625 μs=625 Bitdauern) auf einem Kanal ausgesandt werden, um dann die Frequenz zu wechseln. Wenn nun ein Burst wegen eines schlechten Kanals oder einer Störung nicht fehlerfrei übertragen werden konnte, wird er verworfen und - bei Datenkommunikation - ggf. neu angefordert. Dieses Verfahren hat sich als einfach und sehr robust erwiesen. Auf folgende Parameter ist zu achten:

- Für den Zugriff, also wenn ein Teilnehmer sich bei einem anderen zur Kommunikation anmelden will, müssen spezielle - kurze - Sequenzen oder Kanäle zur Verfügung stehen, so dass beide Teilnehmer sich in möglichst kurzer Zeit treffen können.
- Wenn viele Teilnehmer gleichzeitig kommunizieren wollen, müssen Hop-Sequenzen zur Verfügung stehen, die möglichst wenig Überdeckung haben können, so dass es sehr unwahrscheinlich ist, dass - nach einer Überschneidung - sich zwei Kommunikationsstrecken nochmals nach kurzer Zeit wieder treffen.

Letzteres lässt sich umgehen, indem man den unterschiedlichen Partnern jeweils PN-Sequenzen zum Hopping zuordnet.

8.2 Direct Sequence Spread Spectrum und Code Division Multiple Access

Im letzten Kapitel wurde dargestellt, wie man - jedenfalls meistens - einem schlechten Kanal aus dem Wege geht. Hier wird nun dargestellt, wie man eine Übertragung so gestaltet, dass ein (kleiner) Einbruch an einem Frequenzbereich des Kanals keine Rolle spielt. Die hier beschriebene Verfahren werden in modernen Kommunikationssysteme wie z.B. UMTS (Wideband Code Division Multiple Access (WCDMA)), ZigBee bzw. dem Standard IEEE 802.15.4, und einigen Varianten des WLAN (IEEE 802.11) verwendet. Die grundsätzliche Vorstellung und Argumentation für eine Aufspreizung des Spektrums ist in Abb.8.4 dargelegt. Man geht von einer schmalbandigen Störung des Kanals aus. Das kann ein Signal genauso sein, wie ein Fading, was letztlich dann als Substraktion eines Störers betrachtet werden kann. Ein Einträger-Signal würde, wenn es mit der Störung zusammenfällt, massiv gestört werden, eine Übertragung wird unmöglich. Wird aber das Signal - bei gleichem Informationsgehalt - über einen weiteren Frequenzbereich gespreizt und dann auf den Kanal geschickt, dann erscheint am Empfänger dieses breite Signal überlagert mit dem immer noch schmalbandigen Störer. Im Empfänger findet nun ein Entspreizen statt. Dieses führt effektiv dazu, dass das Nutzsignal rekonstruiert wird, der Störer aber wird nun über das ganze Band verteilt und bleibt so als quasi zusätzliches Rauschen ohne großen Einfluss auf die Übertragung. Grundsätzlich entspricht der Gewinn dieses Prozesses der Aufspreizung des Störers:

$$G \approx \frac{B_c}{B_m} = \frac{R_{ch}}{R_S} \tag{8.1}$$

mit B_c, B_m den relevanten Bandbreiten des gespreizten und ungespreizten Signals, sowie R_{ch}, R_S den Raten der Chips (also gespreizt) und dem Signal.
Nachdem nun der Gewinn klar ist, ist nun zu betrachten, wie man technisch diese Prozesse Spreizung und Entspreizung ausführt. Der Vorgang der Ratenerhöhung ist einfach, anstatt eines Nutzsymbols sendet man viele wechselnde Pegel (sogenannte Chips) (Abb. 8.5). Dieses kann rein reell oder auch bei komplexer Modulation komplex erfolgen. Der eigentliche Spreizprozess in Abb. 8.5 ist die Multiplikation mit $c(t)$ nach (ggf. auch vor) der Modulation mit dem Träger $i(t)$ entscheidend ist, dass sich die Chips $c(t)$ um ein Vielfaches schneller ändern als die Nutzdaten $d(t)$. Das Verhältnis zwischen beiden ist der

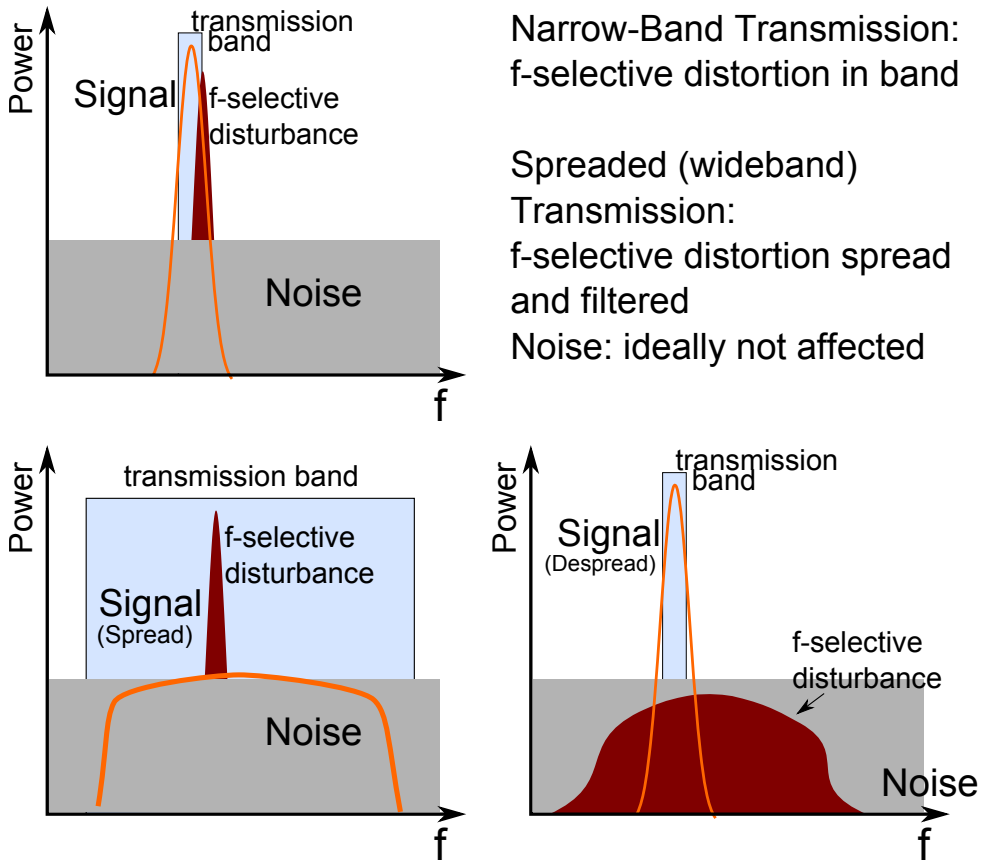


Abbildung 8.4: Darstellung der Vorteile von Spektraler Spreizung bei schmalbandiger Störung.

Spreizfaktor, der z.B. bei UMTS bis zu 512 betragen kann.
Folgende Anforderungen sind an eine Spreizsequenz zu stellen:

- Es muss eine wirkliche Aufspreizung des Spektrum stattfinden. Während die Sequenzen $c(t) = \{1, 1, 1, 1, ...\}$ und $c(t) = \{1, -1, 1, -1, 1, ...\}$ nur eine Erhöhung der Rate, bzw. einfach einen Frequenzversatz bewirken würden, kann eine PN-Sequenz, die ein flaches, breites Spektrum besitzt, tatsächlich die Nutzdaten gleichmäßig über einen weiteren Frequenzbereich verteilen.

- Es soll für verschiedene Anwendungen unterschiedliche Sequenzen zur Spreizung geben, die möglichst wenig miteinander gemein haben, deren Korrelation gering ist.

Besonders gut werden diese Eigenschaften von PN-Sequenzen erfüllt.

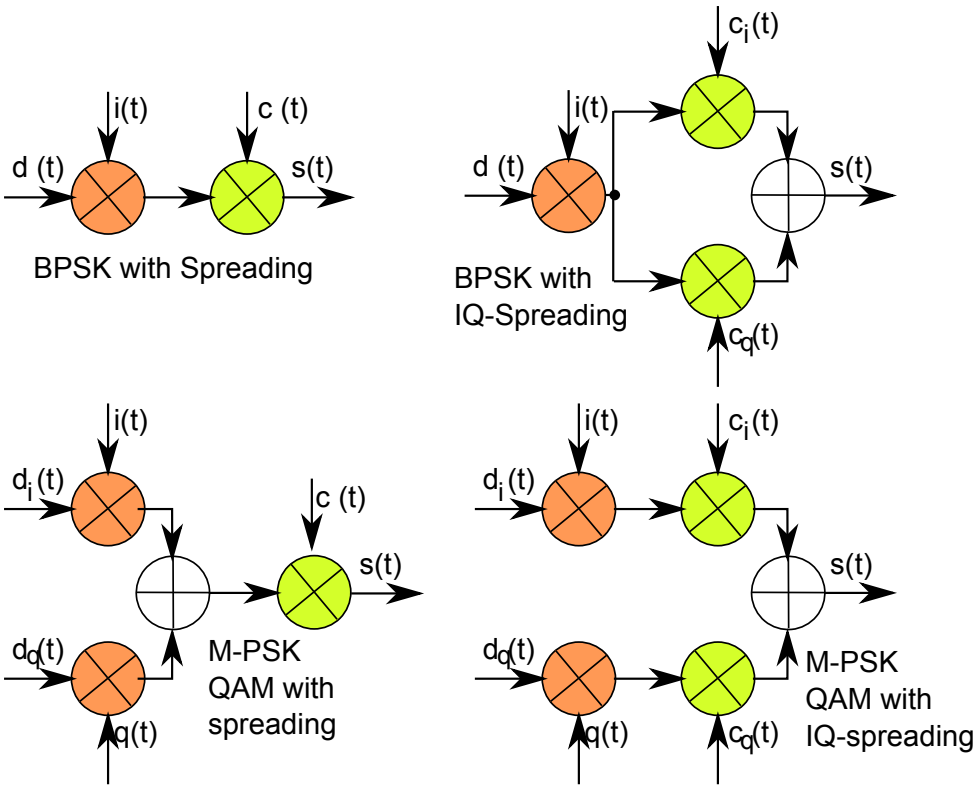


Abbildung 8.5: Verschiedene Arten der Aufspreizung des Spektrums.

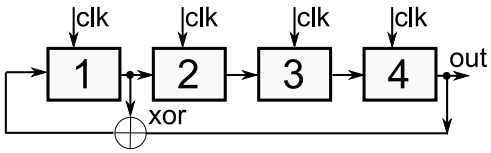


Abbildung 8.6: Schieberegister mit Rückkopplungen.

8.2.1 PN-Sequenzen

PN-Sequenzen oder auch Pseudo Random Bit Sequence (PRBS) sind solche deterministischen Sequenzen, deren Spektrum und Aussehen sich so verhält, als wäre es eine zufällige Folge von Null (Minus Eins) und Eins. Man erzeugt die Sequenzen durch rückgekoppelte Schieberegister (Abb. 8.6). Nach einer Ladung (Seed) des Registers mit mindestens einer Eins liefert das gezeigte Beispiel die Sequenz 000111101011001 der Länge $2^m - 1 = 15$ (m Anzahl der Register) bevor sie sich wiederholt. Der Sequenz ist anzusehen, dass dort schnelle und langsame Wechsel vorkommen, das Spektrum also wohl flach sein wird. Eine Implementation ist im Anhang A.2.1 gegeben.

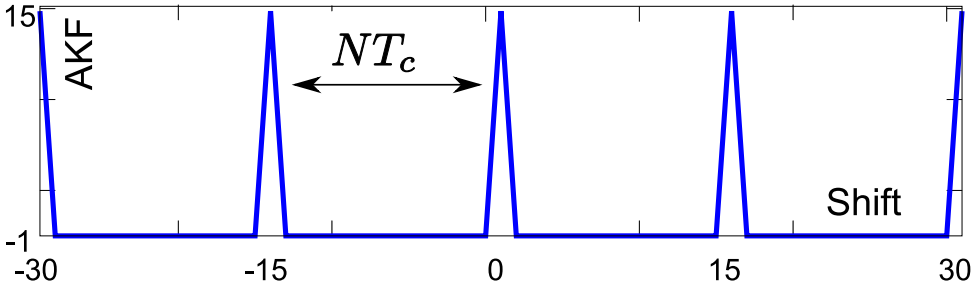


Abbildung 8.7: Autokorrelationsfunktion (AKF) einer PN-Sequenz.

Wichtig sind die Eigenschaften dieser Seqeunzen mit der Autokorrelationsfunktion

$$R_{cc}(\tau) = \frac{1}{T}\int_0^T c(t)c(t+\tau)dt \tag{8.2}$$

mit T der kompletten Periodendauer der ganzen Sequenz. Von Hand berechnet sich diese durch

Verschiebung	Sequenz	Gleichheiten	Ungleichheiten	AKF
	000111101011001			
0	000111101011001	15	0	15
1	001111010110010	7	8	-1
2	011110101100100	7	8	-1
...	111101011001000	7	8	-1

und ist in Abb. 8.7 dargestellt. Die Autokorrelationsfunktion (AKF) ist wichtig, wenn mehrere Nutzer auf der gleichen Frequenz, aber mit unterschiedlichen Daten kommunizieren sollen. Umgekehrt stellt die hohe Korrelation bei exakter Überlagerung die Demodulations(Entspreizung)vorschrift dar: Multiplikation der empfangenen Chipfolge mit der PN-Sequenz und Mittelwertbildung liefert bis auf ein Vielfaches das Nutzbit zurück.
Die Länge einer PN-Sequenz bevor sie sich wiederholt ist maximal $2^m - 1$ bei m Schieberegistern. Dieses liegt daran, dass mit m Registern, also m Bits 2^m verschiedene Zustände dargestellt werden können, aber der Null-Zustand nicht erlaubt ist, da er immer auf Null führen muss. Gleichzeitig folgt aus dieser Überlegung auch, dass alle Zustände, also bspw. auch die mit m Einsen, auftreten müssen. Die Sequenz wird eindeutig durch die Rückkoppelpfade vorgegeben, nicht alle möglichen Kombinationen von Rückkoppelungen führen auf die maximale Länge. Der Start wird durch die Vorladung der Register definiert und bestimmt letztlich nur eine Verschiebung, da er immer irgendeinen gültigen und notwendigen Registerzustand treffen muss.

Tabelle 8.1: Anzahl von möglichen maximal langen PN-Seqeunzen und Beispiel für die Tabs.

m	$\{i_r\}$	No.	m	$\{i_r\}$	No.	m	$\{i_r\}$	No.
2	1,2	1	10	2,10	60	18	7,18	8064
3	1,3	2	11	2,11	176	19	1,2,5,19	27594
4	1,4	2	12	1,4,6,12	144	20	3,20	24000
5	2,5	6	13	1,3,4,13	630	21	2,21	84672
6	1,6	6	14	1,6,10,14	756	22	1,22	120032
7	1,7	18	15	1,15	1800	23	5,23	356960
8	2,3,4,8	16	16	1,3,11,12,16	2048	24	1,2,7,24	276480
9	4,9	48	17	3,17	7710			

In Tab. 8.1 sind einige maximal lange PN-Sequenzen aufgeführt. Für jede Länge ist jeweils ein Beispiel von Rückkopplungstabs gegeben, das auf die maximale Länge der Sequenz führt.

8.2.1.1 Kreuzkorrelation

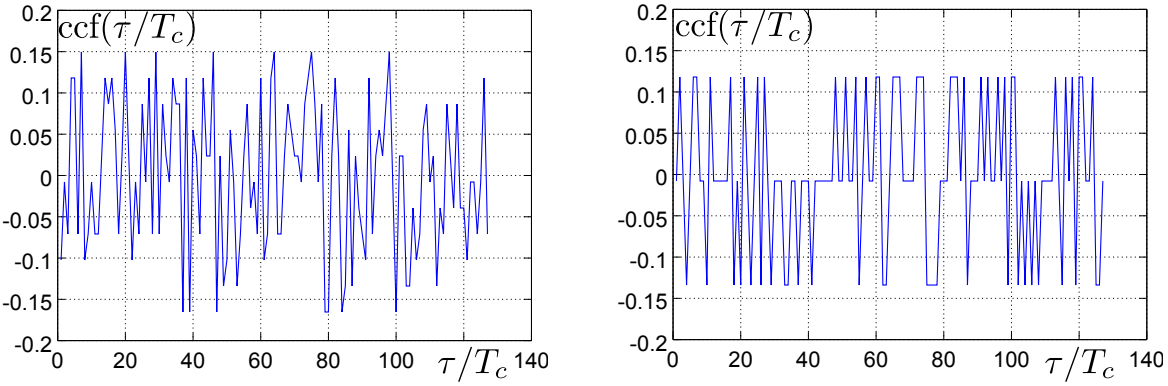
Unterschiedliche Signalquellen sind dadurch zu trennen, dass man ihnen vollständig unterschiedliche Sequenzen zuweist. Diese entstehen durch Verwendung verschiedener Rückkoppel-Schaltungen [44], die in Tabelle 8.2 für einige Varianten ausgeführt sind. Es kann so eine Kreuzkorrelation berechnet werden, bei der zwei Unterschiedliche PN-Sequenzen in verschiedenen Verschiebungen aneinander multipliziert und gemittelt werden:

$$R_{ij}(\tau) = \frac{1}{T}\int_0^T c_i(t)c_j(t+\tau)dt. \tag{8.3}$$

Die gesamte Periodendauer der Sequenz ist hier $T = L_m \times T_c$ mit der Dauer T_c eines Chips und der Gesamtlänge der Sequenz $L_m = 2^m - 1$. Hier zeigt sich, dass die Kreuzkorrelationseigenschaften dieser Sequenzen nicht optimal sind, also immer bzw. oft eine Korrelation $\neq 0$ nach Abb. 8.8 verbleibt. Im Vergleich dazu zeigt die AKF eben nur Werte von 1 oder $1/L_m$. Diese recht hohen Werte machen es schwierig in einem Empfänger die Signale unterschiedlicher Sender sauber zu trennen.

8.2.1.2 Gold-Sequenzen

Durch eine verbesserte Konstruktionsvorschrift können die Eigenschaften von PN-Sequenzen behalten werden und gleichzeitig erheblich mehr und systematischer verschiedene Codes



(a) (b)

Abbildung 8.8: Kreuzkorrelationsfunktion der PN-Sequenzen mit $n = 7$ und Rückkopplung an [6,7] und [1,7] (a) und an [3,7] und [1,2,3,7] (b).

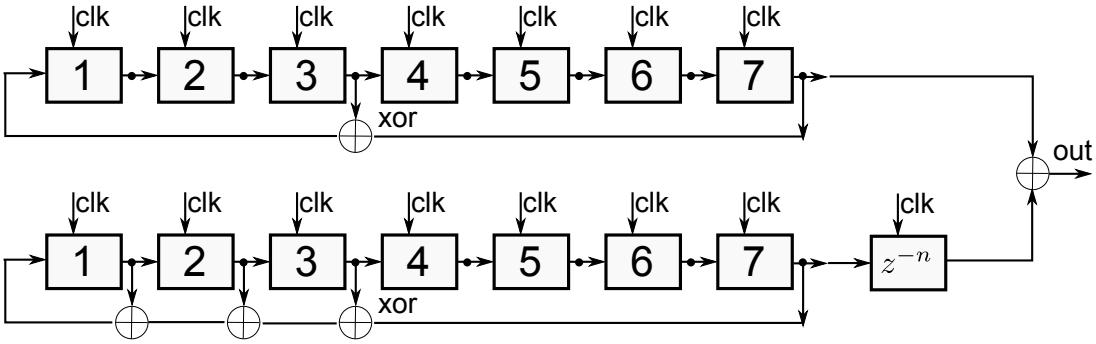
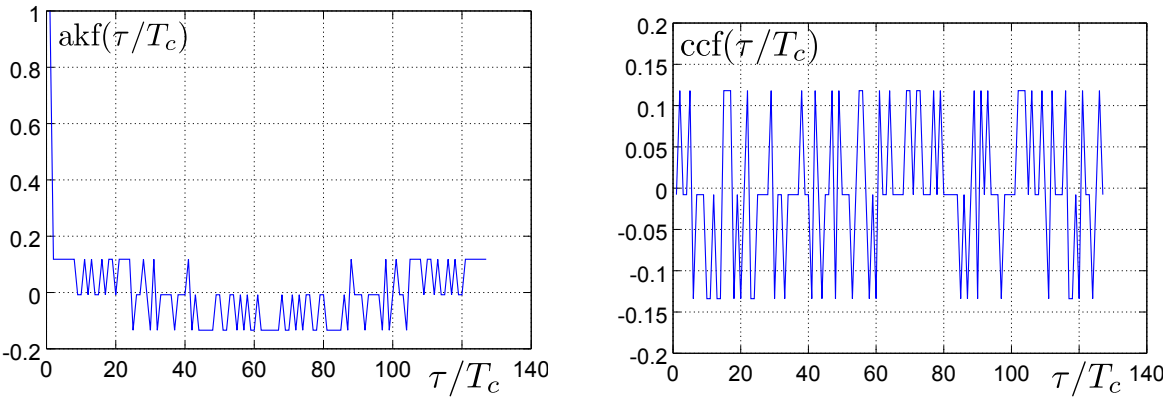


Abbildung 8.9: Erzeugungsstruktur für einen-Gold-Code mit $n = 7$ und Rückkopplung an [3,7] und [1,2,3,7]. Das Element z^{-n} meint nur eine wählbare Verzögerung.

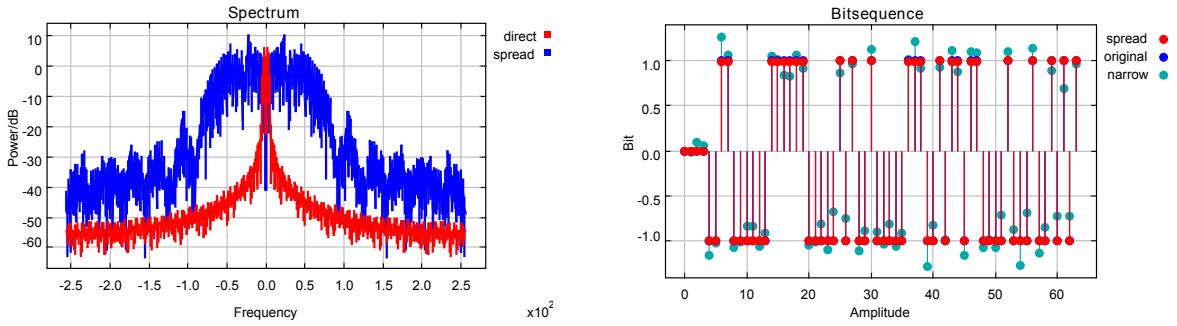


(a) (b)

Abbildung 8.10: Auto- (a) Kreuzkorrelationsfunktion (b) der Gold-Codes $n = 7$ und Rückkopplung an [3,7] und [1,2,3,7] (KKF mit Verschiebung von Code zwei um eins).

Tabelle 8.2: Rückkoppelschaltungen für Bitsequenzen maximaler Länge (unter dem ersten Strich die Spiegelsequenzen aus $[A,B,C,m]$ mit $[m-C,m-B,m-A,m]$).

m	2-Taps	4-Taps	6-Taps
3	2,3		
	1,3		
4	3,4		
	1,4		
5	3,5	2,3,4,5	
		1,3,4,5	
	2,5	1,2,3,5	
		1,2,4,5	
6	5,6	1,4,5,6	
		2,3,5,6	
	1,6	1,2,5,6	
		1,3,5,6	
7	6,7	4,5,6,7	2,3,4,5,6,7
	4,7	2,5,6,7	1,2,4,5,6,7
		1,4,6,7	
		3,4,5,7	
	1,7	1,2,3,7	1,2,3,4,5,7
	3,7	1,2,5,7	1,2,3,5,6,7
		1,3,6,7	
		2,3,6,7	
8		1,6,7,8	2,4,5,6,7,8
		3,5,7,8	1,2,5,6,7,8
		2,3,7,8	
		4,5,6,8	
		2,5,6,8	1,2,3,4,6,8
		1,2,7,8	1,2,3,6,7,8
		1,3,5,8	
		1,5,6,8	
		2,3,4,8	
		2,3,6,8	



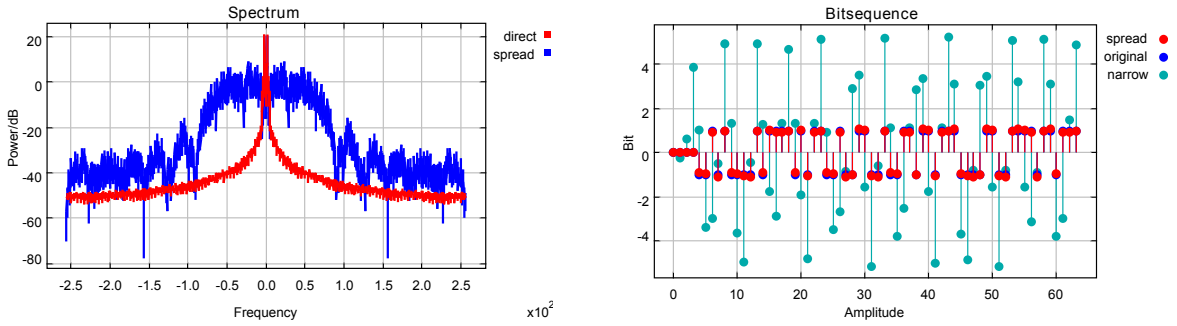
(a) (b)

Abbildung 8.11: Spreizen des Spektrums um den Faktor 15 Spektrale Darstellung (a) und Darstellung der Übertragung (b). Alle Signale unterliegen einer Impulsformung mit einem RRC-Filter mit $\alpha = 0.22$. [es ist mir unklar, warum der direkte Pfad nicht wieder auf 1 fällt, ggf. sind hier die Filterungen noch nicht optimiert].

mit guten Kreuzkorrelationseigenschaften gewonnen werden. Diese wurde 1967 von Robert Gold gefunden. Die Konstruktion erfolgt so, dass zwei "normale" PN-Sequenzen, sogenannte Muttercodes, deren Betrag der Kreuzkorrelationsfunktion (KKF) unter $2^{(n+2)/2}$ bleibt, in verschobenen Versionen miteinander Modulo-2 addiert werden (Abb. 8.9). Daraus entstehen nun $2^n - 1$ Sequenzen (also so viele Sequenzen, wie die ursprünglichen PN-Folgen lang sind), die ebenso gute KKF-Eigenschaften haben, deren KKF nämlich ebenso unter $2^{(n+2)/2} - 1$ für gerade und $2^{(n+1)/2} - 1$ für ungerade n bleibt (Abb. 8.10). Nachteil ist, dass bei diesen Codes die KKF-Eigenschaften sich erheblich verschlechtern, wenn zwei nicht ganzzahlig gegeneinander verschobene Codes aufeinander treffen. Die Verwendung von Gold-Codes erfordert also eine hohe Synchronität und wird daher vorzugsweise auf dem Weg von einem zentralen Sender (Basistation) zum Endgerät und nicht umgekehrt verwandt.

8.2.2 Spread-Spectrum-Modulation

Die Spreizung des Signals erfolgt wie schon in Abb. 8.5 gezeigt. In Abb. 8.11 ist eine Simulation der Spreiztechnik vorgenommen worden. Die Aufweitung des doppelseitigen Spektrums um den gegebenen Faktor ist deutlich zu erkennen. Der Vorteil dieser Technik wird deutlich, wenn – wie in Abb. 8.12 – ein schmalbandiger Störer (hier ein Sinussignal im unteren Frequenzbereich) hinzukommt. Während die Signale, die über den gespreizten Pfad geschickt wurden nahezu ungestört empfangen werden können, ist der direkte Pfad nicht mehr zu gebrauchen.



(a) (b)

Abbildung 8.12: Spreizen des Spektrums um den Faktor 15 Spektrale Darstellung (a) und Darstellung der Übertragung (b). Mit einem schmalbandigen Störer bei $f = 0{,}2$ mit Amplitude $a = 0.5$.

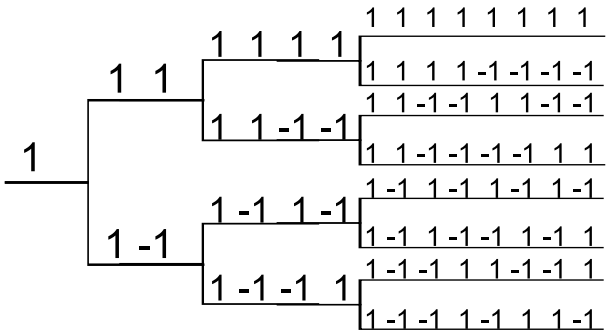


Abbildung 8.13: Baumaufbau der Walsh-Matrizen.

Die simple Aufspreizung der Signale auf einen größeren Frequenzbereich schafft eine erhöhte Resistenz gegen Störungen und Fading um den Preis der Verschwendung von Bandbreite. Tatsächlich wird dieses bei Kurzstreckenfunk im Industry, Scientific, Medicine (ISM)-Band in Kauf genommen. So wird bei WLAN nach dem Standard IEEE 802.11 für die Übertragungsraten von 5,5 bzw. 11 Mbit/s eine Spreizung um den Faktor 11 mit folgender Barker-Sequenz vorgenommen: $c = \{+1, -1, +1, +1, -1, +1, +1, +1, -1, -1, -1\}$. Im 802.15.4 (ZigBee) (900 Mhz-PHY) im wird mit der bipolarisierten Sequenz $c = \{111101011001000\}$ um den Faktor 15 gespreizt. Im UMTS wird jeder Basis- und Mobilstation ein individueller Spreizcode (Scrambling-Code) aus der Serie der Gold-Codes zugewiesen. Wie unten gezeigt wird, ist das nicht die einzige Spreizung in diesem Mobilfunkstandard.

8.2.3 Mehrfachausnutzung des gespreizten Kanals

Die o.g. Verschwendung von Bandbreite ist bspw. in den Mobilfunkstandards des UMTS nicht hinnehmbar, daher geht man den Weg, dass man mehreren Teilnehmern das gleiche Frequenzband zuweist und die Kanäle anhand spezieller orthogonaler Codes unterscheidet. Die Überlegung ist einfach, dass durch spektrale Aufspreizung Redundanz zugefügt wurde, die ausgenutzt werden muss. Hierzu verwendet man Walsh-Codes, die rekursiv nach

$$w(2^0) = 1; \tag{8.4}$$

$$w(2^n) = \begin{pmatrix} w(2^{n-1}) & w(2^{n-1}) \\ w(2^{n-1}) & -w(2^{n-1}) \end{pmatrix} \tag{8.5}$$

gebildet werden. Eine einfache Octave-Implementation ist

```
function w=walsh_codes(n)
  w0=1;
  for i = 1:n
    w=[w0 w0; w0 -w0];
    w0=w;
  end
```

oder als Baum nach Abb. 8.13. Ausgerechnet ergibt sich eine Matrix, wie sie für die Länge 16 (Ordnung vier) in Abb. 8.14 dargestellt ist.
Die wichtigste Eigenschaft ist, dass bei perfekter Synchronisierung die einzelnen Zeilen orthogonal zueinander sind, also

$$wal(2^n,i) \times wal(2^n,j) = \begin{cases} 2^n; & i = j \\ 0; & i \neq j \end{cases} \tag{8.6}$$

gilt. Wie in Abb. 8.15 gilt dieses aber nur bei perfekter Synchronität zweier Ströme, schon bei leichter Abweichung ist die Orthogonalität vollkommen zerstört. Diese Eigenschaften weichen also erheblich von den AKF-Eigenschaften der PN-Sequenz (Abschnitt 8.2.1) ab, die (Abb. 8.7) keine perfekte Orthogonalität zeigen, aber unempfindlich gegen Verschiebungen sind.
Bei perfekter Synchronität sind Datenströme, die mit einer Walsh-Sequenz gespreizt wurden, von solchen, die mit einer anderen multipliziert wurden, leicht trennbar. Es ist nur die AKF zu berechnen, also nur die vermutete Sequenz synchron an den Datenstrom zu multiplizieren und der Mittelwert zu bilden. Im UMTS wird dieses Verfahren ausgenutzt, um im gleichen Frequenzband mit einer Chiprate von 3,84 Mchip/s im Uplink 256 bzw. im Downlink bis zu 512 Kanäle mit Codes zu trennen. Die Spreizfaktoren sind entsprechend.

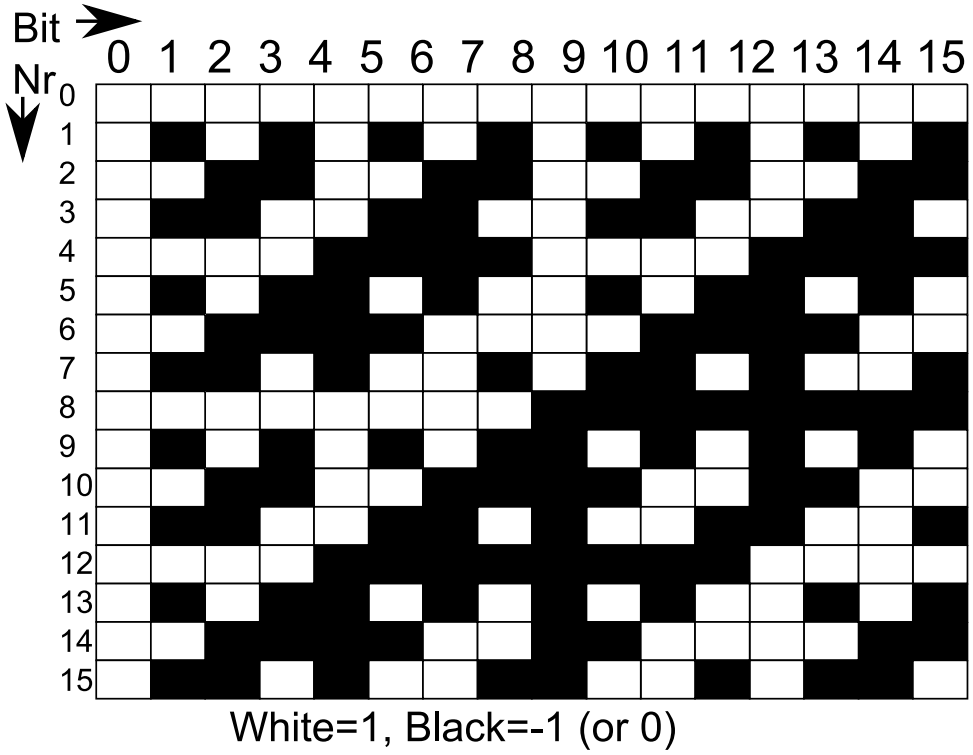


Abbildung 8.14: Walsh-matrix der Länge 16.

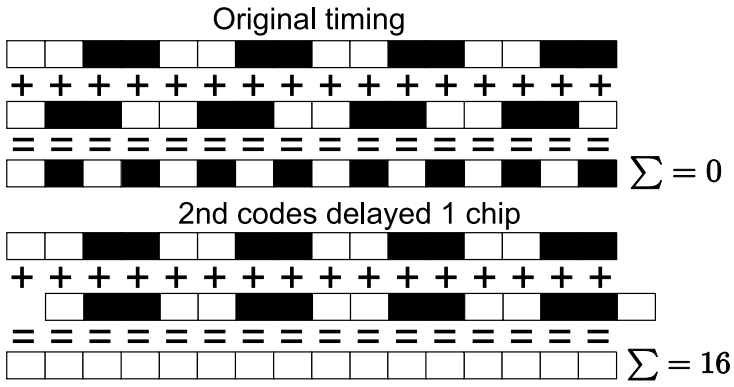


Abbildung 8.15: Zur Orthogonalität der Walsh-Vektoren.

Einzelne oder mehrere Codes können verschiedenen Diensten und Nutzern zugewiesen werden. In Abb. 8.16 ist dieses für einen Nutzer (ein Handy) einmal aufgeführt. Hier ist ebenso gezeigt, dass neben der Kanalisierung der Nutzer mit Walsh-Codes (linke Seite) auch noch die eigentliche Frequenzspreizung mit einem Scrambling Code (rechte Seite) erfolgt. Das System verwendet also zwei Stufen.

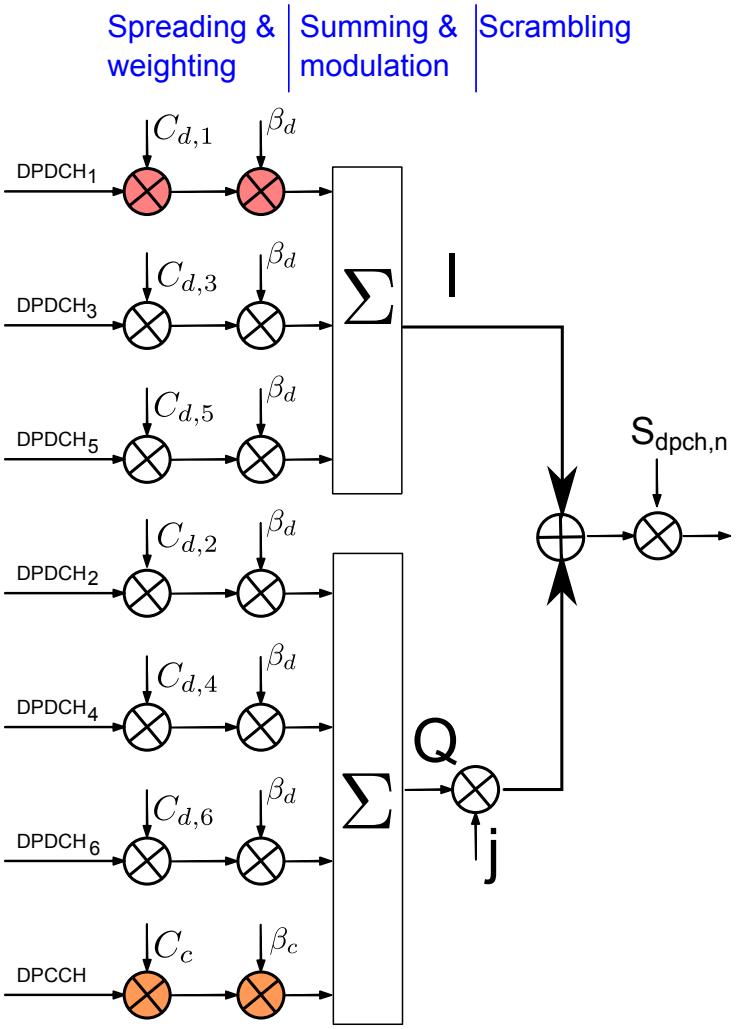


Abbildung 8.16: Offizielle Darstellung der UMTS Uplink-Kanäle für einen Nutzer.

8.2.4 Das Beispiel UMTS

Die Entwicklung und Standardisierung am UMTS wurden als Erweiterung des gerade eingeführten GSM 1992 begonnen. Als Technologie hat sich hier genau die vorstehend beschriebene Technologie des CDMA durchgesetzt. Dieses war nicht selbstverständlich, weitere Technologien standen zur Auswahl. Ob es nun handfeste technische oder ebenso handfeste wirtschaftliche Gründe der Protagonisten waren, die diese Auswahl betrieben haben, ist unklar. Im Sommer 2000, kurz vor dem Platzen der Internetblase wurden für damals ca. 100 Mrd DM die Lizenzen zum Betrieb eines UMTS-Netzes versteigert.
UMTS arbeitet mit einer Chiprate von 3,84 Mchips/s in einem Kanalabstand von 5 MHz. Als Modulation wird QPSK verwendet, wodurch ein I-Kanal und ein Q-Kanal zur Verfügung stehen, die auch (siehe Abb. 8.16) für unterschiedliche logische Kanäle genutzt

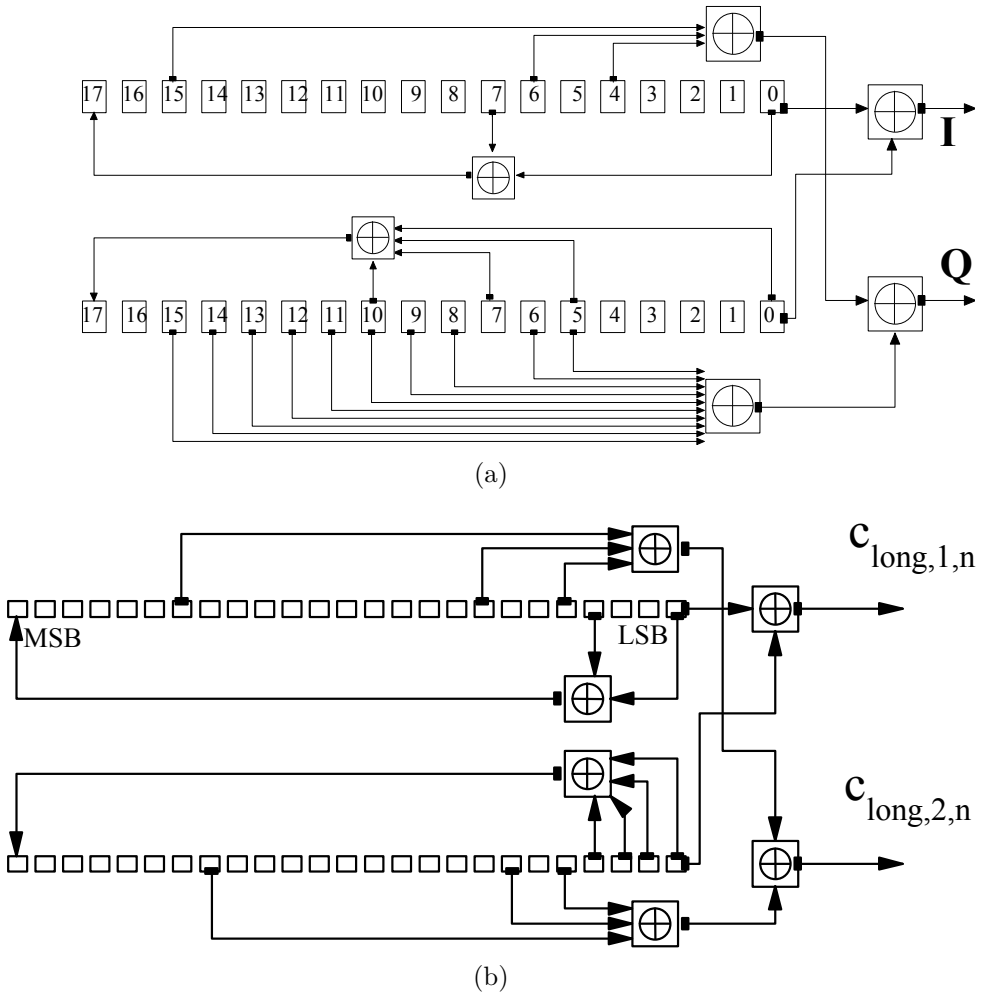


Abbildung 8.17: Erzeugung der Scrambling-Sequenzen im Downlink (a) und im Uplink (b).

werden.
Die eigentliche Spreizung der Daten im UMTS findet mehrfach statt.

- Die Daten werden “gescrambled”, also mit einem Gold-Code aufgespreizt. Dazu stehen im Downlink $2^{18} - 1 = 262\,143$ verschiedene Sequenzen zur Verfügung, von denen 8192 verwendet werden. Die Rückführungen der Generatorpolynome sind bei [1,7,18] und [1,5,7,10,18] (Abb. 8.17a). Die Sequenzen sind sind 512 Sätze zu je 16 Codes eingeteilt und einer Basistation wird ein solcher Satz zugewiesen, der erste Code ist jeweils der “Primary” Code, der in den Pilot-Kanälen im UMTS zu verwenden ist. Die Überlangen Codes werden in Blöcken zu zwei mal 38 400 Chips aufgeteilt (das entspricht einer Zeitdauer von 10 ms) und ein Teil wird für den I-, der

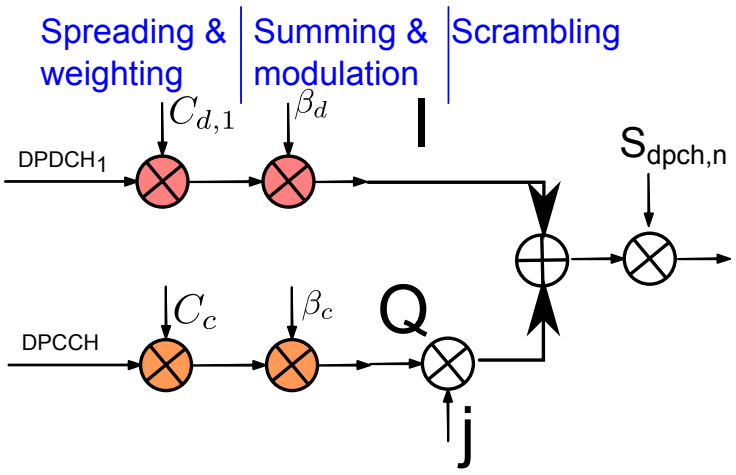


Abbildung 8.18: Zusammenfassung der Spreizporzesse im Uplink.

andere für den Q-Kanal verwendet. Hiermit werden unterschiedliche Basistationen trennbar gemacht.

- Im Uplink existieren $16\,777\,215 = 2^{24} - 1$ Codes mit den Generatorpolynomen $X^{25} + X^3 + 1$ mit Start $\vec{X}_n = \{n_0, n_1, \ldots, n_{23}, 1\}^T$, der die Code-Nummer definiert und $Y^{25} + Y^3 + Y^2 + Y^1 + 1$ mit dem Start $\vec{Y} = \vec{1}$ (Abb. 8.17b). Aus diesen werden zwei long-Sequenzen $c_{long,1,n}$ and $c_{long,2,n}$ erzeugt, die dann schließlich die komplexe Sequenz

$$C_{long,n}(i) = c_{long,1,n}(i)\left(1 + \mathrm{j}(-1)^i c_{long,2,n}(\lfloor i/2 \rfloor)\right) \tag{8.7}$$

  bilden. Hiermit werden unterschiedliche Mobil-Stationen trennbar gemacht. Doch das ist noch nicht genügend, so dass

- die eigentliche Trennung der Kanäle (Nutzer) erfolgt durch die Zuweisung von Walsh-Codes. Hierfür stehen im Downlink 512 und im Uplink 256 orthogonale Vektoren der entsprechenden Läge zur Verfügung. Oft wird dieses als die eigentliche Spreizung bezeichnet, was jedoch nur teilweise korrekt ist.

Der gesamte Spreizprozess ist nochmal für den Downlink in Abb. 8.18 gezeigt.

8.3 Othogonal Frequency Division Multiplexing in 4G, 5G usw.

Als eines der Arbeitspferde der Funkkommunikation, aber nicht nur dort, hat sich mittlerweile das OFDM etabliert. Es ist die Wellenform der Wahl in LTE, 5G NR, Digital

Audio Broadcasting (DAB) und Digital Video Broadcasting (DVB) sowie seit langem dem WiFi in IEEE 802.3 und vielen anderen Funkstandards. Eine etwas abgewandelte Form wird auch kabelgebunden im Digital Subscriber Line (DSL) sowie in der Powerline-Communication verwendet. Grund genug also, sich mit OFDM näher zu befassen.
Die Idee des Verfahrens ist in einfachen Worten schnell beschrieben: Das gesamte zu verwendende Übertragungsband wird in viele schmale Übertragungsbänder aufgeteilt. Das können bis zu mehrere hundert oder sogar mehr als Tausend sein. Die einzelnen Träger werden individuell mit Symbolen moduliert und das Signal wird nach evtl. weiteren Modifikationen über den Kanal übertragen. Daher fällt OFDM auch in die Klasse der Mehrträgerverfahren. Zwangsläufig findet die Übertragung des einzelnen Symbols nun erheblich langsamer statt, die Symboldauer steigt. Es können jedoch mehrere Symbole parallel übertragen werden, so dass, vorausgesetzt man macht es richtig und muss zwischen den einzelnen Frequenzbändern keine Schutzbänder einführen, für das ganze Band die gleiche Übertragungsrate erzielt wie vorher im Einträgerverfahren.
Was sind, wenn man doch nun erst einmal nur nichts verloren hat, die Vorteile dieses Verfahrens?

- Die Symboldauer wird länger, wenn man nun relativ gesehen eine kurze Schutzperiode, entweder eine Wartezeit oder irgendeinen Symbolanteil, zwischen einzelne Symbole einfügt, die länger ist, als die Impulsantwort des Kanals. Man wartet also, bis auch die letzten Mehrwegeanteile des Signals verschwunden sind, kann man sehr einfache Empfänger verwenden, denn man hat keine Intersymbol Interference (ISI) mehr zu verarbeiten.

- Ein sehr frequenzselektiver Kanal würde dazu führen, dass evtl. ein oder mehrere Unterträger stark gestört würden, und ggf. die Symbole dort fehlerbehaftet empfangen und entschieden würden. Bei geschickter Kodierung ist aber die Information und Redundanz der verbleibenden Träger ausreichend, alle Nutzinformationen zu erhalten.

- Geeignete schmalbandige Kanalschätzungen können getätigt werden und der Kanal kann ggf. durch recht einfache Equalizer ausgeglichen werden.

- Im Mehrfachzugriffsverfahren könnten im Down-, oder sogar im Uplink Träger verschiedenen Nutzern zugewiesen werden. Man hat also effektiv ein FDMA, allerdings - was noch zu besprechen sein wird - ohne die sonst notwendigen Schutzbänder

- OFDM lässt sich gut mit Verfahren wie MIMO integrieren.

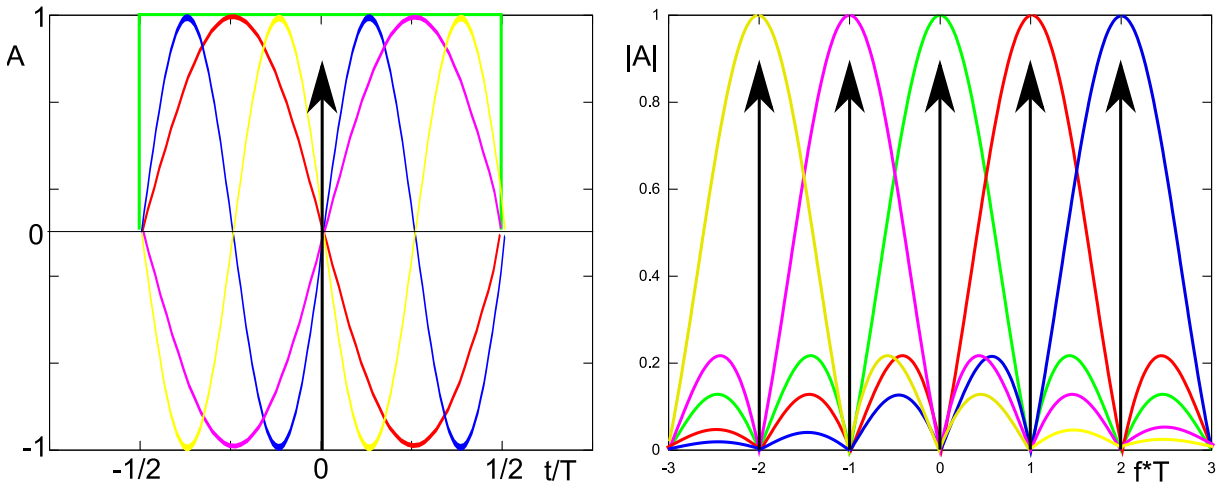


Abbildung 8.19

Evtl. beschränkende oder nachteilige Eigenschaften sind

- Ein erhöhter Aufwand für die Konstruktion der Mehrträger
- Die Symboldauer muss kürzer als die Kohärenzlänge des Kanals sein, über der Dauer eines Symbols dürfen sich die Kanaleigenschaften also nicht zu sehr ändern, da man es ansonsten nicht mehr mit einem zeitinvarianten Verhalten zu tun hat.
- Frequenzverschiebung durch den Dopplereffekt führt bei schmalen Trägern schnell zu Problemen und limitiert den Einsatz von OFDM bzw. muss entsprechend berücksichtigt werden.

Entscheidend für die Anwendung von OFDM sind effektive Implementation und schnelle Algorithmen, bzw. hohe Rechenleistung. Zuerst müssen die Unterträger orthogonal im Zeit- und Frequenzbereich dargestellt werden und sich überlappen dürfen, denn Schutzbänder zwischen einzelnen Unterträgern sind nicht ökonomisch. In Abb. 8.19 ist überlappenden Funktionen im Zeit- und Frequenzbereich dargestellt. Die Symboldauer beträgt T_S und wenn die Frequenz der Unterträger $f_S = \frac{n}{T_S}$ mit ganzen Zahlen n beträgt, sind die darauf gebildeten Funktionen im Zeitbereich und Frequenzbereich orthogonal.
In LTE ist z.B. der Trägerabstand $\Delta f = f_s = 15\,\text{kHz}$ und damit die reine Symboldauer $T_S = 66{,}667\,\mu\text{s}$, in 5G NR kann der Trägerabstand je nach Applikation und Band zwischen $f_s = 15\,\text{kHz}$ und $f_s = 240\,\text{kHz}$ eingestellt werden. Ein gesamter OFDM Sender und Empfänger ist in Abb. 8.20 gezeigt und verdeutlicht auch die anderen Aspekte der Implementation. Zunächst werden im Modulator die evtl. komplexen Modulationssymbole für

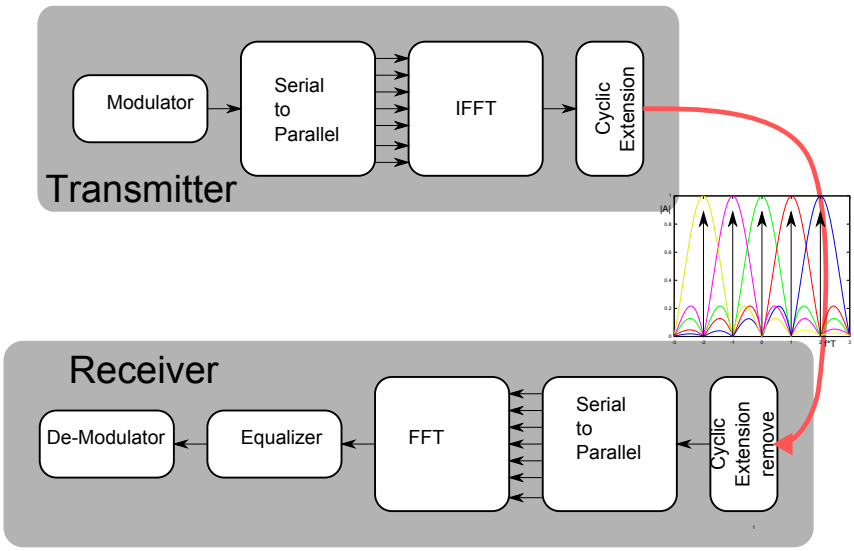


Abbildung 8.20: Prinzipieller Aufbau einer kompletten OFDM-Kette

die ggf. trägerspezifische Modulationsform erzeugt. Diese werden in einer Inversen Fouriertransformation nach Seriell-Parallelwandlung in das zu sendende Signal umgewandelt:

$$s(t) = \sum_{i=N/2}^{N/2-1} s_i \, e^{-\mathrm{j}2\pi\Delta f i t} \, . \tag{8.8}$$

Der eigentliche Hochfrequenzträger liegt hier in der Mitte, und das OFDM-Signal verhält sich symmetrisch um ihn herum. In praktischen Implementation sind im Exponenten noch weitere Parameter z.B. zur Verschiebung in einen bestimmen Frequenzbereich enthalten, die hier aber der Einfachheit entfallen. Die Signalform (Gl. 8.8) lässt sich einfach und schnell durch eine Discrete Fourier Transform (DFT) oder besser noch Fast Fourier Transform (FFT) darstellen, so dass sich insgesamt ein Ablauf mit Signalformen an den unterschiedlichen Stellen wie in Abb. 8.21 ergibt. Zur Illustration wurden links nur die einzelnen Träger ohne Modulation gezeigt, im Gesamtspektrum sind diese dann moduliert. Um die Stärke der OFDM voll aus zu nutzen, nämlich, dass diese Modulationsart gut mit Mehrwegeausbreitung umgehen kann, wird ein Prefix eingeführt (Abb. 8.22) und in dem LTE-System mit realen Parametern in Abb. 8.23 gezeigt.
Im Prinzip ist das Prefix nur eine Schutzzeitzeit zwischen den Symbolen, die so lang ist, dass sämtliche Anteile, die z.B. durch Mehrwegeausbreitung zustande kommen, verschwunden sind. Somit gibt es keine Interferenz zwischen den Symbolen (ISI). Im DAB erlaubt dieses bei guter Synchronität sogar den Aufbau eines bundesweiten Gleichwellennetzes. Im Prinzip sind die Anteile, die evtl. von einem anderen Sender beim Autoradio

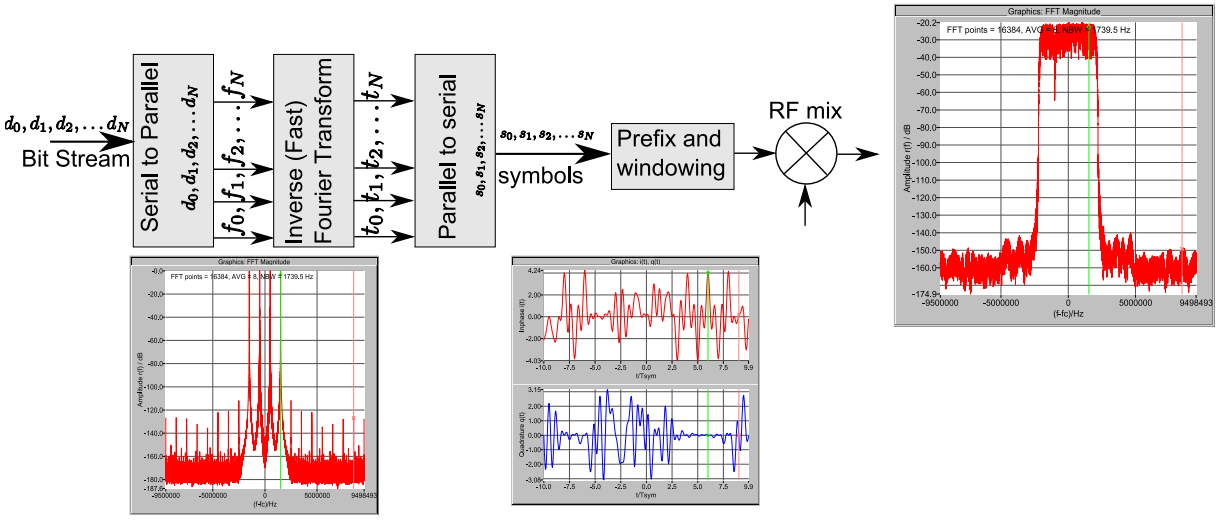


Abbildung 8.21: Struktur eines OFDM Transmitters (ohne explizites Prefix)

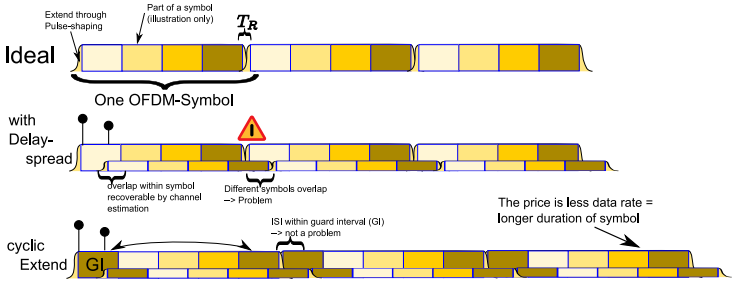
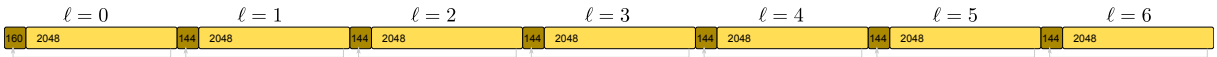


Abbildung 8.22: Zum Prefix, bzw. Cyclic Prefix



$$T_{total} = 15\,360 \times T_S = 0.5\,\text{ms}$$

Abbildung 8.23: ein LTE-Downlink Zeitschlitz

ankommen auch nur verzögerte Anteile, die durch Mehrwegeausbreitung zustande gekommen sein könnten. Mit einer Schutzzeit würden auch diese Anteile keine ISI erzeugen.
Um den Verstärker nicht hoch und runter fahren zu müssen, und weil man auf diese Art noch ggf. gestörte Symbolanteile an anderer Stelle hat, verwendet man anstatt Stille einfach ein sogenanntes Cyclic Prefix. Hier werden also entsprechend der Länge Symbolanteile vom Ende an den Anfang des Symbols gesetzt. Im Empfänger wird dann zunächst der optimale, bzw. notwendige Symbolteil gesucht und die nicht benötigten Anteile werden entfernt und ggf. das Symbol nochmal mit einer Fensterfunktion gewichtet. Insgesamt kann man sich die OFDM-Symbole also wie in Abb. 8.24 vorstellen
Praktische Parameter z.B. aus dem LTE sind eine Zeitbasis von $T_S = \frac{1}{30{,}72\,\text{MHz}} =$

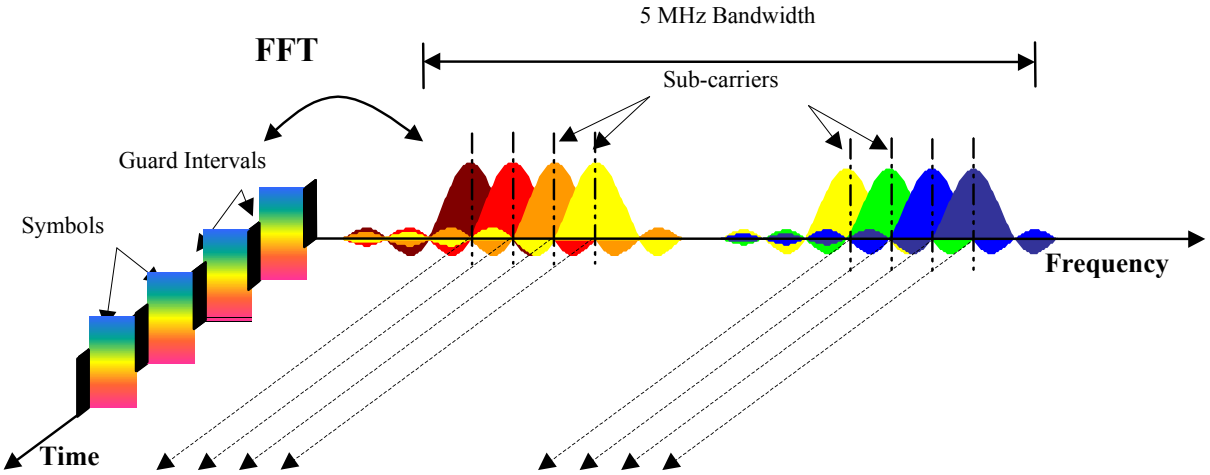


Abbildung 8.24: Die OFDM Situation nach [45].

Tabelle 8.3: Struktur der OFDM-Symbole

		CP Parameters		
CP Config		N_{symb}^{DL}	CP_{len} $\ell = 0$	CP_{len} $\ell \neq 0$
Normal	$\Delta f = 15\,\text{kHz}$	7	160	144
Extended	$\Delta f = 15\,\text{kHz}$	6	512	512
	$\Delta f = 7.5\,\text{kHz}$	3	1024	1024

32,552 ns und eine Länge eines Zeitschlitzes mit $15\,360\,T_S = 0.5\,\text{ms}$ und einem Subframe mit zwei Zeitschlitzen, der dann 1 ms lang ist. Der Trägerabstand beträgt also $\Delta f = 15\,\text{kHz}$ und damit ist die Symbollänge $T_{Symb} = 66{,}67\,\mu\text{s} = 2048\,T_S$. Unterschiedliche Strukturen der Schlitze, die mit unterschiedlichen Prefix-Längen auch unterschiedlichen Ausbreitungsszenarien Rechnung tragen sind in Tab.8.3 aufgeführt. Mit Prefix-Zeiten von 4,6875; 16,67; 33,33 μs können Distanz-Spreizungen von 1.4; 5; 10 km bekämpft werden.

Kapitel 9

Modulation, Demodulation und Trägerrückgewinnung

Im Abschnit 4.3 wurden komplexe, lineare Modulationsformen im Basisband vorgestellt. Dieses ist an dieser Stelle in Verbindung mit der notwendigen Modulation und De-Modulation nach Abschnitt 6 für digitale Modulationsverfahren zu untersuchen.

9.1 Ideale Modulation und De-Modulation

In seiner reinen und kohärenten Form kann ein kompletter Modulations-, Übertragungs-, De-Modulationsprozess wie in Abb. 9.1 dargestellt werden. Die In- und Quadraturphase Komponenten I und Q werden mit einem komplexen Lokaloszillatorsignal

$$u(t) = U_0 \times e^{-j\omega t} \tag{9.1}$$

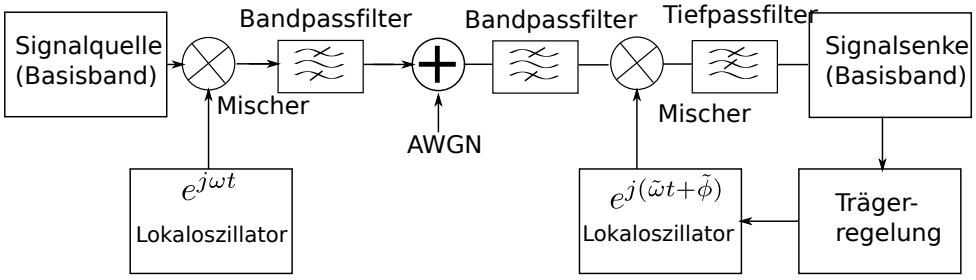


Abbildung 9.1: Modulator, Kanal und Demodulator in idealisierter Form.

der Trägerfrequenz ω umgesetzt. Am Empfänger wird durch phasen- und frequenzrichtige Multiplikation mit dem dort vorhandenen Trägersignal

$$\tilde{u}(t) = \tilde{U}_0 \times e^{-j(\tilde{\omega}t+\tilde{\phi})} \tag{9.2}$$

umgesetzt und danach mit einem Tiefpass gefiltert. Mathematisch liegt also der Prozess (Symbole $s_I + js_Q$ und $U_0 = \tilde{U}_0 = 1$ angenommen) vor:

$$s(t) = s_I \cos \omega t - s_Q \sin \omega t \tag{9.3}$$

und dann im Empfänger mit idealer Annahme $\tilde{\phi} = 0$ und $\tilde{\omega} = \omega$ in den einzelnen Zweigen:

$$\tilde{s}_I(t) = \int_0^{T_s} \left(s_I \cos \omega t - s_Q \sin \omega t\right) \times \cos \omega t \, dt \tag{9.4}$$

$$= \int_0^{T_s} \left(s_I \cos^2 \omega t - s_Q \sin(\omega t) \cos(\omega t)\right) \, dt \tag{9.5}$$

$$\approx a \times s_I \tag{9.6}$$

wobei die Näherung gewählt wurde, weil die Symboldauer T_S nicht unbedingt ein ganzzahliges Vielfaches der Periodendauer $2\pi/\omega$ des Trägers sein muss. a ist einfach ein konstanter Faktor. Die Integration, bzw. Filterung mit einem Tiefpass lässt die Komponenten mit höherer Frequenz nach dem eigentlichen Demodulationsprozess verschwinden. Die Filterung ist daher essentiell und sollte idealerweise an die Impulsform s_I, s_Q angepasst sein. Hier wurde vereinfachend von einem Recheckimpuls ausgegangen. Die Filterform ist daher auch zu einer Rechteckform im Zeitbereich, einer $\sin x/x$-Form im Frequenzbereich gewählt worden. Ähnliche Argumente wie in Gleichung 9.6 gelten auch für die Q-Komponente. Wie sich nun die Bitfehler verhalten, wenn auf dem Kanal einfach nur Rauschen zugefügt wird, ist in Abschnitt 4.4.4 beschrieben. Da hier im Empfänger das Trägersignal frequenz- und phasenrichtig wieder zugefügt wird, spricht man hier von kohärenter Demodulation. Offensichtlich ist, dass diese kohärente Demodulation nur dann gut funktioniert, wenn eben Frequenz und Phase des Modulators im Demodulator bekannt sind. Da diese nicht unmittelbar vorhanden sind, müssen beide Parameter im Empfänger wieder hergestellt werden. Dieses ist in Abb. 9.2 gezeigt.

Bevor aber betrachtet wird, wie der Träger rückgewonnen werden kann, sollen die Auswirkungen eines Phasenfehlers im Träger besprochen werden. Ein abweichende Phase $\tilde{\phi}$ zwischen Sender und Empfänger bei gleichem Träger führt bei einer einfachen ASK mit

$$s(t) = s_I(t) \cos \omega t \tag{9.7}$$

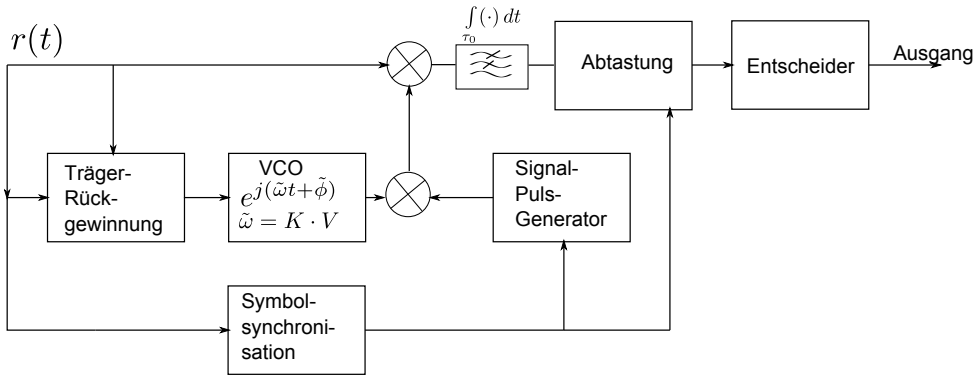


Abbildung 9.2: Blockdiagramm eines eines BPSK Empfängers.

mit also $s_Q = 0$, weil nur die I-Komponenten (also nur die Amplitude) moduliert wird, zu

$$\tilde{s}(t) = s_I(t)\cos(\omega t)\cos(\omega t + \tilde{\phi}) \tag{9.8}$$

$$= \frac{s_I}{2}\cos\tilde{\phi} + \frac{s_I}{2}\cos\left(2\omega t + \tilde{\phi}\right) \tag{9.9}$$

und nach Filterung bleibt dann also

$$\hat{s}(t) = \frac{s_I}{2}\cos\tilde{\phi}, \tag{9.10}$$

was besagt, dass die detektierte Amplitude und damit die Leistung um $\cos\tilde{\phi}$ bzw. $\cos^2\tilde{\phi}$ reduziert werden. Schon dieses ist - in dieser Form - desaströs, wenn der Phasenunterschied 90° beträgt. Dann ist die detektierte Amplitude nämlich Null. Dieses lässt sich allerdings ggf. leicht durch eine Leistungsmessung vor der Heruntermischung auf die Frequenz Null beheben. Keine Chance hat man mehr bei einer Quadratur-Modulation, bei der mit

$$s(t) = s_I(t)\cos\omega t - s_Q\sin\omega t \tag{9.11}$$

nach

$$\tilde{s}_I(t) = \left(s_I(t)\cos\omega t - s_Q\sin\omega t\right)\cos(\omega t + \tilde{\phi}) \tag{9.12}$$

$$= s_I(t)\cos\omega t\cos(\omega t\tilde{\phi}) - s_Q\sin\omega t\cos(\omega t + \tilde{\phi}) \tag{9.13}$$

$$= \frac{s_I(t)}{2}\cos\tilde{\phi} + \frac{s_I(t)}{2}\cos(2\omega t + \tilde{\phi}) - \frac{s_Q(t)}{2}\sin\tilde{\phi} - \frac{s_Q(t)}{2}\sin(2\omega t + \tilde{\phi}) \tag{9.14}$$

gilt. Tiefpassfilterung vernichtet die Komponenten bei doppelter Frequenz:

$$\hat{s_I}(t) = \frac{s_I(t)}{2}\cos\tilde{\phi} - \frac{s_Q(t)}{2}\sin\tilde{\phi} \tag{9.15}$$

und entsprechend:

$$\hat{s_Q}(t) = \frac{s_Q(t)}{2} \cos\tilde{\phi} + \frac{s_I(t)}{2} \sin\tilde{\phi}. \tag{9.16}$$

So dass am Empfänger in der Regel irgendeine Kombination der I und Q-Komponente dem Entscheider zugeführt wird, der dann daraus Fehlentscheidungen ableiten wird. Dieses belegt eindeutig die Notwendigkeit einer phasenrichtigen Trägerrückgewinnung.

9.2 Trägerrückgewinnung

Es existieren verschiedene Verfahren der Trägerrückgewinnung:

1. Zufügen des Trägers als Pilot oder als Trägerrest in der Amplitudenmodulation. Hier kann mit einem schmalbandigen Bandpassfilter der Träger vom modulierten Signal isoliert werden und genutzt werden, um im Empfänger eine PLL anzusteuern.

2. Ableiten des Trägersignals aus dem modulierten Signal (mit unterdrücktem Träger). Dieses ist das weiter verbreitete Verfahren, weil zwar ggf. der Aufwand im Empfänger mindestens prinzipiell steigt, aber dafür sämtliche Sendeleistung in das modulierte Signal gesandt wird und damit zur Erhöhung des Signal Rausch-Verhältnisses zur Verfügung steht. Dieses Verfahren unterteilt sich in

   (a) Demodulationsmethoden, die die Phase des Trägers aus dem demodulierten und entschiedenen Symbol ableiten und

   (b) solche, bei denen die statistische Natur des Signals verwendet wird.

9.2.1 PLL und Trägerrückgewinnung aus dem unmodulierten Träger

Die konzeptionell einfachste Variante zur Trägerrückgewinnung ist eine, bei der der unmodulierte Träger zur Verfügung steht. Dieser könnte z.B. in bei einer Zweiseitenbandmodulation beigesetzt sein und durch ein schmalbandiges Tiefpassfilter isoliert werden (Abb. 9.3). Wird dieser mit einem lokal erzeugten sinusförmigen Signal multipliziert, so ergibt sich

$$e(t) = \cos\omega t \sin(\omega t + \tilde{\phi}) \tag{9.17}$$

$$= \frac{1}{2}\sin\tilde{\phi} + \frac{1}{2}\sin(2\omega t + \tilde{\phi}). \tag{9.18}$$

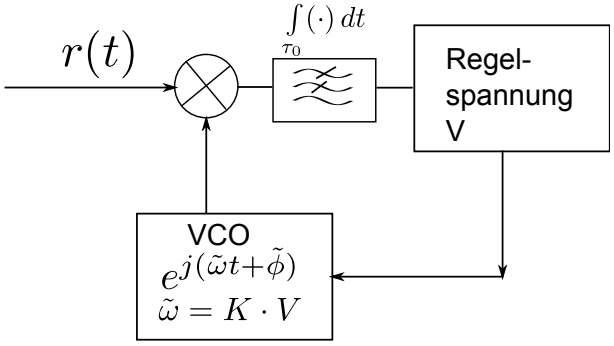


Abbildung 9.3: Generische Gewinnung der Phasendifferenz aus einem unmodulierten Träger.

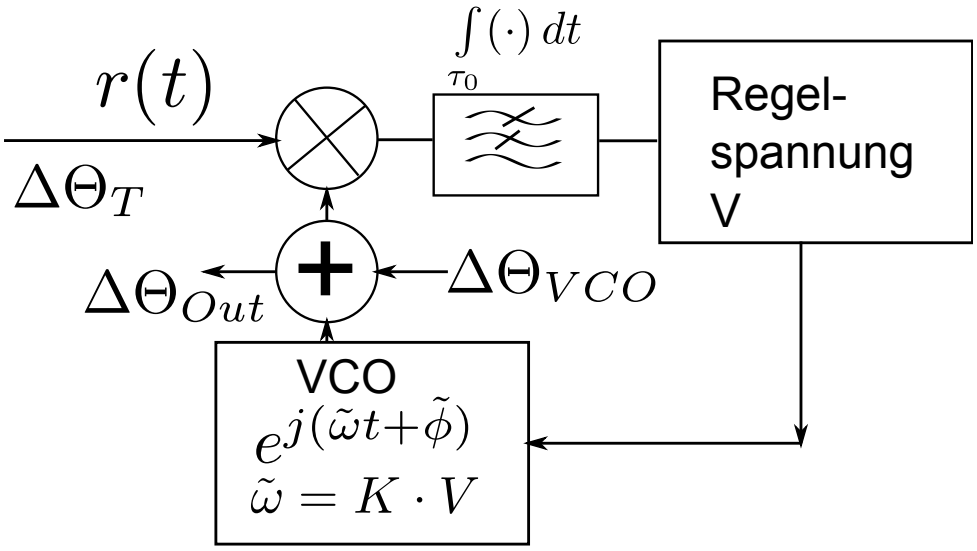


Abbildung 9.4: Generische Darstellung eines Phasenregelkreises (PLL).

Die Tatsache, dass beide Signale in diesem Ausdruck die gleiche Frequenz $\omega/(2\pi)$ aufweisen und sich durch eine potenziell zeitabhängige Phase $\tilde{\phi}$ unterscheiden, soll unterstreichen, dass beide Signale frequenzmäßig mindestens dicht benachbart sind. Kleine Unterschiede in der Frequenz könnten bspw. auch durch einen Phasenterm $\tilde{\phi} = \Delta\omega t + \tilde{\psi}$ ausgedrückt werden und sind somit in obiger Gleichung (9.18) sehr wohl auch erfasst. Der erste Term in Gl. (9.18) ist für kleine Winkelabweichungen $\approx \tilde{\phi}$ und könnte - wenn durch ein Tiefpassfilter von dem restlichen schnell variierenden Term isoliert als Regelgröße in einem Phasenregelkreis verwendet werden (Abb. 9.4). Er besteht eben aus den

1. schon genannten Komponenten zur Bestimmung des Fehlers,

2. dem Schleifenfilter (Loopfilter), das als Tiefpassfilter ausgeführt ist,

3. einer Schleifenverstärkung, die sich auch in VCO und Filter verstecken kann,

4. und dem zu regelnden Glied, hier dem spannungsgesteuerten Oszillator (VCO).

PLL können ganz analog, teilweise analog oder auch ganz digital, ggf. sogar in Programmcode, der nicht einmal in Echtzeit läuft, realisiert sein. Der Voltage Controlled Oscillator (Spannungsgesteuerter Oszillator) (VCO) kann modelliert werden als Oszillator mit einer momentanen Phase, die von der Historie seiner Steuerspannungen abhängt:

$$\Phi(t) = \omega t + \tilde{\phi}(t) = \omega t + K_{VCO} \int_0^t u(\tau) d\tau \tag{9.19}$$

wenn die Phase zum Zeitpunkt $t = 0$ Null war. K_{VCO} ist irgendein konstanter Verstärkungsfaktor. Legt man eine konstante Spannung u_0 an, so wandert natürlich die Phase mit konstanter Geschwindigkeit, und daher ist die Frequenz dann

$$\omega t + K_{VCO} \int_0^t u_0 d\tau = \omega t + K_{VCO} u_0 t = (\omega + K u_0) t \tag{9.20}$$

und $K/(2\pi)$ bestimmt die Abstimmempfindlichkeit in Hz/V. Dieses ist ein wichtiges Charakteristikum eines VCO. Gemäß Gl. 9.19 kann nach Fourier-Transformation der VCO als Integrator mit der Transferfunktion

$$V(j\omega) = \frac{K_{VCO}}{j\omega} \tag{9.21}$$

geschrieben werden. Das Schleifenfilter hat die normale Übertragungsfunktion erster Ordnung

$$G(j\omega) = \frac{1}{1 + j\omega/\omega_g}. \tag{9.22}$$

Damit ergibt sich im Bereich $\tau > 1/\omega$ die Tiefpass-typische um 20 dB pro Dekade abfallende Frequenzcharakteristik. Mit der Annahme der Linearisierung $\sin\tilde{\phi} \approx \tilde{\phi}$ kann die gesamte Schleifenübertragung zu

$$H(j\omega) = \frac{K/(j\omega) \times G(j\omega)}{1 + K/(j\omega) \times G(j\omega)} \tag{9.23}$$

$$= \frac{K}{K + j\omega(1 + j\omega\tau)} \tag{9.24}$$

$$= \frac{K}{K + j\omega - \omega^2\tau} \tag{9.25}$$

erfasst werden, wobei hier K die gesamte Normierung und Verstärkung der Schleife aus allen Komponenten enthält. Dieses beschreibt die Schleifenübertragungsfunktion, die eine Phasenänderung des Trägers ($\Delta\Theta_T$) am Eingang auf den Ausgang $\Delta\Theta_{out}$ auslöst. Eine Störung, also Phasenschwankung des VCO erscheint am Ausgang einfach als

$$H'(j\omega) = \frac{1}{1 + V(j\omega)G(j\omega)} \tag{9.26}$$

$$= \frac{j\omega - \omega^2\tau}{K + j\omega - \omega^2\tau}. \tag{9.27}$$

Zusammengebaut ergibt sich für die Phase am Ausgang nun

$$\Delta\Theta_{out} = \frac{K}{K + j\omega - \omega^2\tau}\Delta\Theta_T + \frac{j\omega - \omega^2\tau}{K + j\omega - \omega^2\tau}\Delta\Theta_{VCO}. \tag{9.28}$$

Bei geringen Frequenzen und großen, bzw. genügenden Verstärkungen K kann der zweite Term vernachlässigt werden und es ist ersichtlich, dass die Ausgangsphase der Phase des Trägers folgt, der VCO selber hat keinen Einfluss auf die Ausgangsphase. Bei hohen Frequenzen aber genügt die Schleifenverstärkung nicht mehr und der zweite Term beginnt irgendwann das Geschehen zu dominieren. Dann wiederum spielt der Träger keine Rolle mehr und die Eigenschaften des VCO sind am Ausgang zu erkennen. Während der erste Term ein Tiefpassfilter zweiter Ordnung darstellt, beschreibt der zweite ein Hochpassfilter. Beiden kann eine obere bzw. untere Grenzfrequenz zugeordnet werden, indem der Punkt berechnet wird, an dem der Betrag des Frequenzganges auf $1/\sqrt{2}$ abgefallen ist, also für den ersten Term

$$\left|1 + \frac{j\omega}{K} - \omega^2\frac{\tau}{K}\right|^2 = 2 \tag{9.29}$$

$$\Leftrightarrow \left(1 + \frac{\omega^2\tau}{K} + \frac{\omega^2}{K^2}\right) = 2 \tag{9.30}$$

gilt. Das Ergebnis ist

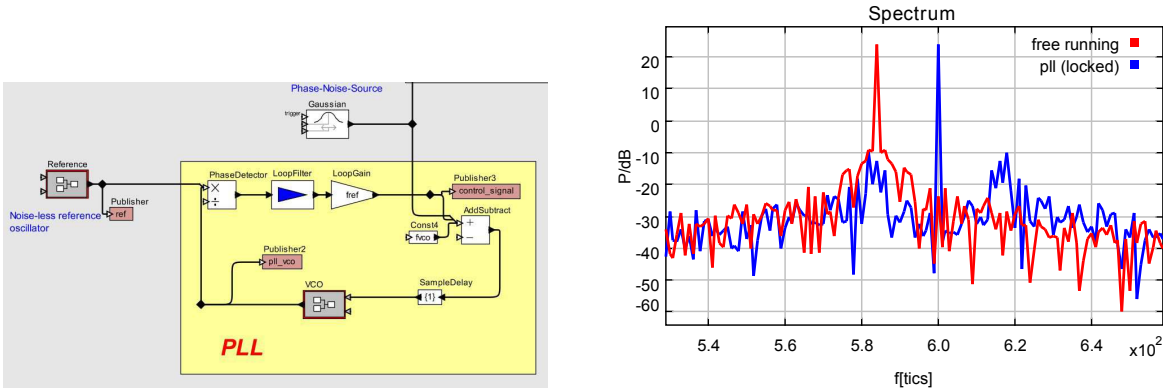
$$\omega_{TP} = \pm\sqrt{\frac{1}{\tau}\left(\sqrt{2K^2 + \frac{K}{\tau} + \frac{1}{4\tau^2}} - K - \frac{1}{2\tau}\right)} \tag{9.31}$$

und mit

$$\frac{\left|\frac{j\omega}{K} - \frac{\omega^2\tau}{K}\right|}{\left|1 + \frac{j\omega}{K} - \omega^2\frac{\tau}{K}\right|} = \frac{1}{\sqrt{2}} \tag{9.32}$$

und schließlich

$$\omega_{HP} = \pm\sqrt{\frac{1}{\tau}\left(K + \frac{1}{2\tau} + \sqrt{2K^2 + \frac{1}{4\tau^2} + \frac{K}{\tau}}\right)} \tag{9.33}$$



(a) (b)

Abbildung 9.5: Implementation einer PLL (a) und Ergebnis der Simulation mit Spektrum des frei laufenden VCO und des eingerasteten (b)..

als untere Grenzfrequenz für das Hochpassfilter. Natürlich liegt die Grenzfrequenz des Hochpassfilters über der des Tiefpassfilters. Je nach Maßgabe des Phasenrauschens des VCO, also der Amplitude von $\Delta\omega_{VCO}$ beginnt dieses irgendwann das Geschehen zu dominieren, da die TP-Charakteristik nicht mehr greifen kann.
Die PLL ist offensichtlich in der Lage, dem Phasensignals des Trägers bis zur Grenzfrequenz des Schleifenfilters zu folgen, vor allem wenn der VCO schon vor langer Zeit der Regelung folgte, die PLL also eingrastet ist. Dieser Bereich ist der *Haltebereich*. Das *Fangverhalten* einer PLL kann mit den o.g. linearen Analysen nicht mehr untersucht werden. Nichtlineares Verhalten ist außerhalb der Betrachtungen dieser Veranstaltung. In Abb. 9.5 ist eine beispielhafte Implementation als Signalflussmodell als Simulation gezeigt, wie auch das Ergebnis, an dem man erkennt, dass der freilaufende VCO auf die Frequenz der Referenz eingerastet hat und auch die PLL in der Nähe des Referenzträgers das Rauschen gegenüber dem freilaufenden VCO deutlich senken konnte.

Einschub: Simulation eines Tiefpassfilters

In einem Phasenregelkreis werden Tiefpassfilter benötigt. Eine einfache analoge Realisierung eines solchen Filters ist beispielsweise einfach ein RC-Glied (Abb. 9.6), bei dem die Ausgangsspannung in der komplexen Wechselstromrechnung gemäß

$$U_2(j\omega) = U_1(j\omega)\frac{1}{1+j\omega RC} = U_1(j\omega)\frac{1}{1+j\omega/\omega_g} \tag{9.34}$$

mit der Grenzkreisfrequenz $\omega_g = 1/(RC)$ des RC-Gliedes von Eingangsspannng abhängt. Um - quasi rückwärts - eine Differentialgleichung abzuleiten wird der Nenner auf die rechte Seite gebracht und

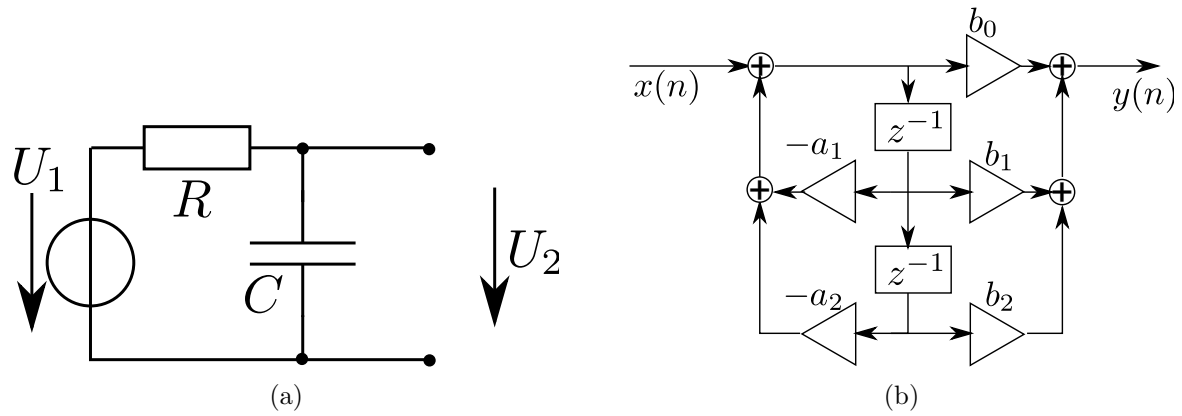


Abbildung 9.6: Analoges Tiefpassfilter Widerstand R und Kapazität C (a) und generische Darstellung als digitales IIR-Filter (b).

es entsteht

$$U_2(1 + j\omega/\omega_g) = U_1. \tag{9.35}$$

Ersetzt man nun einfach $j\omega$ durch das Differential $\frac{d}{dt}$ so hat man die Differentialgleichung (Dieser Schritt ist nicht mathematisch sauber, aber er funktioniert)

$$u_2(t) + \frac{1}{\omega_g}\frac{du_2(t)}{dt} = u_1(t). \tag{9.36}$$

Gleichzeitig wurde die Zeitabhängigkeit der Spannungen eingesetzt. das Differential lässt sich beispielsweise durch eine Differenzengleichung der Form

$$\frac{df}{dt} \approx \frac{f(t) - f(t - \Delta t)}{\Delta t} = f_s\left(f(t) - f(t - \Delta t)\right) \tag{9.37}$$

mit der Abtastfrequenz $f_s = 1/\Delta t$ annähern. Dieses auf obige DGL angewandt führt auf

$$u_2(t) + \frac{f_s}{\omega_g}\left(u_2(t) - u_2(t - \Delta t)\right) = u_1(t) \tag{9.38}$$

$$\Leftrightarrow u_2(t) = \frac{1}{1 + \frac{f_s}{\omega_g}}u_1(t) + \frac{\frac{f_s}{\omega_g}}{1 + \frac{f_s}{\omega_g}}u_2(t - \Delta t). \tag{9.39}$$

Die Spannung (das Signal) am Ausgang, also u_2 oder y hängt vom Eingangssignal u_1 oder x und dem Ausgangssignal am vorherigen Abtastzeitpunkt ab. Das Ausgangssignal am vorangegangen Zeitpunkt kann durch ein Schieberegister erzeugt werden. Die mathematische Notation dafür ist bekanntlich (z-Transformation) z^{-1}. Generisch ist so ein Filter in Abb. 9.6(b) dargestellt. Offensichtlich ist in dieser Schieberegisterdarstellung

$$a_0 y(n) = b_0 x(n) + b_1 x(n-1) + b_2 x(n-2) - a_1 y(n-1) - a_2 y(n-2). \tag{9.40}$$

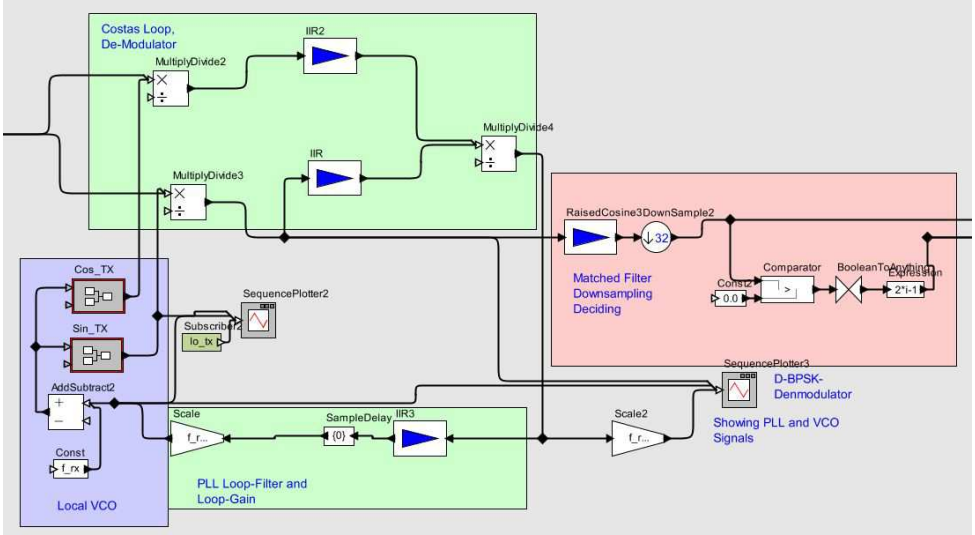


Abbildung 9.7: Costas-Loop zur Rückgewinnung des Trägers bei einem BPSK-Signal.

Die Koeffizienten in Gleichung (9.40) können durch Koeffizientenvergleich mit Gleichung (9.39) gewonnen werden und es ist

$$b_0 = \frac{1}{1 + \frac{f_s}{\omega_g}}, \qquad b_1 = 0, \qquad b_2 = 0 \tag{9.41}$$

$$a_0 = 1, \qquad a_1 = -\frac{\frac{f_s}{\omega_g}}{1 + \frac{f_s}{\omega_g}}, \qquad a_2 = 0. \tag{9.42}$$

Diese Koeffizienten können analytisch in IIR-Filtern zur Simulation vorgegeben werden. Es wird noch darauf hingewiesen, dass die hier verwendete Umwandlung von Differential- in Differenzengleichung für im weiteren zu besprechende Fälle ungenügend ist. Dann wird eine bessere Näherung betrachtet werden.

9.2.2 Costas-Loop und Trägerrückgewinnung aus dem modulierten Signal

Die Rückgewinnung der Phaseninformation aus einem zugesetzten (Rest)Träger birgt den Nachteil in sich, dass ein Träger zugesetzt sein muss. Dieser erfordert Energie, die den Nutzdaten nicht zur Verfügung steht. Die Leistungsfähigkeit einer solchen Implementation leidet in dem Maße, wie eben Energie aus den Symbolen in den Träger abgezogen wird. Eine bessere Methode ist es daher, den Träger aus dem modulierten Signal zu rekonstruieren. Dieses kann für die bipolare Modulation BPSK einfach in Form einer Costas-Loop bewerkstelligt werden. So eine Schleife ist in Abb. 9.7 dargestellt. Hier wird aus dem

modulierten Eingang ein Signal erzeugt, dass der Phasenabweichung zwischen ursprünglichem Träger und lokalem Oszillator-Signal proportional ist. Die Multiplikation des - hier noch allgemeinen - Empfangssignal

$$s(t) = d_I(t)\cos\omega t + d_Q(t)\sin\omega t \tag{9.43}$$

mit nur langsam zeitabhängigen Datengrößen d_I und d_Q wird im Empfänger kohärent heruntergemischt auf

$$\begin{aligned} d_1 &= \text{LPF}\left[\left(d_I(t)\cos\omega t + d_Q(t)\sin\omega t\right) \times \sin(\omega t + \phi)\right] && (9.44)\\ &= \text{LPF}\left[d_I(t)\cos\omega t\sin(\omega t + \phi) + d_Q(t)\sin\omega t\sin(\omega t + \phi)\right] && (9.45)\\ &= \frac{1}{2}\left(d_I(t)\sin\phi + d_Q(t)\cos\phi\right) && (9.46) \end{aligned}$$

und

$$\tag{9.47}$$

$$\begin{aligned} d_2 &= \text{LPF}\left[\left(d_I(t)\cos\omega t + d_Q(t)\sin\omega t\right) \times \cos(\omega t + \phi)\right] && (9.48)\\ &= \text{LPF}\left[d_I(t)\cos\omega t\cos(\omega t + \phi) + d_Q(t)\sin\omega t\cos(\omega t + \phi)\right] && (9.49)\\ &= \frac{1}{2}\left(d_I(t)\cos\phi - d_Q(t)\sin\phi\right). && (9.50) \end{aligned}$$

Mit der Operation LPF ist eine Tiefpassfilterung gemeint, die ideal die Komponenten bei der doppelten Frequenz unterdrückt.
Für ein reines BPSK-Signal mit rechteckförmiger Darstellung der Bits ist natürlich $|d_I| = 1$ und $d_Q = 0$ zu allen Zeiten und es verbleibt

$$d_1 = \frac{d_I}{2}\sin\phi \tag{9.51}$$

$$d_2 = \frac{d_I}{2}\cos\phi. \tag{9.52}$$

Die Multiplikation dieser beiden Größen, die in Abb. 9.7 durchgeführt wird, bringt den Term

$$d_1 \times d_2 = \frac{d_I^2}{4}\sin\phi\cos\phi \approx \phi \tag{9.53}$$

für kleine ϕ, denn $d_I = 1$. Damit liegt eine Regelgröße vor, die der Phasenabweichung zwischen Träger und LO proportional ist. Wie zu sehen ist, tritt das gleiche Regelsignal bei $\phi' = \phi + \pi$ auf. Dann sind sin und cos negativ, es liegt eine Mehrdeutigkeit von 180° vor; die Regelung des LO erfolgt also - willkürlich - auf den Träger oder seine negierte

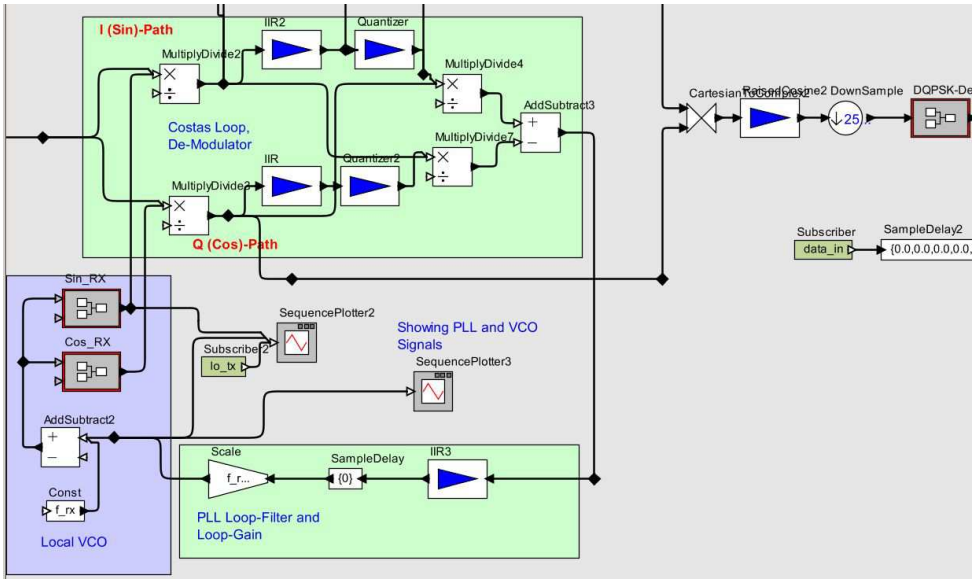


Abbildung 9.8: Erweiterung der Costas-Loop zur Rückgewinnung des Trägers bei einem QPSK-Signal.

Form. Daher hat man nach dem Entscheider entweder (Rauschfreiheit vorausgesetzt) keine oder 100 % Bitfehler. Dieses lässt sich natürlich leicht korrigieren, bzw. durch Verwendung einer differentiellen Modulation von vorneherein vermeiden.
In Abb. 9.8 ist eine Erweiterung der Costas-Loop zur Trägerrückgewinnung von QPSK-Signalen dargestellt. Hier werden in den einzelnen Zweigen zunächst nur die Vorzeichen der Werte gebildet, also

$$d_1' = \mathrm{sgn}\left[d_I \sin\phi + d_Q \cos\phi\right] \tag{9.54}$$

$$d_2' = \mathrm{sgn}\left[d_I \cos\phi - d_Q \sin\phi\right]. \tag{9.55}$$

Für kleine Phasenabweichungen ϕ und $|d_I| = |d_Q|$ (rechteckförmige Signale in reiner Phasenmodulation) ist der cos-Term immer größer als der sin-Term und immer positiv, und nur das Vorzeichen des damit verbundenen Datensignals dominiert das Ergebnis:

$$d_1' \approx \mathrm{sgn}(d_Q)\cos\phi \tag{9.56}$$

$$d_2' \approx \mathrm{sgn}(d_I)\cos\phi. \tag{9.57}$$

Die nun folgenden kreuzweise Multiplikation mit d_1 und d_2 gemäß Gleichung (9.46) und

(9.50) führt auf

$$d_1' d_2 \approx \operatorname{sgn}(d_Q) \cos\phi \left(d_I \cos\phi - d_Q \sin\phi \right) \tag{9.58}$$
$$= -|d_Q| \sin\phi \cos\phi + \operatorname{sgn}(d_Q) d_I \cos^2\phi \tag{9.59}$$
$$d_2' d_1 \approx \operatorname{sgn}(d_I) \cos\phi \left(d_I(t) \sin\phi + d_Q(t) \cos\phi \right) \tag{9.60}$$
$$= |d_I| \sin\phi \cos\phi + \operatorname{sgn}(d_I) d_Q \cos^2\phi \tag{9.61}$$

und die Differenz ist

$$-d_1' d_2 + d_2' d_1 \approx (|d_I| + |d_Q|) \sin\phi \cos\phi + (\operatorname{sgn}(d_I) d_Q - \operatorname{sgn}(d_Q) d_I) \cos^2\phi \tag{9.62}$$
$$\Rightarrow \overline{-d_1' d_2 + d_2' d_1} \approx 2 \sin\phi \cos\phi \tag{9.63}$$

wobei verwendet wurde, dass bei reiner Phasenmodulation $|d_I| = |d_Q| = 1$ ist und im Mittel positive wie negative Datenwerte gleich wahrscheinlich sind daher der zweite Term nach Tiefpassfilterung durch das Schleifenfilter verschwindet. Wiederum ist ein Signal entstanden, das der Phasenabweichung ϕ näherungsweise proportional ist. Die Mehrdeutigkeit ist hier $2\pi/N$ mit der Wertigkeit der Modulation N. Bei einer QPSK ist die Phasen also nur mit einer Unsicherheit in 90°-Schritten regelbar. Wiederum ist diese Mehrdeutigkeit durch grobe Vor- oder Nachsynchronisation oder durch eine differentielle Kodierung aufzuheben.
BPF und M-ary-Loop einfügen

9.2.3 Trägerrückgewinnung für allgemeine Modulationen

Die bisher dargestellten Verfahren zur Trägerrückgewinnung erforderten entweder den Zusatz des Trägers selber, oder machten Einschränkungen bezüglich der Modulation, nämlich, dass die zulässigen Werte auf dem Einheitskreis liegen. Allgemein kann man natürlich auch die Phasendifferenz ϕ ermitteln, indem man die Abweichung zwischen dem empfangenen und dem idealen Konstellationspunkt misst (Abb. 9.9). Hierzu muss an einer Stelle in der Loop eine Entscheidung gefällt werden, welcher Zustand (mutmaßlich) gesendet wurde. Bei geringer Bitfehler- bzw. Symbolfehlerrate trifft dieser dann mit hoher Wahrscheinlichkeit den wirklich gesendeten Zustand. Es kann die Phasendifferenz zwischen beiden als Regelgröße für den VCO verwendet werden.
Vorteil dieses Verfahrens ist die Allgemeingültigkeit, man muss nur wissen, welche Modulation zugrunde lag, und das muss man sowieso wissen. Nachteil ist die lange und langsame Schleife, so steht im UMTS mit 3,84 Mchips (MSymbols pro Sekunde) nur alle 260 ns ein Schätzwert zur Verfügung. Das entspricht aber über 500 Perioden des 2 GHz-Trägers.

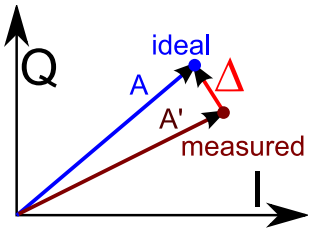


Abbildung 9.9: Abweichung eines gemessenen vom idealen (entschiedenen) Konstellationspunkt.

Rechnet man noch weitere Verzögerungen, die alle im Symboltakt erfolgen, hinzu, können durchaus einige tausend Perioden vergehen, ehe ein Regelsignal zur Phasenkorrektur vorhanden ist. Entsprechend darf natürlich die Phase in dieser Zeit nicht weggelaufen sein (bei QPSK - im UMTS) nicht mehr als $\pi/4$ in - abgeschätzt - 1000 Perioden, also 500 ns, was einer maximalen Frequenzabweichung von $1/(8 \times 500\,\text{ns}) = 250\,\text{kHz} = 125\,\text{ppm}$ des freilaufenden VCO vom Idealwert entspricht. Die Regelbreite ist also begrenzt, ggf. sind andere Verfahren zur Kanalschätzung hier erfolgreicher.

9.3 Taktrückgewinnung

Im gleichen Maße, wie die Phase des Trägers am Sender unbekannt ist, ist auch der ideale Entscheidungs- bzw. Abtastzeitpunkt unbekannt und muss rekonstruiert werden. Die Aufgabenstellung ist hier die gleiche, wie die der Trägerrückgewinnung. Deshalb, und weil es ansonsten zu weit führt, werden Verfahren hierzu derzeit nicht vorgestellt. Prinzipiell eignet sich eine Struktur, wie die Costas-Loop dazu, die Abtastphase zurückzugewinnen.

Kapitel 10

Kanalkodierung: Erweiterung für Funkkanäle

Die Notwendigkeit und die Grundlagen der Kanalkodierung sind schon in Kap. 5 dargelegt worden. In diesem Kapitel wird eine Erweiterung auf die vor allem bei drahtloser Übertragung wichtigen Faltungscodes geliefert. Natürlich sind die Codes zur Fehlerkorrektur allgemein auf beliebigen Kanälen an zu wenden, die Faltungscodes jedoch werden schwerpunktmäßig zur Datensicherung in Funkkanälen eingesetzt, jedenfalls im WLAN [46], und der Mobilkommunikation der zweiten (GSM) [47], dritten (UMTS) [48] und vierten (LTE) [49] Generation. In der fünften Generation hat man Faltungscodes fallen gelassen [50] und setzt stattdessen auf Polarcodes und LDPC (siehe Abschnitt 5.2.6.3). Als neue Codeklasse sollen die Faltungscodes, die auch lineare Codes sind, beleuchtet werden. Der Kodierer ist einfach über Schieberegister und Xor-Glieder zu implementieren (Abb. 10.1).

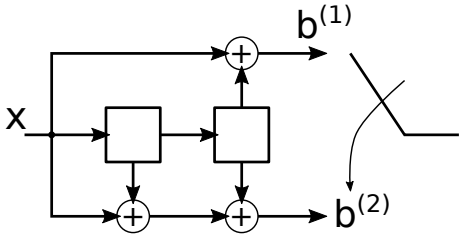


Abbildung 10.1: Unsystematischer Faltungscode $C(2,1,2)$

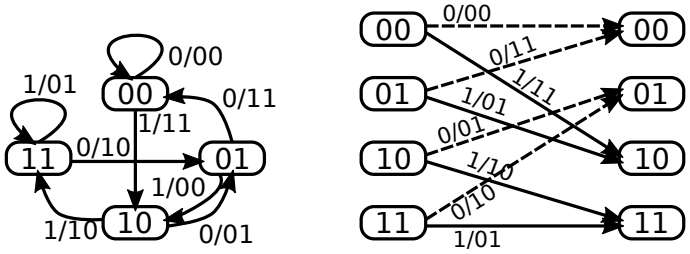


Abbildung 10.2: Zustandsdiagramm (links) und Übergangsdiagramm (rechts) für den Faltungscode $C(2,1,2)$ mit Polynomen $g^{(1)} = 5$ und $g^{(2)} = 7$.

10.1 Faltungscodes

In jedem Schritt werden zwei Bits b an den Ausgang gegeben, es wird dem Kodierer ein Bit x geliefert und es gibt zwei Speicher, so dass die Bezeichnung $C(2,1,2)$ zustande kommt. Die Ausgangsbits sind logische Verknüpfungen des Eingangs mit den Speicherstellen, mit den Generatorpolynomen

$$g^{(1)}(x) = x^2 + 1 \tag{10.1}$$

$$g^{(2)}(x) = x^2 + x + 1 \tag{10.2}$$

$$\tag{10.3}$$

auch dargestellt als die Zahlen $g^{(1)} = 5$ und $g^{(2)} = 7$. Die zwei Speicherstellen bedeuten, dass der Kodierer vier unterschiedliche Zustände aufweist. Leicht kann man die verschiedenen Kombinationen durchprobieren und kommt so auf eine Darstellung im Zustandsdiagramm, oder im Übergangsdiagramm nach Abb. 10.2, an der man die Ausgaben des Kodierers in Abhängigkeit vom Eingangsbit und dem Zustand unmittelbar ablesen kann. Die Bezeichnung Faltungscodes rührt daher, dass der Ausgang eines Kodierers als eine Faltung des Eingangssignals mit dem jeweiligen Generatorpolynom geschrieben werden kann:

$$c^{(i)}(x) = d(x) * g^{(i)}(x)\,. \tag{10.4}$$

Der Code kann somit fortlaufend auf unendliche Sequenzen angewandt werden, was in der Regel aber nicht geschieht, sondern man verwendet auch hier Blöcke. Der Vorteil der Faltungscodes liegt darin, dass sie durch geringe Coderaten (in dem o.g. Beispiel hat man die Rate $r = 1/2$) eine hohe Möglichkeit der Fehlerkorrektur bieten. Im Mobilfunk

werden sie oft zur Fehlerkorrektur als innerer Codes (also direkt am Kanal) eingesetzt, und Blockcodes werden zur Restfehlererkennung nachgeschaltet.
Da auch Faltungscodes lineare Codes sind, kann man sie als lineare Verknüpfung, also Matrixgleichung schreiben. Bei dem o.g. Beispiel mit zwei Ausgängen pro Eingangsbit werden die Generatorfolgen untereinander und dann jeweils um eine Stelle versetzt aufgeschrieben. Die Generatorfolgen selbst sind die Impulsantworten des Kodierers,

$$\vec{g}^{(1)T} = 1\,0\,1\,... \tag{10.5}$$

$$\vec{g}^{(2)T} = 1\,1\,1\,... \; . \tag{10.6}$$

Die theoretisch unendlich lange Folge von Nullen wurde einfach durch die ... angedeutet. Die Länge des Impulses, auch Constraint-Length genannt, ist also drei, oder allgemein einer mehr als die Anzahl der Speicher. Die Generatormatrix ist also

$$\vec{b}^T = \vec{x}^T \overline{\overline{G}} \tag{10.7}$$

$$\overline{\overline{G}} = \begin{pmatrix} \vec{g}^{(1)T} & ... & & \\ \vec{g}^{(2)T} & ... & & \\ \vdots & \vec{g}^{(1)T} & ... & \\ & \vec{g}^{(2)T} & ... & \\ & \vdots & \vec{g}^{(1)T} & ... \\ & & \vec{g}^{(2)T} & ... \\ & & \vdots & \ddots \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 & 1 & ... & & \\ 1 & 1 & 1 & 1 & ... & & \\ \vdots & 1 & 0 & 1 & 1 & ... & \\ & 1 & 1 & 1 & 1 & ... & \\ & \vdots & 1 & 0 & 1 & 1 & ... \\ & & 1 & 1 & 1 & 1 & ... \\ & & \vdots & & & & \ddots \end{pmatrix} \tag{10.8}$$

Praktisch verwendet wird diese Darstellung eigentlich nicht, sie zeigt aber den Zusammenhang zu den linearen Codes aus vorhergehenden Kapiteln. Kodierung und vor allem die Dekodierung kann mit den Übergangsdiagrammen, die man zu einer Kette zu sogenannten Trellis-Diagramm ergänzt. Trellis steht für Spalier und soll das Aussehen beschreiben. Die Codes werden danach auch manchmal Trellis-Codes genannt. Insgesamt sorgt das natürlich für eine Namensverwirrung.
Mit dem Trellisdiagramm kann man eine Bitfolge kodieren (Abb. 10.3). Man sieht auch den Aufbau des Trellisdiagramms, in dem die Übergänge nach und nach gezeigt sind. In den Zeilen sieht man die einzelnen Zustände und die Spalten stellen die Zeitpunkte dar. Zwischen den Zuständen werden die Übergänge gezeigt und dabei, man beobachte den roten Pfad, auch die Ausgaben gezeigt sind. Die Implementation würde man wohl kaum so gestalten, sondern auf die einfache Realisierung mit Schieberegistern (Abb. 10.1) zurück greifen.

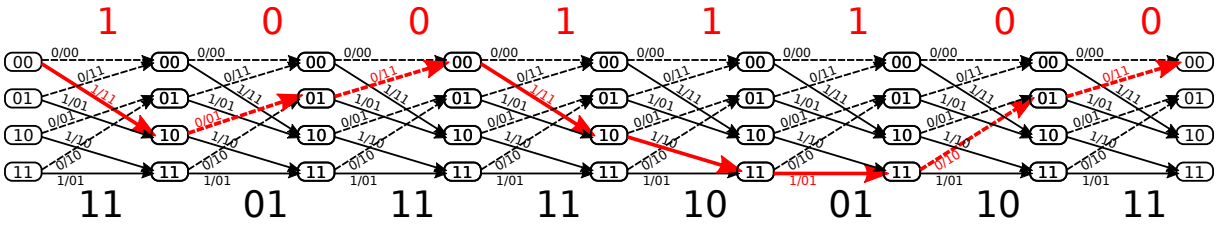


Abbildung 10.3: Trellisdiagramm mit Beispiel zur Kodierung.

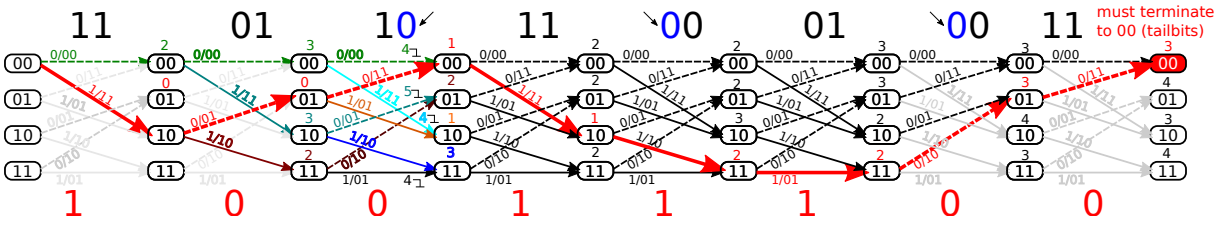


Abbildung 10.4: Trellisdiagramm mit Beispiel zur Dekodierung.

Zur Dekodierung ist so ein Trellisdiagramm aber gut geeignet (Abb. 10.4). Die empfangene Signalfolge (hier nun oben) ist Eingang in das Diagramm. Von Schritt zu Schritt wird nun für jeden Übergang die Anzahl der Fehler ermittelt. Das ist hier einfach die Anzahl der Nicht-Übereinstimmungen zwischen Empfangsfolge und Übergangsfolge. Die Anzahl der Fehler wird dem jeweiligen Übergang zugeordnet und an den Zielzustand geschrieben. Da im ersten Schritt im Nullzustand gestartet wurde, sind hier einige Übergänge ausgegraut, da sie gar nicht stattgefunden haben können. Im zweite Schritt wird das Verfahren wiederholt und es werden dem Zielzustand die Anzahl der Fehler zugeordnet, die sich auf dem Pfad bis zu ihm insgesamt (also in Summe) ereignet haben müssten. Bitfehler (blau, bzw. mit Pfeil) werden zunächst nicht gesondert beachtet, sondern es werden einfach die kumulierten Fehler weiter gezählt.

Irgendwann werden unweigerlich zwei mögliche Pfade auf einen Zustand treffen. Es wird danach nur noch der Pfad weiter verfolgt, der den geringeren Fehler bis zu diesem Punkt hatte. Das macht auch Sinn, denn ab da kann der schlechtere Pfad gar nicht mehr auf holen, beide sind ja im gleichen Zustand. Haben zwei Pfade die gleiche Anzahl Fehler, nimmt man einfach einen der beiden. Es ist bis hier nicht unterscheidbar, welcher Pfad der bessere war, jedenfalls nicht, wenn man nur hart entschiedene Eingangsfolgen verwendet. Besondere Beachtung verdient noch das Ende des Trellis. Diese Bitfolge ist bewusst so gewählt, dass sie im Nullzustand endet. Das macht man praktisch häufig so, aber es gäbe noch Alternativen. Wie am Anfang sind nun einige Übergänge ausgegraut, da sie gar nicht stattgefunden haben können, sie führten ja nicht in den Nullzustand. In diesem Beispiel sieht man auch den Vorteil: Hätte man die Folge nicht in Null terminiert, so wäre auch

der Zustand 10 mit gleicher Fehlerzahl erlaubt gewesen. Die dekodierte Bitfolge erhält man nun, wenn man den Pfad mit geringstem Fehler zurück geht, er ist hier wieder in rot und fett gedruckt gezeigt.
Alternativ zur Terminierung im Nullzustand kann man natürlich den Code auch einfach abschneiden, dass das zu Problemen führt, ist schon gezeigt, oder man kann den Code so strukturieren, dass er im Endzustand startet. Das sind sog. Tail-Biting Codes. Hierfür muss man aber schon am Anfang das Ende des Blockes kennen und der theoretische Vorteil eines fortlaufenden Codes wäre dahin.
Für die Rückverfolgung muss man im Vorwärtsprozess für jeden erreichten Zustand speicher, welche Anzahl Fehler vorgefunden wird, und aus welchem vorhergehenden Zustand der Pfad kam. Der gesamte Algorithmus nennt sich nach seinem Erfinder Viterbi Algorithmus.
Im oben gezeigten Beispiel wurden als Eingang zum Viterbi Algorithmus hart entschiedenen Bitfolgen verwendet, so dass dieses der Hard-Decision Viterbi ist. Stattdessen können auch differenziertere Eingangsgrößen verwendet werden, so dass man den Soft-Decision Viterbi vorliegen hat. In Abb. 10.5 ist ein solcher Schritt dargestellt. Neben der empfangenen Bitfolge wird die Wahrscheinlichkeit der Bits als Eingang verwendet. Man kann hier z.B. verwenden

$$p(b) = p(0) - p(1) \tag{10.9}$$

die Wahrscheinlichkeit mit der man meint, eine Eins wurde gesendet wird also von der Wahrscheinlichkeit der Null abgezogen. Die Metrik für den Übergang berechnet man nach

$$m_i = (-1)^{b_i^{(1)}} \cdot p(b_i'^{(1)}) + (-1)^{b_i^{(2)}} \cdot p(b_i'^{(2)}) \tag{10.10}$$

oder im Klartext: Das jeweilige Bitvertrauen wird positiv gezählt, wenn die dazu assoziierte Übergangssequenz eine Null war, sonst negativ. Die gesamte Metrik für den Übergang ist dann die Summe daraus und die Metrik für den erreichten Zustand ist wieder die gesamte negative Summe (Differenz) aller Übergänge bis des Pfades bis hier her. Der Vorteil ist natürlich wieder, dass mehr und feinere Informationen berücksichtigt werden können, so hier zum Beispiel, dass die fehlerhafte Null mit einem nur geringen Vertrauen bewertet wird, was praktisch wohl der Fall sein wird.

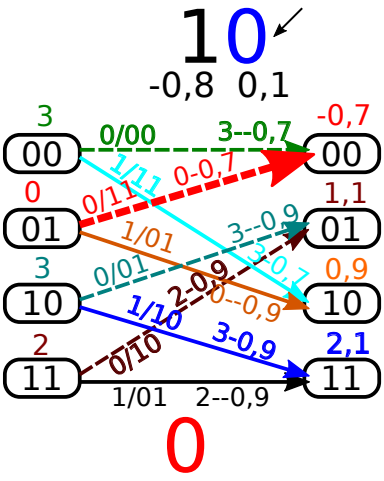


Abbildung 10.5: Ein beispielhafter Schritt für Soft-Decision Viterbi Dekodierung.

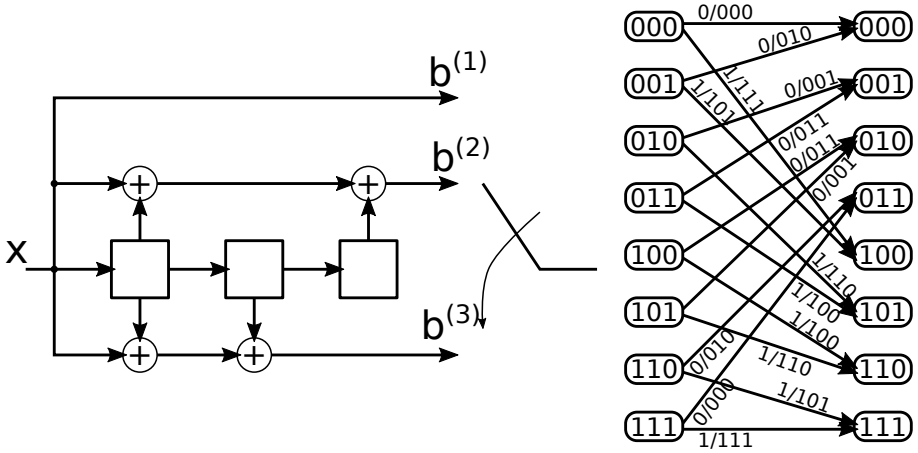


Abbildung 10.6: Faltungscode $C(3,1,3)$ mit $g^{(1)} = 10_8$, $g^{(2)} = 15_8$, $g^{(1)} = 16_8$ (jeweils oktal angegeben) mit Schiebregisterrealisierung (links) und Übergängen (rechts).

10.1.1 Weitere Beispiele für Faltungscodes

Vorstehend wurde nur ein sehr einfacher, kurzer und praktisch nicht verwendeter Faltungscode erläutert, bzw. als Beispiel verwendet. Dieser hatte zwei Speicherstellen, also eine Constraint Length von drei, zwei Ausgänge bei einem Eingang, und er war unsystematisch. Weiterhin war aufgrund der Struktur die Impulsantwort endlich, man hat es also mit einem Finite Impulse Response (FIR)-System zu tun.

Diese Parameter kann man alle anders wählen. Gerne verwendet werden systematische Faltungscodes (Abb. 10.6): Hier ist eines der Codebits das Datenbit und damit ist dem Codes unmittelbar die Bitfolge zu entnehmen. Der dargestellte Code hat eine Constraint Length von vier mit Impulsantwort 100 011 001 010 und eine Coderate von $r = 1/3$.

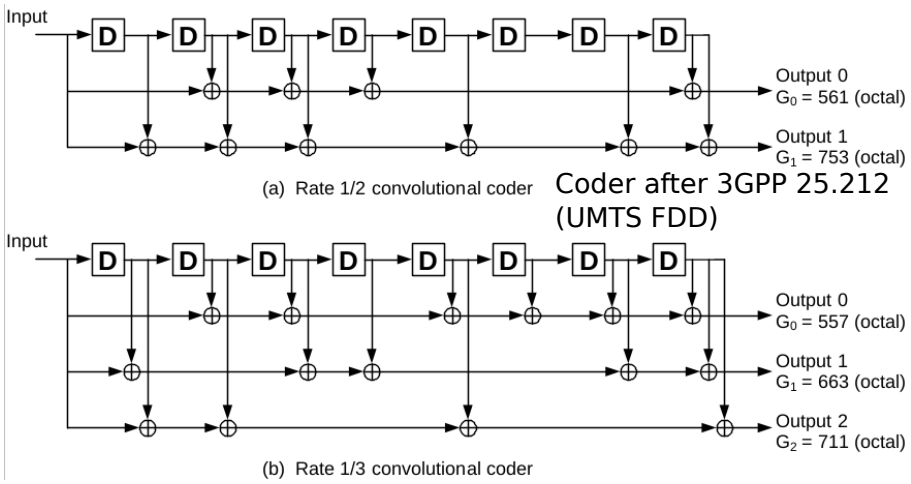
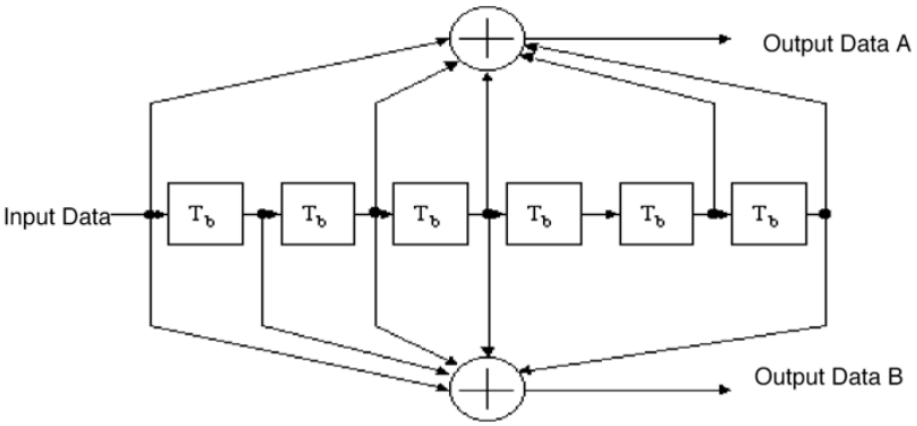


Abbildung 10.7: Faltungscodes nach [48] für das UMTS.

Im UMTS werden die in Abb. 10.7 verwendet, die beide unsystematisch mit acht Speicherstellen sind. Damit liegen 256 Zustände vor. Je nach Kanalbedingungen oder Kanal kann der Code mit der Rate $r = 1/2$ oder der mit der Rate $r = 1/3$ verwendet werden. Bislang sind nur Codes behandelt worden, die eine Rate von $r = 1/n$ zulassen. Dieses ist nicht immer praktikabel. Um andere Coderate zu erzielen kann man auf zwei Arten vorgehen: Entweder man verwendet in jedem Schritt mehrere (m) Eingangsbits bei den n Ausgängen, so dass man Raten von $r = m/n$ bekommt. Natürlich macht nur $r < 1$ also $n > m$ Sinn. Oder als Alternative punktiert man den Code, d.h. es werden nach fest vorgegebenem und in Sender und Empfänger gleichermaßen bekanntem Schema Bits weggelassen. Diese Lücken werden im Empfänger wieder aufgefüllt, entweder als Dummy-Bits, oder mit Werten sehr hoher Unzuverlässigkeit, denn letztlich sind punktierte Bits auch nur Übertragungsfehler. Und der Rest des Codewortes kann dann durch den Viterbi Algorithmus rekonstruiert werden. Im WLAN [46] verwendet man genau dieses Vorgehen, um Coderaten von $R = 1/2$, $2/3$, $3/4$ zu erzielen. Abb. 10.8 zeigt den Basiscode und in Abb. 10.9 ist genau das Schema der Punktierung mit gestohlenen und zugesetzten Stellen erläutert.

10.1.2 Leistungsfähigkeit von Codes

In Abschnitt 5.2.1 wurde die Hammingdistanz als Maß für die Fehlererkennungs- und Korrektureigenschaften der Blockcodes eingeführt. Ein ähnliches Maß soll nun auch für Faltungscodes entwickelt werden. Hier heißt das dann die „freie Distanz“ und sie lässt sich etwas schwieriger berechnen. Es wird das Übergangsdiagramm Abb. 10.2 oder das



after WLAN (IEEE 802.11-2016)

Figure 17-8—Convolutional encoder (k = 7)

Abbildung 10.8: Faltungscode nach [46] für das WLAN.

Trellisdiagramm Abb. 10.3 des sehr einfachen Beispielcodes betrachtet. Da man weiß, dass der Nullvektor ein gültiges Codewort ist, muss man nun unter allen Möglichen den Vektor finden, der den kleinsten Abstand von diesem Nullvektor hat. Bei der zu betrachtenden Sequenz würde man also im Nullzustand starten und muss ihn mindestens einmal verlassen. Das geht nur mit einer Eins in den Zustand 10. Es ist klar, dass bei 2^k Zuständen (also hier vier) man höchstens jeden Zustand einmal durchlaufen muss, um wieder den Nullzustand zu erreichen, denn ansonsten ist man in einer Schleife, und die könnte man unterlassen, wenn man die wirklich kürzeste Folge sucht.
Möglcihe Folgen von und zum Nullzustand sind also

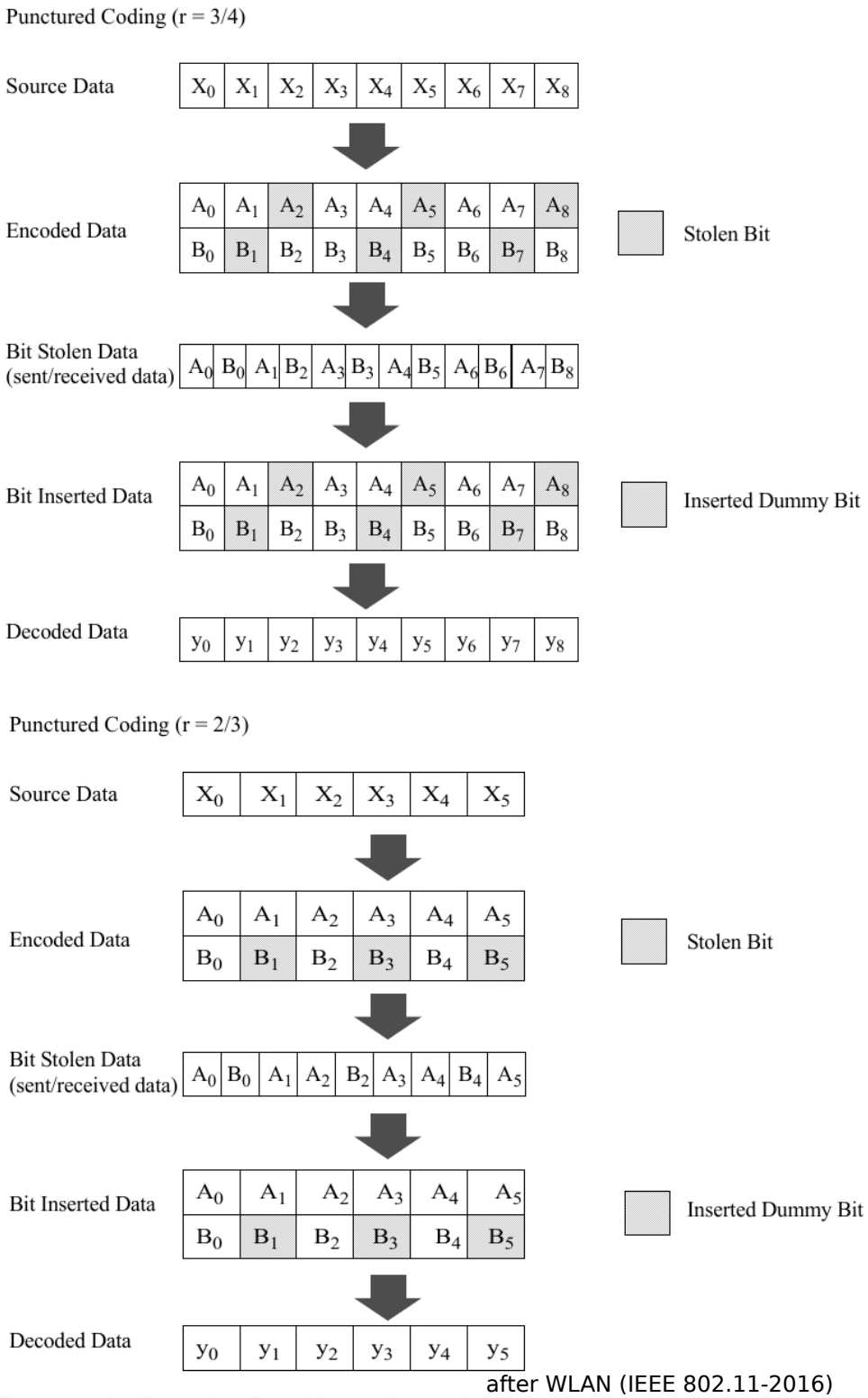
$$00 \overset{11}{\to} 10 \overset{01}{\to} 01 \overset{11}{\to} 00 \tag{10.11}$$

$$00 \overset{11}{\to} 10 \overset{10}{\to} 11 \overset{01}{\to} 01 \overset{11}{\to} 00 \tag{10.12}$$

eine Sequenz führt zu fünf Einsen, die andere zu sechs, und damit haben die jeweils kürzesten Vektoren, die nicht der Nullvektor sind, das Gewicht fünf.

10.1.3 Turbocodes und IIR-Faltungscodes

Ineinander verschachtelte Codes werden in der Mobilkommunikation häufig verwendet. Bislang waren das eine Cyclic Redundancy Check (CRC)-Überprüfung als äußerer Code, die Restfehler erkennen sollte und einen evtl. fehlerhaften Rahmen verwerfen sollte, und ein anderer, oft Faltungscode als innerer Code. Dieser sollte die Fehlerrate der Übertragung möglichst senken. Solche verschachtelten Codes bringen gegenüber einem einfachen



after WLAN (IEEE 802.11-2016)

Figure 17-9—Example of the bit-stealing and bit-insertion procedure (r = 3/4, 2/3)

Abbildung 10.9: Punktierungsschema nach [46] für das WLAN.

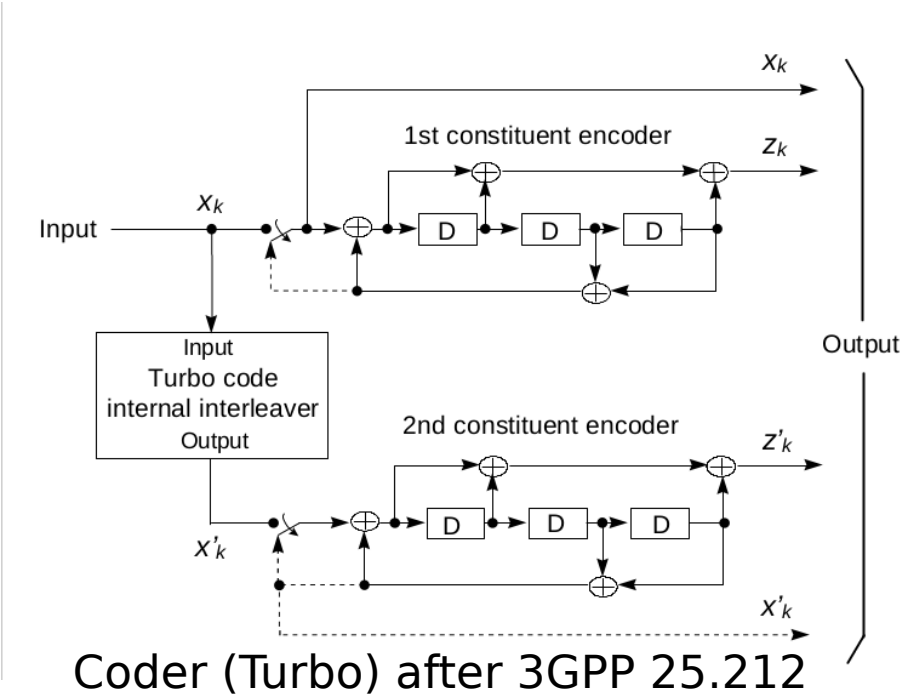


Abbildung 10.10: Turbo-Kodierer für UMTS nach [48] (Gestrichelte Linien zur Terminierung des Codes).

Code Vorteile. Bei den Turbocodes verschachtelt man nun zwei Faltungscodes miteinander (Abb. 10.10).
Es werden zwei Faltungskodierer verwendet, die zwar wenig Speicherelemente (drei) haben, aber durch die Rückkkopplung eine unendlich Impulsantwort aufweisen. Diese einzelnen Kodierer nennt man auch Infinite Impulse Response (IIR)-Faltungskodierer. Weiterhin arbeiten die beiden Kodierblöcke parallel. Man hat es also mit einem Parallel Concatenated Convolutional Code (PCCC) zu tun (Gegensatz wäre Hintereinanderschaltung als seriell verketteter Code). Der Eingang des zweiten Blockes ist aber gegenüber dem ersten verwürfelt mit einem Interleaver. Das heißt, die Information des gleichen Eingangsbits tritt am Ausgang drei Mal auf: Einmal direkt, weil es ein systematischer Code ist, einmal recht schnell, weil ein Kodierer folgt und einmal zeitlich weit entfernt, weil zunächst der Interleaver die Bits möglichst zufällig und möglichst weit voneinander entfernt. In der Annahme, dass der Kanal nur kurzzeitig schlecht wird, kann man darauf hoffen, dass ein oder zwei Einflüsse des Eingangsbits gestört werden, aber niemals alle drei und damit kann man die eigentliche Information immer zurück gewinnen.
Die Dekodierung erfolgt wiederum in einem iterativen Prozess: Zuerst werden die Informationen der Kodierung des ersten Coders ausgewertet, diese werden als Zusatzinforma-

tionen in die Kodierung des zweiten eingespeist und nun wird diese Ausgangsinformation nochmal in den ersten Decoder gegeben und so weiter. Irgendwer hat nun dieses mit dem Turbolader eines Verbrennungsmotors verglichen, wo der Abgasstrom genutzt wird, eine Turbine an zu treiben, um in den Motor am Einlass mehr brennbares Gemisch hinein zu pumpen. Daher rührt der marketingtechnisch geschickt gewählte Name, der Prozess ist überhaupt nicht vergleichbar.
Turbocodes wurden erst Anfang der 1990er Jahre gefunden, wurden zurecht viel beachtet und haben quasi sofort Einzug in den 3GPP-Standard für UMTS gefunden. Warum? Turbocodes erreichen bis auf wenige Zehntel dB die Shannongrenze. Heute kennt man weitere solcher Codes, weshalb die Turbocodes wieder auf dem Rückzug sind. Der Grund hierfür ist, dass der iterative Prozess und vor allem der Interleaver, der Bits zeitlich lange voneinander trennen soll, zu großen Latenzzeiten führt, die man für Echtzeit- und zeitkritisch Kommunikation nicht gebrauchen kann.

10.2 Interleaver

In Abschnitt 7.2 und insbesondere in Abschnitten 7.2.1 und 7.2.2 wurde verdeutlicht, dass Mobilfunkkanäle zeitlichen Änderungen unterliegen, und es sein kann, dass ein Kanal kurzzeitig ein schlechtes Signal-Rauschverhältnis präsentiert, oder sogar ganz verschwindet. Damit ergibt sich eine hohe Wahrscheinlichkeit, dass benachbarte Symbole betroffen werden. Es treten sogenannte Büschelfehler auf. Dieses ist für einen Kanalkodierer üblicherweise ein ungünstiger Fall, besser wäre es, wenn Fehler möglichst weit voneinander entfernt auftreten und möglichst unkorreliert sind. Dieses erreicht man am einfachsten dadurch, dass man die Nachbarschaftsverhältnisse aufbricht und die Symbole einfach zeitlich besser verteilt. Eine einfache Variante ist der Zeilen-, Spalten-Interleaver, der in Abb. 10.11. Während ohne Interleaver das Ergebnis „Kanalkodierung behebt XXXXXXX“ wäre, bekommt man mit Interleaver „KanalkXdieruXX beheXX FehlXX“, was mit gutem Willen ein Mensch noch wieder richtig rekonstruieren kann. Bei den Maschinen ist es ähnlich.
Über den Zeilen-Spalten Interleaver hinaus hat man verschiedene Methoden untersucht. Am besten arbeiten solche Verwürfelungsmechanismen, die die benachbarten Symbole möglichst weit von einander trennen, und tatsächlich ein zufallsähnliches Muster erzeugen. Insbesondere die weite Trennung erfordert große bzw. lange Blöcke, die gelesen werden. Dieses führt zu langen Latenzzeiten, denn der Block muss erst einmal zusammengestellt werden, bevor er gesendet werden kann und im Empfänger muss man den Block zunächst komplett empfangen, bevor der eigentliche Fehlerkorrekturmechanismus anspringen kann.

K	a	n	a	l	k	o
d	i	e	r	u	n	g
	b	e	h	e	b	t
	F	e	h	l	e	r

K	a	n	a	l	k	o
d	i	e	r	u	n	g
	b	e	h	e	b	t
X	X	X	X	X	X	X

K	d		
a	i	b	F
n	e	e	e
a	r	h	h
l	u	e	l
k	n	b	e
o	g	t	r

K	d		
a	i	b	F
n	e	e	e
a	r	h	h
l	u	e	l
k	X	X	X
X	X	X	X

Abbildung 10.11: Zeilen Spalten Interleaver und verschiedenen Fehlerauswirkungen, Übertragung erfolgt immer von links nach rechts und dann Zeile für Zeile, in beiden Fällen sind die letzten sieben Symbole fehlerhaft.

Gerade für Echtzeit und Machine-Machine Kommunikation (Stichwort Ultra Reliable Low Latency Communication (URLLC)) ist das nicht wünschenswert. Als ein Beispiel aus einer Zeit, in der man noch langsame Datenübertragung gewohnt war, ist in Abb. 10.12 [47]. Hier werden 456 zusammenhängende Bits (jeweils gleiche Farbe) auf dann insgesamt 22 Rahmen verteilt. Bei einer Rahmenlänge von 577 μs kommt man also auf eine minimale Latenz von über 25 ms (beinhaltet die Wartezeit in Sender und Empfänger) nur für die Verteilung der Symbole. Auf der horizontalen Achse ist die Position des Symbols in einem Rahmen aufgetragen und man sieht, dass zu Beginn und Ende (Rahmen 1 und 22) jeweils 6 Bit gleichverteilt sind, dann (2 und 23) 12, in 3 und 22 18 und schließlich 24 Bit. Die Übertragung erfolgt Zeile für Zeile von links nach rechts. Es sei darauf hingewiesen, dass Latenzzeit und Übertragungsrate bzw- geschwindigkeit zwei unterschiedliche Größen sind, die bei der Betrachtung hier wenig miteinander zu tun haben.
Entscheidend ist die Auswahl des richtigen Interleaver bei den Turbo-Codes (Abschnitt. 10.1.3), die aber in diesem Skript nur kurz angesprochen wurden. Daher wird dieses Thema hier auch nicht weiter vertieft.

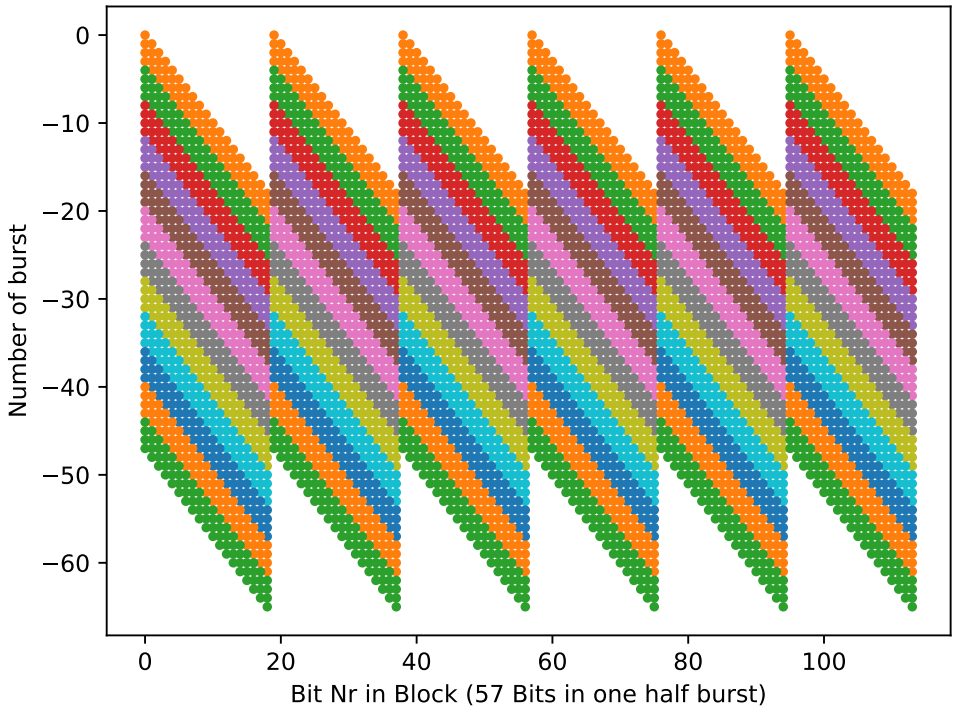


Abbildung 10.12: Interleavermuster für den GSM FullRate 12,2kbps Dienst nach [47] gleiche Farben gehören zu einem ursprünglichen 456 Bit-Paket

Kapitel 11

Kanal-Zugriffstechniken und Spektrumsausnutzung

11.1 Einführung und Überblick

Soll ein Medium von mehreren Sendern und Empfängern mehr oder weniger gleichzeitig genutzt werden, sollen also in dem Medium mehrere Kanäle aufgebaut werden, so ist eine Verabredung zum Zugriff auf den Kanal notwendig. Dieses gilt schon bei einfachster (menschlicher) Sprachkommunikation. Mehrere Gespräche können gleichzeitig nicht stattfinden. Lösungen sind, dass man in verschiedene Räume geht und die Türen schließt (SDMA), oder erst dann spricht, wenn der Kanal frei ist (TDMA mit Carrier Sense Multiple Access (CSMA)), also gerade von niemandem gesprochen wird. Hält man letzteres nicht ein, so kommt es zu einer Kollision, die Gespräche sind beide nicht mehr zu verstehen, ggf. bittet der Empfänger um nochmalige Aussendung der Information (Re-Transmission), es handelt sich um das ALOHA-Protokoll. Natürlich läßt man einander ausreden und antwortet erst, wenn der Partner ausgesprochen hat (Time Division Duplex (TDD)). Im Orchester oder Chor sind einzelne Stimmen herauszuhören, wenn sie sich in der Frequenz (deutlich) unterscheiden, wie z.B. Sopran und Bass (FDMA), oder sich in der Klangfarbe unterscheiden, wie z.B. Cello und Posaune, die zwar im gleichen Frequenzbereich liegen, aber sehr unterschiedlich klingen (CDMA). Gerade bei letzterem ist gute Synchronisation von Vorteil.
Diese trivial anmutenden nicht-technischen Erläuterungen werden nun auf eine technische Ebene gehoben und dort genauer untersucht.

f/MHz	RDS	Sender	P_S/kW	Entf./km	Plg.
88.8	SWR2 BW	Heidelberg/Königstuhl, SWR	100	51	29°
89.9	SWR1 RP	Weinbiet	25	46	335°
90.4	KLASSIK	Karlsruhe/EnBW-Kraftwerk	2	6	277°
92	SWR2 RP	Donnersberg	60	77	334°
92.9	SWR1 BW	Langenbrand	5	28	141°
93.5	SWR1 BW	Hornisgrinde/SWR	80	46	197°
94	SWR4 OG	Hornisgrinde/SWR	5	46	197°
95.9	SWR4 LU	Weinbiet	25	46	335°
96.2	SWR2 BW	Hornisgrinde/SWR	80	46	197°
96.6	DKULTUR	Wörth	0.2	8	303°
97	SWR4 KA	Ettlingen	8	8	157°
97.8	SWR1 BW	Heidelberg/Königstuhl, SWR	100	51	29°
98.4	SWR3	Hornisgrinde/SWR	80	46	197°
99.1	SWR1 RP	Donnersberg	60	77	334°
99.9	SWR3	Heidelberg/Königstuhl, SWR	100	51	29°
100.4	REGNBOGN	Hornisgrinde/SWR	80	46	197°
100.7	ENERGY	Güglingen	20	45	86°
101.1	SWR3	Donnersberg	60	77	334°
101.8	neue	Karlsruhe-Grünwettersbach	25	7	138°
102.2	SWR2 RP	Weinbiet	25	46	335°
102.8	REGNBOGN	Heidelberg/Königstuhl, DTK	10	51	29°
103.6	RPR1.	Kalmit	25	41	328°
104.1	SWR4 MA	Heidelberg/Königstuhl, SWR	50	51	29°
104.8	Querfunk	Karlsruhe-Grünwettersbach	1	7	138°
105.2	big FM	Langenbrand	20	28	141°
105.6	SWR4 KL	Donnersberg	60	77	334°
106.3	DLF	Hornisgrinde/SWR	80	46	197°
106.7	big FM	Kalmit	25	41	328°

Tabelle 11.1: UKW-Sendertabelle (Auszug) für den Raum Karlsruhe (alle horizontal polarisiert) mit höheren prognostizierten Empfangsleistungen nach http://fmscan.org/ (Dez. 2011)

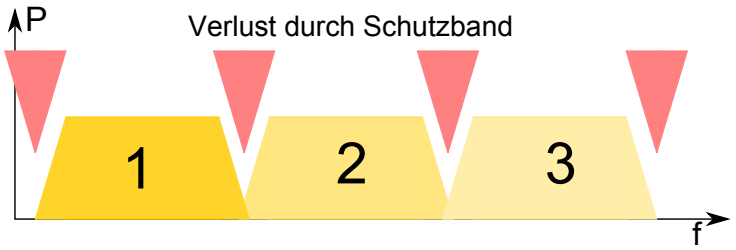


Abbildung 11.1: Schematisch: Frequenzbänder

11.2 Frequency Division Multiple Access

Wie im Kapitel 6 dargestellt, ist es ein Leichtes, ein Signal auf eine andere Frequenz zu modulieren bzw. umzusetzen. Dieses nutzt man technisch aus, indem man verschiedenen Kanälen einfach unterschiedliche Frequenzen zuweist, man spricht hier vom FDMA. Täglich nutzt dieses jeder beim UKW-Rundfunk (Tab. 11.1), und es ist selbstverständlich, dass für den Empfang eines anderen Rundfunkprogramms die Empfangsfrequenz des Empfängers umgestellt werden muss. Nach Frequenznutzungsplan (Teilplan 185, Eintrag 185003) der Bundesnetzagentur für diesen Frequenzbereich von 87,5 bis 108 MHz liegen hier ein Kanalraster von 100 kHz und eine Kanalbreite von 300 kHz vor. Jedenfalls in obiger Tabelle ist dieses eingehalten.
Dasselbe Zugriffsverfahren wird bei vielen weiteren Diensten verwendet, so unterscheiden sich die terrestrischen wie auch die vom Satelliten ausgestrahlten Fernsehprogramme ebenfalls in der Frequenz. Bei analoger Ausstrahlung ist dieses neben der Polarisation der Aussendung (Satelliten-TV) die einzige Unterscheidung, die neueren digitalen Übertragungstechniken ermöglichen eine größere Vielfalt von Unterscheidungskriterien. Auch beim Kabel-TV (DVB-C oder Analog) findet man die “Sender” bei unterschiedlichen Frequenzen bzw. Kanälen. Die ADSL-Techniken nutzen dieses ebenso, indem dem eigentlichen ISDN oder POTS (analog)-Signal das Datensignal im höheren Frequenzbereich überlagert wird.
Die Frequenz und letztlich das Frequenz-Multiplexing ist die dominierende Methode bei der Frequenzregulierung der Bundesnetzagentur, die in ihren Frequenznutzungsplänen sämtliche Dienste zwischen 9 kHz und 275 GHz nach ihren Frequenzbändern sortiert und so auch reguliert.
Es ist technisch sehr schwierig und erfordert ein hohes Maß an Synchronität, einzelne Funkdienste oder auch nur unterschiedliche Programme und damit unterschiedliche Kanäle nahtlos ineinander über gehen zu lassen. Daher sind neben jedem Frequenzkanal Schutzbänder erforderlich, es wird ein Teil des Spektrums verschwendet (Abb. 11.1). So

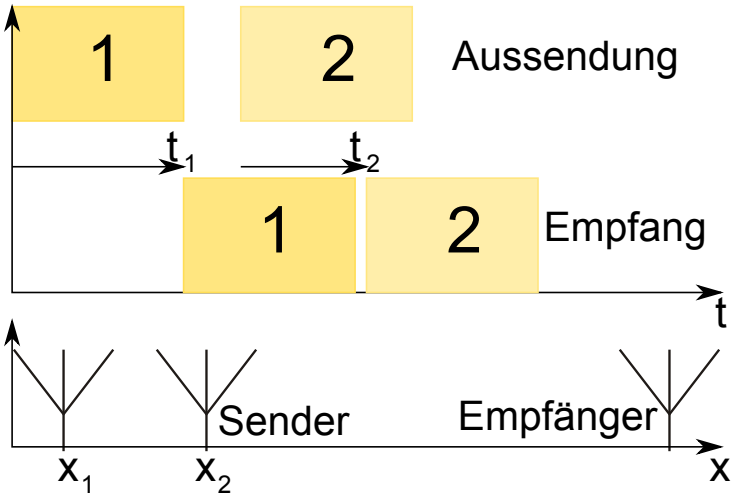


Abbildung 11.2: Prinzip des TDMA

wird effektiv das nutzbare Spektrum verringert, bzw. eben nicht voll ausgenutzt. Eine genauere Betrachtung wird beispielhaft am GSM-System weiter unten ausgeführt. Die Hauptursache besteht in den endlich ausgedehnten Spektren modulierter Träger. Die Spektren können mit vertretbarem Aufwand nur endlich stark zu den Seiten hin unterdrückt, ausgefiltert werden. Zwar sind steile Filter möglich, doch stellen sie einen sehr hohen Aufwand und damit hohe Kosten dar. Auch erfordern, wenn man sich eine ideale Puls und Spektrumsformung im digitalen Bereich vorstellt, die gewünschten steil abfallenden Flanken unrealistisch lineare Verstärker, die deshalb auch noch eine sehr schlechte Leistungseffizienz aufweisen müssen.

11.3 Time Division Multiple Access

Eine Alternative zum FDMA ist die Zuweisung einer Ressource für einen bestimmten Zeitschlitz, man spricht dann vom TDMA. Dieses kommt wahrscheinlich der menschlichen Kommunikation am nächsten, es wird nur gesprochen, wenn die anderen schweigen. Diese Zugriffsvariante hat sich historisch für analoge Übertragung von kontinuierlichen Datenströmen (meist Sprache oder bewegte Bilder) nicht angeboten, denn man hätte diese Ströme dann zerhacken müssen, und wer hätte das ertragen? Im Zeitalter der digitalen Kommunikation sieht dieses grundsätzlich anders aus, hier kann man sogar kontinuierliche Ströme zerhacken, in Paketen senden und später wieder zu einem Strom zusammen setzen. Einzig die durchschnittliche Datenrate muss stimmen.
In Abb. 11.2 ist das TDMA-Prinzip verdeutlicht und ist dargelegt, dass auch hier ein

Schutzintervall erforderlich ist. In einem System mit hier bspw. zwei Sendern in unterschiedlichem Abstand von einem Empfänger, wird dem ersten Sender zuerst die Aussendung erlaubt. Dieser befindet sich in einer weiteren Entfernung vom Empfänger als Sender zwei, der zu einem späteren Zeitpunkt senden darf. Das erste Signal hat also eine längere Laufzeit als das Zweite; die Signale erscheinen daher am Empfänger näher aneinander gerückt. Für jeden Meter benötigt die elektromagnetische Welle 3,33 ns, so dass bei einem Kilometer Entfernungsdifferenz ein Zusammenrücken um ca. 3 μs stattfindet. Um Überschneidungen zu vermeiden, muss auch hier ein Schutzintervall eingefügt werden. Dieses richtet sich nach den obigen Ausführungen danach, welche maximale Entfernungsdifferenz zwischen zwei Sendern zu erwarten ist. Hinzukommen aber noch Einflüsse dadurch, dass die beiden Sender natürlich nicht genau synchron laufen und es daher zu Unschärfen in den Aussendungen kommt, und in der Regel die Sender ihre Endverstärker zunächst einschalten und hochfahren müssen, bzw. am Ende wieder definiert abschalten. Auch dieses ist mit Unsicherheiten verbunden und benötigt Zeit, die sich in der Definition eines Schutzintervalls niederschlägt. In der Mobilkommunikation kommen Mehrwegeausbreitungen hinzu, durch die ein Signal direkt oder auch über Umwege seinen Weg zum Empfänger findet. Dieses ist Bestandteil von Kap. 7.2.1 und führt ebenfalls zur Notwendigkeit eines Schutzintervalls.
Wie auch im FDMA kann auch hier der Kanal durch den Schutz nicht zu 100 % ausgenutzt werden.

11.4 Access gegenüber Duplex

In vielen Kommunikationssystemen sind bidirektionale Kanäle erforderlich. Hier kommunizieren zwei Stationen in beide Richtungen miteinander. Prinzipiell sind also auch zwei Kanäle erforderlich, deren Endpunkte Sender und Empfänger aber örtlich sehr dicht beieinander realisiert werden müssen. Praktisch nimmt dieses also eine Sonderstellung ein. Das einfachste, bei der kabelgebundenen Kommunikation ist natürlich, beide Kanäle durch unterschiedliche Kabel zu trennen. Dieses geschieht im SDH/ PDH sowohl elektrisch als auch optisch durch dediziert für eine Richtung vorhandene LWL oder Koaxkabel, oder auch - z.B. im Ethernet (Twisted Pair) - durch die Zuweisung eines Adernpaares für jede Richtung.
Aber auch hier sind die o.g. Varianten FDMA und TDMA möglich, indem entweder zu einem anderen Zeitpunkt Station 1 sendet und 2 empfängt und umgekehrt (man spricht von TDD) oder eben beide Aussendungen auf unterschiedlichen Frequenzen mit einem deutlichen Schutzband dazwischen stattfinden (Frequency Division Duplex (FDD)). Im

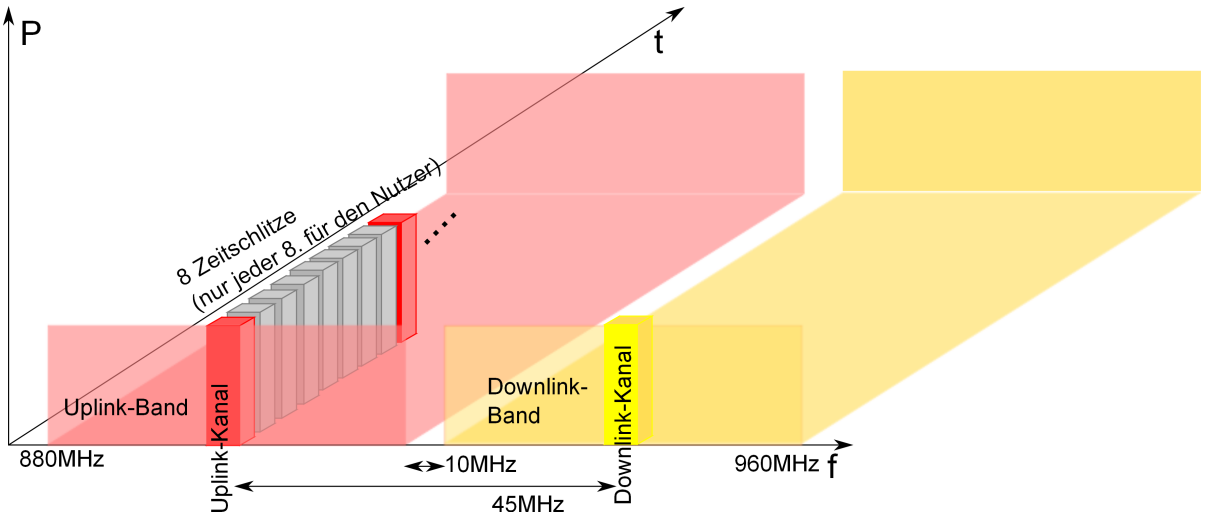


Abbildung 11.3: Parameter des GSM (nicht maßstäblich) im Überblick.

letzteren Fall ist sogar voll-Duplex möglich, d.h. beide Stationen können gleichzeitig Senden und Empfangen, was in der analogen (Funk)Kommunikation ein sehr schönes Leistungsmerkmal ist.

11.5 Beispielhafte Untersuchung des GSM-Systems

11.5.1 Datenrate und Spektrale Ausnutzung

Die Effizienz der beiden besprochenen Zugriffsverfahren sollen anhand des GSM-Systems untersucht werden. GSM bietet sich deshalb an, weil es, anders als modernere Mobilkommunikationsverfahren, noch reinrassig auf TDMA und FDMA als Zugriffsverfahren setzt. Für das hier betrachtete E-GSM im 900 MHz-Band stehen das

- Uplink-Band von 880-915 MHz für die Kommunikation vom Handset zur Basisstation und das
- Downlink-Band von 925-960 MHz für die Kommunikation von der Basisstation zum Endteilnehmer zur Verfügung.
- Die Lücke beträgt 10 MHz, während der
- Duplexabstand zwischen dem zugewiesenen Uplink- und Downlink-Kanal 45MHz beträgt.
- Die Kanäle sind im Abstand von 200 kHz zueinander angeordnet.

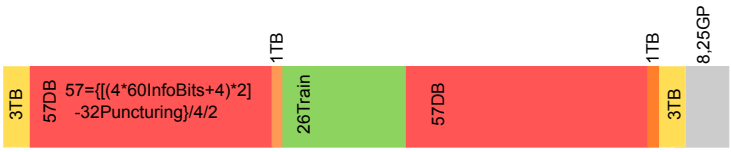


Abbildung 11.4: Der 12 kbps bzw. 9,6 kbps Datenservice Parameter des GSM (nicht maßstäblich) im Überblick.

Desweiteren werden im GSM auch Zeit-Multiplexverfahren einsetzt. So ist ein Rahmen acht Zeitschlitze mit ca. 5,769 ms lang. Im Normalbetrieb, d.h. bei Sprachübertragung oder bei normaler, vollratiger Datenübertragung hat ein Nutzer nur jeden achten Zeitschlitz zur Verfügung, die anderen werden anderen Nutzern zugewiesen. Alle diese Parameter sind zentral vom Netzwerk gesteuert (vgl. Abb. 11.3)
Die weiteren Parameter des GSM sind:

- Modulation ist eine Frequency Shift Keying (FSK) mit einem Bandbreite-Perioden Produkt von $B \times T = 0{,}5$, womit eine Datenrate von $1/T = 2B$ erreicht wird, in einer Symbolperiode also - wie bei der QPSK - zwei Bit gesendet werden können.
- Der Aufbau eines Zeitschlitzes für eine Datenrate von 12 bzw. 9.6 kbps ist in Abb. 11.4 dargestellt und wie folgt Parametrisiert:
  - Gesamtlänge 156,25 Perioden mit einer Datenrate von

$$\frac{1}{T} = \frac{1625}{6\,\text{ms}} = 270{,}833\,\text{kbit/s}, \tag{11.1}$$

    darin
  - $2 \times 57 = 114$ Nutzbits (kodiert mit Rate $R_C = 1/2$ und punktiert mit $R_P = 57/61$)
  - 26 in der Mitte zur Kanalschätzung etc. (Midamble)
  - 8,25 Periodendauern als Schutzintervall (Guardperiod)
  - 2×3 Bits am Anfang zum ein und Ausschalten der Verstärker (Tail-bits)
  - 2×1 Bit am Ende jeder 57 Bits ebenfalls als Tail-bits

Einschub Aufbau des Frames

Der Frame wird wie folgt aufgebaut: Ein Datensatz von $4 \times 60 = 240$ Informationsbits wird mit jeweils einem, also vier Tail-Bits versehen und dann mit einem Faltungskodierer mit der Rate $R_C = 1/2$ auf insgesamt 488 Bits aufkodiert.

Zur Ratenanpassung werden 32 Bits wegfallen gelassen (Punktierung mit der Rate $R_P = 57/61$), so dass sich 456 Bits nach Kodierung und Punktierung ergeben. Diese Bits werden mit einem Interleaver schließlich auf insgesamt 22 Schlitze verteilt. Dieses geschieht aber überlappend für andere Datenblöcke so, dass effektiv eine Verteilung auf 4 Schlitze (also $2 \times 4 = 8$-$57-$Bit-Blöcke gerechnet werden kann).
Effektiv werden also hier vier Zeitschlitze der zeitlichen Länge 5,769 ms verwendet und mit $456 = 4 \times (60 + 1) \times R_P/R_C$ Datenbits belegt. Dieses entspricht einer kodierten Datenrate von 19,76 kbps bzw. 10,4 kbps nach der Kodierung. Die Nomenklatur des Dienstes (12 kbps bzw. 9,6 kbps) suggeriert eine andere Datenrate, die daher rührt, dass von einem Datenstrom von 60 Bit alls 5 ms ausgegangen wird, und neben den Informationsbits auch noch Kontrollbits übertragen werden müssen, die hier ggf. vom FACCH "gestohlen" werden. Das wird hier nicht betrachtet.

Nach dieser Darstellung der Technik soll nun die Effizienz untersucht werden. Am einfachsten fällt dieses bei der Betrachtung im Frequenzbereich, hier ist für die Datenrate nach (Gl. 11.1) mit

$$B \times T = 0{,}5 \tag{11.2}$$

$$\Leftrightarrow B = \frac{0{,}5}{T} = 135{,}42\,\text{kHz} \tag{11.3}$$

nötig. Es stehen aber 200 kHz zur Verfügung, wodurch sich ein Guardband von $64{,}582\,\text{kHz} = 2 \times 32{,}29\,\text{kHz}$ ergibt. Das Spektrum wird also nur mit einer Effizienz von

$$\eta_{ch} = \frac{135{,}42}{200} = 67{,}71\,\% \tag{11.4}$$

ausgenutzt. Weitere, vor allem netzwerkbezogene, Effizienzbetrachtungen folgen weiter unten.
Im Zeitbereich ist die Lage komplizierter, doch es kommt letztlich zu der Aussage, dass in einem Zeitschlitz mit 156,25 Perioden Länge 114 unkodierte Bit transportiert werden:

$$\eta_{slot} = \frac{114}{156{,}25} = 72{,}96\,\%. \tag{11.5}$$

Betrachtet man noch die Kodierung und Punktierung, so ergibt sich hier ein Effizienzfaktor von

$$\eta_{cod} = \frac{R_C}{R_P} = \frac{61}{57 \times 2} = 53{,}51\,\%, \tag{11.6}$$

und insgesamt folgt also

$$\eta = \eta_{ch} \times \eta_{slot} \times \eta_{cod} = 26{,}435\,\%. \tag{11.7}$$

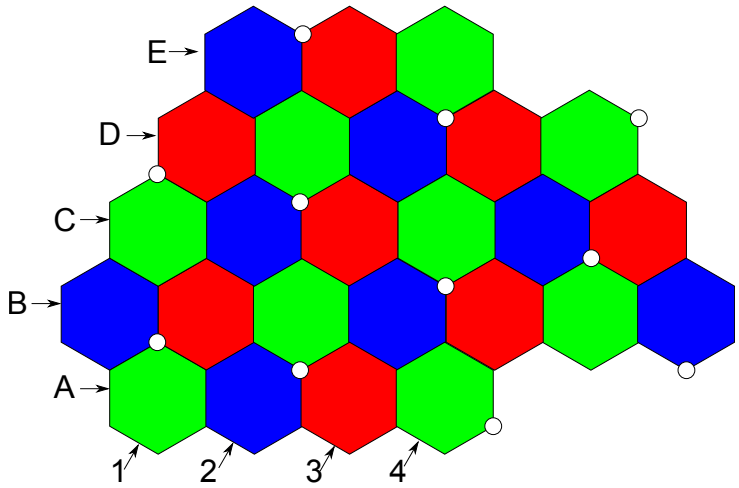


Abbildung 11.5: Wabenförmige Netzstruktur mit drei Frequenzen. Antennenstandorte sind in den weißen Kreisen.

Oder anders ausgedrückt werden bei diesem Dienst

$$B_{12} = 0{,}26 \frac{\text{bit}}{\text{Hz} \cdot \text{s}} \tag{11.8}$$

übertragen. Das klingt nicht viel, und zeigt auch, warum das GSM System in seiner ursprünglichen Form nicht für hohe Datenraten geeignet ist; es war dafür auch nicht ausgelegt.

11.5.2 Netzwerkbetrachtungen

Außer Acht gelassen wurde bislang der Duplex Abstand von 10 MHz zwischen Up- und Downlink bei einem Gesamtband von 70 MHz, der weiterhin die spektrale Ausnutzung um den Faktor

$$\eta_{dup} = \frac{960 - 925 + 915 - 880}{960 - 880} = 87{,}5\,\% \tag{11.9}$$

reduziert. Weiterhin müssen Netzwerkaspekte berücksichtigt werden. So können Frequenzen bzw. Kanäle sicherlich nicht in benachbarten Zellen verwendet werden. Ein Nutzer an der Grenze zwischen zwei Zellen mit gleicher Frequenz würde massiv durch die Nachbarn gestört werden, ein Empfang wäre unmöglich. In Abb. 11.5 ist eine einfache Situation gezeigt, in der davon ausgegangen wird, dass eine Frequenz jeweils in der übernächsten Zelle wieder verwendet werden kann. Dieses ist schon der beste anzunehmende Fall. Üblicherweise geht man von einer sechseckigen Wabenstruktur aus, auch wenn Funkfelder natürlich praktisch nicht so aussehen. Die weißen Kreise beschreiben die Standorte von

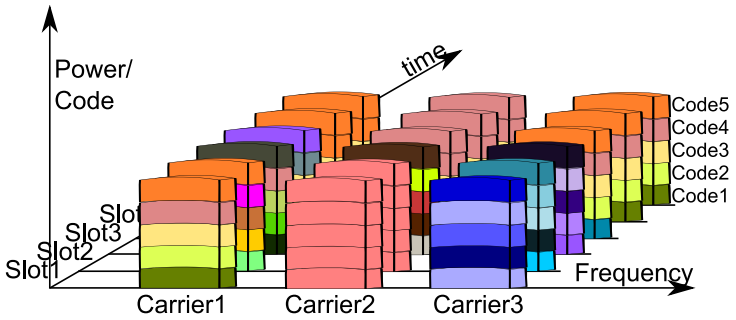


Abbildung 11.6: Die drei Dimensionen des Zugriffs auf einen Funkkanal.

Basistationen mit 3-Sektoren-Antennen, die in die jeweils notwendige Richtung gezielt und gerichtet abstrahlen können. Unter dem Strich ergibt diese Betrachtung, dass in einer Zelle nur jeder dritte Kanal genutzt werden kann. Der Wiederverwendungsfaktor von

$$\eta_{reuse} = 33{,}3\,\% \tag{11.10}$$

stellt in diesem System also den größten Einzelfaktor zur Verschlechterung des Ausnutzungsgrades dar. Dieses ist eine überaus positive Abschätzung, in der Regel geht man von einer Wiederverwendung der Frequenz erst in der übernächsten Zelle aus, der Wiederverwendungsgrad sinkt dann auf $1/7 \approx 15\,\%$. Eine Verbessung der Situation ist im Abschnitt 11.6 präsentiert.

11.6 Code Division Multiple Access

Mit den in Abschnitt 8.2 und insbesondere Abschnitt 8.2.3 lässt sich die oben beschriebene Situation der schlechten Frequenzausnutzung deutlich verbessern: Weist man jeder Zelle eine unterschiedliche Scrambling-Sequenz zu, so kann man Signale unterschiedlicher Basisstationen aufgrund der nahezu Orthogonalität der Sequenzen trennen (Abb. 8.7). Dieses gilt auch, wenn die beiden Zellen auf der gleichen Frequenz operieren. Damit ist ein Wiederverwendungsfaktor von $\eta_{reuse} = 100\,\%$ erzeugt worden.

Noch soft-handover???

zudem teilen sich durch Trennung mit Walsh-Codes mehrere Nutzer das gleiche Frequenzband. Hiermit kann man neben Zeit und Frequenz eine dritte Dimension der Zugriffe, wie in Abb. 11.6 gezeigt, betrachtet werden. Unterschiedliche Nutzer und Dienste unterschieden sich somit weder in Frequenz, noch in Zeit, dafür aber im verwendeten Code. Natürlich kommt es zu Interferenzen, die durch geeignete Emfangsmechanismen minimiert werden können: Werden nur die gewünschten Signale des Zielnutzers entspreizt, dann sind alle anderen Signale Rauschen, werden aber alle Nutzersignale entspreizt, so machen auch diese

Signale Sinn und können vom Zielsignal abgezogen werden. Diese Signale sind damit nicht mehr vorhanden. Man kann sich leicht vorstellen, dass es optimal ist, wenn alle Codes mit der gleichen Leistung empfangen werden. Sollte umgekehrt ein Code deutlich stärker empfangen werden als der gesuchte, dann wird es zu erheblichen Störungen kommen, denn das Rauschniveau steigt unweigerlich an. Alle an der Kommunikation beteiligten Stationen müssen also ihre Leistung so Regeln, dass sie mit in etwa gleicher Amplitude an der Basisstation zu empfangen sind. Aufgrund der schnell ändernden Empfangssituation hat dieses entsprechend schnell zu erfolgen. Im UMTS ist hierfür ein Dynamikbereich von über 70 dB (minimale Leistung -50 dBm bis maximal 23 dBm [Mobilstationsabhängig]) mit einer Genauigkeit von 1 dB und einer Rate von 1500 Hz vorgesehen.

11.7 Orthogonal Frequency Division Multiple Access

Kommt nur, wenn hierfür noch Zeit ist, ist das Zugriffsverfahren bei LTE.

11.8 Ungesteuerter Zugriff: ALOHA und CSMA

Nur bei Bedarf, da es schon in Vorlesungen vorher gemacht wurde und auch nicht so schwierig ist, als dass man es in einer Vorlesung wirklich vertieft machen müsste.

Kapitel 12

Sender und Empfänger

12.1 Empfängerstrukturen

In diesem Abschnitt werden unterschiedliche Emfängertypen vorgestellt. Da häufig Signale nicht direkt im Basisband, sondern auf einen hochfrequenten Träger aufmoduliert übertragen werden, muss ein Empfänger zwei wesentliche Aufgaben erfüllen:

- Selektion des gewünschten Signals und gleichzeitige Unterdrückung unerwünschter Signale und weiterer Störungen,
- Transformation der Signale in das Basisband und ggf. weitere Rekonstruktion und Demodulation

Gängige Typen von Empfängern sind in Abb. 12.1 dargestellt. Vergleiche auch [41, Q] für eine zwar alte, aber dennoch aktuell verwendbare und exzellente Darstellung der Materie.

Tabelle 12.1: Empfängerfamilien

HF/ZF-Verstärkung	Frequenz Umsetzung	Homo-/ Heterodyn	(A)-Synchron	Name
Nein	Nein		Asynch.	Quadratischer Detektor
Ja	Nein		Asynch.	Geradeausempfänger
Ja	Ja	Beides	Beides	Überlagerungsempfänger
Ja	Ja	Heterodyn	Beides	Superhet
Ja	Ja	Homodyn	Asynch.	Quadraturempfänger
Ja	Ja	Homodyn	Synch.	Synchronempfänger

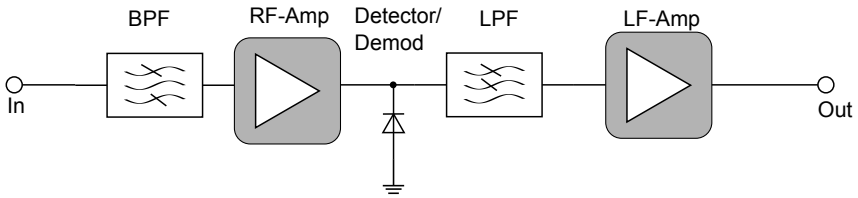


Abbildung 12.1: Geradeausempfänger oder bei Wegfall der Verstärker quadratischer Detektor.

12.1.1 Quadratischer Detektor und Geradeausempfänger

Das entsprechende prinzipielle Schaltbild ist in Abb. 12.1 gezeigt. Das Signal wird durch ein (HF) Bandpassfilter selektiert und ggf. nach Verstärkung in das Basisband umgesetzt. Der Vorteil ist die Einfachheit der Schaltung, es werden nur sehr wenige insbesondere aktive Elemente benötigt. Da nur die Leistung (Amplitude) detektiert wird, eignen sich Quadratischer Detektor und Geradeausempfänger nur für analoge (amplitudenmodulierte) oder einfachste digitale Modulationsverfahren ohne Notwendigkeit der Auswertung der Phaseninformation.
Ein wichtiger Nachteil ist, dass die Selektion auf der HF-Ebene, also ggf. einer recht hohen Frequenz, stattfinden muss. Dieses kann mit verlustbehafteten Elementen konzentriert oder verteilt sehr schwierig werden, insbesondere dann, wenn die Selektion nicht nur scharfbandig, sondern auch noch auf verschiedene Frequenzen abstimmbar sein soll.
In einer Zeit, in der aktive Bauelemente teuer und aufwändig waren, ggf. Elektronenröhren verwendet werden mussten, boten sich solche Schaltungen an. In der elektrischen Empfangstechnik ist dieser Empfängertyp heutzutage verdrängt worden. Als Geradeausempfänger, bzw. Audion [51, 52] hat er jedoch zum Durchbruch des Rundfunks geführt und seinen Erfinder Lee de Forest als einen der ganz Großen der Elektrotechnik zu Weltruhm gebracht. Das Audion besteht aus einer Elektronenröhre, die verstärkt und gleichzeit in der Sättigung betrieben wird, so dass das amplitudenmodulierte Nutzsignal an der Nichtlinearenkennlinie mit einem evtl. vorhandem Trägerrest heruntergemischt wird.
Heute funktionieren in aller Regel optische Empfänger nach dem Prinzip des Quadratischen Detektors . Die HF ist dann das optische Signal, die NF das opto-elektrische. Die Detektion/ Demodulation übernimmt eine Photodiode (Abschnitt 12.2.2.1), gefiltert wird grob oder gar nicht.

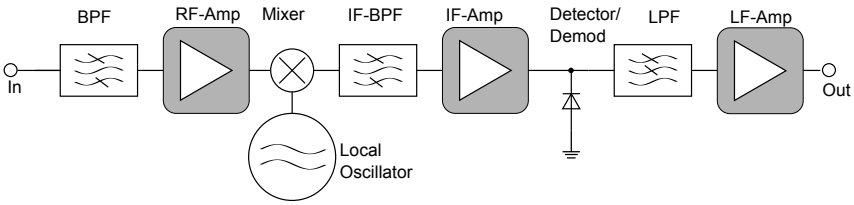


Abbildung 12.2: Super-Heterodyn-Empfänger (Superhet) mit einer Zwischenfrequenz.

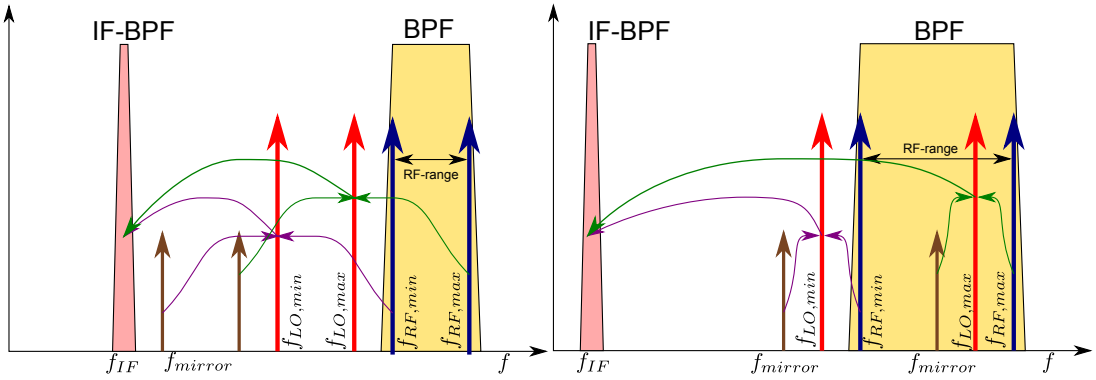


Abbildung 12.3: Frequenzverhältnisse im Superhet mit guter (a) und schlechter (b) Wahl der Zwischenfrequenz.

12.1.2 Heterodyner Empfang (Superhet)

Der wesentliche und entscheidende Unterschied ist die Einführung einer Zwischenstufe auf einer vom Entwickler wählbaren Frequenz. Alle Filter können auf eine feste Frequenz abgestimmt werden, der Lokaloszillator (LO) setzt das HF-Signal auf die ZF um:

$$f_{ZF} = |m f_{HF} \pm n f_{LO}| \,. \tag{12.1}$$

Die Ordnungszahlen m und n, die ganze positive Zahlen sind, sind in den meisten Fällen beide eins. Selbst in dem Fall ergeben sich durch die Betragsbildung zwei Lösungen, einmal für die gewünschte Mischung und dann noch für die sogenannte Spiegelfrequenz (Abb. 12.3). In Abb. 12.3(b) ist dieses Problem nochmals dargestellt: Es kann zu Situationen kommen, in denen die Spiegelfrequenz im Durchlassband des HF-Bandpassfilters liegt. Dann wird dieses Signal durch den Lokaloszillator in das Zwischenfrequenzband gemischt und kann durch das dortige ZF-Filter nicht gesperrt werden. Die Selektion versagt. Daher sind die Frequenzen zu wählen wie in Abb. 12.3(a) gezeigt, die ZF muss also immer größer sein, als das betrachtete HF-Band breit ist. Dieses kann bedeuten, dass eine zweite ZF eingeführt werden muss und in mehreren Stufen das Signal herabgesetzt wird, oder dass

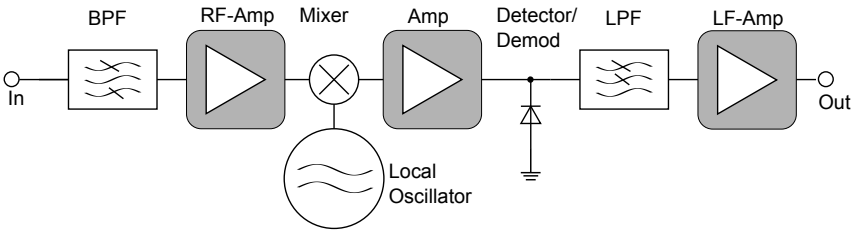


Abbildung 12.4: Homodyner Empfänger, Struktur.

die ZF oberhalb des HF-Bandes liegen muss. Letztere Wahl ist nicht unüblich und bspw. in HF-Spektrumanalysatoren zur Vermeidung der o.g. Mehrdeutigkeiten ausgeführt.
Richtig ausgeführt bietet der Superhet aber so viele Vorteile, dass er der Standardempfänger im analogen Rundfunk geworden ist:

- HF und ZF-Bandpassfilter können fest implementiert werden, die Auswahl der genauen HF-Frequenz wird durch Verstimmung des Lokaloszillators erreicht. Dieses ist vergleichsweise einfach durch z.B. PLL und VCO zu implementieren.
- Die schmalbandige Selektion kann auf einer (mehr oder weniger frei) gewählten ggf. niedrigeren Frequenz ausgeführt werden. Hier kann eine ZF gewählt werden, auf der Filter leicht implementierbar oder in bestimmten Technologien leicht verfügbar sind.

Der Preis ist der höhere Aufwand durch den

- Lokaloszillator,
- Weiteres Filter
- Mischer,
- Weitere Verstärkerstufen.

12.1.3 Homodyner Empfang

Dem Problem der Mehrdeutigkeit durch Spiegelfrequenzen geht man beim homodynen Empfänger wenigstens scheinbar aus dem Weg. Hier ist der Lokaloszillator möglichst genau auf die Mittenfrequenz des zu selektierenden HF-Kanals abgestimmt. Andere Kanäle können einfach durch Veränderung der LO-Frequenz ausgewählt und ins Basisband gemischt werden, auf der HF-Ebene muss das HF-Band ausgewählt werden, um fremde

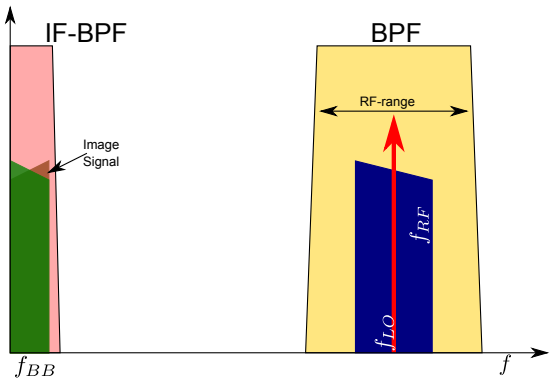


Abbildung 12.5: Frequenzverhältnisse im homodynen Empfang.

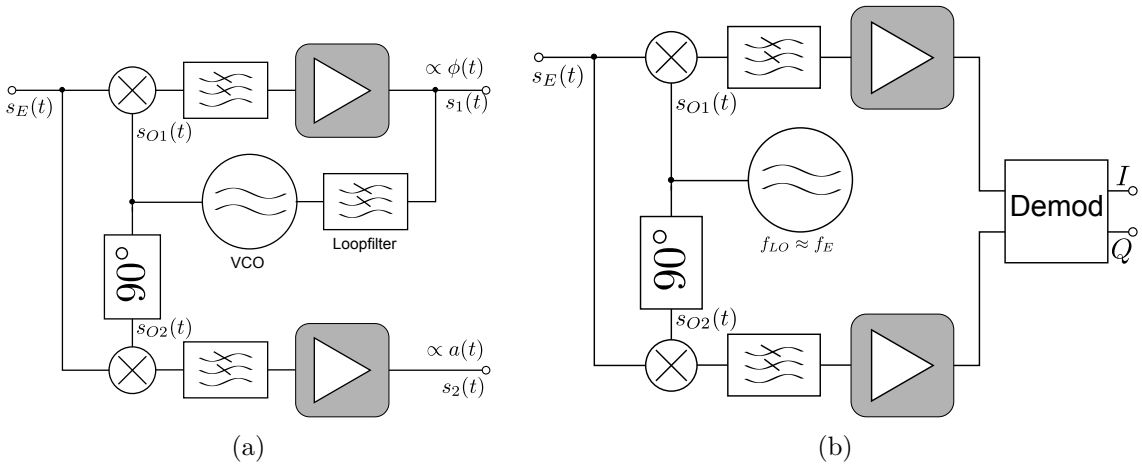


Abbildung 12.6: Homodyner Synchronempfänger (a) und Quadraturempfänger (b).

Störer und Rauschen fern zu halten, Selektion findet allein durch die Kombination LO und BB-Filter statt.
Herausforderungen treten auf, weil nach (Gl. 12.1) das Signal sowohl in Gleich- als auch in Kehrlage auftritt. Mathematisch ergibt sich das daraus (wie auch die Spiegelfrequenz), dass in einem Mischer das eigentlich komplexe Signal mit einem reellen Lokaloszillator umgesetzt wird. Dieses ist ein $\sin(\cdot)$ oder $\cos(\cdot)$, der bekanntlich nach

$$\cos(\omega t) = \frac{1}{2}\left(e^{j\omega t} + e^{-j\omega t}\right) \tag{12.2}$$

positive und negative Frequenzen enthält. Wenn beide (Gleich- und Kehrlage) richtig und genau zusammenfallen, dann lässt sich das gesuchte Signal detektieren. Dazu ist in jedem Fall eine Trägerregelung (PLL) notwendig.

Synchronempfänger Beim Synchronempfänger (Abb. 12.6(a)) wird ein Zeitsignal

$$s_E(t) = \mathrm{Re}\left\{[1 + a(t)]\, e^{j\phi(t)} e^{j\omega_E t}\right\} \tag{12.3}$$

im Mischer mit dem LO und dessen um 90° gedrehter Variante gemischt:

$$s_{O1}(t) = \sin(\omega_E t) = \frac{1}{2j}\left(e^{j\omega t} - e^{-j\omega t}\right) \tag{12.4}$$

$$s_{O2}(t) = \cos(\omega_E t) = \frac{1}{2}\left(e^{j\omega t} + e^{-j\omega t}\right) \tag{12.5}$$

und es ergibt sich insgesamt auch nach Tiefpassfilterung

$$s_1(t) = \mathrm{Re}\left\{s_E(t) S_{O1}(t)\right\} = \mathrm{Re}\left\{\frac{j}{2}\left[1 + a(t)\right] e^{j\phi(t)}\left(-e^{j2\omega_E t} + 1\right)\right\} \tag{12.6}$$

$$\propto [1 + a(t)]\sin(\phi(t)) \tag{12.7}$$

$$s_2(t) = \mathrm{Re}\left\{s_E(t) S_{O2}(t)\right\} = \mathrm{Re}\left\{\frac{1}{2}\left[1 + a(t)\right] e^{j\phi(t)}\left(e^{j2\omega_E t} + 1\right)\right\} \tag{12.8}$$

$$\propto [1 + a(t)]\cos(\phi(t)) \tag{12.9}$$

$$\tag{12.10}$$

wobei das erste Signal bei kleinen Phasenschwankungen proportional zur Phase des Trägers ist und das zweite im Wesentlichen proportional zur Amplitude. Es ist wichtig, dass die Phasenschwankungen klein sind, daher muss der Lokaloszillator mittels PLL geregelt werden.

Quadraturempfänger Beim Quadraturempfänger nach Abb. 12.6(b) verzichtet man auf die exakte Regelung des Lokaloszillators und stellt ihn nur so ein, dass das herunter gemischte Signal nahe Null zu liegen kommt. Nach Selektion in den Tiefpassfiltern führt man es einen Demodulator zu. Dieser ist entsprechend komplex und beinhaltet mittelbar die Trägerregelung und die Demodulation oder gleicht diese Funktionen durch aufwändige mathematische Operationen aus, um den Betrag und die Phase nach

$$\sqrt{x^2 + y^2}, \arctan\left(\frac{y}{x}\right) \tag{12.11}$$

zu bestimmen. Als Demodulator kann aber z.B. auch eines der in Kapitel 9 insbesondere in Abschnitt 9.2 und 9.2.2 dargestellten Verfahren verwendet werden.

12.2 Optische Strukturen

Zum Überblick über ein optisches System ist in Abb. 12.7 für eine einfache unidirektionale Übertragungsstrecke dargestellt. Die Komponenten sind offensichtlich Sender (eine

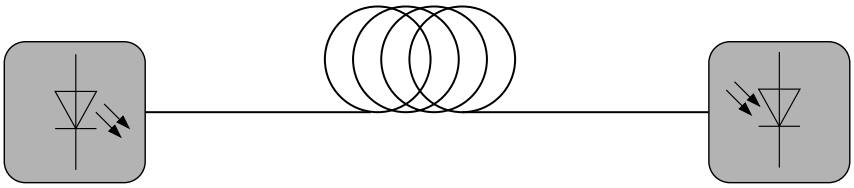


Abbildung 12.7: Einfache optische Übertragungsstrecke.

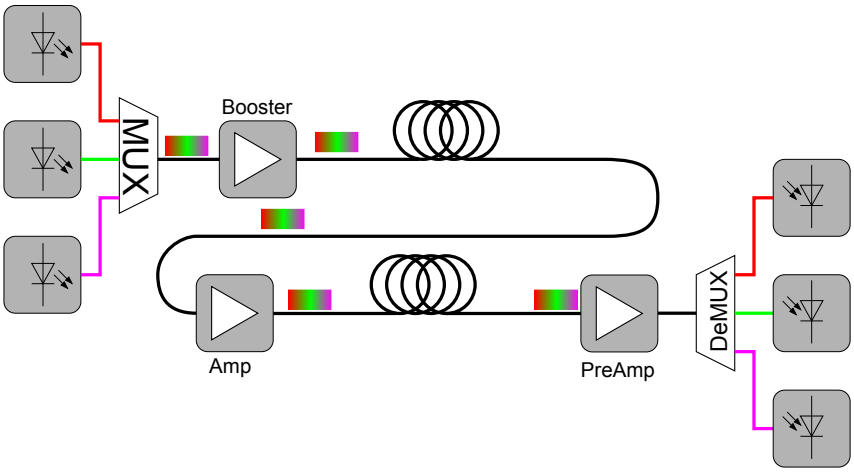


Abbildung 12.8: Optische Übertragungsstrecke mit Wellenlängen Multiplex.

Laserdiode oder LED) die Glasfaser und ein Empfänger, eine Photodiode. Eine in Weitverkehrsstrecken mit hoher Datenrate heute gebräuchliche Erweiterung ist in Abb. 12.8 gezeichnet. Hier nutzt man aus, dass Licht sich in linearen Medien nicht gegenseitig beeinflusst, und daher Information auf unterschiedlichen Wellenlängen (Frequenzen, auch gerne „Farben“ genannt) parallel übertragen werden kann. Der Vorteil ist, dass mehrere Sender unterschiedlicher Wellenlänge über einen MUX eine Faser speisen können, die dann gemeinsam genutzt wird. Auf der Empfängerseite wird das Signal wieder in seine einzelnen Farben im DeMUX zerlegt und der Empfängerdiode zugeführt. Da es sich hier oft um Weitverkehrstrecken handelt, erfahren die Signale auf der Strecke eine hohe Dämpfung und müssen verstärkt werden. Dieses geschieht durch optische Faserverstärker auf rein optischem Wege, das Signal muss also nicht zurück in ein elektrisches gewandelt und dann aufbereitet werden. Die optische Verstärkung kann, muss aber nicht, an typischerweise drei Stellen erfolgen:

- Als Booster am Anfang der Strecke, um ein recht hohes Signal von vorherein auf die Strecke zu schicken,
- als Pre-Amplifier am Ende der Strecke, um damit effektiv die Empfindlichkeit der

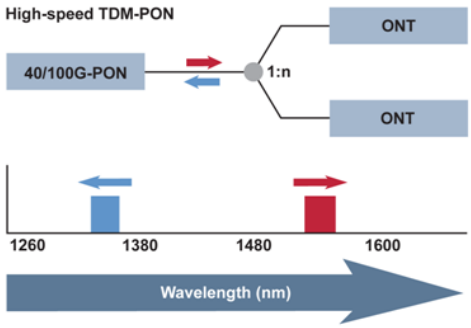


Abbildung 12.9: High-Speed TDM passive optical Network Struktur nach FTTH-Council.

Empfängereinheit zu erhöhen, und/oder

- als Streckenverstärker an verschiedenen Stellen der Strecke, um dort wo nötig das Signal vor dem Versinken im Rauschteppich zu bewahren. Dieses kann an keiner, einer oder auch mehreren Positionen geschehen.

12.2.0.1 Wellenlängenmultiplex

Das Konzept des Wellenlängenmultiplex ist oben erläutert. In seiner Ausführung gibt es allerdings viele verschiedene Geschmacksrichtungen, die sich in Anwendung, Datenrate und Farben deutlich unterscheiden. Zunächst ist eine einfache Übertragung auf zwei sehr unterschiedlichen Wellenlänge (z.B. 1310 nm und 1550 nm) möglich. Durch den großen Unterschied der Wellenlängen sind die MUX und DeMUX-Einheiten recht einfach auszuführen, es können auch einfache Sendeeinheiten für diese Standard-Wellenlängen verwendet werden. Allerdings sind hier nur zwei unterschiedliche Kanäle aufgebaut und damit wird auch nur eine Verdopplung der Datenrate angeboten. Dieses Verfahren eignet sich beispielsweise auch, um auf einer Faser einen Hin- und Rückweg zu bewerkstelligen (Abb. 12.9).
Weitere Varianten sind sogenannte „Overlay Networks “ kombinieren in einer Faser einen Datenstrom bei 1550 nm und einen beim 1310 nm.
Höhere Datenraten sind mit der Nutzung mehrerer Wellenlänge verbunden, die z.B. im CWDM nach [53] definiert sind (Tab. 12.2), das 18 verschiedenen Wellenängen in einem recht groben Raster mit einen Abstand von jeweils 20 nm definiert. Nicht alle der Wellenlängen, insbesondere im Bereich 1270-1470 nm, sind wirklich nutzbar, da sie auf das Absorptionsmaximum des Wassers in den Glasfasern fallen.
Noch höhere Datenraten werden im DWDM zur Verfügung gestellt. In der Betrachtung wechselt man hier zur Frequenzdarstellung, da nur diese - im Gegensatz zur Wellenlängen-

Tabelle 12.2: Coarse Wavelength Division Multiplexing (CWDM) Wellenlängengitter (Abstand 20 nm) nach [53]

1271	1291	1311
1331	1351	1371
1391	1411	1431
1451	1471	1491
1511	1531	1551
1571	1591	1611

Tabelle 12.3: Dense Wavelength Division Multiplexing (DWDM) Wellenlängengitter (Beispiele) nach [54]

12,5 GHz	25 GHz	50 GHz	100 GHz	λ/nm
195,9	195,9	195,9	195,9	1530,3341
195,8875				1530,4318
195,875	195,875			1530,5295
195,8625				1530,6271
195,85	195,850	195,85		1530,7248
195,8375				1530,8225
195,825	195,825			1530,9203
195,8125				1531,0180
195,8	195,8	195,8	195,8	1531,1157

berechnung - jedem Kanal die gleiche Bandbreite zur Verfügung stellt. DWDM ist (ohne das flexible Gitter) mit Abständen von 100 GHz, 50 GHz, 25 GHz und 12,5 GHz definiert [54]. Die Kanalallokation mit einer jeweils immer natürlich Zahl (mit Null) n ist

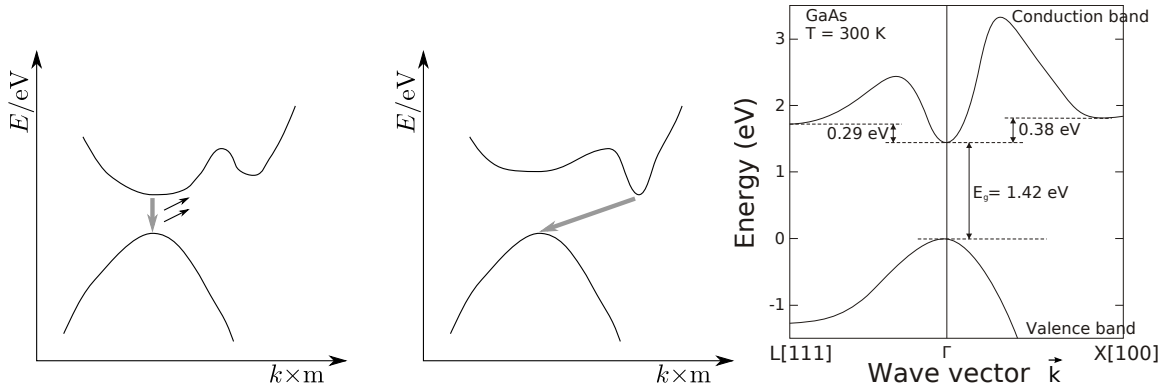
$$12{,}5\,\text{GHz}: \quad f = (193{,}1 + n \times 0{,}0125)\text{THz} \tag{12.12}$$

$$25\,\text{GHz}: \quad f = (193{,}1 + n \times 0{,}025)\text{THz} \tag{12.13}$$

$$50\,\text{GHz}: \quad f = (193{,}1 + n \times 0{,}050)\text{THz} \tag{12.14}$$

$$100\,\text{GHz}: \quad f = (193{,}1 + n \times 0{,}100)\text{THz}. \tag{12.15}$$

Die Wellenlänge ergibt sich dann durch Division der Lichtgeschwindigkeit $c = 2{,}99792458 \times 10^8$ m/s durch die Frequenz (Tab. 12.3). Natürlich sind bei dieser dichten Packung der Übertragungskanäle hohe Anforderungen an Sender, MUX und DeMUX zu stellen. Praktisch häufig genutzt werden das 100 GHz und das 50 GHz-Gitter, die - je nach Hersteller



(a) (b)

Abbildung 12.10: Direkter und indirekter Halbleiter im Impuls - Energie - Diagramm generisch (a) und Ga As (b) nach `http://en.wikipedia.org/wiki/File:Bandstruktur_GaAs_en.svg`.

und System - 32, 40, 64, 80 oder im 25 GHz-Gitter sogar 160 Kanäle [55] ermöglichen. Dieses skaliert die in einer Faser übertragbare Datenrate natürlich ganz erheblich.
Auf optische Bänder (L,C)-Band und Red und Blue-Band, sowie den Optical Supervisory Channel ist nicht eingegangen worden.

12.2.1 Optische Senderstruktur

Es fehlen: Das Laserprinzip und im Halbleiterlaser die Berechnung der Laserbedingung aus Verlusten durch die Reflexion und die verteilte Verstärkung in der aktiven Zone nach $e^{\alpha L}$.

12.2.1.1 Halbleiter

Optische Sender sind heute in der Regel Halbleiter Laser-Dioden, in seltenen Fällen und nur bei geringen zu übertragenden Datenraten, kurzen Strecken und hohem Preisdruck sind LED verwendet. Die Grundstruktur beider Elemente ist gleich. In beiden Fällen liegt ein direkter Halbleiter zugrunde, der einen direkten und damit strahlenden Übergang ermöglicht. Das Material ist daher meist im Al Ga As P oder In Ga As P-System zu finden. Hierin können die Bandabstände zwischen Leitungs- und Valenzband und damit die emittierten Wellenlängen im Bereich von 0,5 bis 2,3 eV bzw. 600 bis 3000 nm eingestellt werden.

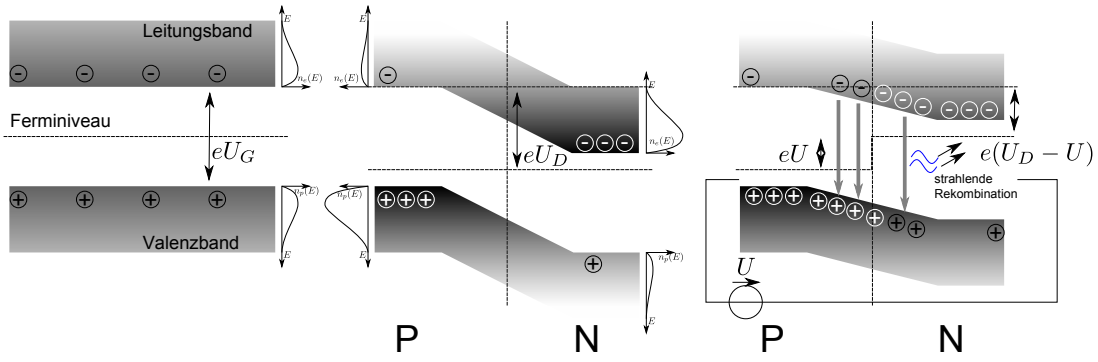


Abbildung 12.11: Halbleiter, PN-Übergang ohne und mit Vorspannung (in Flussrichtung).

12.2.1.2 PN-Übergang und LED

Die Grundstruktur ist der PN-Übergang (Abb. 12.11), in der in Flussrichtung vorgespannt wird. Im einfachen Halbleiter mit Ferminineau (Mittellinie) in der Mitte sind mit einer gewissen Wahrscheinlichkeit Löcher im Valenzband und Elektronen im Leitungsband. Deren Anzahl wird im wesentlichen durch die Bandlücke eU_G und die Temperatur bestimmt. Durch positive oder negative Dotierung einzelner Bereiche des Halbleiters werden die Besaetzungswahrscheinlichkeiten der Löcher bzw. Elektronen verändert, was man durch eine Verschiebung des Ferminineaus zum Valenzband (p-dotiert) oder Leitungsband (n-dotiert) ausdrückt. Bei Verbindung zweier unterschiedlich dotierter Hablleiter ergibt sich der PN-Übergang, in dessen unmittelbarer Nähe eine Verarmungszone auftritt, aus der aufgrund der Drift sowohl Löcher als aus Elektronen getrieben werden. Das Ferminniveau bleibt stetig, so dass sich zwischen den beiden Bereich des PN-Übergangs eine Driftspannung von U_D einstellt. Mittels Vorspannung U kann man nun zum einen das Ferminniveau auf beiden Seiten verschieben als auch dadurch Ladungsträger in die Verarmungszone treiben. Hier liegen nun Löcher und Elektronen vor. Stromleitung kommt zustande, wenn die Löcher aus der P-Zone mit den Elektronen der N-Zone rekombinieren, so dass entsprechende Ladungsträger nachgeliefert werden müssen und können. Bei direkten Halbleitern erfolgt diese Rekombination strahlend, so dass Photonen der Energie (Wellenlänge) der Bandlücke freigesetzt werden. Sowohl Leitungs- und Valenzband bieten quasi kontinuierliche Energieniveaus an, auf denen sich die Ladungsträger befinden können. Allerdings können nicht unendlich viele Ladungsträger direkt an der Bandkante sitzen, sondern werden sich auf unterschiedliche Energieniveaus verteilen. So entspricht die bei der Rekombination ausgesandte Wellenlänge nicht genau dem Energieniveau der Bandlücke, sondern ist statistisch etwas größer und bei spontaner Emission auch für jede Rekombination immer etwas anders. Es ergibt sich eine spektrale Streuung (Breite) der

Emissionen der Diode.
Bild (ggf. auch Formel) der Emissionbreite einer LED angeben.
Eine LED bringt damit folgende Nachteile mit sich:

- Die spektrale Breite der Emission ist groß, so dass viele verschiedene Wellenlängen in der Faser unterwegs sind, was durch Dispersion zu einer großen Aufweitung der Pulse führt, und somit nur geringe Datenraten ermöglicht.
- Die Strahlung aus dem Halbleiter heraus ist nur wenig aus gerichtet, ggf. nur in eine Halbebene. Damit ist es schwierig in eine Faser mit geringer numerischer Apertur (also einer SMF) einzukoppeln.

12.2.1.3 Laserdiode und Resonator

Die genannten Nachteile sind schnell in Forderungen für eine gute Quelle umzusetzen:

- Geringe spektrale Dichte,
- gerichtete Strahlung mit geringem Öffnungswinkel.

Leicht erfüllbar ist dieses durch Laserstrukturen. Das Grundprinzip der „Light Amplification by Stimulated Emission of Radiation“ ist die Rückkopplung der spontan emittierten Photonen in das aktive Medium (PN-Übergang), wo sie weitere Emission von Photonen stimulieren. Diese Photonen haben nun genau die Wellenlänge der stimulierenden. Durch wiederholte Rückkopplung ergibt sich eine weitere Verstärkung und damit eine weitere Einengung des Spektrums. Nur eine (oder wenige) spektral Linie(n) wachsen.
Die Rückkopplung erfolgt durch einen Resonator, was im optischen nur zwei zueinander parallel Spiegel sind. Diese Spiegel sind in Kristallstrukturen auch noch sehr einfach herzustellen, indem man die Kristallkanten einfach sauber abbricht. Beim Übergang vom Kristall zur Luft (oder einen anderen Medium) ergibt sich in der Regel ein großer Brechzahlsprung und damit eine genügend große Reflexion

$$R = r^2 = \frac{(n-1)^2}{(n+1)^2} \tag{12.16}$$

um zusammen mit der Verstärkung durch stimulierte Emission die Laserstrahlung zu entfachen. Der Brechungsindex von $\mathrm{Al}_x\,\mathrm{Ga}_{1-x}\,\mathrm{As}$ ist [56, S.715]

$$n(x) = 3{,}59 - 0{,}71x + 0{,}091x^2 (\approx 3{,}6). \tag{12.17}$$

Es bilden sich longitudinale Moden in dem Resonator aus, die in Abb. 12.12 dargestellt sind. Vereinfachend wurde hier ein Phasensprung von 180° an jedem Ende angenommen,

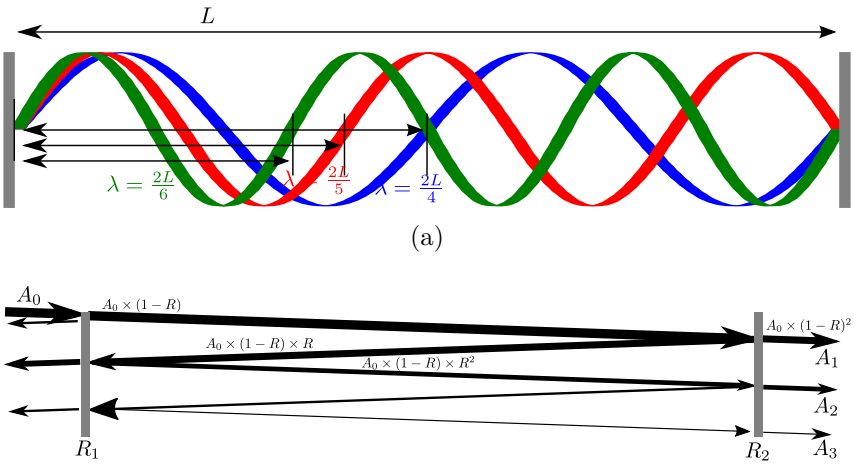


Abbildung 12.12: Einfache Darstellung des Fabry Pierot Resonators (a) und zur Herleitung der Airy-Formel (b).

so dass die gezeigten Feldstärken eine Null zeigen müssen. Unter diesen Bedingungen ist die Resonanzbedingung

$$L = n \times \frac{\lambda}{2} \qquad\qquad \lambda = \frac{2L}{n}, \tag{12.18}$$

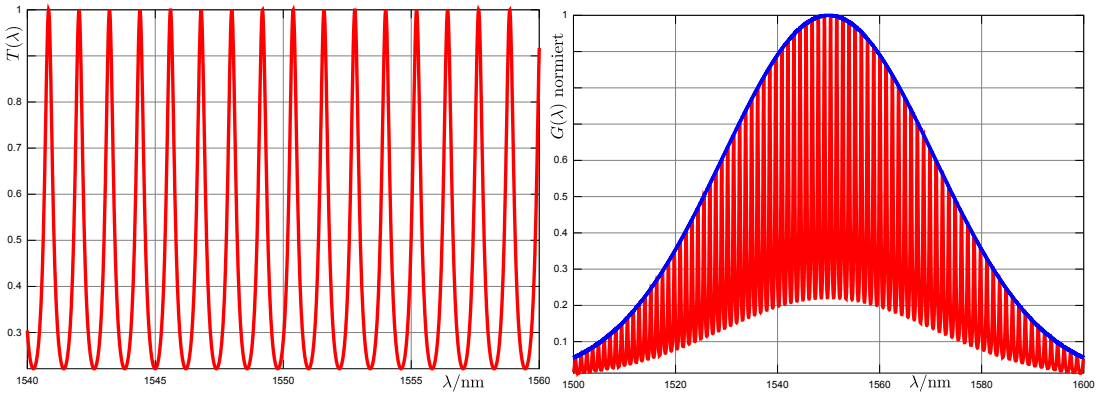
wobei n eine ganze Zahl sein muss. Bei Resonantoren, die lang gegenüber der Wellenlänge sind, ergeben sich also für relativ große longitudinale Ordnung n viele, dicht bei einander liegende erlaubte Lösungen und damit Resonanzwellenlängen. Berücksichtigt man Mehrfachreflexionen und endliche Reflexion $R_1 = R_2 = R$ an beiden Seiten, dann ergibt sich für die Transmission eines Fabry-Pierot-Resonators [7]

$$T = \frac{1}{1 + \frac{4R}{(1-R)^2} \sin^2\left(2\pi \frac{L}{\lambda}\right)}. \tag{12.19}$$

Bei den üblichen recht langen (gegenüber der Wellenlänge) Resonatoren erhält man viele eng benachbarte Linien **evtl. noch den Linienabstand berechnen** (Abb. 12.13)(a). Es können die Linien wachsen, für die die Verstärkung (G) durch das Medium die Verluste durch die endliche Reflexion (R) an den Enden mindestens aufwiegt:

$$R \times G = R \times e^{\alpha L} > 1. \tag{12.20}$$

Hier ist α der Realteil des Ausbreitungskoeffizienten in dem aktiven Medium. In einfachen Leitungen und Medien ist so ein Exponent immer kleiner als Null, so dass die Leistung mit der Strecke exponentiell abfällt. Im Lasermedium jedoch muss der Exponent positiv



(a) (b)

Abbildung 12.13: Modenfolge eines einfache Fabry-Pierot-Resonators $L \times n = 1\,\text{mm}$ und Reflexion $R = 0{,}35$ (Brechungsindex $n = 2$) (a) und (b) mit einer Verstärkungscharakteristik mit Maximum bei $\lambda = 1550\,\text{nm}$ und einer Halbwertsbreite von $\Delta\lambda = 50\,\text{nm}$.

werden, so dass durch stimulierte Emission die Leistung eben ansteigt. Die Energie kommt - wie oben schon angedeutet - aus der angelegten Gleichspannung.
Welche der vielen möglichen Linien genau nun die Stärkste wird, hängt von der Verstärkungskurve (siehe auch Abb. 12.13(b)) des Laser-Materials ab. In jedem Fall ergibt sich ein zwar eingeengtes, aber doch nicht monochromatisches Spektrum des Lasers. Durch die Sättigung der optischen Verstärkung im Laser wird erreicht, dass die bei weitem meiste Leistung in eine Linie konzentriert wird. Jedoch bleibt immer noch Restenergie in Seitenlinien.
Da das Spektrum eines Fabry-Pierot-Lasers noch auch mehreren Linien besteht, wird sich bei langen Übertragungsstrecken immer noch eine merkliche Aufweitung der Pulse durch Dispersion ergeben, und kann ein solcher Laser nicht verwendet werden, um eng benachbarte DWDM-Kanäle in einer Faser zu nutzen. Eine weitere Einengung des Spektrum kann erreicht werden, indem die Resonatorlänge verkürzt wird. Dieses kann technologisch sinnvoll durch einen Vertical Cavity Surface Emitting Laser (VCSEL) erreicht werden, der auch noch den weiteren Vorteil einer einfachen Fertigbarkeit durch die vertikale Struktur bietet. Die kürzere aktive Zone erfordert eine örtlich höhere Verstärkung und damit auch eine höhere Stromdichte und Leistungsdichte.
Der Schlüssel zum monochromatischen Licht liegt aber in einer monochromatischen Spiegelstruktur an beiden Seiten. Diese Reflektoren sind periodische Änderungen des Brechzahlindexes, durch z.B. unterschiedlich dotierte Materialien, oder auch periodische Verdickungen/ dünnungen des Wellenleiters. Genannt werden sie Bragg-Reflektoren. In jeder

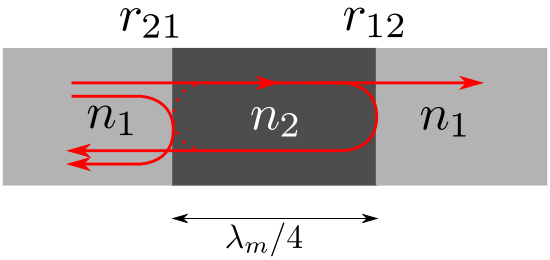
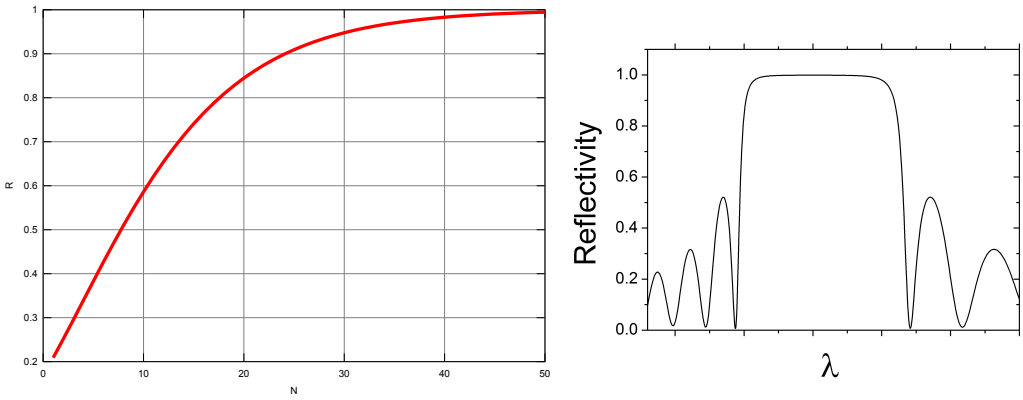


Abbildung 12.14: Eine (symmetrische) Braggzelle.



(a) (b)

Abbildung 12.15: Reflexion für unterschiedliche Ordnung des Bragg Reflektors (a) und Reflexion als Funktion der Wellenlänge (b) nach `http://de.wikipedia.org/wiki/Bragg-Spiegel` (b).

Zelle, die in Abb. 12.14 gezeigt ist, werden an jedem Ende Signalanteile reflektiert. Es gilt

$$r_{12} = \frac{n_1 - n_2}{n_2 + n_1} < 0, \qquad r_{21} = \frac{n_2 - n_1}{n_2 + n_1} > 0, \qquad \text{Annahme } n_2 > n_1. \tag{12.21}$$

Die optische (elektrische) Länge der einzelnen Brechzahlzone von $\lambda/4$ resultiert in einem Umweg der reflektierten Wellen von 180°. Mit der ebenfalls 180° Phasenverschiebung an einer der Grenzschichten folgt dann der Umweg einer ganzen Wellenlänge, so dass sich bei der richtigen Wellenlänge konstruktive Überlagerung in Reflexionsrichtung ergibt. Viele dieser Zellen hintereinander erlauben eine weitere Erhöhung der Reflexion. Bei einer Struktur aus Ga As und $Al_{0,3}\,Ga_{0,7}\,As$ ergeben sind nach (Gl. 12.17) Brechungsindices von $n_1 = 3{,}59$ und $n_2 = n(0{,}3) = 3{,}39$. Mit Ga As ($n_s = 3{,}59$) und einer Glasfaser ($n_0 = 1{,}5$) auf den beiden Seiten ergibt sich mit

$$R = \left[\frac{n_0 (n_2)^{2N} - n_s (n_1)^{2N}}{n_0 (n_2)^{2N} + n_s (n_1)^{2N}}\right]^2 \tag{12.22}$$

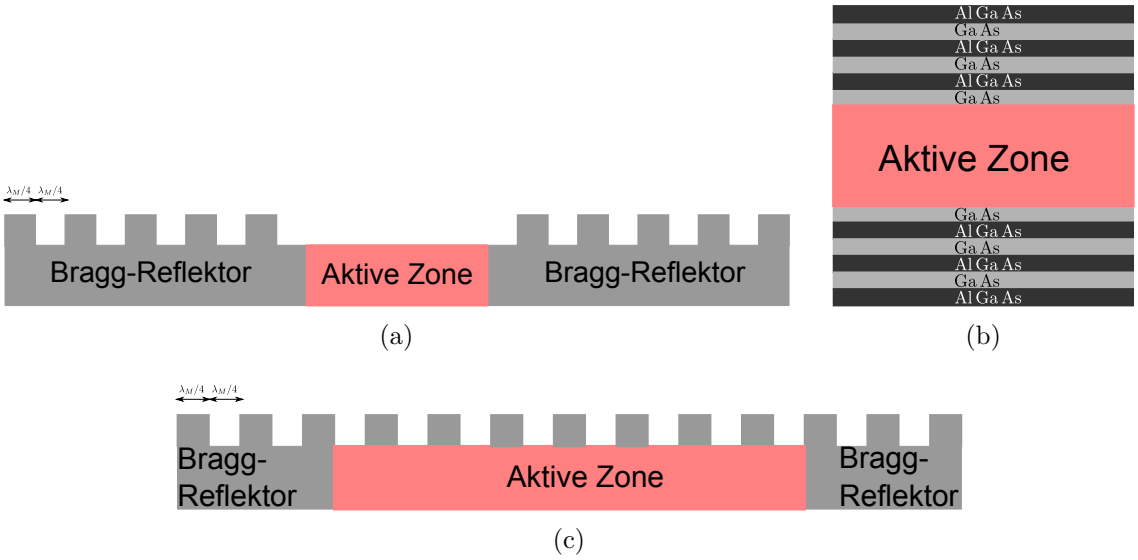


Abbildung 12.16: Unterschiedliche Laser-Reflektor-Strukturen Distributed Bragg Reflector (DBR) Struktur horizontal für Kantenemitter (a), vertikal als VCSEL (b) und als verteilte Struktur auch in der aktiven Zone (c).

die Reflexion nach Abb. 12.15(a). Für lange Reflektoren, also hohe Ordnungen N, kann die Reflexion somit beliebig hoch geschraubt werden (Abb. 12.16(a)). Zudem bringt der Bragg-Reflektor eine gewisse Bandpasscharakteristik mit sich. Vor allem ersteres führt dazu, dass man die Länge der aktiven Zone verkürzen kann (es ist weniger Gesamtverstärkung nötig), wie etwa in einer vertikalen Struktur im VCSEL (Abb. 12.16(b)). Dadurch wird der Abstand der longitudinalen Moden voneinander größer und der Laser trägt weniger Moden oder sogar nur noch eine Linie. Lässt man - wie in Abb. 12.16(c) den Reflektor verteilt in die aktive Zone hineingleiten, dann bekommt man so über einen Distributed Feedback Laser (DFB-Laser) nur noch eine Laserlinie.
Die Eigenschaften eines solchen Lasers können wie folgt genähert werden [7, S.86]: Jede einzelne Zelle in der aktiven Zone kann man als einen Bereich der Länge $\lambda/2$ sehen, der zu beiden Seiten mit einem Bragg-Reflektor abgeschlossen ist. Damit ersetzt man in (Gl. 12.19) die Länge des Reflektors L durch die Länge eines Reflektors, und passt die Reflexion entsprechend auf die Reflexion der Bragg-Reflektoren zu beiden Seiten an. Das Ergebnis dieser näherungsweisen Berechnung ist eine scharfe Linie, die in Abb. 12.17 gezeigt ist. Insbesondere für Laser mit hoch reflektiven Bragg-Reflektoren ergeben sich sehr schmale Emissionslinien. Man muss hier allerdings bemerken, dass dieses alles lineare Überlegungen sind. Die Laser-Funktion ist allerdings hochgradig nichtlinear, d.h. die Verstärkung muss sättigen, sonst würde der Laser unendlich viel Energie ziehen und ab-

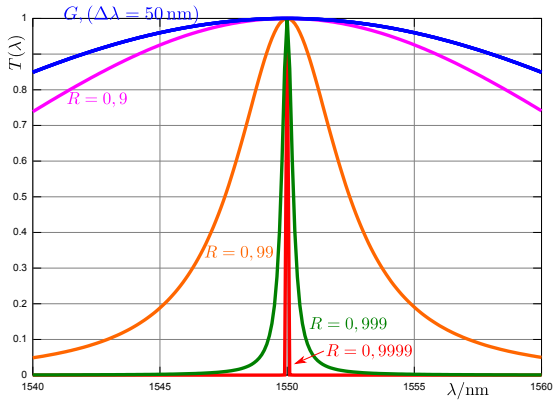
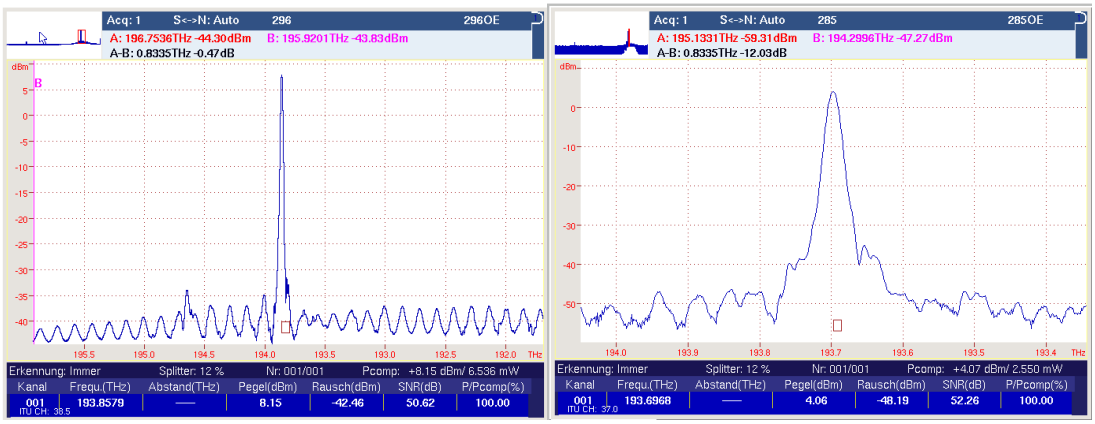


Abbildung 12.17: Näherungsweises spektrales Verhalten eines DFB-Lasers mit Bragg-Reflektoren unterschiedlicher Reflexion. Mittenfrequenz immer 1550 nm.



(a) (b)

Abbildung 12.18: Fabry-Pierot-Laser (?) einer Marconi STM-4 Maschine (a) und DFB-Laser einer Alcatel DWDM Maschine (b).

geben. Daher sind die realen Verhältnisse, was insbesondere Linienbreite und Rauschen anbelangt deutlich komplexer.

ggf. noch etwas zu $\lambda/4$-Sprung in den Bragg-Gittern schreiben

Im Endeffekt ergeben sich Laser, die in der Lage sind, eine schmale Linie zu emittieren. Diese können verwendet werden, um hohe Datenraten trotz dispersiver Fasern zu übertragen und ermöglichen auch die Übertragung mehrerer, spektral dicht benachbarter Lichtsignale in einer Faser (DWDM). Optische Spektren dieser Laser sind in Abb. 12.18. Eine grundsätzliche Laserkennlinie ist in Abb. 12.19 gezeigt. Deutlich sichtbar ist die sogenannte Laserschwelle, der Übergang von der spontanen Emission zum stimulierten, also dem Laserbetrieb.

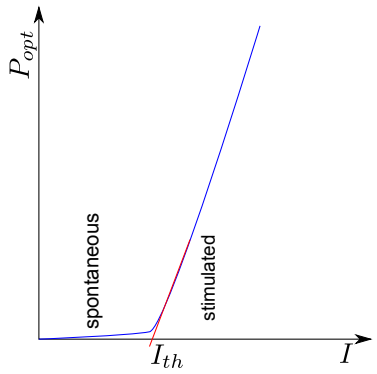


Abbildung 12.19: Generische Kennlinie eines Lasers.

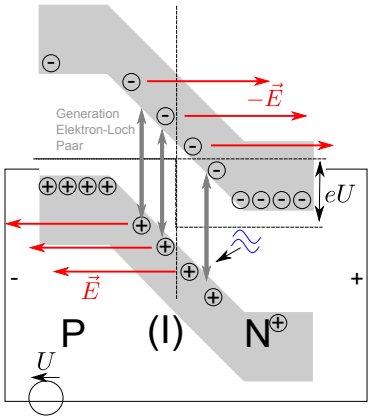


Abbildung 12.20: Bändermodell des generischen PN-Übergangs in einer Photodiode.

Hier könnte ggf. noch etwas zur internen und externen Modulation der LD kommen

12.2.2 Optische Empfängerstruktur

12.2.2.1 Photodioden

Die Interaktion von Licht mit angeregten Elektronen, wie sie bei der LED oder der Laser-Diode ausgenutzt wurde, lässt sich umkehren, um optische Signale zu empfangen. Hierzu wird im Prinzip nur ein PN-Übergang in Sperr-Richtung vorgespannt, wie in Abb. 12.20 [57, 7, 58]. Eigentlich fließt bei der dargestellten Vorspannung kein Strom, die Verarmungszone ist ausgeräumt. Trifft jedoch ein Photon mit einer Frequenz und damit Energie $hf > eU_G$, größer als die Bandlücke, dann wird ein Elektron aus dem Valenzband heraus in das Leitungsband angeregt. Ein Elektronen-Loch-Paar entsteht. In dem elektrischen Feld werden die Ladungsträger getrennt und es kommt zu einem Stromfluss, der

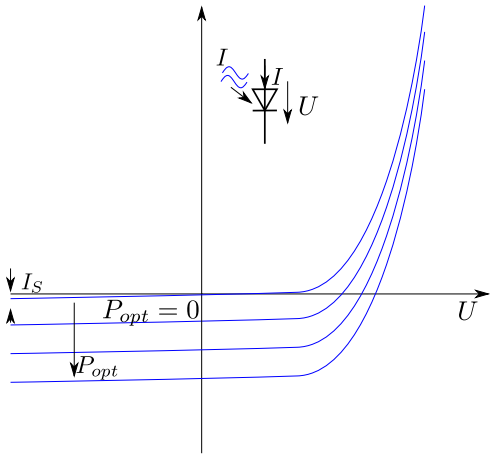


Abbildung 12.21: Kennlinienfeld der Photodiode.

detektiert werden kann. Alle Photonen, die genügend Energie haben, die Elektronen und Löcher zu trennen werden detektiert. Überschüssige Energie wird durch Gitterstöße abgebaut. Daher ist ein Photodetektor in der Regel ohne weitere Maßnahmen auch anders als LED oder Laser-Dioden breitbandig, wenn auch ggf. für eine bestimmte Wellenlänge optimiert.
Die Kennlinie der Photodiode ist in Abb. 12.21 dargestellt und verläuft nach

$$I(U,P_{opt}) = I_S \times \left(e^{-\frac{eU}{kT}} - 1\right) - I_P, \tag{12.23}$$

worin I_P den von der Lichtleistung abhängigen Photostrom darstellt, der die Kennlinie nach unten verschiebt. Abb. 12.21 legt auch nahe, dass man die Diode ohne irgendeine Vorspannung als Photozelle verwenden kann, was zunächst vorteilhaft, weil sehr einfach, erscheint. Die Vorspannung in Sperrichtung bringt jedoch mehr Vorteile, weil

- das elektrische Feld die Ladungsträger beschleunigt und daher die Ansprechzeitverringert und die Geschwindigkeit erhöht,
- die Sperrspannung die Verarmungszone vergrößert und daher die Kapazität senkt, was wiederum die Reaktionsgeschwindigkeit erhöht,
- die größere Verarmungszone zu einem größeren Bereich führt, in dem die Ladungsträger zum Photostrom beitragen und dadurch die Empfindlichkeit erhöht.

P-I-N-Photodiode Die Vorteile der PN-Schicht in Sperrichtung können erweitert werden, wenn man technologisch die Verarmungszone weiter durch eine intrinsische Halbleiterschicht von 50-100 μm Dicke vergrößert. Dieses führt zu einer weiter verkleinerten Kapazität und damit höheren Geschwindigkeit. Der Aufbau der Diode ist in Abb. 12.22(c)

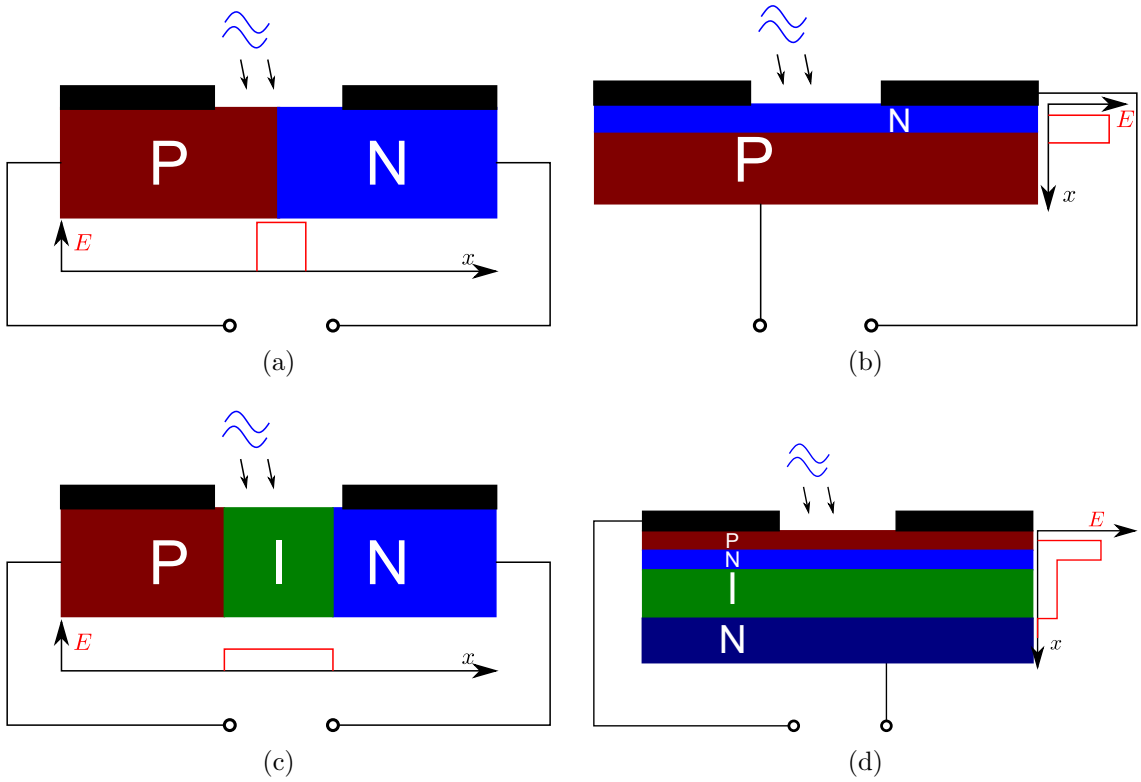


Abbildung 12.22: Schematischer Aufbau einer PN-Photodiode (a) auch horizontal (b) und der PIN-Diode **evtl auch planar zeichnen [58, S.303]** (c) sowie der durch die P-Zone bestrahlte Avalanche Photo Diode (APD) (d). Feldstärkeverläufe und alle anderen Dimensionen sind nicht maßstabsgetreu.

gezeigt. Außer der Verringerung der Kapazität erhält man auch einen größeren photoempfindlichen Bereich. Die große I-Schicht führt allerdings zu großen Transitzeiten der Ladungsträger, so dass hier Optimierungsbedarf besteht. Eine Änderung des Materialsystems vom Silizium mit seinen relativ geringen Beweglichkeiten hin zu In Ga As oder In P mit entsprechend höherer Beweglichkeit der Ladungsträger ist die Lösung.

APD Photodiode Eine weitere Erhöhung der Sperrspannung an der Diode führt zu einer weiteren Erhöhung des elektrischen Feldes [57, S.414]. Dadurch werden die in der photoempfindlichen Schicht erzeugten Elektronen und Löcher so stark beschleunigt, dass ihre kinetische Energie ihrerseits ausreicht, weitere Elektronen aus dem Valenzband heraus zu schlagen. Diese werden nun wiederum beschleunigt, bis der Prozess sich wiederholt, oder das Feld am Ende der aktiven Schicht verschwindet. Aus jedem einzelnen Elektronen-Loch-Paar werden so also durch einen Lawinenprozess viele (Abb. 12.23).

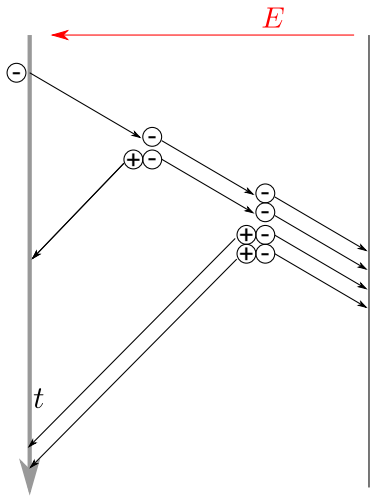
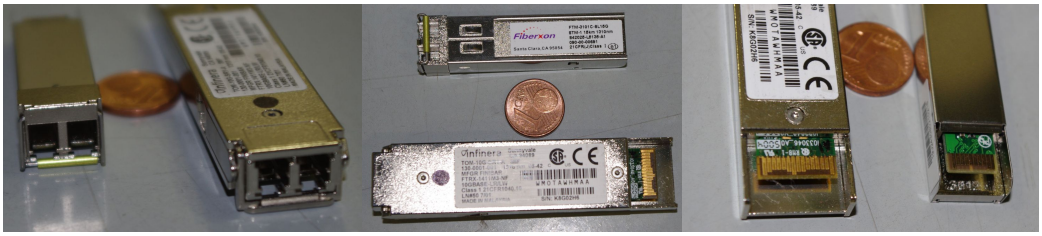
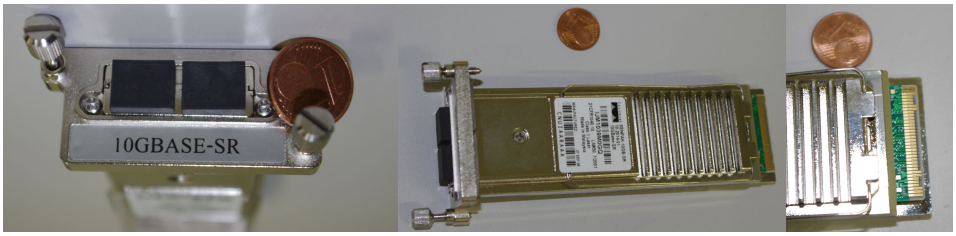


Abbildung 12.23: Elektronenvervielfachung (Löcher führen nicht zu neuen Paaren) in einer APD. Löcherbeweglichkeit geringer als Elektronenbeweglichkeit.



(a) (b) (c)



(d) (e) (f)

Abbildung 12.24: Obere Reihe: Standard SFP (der kleinere) [59, 60] und XFP aus verschiedenen Positionen und unten Xenpak [61]. (Fotos: selbst)

Die Emfindlichkeit der APD ist hoch, weshalb sie auch gerade im Weitverkehr eingesetzt wird, wenn nach einer langen Strecke die optische Leistung gering sein kann.

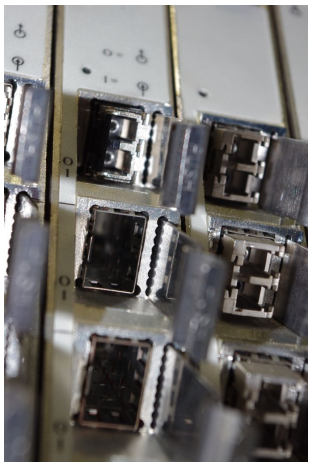
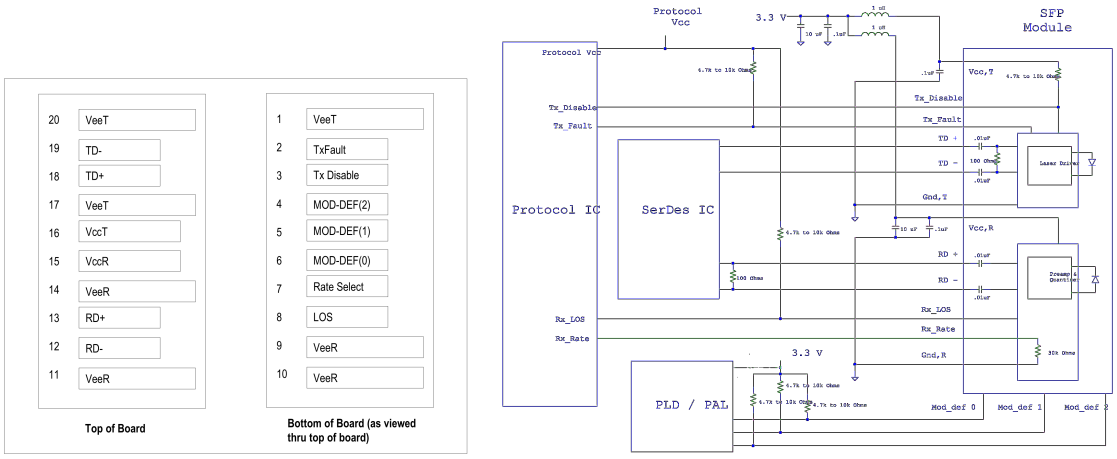


Abbildung 12.25: SFP in einem Board (links oben und rechts) und leere Steckplätze.



(a) (b)

Abbildung 12.26: Interface (a) und Beschaltung (b) von SFP.

12.2.3 SFP-Module

Laser-Dioden werden heute meist als SFP Module mit verschiedenen Eigenschaften geliefert [59]. Diese Module haben einen optischen oder auch elektrischen kompletten Transceiver „on board“, der auf ein standardisiertes Interface umgesetzt ist. Verschiedene Varianten dieser Einschübe sind in Abb. 12.24 gezeigt. Der Vorteil ist, dass man den direkten physikalischen Zugriff auf das Medium von einer weitgehend herstellerabhängigen Implementierung getrennt hat. Ein Netzelement kann so mit einem Hauptboard verschiedene Interfaces bedienen, die einfach in standardisierte Steckplätze gesteckt werden (Abb. 12.25). Diese können beispielsweise elektrische (10-Base-T bis 1000Base-T) sein, oder auch op-

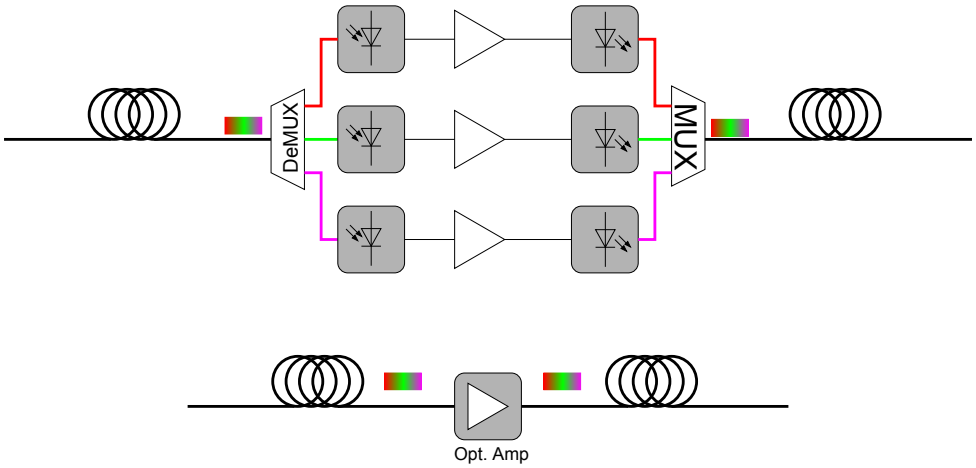
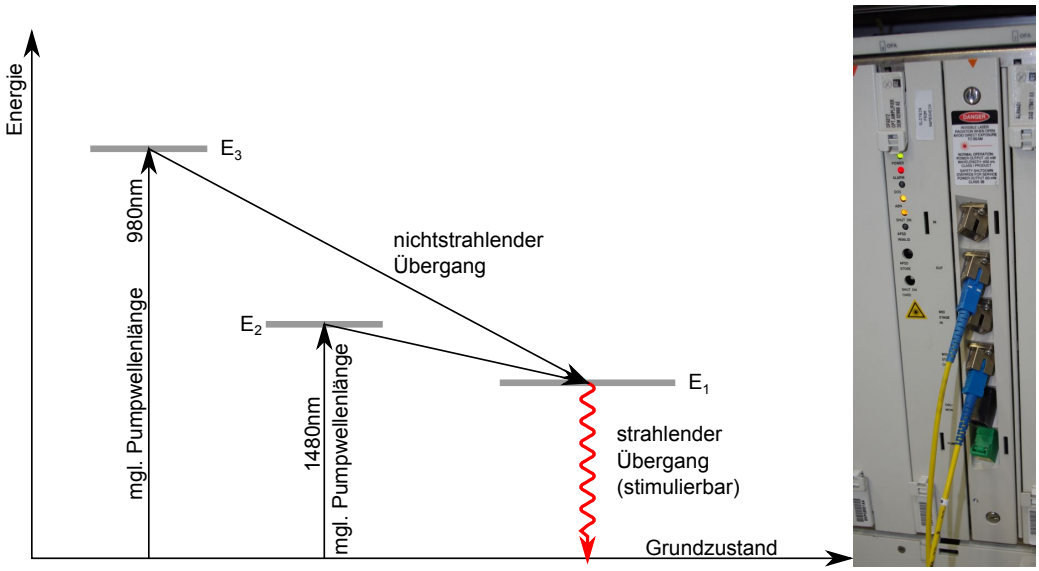


Abbildung 12.27: Signal-Regeneration ohne optischen Verstärker (oben) und mit (unten).

tische in verschiedenen Wellenlängenbereichen bis hin zu verschiedenen Wellenlängen im WDM.
Noch mehr zu Performance und Beschaltung der SFP einfügen.

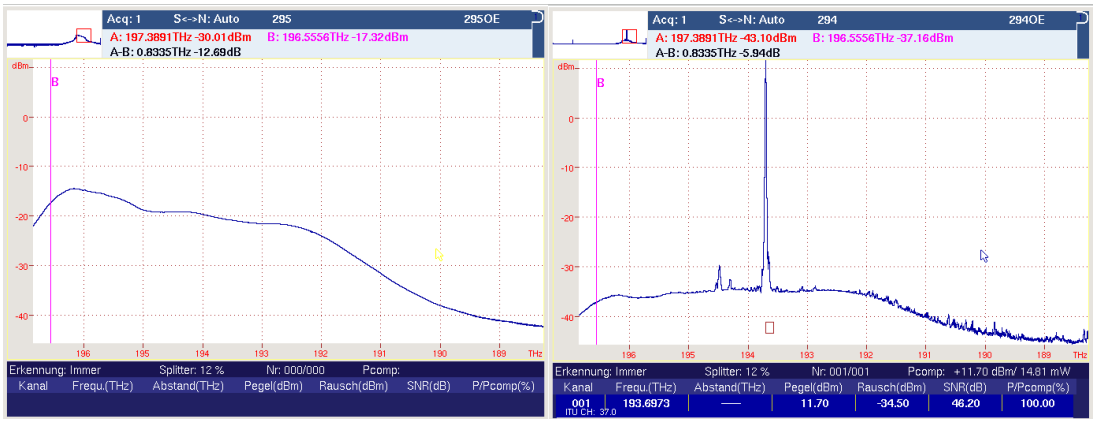
12.2.4 Optische Verstärker (EDFA)

Das wohl entscheidende Element im Wellenlängenmultiplex ist der optische Verstärker. Erst mit ihm ist das WDM sinnvoll möglich geworden. Der optische Verstärker biete nämlich die Möglichkeit in einer Baugruppe die optischen Signale alle gleichzeitig zu verstärken, ohne eine Aufwändige optisch-elektrische Wandlung, Regeneration, und wiederum elektro-optische Wandlung für alle Kanäle vollziehen zu müssen (Abb. 12.27).
Das Prinzip des Optical Fiber Amplifier (OFA) ist dem des Lasers vergleichbar, jedoch ist das aktive Medium ein anderes. Hier handelt es sich um eine Glasfaser von einigen 10 m Länge mit einem Kern vom Durchmesser von ca. 2 μm. Dieser Kern ist ist mit einen Seltenen-Erd-Material (in der Regel) Erbium (Er^{3+}) dotiert. Die Konzentration des Erbium beträgt in etwa $10^{24}/m^3$. Dieser Aufbau gibt dem OFA auch den Namen EDFA. Die praktisch wesentlichen Energiezustände in einem EDFA sind in Abb. 12.28 dargestellt. Durch einen Pumplaser (optisches Pumpen) bei 1480 nm oder 980 nm Wellenlänge, der Leistungen im 100 mW bereitstellt, werden die Atome auf die Niveaus E_2 bzw. E_3 angeregt. Durch nichtstrahlende Übergänge fallen die Elektronen auf den Zustand E_1 zurück, der etwas unterhalb des zweiten Niveaus liegt. Der jetzt erreichte Zustand ist relativ stabil. Ein Übergang von hier zum Grundzustand erfolgt strahlend, so dass es zu Amplified Spontaneous Emission (ASE) kommt (Abb. 12.29(a)). Dieses ist im System-Sinne natürlich Rauschen und daher unerwünscht. Wird ein optisches Signal angelegt, so sti-



(a) (b)

Abbildung 12.28: Genutzte Energieniveaus in einem Erbium Doped Fiber Amplifier (EDFA) (a) und eine solche Baugruppe (b) im Test.



(a) (b)

Abbildung 12.29: Ausgangsspektrum eines EDFA ohne (a) und mit (b) Signal. Das ASE ist gut zu erkennen. (Justage der Leistungsachse gemäß Splitter 12 % notwendig).

mulieren diese eingespeisten Photonen den strahlenden Übergang, wie im Laser, so dass das eingespeiste Signal den OFA verstärkt verlässt (Abb. 12.29(b)). Da die Stimulation die gewünschten Übergänge gegenüber den spontanen bevorzugt, geht das ASE-Rauschen dadurch signifikant zurück.

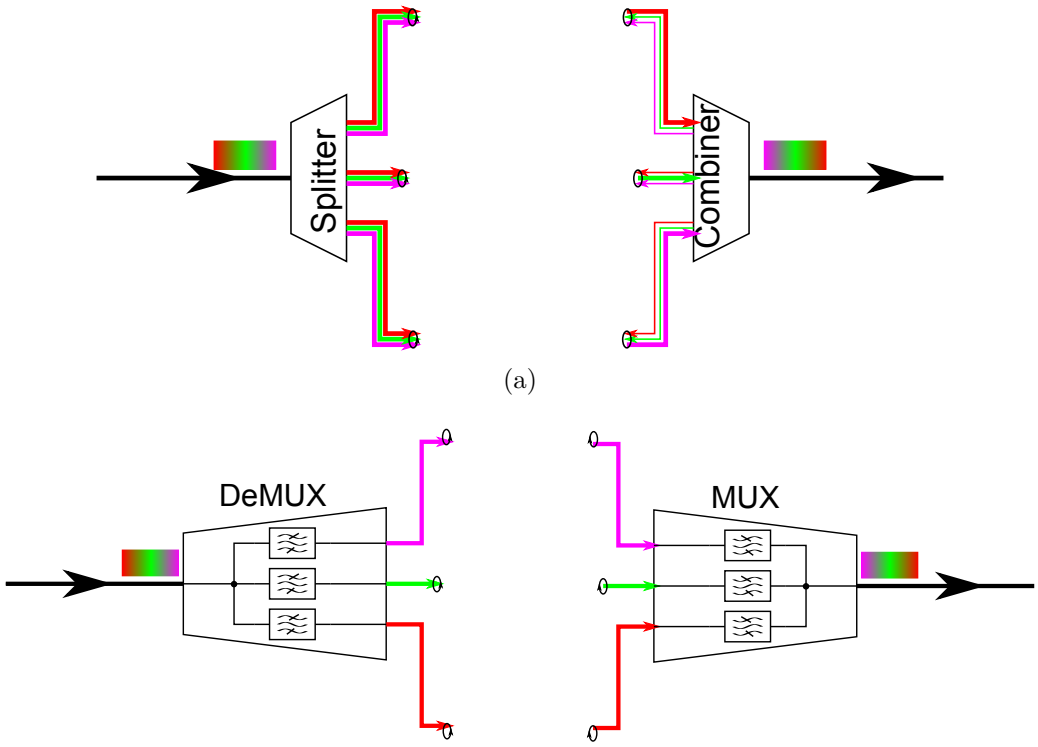


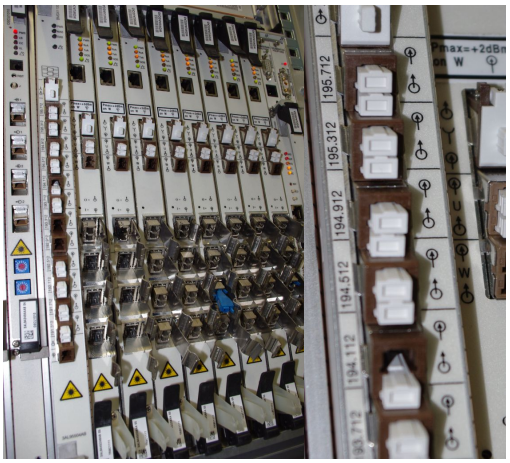
Abbildung 12.30: Combiner und Splitter (a) und selektive Multiplexer und De-Multiplexer (b) für WDM Systeme.

Die Bandbreite eines EDFA beträgt in etwa 4 THz, was einem Wellenlängenbeiche von 1530 nm ($f \approx 192$ THz) bis 1565 nm ($f \approx 196$ THz) entspricht, nach 12.3 also über fünfzig 50 GHz Kanälen.

12.2.5 Optische Filter (MUX, DeMUX)

Laser können immer nur ein Signal bei einer Wellenlänge erzeugen. Am Ausgang einer einfachen optischen Baugruppe liegt also ein durchaus breitbandiges Signal bei einer Wellenlänge (Farbe) vor (siehe auch Abb. 12.18(b)). Wie in der Funktechnik auch, muss dieses Signal auf das physikalische Medium, die Faser, gebracht werden, ohne die Signale bei anderen Farben zu beeinflussen. Da eine aktive Baugruppe in der Regel immer nur das Signal bei einer Wellenlänge verarbeiten kann, muss das mehrfarbige Signal auch wieder möglichst verlustfrei vom Medium abgenommen werden können.
Grundsätzlich kann dieses in zwei unterschiedlichen Technologien erfolgen:

- Die Signale werden breitbandig kombiniert und verteilt (Abb.12.30(a)). Die eigent-



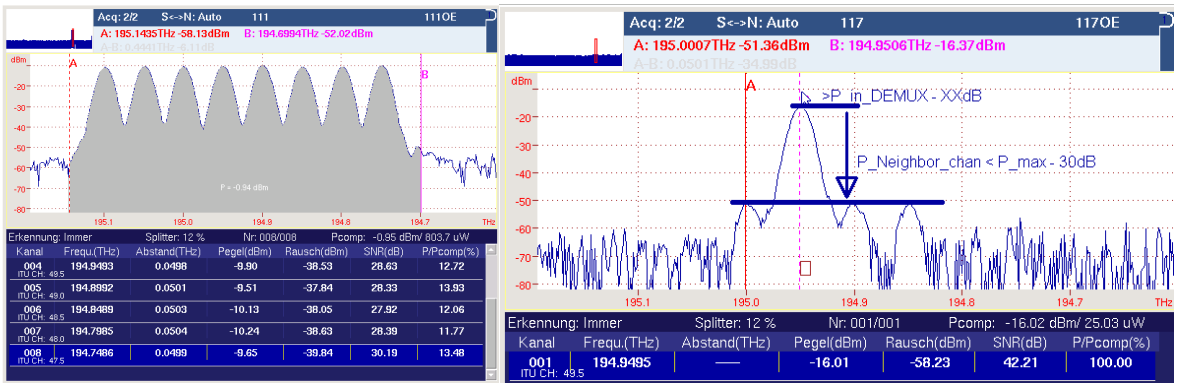
(a) (b)

Abbildung 12.31: MUX-DeMUX Baugruppe (links) im Feld der Elektro-optischen Baugruppen (nicht gesteckt) (a) und vergrößert zur Vorstellung der Wellenlängenbereiche (b)

liche Baugruppe (Combiner oder Splitter) ist sehr einfach aufzubauen. Der Nachteil ist, dass die Wellenlängenselektion beim Empfang in der nachgeschalteten aktiven elektro-optischen Baugruppe erfolgen muss, und das insbesondere beim Spitter empfangsseitig ede Farbe in jeden Zweig aufgespalten werden muss. Selbst eine verlustfreie Baugruppe fügt also dem Signal eine Dämpfung von $10\log(N)$ zum was schon bei 1:8 auf 9 dB Dämpfung führt (optische Verluste nicht berücksichtigt, jede Verdopplung von N fügt 3 dB mehr Dämpfung zu). Trotz aller Einfachheit werden solche Systeme heute nicht mehr oder nur noch selten eingesetzt.

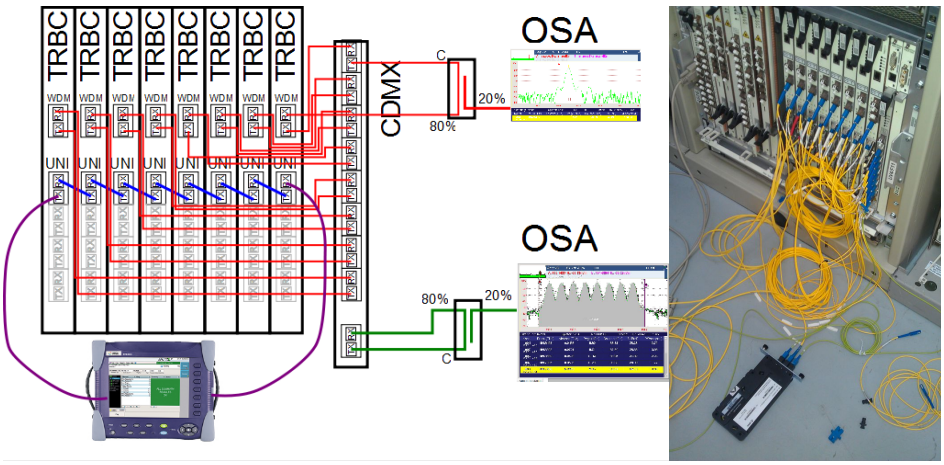
- Echte Wellenlängenselektive Multiplexer und De-Multiplexer (Abb. 12.30(b)). Diese bestehen aus N Filtern, die das Signal auf einen Punkt führen, bzw. wellenlängenselektiv aufspalten. Der Vorteil ist, dass die jeweils nicht im Pfad befindlichen Farben nicht beeinflusst werden, und somit in jedem betrachteten Pfad nur die optischen Verluste des Filters auftreten, keine systembedingten Verteilungsdämpfungen. Nachteil ist, dass solche MUX/ DeMUX nur für die entwickelte Wellenlängenkombinations verwendet werden können, und sie natürlich weit aufwändiger sind.

Eine Baugruppe des Systems Alcatel 1626 ist in Abb. 12.31 gezeigt. Dort ist die Multiplexer-Demultiplexer Einheit (eine Steckkarte) in einem Feld der Elektro-optischen Baugruppen gezeigt und in einer Vergrößerung werden die einzelnen Ports den Wellenlängen zugeordnet.
Messergebnisse für DWDM-Signale vor und nach dem MUX und DeMUX sind in Abb. 12.32



(a) (b)

Abbildung 12.32: Acht-Kanaliges DWDM-Signal im 50 GHz-Raster bei einem Alcatel 1626-System (a) nach dem Multiplexer und (b) dasselbe Signal wieder nach dem De-Multiplexer,



(a) (b)

Abbildung 12.33: Möglicher Testaufbau schematisch (a) und als Verkabelung (b) eines DWDM-Systems im Labor.

gezeigt. Die Wirkung der DeMUX-Einheit ist klar zu erkennen, die Nachbarkanäle werden sehr wirkungsvoll unterdrückt. Ein möglicher Testaufbau zur Überprüfung der spektralen Reinheit von Baugruppen, der Wirkung der Filtereinheiten sowie der Bitfehlerrate ist in Abb. 12.33 gezeigt.

Hier fehlen noch die einzelnen Technologien zur Aufspaltung der Signale. Insgesamt ist die detaillierte technische Beschreibung der optischen Verfahren und physikalischen Effekte kurz gekommen, es wurde der Systemgedanke zu-

nehmend in den Vordergrund gestellt.

12.3 Optische Übertragungsverfahren und 100 GBit/s

Hier können noch Punkte zur Übertragung sehr hoher Bitraten folgen, die unter dem Stichpunkt 100 GBit/s zusammengefasst werden können. Stichpunkte sind

- Datenstruktur (SDH/PDH und die ODU/ODP)
- Übertragung auf mehreren Farben und auf einer Farbe
- Kohärenter optischer Empfang, Technik und Gewinn, damit einhergehend
  - Modulationsverfahren (QPSK) auch bei Optik
  - Polariationsmultiplexing

  Womit dann 4 Bit pro Symbol erreicht werden
- evtl. noch ausgefeilte Techniken der Glasfasern
- Referenzen zur möglichen Kanalkapazität, dass nämlich die Kapazität nach Shannon bei sicheren SNR und hohen Bandbreiten (50 GHz-Kanalraster) lange nicht ausgenutzt werden.

12.3.1 Optische Filter im DWDM (ROADM)

Erweiterung für weitere Semester kann die Vorstellung rein optischer Netzelemente wie ROADM sein

Anhang A

Simulationstechniken

In diesem Abschnitt werden wir auf Simulationstechniken eingehen und beispielhaft einige Simulationen durchführen und selber programmieren.

A.1 Simulationsumgebung

Am Beginn, aber natürlich nach der Problemanalyse, steht die Auswahl einer geeigneten Plattform und der Simulationsumgebung. Grundsätzlich gibt es (mindestens) drei verschiedene Herangehensweisen:

1. Textuelle Programmierung (z.B. C/C++, Java, Matlab, Python u.ä.)
2. Grafische Programmierung (Simulink, Ptolemy, NI-Labview usw.)
3. Spezielle Simulationstools (WinIQ-SIm, ADS usw.)

auf einige Beispiel wird nun im folgenden eingegangen:

A.1.1 Matlab und Clone Gnu-Octave

Basis einer Matlab-Simulation ist eine textuell erstellte .m-Datei. Matlab stellt einen Interpreter dar. Die Programmiersprache, also die .m-Dateien sind quasi standardisiert, in vielen wissenschaftlichen Einrichtungen und vielen Unternehmen ist Matlab verbreitet. Matlab selber steht für "Matrix Laboratory" und unterstützt in diesem Sinne vor allem die Arbeit mit Matrizen, die ohnehin die Basis jeder numerischen Simulation sind. Matlab hat wenig Optionen zur symbolischen Rechnung, ist aber eben sehr effektiv wenn es um Matrix-Manipulationen geht. In Matlab sind einzelne Funktionen in einzelnen Dateien abzulegen. Da diese in der Regel sehr kurz sind, kommt es bei komplexeren Matlab

Projekten sehr schnell zu einer Unzahl von Dateien. Beachtet man die Matrix-Orientierte Struktur nicht, und weicht stattdessen auf Schleifenprogrammierung aus, wird Matlab sehr langsam. Matlab kommt mit einer eigenen, sehr mächtigen Entwicklungsumgebung und biete viele grafische Möglichkeiten, von einfachen Plots hin zu komplexen Grafischen Nutzer Oberflächen. Viele, auch sehr mächtige und spezialisierte Funktionen und Pakete sind für Matlab verfügbar, so dass man auf einem sauberen und erprobten Standard aufsetzen kann.
Größter Nachteil von Matlab ist die - außerhalb der Bildungseinrichtungen - assoziale Lizensierungspolitik von "The Mathworks", so dass eine Matlabinstallation schnell sehr teuer werden kann und "Wartungskosten" verschlingt. Daher kann die frei verfügbare Alternative "Gnu-Octave" verwendet werden. Dieses Paket kann einfach installiert werden. Gnu-Octave unterstützt .m-Code und so sind Matlabprogramme oft ohne jede Änderung in Octave lauffähig und umgekehrt. Folgende Nachteile gegenüber dem kommerziellen Bruder sind aber in Kauf zu nehmen:

- Nicht alle Pakete wie in Matlab sind vorhanden oder in der gleichen Qualität vorhanden
- GUI und objektorientierte Programmierung sind mit Gnu-Octave nicht möglich, es bleiben also nur die "Plot"-Funktionen, die aber auch hier sehr mächtig sind
- Octave ist langsamer (weniger effizient), das kann den Faktor 2-5 ausmachen.
- Octave ist rein textuell und die Programme werden nur in einem Texteditor geschrieben, es wird keine IDE mit geliefert, es ist kein sinnvoller Debugger vorhanden.

Verfügbarkeit:

- Matlab ist unter http://www.mathworks.de für viel Geld und vollständig verfügbar
- Gnu-Octave ist unter (http://www.gnu.org/software/octave/) ebenso vollständig, aber für kein Geld verfügbar. Ein Texteditor ist noch notwendig, z.B. der Emacs (http://www.gnu.org/software/emacs/) oder auch Notepad++ (http://notepad-plus-plus.org/)

A.1.2 C/C++ und Java

Dieses sind vollständige und lange und gut etablierte Programmiersprachen für eigentlich alle denkbaren Anwendungen. C++ und Java unterstützen vollständig die objektorientierte Programmierung, während das ANSI-C eine einfache strukturierte Programmiersprache

ist, die gerne in der Programmierung von Embedded Systemen verwendet wird. Alle Varianten sind Compiler-Sprachen, während Java bekanntlich nur auf einen Zwischencode compiliert wird, um dann auf einer virtuellen Maschine ausgeführt zu werden. Dieses ist die Stärke von Java; ein Java-Programm ist plattformunabhängig und kann so leicht zwischen allen denkbaren Betriebssystemen transferriert werden.
Entwickelt werden die Programme in diesen Sprache mit Hilfe von Integrated Development Environment (IDE) Integrierten Entwicklungsumgebungen, z.B. eignet sich besonders Netbeans (http://netbeans.org/) für Java, während Eclipse (http://eclipse.org/) eine mögliche Umgebung mit vielen Plug-Ins für C/C++ darstellt. Beide Varianten inkl. der zugehörigen Compiler sind frei verfügbar und in großen Teilen auch quellenoffen. Natürlich gibt es auch kommerzielle Pakete, insb. für C/C++ wie z.B. von Microsoft oder Borland. Mit den o.g. offenen C/C++-Varianten gestaltet sich die grafische Darstellung und die Programmierung von GUI etwas schwieriger, Java ist hier mit Netbeans vorbildlich einfach, aber das ist ggf. auch Geschmackssache und hängt stark von den zusätzlich verwendeten Paketen ab.
Als allgemeine Programmiersprachen lassen C/C++ und Java natürlich prinzipiell jede Art von Programm/ Simulation zu, unterstützen aber nicht unbedingt die Entwicklung von numerischen Simulationen. So fehlen z.B. in allen drei Sprachen die native Implementierungen von komplexen Zahlen, etwas was 1977 schon in FORTRAN selbstverständlich war. Aber natürlich lassen sich diese mit geeigneten Bibliotheken und Paketen nachträglich besorgen. Ebenso müssen numerische und kommunikationstechnische Funktionen entweder selbst geschrieben werden oder über spezielle Bibliotheken eingebunden werden. Dieses ist in viele Fällen möglich, wenn auch nicht immer durchgestochen und durchgängig verfügbar.
Hauptvorteil dieser echten, compilerorientierten Programmiersprachen ist die hohe Effizienz und Geschwindigkeit in der Ausführung. Der Nachteil ist die Notwendigkeit einer sehr kleinteiligen Entwicklung und damit ggf. eine hohe Fehlermöglichkeit und damit eine recht langsame und langwierige Entwicklung der Programme.

Verfügbarkeit

C/C++ Wer Linux verwendet ist fein raus und hat schon fast alles... Unter Windows ist mehr Arbeit nötig:

- Als Compiler bietet sich der gcc an, den es entweder (dann als Linux-Emulation unter Windows) in Cygwin (http://cygwin.org/) oder als Windows-Variante in

MinGW (http://mingw.org/) gibt. Beide Compiler sind tauglich und können mit der nun kommenden IDE bedient werden.

- Als vollständige IDE bietet sich Eclipse CDT (C Developers Toolkit) (http://eclipse.org/) an. Das zusammen ist dann auch quasi vollständig und eignet sich schon um Konsolen-Programme zu entwickeln.
- GUI-Entwicklung setzt eine Erweiterung, z.B. Qt (http://qt.nokia.com/) voraus, doch dieses steht hier nicht um Fokus
- Numerische Bibiotheken für C/C++ sind tlw. seit Jahrzehnten verfügbar, doch erfordert ihre Intergration/ Installation ggf. eine eigene Compilation und demenstsprechend einigen Aufwand:
  - Basis für lineare Algebra: http://www.netlib.org/blas
  - Lineare Algebra (Matrizen, Gleichung etc): http://www.netlib.org/lapack
  - Erweiterte Bibiotheken "Armadillo" unter http://arma.sourceforge.net/
  - Kommunikationstechnische Programme wie Kodierer/Dekodierer gibt es unter http://sourceforge.net/apps/wordpress/itpp/
  - Oft notwendige Fouriertransformationen findet man als "Fastest Fourier Transform of the West" in http://www.fftw.org/

  An der Vielzahl und Komplexität dieser Liste sieht man schon: Effizienz (des hinterher erstellten Programmes) hat seinen Preis, der leider meist nicht lohnt

Java

- Java und das Java Development Kit (JDK) findet man unter `http://www.oracle.com/technetwork/java/javase/downloads/index.html` bzw. als allgemeine Einstiegsseite `http://www.java.com`
- die IDE Netbeans gibt es unter `http://netbeans.org`
- Zum Zeichnen (von Kurven etc.) bietet sich ggf. JFreeChart (http://www.jfree.org/jfreechart/) an
- Numerische Bibiotheken sind ebenfalls verfügbar, allerdings kann und will ich hier keinen eindeutigen Hinweis geben, aus meiner Sicht sind diese für hohe numerische Ansprüche wenig genügend.

A.1.3 Python

Dieses ist eine frei und quellenoffen verfügbare Skript-Sprache, die vollständig objektorientierte Entwicklung ermöglicht. Damit nimmt Python eine Zwischenstellung zwischen

Matlab/GNU-Octave auf der einen und C/C++ und Java auf der anderen Seite ein. Python biete - tlw. über Pakete - viele Bibliotheken zur numerischen Mathematik und zur grafischen Ausgabe/ Unterstützung an und kann als Interpretersprache interaktiv ausgeführt werden. Python, und das ist ein Vorteil gegenüber Gnu-Octave, wird von einer richtigen Entwicklungsumgebung unterstützt, der Eclipse und lässt sich so recht effizient entwickeln. Ein Debuggen ist durch Debuggerunterstützung (GDB) leicht möglich. Python unterstützt natürlich die sequentielle Programmierung a la C/C++ und Java aber auch tlw. die Vektorisierung wie in Matlab, letztere aber lange nicht so konsequent, wie Matlab das tut.
Die Effizienz der Python-Simulationen liegt irgendwo zwischen Matlab und Gnu-Octave.

Verfügbarkeit:

- Python selber gibt es unter http://www.python.de/ oder http://www.python.org/ . Damit wird nur der textbasierte Interpreter installiert.
- Mit Eclipse (s.o.) steht eine Entwicklungsumgebung zur Verfügung, die mittels PyDev (http://pydev.org/) auch Python unterstützt.
- Grafische Unterstützung erfolgt durch Matplotlib http://matplotlib.sourceforge.net/
- und numerische durch NumPy (http://numpy.scipy.org/), die beide noch nachinstalliert werden müssen, und in die Programme eingebunden werden müssen, was aber kein Problem ist.

A.1.4 Ptolemy

Wer gerne die Maus schubst und keine Texte tippen möchte, sondern grafisch und am Datenfluss orientiert arbeitet, der findet mit dem ptolemy-Projekt der University of California at Berkely sein Objekt der Begierde. Auf einer Java-Plattform werden verschiedene Datenflussmodelle zur Verfügung gestellt.

Verfügbarkeit

- Das gesamte Paket ist unter `http://ptolemy.berkeley.edu/index.htm` zu bekommen und mit vielen - auch kommunikationstechnischen Beispielen ausgestattet

A.1.5 LabView

Eine kommerzielle, sehr auf die Automatisierung von Testprozessen ausgerichtete und professionelle (innerhalb und außerhalb der Hochschule auch teure) grafische Programmierumgebung ist National Instruments Labview. Auch hiermit lassen sich am Datenfluss orientiert "Programme" mittels Klick erstellen. Dieses ist recht effektiv. Für Labview sind viele Pakete, auch für die Kommunikationstechnik, verfügbar. Der Vorteil der visuellen Programmierung wird erkauft mit - wenn schlecht geschrieben - schlechter Übersichtlichkeit und tlw. tiefer Verschachtelung in Sub-Vi (VI=Virtuelle Instrumente). Weiterer Nachteil ist, dass mit LabView und seinen Baukästen alles das, was vorgesehen ist gut geht, aber eine allgemeine Problemlösung, vor allem in komplizierten und kleinen Details, dann doch textuell, z.B. in eingebundenem C/C++ geschehen muss. Natürlich wiegt der Vorteil der von vorneherein grafischen Oberfläche und grafischen Ausgabe stark.

Verfügbarkeit

- LabView (als Testversion oder zum Kauf) ist unter http://www.ni.com/labview/d/ verfügbar.

A.2 Simulation

A.2.1 Datenerzeugung, die PN-Sequenz

Basis der Simulation von kommunikationstechnischen Systemen ist oft ein Datenstrom von Bits. Dieser Strom sollte möglichst die statistischen und spektralen Eigenschaften von zufälligen Daten haben, also nicht etwa periodisch sein, oder schlimmer noch nur als Nullen oder Einsen bestehen. Auf der anderen Seite wäre es aber trotzdem gut, wenn der Datenstrom fest und bekannt wäre, so dass man mit wenigen Parametern den Strom eindeutig beschreibt.
Nun, so etwas ist als Pseudo Noise (PN)-Sequenz oder auch als PRBS bekannt und wird durch eine einfache Anordnung rückgekoppelter Schieberegister erzeugt (Abb. A.1). Die eindeutige Definition der entstehenden Sequenz ergibt sich aus

- Anzahl der Register (Länge) m
- Position der Rückkopperlungswege und dem daraus resultierenden binären Wort
- Start (Seed) des Schieberegisters, das natürlich niemals Null sein darf. Auch dieses ist ein binäres Wort.

Die Implementation in Gnu-Octave ist bspw. wie folgt gegeben, wobei ein Bit als eine komplette Zahl (ein Element des Ausgabefeldes pn) dargestellt wurde.

```
function pn=pn_sequence(n,m,s,fb)
% Simulates a PN-Sequence generator with
% n: length of the generator
% m: number of states to be simulated
% s: Start loading (must be vector of length n)
% fb: feedback (must be vector of length m)
r=s;              % set the initial contents
pn=zeros(1,m);
for i = 1:m
   f=rem(sum(r.*fb),2);              \% calculate feedback value
   pn(i)=r(n);
   r(2:n)=r(1:n-1);
   r(1)=f;
end
```

Die Implementation in Python dagegen ist

```
import numpy as np

## Function pn_sequence, generates a pn_sequence and
# returns the bit-pattern, where
# effectively one bit is represented in one integer,
# so this might be optimized...
#
# Examples for a complete PN9 and a PN7 sequence
#     s=np.array([1,0,0,0,0,0,0,0,0], dtype=np.byte)
#     fb=np.array([0,0,0,1,0,0,0,0,1], dtype=np.byte)
#     pn=pn_sequence(9,511,s,fb])
#
#     s=np.array([1,0,0,0,0,0,0,], dtype=np.byte)
#     fb=np.array([1,0,0,0,0,0,1], dtype=np.byte)
#     pn=pn_sequence(7,127,s,fb)
#
# @param[in] n length of the generator
```

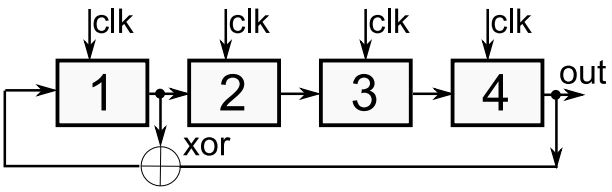


Abbildung A.1: Rückgekoppeltes Schieberegister zur Erzeugung einer PRBS

```
# @param[in]m: number of states to be simulated
# @param[in] s: Start loading (must be vector of length n)
# @param[in] fb: feedback (must be vector of length m)
def pn_sequence(n,m,s,fb):
    # set the initial contents
    r=s
    pn=np.zeros(m,dtype=type(s[0]))
    for ii in range(0,m):
        # calculate feedback value
        f=np.sum(r*fb)%2
        pn[ii]=r[n-1]
        r[1:n]=r[0:n-1]
        r[0]=f
    return pn
```

Diese erscheint vor allem wegen der langen Kommentare etwas länger, im Kern ist es aber in etwa derselbe Aufwand. Die Einbindung von "numpy" als numerischer Bibliothek dient nur dazu, hier schon die effektiveren Arrays verwenden zu können, es wäre hier noch nicht nötig gewesen.

A.2.1.1 Eigenschaften von PN-Sequenzen

Mit einem Schieberegister der Länge N, also einem ebenso langem Binärwort, kann man insgesamt 2^N Zustände darstellen. Sinnvoll nutzbar sind jedoch nur $2^N - 1$, da der "Nullzustand" ungültig ist, denn er würde immer wieder auf sich selber abgebildet werden. Mit der o.g. Generation durch ein rückgekoppeltes Schieberegister ist damit die maximale Länge einer PRBS bis sie sich zwangläufig wiederholen muss $2^N - 1$. Diese maximale Menge kann nur für bestimmte Rückkoppelungszweige erreicht werden, oder anders ausgedrückt: Nicht jede beliebige Lage der Rückkoppelungs-Taps führt auch tatsächlich auf das maximal Erreichbare. In Tab. A.1 sind nach [5] jeweils die Konfigurationen für Rückkoppelungen in Schieberegistern aufgeführt, die zu der maximale langen Sequenz führen

Tabelle A.1: Tabelle minimal möglicher Rückkoppelungspolynome für rückgekoppelte Schieberegister der Länge N. Lage der Rückkopplerungen $\{i_r\}$ und darüber hinaus Anzahl möglicher verschiedener PN-Sequenzen (bei veränderten Rückkoppelungen).

N	$\{i_r\}$	Anzahl	N	$\{i_r\}$	Anzahl	N	$\{i_r\}$	Anzahl
2	1,2	1	10	2,10	60	18	7,18	8064
3	1,3	2	11	2,11	176	19	1,2,5,19	27594
4	1,4	2	12	1,4,6,12	144	20	3,20	24000
5	2,5	6	13	1,3,4,13	630	21	2,21	84672
6	1,6	6	14	1,6,10,14	756	22	1,22	120032
7	1,7	18	15	1,15	1800	23	5,23	356960
8	2,3,4,8	16	16	1,3,11,12,16	2048	24	1,2,7,24	276480
9	4,9	48	17	3,17	7710			

und die minimale Anzahl von Rückkoppelungspfaden haben. Daneben existieren noch weitere Konfigurationen (mit mehr und anderen) Rückkoppelungszweigen, die ebenso zu einer maximalen Länge der Sequenz führen. Die Anzahl verschiedener Umsetzungen ist angegeben.

A.2.1.2 Aufgaben

1. Programmieren Sie eine Schieberegisterdarstellung wie oben.

2. Experimentieren Sie mit verschiedenen Längen von Registern und verschiedenen Startwerten.

3. Bestimmen und zeichnen Sie für eine bipolare Umsetzung der PN-Sequenz mit $N = 5$ (also $0 \to -1$, $1 \to 1$) die Autokorrelationsfunktion

$$\mathrm{AKF}\left(p(\tau)\right) = \frac{1}{T}\int_0^T p(t)p(t+\tau)\,d\tau. \tag{A.1}$$

   Die AKF müssen Sie natürlich noch programmieren, die Zeiten $t = n\Delta t$ bzw. τ diskret auszuführen sind!

4. Bewerten Sie die Plots der AKF, was fällt Ihnen auf?

A.2.2 Übertragungskanal und Bitfehlerrate

A.2.2.1 Kanal mit Arbitrary Whit Gaussian Noise (AWGN)

Als nächstes ist nun ein Kanal zu simulieren. Dazu muss der vorstehend erzeugte Signalstrom

- moduliert werden, also umgesetzt werden auf eine bipolare, QPSK oder gar noch höherwertige Modulationsform,
- mit Rauschen überlagert werden, es wird also ein zufälliges, der Gauss'schen Statistik genügendes, ggf. komplexes Störsignal addiert.
- detektiert werden,
- mit dem ursprünglichen Strom verglichen werden.

Für die Erzeugung des Störrauschens stehen in Octave und Python die Funktionen **randn** bzw. aus dem Random-Paket in Python **import** random;random.gauss zur Verfügung, die genaue Syntax und die Parameter recherchieren Sie bitte selber. Das Signal-Rauschverhältnis ist der entscheidende Parameter und mit einer Signalenergie E_S von eins, die angenommen werden kann, wenn in einer BPSK die Amplituden "1" bzw. "-1" sind, bleibt noch die Energie, also die Standardabweichung σ des weissen Rauschens. Dieses ist $E_N = \sigma^2 = \mathrm{SNR} \times E_S$, also $\sigma = \sqrt{\mathrm{SNR}}$.
Die Entscheidung über ein empfangenes Bit wird hier, da noch keine Kanalkodierung verwendet wurde, natürlich hart getroffen. Bei einer BPSK ist also jede Amplitude > 0 eine "1" und jede andere eine "0" (bzw. "-1"). Im Vergleich zum bekannten, erzeugten Bitstrom kann nun die Anzahl der Bitfehler bestimmt werden.

A.2.2.2 Aufgaben

1. Bestimmen Sie die Bitfehlerrate einer BPSK in einem AWGN-Kanal (gauß'sches Rauschen) in Abhängigkeit vom Signal-Rausch-Abstand. Verwenden Sie dazu eine PN-9-Sequenz. Achten Sie darauf, dass Bitfehlerraten bis zu 10^{-4} sicher dargestellt werden können und nicht von Simulationslauf zu Simulationslauf zu großen Schwankungen unterliegen.
2. Stellen Sie das Ergebnis grafisch dar.
3. Verändern Sie die Simulation so, dass eine QPSK verwendet werden kann. Führen Sie die Zuordnung der Konstellationspunkte mit Gray-Kodierung durch.

A.2.2.3 Bitfehlerrate im Schwundkanal

In diesem Abschnitt soll ein Schwundkanal, bei dem keine direkte Sichtverbindung besteht simuliert werden. Hierzu ist ein Rayleighprozess nach aufzusetzen, der in Abschnitt 7.2.2 und Abb. 7.7 erläutert wurde. Die Umsetzung in eine Kanalsimulation ist ebenso dort dargestellt. Eine Darstellung des der spektralen Überhöhungen aufgrund des Dopplereffektes wird hier vernachlässigt.

A.2.2.4 Aufgaben

1. Kombinieren Sie zwei gauß'sche Rauschprozesse im Komplexen so, dass für den Betrag eine Rayleighverteilung ensteht. Konkret: Setzen Sie Abb. 7.7(a) in ein Programm um, stellen Sie das Ergebnis grafisch, auch als Histogramm dar!

2. Simulieren Sie - unter Vernachlässigung der Dopplerverschiebung und Idealisierung der Tiefpassfilterung - die Bitfehlerrate in einem Schwundkanal nach Abb. 7.7(b). Gehen Sie

   (a) bei einem schnellen Rayleigh-Kanal von jeweils einem neuen Wert pro gesandtes Bit aus, und

   (b) bei einem langsamen Rayleigh-Kanal, wenn sich nur bei jedem zehnten/ hundertsten/ tausendsten Bit ein neuer Wert des Rayleigh-Fadings. Beachten Sie, dass zwischen den Fading-werten linear zu interpolieren ist!

3. Stellen Sie die Bitfehlerraten für unterschiedliche und sinnvolle Signal-Rausch-Verhältnisse dar!

An dieser Stelle sollte sich ergeben, dass aufgrund der gleichförmigen Verteilung der Phase des Rayleigh-Prozesses überhaupt kein Empfang möglich ist. Natürlich können Sie bei den langsamen Prozessen einmal Glück haben und ein Bitstrom wird fehlerfrei übertragen, aber das ist in keiner Weise sicher. Prinzipiell müssen Sie die Phase des Übertragungskanals schätzen und dann den Effekt entsprechend kompensieren.

Literatur

[1] John W. Eaton. *Gnu-Octave*. 2011. URL: http://www.gnu.org/software/octave/.

[2] Rohde and Schwarz. *R&S® WinIQSIM™ Simulation Software*. 2011. URL: http://www2.rohde-schwarz.com/product/WinIQSIM.html.

[3] Rohde and Schwarz. *R&S® WinIQSIM2™ Simulation Software*. 2011. URL: http://www2.rohde-schwarz.com/product/WinIQSIM2.html.

[4] Edward A. Lee. *The Ptolemy Project*. 2013. URL: http://ptolemy.berkeley.edu/ (besucht am 16. 01. 2013).

[5] Karl-Dirk Kammeyer. *Nachrichtenübertragung*. 5. Aufl. Stuttgart: Vieweg+Teubner, 2011.

[6] Carsten Roppel. *Grundlagen der digitalen Nachrichtenübertragung*. Carl Hanser Verlag, 2006.

[7] Volkmar Brückner. *Elemente Optischer Netze*. 2. Aufl. Vieweg und Teubner Verlag, 2011.

[8] Jörg Rech. *Ethernet*. 2. Aufl. Heise Zeitschriften Verlag, 2008.

[9] John G. Proakis. *Digital Commnication*. 4. Aufl. Mc. Graw Hill, 2000.

[10] ITU. *History of the ITU*. URL: http://www.itu.int/en/history/Pages/default.aspx (besucht am 01. 07. 2011).

[11] Andrew S. Tanenbaum. *Computernetzwerke*. 3. Aufl. Prentice Hall, 1998.

[12] Community. *Ethernet*. 2011. URL: http://de.wikipedia.org/wiki/Ethernet.

[13] IEEE. *Std. 802.3: Carrier sense multiple access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*. Section 1. 10 Mb/s Networks. 2008.

[14] IEEE. *Std. 802.3: Carrier sense multiple access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*. Section 2. 100 Mb/s Networks. 2008.

[15] IEEE. *Std. 802.3: Carrier sense multiple access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*. Section 3. 1000 Mb/s Networks. 2008.

[16] IEEE. *Std. 802.3: Carrier sense multiple access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*. Section 4. 10 Gb/s Networks. 2008.

[17] IEEE. *Std. 802.3: Carrier sense multiple access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*. Section 5. Ethernet for Backplanes and Subscriber Access. 2008.

[18] IEEE. *Std. 802.3ba: Carrier sense multiple access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*. Section 2. 40…100 Gb/s Networks. 2008.

[19] Community. *PCI-Express*. http://en.wikipedia.org/wiki/PCI_Express#PCI_Express_3.0. 2011.

[20] Industry Interest Group. *Universal Serial Bus 3.0 Specification*. Techn. Ber. USB Implementers Forum, 2011.

[21] ITU. *G.783 Characteristics of synchronous digital hierarchy (SDH) equipment functional blocks*. März 2006.

[22] ITU. *G.803 Architecture of transport networks based on the synchronous digital hierarchy (SDH)*. März 2000.

[23] ITU. *G.983.1 Broadband optical access systems based on Passive Optical Networks (PON)*. Jan. 2005.

[24] *Bespulte Leitung*. 2011. URL: `http://de.wikipedia.org/wiki/Bespulte%5C_Leitung`.

[25] R. E. Collin. *Field Theory of Guided Waves*. 2. Aufl. IEEE Press, 1992.

[26] J. A. Stratton. *Electromagnetic Theory*. McGraw Hill, 1941.

[27] H.-G. Unger. *Optische Nachrichtentechnik Teil I: Optische Wellenleiter*. 2. Aufl. Heidelberg: Hüthig Verlag, 1989.

[28] ITU. *G.651.1 Characteristics of a 50/125 µm multimode graded index optical fibre cable for the optical access network*. Juli 2007.

[29] ITU. *G.652 Characteristics of a single-mode optical fibre and cable*. Nov. 2009.

[30] Sian Chong Jeffrey Lee. *Discrete Multitone Modulation for Short-Range Optical Communications*. Eindhoven: Technische Universiteit Eindhoven, 2009.

[31] Wikipedia, Hrsg. *LWL-Steckverbinder*. 2013. URL: `http://de.wikipedia.org/wiki/LWL-Steckverbinder` (besucht am 31. 01. 2013).

[32] Roland Kiefer und Peter Winterling. *DWDM, SDH & Co.* 2. Aufl. Hüthig Verlag, 2002.

[33] ITU. *G.653 Characteristics of a dispersion-shifted, single-mode optical fibre and cable*. Juli 2010.

[34] ITU. *Definition: B3ZS (also designated HDB2) and HDB3*. Techn. Ber. International Telcommunication Union, 2011. URL: `http://www.itu.int/sancho`.

[35] J. W. Goodman. *Introduction to Fourieroptics*. McGraw Hill, 1968.

[36] W. H. Press u. a. *Numerical Recipes*. 3. Aufl. Camebridge University Press, 2007.

[37] Robert G. Gallager. "Low-Density Parity-Check Codes". Diss. Cambridge, MA, USA: Massachusetts Institute of Technology (MIT), 1963.

[38] Technical Specification Group Radio Access Network. *Physical channels and modulation*. Technical Specification. 3rd Generation Partnership Project, 2019.

[39] David J. C. MacKay. "Good Error-Correcting Codes based on Very Sparse Matrices". In: *IEEE Transactions on Information Theory* 45.2 (1999). download under http://www.inference.phy.cam.ac.uk/mackay/CodesGallager.html, S. 399–431.

[40] Tom Richardson und Ruediger Urbanke. *Modern Coding Theory*. New York, NY, USA: Cambridge University Press, 2008. ISBN: 0521852293, 9780521852296.

[41] H. Meinke und F.W. Grundlach. *Taschenbuch der Hochfrequenztechnik*. 5. Aufl. Berlin: Springer, 1992.

[42] ITU. *Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000*. International Telecommunications Union. 2003. URL: `http://www.itu.int`.

[43] ITU. *IMT Vision - Framework and overall objectives of the future development of IMT for 2020 and beyond*. International Telecommunications Union. 2015. URL: `http://www.itu.int`.

[44] *Linear Feedback Shift Registers – Implementation, M-Sequence Properties, Feedback Tables*. 2012. URL: http://www.newwaveinstruments.com/resources/articles/m%5C_sequence%5C_linear%5C_feedback%5C_shift%5C_register%5C_lfsr.htm.

[45] Technical Specification Group Radio Access Network. *TS 25.892 V6.0.0: Feasibility Study for Orthogonal Frequency Division Multiplexing (OFDM) for UTRAN enhancement*. Technical Specification. 3rd Generation Partnership Project, 2004.

[46] IEEE Computer Society. *Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*. IEEE Standard for Information Technology 802.11. IEEE, 2007.

[47] 3GPP. *GSM/EDGE Channel coding*. Techn. Ber. 3GPP, 2019.

[48] 3GPP. *Multiplexing and channel coding (FDD)*. Techn. Ber. 3GPP, 2019.

[49] 3GPP. *Evolved Universal Terrestrial Radio Access (E-UTRA); Multiplexing and channel coding*. Techn. Ber. 3GPP, 2019.

[50] Technical Specification Group Radio Access Network. *NR; Multiplexing and channel coding*. Technical Specification. 3rd Generation Partnership Project, 2019.

[51] Lee de Forst. "Space Telegraphy". US 879.532. 1908.

[52] Wikipedia, Hrsg. *Audion*. 2013. URL: http://de.wikipedia.org/wiki/Audion (besucht am 31. 01. 2013).

[53] ITU. *G.694.2 Spectral grids for WDM applications: CWDM wavelength grid*. International telecommunications Union (ITU)-T. März 2003.

[54] ITU. *G.694.1 Spectral grids for WDM applications: DWDM frequency grid*. International telecommunications Union (ITU)-T. Feb. 2012.

[55] Infinera. *Infinera Line System (ILS)*. URL: http://www.infinera.com/products/ILS.html (besucht am 22. 01. 2013).

[56] S. M Sze. *Physics of Semiconductor Devices*. John Wiley and Sons, 1981.

[57] Karl Joachim Ebeling. *Integrierte Optoelektronik*. 2. Aufl. Springer Verlag, 1992.

[58] Jürgen Jahns. *Photonik*. Oldenbourg Verlag, 2001.

[59] SFP-Committee, Hrsg. *SFP (Small Formfactor Pluggable) Transceiver*. 2001. URL: ftp://ftp.seagate.com/sff/INF-8074.PDF (besucht am 26. 01. 2013).

[60] Wikipedia, Hrsg. *Small form-factor pluggable transceiver*. 2013. URL: http://en.wikipedia.org/wiki/Small_form-factor_pluggable_transceiver (besucht am 31. 01. 2013).

[61] Wikipedia, Hrsg. *XENPAK*. 2013. URL: http://en.wikipedia.org/wiki/XENPAK (besucht am 31. 01. 2013).

Glossar

3GPP Third Generation Partnership Project. 9, 11, 147, 148, 210

5G NR 5.th Generation Mobile Communication New Radio. 148, 180, 182

ADC Analog Digital Converter. 102

ADSL Asynchronous Digital Subscriber Line. 2

AKF Autokorrelationsfunktion. 170, 176

ALOHA beschreibt ein Zugriffsverfahren der unteren Schichten für Computernetze. Der Begriff geht auf das in Hawaii entwickelte ALOHA-Net in den 70er-Jahren zurück, später bildete ALOHA das Zugriffsverfahren des Ethernets.. 213

AMI Alternate Mark Inversion. 63

ANSI American National Standards Institute. 11

APD Avalanche Photo Diode. 243, 244

ASE Amplified Spontaneous Emission. 246, 247

ASK Amplitude Shift Keying. 78, 80

AWGN Additive White Gaussian Noise. 82

BP Backplane. 14

BPSK Binary Phase Shift Keying. 62, 80, 86, 195, 196

Bundesnetzagentur Ausgeschrieben: Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen, erreichbar unter http://www.bundesnetzagentur.de regelt in Deutschland die Nutzung von Frequenzen für verschiedene

Dienste und überwacht und reguliert den fairen Zugang zu (Kommunikations-)Netzen. 9

CDMA Code Division Multiple Access. 3, 178, 213

CEPT European Conference of Postal and Telecommunications Administration. 9

CRC Cyclic Redundancy Check. 207

CSMA Carrier Sense Multiple Access. 213

CWDM Coarse Wavelength Division Multiplexing. 231, 232

DAB Digital Audio Broadcasting. 180, 183

DBR Distributed Bragg Reflector. 239

DeMUX De-Multiplexer. v, 230–232, 248–250

DFB-Laser Distributed Feedback Laser. 239

DFT Discrete Fourier Transform. 183

DIN Deutsches Institut für Normung. 9

DSL Digital Subscriber Line. 181

DVB Digital Video Broadcasting. 181

DWDM Dense Wavelength Division Multiplexing. 231, 232, 240, 249

ECO European Communications Office. 9

EDFA Erbium Doped Fiber Amplifier. 246–248

EDGE Enhanced Data Rates for GSM Evolution. 147

EGPRS Enhanced General Packet Radio System. 147

EMV Elektromagnetische Verträglichkeit. 10

EN European Norm. 9

ESB-AM Einseitenband-Amplitudenmodulation. 140

ETSI European Telecommunications Standards Institute. 9–11, 147

FCC Federal Communications Commission. 10, 11

FDD Frequency Division Duplex. 217

FDMA Frequency Division Multiple Access. 6, 181, 213, 215–218

FEC Forward Error Correction. 18

FFT Fast Fourier Transform. 183

FHSS Frequency Hopping Spread Spectrum. 165, 166

FIR Finite Impulse Response. 205

FM Frequenzmodulation. 143

FSK Frequency Shift Keying. 219

GERAN GSM/EDGE Radio Access Network. 9

GPRS General Packet Radio System. 147

GSM Global System for Mobile Telecommunication. 9, 146, 147, 178, 200

HDB3 High Densitiy Bipolar (3 consecutive zeros). 63–65

HSCSD High Speed Circuit Switched Data. 147

HSDPA High Speed Downlink Packet Access. 147

HSPA High Speed Packet Access. 147, 148

HSUPA High Speed Uplink Packet Access. 147

IDE Integrated Development Environment. 254, 255

IEEE Institute of Electrical and Electronics Engineers. 10, 11

IIR Infinite Impulse Response. 209

IMT International Mobile Telecommunications. 147

ISI Intersymbol Interference. 181, 183, 184

ISM Industry, Scientific, Medicine. 175

ISO International Standardization Organization. 11

ITU International Telecommunication Union. 7, 8

KKF Kreuzkorrelationsfunktion. 174

LAN Local Area Network. 15

LDPC Low Density Parity Check Codes. 109, 115, 200

LED Light Emitting Diode [Leuchtdiode]. v, 52, 230, 233–235, 241

LLC Logical Link Control. 11

LLR Logarithmic Likelihood Ratio. 113, 115

LR Likelihood Ratio. 114

LTE Long Term Evolution. 2, 9, 148, 180, 182, 184, 200

LWL Lichtwellenleiter. 14, 36, 38, 42, 44, 52, 57, 217

MAC Medium Access Control. 11, 12, 16

MII Medium Independent Interface. 16, 17

MIMO Multiple Inputs Multiple Outputs. 148, 181

MMF Multi Mode Fiber [Mehrmodenfaser]. 45–48, 52, 53

MUX Multiplexer. v, 230–232, 248, 249

NRZ Non Return (to) Zero. 61, 63, 64

NRZI Non Return to Zero, Inversion on one. 63, 66

OFA Optical Fiber Amplifier. 246, 247

OFDM Orthogonal Frequency Division Multiplexing. 3, 180–184

OOK On Off Keying. 61

OSB Oberes Seitenband. 137

OSI Open System Interconnection. 11

Oversampling Überabtastung eines Signals um einen bestimmten Faktor, die Oversampling-Rate. 60

P2MP Point to Multipoint. 14

P2P Point to Point. 14

PCCC Parallel Concatenated Convolutional Code. 209

PCS Physical Coding Sublayer. 18

PLL Phase Locked Loop (Phasenregelkreis). iv, 189, 193, 227–229

PM Phasenmodulation. 142

PMA Physical Medium Attachment. 18

PMD Physical Medium Dependent. 18

PN Pseudo Noise. 257

POF Polymer Optical Fiber. 48

PON Passive Optical Networks. 15

PRBS Pseudo Random Bit Sequence. 169, 257, 259

PSK Phase Shift Keying. 86

QAM Quadrature Amplitude Modulation. 80

QPSK Quadrature Phase Shift Keying. 80, 86, 178

RS Reconciliation Sublayer. 16

Rx Receiver oder Receive. 58

RZ Return (to) Zero. 64

SDH Synchronous Digital Hierarchy. 2

SDMA Space Division Multiple Access. 6, 213

SFP Small Formfactor Plugable. v, 245

SMF Single Mode Fiber [Einmodenfaser]. 15, 45, 46, 48, 52, 53, 235

SNR Signal Noise Ratio. 85

SONET Synchronous Optical Network. 2

STP Shielded Twisted Pair (Cable). 14

TDD Time Division Duplex. 213, 217

TDMA Time Division Multiple Access. 6, 213, 216–218

TP Twisted Pair (Cable). 14

Tx Transceiver oder Transmit. 58

UMTS Universal Mobile Communication System. 2, 9, 147, 148, 167, 168, 175, 176, 178, 179, 200, 206, 210, 223

URLLC Ultra Reliable Low Latency Communication. 211

USB Unteres Seitenband. 137

UTP Unshielded Twisted Pair (Cable). 14

VCO Voltage Controlled Oscillator (Spannungsgesteuerter Oszillator). 191–193, 198, 227

VCSEL Vertical Cavity Surface Emitting Laser. 237, 239

VDE Verband der Elektrotechnik Elektronik Informationstechnik e.V.. 9, 10

WAN Wide Area Network. 2, 15

WCDMA Wideband Code Division Multiple Access. 167

WLAN Wireless Local Area Network. 3, 167

ZSB-AM Zweiseitenband-Amplitudenmodulation. 137

ZSB-AM-uT Zweiseitenband-Amplitudenmodulation mit unterdrücktem Träger. 138

Index

A

Ähnlichkeitstheorem, 68
Autokorrelation, 74
Autokorrelationsfunktion, 170

B

Bandbreite-Länge-Produkt, 54
Basisband, 58
Bel, 20
Bitübertragungsschich, 11
Booster, 230
Bragg-Reflektoren, 237
Brechungsgesetz, 38
Brechungsindex
    effektiver, 43
Bundesnetzagentur, 10, 11
Büschelfehler, 210

C

Carrier-Ethernet, 12
Cat 5 Kabel, 14
Cat7-Kabel, 15
Checksumme, 59, 117
Chip, 167
CLGMII, 17
Code
    dualer, 99
    linearer, 96, 99
    orthogonaler, 176
    systematischer, 97
Coderate, 96, 102
Costas-Loop, 195

D

Dekodierung, 102
    Hard-Decision, 102
    Soft-Decision, 104
Demodulation
    kohärente, 187
De-Moivre'sche Formel, 79
Dezi-Bel, 20
Dirac-Stoß, 73
Dispersion, 27, 31, 52
Distanz
    freie, 206

E

Entspreizung, 167
Entzerrer, 60
Ethernet, 12, 16, 120

F

Faltungscode, 200
Faltungstheorem, 73
Faserparameter, 45
Fehlerkorrektur, 12
Filmwellenleiter, 39, 42

G
Gallager-Codes, 109
Generatormatrix, 96
Geradeausempfänger, 225
Glasfaser, 39, 45
Glasfaserspleißen, 49
Gold-Code, 171
Gold-Codes, 175
Goos-Haenchen-Effekt, 41, 43
Gradientenfaser, 52
Gruppengeschwindigkeit, 34

H
Hadamard-Code, 107
Hamming Code, 98
Hamming-Codes, 106
Hammingdistanz, 95, 102, 206
Homodyner Empfänger, 227

I
IEEE 802.11, 167
IEEE 802.15.4, 167
Impulsformung, 91
Interleaver, 210
ITU-D, 7
ITU-R, 7
ITU-T, 7

K
Kanal
    frequenzselektiv, 153
    Rayleigh, 154
    Rice, 159
Kanalkodierer, 59
Kanalzugriff, 12
Kernbrechungsindex, 43
Kodiergewinn, 102
Kodierung
    4B/5B, 64
    Manchester, 63, 64
Kommunikationskanal, 3
Konforme Abbildungen
    Methode der, 30
Konstellationsdiagramm, 80, 81
Kreuzkorrelation, 171

L
Laser, 60, 235
Laserdiode, 230, 235
Laserschwelle, 240
LDPC
    irreguläre, 111
    reguläre, 111
Leitungskodierer, 59
Linearitatstheorem, 67
LWL-Verbinder
    DIN/ LSA, 51
    E2000, 51
    FCPC, 51
    LC, 51
    MTRJ, 51
    SC, 51

M
Materialdispersion, 53
Midamble, 219
Moden, 42
Modendispersion, 52, 53
Modulationstheorem, 70

N
Netzelementes, 16
Numerische Apertur, 41, 44

P
Paritätsbits, 96
Parsevals Theorem, 72
Phasengeschwindigkeit, 32
Phasenkoeffizient, 43
Photodiode, 230
Physical Layer, 11, 12
P-I-N-Photodiode, 242
PN-Sequenz, 168, 170, 176
PN-Sequenzen, 169
Polarcodes, 200
Polarisation, 42
Pre-Amplifier, 230
Prüfmatrix, 99
Pulsformung, 59
Pupinspule, 28

Q
Quadratischer Detektor, 225
Quadraturempfänger, 229

R
Redundanzbits, 96
Referenzmodelle, 11
Resonator, 235

S
Scrambler, 59
Scrambling-Sequenz, 222
Sicherungsschicht, 11
spektrale Dichte, 77
Spiegelfrequenz, 226
Spreizfaktor, 168
Streckenverstärker, 231
Stufenfaser, 45
Superhet, 226, 227
Synchronempfänger, 229
Synchronisation, 60
Syndrom, 100

T
Ternares Signal, 63
thermisches Rauschen, 82
Totalreflexion, 38, 39, 41, 44
Trägerrückgewinnung, 189
Turbocode, 115

U
Uberabtastung, 60
Ubertragungskanal, 58
Übertragungsrahmen, 120
Übertragungssystem, 5

V
Verkürzungsfaktor, 29
Verschiebungstheorem, 69
Verstarker, 60
Videoband, 58
Viterbi Algorithmus, 204

W
Wahrscheinlichkeitsverhältnis, 114
Walsh-Code, 222
Walsh-Codes, 176
Wandler, 60
Wellenimpedanz, 28
Wellenlängenmultiplex, 231
Wellenleiter, 39

X
XLGMII, 17

Z
ZigBee, 167
Zwischenfrequenz, 226