ulm university universität
**u**ulm

# Vorlesungsskript

# Vorlesungsskript Physikalische Elektronik und Messtechnik

Othmar Marti
Institut für Experimentelle Physik
Universität Ulm

11. Oktober 2012

# Inhaltsverzeichnis


**1 Einleitung** **9**

**2 Mathematische Grundlagen** **11**
2.1 Darstellung von elektronischen Messsystemen: Blockschematas . . . 11
2.1.1 Symbole für Blockschematas . . . 11
2.1.2 Rechnen mit Blockschematas . . . 14
2.2 Darstellung von elektronischen Messsystemen: Signalflussdiagramme 16
2.2.1 Grundlagen . . . 17
2.2.2 Begriffe aus der Theorie der Signalflussdiagramme . . . 19
2.2.3 Allgemeine Formel für Signalflussdiagramme . . . 20
2.3 Übertragungsfunktionen . . . 21
2.4 Kontinuierliche und diskrete Signale . . . 24
2.4.1 Signale . . . 24
2.4.2 Fourier-Transformationen . . . 25
2.4.3 Laplace-Transformationen . . . 30
2.4.4 z-Transformationen . . . 32
2.4.5 Anwendung der Transformationen auf Einschaltvorgänge . . . 37
2.4.6 Digitale Signale . . . 41
2.5 Vierpole und Vierpoltheorie . . . 52
2.5.1 Zusammenschaltung von Vierpolen . . . 53
2.5.2 Übertragungsfunktion eines Vierpols . . . 56
2.5.3 Ersatzstrukturen für Vierpole . . . 58
2.6 Filter . . . 60
2.6.1 Analogfilter . . . 60
2.6.2 Digitalfilter . . . 67
2.7 Modulationstheorie . . . 86
2.8 Rauschen . . . 89
2.8.1 Widerstandsrauschen . . . 89
2.8.2 Weitere Rauschquellen . . . 95
2.8.3 Einfluss von Filtern auf das Rauschen . . . 95
2.9 Digitale Signalprozessoren (DSP) . . . 97
2.9.1 Klassische Rechner . . . 97
2.9.2 Digitale Signalprozessoren . . . 99

**3 Bauelemente und Schaltungstechnik** **105**
3.1 Halbleiter–Grundlagen . . . 105
3.1.1 Grundlagen . . . 105
3.1.2 Intrinsischer Halbleiter . . . 110
3.1.3 Dotierung von Halbleitern . . . 114

3.1.4 Ladungsträgerdichten im dotierten Halbleiter . . . . . 116
3.1.5 Leitfähigkeit in Abhängigkeit von Dotierkonzentration und Temperatur . . . . . 117
3.1.6 Rekombinationsprozesse und Ladungsträgertransport: Grundgleichungen zur Funktion von Halbleiter–Bauelementen . . . 121
3.2 Phänomene elektrischer Kontakte . . . . . 124
3.2.1 Grundlagen . . . . . 124
3.2.2 p–n–Übergänge . . . . . 125
3.2.3 Vorgespannte p–n–Übergänge — gleichrichtende Dioden . . 129
3.2.4 Hetero–Übergänge . . . . . 132
3.2.5 Metall–Halbleiter–Kontakte (Ohmsche Kontakte, Schottky–Dioden) . . . . . 136
3.2.6 Metall–Isolator–Halbleiterkontakte (MIS– und MOS–Dioden) 142
3.3 Wichtige Halbleiter–Bauelemente (Aufbau, Funktion, Technologie) . 148
3.3.1 Ladungsgekoppelte Bauelemente (CCD charge coupled devices) . . . . . 148
3.3.2 Feldeffekt–Transistoren (Unipolare Transistoren) . . . . . 154
3.3.3 CMOS–Technologie und Halbleiter–Speicher . . . . . 161
3.3.4 Bipolare Transistoren (BJT Biplor Junction Transistor, Injektionstransistoren) . . . . . 169
3.3.5 Einige Optoelektronische Bauelemente . . . . . 175
3.3.6 Ausblick . . . . . 180
3.4 Grundschaltungen . . . . . 181
3.4.1 Lineare passive Bauelemente . . . . . 181
3.4.2 Dioden . . . . . 184
3.4.3 Bipolartransistoren . . . . . 188
3.4.4 Feldeffekttransistoren . . . . . 200
3.4.5 Einige Grundschaltungen mit Transistoren . . . . . 205
3.5 Operationsverstärker . . . . . 208
3.5.1 Grundlagen, Grundtypen, Rückkopplung . . . . . 208
3.5.2 Standard–Operationsverstärker (VV–OPV) . . . . . 211
3.5.3 Transkonduktanz–Verstärker (VC–OPV) . . . . . 216

**4 Sensoren und Messverfahren 217**
4.1 Basismessverfahren . . . . . 217
4.1.1 Strom . . . . . 217
4.1.2 Spannung . . . . . 222
4.1.3 Wechselstrom und Wechselspannung . . . . . 225
4.1.4 Ladung . . . . . 238
4.1.5 Widerstand . . . . . 241
4.1.6 Messung von L und C . . . . . 249
4.1.7 Brückenschaltungen . . . . . 251
4.1.8 Wandlerschaltungen . . . . . 259
4.1.9 Lock-In Verstärker am Beispiel des AD630 Chips . . . . . 281
4.2 Messung weiterer physikalischer Grössen . . . . . 285
4.2.1 Frequenzmessung . . . . . 285
4.2.2 Magnetfelder . . . . . 295
4.2.3 Dielektrische Funktion . . . . . 305

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

4.2.4 Temperaturmessungen . . . 306
4.2.5 Licht . . . 316
4.3 Leitungen . . . 322
4.3.1 Leitungsgleichungen . . . 322
4.3.2 Elektrische Leitungen bei hohen Frequenzen . . . 335
4.3.3 Optische Leitungen . . . 340
4.4 Messungen kleiner Pegel . . . 359
4.4.1 Testfelder . . . 359
4.4.2 Spannungen . . . 364
4.4.3 Ströme . . . 366
4.4.4 Techniken zur Verhinderung von Fehlmessungen . . . 368
4.5 Lichtquellen für optische Messverfahren . . . 376
4.5.1 Grundlagen der Lasertechnik . . . 376
4.5.2 Kurzzeitlaser . . . 390
4.6 Optische Messverfahren . . . 406
4.6.1 Absorptionsmessung . . . 406
4.6.2 Reflexionsmessung . . . 408
4.6.3 Polarisationsmessung . . . 409
4.6.4 Spektrometer und Polychromatoren . . . 409
4.6.5 Messverfahren für kurze Zeiten . . . 416
4.7 Elektrooptische Messverfahren für kurze Zeiten . . . 420
4.8 Elektrische Messverfahren für kurze Zeiten . . . 421
4.9 Elektrische Spektralanalyse und Netzwerkanalyse . . . 423
4.10 Messung mit Elektronen . . . 433
4.10.1 Tunneleffekt, Bänderstruktur . . . 433
4.10.2 Tunneltheorie von Simmons . . . 435
4.10.3 The Transfer Hamiltonian Method . . . 439
4.10.4 Rastertunnelmikroskopie (STM) . . . 441
4.10.5 Resolution of a Scanning Tunneling Microscope . . . 457
4.10.6 Tunnelspektroskopie . . . 462
4.10.7 *Graphite* . . . 462
4.10.8 Low Temperature Experiments . . . 464
4.10.9 Related Techniques . . . 464
4.10.10 Serieschaltung von Tunneldioden . . . 469
4.10.11 Single Elektron Transistor . . . 473
4.10.12 Feldemissionsmikroskopie, Feldionenmikroskopie . . . 476
4.10.13 Projektionselektronenmikroskopie . . . 481
4.10.14 Rasterelektronenmikroskopie . . . 483
4.10.15 Transmissionselektronenmikroskopie(TEM) . . . 517
4.10.16 Elektronenbeugung . . . 520
4.10.17 Scanning Force Microscopy . . . 534
4.11 Rechnergestützte Messtechnik . . . 558
4.11.1 Verwendung externer Geräte . . . 558
4.11.2 In Rechner eingebaute Messdatenerfassung . . . 564
4.11.3 Übersicht über Programme zur Datenerfassung . . . 565

**A Physikalische Grundlagen** **567**
A.1 Maxwellsche Gesetze . . . 567

A.2 Kirchhoffsche Gesetze . . . 567
A.3 Komplexe Spannungen und Ströme . . . 568
A.4 Ebene Wellen . . . 569

**B Berechnung von Schaltungen 571**
B.1 Brückenschaltung mit Widerständen . . . 571

**C Tabellen 573**
C.1 Tabelle der Laplacetransformationen . . . 573
C.2 Tabelle der Carson-Heaviside-Transformation . . . 574
C.3 Tabelle der z-Transformationen . . . 575
C.4 Einstellzeiten und Zeitkonstanten . . . 576

**D Vergleich der Kenngrössen von Bauarten analoger Filter 577**

**E Diagramme der Filterübertragungsfunktionen 579**
E.1 Tiefpassfilter . . . 579
E.2 Hochpassfilter . . . 585
E.3 Bandpassfilter . . . 591
E.4 Bandsperrenfilter . . . 597
E.5 Allpassfilter . . . 603
E.6 Schwingkreis . . . 605

**F Filterkoeffizienten 607**

**G Maple V Texte 615**
G.1 Ortskurve in der komplexen Ebene . . . 615
G.2 Definitionen der Filterfunktionen . . . 615
G.2.1 Kritische Filter . . . 615
G.2.2 Butterworth . . . 615
G.2.3 Bessel . . . 615
G.2.4 Tschebyscheff 1dB . . . 616
G.2.5 Tschebyscheff 3dB . . . 616
G.2.6 Allpass . . . 616
G.2.7 Schwingkreis . . . 616
G.3 Darstellung der Filter . . . 616
G.3.1 Tiefpass-Hochpasstransformation . . . 616
G.3.2 Tiefpass-Bandpasstransformation . . . 616
G.3.3 Tiefpass-Bandsperrentransformation . . . 616
G.3.4 Beispiel:Butterworth Tiefpässe . . . 617
G.4 Smith-Chart . . . 618
G.5 Smith-Charts mit Gnuplot 4.2 . . . 624
G.6 Daten auf Smith-Charts zeichnen mit Gnuplot 4.2 . . . 626
G.7 Polar-Plots mit Gnuplot 4.2 . . . 627

**H Leistungen eines DSPs 633**

**I Materialeigenschaften 635**
I.1 Eigenschaften von Isolationsmaterialien . . . 635
I.2 Thermolelektrische Koeffizienten . . . 635

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

I.3 Seebeck-Koeffizienten . . . . . 635
I.4 Debye-Temperatur und Temperaturkoeffizient des Widerstandes . . 635

**J Image Processing: an Introduction 637**
J.1 Why Image Processing? . . . . . 637
J.2 Correcting Distorted Images . . . . . 638
J.3 Filtering and Data Analysis in Real Space . . . . . 639
J.4 Filtering and Data Analysis in the Spatial Frequency Domain . . . 641
J.5 Viewing the Data . . . . . 644
J.6 Background Plane Removal . . . . . 647

**K Correction of Linear Distortions in Two and Three Dimensions 649**

**L Beschreibung periodischer Oberflächen 651**
L.1 Mathematische Beschreibung . . . . . 651
L.1.1 Bravais-Netze . . . . . 652
L.1.2 Überstrukturen, Rekonstruktionen . . . . . 653

**M Symbole 657**

**N Hilfsprogramme 659**

**Abbildungsverzeichnis 660**

**Tabellenverzeichnis 678**

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

# 1. Einleitung

Die Vorlesung **Physikalische Elektronik und Messtechnik** behandelt die theoretischen und praktischen Grundlagen des Messens mit elektronischen Hilfsmitteln aller Art. Zwar wird das Schwergewicht auf Elektronik gelegt. Die modernen Entwicklungen gerade in der optischen Messtechnik sollen bei der Behandlung nicht ausgespart bleiben.

Die Vorlesung beginnt mit einer Darstellung der mathematischen Grundlagen. In Kapitel 2. Zuerst wird in die Sprache der Blockschaltbilder eingeführt.

Im nächsten Kapitel 3 werden Halbleiterschaltungen behandelt. Ausgehend von der Physik der Halbleitermaterialien werden die einfachsten Bauelemente, nämlich Transistoren und Dioden, besprochen. Es folgt eine Darstellung der Grundschaltungen dieser Bauelemente. Kombinationen der Grundschaltungen sind die Differenzverstärker und letztlich auch die Operationsverstärkerschaltungen.

Das letzte Kapitel 4 ist der Diskussion von elektronischen Messverfahren sowie der Messung von Eigenschaften mit Elektronen gewidmet. Nach einer Beschreibung der grundlegenden Messverfahren werden unter anderem auch die Rastertunnelmikroskopie, die Elektronenmikroskopie und verschiedene, auf Elektronen basierende Verfahren der Oberflächenphysik besprochen.

# 2. Mathematische Grundlagen

## 2.1. Darstellung von elektronischen Messsystemen: Blockschematas

Elektronische Schaltungen wie auch ganze elektronische *Messgeräte* können als Systeme betrachtet werden. Die einzelnen Baublöcke sind entweder grundlegende Systeme, oder sie können als Zusammenfassung von verschiedenen einfacheren Systemen betrachtet werden. Je nach Tiefe der Betrachtung ist zum Beispiel ein Lock-In Verstärker ein grundlegendes System mit einer, durchaus nicht trivialen Beziehung zwischen Ausgangs- und Eingangssignalen. Alternativ kann er aber auch als Zusammensetzung der folgenden Baugruppen aufgefasst werden:

1. Eingangsverstärker
2. Referenzoszillator
3. Phaserschieber
4. Mischer
5. Tiefpassfilter

Diese Liste könnte, wenn man wollte, noch weiter unterteilt werden.

### 2.1.1. Symbole für Blockschematas

Es ist üblich geworden, die folgenden Symbole für die Darstellung von Systemen zu verwenden[**?**].

Abbildung 2.1.: Ein grundlegender Systemblock

Die Abbildung 2.1 zeigt ein einen solchen grundlegenden Systemblock. Das Wort Block in dieser Darstellung wird, je nach Typ oder Übertragungsfunktion ausgewechselt. Die Anschlüsse werden mit Pfeilen versehen, um den Signalfluss darzustellen. So würde man zum Beispiel eine Differentiation wie in Abbildung 2.2 darstellen.
Eine Integration könnte wie in Abbildung 2.3 aussehen.
In allen Darstellungen ist es optional, die Bezeichnungen Eingang und Ausgang zu verwenden. Sie sollten benutzt werden, wenn der Signalfluss aus der Darstellung nicht eindeutig abgelesen werden kann. Die einzelnen Blöcke werden mit Linien

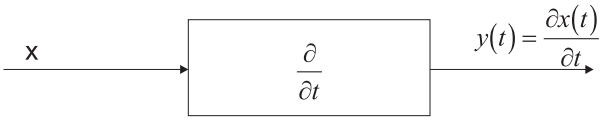



Abbildung 2.2.: Ein Differentialoperator in der Blockschaltbildschreibweise

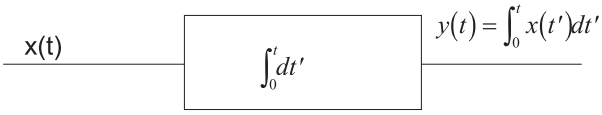



Abbildung 2.3.: Ein Integraloperator in der Blockschaltbildschreibweise

verbunden. Sollten ein Ausgang eines Blocks auf mehrere Eingänge aufgeteilt werden, so werden Abnahmepunkte wie in Abbildung 2.4 verwendet.
Abnahmepunke dienen nicht nur dazu, Signale nach vorne zu leiten. Wie Abbildung 2.5 zeigt, können auch Rückkoppelungen mit dieser Formensprache gehandhabt werden.
Um allgemeine Signalflüsse darstellen zu können, sind noch Summationspunkte notwendig. Sie werden wie in Abbildung 2.6 dargestellt. Es können zwei oder mehr Summationseingänge verwendet werden.
Wir können nun diese Formensprache verwenden, um die Differentialgleichung

$$\frac{\partial y\left(t\right)}{\partial t} + ky\left(t\right) = x\left(t\right) \tag{2.1.1}$$

darzustellen. Wir schreiben die Gleichung (2.1.1) um, so dass wir sie in die Blockschaltbildform bringen können. Zuerst isolieren wir die Ableitung.

$$x\left(t\right) - ky\left(t\right) = \frac{\partial y\left(t\right)}{\partial t} \tag{2.1.2}$$

Schliesslich integrieren wir die Gleichung (2.1.2) und erhalten

$$\int_{t_0}^{t} x\left(t'\right) - ky\left(t'\right)dt' = y\left(t\right) \tag{2.1.3}$$

Die Gleichung (2.1.1) in der Form (2.1.3) kann nun wie in Abbildung 2.7 dargestellt werden.

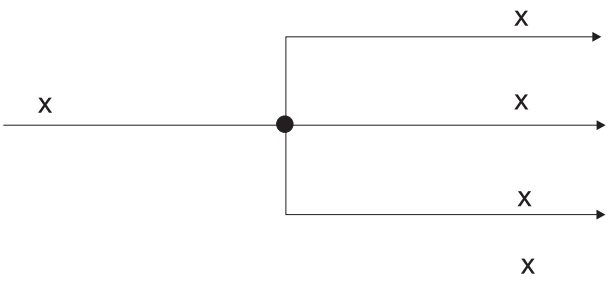



Abbildung 2.4.: Ein Abnahmepunkt. Das Eingangssignal wird nach rechts auf drei Zweige aufgeteilt.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

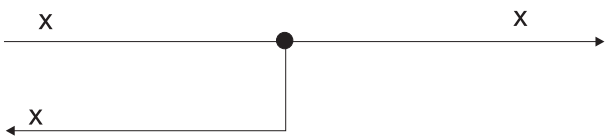



Abbildung 2.5.: Ein Abnahmepunkt. Hier wird ein Teil des Signals rückgekoppelt

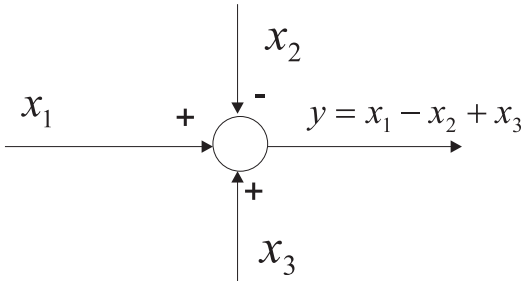



Abbildung 2.6.: Ein Summationspunkt. Die Operatoren + und - geben an, ob addiert oder subtrahiert werden soll

Die Umstellung in der Gleichung musste durchgeführt werden, um $y(t)$ zu isolieren. Eine alternative Art der Umformung ist

$$x\left(t'\right) - \frac{\partial y\left(t\right)}{\partial t} = y\left(t\right) \tag{2.1.4}$$

Das entsprechende Blockschaltbild ist in der Abbildung 2.8 zu sehen.
Wenn man die Abbildungen 2.7 und 2.8 vergleicht, sieht man, dass die gleiche Differentialgleichung auf zwei verschiedene Arten dargestellt werden kann. Es gibt offensichtlich Regeln, die einem ermöglichen, die Umstellung auf rein formalem Wege zustande zu bringen. Der Vergleich der beiden Abbildungen sagt zum Beispiel, dass wenn $k$ in einem nach links gerichteten Zweig vorkommt, wir $\frac{1}{k}$ in einen nach rechts gerichteten Zweig einsetzen müssen. Ebenso sind die Integration und die Differentiation ein Paar, wenn wir in einem Signalzweig die Signalflussrichtung wechseln. Im nächsten Abschnitt 2.1.2 werden die Rechenregeln für Blockschematas dargestellt.
Zum Schluss dieses Abschnittes sei darauf hingewiesen, dass für Operationsverstärker die genau gleichen Regeln gelten: Ein Bauelement, das differenziert eingebaut in die Rückkoppelschleife, bewirkt, dass die Gesamtschaltung integriert. Mit diesem Konzept, das im Kapitel 3 besprochen wird, können unter anderem grosse

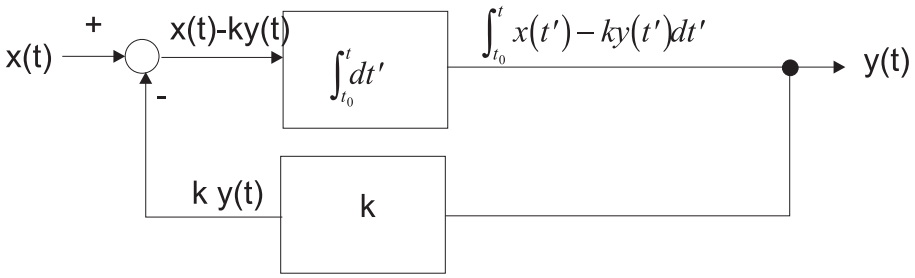



Abbildung 2.7.: Das Blockschaltbild der Differentialgleichung (2.1.1) in der Form (2.1.3).

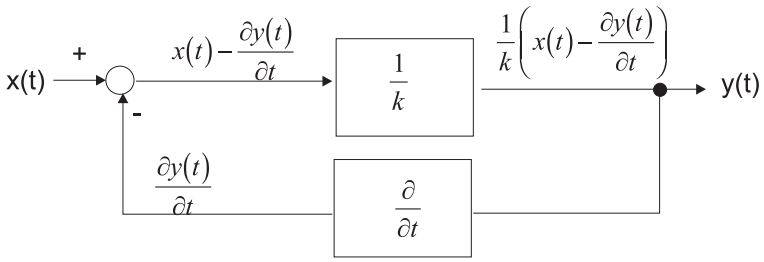



Abbildung 2.8.: Das Blockschaltbild der Differentialgleichung (2.1.1) in der Form (2.1.4).

Impedanzen oder Zirkulatoren realisiert werden.

## 2.1.2. Rechnen mit Blockschematas

Das Rechnen mit Blockschematas erlaubt, auf eine standardisierte Weise die Umorganisation und die Berechnung von Schaltungen. Diese Rechnungen werden benötigt, um Schaltungen zu vereinfachen oder um sie, bei gleicher Funktion, anders zu strukturieren. Dies kann nötig sein, weil Toleranzen und Fehler der Bauteile nicht bei jeder Konfiguration sich gleich auswirken.

### 2.1.2.1. Kaskadierung (Reihenschaltung) von Blöcken

Wenn zwei Blöcke mit den Transferfunktionen $G_1$ und $G_2$ hintereinander geschaltet sind, dann können diese durch einen Block mit der Transferfunktion $G_1G_2$ ersetzt werden (Abbildung 2.9).

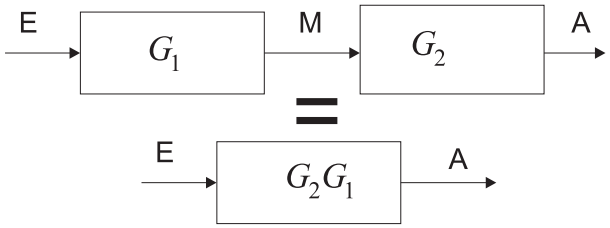



Abbildung 2.9.: Kaskade von zwei Blöcken

### 2.1.2.2. Kommutativgesetz für die Kaskadierung

Abbildung 2.10.: Das Kommutativgesetz für einen Block.

Für lineare Systeme ist die Reihenfolge, in der zwei Blöcke mit kommutativen Operatoren (Multiplikation, aber auch die Ableitung einer komplexen Funktion) angeordnet werden, unerheblich. Dies wird mit dem Kommutativgesetz $G_1G_2 = G_2G_1$ beschrieben (Abbildung 2.10). Während diese Aussage mathematisch gesehen korrekt ist, kommt es bei der Realisierung durchaus auf die Reihenfolge an. Da reale

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Baublöcke **immer** nichtlinear sind (Begrenzung, Rauschen, Wechselwirkung) kann die Platzierung darüber entscheiden, ob ein Design gut oder schlecht ist.

### 2.1.2.3. Transformationen

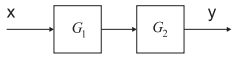
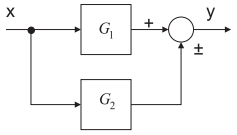
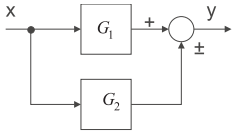
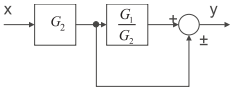
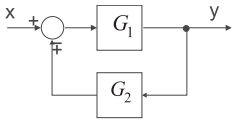
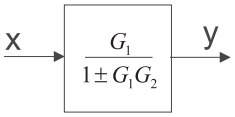
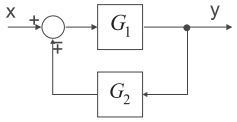
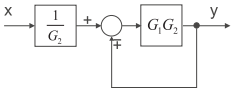
| | Transformation | Gleichung | Ausgangsdiagramm | Äquivalentes Diagramm |
|---|---|---|---|---|
| 1 | Reihenschaltung | $y = (G_2 G_1)x$ |  | |
| 2 | Parallelschaltung | $y = G_1 x \pm G_2 x$ |  | |
| 3 | In Vorwärtsrichtung parallelgeschalteten Block entfernen | $y = G_1 x \pm G_2 x$ |  |  |
| 4 | Block in der Rückkopplungsleitung entfernen | $y = G_1 (x \mp G_2 y)$ |  |  |
| 5 | Block aus Rückkoppelleitung verschieben | $y = G_1 (x \mp G_2 y)$ |  |  |

Tabelle 2.1.: Algebra mit Blockdiagrammen: Kombination von Blöcken

Ein rückgekoppeltes System, eine der am häufigsten vorkommenden Strukturen in der Physikalischen Elektronik und Messtechnik, sieht wie in Abbildung 2.11 aus. Wenn man die Konventionen aus der Abbildung 2.11 verwendet und insbesondere beachtet, dass das obere Vorzeichen des Zweiges $B$ eine Gegenkopplung bedeutet, so erhält man die folgenden universellen Beziehungen:

$$\frac{A}{E} = \frac{G}{1 \pm GH} \tag{2.1.5}$$

$$\frac{F}{E} = \frac{1}{1 \pm GH} \tag{2.1.6}$$

$$\frac{B}{E} = \frac{GH}{1 \pm GH} \tag{2.1.7}$$

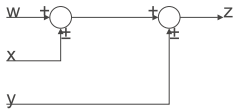
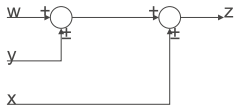
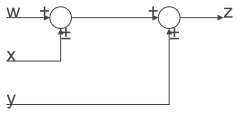
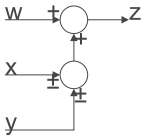
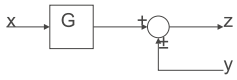
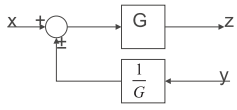
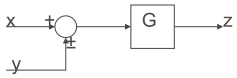
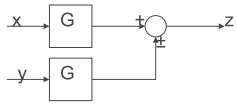
| | Trans-formation | Gleichung | Ausgangs-diagramm | Äquivalentes Diagramm |
|---|---|---|---|---|
| 6a | Summa-tionspunkte verschieben | $z = w \pm x \pm y$ |  |  |
| 6b | Summa-tionspunkte verschieben | $z = w \pm x \pm y$ |  |  |
| 7 | Summa-tionspunkt vor Block schieben | $z = Gx \pm y$ |  |  |
| 8 | Summa-tionspunkt nach Block schieben | $z = G\,(x \pm y)$ |  |  |

Tabelle 2.2.: Algebra mit Blockdiagrammen: Summationspunkte verschieben

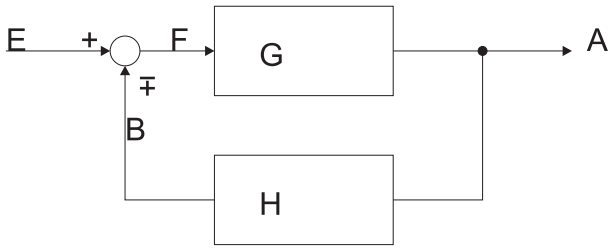



Abbildung 2.11.: Blockschaltbild eines rückgekoppeltes Systems

Hier ist $E$ das Eingangssignal, $A$ das Ausgangssignal und $B$ das Rückkoppelsignal vor dem Summationspunkt und $F$ das Fehlersignal. Eine Analyse der obigen Gleichungen zeigt, dass wenn der Betrag von H gross ist bei einer negativen Rückkoppelung, also bei einem positiven Vorzeichen, dass $\frac{F}{E}$ beliebig klein wird. $B$ wird dann gleich dem negativen Eingangssignal $E$.

## 2.2. Darstellung von elektronischen Messsystemen: Signalflussdiagramme

Eine weitere Möglichkeit, die Struktur einer Schaltung darzustellen bieten die Signalflussdiagramme. Der Hauptvorteil der Signalflussdiagramme liegt darin, dass sie sich einfacher zeichnen lassen.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

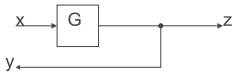
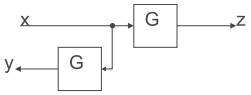
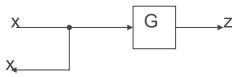
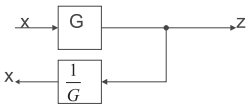
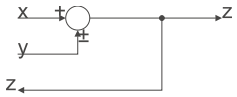
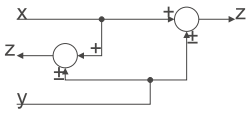
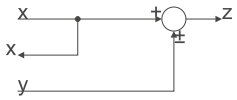
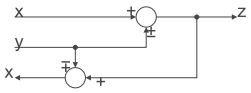
| | Trans-formation | Gleichung | Ausgangs-diagramm | Äquivalentes Diagramm |
|---|---|---|---|---|
| 9 | Abnahme-punkt vor Block schieben | $y = Gx$ |   |   |
| 10 | Abnahme-punkt nach Block schieben | $y = Gx$ |   |   |
| 11 | Abnahme-punkt vor einen Summati-onspunkt schieben | $z = x \pm y$ |   |   |
| 12 | Abnahme-punkt nach einen Summati-onspunkt schieben | $y = x \pm y$ |   |   |

Tabelle 2.3.: Algebra mit Blockdiagrammen:Abnahmepunkte verschieben

### 2.2.1. Grundlagen

Abbildung 2.12 zeigt einen einfachen Signalflussgraphen. Es wird die Gleichung

$$Y_i = A_{ij} X_j \tag{2.2.1}$$

dargestellt. Die Variablen $X_i$ und $Y_j$ sind jeweils mit einem Punkt dargestellt. Diese Punkte heissen **Knoten**. jede Variable hat in einem Signalflussdiagramm einen Knoten. Die Verknüpfung von zwei Variablen erfolgt durch Zweige. Dabei ist immer in Flussrichtung (gegeben durch den Zweig) die Übertragungsfunktion auf die Ausgangsvariable anzuwenden. Hier steht bewusst Übertragungsfunktion: ein Integral oder eine Ableitung sind auch denkbar.
Eine Summation von Werten wird wie in Abbildung 2.13 dargestellt. Soll ein konstanter Wert dazugezählt werden, so wird der Wert als Variable mit der Übertragungsfunktion 1 geführt.

$$Y_j = \sum_{i=1}^{n} A_{ji} X_i \tag{2.2.2}$$

Der Wert einer Variablen wird auf alle von dem entsprechenden Knoten ausgehenden Variablen übertragen, wie in Abbildung 2.14 dargestellt.

$$X_j = A_{jk} X_k \quad \text{für } (k = 1, 2, \ldots, n) \tag{2.2.3}$$

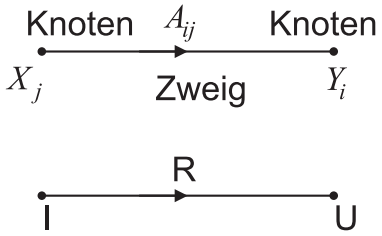



Abbildung 2.12.: Signalflussdiagramm. Oben ist ein allgemeiner Zweig dargestellt, bei dem gilt: $Y_i = A_{ij}X_j$. Als Beispiel ist unten das Ohm'sche Gesetz $U = RI$ gezeigt.

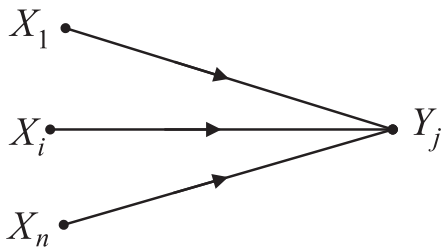



Abbildung 2.13.: Summation in einem Signalflussdiagramm: $Y_j = \sum_{i=1}^{n} A_{ji}X_i$.

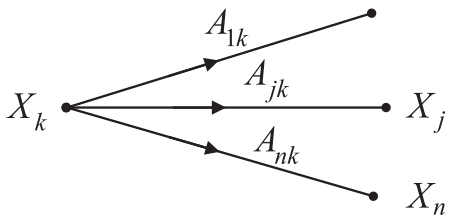



Abbildung 2.14.: Übertragungsregeln in einem Signalflussdiagramm: $X_j = A_{jk}X_k$ für $(k = 1, 2, \ldots, n)$.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

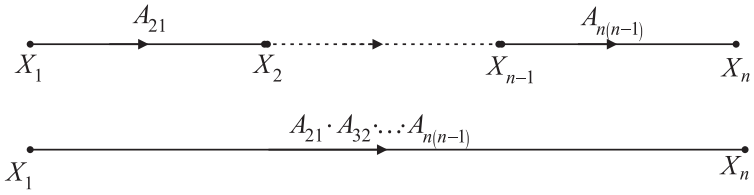



Abbildung 2.15.: Multiplikationsregeln für Signalflussdiagramme: $X_n = \left(\prod_{i=2}^{n} A_{i(i-1)}\right) X_1$.

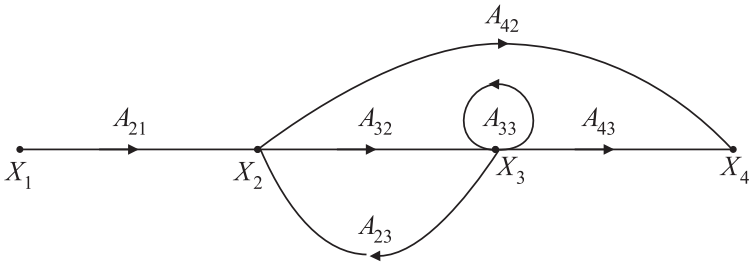



Abbildung 2.16.: Signalflussdiagramm zur Erklärung der Definitionen

Gilt für eine Kette, dass

$$X_j = A_{j(j-1)} X_{j-1} \quad \text{für } (j = 2, \ldots, n) \tag{2.2.4}$$

dann können diese Gleichungen zusammengefasst werden zu

$$X_n = A_{21} \cdot A_{32} \cdot \ldots \cdot A_{n(n-1)} X_1 = \left(\prod_{i=2}^{n} A_{i(i-1)}\right) X_1 \tag{2.2.5}$$

Dies wird im Signalflussdiagramm wie in Abbildung 2.15 dargestellt.

## 2.2.2. Begriffe aus der Theorie der Signalflussdiagramme

Die folgenden Definitionen sind in Abbildung 2.16 abgebildet:

**Pfad** Ein Pfad ist ein zusammenhängender, in eine Richtung zeigende Abfolge von Verbindungen zwischen Knoten. In Abbildung 2.16 ist $(X_1 \to X_2 \to X_3 \to X_4)$, $(X_2 \to X_3 \to X_2)$, $(X_3 \to X_3)$ und $(X_1 \to X_2 \to X_4)$ Pfade.

**Eingangsknoten** Ein Eingangsknoten ist ein Knoten, von dem nur Pfade ausgehen. Beispiel: $X_1$.

**Quelle** Siehe Eingangsknoten (der Begriff muss vom Diagramm her verstanden werden, nicht von der Aussenwelt)

**Ausgangsknoten** Ein Ausgangsknoten ist ein Knoten, bei dem nur einlaufende Pfade auftreten. Beispiel: $X_4$. Gibt es keine solchen Knoten, fügt man einen Zusatzknoten mit einer Übertragungsfunktion $A = 1$ hinzu.

**Senke** Siehe Ausgangsknoten.

**Vorwärtspfad** Ein Vorwärtspfad ist ein Pfad, der vom Eingangsknoten zum Ausgangsknoten führt. Beispiel: $(X_1 \to X_2 \to X-3 \to X_4)$, oder $(X_1 \to X_2 \to X_4)$.

**Rückwärtspfad** Ein Rückwärtspfad ist ein Pfad, ist ein Pfad dessen Anfangs- und Endknoten gleich sind. Beispiel: $(X_2 \to X_3 \to X_2)$ und $(X3 \to X_3)$.

**Rückkopplungsschleife** Siehe Rückwärtspfad.

**Selbstbezogene Schleife** $(X3 \to X_3)$ ist eine selbstbezogene Schleife.

**Verstärkung eines Zweiges** Die Verstärkung eines Zweiges ist der Faktor, mit dem bei diesem Zweig multipliziert werden muss. Beispiel: $A_{33}$ ist die Verstärkung der selbstbezogenen Schleife.

**Verstärkung eines Pfades** Die Verstärkung eines Pfades ist der Operator, der entsteht wenn man alle Teiloperatoren hintereinander anwendet. Im Falle rein multiplikativer Verstärkungen ist dies Das Produkt der einzelnen Pfadverstärkungen. Beispiel. Der Pfad $(X_1 \to X_2 \to X_4)$ hat die Verstärkung $A_{21}A_{32}A_{42}$.

**Schleifenverstärkung** Die Schleifenverstärkung ist die Verstärkung einer Rückkoppelungsschleife. Beispiel: $(X_2 \to X_3 \to X_2)$hat die Verstärkung $A_{32}A_{23}$.

### 2.2.3. Allgemeine Formel für Signalflussdiagramme

bezeichnet man mit $T$ das Verhältnis zwischen Dem *Signal* am Ausgangsknoten und dem am Eingangsknoten, so gilt

$$T = \frac{\sum\limits_i P_i \Delta_i}{\Delta} \tag{2.2.6}$$

Dabei ist

- $P_i$ die Pfadverstärkung des $i$-ten Vorwärtspfades.
- $P_{jk}$ das $j$-te mögliche Produkt von $k$ sich nicht berührenden Rückkoppelschleifen.
- $\Delta = 1 - (-1)^{k+1} \sum\limits_k \sum\limits_j P_{jk} = 1 - \sum\limits_j P_{j1} + \sum\limits_j P_{j2} - \sum\limits_j P_{j3} + \ldots =$ 1 - (Summe aller Schleifenverstärkungen)+ (Summe aller Verstärkungsprodukte von je zwei sich nicht berührenden Rückkoppelschleifen) - (Summe aller Verstärkungsprodukte von je drei sich nicht berührenden Rückkoppelschleifen) + ...
- $\Delta_i = \Delta$ berechnet unter Weglassung aller Rückkoppelschleifen, die den Pfad $P_i$ berühren.

Zwei Pfade heissen **nichtberührend**, wenn sie keine gemeinsamen Knoten haben. $\Delta$ heisst die Determinante des Signalflussgraphen oder seine charakteristische Funktion. Signalflussgraphen treten auch als Feynmansche Pfaddarstellungen auf. Auch in der Quantenelektrodynamik gelten analoge Rechenregeln für Pfade.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

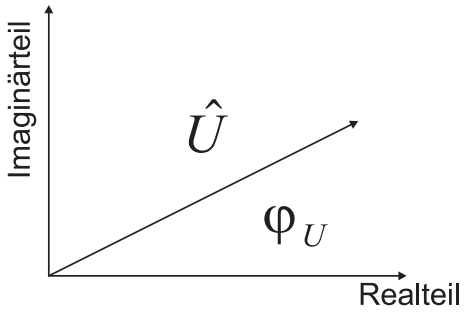



Abbildung 2.17.: Zeigerdiagramm für die Spannung $\underline{U}$

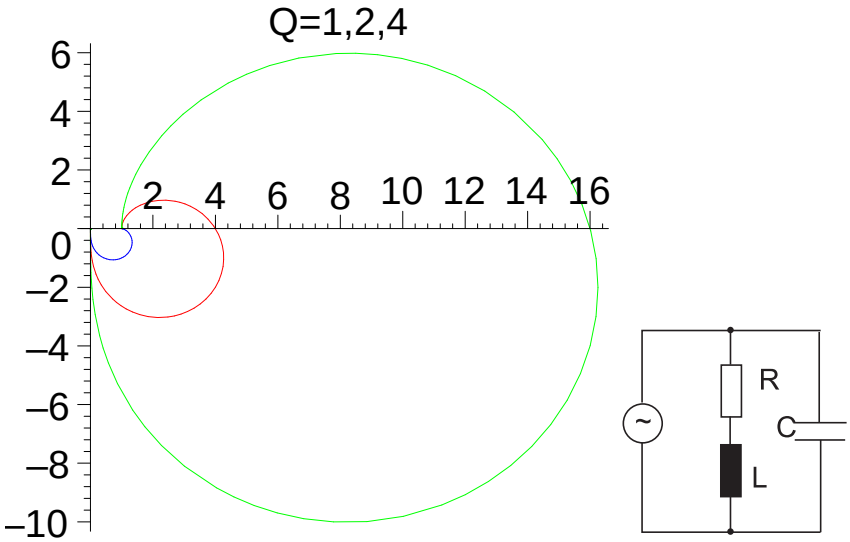



Abbildung 2.18.: Ortskurve für einen realen *Parallelschwingkreis*. Die Impedanz ist mit $R$ skaliert. Rechts befindet sich eine Skizze dieses Schwingkreises.

## 2.3. Übertragungsfunktionen

Bei der Besprechung von Übertragungsfunktionen gehen wir von der komplexen Darstellung aus[?]. Eine Einführung in die Materie kann in A gefunden werden. Ausgehend von komplexen Amplituden, kann man eine Zeigerdarstellung für den Real- und Imaginärteil angeben.

Wenn nun bei einer komplexen Impedanz $\underline{Z}$ diese von einem Parameter p abhängt, so kann man eine Ortskurve zeichnen. Üblicherweise ist $p = \omega$, es kann aber auch eine andere Grösse, z.B. die Kapazität bei einer Serienschaltung von Widerstand und Kondensator, sein. Bei Parallelschaltungen empfiehlt es sich, mit Leitwerten zu rechnen.

Ortskurven sind für Niederfrequenzanwendungen vielfach zu aufwendig zum berechnen. Es hat sich aber im Laufe der Jahre eingebürgert, dass Hochfrequenzanwendungen fast nur mit Hilfe von Ortskurven charakterisiert werden. Dies gilt z.B. auch für die Angabe des Frequenzverhaltens von Transistoren.

Abbildung 2.18 zeigt die Ortskurve für einen realen *Parallelschwingkreis*, bestehend aus einem Widerstand $R$, einer Spule $L$ und einem Kondensator $C$. Die Impedanz der Schaltung aus 2.18 berechnet sich am einfachsten über den komplexen

Leitwert $\underline{Y}$.

$$\underline{Y} = \frac{1}{R + j\omega L} + j\omega C = \frac{1 - \omega^2 LC + j\omega RC}{R + j\omega L} \tag{2.3.1}$$

Man könnte mit dem Leitwert $\underline{Y}$ genau so gut eine Ortskurve darstellen (zum Teil ist dies bei Hochfrequenzanwendungen üblich), aber wir wollen hier die Impedanz $\underline{Z}$ verwenden.

$$\underline{Z} = \frac{R + j\omega L}{1 - \omega^2 LC + j\omega RC} \tag{2.3.2}$$

Mit den üblichen Abkürzungen $\omega_0 = 1/\sqrt{LC}$ und $\Omega = \omega/\omega_0$ sowie $Q = \frac{1}{R}\sqrt{\frac{L}{C}}$ sowie nach der Normierung von $\underline{Z}$ mit $R$ wird

$$\frac{\underline{Z}}{R} = \underline{z} = \frac{1 + j\Omega\left[\left(1 - \Omega^2\right)Q - \frac{1}{Q}\right]}{\left(1 - \Omega^2\right)^2 + \left(\frac{\Omega}{Q}\right)^2} \tag{2.3.3}$$

Real- und Imaginärteile sind dann

$$\begin{aligned} \mathrm{Re}\left(\underline{z}\left(\Omega\right)\right) &= \frac{1}{\left(1 - \Omega^2\right)^2 + \left(\frac{\Omega}{Q}\right)^2} \\ \mathrm{Im}\left(\underline{z}\left(\Omega\right)\right) &= \frac{\Omega\left[\left(1 - \Omega^2\right)Q - \frac{1}{Q}\right]}{\left(1 - \Omega^2\right)^2 + \left(\frac{\Omega}{Q}\right)^2} \end{aligned} \tag{2.3.4}$$

Für grosse $Q$ kann die Abbildung verbessert werden, wenn man sowohl Real- wie auch Imaginärteil durch $Q^2$ teilt.

$$\begin{aligned} \mathrm{Re}\left(\frac{\underline{z}\left(\Omega\right)}{Q^2}\right) &= \frac{1}{\left(1 - \Omega^2\right)^2 Q^2 + \Omega^2} \\ \mathrm{Im}\left(\frac{\underline{z}\left(\Omega\right)}{Q^2}\right) &= \frac{\Omega\left[\left(1 - \Omega^2\right)Q - \frac{1}{Q}\right]}{\left(1 - \Omega^2\right)^2 Q^2 + \Omega^2} \end{aligned} \tag{2.3.5}$$

Die normierte Abbildung 2.19 zeigt schön, dass für grössere $Q$ die Ortskurve zu einem Kreis wird. Es ist dem Leser überlassen, Ortskurven komplizierterer Schaltungen zu berechnen.
Wenn wir eine Schaltung mit Blöcken entsprechend dem Kapitel 2.1 haben, mit $h(t)$ der Antwort des Systems auf einen Diracschen $\delta$-Impuls ist, dann ist die Antwort auf eine allgemeine Anregung $x(t)$ durch eine Faltung gegeben.

$$y\left(t\right) = \int_{-\infty}^{\infty} h(t')x\left(t - t'\right)dt' = h\left(t\right) * x\left(t\right) \tag{2.3.6}$$

**Beispiel (Maple-Datei)**: Ortskurven

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

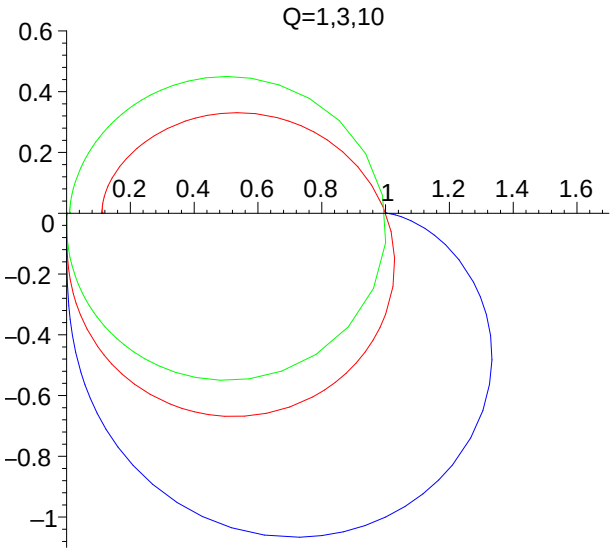



Abbildung 2.19.: Ortskurve für einen realen *Parallelschwingkreis*. Hier wurde sowohl durch R wie auch durch $Q^2$ geteilt.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

# 2.4. Kontinuierliche und diskrete Signale

Signale und Signalformen sind wesentlich zum Verständnis physikalischer Messsysteme. Generell werden Signale in zwei Kategorien aufgeteilt:

1. Kontinuierliche Signale
2. Zeitbegrenzte Signale

Die erste Kategorie von Signalen kann sowohl in der Zeitdomäne wie auch in der Frequenzdomäne behandelt werden. Die zweite Kategorie wird bevorzugt in der Zeitdomäne diskutiert. Genau genommen gibt es keine periodischen Signale, da nie eine unendliche Messzeit möglich ist. In der Frequenzdomäne wird vorwiegend mit der Fourier-Transformation gearbeitet. Die Fourier-Transformation setzt unendlich dauernde Signale voraus. Diese Signale verletzen aber die Kausalität. Um dieses Problem zu lösen verwendet man in der Regel in der Elektronik die Laplace-Transformation.

## 2.4.1. Signale

### 2.4.1.1. Periodische Signale

Eine erste Gruppe von Signalen sind die periodischen Signale. Diese können auf Summen von Sinus- oder Cosinus-Funktionen zurückgeführt werden. Typische, in der Elektronik vorkommende periodische Signale sind:

- Harmonische Funktionen:
  $\sin(\omega t)$ und $\cos(\omega t)$
- Rechteckfunktion:
  $$f(t) = \left\{ \begin{array}{ll} 1 & \text{für} \quad nT \leq t < \left(n+\frac{1}{2}\right)T \\ 0 & \text{für} \quad \left(n+\frac{1}{2}\right)T \leq t < (n+1)\,T \end{array} \right\} n \in \mathcal{Z}$$
- Impulsfunktion:
  $$f(t) = \left\{ \begin{array}{ll} 1 & \text{für} \quad nT \leq t < (n+\alpha)\,T \\ 0 & \text{für} \quad (n+\alpha)\,T \leq t < (n+1)\,T \end{array} \right\} n \in Z; \quad 0 \leq \alpha < 1$$
- Dreiecksfunktion:
  $$f(t) = \left\{ \begin{array}{ll} 2A\left(\frac{t}{T} - n - \frac{1}{4}\right) & \text{für} \quad nT \leq t < \left(n+\frac{1}{2}\right)T \\ 2A\left(n+1-\frac{t}{T}-\frac{1}{4}\right) & \text{für} \quad \left(n+\frac{1}{2}\right)T \leq t < (n+1)\,T \end{array} \right\} n \in Z$$
- Sägezahnfunktion:
  $$f(t) = A\left(\frac{t}{T} - n\right) \qquad \text{für} \quad nT \leq t < (n+1)\,T; n \in Z$$

### 2.4.1.2. Einmalige Signale, Einschalt- und Ausschaltvorgänge

Einmalige Funktionsverläufe treten auf, wenn ein Gerät eingeschaltet oder ausgeschaltet wird. Als Beispiel kann man die Ladekurve eines Kondensators $C$ betrachten, wenn er über einen Widerstand $R$ an eine *Spannungsquelle* $U$ angeschlossen

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

wird. Zur Zeit $t = 0$ soll die Verbindung eingeschaltet werden. Wir haben dann die folgenden Situation für $U_C$, die Spannung am Kondensator.

$$\begin{aligned} U_C(t) &= 0 && \text{für} \quad t < 0 \\ U_C(t) &= U\left(1 - e^{-\frac{t}{RC}}\right) && \text{für} \quad t \geq 0 \end{aligned} \tag{2.4.1}$$

Die resultierende Funktion ist klar nicht periodisch. Genau genommen gibt es nur nichtperiodische Signale, da alle periodischen eine unendlich lange Dauer haben müssten.

#### 2.4.1.3. Diskrete Signale

Ein typisches Beispiel für diskrete Signale ist die Treppenfunktion. Sie ist wie folgt definiert:

$$\begin{aligned} f(t) = f(nT) = \quad f_n \quad \textit{für} \quad & bT \leq t < (n+1)T \\ & (n = 0, 1, 2, \ldots; T > 0, T = const \end{aligned} \tag{2.4.2}$$

Die Treppenfunktion generiert aus der Folge $\{f_n\}$ kann in realen Schaltungen gut implementiert werden. Mathematisch einfacher zu handhaben ist jedoch der Dirac-Kamm:

$$\begin{aligned} f(t) = \quad \delta_n(nT); \qquad \textit{für} \quad & bT \leq t < (n+1)T \\ & (n = 0, 1, 2, \ldots; T > 0, T = const \end{aligned} \tag{2.4.3}$$

Der Dirac-Kamm erlaubt ein einfaches rechnen, da eine Integration mit Hilfe der $\delta$-Funktion sofort gelöst werden kann.

#### 2.4.1.4. Digitale Signale

Eine Sonderklasse der diskreten Signale sind die digitalen Signale, wie sie in der Computertechnik vorkommen. Die digitalen Signale haben zwei Werte, 0 oder 1. Diese Werte sind in Logikpegel kodiert. So ist bei der TTL-Logik der Nullwert $0V < x < 0.8V$ und der 1-Pegel $2V < x < 5V$. *Digitale Signale* werden mit logischen Schaltungen verknüpft. Ihre Schaltpegel sind definiert, eine Schaltung für digitale Signale darf nur mit den entsprechenden Pegelwerten betrieben werden, so dass keine undefinierten Zustände auftreten.
Heute werden in ausgewählten Anwendungsbereichen Logiken mit weichen Übergängen zwischen den einzelnen Zuständen verwendet. Diese *Fuzzy-Logik*en ermöglichen Aussagen wie: Es ist ein bisschen kalt, oder, es ist ein wenig zu warm. Gleitend definierte Übergänge zwischen Schaltzuständen sind vor allem bei schlecht in Zahlen fassbaren Problemen von Vorteil.

### 2.4.2. Fourier-Transformationen

Periodische Signale $f(t) = f(t + T)$ können als Reihenentwicklung

$$f(t) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos n\omega_0 t + b_n \sin n\omega_0 t\right) \tag{2.4.4}$$

geschrieben werden. Die Koeffizienten der Reihenentwicklung können wie folgt berechnet werden:

$$a_n = \frac{2}{T} \int_{t_0}^{t_0+T} f(t) \cos(n\omega_0 t) dt$$
$$b_n = \frac{2}{T} \int_{t_0}^{t_0+T} f(t) \sin(n\omega_0 t) dt \tag{2.4.5}$$

Alternativ kann eine komplexe Darstellung gewählt werden. Die Funktion heisst dann:

$$f(t) = \frac{1}{2} \sum_{n=-\infty}^{+\infty} c_n e^{jn\omega_0 t} \tag{2.4.6}$$

Auch hier können die $c_n$ mit einer Integralformel berechnet werden:

$$c_n = \frac{2}{T} \int_{t_0}^{t_0+T} f(t) e^{-jn\omega_0 t} dt = \begin{cases} \frac{1}{2}(a_n - jb_n) & \text{für } n > 0 \\ \frac{1}{2}(a_{-n} + jb_{-n}) & \text{für } n < 0 \end{cases} \tag{2.4.7}$$

Die Fourierkoeffizienten einer Funktion heissen das Amplitudenspektrum. Da die Sinus- und Cosinusfunktionen der Frequenz $\omega_0$ zusammen ein orthogonales Funktionensystem bilden, kann jede periodische Funktion dieser Frequenz eindeutig dargestellt werden. Die Amplitudenspektren haben die folgenden Eigenschaften:

- Je schnellere Änderungen des Signals auftreten, desto grösser sind die höheren Fourierkoeffizienten.
- Eine Funktion $f(t)$ wird durch eine trigonometrische Reihe $s_m(t) = \frac{\alpha_0}{2} + \sum_{n=1}^{m} \alpha_n \cos n\omega_0 t + \sum_{n=1}^{m} \beta_n \cos n\omega_0 t$ approximiert. Dann ist der quadratische Fehler $\delta^2 = \frac{1}{T} \int_0^T [f(t) - s_m(t)]^2 dt$ minimal, wenn die Koeffizienten $\alpha_n$ und $\beta_n$ die Fourierkoeffizienten sind.
- Für jede beschränkte und im Intervall $0 < t < T$ stückweise stetige Funktion konvergiert die Fourierreihe im Mittel gegen die gegebene Funktion.
- Ist eine Funktion einschliesslich ihrer k-ten Ableitung stetig, dann streben für $n \to \infty$ auch $a_n n^{k+1}$ und $b_n n^{k+1}$ gegen null.
- Ist $f(t)$ eine gerade Funktion, das heisst $f(-t) = f(t)$, so gilt: $b_n = 0 \quad (n = 0, 1, 2, \ldots)$. Dies ist die **Symmetrie erster Art**.
- Wenn $f(t)$ ungerade ist, das heisst, $f(-t) = -f(t)$, dann gilt: $a_n = 0 \quad (n = 0, 1, 2, \ldots)$. Dies ist die **Symmetrie zweiter Art**.
- Gilt $f\left(t + \frac{T}{2}\right) = -f(t)$, dann sind $a_{2n} = b_{2n} = 0 \quad (n = 0, 1, 2, \ldots)$. Dies ist die **Symmetrie dritter Art**.
- Besitzt eine ungerade Funktion die Symmetrie dritter Art (**Symmetrie vierter Art**), dann gilt $a_n = b_{2n} = 0 \quad (n = 0, 1, 2, \ldots)$.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

- Besitzt eine gerade Funktion die Symmetrie dritter Art (**Symmetrie vierter Art**), dann gilt $a_{2n} = b_n = 0 \quad (n = 0, 1, 2, \ldots)$.

Mit Hilfe der oben gezeigten Symmetrien kann sehr schnell der Oberwellengehalt einer Funktion abgeschätzt werden.
Für nichtperiodische Signale oder für Ausschnitte aus periodischen Signalen verwendet man die *Fouriertransformation* anstelle der Fourierreihe. Die *Fouriertransformation* und ihre Rücktransformation sind wie folgt definiert:

$$f(t) = \int_{-\infty}^{+\infty} \underline{F}(\omega)\, e^{j\omega t} d\omega \tag{2.4.8}$$

$$\underline{F}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(t)\, e^{-j\omega t} dt \tag{2.4.9}$$

$\underline{F}(\omega)$ ist die spektrale VerteilungsfunktionKreisfrequenz eines Signals. Damit sie existiert, muss das Integral

$$\int_{-\infty}^{+\infty} |f(t)|\, dt \tag{2.4.10}$$

endlich sein. Als Beispiel berechnen wir das Spektrum eines Rechteckimpulses. Der Impuls ist gegeben durch

$$f(t) = \begin{cases} 0 & für \quad t < -\frac{t_p}{2} \\ A & für \quad -\frac{t_p}{2} \leq t \leq \frac{t_p}{2} \\ 0 & für \quad t > \frac{t_p}{2} \end{cases} \tag{2.4.11}$$

Das Spektrum wird dann

$$\underline{F}(\omega) = \int_{-\frac{t_p}{2}}^{\frac{t_p}{2}} A e^{-j\omega t} dt \tag{2.4.12}$$

$$= j\frac{A}{\omega}\left(e^{-j\omega\frac{t_p}{2}} - e^{j\omega\frac{t_p}{2}}\right) \tag{2.4.13}$$

$$= t_p A \frac{\sin\left(\omega\frac{t_p}{2}\right)}{\omega\frac{t_p}{2}} \tag{2.4.14}$$

Das Spektrum $\underline{F}$ ist reell. Dies ist eine Konsequenz der Tatsache, dass $f(t)$ eine gerade Funktion ist. Wäre $f(t)$ eine ungerade Funktion, dann wäre das Spektrum rein imaginär. Die grössten Amplituden in $\underline{F}$ sind auf den Bereich $0 \leq f = \frac{\omega}{2\pi} \leq \frac{1}{t_p}$ beschränkt. $B = \frac{1}{t_p}$ heisst die Bandbreite des Impulses. Allgemein gilt für Pulse

$$Bt_p = 1 \tag{2.4.15}$$

Je kürzer also ein Puls ist,desto grösser ist seine Bandbreite. Für einen unendlich scharfen Puls, einen Dirac-$\delta$-Puls bedeutet dies, dass seine Spektralfunktion konstant ist. Dieses Gesetz hat eine Ähnlichkeit mit den Unschärferelationen der Quantenmechanik.

**2.4.2.0.1. Wiener-Khintchine-Relationen** Die *Wiener-Khintchine-Relation*en verknüpfen die Autokorrelationsfunktion mit dem Leistungsspektrum eines Signals. Wir definieren die Korrelationsfunktion $K(s)$ der Funktion $y(t)$ als das Ensemblemittel

$$K\left(s\right)=\left\langle y\left(t\right)y\left(t+s\right)\right\rangle \tag{2.4.16}$$

Die Grösse

$$K\left(0\right)=\left\langle y^{2}\left(t\right)\right\rangle \tag{2.4.17}$$

ist offensichtlich die Varianz von $y(t)$, wenn $s=0$ ist. Wie jede Funktion kann auch $K(s)$ als Fourierintegral geschrieben werden

$$K\left(s\right)=\int_{-\infty}^{\infty}J\left(\omega\right)e^{j\omega s}d\omega \tag{2.4.18}$$

$J(\omega)$ ist das Leistungsspektrum oder die Spektrale Dichte der Funktion $y(t)$. Die *Fouriertransformation* in Gleichung (2.4.18) kann umgekehrt werden.

$$J\left(\omega\right)=\frac{1}{2\pi}\int_{-\infty}^{\infty}K\left(s\right)e^{-j\omega s}ds \tag{2.4.19}$$

Die Gleichungen (2.4.18) und (2.4.18) Sind die *Wiener-Khintchine-Relation***en**. Die Relation kann bewiesen werden, indem man in Gleichung (2.4.18) von rechts mit $e^{-j\omega' s}$ multipliziert und über $s$ integriert.

$$\begin{aligned}\int_{-\infty}^{\infty}ds\;K\left(s\right)e^{-j\omega' s} &= \int_{-\infty}^{\infty}ds\int_{-\infty}^{\infty}d\omega J\left(\omega\right)e^{j(\omega-\omega')s}\\ &= 2\pi\int_{-\infty}^{\infty}d\omega J\left(\omega\right)\delta\left(\omega-\omega'\right)\\ &= J\left(\omega'\right)\end{aligned} \tag{2.4.20}$$

Die Korrelationsfunktion $K(s)$ ist reell und gerade, also eine Symmetrie erster Art. Deshalb würden bei einer Fourierreihe nur cos-Terme auftreten. Hier bedeutet dies, dass $J(\omega)$ auch reell und gerade ist, also

$$\begin{aligned}K^{*}\left(s\right) &= K\left(s\right)\\ K\left(-s\right) &= K\left(s\right)\\ J^{*}\left(\omega\right) &= J\left(\omega\right)\\ J\left(-\omega\right) &= J\left(\omega\right)\end{aligned} \tag{2.4.21}$$

Die beiden Relationen können bewiesen werden, indem man Gleichung (2.4.19 anwendet und dann die Integrationsvariable von $s$ nach $-s$ ändert. Gleichung (2.4.18) kann umgeschrieben werden

$$\begin{aligned}\left\langle y^{2}\right\rangle =K\left(0\right) &= \int_{-\infty}^{\infty}J\left(\omega\right)d\omega=\int_{0}^{\infty}J_{+}\left(\omega\right)d\omega\\ J_{+}\left(\omega\right) &\equiv 2J\left(\omega\right)\end{aligned} \tag{2.4.22}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Die Fourierintegrale können wegen den Symmetrieeigenschaften auch als cos-Transformationen geschrieben werden. Man setzt $e^{\pm j\omega s} = \cos\omega s \pm j \sin\omega s$ und erhält (da sin ungerade ist)

$$\begin{aligned} K(s) &= \int_{-\infty}^{\infty} J(\omega)\cos\omega s\, d\omega = 2\int_{0}^{\infty} J(\omega)\cos\omega s\, d\omega \\ J(\omega) &= \tfrac{1}{2\pi}\int_{-\infty}^{\infty} K(s)\cos\omega s\, ds = \frac{1}{\pi}\int_{0}^{\infty} K(s)\cos\omega s\, ds \end{aligned} \tag{2.4.23}$$

$K(s)$ und $J(\omega)$ können direkt aus den Fourierkoeffizienten von $y(t)$ berechnet werden. Wenn $y(t)$ stationär und ergodisch ist, ist $K(s)$ zeitunabhängig und das Ensemblemittel $\langle y\rangle$ kann durch das Zeitmittel $\{y\}$ ersetzt werden (Die Definitionen finden Sie in Reif [?] Kapitel 15.14 oder im Anhang M). Dies gilt für periodische Funktionen, muss aber für statistisch schwankende Funktionen wie das Rauschen gefordert werden. Gleichung (2.4.18) kann nun geschrieben werden als

$$K(s) = \frac{1}{2\Theta}\int_{-\Theta}^{\Theta} y(t)\, y(s+t)\, dt \tag{2.4.24}$$

Wir setzen

$$y_\Theta(t) \equiv \begin{cases} y(t) & \text{für } -\Theta < t < \Theta \\ 0 & \text{sonst} \end{cases} \tag{2.4.25}$$

und erhalten für $K(s)$

$$K(s) = \frac{1}{2\Theta}\int_{-\infty}^{\infty} y_\Theta(t)\, y_\Theta(s+t)\, dt \tag{2.4.26}$$

Durch die Ersetzung von $y(t)$ mit $y - \Theta(t)$ führt man einen Fehler der Grössenordnung $\frac{s}{\Theta}$ ein, der für $\Theta \to \infty$ verschwindet. Mit

$$y(t) = \int_{-\infty}^{\infty} C(\omega)\, e^{j\omega t} d\omega \tag{2.4.27}$$

($C(\omega)$ ist die *Fouriertransformation* von $y(t)$) erhält man

$$\begin{aligned} K(s) &= \frac{1}{2\Theta}\int_{-\infty}^{\infty} dt \int_{-\infty}^{\infty} d\omega C(\omega) e^{j\omega t} \int_{-\infty}^{\infty} d\omega' C(\omega') e^{j\omega'(s+t)} \\ &= \frac{1}{2\Theta}\int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' C(\omega) C(\omega') e^{j\omega' s} \int_{-\infty}^{\infty} dt e^{j(\omega+\omega')t} \\ &= \frac{1}{2\Theta}\int_{-\infty}^{\infty} d\omega \int_{-\infty}^{\infty} d\omega' C(\omega) C(\omega') e^{j\omega' s} \left[2\pi\delta(\omega+\omega')\right] \\ &= \frac{\pi}{\Theta}\int_{-\infty}^{\infty} d\omega C(\omega) C(-\omega) e^{j\omega s} \\ &= \frac{\pi}{\Theta}\int_{-\infty}^{\infty} d\omega |C(\omega)|^2 e^{j\omega s} \end{aligned} \tag{2.4.28}$$

Setzt man

$$J(\omega) = \frac{\pi}{\Theta}|C(\omega)|^2 \tag{2.4.29}$$

so erhält man Gleichung (2.4.18). Durch diese Rechnung wird klar, dass $J(\omega)$ das Leistungsspektrum ist. Die Wiener-Khintchine-Relationen lassen sich wie folgt zusammenfassen:

**Die Autokorrelation und das Leistungssprektrum sind Fouriertransformierte**

Zum Schluss sei angemerkt, dass

$$\left\langle y^2 \right\rangle = K(0) = \frac{\pi}{\Theta}\int_{-\infty}^{\infty} |C(\omega)|^2 d\omega \tag{2.4.30}$$

ist.

## 2.4.3. Laplace-Transformationen

Die *Fouriertransformation* im vorangegangenen Kapitel kann nur gelöst werden, wenn das Integral über den Betrag der Zeitfunktion endlich ist. Weiter müssen die Funktionswerte zu allen früheren, aber auch zu allen späteren Zeiten bekannt sein. Damit ist die *Fouriertransformation* akausal. Die Kausalität verlangt nun, dass ein *Signal* nur von seiner Vorgeschichte, nicht aber von seiner Zukunft abhängen kann. Eine Konsequenz der Kausalität ist, dass es keine beliebig scharfen Filter geben kann.
Mit der *Laplace-Transformation* kann insbesondere sehr elegant das Problem der Berechnung von Faltungsintegralen gelöst werden. Dieses Problem taucht immer dann auf, wenn Ein Ausgangssignal bei bekannter Impulsantwort aus dem Eingangssignal berechnet werden muss.
Die *Laplace-Transformation* ist nun definiert durch

$$F(p) = \int_{0}^{\infty} f(t) e^{-pt} dt \tag{2.4.31}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Hier ist $p = x + j\omega$ eine komplexe Funktion. Wenn $f(t) = 0$ ist für $t < 0$ dann ist in vielen Fällen die *Fouriertransformation* und die *Laplace-Transformation* äquivalent. Vielfach schreibt man für die Laplace-Transformation auch

$$F(p) = L(f(t)) \tag{2.4.32}$$

Die Umkehrfunktion der *Laplace-Transformation* ist nicht so einfach wie die der *Fouriertransformation*. Während bei dieser ein Vorzeichen gewechselt werden muss, benötigt die *Laplace-Transformation* eine Integration in der komplexen Ebene.

$$f(t) = \frac{1}{2\pi j} \lim_{r\to\infty} \int_{s-jr}^{s+jr} e^{pt} \frac{F(p)}{p} dp \tag{2.4.33}$$

dabei muss der Integrationsweg $s$ so gewählt werden, dass alle singulären Punkte des Integranden links von der Geraden $\Re p = s$ liegen. Die wichtigsten Eigenschaften der Laplace-transformierten Funktion sind:

- $L\left(\frac{df(t)}{dt}\right) = pF(p) - f(0)$
- $L\left(\frac{d^2 f(t)}{dt^2}\right) = p^2 F(p) - pf(0) - f'(0)$
- $L\left(\frac{d^n f(t)}{dt^n}\right) = p^n F(p) - p^{(n-1)} f(0) - p^{(n-2)} f'(0) - \ldots - f^{(n-1)}(0)$
- $L\left(\int_0^t f(t)\, dt\right) = \frac{1}{p} F(p)$
- $L(f(at)) = \frac{1}{a} F\left(\frac{p}{a}\right)$
- Wenn $L(f_1(t)) = F_1(p)$ und $L(f_2(t)) = F_2(p)$ ist, dann ist auch $L(a_1 f_1(t) + a_2 f_2(t)) = a_1 F_1(p) + a_2 F_2(p)$
- Dämpfungssatz: Wenn $L(f(t)) = F(p)$, dann ist $L(e^{-at} f(t)) = F(p + a)$
- Verschiebungssatz nach rechts (Retardation): Wenn $L(f(t)) = F(p)$ und $\lambda > 0$, dann ist $L(f(t)) = e^{-\lambda p} F(p)$
- Verschiebungssatz nach links: Wenn $L(f(t)) = F(p)$ und $\lambda > 0$, dann ist $L(f(t)) = e^{\lambda p}\left[F(p) - \int_0^\lambda e^{-pt} f(t) dt\right]$
- Satz von Borel: Wenn $L(f_1(t)) = F_1(p)$ und $L(f_2(t)) = F_2(p)$ ist, dann ist auch $L\left(\int_0^t f_1(t-\tau) f_2(\tau)\, d\tau\right) = \frac{1}{p} F_1(p) F_2(p)$
- Die Ausgangsfunktion zu $F(p) = p$ ist die Dirac-$\delta$-Funktion

Die Laplace-Transformation wird eingesetzt, um Differentialgleichungssysteme zu lösen. In der Elektronik wird Sie zur Berechnung von Frequenzgängen verwendet. Einige Funktionen und ihre Laplacetransformierten sind in der Tabelle C.1 im Anhang angegeben.
**Beispiel (Maple-Datei)**: Laplace-Transformation (Maple)

$$\{x(t_\mu)\} \rightarrow \boxed{T_a} \rightarrow \{y(t_\mu)\} = \{x(t_{\mu-1})\}$$

Abbildung 2.20.: Digitale Übertragungskette im Zeitbereich

$$X(z) \rightarrow \boxed{z^{-1}} \rightarrow Y(z)$$

$$Y(z) = z^{-1}X(z) = e^{-j2\pi\frac{f}{f_a}}X(z)$$

Abbildung 2.21.: Digitale Übertragungskette im Frequenzbereich

## 2.4.4. z-Transformationen

Die obigen Transformationen, die *Fouriertransformation* (Abschnitt 2.4.2) und die *Laplace-Transformation* (Abschnitt 2.4.3), können nur auf kontinuierliche Signale angewandt werden. Digitale Signalverarbeitung funktioniert aber nur mit zeit- und amplitudendiskreten Messwerten. Die hier besprochene **z-Transformation** ist die für dieses Problem angepasste Transformation. Die z-Transformation und die im Abschnitt 2.6.2 besprochenen Digitalfilter und -techniken können auch auf die Datenanalyse im Computer angewandt werden. Während die Laplace-Transformation und die Fourier-Transformation zur Lösung von Differentialgleichungen und -gleichungssystemen verwendet werden können, wird die z-Transformation zur Berechnung von **Systemen von Differenzengleichungen** verwendet.
Wir betrachten nun eine Übertragungskette für diskrete Signale (Abbildung 2.20 und 2.21).
Hier ist die Funktion $f(t)$ für die Zeiten $0 < t < \infty$ nur für diskrete Argumente $t_n = nT_a$ $(n = 0, 1, 2, \ldots; T_a > 0, T_a \text{ const})$ definiert. Die Amplitudenwerte an den diskreten Zeitwerten sind ebenfalls diskret. Die Folge $\{f_n\}$ und die an diskreten Zeitwerten definierte Funktion $f(nT_a)$ sind äquivalent.
Die z-Transformation $F(z)$ der Folge $\{f_n\}$ ist definiert durch

$$F(z) = \sum_{n=0}^{\infty} f_n z^{-n} \tag{2.4.34}$$

Die Folge $\{f_n\}$ heisst z-transformierbar, wenn die Summe in Gleichung (2.4.34) konvergiert. Als Kürzel kann man auch schreiben

$$F(z) = Z\{f_n\} \tag{2.4.35}$$

$\{f_n\}$ ist die Originalfolge, $F(z)$ die Bildfolge.

**2.4.4.0.1. Ein Beispiel** Sei $f_n = 1$, $(n = 0, 1, 2, 3, \ldots)$. Die z-Transformation ist

$$F(z) = \sum_{n=0}^{\infty} z^{-n} \tag{2.4.36}$$

Die Summe in Gleichung (2.4.36 ist bezüglich $\frac{1}{z}$ eine geometrische Reihe. Sie konvergiert gegen $\frac{z}{z-1}$, wenn $\frac{1}{z} < 1$ ist. Das heisst aber, dass die Folge $\{f_n\}$ z-transformierbar ist für alle $z$-Werte ausserhalb des Einheitskreises $|z| > 1$.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

#### 2.4.4.0.2. Eigenschaften

- Für jede z-transformierbare Folge $\{f_n\}$ ist $F(z)$ eine Potenzreihe bezüglich $z^{-1}$. Es existiert eine reelle Zahl $R$ als Konvergenzradius der durch $F(z)$ gegebenen Potenzreihe. $\{f_n\}$ ist dann für $|z| > R$ z-transformierbar.
- Wenn $\{f_n\}$ z-transformierbar ist, dann ist $F(z)$ eine analytische Funktion. und gleichzeitig die einzige Bildfunktion von $\{f_n\}$.
- Wenn $F(z)$ für $|z| > \frac{1}{R}$ analytisch ist und für $z \longrightarrow \infty$ regulär ist (d.h. $F(z)$ ist eine Potenzreihe analog zur Gleichung (2.4.36) und $F(\infty) = f_0$), dann gibt es genau eine Folge $\{f_n\}$.
- Wenn $F(z) = \mathrm{Z}\{f_n\}$ existiert, dann ist

$$f_0 = \lim_{z\to\infty} F(z) \tag{2.4.37}$$

Dabei kann z auf der reellen Achse oder längs eines beliebigen Weges nach $\infty$ laufen. Da

$$z\{F(z) - f_0\} = f_1 + f_2\frac{1}{z} + f_3\frac{1}{z^2} + \cdots \tag{2.4.38}$$

und

$$z^2\left\{F(z) - f_0 - f_1\frac{1}{z}\right\} = f_2 + f_3\frac{1}{z} + f_4\frac{1}{z^2} + \cdots \tag{2.4.39}$$

auch z-Transformierte sind,bekommt man

$$f_1 = \lim_{z\to\infty} z\{F(z) - f_0\} \qquad f_2 = \lim_{z\to\infty} z^2\left\{F(z) - f_0 - f_1\frac{1}{z}\right\}\ldots \tag{2.4.40}$$

**Auf die oben gezeigte Art und Weise kann aus der Bildfunktion $F(z)$ die Originalfolge $\{f_n\}$ rekonstruiert werden.**

- Wenn $\lim_{n\to\infty} f_n$ existiert, dann ist

$$\lim_{n\to\infty} f_n = \lim_{z\to 1+0} (z-1)F(z) \tag{2.4.41}$$

Um diesen Grenzwertsatz anwenden zu können, muss man wissen, dass der Grenzwert existiert. Der Satz ist nicht umkehrbar.

### 2.4.4.1. Rechenregeln für die z-Transformation

Wir gehen von der Folge $\{f_n\}$ und ihrer z-Transformierten $\mathrm{Z}\{f_n\} = F(z)$ aus

#### 1. Verschiebungssatz

$$\mathrm{Z}\{f_{n-k}\} = z^{-k}F(z) \qquad \textit{für } k = 0, 1, 2, \ldots \tag{2.4.42}$$

#### 2. Verschiebungssatz

$$\mathrm{Z}\{f_{n+k}\} = z^k\left[F(z) - \sum_{\nu=0}^{k-1} f_\nu z^{-\nu}\right] \qquad \textit{für } k = 1, 2, 3, \ldots \tag{2.4.43}$$

**Summation** Für $|z| > \max\left(1, \frac{1}{R}\right)$ gilt:

$$\mathrm{Z}\left\{\sum_{\nu=0}^{n-1} f_\nu\right\} = \frac{1}{z-1} F(z) \tag{2.4.44}$$

**Differenzenbildung** Für die Differenzen

$$\Delta f_n = f_{n+1} - f_n, \qquad \Delta^m f_n = \Delta\left(\Delta^{m-1} f\right) \qquad \left(m = 1, 2, \ldots; \Delta^0 f_n = f_n\right)$$

gilt:

$$\begin{array}{lcl} \mathrm{Z}\left\{\Delta f_n\right\} & = & (z-1)\, F(z) - f_0 \\ \mathrm{Z}\left\{\Delta^2 f_n\right\} & = & (z-1)^2\, F(z) - z\,(z-1)\, f_0 - z\Delta f_0 \\ \vdots & = & \vdots \\ \mathrm{Z}\left\{\Delta^k f_n\right\} & = & (z-1)^k\, F(z) - z \sum\limits_{\nu=0}^{k-1} (z-1)^{k-\nu-1}\, \Delta^\nu f_0 \end{array} \tag{2.4.45}$$

**Dämpfung** für $\lambda \neq 0$ beliebig komplex und $|z| > \frac{|\lambda|}{R}$ gilt

$$\mathrm{Z}\left\{\lambda^n f_n\right\} = F\left(\frac{z}{\lambda}\right) \tag{2.4.46}$$

**Faltung** Als Faltung der Folgen $\{f_n\}$ und $\{g_n\}$ bezeichnet man

$$f_n * g_n = \sum_{\nu=0}^{n} f_\nu g_{n-\nu} \tag{2.4.47}$$

Existieren $\mathrm{Z}\left\{f_n\right\} = F(z)$ für $|\lambda| > \frac{1}{R_1}$ und $\mathrm{Z}\left\{g_n\right\} = G(z)$ für $|\lambda| > \frac{1}{R_2}$ dann ist

$$\mathrm{Z}\left\{f_n * g_n\right\} = F(z)G(z) \tag{2.4.48}$$

Die in Gleichung (2.4.48 definierte *Faltung* konvergiert für $|z| > \max\left(\frac{1}{R_1}, \frac{1}{R_2}\right)$. Dieser *Faltungssatz* ist analog zu den Faltungssätzen für die *Fouriertransformation* und die Laplace-Transformation.

**Differentiation der Bildfunktion**

$$\mathrm{Z}\left\{n f_n\right\} = -z \frac{dF(z)}{dz} \tag{2.4.49}$$

Höhere Ableitungen lassen sich analog bilden.

**Integration der Bildfunktion** Falls $f_0 = 0$ gilt

$$\mathrm{Z}\left\{\frac{f_n}{n}\right\} = \int_z^\infty \frac{F(\xi)}{\xi} d\xi \tag{2.4.50}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

#### 2.4.4.2. z-Transformation und Laplace-Transformation

Die diskrete Funktion $f(t)$ kann auch als Treppenfunktion geschrieben werden:

$$f(t) = f(nT_A) = f_n \quad \text{für} \quad nT_A \le t < (n+1)T_A \quad (n = 0, 1, 2, \ldots; T_A > 0,\ T_A = \text{const} \tag{2.4.51}$$

Hier ist $T_A$ die Abtastzeit.
Die Laplace-Transformierte dieser Funktion ist (Siehe auch Gleichungen (2.4.31) und (2.4.32)).

$$\begin{aligned} L\{f(t)\} &= F(p) \\ &= \sum_{n=0}^{\infty} \int_{nT_A}^{(n+1)T_A} f_n e^{-pt} dt \\ &= \sum_{n=0}^{\infty} f_n \frac{e^{-nT_A p} - e^{-(n+1)T_A p}}{p} \\ &= \frac{1 - e^{-pT_A}}{p} \sum_{n=0}^{\infty} f_n e^{-nT_A p} \end{aligned} \tag{2.4.52}$$

Die unendliche Folge wird auch als *diskrete Laplace-Transformation* bezeichnet.

$$\mathrm{D}\{f(t)\} = \mathrm{D}\{f_n\} = \sum_{n=0}^{\infty} f_n e^{-nT_A p} \tag{2.4.53}$$

Ersetzt man in der Gleichung (2.4.53) $e^{T_A p}$ durch $z$ (dies ist der Ursprung der Bezeichnung z-Transformation) erhält man für Treppenfunktionen die Beziehung

$$pF(p) = \left(1 - \frac{1}{z}\right) F(z) \tag{2.4.54}$$

$$p\mathrm{L}\{f(t)\} = \left(1 - \frac{1}{z}\right) \mathrm{Z}\{f(t)\} \tag{2.4.55}$$

Mit Hilfe der Beziehungen in Gleichungen (2.4.54) und (2.4.55) kann man aus den Laplace-Transformationen in Tabelle C.1 die entsprechenden z-Transformationen ausrechnen.

#### 2.4.4.3. Rücktransformation

Die Rücktransformation

$$\mathrm{Z}^{-1}\{F(z)\} = \{f_n\} \tag{2.4.56}$$

der z-Transformation kann mit vier verschiedenen Methoden berechnet werden.

1. Benutzung von Tabellen
2. Berechnung der Laurent-Reihe von $F(z)$
3. Berechnung der Taylor-Reihe von $F\left(\frac{1}{z}\right)$. Es ist

$$f_n = \frac{1}{n} \frac{d^n}{dz^n} F\left(\frac{1}{z}\right)\Big|_{z=0} \qquad (n = 0, 1, 2, \ldots) \tag{2.4.57}$$

4. Anwendung eines Grenzwertsatzes

**2.4.4.3.1. Beispiel** Es soll die zu $F(z) = \frac{2z}{(z-2)(z-1)^2}$ gehörige Folge berechnet werden.

1. Aus der Partialbruchzerlegung von $\frac{F(z)}{z}$ erhält man

$$F(z) = \frac{2z}{z-2} - \frac{2z}{(z-1)^2} - \frac{2z}{z-1} \tag{2.4.58}$$

und somit

$$\{f_n\} = 2\left(2^n - n - 1\right) \quad \text{für } n \geq 0 \tag{2.4.59}$$

2. Entwicklung in Potenzen von $\frac{1}{z}$

$$F(z) = 2z^{-2} + 8z^{-3} + 22z^{-4} + 52z^{-5} + \ldots \tag{2.4.60}$$

Daraus kann man direkt die Folge $\{f_n\}$ ablesen. Einen geschlossenen Ausdruck erhält man jedoch nicht.

3. Man berechne die Ableitungen von $F\left(\frac{1}{z}\right)$

$$\begin{aligned}
F\left(\frac{1}{z}\right) &= \tfrac{2}{1-2z} - \tfrac{2z}{(1-z)^2} - \tfrac{2}{1-z} && \textit{also} \quad F\left(\frac{1}{z}\right)\Big|_0 = 0 \\
\frac{dF\left(\frac{1}{z}\right)}{dz} &= \tfrac{4}{(1-2z)^2} - \tfrac{4z}{(1-z)^3} - \tfrac{4}{(1-z)^2} && \textit{also} \quad \frac{dF\left(\frac{1}{z}\right)}{dz}\Big|_0 = 0 \\
\frac{d^2F\left(\frac{1}{z}\right)}{dz^2} &= \tfrac{16}{(1-2z)^3} - \tfrac{12z}{(1-z)^4} - \tfrac{12}{(1-z)^3} && \textit{also} \quad \frac{d^2F\left(\frac{1}{z}\right)}{dz^2}\Big|_0 = 4 \\
\frac{d^3F\left(\frac{1}{z}\right)}{dz^3} &= \tfrac{96}{(1-2z)^4} - \tfrac{48z}{(1-z)^5} - \tfrac{48}{(1-z)^4} && \textit{also} \quad \frac{d^3F\left(\frac{1}{z}\right)}{dz^3}\Big|_0 = 48
\end{aligned} \tag{2.4.61}$$

Berücksichtigt man, dass die Koeffizienten der Taylor-Entwicklung noch mit $n!$ normiert sind, erhält man ein konsistentes Resultat für $\{f_n\}$.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

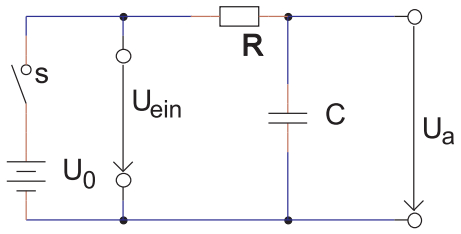



Abbildung 2.22.: Einschalten einer Spannung an einem RC-Glied

4. Aus den Grenzwertsätzen erhält man

$$\begin{aligned}
f_0 &= \lim_{z\to\infty} F(z) \\
&= \lim_{z\to\infty} \frac{2z}{z^3-4z^2+5z-2} = 0 \\
f_1 &= \lim_{z\to\infty} z\,(F(z)-f_0) \\
&= \lim_{z\to\infty} \frac{2z^2}{z^3-4z^2+5z-2} = 0 \\
f_2 &= \lim_{z\to\infty} z^2\left(F(z)-f_0-\frac{f_1}{z}\right) \\
&= \lim_{z\to\infty} \frac{2z^3}{z^3-4z^2+5z-2} = 2 \\
f_3 &= \lim_{z\to\infty} z^3\left(F(z)-f_0-\frac{f_1}{z}-\frac{f_2}{z^2}\right) \\
&= \lim_{z\to\infty} z^3\left(\frac{2z}{z^3-4z^2+5z-2}-\frac{2}{z^2}\right) = 8
\end{aligned} \tag{2.4.62}$$

## 2.4.5. Anwendung der Transformationen auf Einschaltvorgänge

Um Einschaltvorgänge zu untersuchen werden entweder die Stoss- oder Sprungantwort untersucht. Die Sprungantwort ist die Antwort des Systems auf die Sprungfunktion

$$U_{sp}(t;t_0;U_0) = \begin{cases} 0 & \text{für} \quad t < t_0 \\ U_0 & \text{für} \quad t \geq t_0 \end{cases} \tag{2.4.63}$$

am Eingang. Als erstes Beispiel berechnen wir die Ausgangsspannung an einem RC-Glied, wenn am Eingang eine Sprungfunktion angelegt wird (Abbildung 2.22). Unter der Voraussetzung, dass zur Zeit $t = 0$ der Kondensator entladen ist, erhält man mit $U_{ein} = IR + U_a$ und der Beziehung $I = \frac{dQ}{dt} = C\frac{dU_a}{dt}$ die Differentialgleichung

$$\frac{dU_a}{dt} + \frac{1}{RC}\,(U_a - U_{ein}) = 0 \tag{2.4.64}$$

Die Lösung dieser elementaren Differentialgleichung unter Berücksichtigung der Anfangsbedingungen ist

$$U_a(t) = U_0\left(1 - e^{-t/RC}\right) \tag{2.4.65}$$

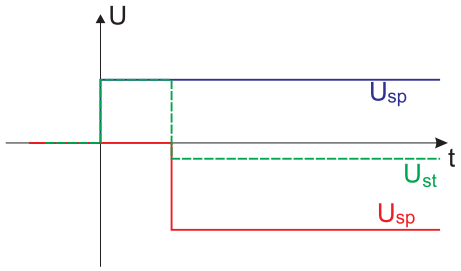



Abbildung 2.23.: Kombinierte Stoss- und Sprungfunktion $U_{st}(t, \Delta t, U_{st,0}, U_\infty)$, bestehend aus $U_{sp}\left(t, 0, U_{st,0}\right)$ (blau) und $U_{sp}\left(t, \Delta t, U_\infty - U_{st,0}\right)$ (rot)

Die Lösung, und damit die zeitliche Übertragungsfunktion oder Einheitsschrittantwort $\frac{U_a}{U_0}$ ist die bekannte Exponentialfunktion.
Im Allgemeinen besteht das Einschaltsignal aus einer Kombination einer Stossfunktion und einer Sprungfunktion, wie sie Abbildung 2.23 zeigt. Wenn wir annehmen, dass für $t < 0$ $U_e = 0$ gilt, dann für $0 \geq t < \Delta t$ der Wert $U_e = U_{st}$ und für $t \geq \Delta t$ $U_e = U_\infty$ ist, kann die Eingangsfunktion mit Gleichung (2.4.63 als

$$U_{st}(t, \Delta t, U_{st,0}, U_\infty) = U_{sp}\left(t, 0, U_{st,0}\right) + U_{sp}\left(t, \Delta t, U_\infty - U_{st,0}\right) \tag{2.4.66}$$

geschrieben werden. Unter der Voraussetzung, dass

- die Schaltung stabil ist
- das Überlagerungsgesetz gilt und
- die Sprungantwort unabhängig vom Zeitpunkt des Sprunges ist

lässt sich die Stossantwort mit $U_\infty = 0$ berechnen.

$$\begin{aligned}
U_a\left(t\right) &= U_a\left(U_{sp}\left(t, 0, U_{st,0}\right), t\right) + U_a\left(U_{sp}\left(t, \Delta t, -U_{st,0}\right), t\right) \\
&= \begin{cases} 0 & \text{für} \quad t < 0 \\ U_{sp}\left(1 - e^{-t/RC}\right) & \text{für} \quad 0 \leq \mathrm{t} < \Delta t \\ U_{sp}\left[\left(1 - e^{-t/RC}\right) - \left(1 - e^{-(t-\Delta t)/RC}\right)\right] & \text{für} \quad \Delta t \leq t \end{cases} \\
&= \begin{cases} 0 & \text{für} \quad t < 0 \\ U_{st,0}\left(1 - e^{-t/RC}\right) & \text{für} \quad 0 \leq \mathrm{t} < \Delta t \\ U_{st,0} e^{-t/RC}\left(e^{\Delta t/RC} - 1\right) & \text{für} \quad \Delta t \leq t \end{cases} \\
&= \begin{cases} 0 & \text{für} \quad t < 0 \\ U_{st,0}\left(1 - e^{-t/RC}\right) & \text{für} \quad 0 \leq \mathrm{t} < \Delta t \\ U_{st,0}\frac{\Delta t}{RC} e^{-t/RC} & \text{für} \quad \Delta t \leq t,\ \Delta t \to 0 \end{cases}
\end{aligned} \tag{2.4.67}$$

Die obige Gleichung zeigt, dass bei einer stossförmigen Anregung immer eine sofortige Antwort sowie eine langzeitliche Antwort vorhanden ist.
Wenn die Eingangsfunktion mathematisch nicht trivial ist, kann man sie in eine Folge von Stossfunktionen geschrieben werden.

$$U_e\left(t\right) = \sum_{n=0}^{\infty} U_{st}\left(t - n\Delta t, \Delta t, U\left(n\Delta t\right), 0\right) \tag{2.4.68}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Wir modifizieren die obige Gleichung

$$U_e(t) = \sum_{n=0}^{\infty} \frac{U_{st}(t - n\Delta t, \Delta t, U(n\Delta t), 0)}{\Delta t} \cdot \Delta t$$

Durch den Grenzübergang $\Delta t \to 0$ wird die obige Gleichung zu einem Integral, wobei $\frac{U_{st}}{\Delta t}$ die Diracsche Deltafunktion $\delta(t)$ approximiert

$$U_e(t) = \int_0^{\infty} U_{st}(\tau)\delta(t - \tau)\, d\tau = U_{st}(t) \tag{2.4.69}$$

einem Faltungsintegral, das gelöst werden kann. Aus der Stossantwort $U_a(t, n)$ einer einzelnen Stossfunktion erhält man die Antwort des gesamten Systems

$$U_a(m\Delta t) = \sum_{n=0}^{m} U_a(t, n) \tag{2.4.70}$$

Wenn $A(t) = \lim_{\Delta t \to 0} \frac{U_a(t)}{U_{st,0}}$ die zeitliche Übertragungsfunktion eines Einheitsschrittes ist, muss man, um zur Stossantwort (der Reaktion des Systems auf einen Dirac'schen Deltapuls $\delta(t)$ zu kommen, die folgende Umrechnung machen.
Wenn $\Theta(t)$ die *Heaviside-Funktion* ist, also ein Einheitssprung, gilt, dass

$$\delta(t) = \frac{d\Theta(t)}{dt}$$

wenn also $A(t)$ die Systemantwort auf eine Sprungfunktion ist, dann ist $\frac{dA(t)}{dt}$ bei einem *linearen System* die Systemantwort auf einen Deltapuls. Also wird die Sytemantwort auf einen beliebigen Eingang das Faltungsintegral

$$U_a(t) = U_e(0)\, A_0(0) + \int_0^t U_e(\tau) \frac{dA(t - \tau)}{dt} d\tau \tag{2.4.71}$$

Legt man beispielsweise an die Schaltung aus Abbildung 2.22 eine exponentiell ansteigende Spannung $U_e(t) = U_0\left(1 - e^{-t/\tau_0}\right)$ an und berücksichtigt, dass aus Gleichung (2.4.65)

$$\begin{aligned} A_0(0) &= 0 \\ A(t) &= 1 - e^{-t/RC} \\ A(t - \tau) &= 1 - e^{-\frac{t-\tau}{RC}} \\ \frac{dA(t - \tau)}{dt} &= \frac{1}{RC} e^{-\frac{t-\tau}{RC}} \end{aligned} \tag{2.4.72}$$

Die Antwortfunktion ist

$$U_a(t) = \frac{U_0}{RC} \int_0^t \left(1 - e^{-\frac{\tau}{\tau_0}}\right) e^{-\frac{t-\tau}{RC}} d\tau \tag{2.4.73}$$

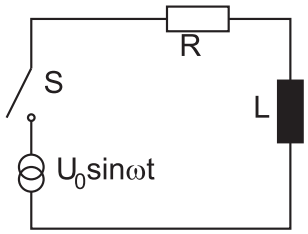



Abbildung 2.24.: Anlegen einer Wechselspannung an eine Spule

Wenn die Eingangsspannung die gleiche Zeitkonstante wie die RC-Schaltung hat, also $RC = \tau_0$, dann ergibt sich

$$U_a(t) = \frac{U_0}{\tau_0} e^{-\frac{t}{\tau_0}} \int_0^t \left(e^{\frac{\tau}{\tau_0}} - 1\right) d\tau = U_0 \left[1 - \left(1 + \frac{t}{\tau_0}\right) e^{-\frac{t}{\tau_0}}\right] \tag{2.4.74}$$

Es gibt die einfache Beziehung zwischen der *Fouriertransformation* und einem Faltungsintegral:

$$\begin{aligned} U_a(t) &= \int_0^\infty U_e(\tau)\, h(t-\tau)\, d\tau \\ &\Updownarrow \quad \text{Fouriertransformation} \\ U_a(\omega) &= h(\omega)\, U_e(\omega) \end{aligned} \tag{2.4.75}$$

Ein weiteres illustratives Beispiel ist das Anlegen einer Wechselspannung $U_0 \sin \omega t$ an eine Spule (Siehe Abbildung 2.17). Für $t > 0$ gilt die Differentialgleichung

$$L\frac{dI}{dt} + RI - U_{e,0} = 0 \tag{2.4.76}$$

Die Lösung dieser Gleichung, sowie die zeitliche Übertragungsfunktion sind

$$\begin{aligned} I(t) &= \frac{U_{e,0}}{R}\left(1 - e^{-\frac{R}{L}t}\right) \\ A(t) &= \frac{1}{R}\left(1 - e^{-\frac{R}{L}t}\right) \end{aligned} \tag{2.4.77}$$

Mit

$$\begin{aligned} A_0(0) &= 0 \\ A(t) &= \frac{1}{R}\left(1 - e^{-\frac{R}{L}t}\right) \\ A(t-\tau) &= \frac{1}{R}\left(1 - e^{-\frac{R}{L}(t-\tau)}\right) \\ \frac{dA(t-\tau)}{dt} &= \frac{1}{L} e^{-\frac{R}{L}(t-\tau)} \end{aligned} \tag{2.4.78}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

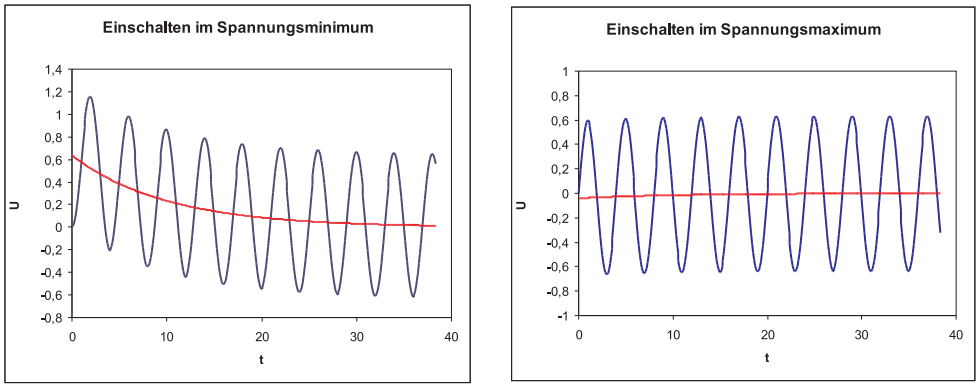



Abbildung 2.25.: Resultierende Ströme beim Anlegen einer Wechselspannung an eine Spule. **Links:** Einschalten im Nulldurchgang, **Rechts:** Einschalten bei Maximalspannung.

bekommt man für den Strom

$$\begin{aligned} I(t) &= \int_0^t U_0 \sin\omega t \frac{1}{L} e^{-\frac{R}{L}(t-\tau)} d\tau \\ &= \frac{U_0}{L} e^{-\frac{R}{L}t} \int_0^t \sin\omega t e^{\frac{R}{L}\tau} d\tau \\ &= \frac{U_0}{R^2+\omega^2 L^2}\left(R\sin\omega t - \omega L\cos\omega t + \omega L e^{-\frac{R}{L}t}\right) \end{aligned} \tag{2.4.79}$$

Schaltet man bei der Maximalspannung ein, ergibt sich

$$I(t) = \frac{U_0}{R^2+\omega^2 L^2}\left(\omega L\sin\omega t + R\cos\omega t - R e^{-\frac{R}{L}t}\right) \tag{2.4.80}$$

Wie aus Abbildung 2.25 (Excel-Tabelle) ersichtlich, wird der Dauerzustand sehr viel schneller erreicht, wenn man bei der Maximalspannung eine Spule an eine Wechselspannung als wenn man im Nulldurchgang schaltet. Der Grund ist der Folgende: Die Anfangsbedingung der allgemeinen Lösung hängt von der Grösse der speziellen Lösung zum Anfangszeitpunkt ab. Da bei einer Spule der Strom im Dauerzustand $90^0$ ausser Phase ist, muss bei einem Einschalten im Nulldurchgang die Anfangsbedingung der allgemeinen Lösung den maximalen Strom kompensieren. Wenn bei der maximalen Spannung eingeschaltet wird, ist der Strom null, die Anfangsbedingung der allgemeinen Lösung ist also auch null. Je weniger die Spule durch den Quellwiderstand gedämpft wird, desto ausgeprägter ist der Effekt, dass die Zeit zum Erreichen des Gleichgewichtszustandes im ersten Fall grösser ist.
**Beispiel (HTML-Datei)**: Anwendung der z-Transformation
**Beispiel (PDF-Datei)**: Anwendung der z-Transformation
**Beispiel (Maple-Datei)**: Berechnung der z-Transformation

## 2.4.6. Digitale Signale

Digitale Signale, also 0 oder 1, werden mit logischen Schaltungen verknüpft. Digitale Signale werden mit Hilfe von *Zahlensystemen* auf unsere Werteskale der natürlichen Zahlen abgebildet. Gebräuchlich sind:

| Dual (binär) | Dezimal | Oktal | Hexadezimal |
|---|---|---|---|
| 000 000 = 0000 | 0 | 0 | 0 |
| 000 001 = 0001 | 1 | 1 | 1 |
| 000 010 = 0010 | 2 | 2 | 2 |
| 000 011 = 0011 | 3 | 3 | 3 |
| 000 100 = 0100 | 4 | 4 | 4 |
| 000 101 = 0101 | 5 | 5 | 5 |
| 000 110 = 0110 | 6 | 6 | 6 |
| 000 111 = 0111 | 7 | 7 | 7 |
| 001 000 = 1000 | 8 | 10 | 8 |
| 001 001 = 1001 | 9 | 11 | 9 |
| 001 010 = 1010 | 10 | 12 | A |
| 001 011 = 1011 | 11 | 13 | B |
| 001 100 = 1100 | 12 | 14 | C |
| 001 101 = 1101 | 13 | 15 | D |
| 001 110 = 1110 | 14 | 16 | E |
| 001 111 = 1111 | 15 | 17 | F |

Tabelle 2.4.: Vergleich der *Zahlensysteme*

(1)

| A | B |
|---|---|
| 0 | 0 |
| 1 | 0 |

(2)

| A | B |
|---|---|
| 0 | 1 |
| 1 | 1 |

(3)

| A | B |
|---|---|
| 0 | 0 |
| 1 | 1 |

(4)

| A | B |
|---|---|
| 0 | 1 |
| 1 | 0 |

Tabelle 2.5.: Logische Verknüpfungen mit einer Eingangsvariablen

- Das binäre Zahlensystem, bestehend aus Dualzahlen. 10110B, "B"nachgestellt.
- Das Oktalsystem. 234O oder 234Q , O (der Buchstabe, Verwechslungsgefahr) oder Q nachgestellt
- das Hexadezimalsystem. 3FH oder $3F, H nachgestellt oder $ vorangestellt
- das Dezimalsystem. 123D, D nachgestellt

In Tabelle 2.4 haben wir, wie üblich für die Ziffern "10" bis "15" die Buchstaben $A \ldots F$ verwendet. Eine schöne Übersicht über die Grundlagen der Digitaltechnik findet man im Buch von Häßler und Straub[**?**].

### 2.4.6.1. Logische Verknüpfungen

**2.4.6.1.1. Nicht-Gatter** Die Grundverknüpfungen mit einer Variable sind in Tabelle 2.5 angegeben. Die Verknüpfungen (1) und (2) haben keinen praktischen Nutzen. (3) ist die Funktion eines *Buffer*s während (4) eine Negation darstellt. Die Logiktafel der Negation sowie die Schaltbilder und die Signalformen sind in Abbildung 2.26 dargestellt. Man schreibt die Negation üblicherweise

$$Z = \neg A = \bar{A} \tag{2.4.81}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

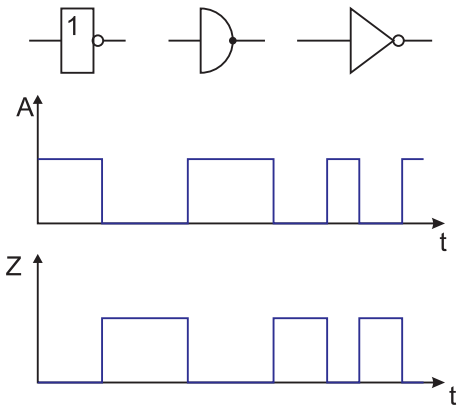



Abbildung 2.26.: *Negation*, *Nicht-Gatter*, *Inverter* und *NOT*

**2.4.6.1.2. Und-Gatter** Die erste der zweiwertigen Funktionen ist das *Und-Gatter*. Sein Ausgang ist eins, genau wenn beide Eingänge auf eins sind. Die Logiktafel des Und-Gatters sowie die Schaltbilder und die Signalformen sind in Abbildung 2.27 dargestellt. In Formeln stellt man die Konjunktion (Und) wie folgt dar:

$$\begin{array}{lcll} Z & = & A \wedge B & \text{genormt} \\ Z & = & A * B & \text{veraltet} \\ Z & = & A\&B & \text{selten(Schreibmaschine!)} \end{array} \tag{2.4.82}$$

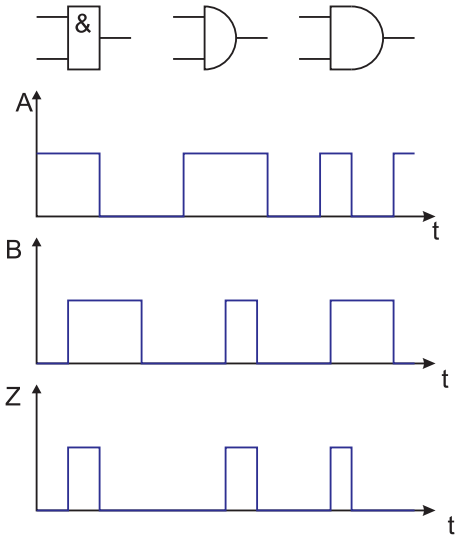



Abbildung 2.27.: *Und*, Und-Gatter, *Konjunktion* und *AND*

**2.4.6.1.3. Oder-Gatter** Die zweite der zweiwertigen Grundfunktionen ist das *Oder-Gatter*. Sein Ausgang ist eins, wenn mindestens einer der beiden Eingänge auf eins ist. Die Logiktafel des Oder-Gatters sowie die Schaltbilder und die Signalformen sind in Abbildung 2.28 dargestellt. In Formeln stellt man die Disjunktion (Oder) wie folgt dar:

$$\begin{aligned} Z &= A \vee B \quad \text{genormt} \\ Z &= A + B \quad \text{veraltet} \end{aligned} \tag{2.4.83}$$

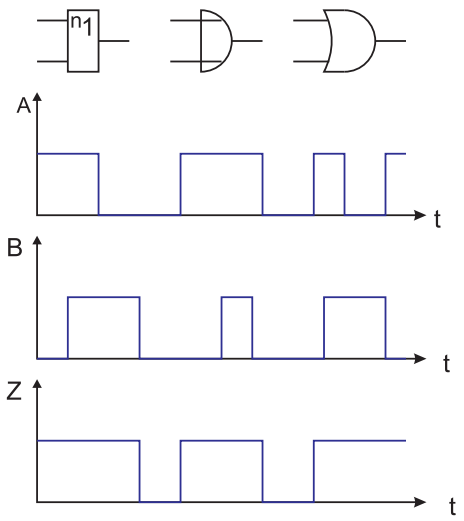



Abbildung 2.28.: Oder, Oder-Gatter, Disjunktion, OR

Mit den Verknüpfungen **UND**, **ODER** und **NICHT** können alle logischen Verknüpfungen erzeugt werden. Es hat sich aber herausgestellt, dass einige andere abgeleitete Verknüpfungen einfacher in Silizium zu bauen sind. Mit einer Auswahl abgeleiteter Funktionen lassen sich ebenso alle Verknüpfungen herstellen.

**2.4.6.1.4. NAND-Gatter** Eine häufig verwendete abgeleitete Verknüpfung ist das NAND-Gatter. Sein Name ist Abgeleitet aus NOT und AND. Sein Ausgang ist eins, wenn einer der beiden Eingänge nicht auf eins ist. Die Logiktafel des NAND-Gatters sowie die Schaltbilder und die Signalformen sind in Abbildung 2.29 dargestellt. In Formeln stellt man die NAND-Funktion wie folgt dar:

$$\begin{aligned} Z &= \overline{A \wedge B} \quad \text{genormt} \\ Z &= \overline{A * B} \quad \text{veraltet} \end{aligned} \tag{2.4.84}$$

**2.4.6.1.5. NOR-Gatter** Eine weitere häufig verwendete abgeleitete Verknüpfung ist das NOR-Gatter. Sein Name ist abgeleitet aus NOT und OR. Sein Ausgang ist eins, wenn beide Eingänge nicht auf eins sind. Die Logiktafel des NOR-Gatters sowie die Schaltbilder und die Signalformen sind in Abbildung 2.30 dargestellt. In Formeln stellt man die NAND-Funktion wie folgt dar:

$$\begin{aligned} Z &= \overline{A \vee B} \quad \text{genormt} \\ Z &= \overline{A + B} \quad \text{veraltet} \end{aligned} \tag{2.4.85}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

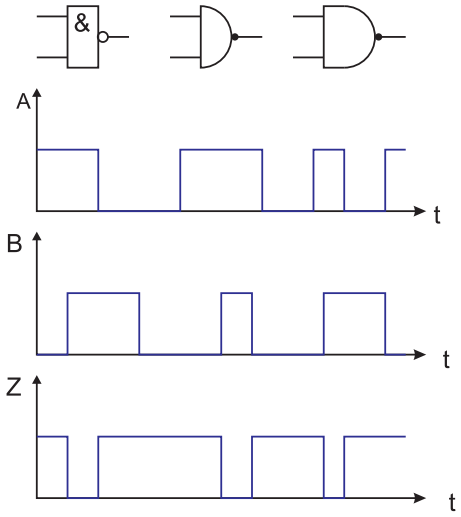



Abbildung 2.29.: NAND, NAND-Gatter

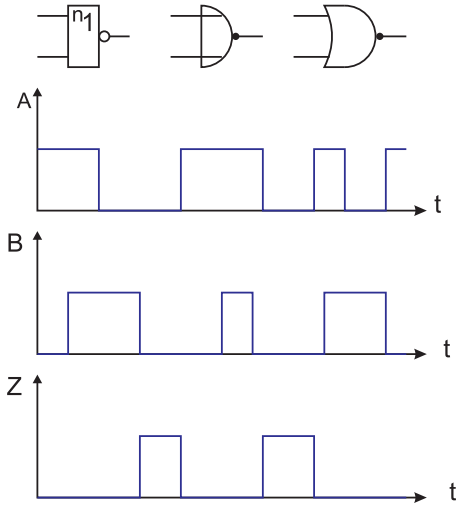



Abbildung 2.30.: NOR, NOR-Gatter

**2.4.6.1.6. Äquivalenzgatter** Eine weitere abgeleitete Verknüpfung ist das Äquivalenz-Gatter. Sein Ausgang ist eins, wenn beide Eingänge auf gleichem Pegel sind. Die interne Funktion kann wie in Tabelle 2.6 gezeigt, abgeleitet werden. Die Logiktafel des Äquivalenz-Gatters sowie die Schaltbilder und die Signalformen sind in Abbildung 2.31 dargestellt. In Formeln stellt man die Äquivalenz-Funktion wie folgt dar:

$$\begin{aligned} Z &= (A \wedge B) \vee \left(\overline{A} \wedge \overline{B}\right) \qquad \text{genormt} \\ Z &= (A * B) + \left(\overline{A} * \overline{B}\right) \qquad \text{veraltet} \end{aligned} \tag{2.4.86}$$

**2.4.6.1.7. Antivalenzgatter oder XOR** Eine weitere abgeleitete Verknüpfung ist das Antivalenz-Gatter, auch XOR-Gatter genannt. Der Name XOR ist eine amerikanisch prägnante Abkürzung für e**X**clusive **OR**. Sein Ausgang ist eins, wenn beide Eingänge auf verschiedenem Pegel sind. Die Logiktafel des Antivalenz-

| $A$ | $B$ | $\overline{A}$ | $\overline{B}$ | $Q = A \wedge B$ | $S = \overline{A} \wedge \overline{B}$ | $Z = Q \vee S$ |
|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| 0 | 1 | 1 | 0 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 1 | 0 | 1 |

Tabelle 2.6.: Wahrheitstabelle des Äquivalenzgatters

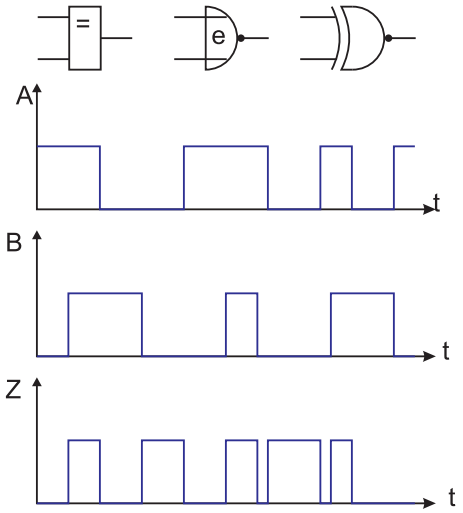



Abbildung 2.31.: Äquivalenz, Exklusiv-NOR, XNOR

Gatters sowie die Schaltbilder und die Signalformen sind in Abbildung 2.32 dargestellt. In Formeln stellt man die Antivalenz-Funktion wie folgt dar:

$$\begin{aligned} Z &= \overline{(A \wedge B) \vee \left(\overline{A} \wedge \overline{B}\right)} && \text{genormt} \\ Z &= \left(A \wedge \overline{B}\right) \vee \left(\overline{A} \wedge B\right) && \text{umgeformt, genormt} \\ Z &= \left(A * \overline{B}\right) + \left(\overline{A} * B\right) && \text{veraltet} \end{aligned} \tag{2.4.87}$$

### 2.4.6.2. Boolesche Algebra, Schaltalgebra

Die Reihenschaltung von zwei Schaltern führt auf die Verknüpfung **UND**. Die Parallelschaltung ergibt demnach **ODER**. Dabei wird angenommen, dass der Schaltzustand 1 dem geschlossenen Schalter entspricht.
In Tabelle 2.7 sind die Grundrechenregeln für Konstanten zusammengefasst. Tabelle 2.8 zeigt die Theoreme der Schaltalgebra. Dabei wird nun eine Variable, $A$, eingeführt, deren Wert beliebig ist.

| | | | | |
|---|---|---|---|---|
| UND | $0 \wedge 0 = 0$ | $0 \wedge 1 = 0$ | $1 \wedge 0 = 0$ | $1 \wedge 1 = 1$ |
| ODER | $0 \vee 0 = 0$ | $0 \vee 1 = 1$ | $1 \vee 0 = 1$ | $1 \vee 1 = 1$ |
| NICHT | $\overline{0} = 1$ | $\overline{1} = 0$ | | |

Tabelle 2.7.: Postulate der Schaltalgebra

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

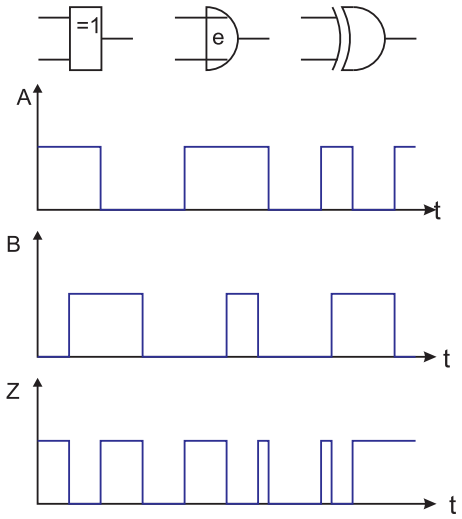



Abbildung 2.32.: Antivalenz, XOR-Gatter,

| | | | | |
|---|---|---|---|---|
| UND | $A \wedge 0 = 0$ | $A \wedge 1 = A$ | $A \wedge A = A$ | $A \wedge \overline{A} = 0$ |
| ODER | $A \vee 0 = A$ | $A \vee 1 = 1$ | $A \vee A = A$ | $A \vee \overline{A} = 1$ |
| NICHT | $\overline{\overline{0}} = 0$ | $\overline{\overline{1}} = 1$ | | |

Tabelle 2.8.: Theoreme der Schaltalgebra

Wie bei den natürlichen oder ganzen Zahlen macht das Kommutativgesetz eine Aussage über die Vertauschbarkeit von Variablen bei der UND- oder ODER-Verknüpfung.

$$\begin{aligned} Z = &\quad A \wedge B \wedge C = \quad C \wedge B \wedge A \\ Z = &\quad A \vee B \vee C = \quad C \vee B \vee A \end{aligned} \tag{2.4.88}$$

Analog gibt es auch ein Assoziativgesetz, das aussagt, dass die Reihenfolge der Verknüpfung beliebig ist.

$$\begin{aligned} Z = &\quad A \wedge (B \wedge C) = \quad (A \wedge B) \wedge C \\ Z = &\quad A \vee (B \vee C) = \quad (A \vee B) \vee C \end{aligned} \tag{2.4.89}$$

Auch für die Schaltalgebra gibt es Distributivgesetze. Man unterscheidet das konjunktive Distributivgesetz

$$Z = A \wedge (B \vee C) = (A \wedge B) \vee (A \wedge C) \tag{2.4.90}$$

und das disjunktive Distributivgesetz

$$Z = A \vee (B \wedge C) = (A \vee B) \wedge (A \vee C) \tag{2.4.91}$$

Zusätzlich zu den oben ausgeführten Gesetzen, die auch von den üblichen Zahlensystemen her bekannt sind, gibt es die deMorganschen Gesetze. Das erste DeMorgansche Gesetz lautet

$$Z = \overline{A \wedge B} = \overline{A} \vee \overline{B} \tag{2.4.92}$$

| $A$ | $B$ | $A \wedge B$ | $\overline{A \wedge B}$ | $\overline{A}$ | $\overline{B}$ | $\overline{A} \vee \overline{B}$ |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| 0 | 1 | 0 | 1 | 1 | 0 | 1 |
| 1 | 0 | 0 | 1 | 0 | 1 | 1 |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 |

Tabelle 2.9.: Wahrheitstabelle des ersten DeMorganschen Gesetzes

| $A$ | $B$ | $A \vee B$ | $\overline{A \vee B}$ | $\overline{A}$ | $\overline{B}$ | $\overline{A} \wedge \overline{B}$ |
|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 1 | 1 | 1 | 1 |
| 0 | 1 | 1 | 0 | 1 | 0 | 0 |
| 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| 1 | 1 | 1 | 0 | 0 | 0 | 0 |

Tabelle 2.10.: Wahrheitstabelle des zweiten DeMorganschen Gesetzes

Die Gültigkeit dieses Gesetzes kann mit der Tabelle 2.9 gezeigt werden.
Das zweite DeMorgansche Gesetz lautet:

$$Z = \overline{A \vee B} = \overline{A} \wedge \overline{B} \tag{2.4.93}$$

Die Gültigkeit dieses Gesetzes kann mit der Tabelle 2.10 gezeigt werden.
Analog zum Rechnen mit ganzen Zahlen (als Beispiel) wird definiert, dass UND stärker bindet als ODER (Punkt vor Strich). damit erreicht man, dass nicht immer Klammern gesetzt werden müssen, um die Reihenfolge der Ausführung von Operationen festzulegen. Aus den DeMorganschen Gesetzen folgt, dass jede UND-Verknüpfung mit ODER- und NICHT-Verknüpfungen realisiert werden kann. Da man immer eine NICHT-Verknüpfung aus einer NOR-Verknüpfung erzeugen kann (entweder man legt einen Eingang auf null, oder man verbindet beide Eingänge) benötigt man, im Prinzip, nur NOR-Gatter, um eine gesamte Logik aufzubauen. Diese Aussage mag, wenn man an integrierte Schaltungen wie die 70LSxx-Reihe denkt, übertrieben klingen. Wenn man eine grössere logische Schaltung jedoch mit programmierbaren Logik-Array (PAL) aufbaut, dann hilft einem die obige Aussage, um mit einem Typ Schaltungen alles aufzubauen. Analog kann man auch zeigen, dass alle logischen Schaltungen aus NAND-Verknüpfungen aufgebaut werden können. Welche Verknüpfung man bevorzugt, hängt unter anderem auch vom inneren Aufbau der Logikfamilien ab.

**2.4.6.2.1. Normalformen** Eine digitale Schaltung ist eindeutig durch ihre Wahrheitstabelle gegeben. Aus der Wahrheitstabelle können zwei Normalformen abgelesen werden.

**ODER-(disjunktive) Normalform (DNF)** Eine DNF ist eine Oderverknüpfung von Vollkonjunktionen (nur UND, jede Variable kommt nur einmal vor). Ziel sind die Zustände 1. Die einzelnen Terme heissen **Minterm**.

**UND-(konjunktive) Normalform (KNF)** Eine KNF ist eine Undverknüpfung von Volldisjunktionen (nur ODER, jede Variable kommt nur einmal vor). Ziel sind die Zustände 0. Die einzelnen Terme heissen **Maxterm**.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

| $A$ | $B$ | $C$ | $Z$ | | |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 1 | $\Longrightarrow$ | $\overline{A} \wedge \overline{B} \wedge \overline{C}$ |
| 0 | 0 | 1 | 0 | | |
| 0 | 1 | 0 | 1 | $\Longrightarrow$ | $\overline{A} \wedge B \wedge \overline{C}$ |
| 0 | 1 | 1 | 0 | | |
| 1 | 0 | 0 | 0 | | |
| 1 | 0 | 1 | 1 | $\Longrightarrow$ | $A \wedge \overline{B} \wedge C$ |
| 1 | 1 | 0 | 0 | | |
| 1 | 1 | 1 | 1 | $\Longrightarrow$ | $A \wedge B \wedge C$ |

$$Z = \left(\overline{A} \wedge \overline{B} \wedge \overline{C}\right) \vee \left(\overline{A} \wedge B \wedge \overline{C}\right) \vee \left(A \wedge \overline{B} \wedge C\right) \vee (A \wedge B \wedge C)$$

Tabelle 2.11.: Erzeugung einer DNF aus der Wahrheitstabelle

| $x_1$ | $x_2$ | y |
|---|---|---|
| 0 | 0 | 0 |
| 0 | 1 | 0 |
| 1 | 0 | 0 |
| 1 | 1 | 1 |

| $x_2 \backslash x_1$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 1 |

Abbildung 2.33.: Vergleich einer Wahrheitstabelle mit einem Karnaugh-Diagramm

Eine DNF sieht dann so aus

$$Z = \left(A \wedge \overline{B} \wedge C\right) \vee \left(\overline{A} \wedge \overline{B} \wedge C\right) \vee (\ldots) \vee \ldots \tag{2.4.94}$$

Entsprechend sieht eine KNF aus.

$$Z = \left(A \vee \overline{B} \vee C\right) \wedge \left(\overline{A} \vee \overline{B} \vee C\right) \wedge (\ldots) \wedge \ldots \tag{2.4.95}$$

Tabelle 2.11 zeigt, wie man aus der Wahrheitstabelle die DNF erzeugt. Man muss nur diejenigen Terme aufschreiben, bei denen als Resultat in der Wahrheitstabelle eine 1 steht. Je weniger Einsen eine Wahrheitstabelle hat, desto effizienter ist die DNF. Die KNF andererseits ist dann anzuwenden, wenn im Ausgangsfeld nur wenige Nullen sind. Tabelle 2.12 zeigt das entsprechende Vorgehen. In der Tabelle wird gezeigt, dass man die DNF auf die negierte Ausgangsvariable $z$ anwendet. Die resultierende Form wird negiert. Schliesslich werden die DeMorganschen Gesetze angewendet. Die KNF wird also erhalten, indem man die Zeilen heraus sucht, die $z = 0$ haben. Für jede dieser Zeilen wird eine Disjunktion (M̈axterm) hingeschrieben, wobei jede Variable mit ïäls $\overline{A}$, jede mit öäls $A$ geschrieben wird.
Eine weitergehende Übersicht über das Arbeiten mit logischen Schaltungen kann in der Referenz [**?**] gefunden werden.

### 2.4.6.3. Karnaugh-Diagramme

Karnaugh-Diagramme bieten eine weitere Vereinfachungsmöglichkeit für logische Schaltnetzwerke. Wie Abbildung 2.33 zeigt, kann aus einer Wahrheitstabelle das Karnaugh-Diagramm abgeleitet werden. Dabei sind die folgenden Regeln zu beachten:

| $A$ | $B$ | $Z$ | $\overline{Z}$ | | |
|---|---|---|---|---|---|
| 0 | 0 | 0 | 1 | $\Longrightarrow$ | $\overline{A} \wedge \overline{B}$ |
| 0 | 1 | 1 | 0 | | |
| 1 | 0 | 0 | 1 | $\Longrightarrow$ | $A \wedge \overline{B}$ |
| 1 | 1 | 1 | 0 | | |

$$\overline{Z} = \left(\overline{A} \wedge \overline{B}\right) \vee \left(A \wedge \overline{B}\right) \qquad \text{aus Tabelle}$$

$$Z = \overline{\left(\overline{A} \wedge \overline{B}\right) \vee \left(A \wedge \overline{B}\right)} \qquad \text{Negation}$$

$$Z = (A \vee B) \wedge \left(\overline{A} \vee B\right) \qquad \text{DeMorgan}$$

Tabelle 2.12.: Erzeugung einer KNF aus der Wahrheitstabelle

- Die Wahrheitstabelle wird zweidimensional angeordnet. Bei einer geraden Anzahl von Eingangsvariablen enthalten Zeilen und Spalten je die Hälfte der Variablen, sonst muss z.B. die Spalten mehr Variablen enthalten.
- Die Variablen werden so angeordnet, dass sich von einer Spalte (Zeile) zur nächsten nur eine Variable ändert. Bemekung. Dies ist der Gray-Code.
- In die Felder werden die Werte der Resultatvariablen $y$ eingetragen.

Wir betrachten in Abbildung 2.34 die beiden Ausgangszellen oben links. Hier steht, dass für die Eingangsvektoren 0000 und 0100 das Ausgangssignal jeweils eins ist. Bei der Berechnung der disjunktiven Normalform ergibt sich für die beiden Zeilen die Konjunktionen

$$K_1 = \bar{x}_1 \wedge \bar{x}_2 \wedge \bar{x}_3 \wedge \bar{x}_4 \tag{2.4.96}$$
$$K_2 = \bar{x}_1 \wedge x_2 \wedge \bar{x}_3 \wedge \bar{x}_4 \tag{2.4.97}$$

Die dann zu bildende Disjunktion liefert unter anderem den Term

$$\begin{aligned} K_1 \vee K_1 &= (\bar{x}_1 \wedge \bar{x}_2 \wedge \bar{x}_3 \wedge \bar{x}_4) \vee (\bar{x}_1 \wedge x_2 \wedge \bar{x}_3 \wedge \bar{x}_4) \\ &= (\bar{x}_1 \wedge \bar{x}_3 \wedge \bar{x}_4) \wedge (\bar{x}_2 \vee x_2) \\ &= \bar{x}_1 \wedge \bar{x}_3 \wedge \bar{x}_4 \end{aligned} \tag{2.4.98}$$

Das Beispiel zeigt, dass jedesmal, wenn $2, 4, 8, 16, \ldots$ Zellen in einer kompakten Gruppe mit eins belegt sind, dass dann in der disjunktiven Normalform nur diejenigen Variablen auftauchen, die sich in der den Zeilen oder Spalten, über welche die Gruppe geht, nicht verändern.
**Aus einem Karnaugh-Diagramm konstruiert man die Verknüpfungen, indem man alle Felder mit einsen in möglichst grossen Gruppen zu** $2, 4, 8, 16, \ldots$ **Feldern zusammenfasst und für jede Gruppe die Konjunktion der unveränderlichen Variablen bildet. Dabei muss das Diagramm**

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

| $x_1$ | $x_2$ | $x_3$ | $x_4$ | y |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 1 |
| 0 | 0 | 0 | 1 | 1 |
| 0 | 0 | 1 | 0 | 1 |
| 0 | 0 | 1 | 1 | 1 |
| 0 | 1 | 0 | 0 | 1 |
| 0 | 1 | 0 | 1 | 0 |
| 0 | 1 | 1 | 0 | 0 |
| 0 | 1 | 1 | 1 | 0 |
| 1 | 0 | 0 | 0 | 1 |
| 1 | 0 | 0 | 1 | 0 |
| 1 | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 1 | 1 |
| 1 | 1 | 0 | 0 | 0 |
| 1 | 1 | 0 | 1 | 0 |
| 1 | 1 | 1 | 0 | 1 |
| 1 | 1 | 1 | 1 | 1 |

| $x_3x_4 \backslash x_1x_2$ | 00 | 01 | 11 | 10 |
|---|---|---|---|---|
| 00 | 1 | 1 | 0 | 1 |
| 01 | 1 | 0 | 0 | 0 |
| 11 | 1 | 0 | 1 | 1 |
| 10 | 1 | 0 | 1 | 1 |

Abbildung 2.34.: Erweitertes Beispiel zum Vergleich einer Wahrheitstabelle mit einem Karnaugh-Diagramm

**in jeder Richtung als periodisch Fortgesetzt betrachtet werden. Zum Schluss wird die entsprechende Disjunktion gebildet.** In Abbildung 2.34 ergeben sich also die Terme

$$K_A = \bar{x}_2 \wedge \bar{x}_4 \tag{2.4.99}$$
$$K_B = \bar{x}_1 \wedge \bar{x}_3 \wedge \bar{x}_4 \tag{2.4.100}$$
$$K_C = x_1 \wedge x_3 \tag{2.4.101}$$
$$K_D = \bar{x}_1 \wedge \bar{x}_2 \tag{2.4.102}$$

Das Schlussresultat ist dann

$$K = K_A \vee K_B \vee K_C \vee K_D \tag{2.4.103}$$
$$= (\bar{x}_2 \wedge \bar{x}_4) \vee (\bar{x}_1 \wedge \bar{x}_3 \wedge \bar{x}_4) \vee (x_1 \wedge x_3) \vee (\bar{x}_1 \wedge \bar{x}_2) \tag{2.4.104}$$

## 2.5. Vierpole und Vierpoltheorie

Ein Vierpol ist ein elektrisches Schaltteil (einfach oder zusammengesetzt), das von aussen mit vier Klemmen angesteuert wird[**?**]. Zwei der Klemmen dienen als Eingang, zwei als Ausgang. Wenn nun am Eingang eine Spannung angelegt wird, so fliest ein Strom, der aber auch von der Belastung am Ausgang abhängt. Genauso kann der Ausgang auf den Eingang rückwirken. Ebenso gibt es Kopplungen vom Eingang auf den Ausgang.
Die Vierpoltheorie beschreibt in einer linearen Näherung um den Arbeitspunkt die Wirkung einer Schaltung. Im Gegensatz zu der Anwendung von Blockschaltbildern wird hier die gegenseitige Beeinflussung von Schaltungen berücksichtigt.

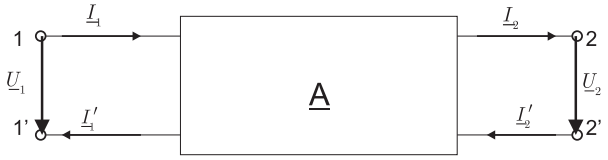



Abbildung 2.35.: Anschlüsse, Ströme und Spannungen bei einem Vierpol

Die Ströme an den Klemmen 1 und 1' sowie 2 und 2' sind jeweils gleich. Für lineare, zeitinvariante passive Vierpole gibt es sechs Möglichkeiten, die gegenseitigen Beeinflussungen in einem Gleichungssystem zu beschreiben. So könnte man zum Beispiel schreiben:

$$\underline{U}_1 = \underline{z}_{11}\underline{I}_1 + \underline{z}_{12}\underline{I}_2 \tag{2.5.1}$$
$$\underline{U}_2 = \underline{z}_{21}\underline{I}_1 + \underline{z}_{22}\underline{I}_2 \tag{2.5.2}$$

Die $\underline{z}_{ij}$ sind komplexwertige Koeffizienten, die wie folgt definiert sind:

$$\underline{z}_{11} = \frac{\partial \underline{U}_1}{\partial \underline{I}_1}\Big|_{\underline{I}_2=0} \quad \text{Leerlaufeingangsimpedanz} \tag{2.5.3}$$
$$\underline{z}_{12} = \frac{\partial \underline{U}_1}{\partial \underline{I}_2}\Big|_{\underline{I}_1=0} \quad \text{negativer Rückwirkungswiderstand} \tag{2.5.4}$$
$$\underline{z}_{21} = \frac{\partial \underline{U}_2}{\partial \underline{I}_1}\Big|_{\underline{I}_2=0} \quad \text{Kernwiderstand vorwärts} \tag{2.5.5}$$
$$\underline{z}_{22} = \frac{\partial \underline{U}_2}{\partial \underline{I}_2}\Big|_{\underline{I}_1=0} \quad \text{negative Leerlaufausgangsimpedanz} \tag{2.5.6}$$

Die obigen Gleichungen geben auch die Messvorschrift für diese Impedanzen wieder. Um $\underline{z}_{11}$ zu bestimmen, speist man bei offenem Ausgang den Strom $\underline{I}_1$ ein und misst die resultierende Spannung $\underline{U}_1$. Die Gleichungen können kompakt als Matrix geschrieben werden, eine Tatsache die die Rechenarbeit sehr erleichtert.

$$\begin{pmatrix} \underline{U}_1 \\ \underline{U}_2 \end{pmatrix} = \begin{pmatrix} \underline{z}_{11} & \underline{z}_{12} \\ \underline{z}_{21} & \underline{z}_{22} \end{pmatrix} \begin{pmatrix} \underline{I}_1 \\ \underline{I}_2 \end{pmatrix} = \underline{\mathbf{Z}} \begin{pmatrix} \underline{I}_1 \\ \underline{I}_2 \end{pmatrix} \tag{2.5.7}$$

Die Matrix $\underline{\mathbf{Z}}$ heisst die Widerstandsmatrix. Durch Permutation können die anderen möglichen Darstellungen erhalten werden. Üblich sind:

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

**Widerstandsmatrix**

$$\begin{pmatrix} \underline{U}_1 \\ \underline{U}_2 \end{pmatrix} = \begin{pmatrix} \underline{z}_{11} & \underline{z}_{12} \\ \underline{z}_{21} & \underline{z}_{22} \end{pmatrix} \begin{pmatrix} \underline{I}_1 \\ \underline{I}_2 \end{pmatrix} = \underline{\mathbf{Z}} \begin{pmatrix} \underline{I}_1 \\ \underline{I}_2 \end{pmatrix} \tag{2.5.8}$$

**Leitwertform**

$$\begin{pmatrix} \underline{I}_1 \\ \underline{I}_2 \end{pmatrix} = \begin{pmatrix} \underline{y}_{11} & \underline{y}_{12} \\ \underline{y}_{21} & \underline{y}_{22} \end{pmatrix} \begin{pmatrix} \underline{U}_1 \\ \underline{U}_2 \end{pmatrix} = \underline{\mathbf{Y}} \begin{pmatrix} \underline{U}_1 \\ \underline{U}_2 \end{pmatrix} \tag{2.5.9}$$

**Kettenform**

$$\begin{pmatrix} \underline{U}_1 \\ \underline{I}_1 \end{pmatrix} = \begin{pmatrix} \underline{a}_{11} & \underline{a}_{12} \\ \underline{a}_{21} & \underline{a}_{22} \end{pmatrix} \begin{pmatrix} \underline{U}_2 \\ \underline{I}_2 \end{pmatrix} = \underline{\mathbf{A}} \begin{pmatrix} \underline{U}_2 \\ \underline{I}_2 \end{pmatrix} \tag{2.5.10}$$

**Hybridform (Reihen-Parallel-Form)**

$$\begin{pmatrix} \underline{U}_1 \\ \underline{I}_2 \end{pmatrix} = \begin{pmatrix} \underline{h}_{11} & \underline{h}_{12} \\ \underline{h}_{21} & \underline{h}_{22} \end{pmatrix} \begin{pmatrix} \underline{I}_1 \\ \underline{U}_2 \end{pmatrix} = \underline{\mathbf{H}} \begin{pmatrix} \underline{I}_1 \\ \underline{U}_2 \end{pmatrix} \tag{2.5.11}$$

Die Matrix $\underline{\mathbf{H}}$ ist besonders beliebt zur Angabe der Vierpolparameter von Transistoren. Bei Transistoren, inherent nichtlinearen Bauteilen, werden die Vierpolparameter am Arbeitspunkt angegeben, es sind also differentielle Parameter. Auch gebräuchlich für Transistoren ist die $\underline{\mathbf{Y}}$-Matrix. Die Vierpolparameter können wie in Tabelle 2.13 angegeben ineinander umgerechnet werden.

## 2.5.1. Zusammenschaltung von Vierpolen

Die Vierpoltheorie erlaubt, das Zusammenschalten einzelner Bauelemente unter Berücksichtigung von Eingangs- und Ausgangswiderständen einfach zu berechnen. Kabel und Leitungen können mit Ketten von Vierpolen modelliert werden.

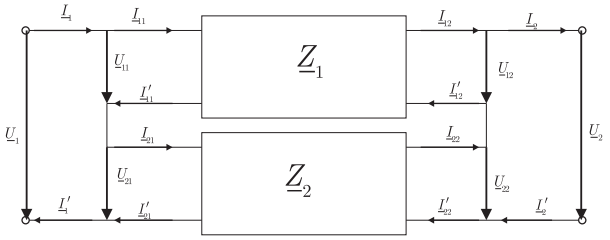



Abbildung 2.36.: Serienschaltung zweier Vierpole

Die Serienschaltung in Abbildung 2.36 kann mit folgenden Bedingungsgleichungen berechnet werden:

$$\begin{aligned} \underline{U}_{11} + \underline{U}_{21} &= \underline{U}_1 \\ \underline{U}_{12} + \underline{U}_{22} &= \underline{U}_2 \\ \underline{I}_{11} = \underline{I}'_{21} &= \underline{I}_1 \\ \underline{I}_{12} = \underline{I}'_{22} &= \underline{I}_2 \end{aligned} \tag{2.5.12}$$

| | $\underline{\mathbf{A}}$ | $\underline{\mathbf{Z}}$ | $\underline{\mathbf{Y}}$ | $\underline{\mathbf{H}}$ |
|---|---|---|---|---|
| $\underline{\mathbf{A}}$ | $\begin{matrix}\underline{a}_{11} & \underline{a}_{12}\\ \underline{a}_{21} & \underline{a}_{22}\end{matrix}$ | $\begin{matrix}\frac{\underline{z}_{11}}{\underline{z}_{21}} & -\frac{\Delta\underline{z}}{\underline{z}_{21}}\\ \frac{1}{\underline{z}_{21}} & -\frac{\underline{z}_{22}}{\underline{z}_{21}}\end{matrix}$ | $\begin{matrix}-\frac{\underline{y}_{22}}{\underline{y}_{21}} & \frac{1}{\underline{y}_{21}}\\ -\frac{\Delta\underline{y}}{\underline{y}_{21}} & \frac{\underline{y}_{11}}{\underline{y}_{21}}\end{matrix}$ | $\begin{matrix}-\frac{\Delta\underline{h}}{\underline{h}_{21}} & \frac{\underline{h}_{11}}{\underline{h}_{21}}\\ -\frac{\underline{h}_{22}}{\underline{h}_{21}} & \frac{1}{\underline{h}_{21}}\end{matrix}$ |
| $\underline{\mathbf{Z}}$ | $\begin{matrix}\frac{\underline{a}_{11}}{\underline{a}_{21}} & -\frac{\Delta\underline{a}}{\underline{a}_{21}}\\ \frac{1}{\underline{a}_{21}} & -\frac{\underline{a}_{22}}{\underline{a}_{21}}\end{matrix}$ | $\begin{matrix}\underline{z}_{11} & \underline{z}_{12}\\ \underline{z}_{21} & \underline{z}_{22}\end{matrix}$ | $\begin{matrix}\frac{\underline{y}_{22}}{\Delta\underline{y}} & -\frac{\underline{y}_{12}}{\Delta\underline{y}}\\ -\frac{\underline{y}_{21}}{\Delta\underline{y}} & \frac{\underline{y}_{11}}{\Delta\underline{y}}\end{matrix}$ | $\begin{matrix}\frac{\Delta\underline{h}}{\underline{h}_{22}} & \frac{\underline{h}_{12}}{\underline{h}_{22}}\\ -\frac{\underline{h}_{21}}{\underline{h}_{22}} & \frac{1}{\underline{h}_{22}}\end{matrix}$ |
| $\underline{\mathbf{Y}}$ | $\begin{matrix}\frac{\underline{a}_{22}}{\underline{a}_{12}} & -\frac{\Delta\underline{a}}{\underline{a}_{12}}\\ \frac{1}{\underline{a}_{12}} & -\frac{\underline{a}_{11}}{\underline{a}_{12}}\end{matrix}$ | $\begin{matrix}\frac{\underline{z}_{22}}{\Delta\underline{z}} & -\frac{\underline{z}_{12}}{\Delta\underline{z}}\\ -\frac{\underline{z}_{21}}{\Delta\underline{z}} & \frac{\underline{z}_{11}}{\Delta\underline{z}}\end{matrix}$ | $\begin{matrix}\underline{y}_{11} & \underline{y}_{12}\\ \underline{y}_{21} & \underline{y}_{22}\end{matrix}$ | $\begin{matrix}\frac{1}{\underline{h}_{11}} & -\frac{\underline{h}_{12}}{\underline{h}_{11}}\\ \frac{\underline{h}_{21}}{\underline{h}_{11}} & \frac{\Delta\underline{h}}{\underline{h}_{11}}\end{matrix}$ |
| $\underline{\mathbf{H}}$ | $\begin{matrix}\frac{\underline{a}_{12}}{\underline{a}_{22}} & \frac{\Delta\underline{a}}{\underline{a}_{22}}\\ \frac{1}{\underline{a}_{22}} & -\frac{\underline{a}_{21}}{\underline{a}_{22}}\end{matrix}$ | $\begin{matrix}\frac{\Delta\underline{z}}{\underline{z}_{22}} & \frac{\underline{z}_{12}}{\underline{z}_{22}}\\ -\frac{\underline{z}_{21}}{\underline{z}_{22}} & \frac{1}{\underline{z}_{22}}\end{matrix}$ | $\begin{matrix}\frac{1}{\underline{y}_{11}} & -\frac{\underline{y}_{12}}{\underline{y}_{11}}\\ \frac{\underline{y}_{21}}{\underline{y}_{11}} & \frac{\Delta\underline{y}}{\underline{y}_{11}}\end{matrix}$ | $\begin{matrix}\underline{h}_{11} & \underline{h}_{12}\\ \underline{h}_{21} & \underline{h}_{22}\end{matrix}$ |
| $\Delta\underline{a}$ | $\underline{a}_{11}\underline{a}_{22} - \underline{a}_{12}\underline{a}_{21}$ | $-\frac{\underline{z}_{12}}{\underline{z}_{21}}$ | $-\frac{\underline{y}_{12}}{\underline{y}_{21}}$ | $\frac{\underline{h}_{12}}{\underline{h}_{21}}$ |
| $\Delta\underline{z}$ | $-\frac{\underline{a}_{12}}{\underline{a}_{21}}$ | $\underline{z}_{11}\underline{z}_{22} - \underline{z}_{12}\underline{z}_{21}$ | $\frac{1}{\Delta\underline{y}}$ | $-\frac{\underline{h}_{11}}{\underline{h}_{22}}$ |
| $\Delta\underline{y}$ | $-\frac{\underline{a}_{21}}{\underline{a}_{12}}$ | $\frac{1}{\Delta\underline{z}}$ | $\underline{y}_{11}\underline{y}_{22} - \underline{y}_{12}\underline{y}_{21}$ | $\frac{\underline{h}_{22}}{\underline{h}_{11}}$ |
| $\Delta\underline{h}$ | $-\frac{\underline{a}_{11}}{\underline{a}_{22}}$ | $\frac{\underline{z}_{11}}{\underline{z}_{22}}$ | $\frac{\underline{y}_{22}}{\underline{y}_{11}}$ | $\underline{h}_{11}\underline{h}_{22} - \underline{h}_{12}\underline{h}_{21}$ |

Tabelle 2.13.: Umrechnung der Vierpolparameter

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Aus Gleichungen (2.5.8) und (2.5.12) kann die Matrix-Form der Serieschaltung berechnet werden:

$$\begin{aligned}\begin{pmatrix} \underline{U}_1 \\ \underline{U}_2 \end{pmatrix} &= \begin{pmatrix} \underline{z}_{111}+\underline{z}_{211} & \underline{z}_{112}+\underline{z}_{212} \\ \underline{z}_{121}+\underline{z}_{221} & \underline{z}_{122}+\underline{z}_{222} \end{pmatrix} \begin{pmatrix} \underline{I}_1 \\ \underline{I}_2 \end{pmatrix} \\ &= (\underline{Z}_1+\underline{Z}_2) \begin{pmatrix} \underline{I}_1 \\ \underline{I}_2 \end{pmatrix} \end{aligned} \tag{2.5.13}$$

Die Notation $\underline{z}_{abc}$ bedeutet, dass das Element $\underline{z}_{bc}$ aus der Matrix $\underline{Z}_a$ gemeint ist. Die Matrizen der einzelnen Vierpole addieren sich also bei einer Serienschaltung.

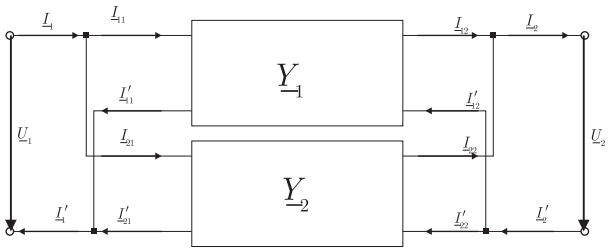



Abbildung 2.37.: Parallelschaltung zweier Vierpole

Bei der Parallelschaltung findet man analog:

$$\begin{aligned}\begin{pmatrix} \underline{I}_1 \\ \underline{I}_2 \end{pmatrix} &= \begin{pmatrix} \underline{y}_{111}+\underline{y}_{211} & \underline{y}_{112}+\underline{y}_{212} \\ \underline{y}_{121}+\underline{y}_{221} & \underline{y}_{122}+\underline{y}_{222} \end{pmatrix} \begin{pmatrix} \underline{U}_1 \\ \underline{U}_2 \end{pmatrix} \\ &= (\underline{Y}_1+\underline{Y}_2) \begin{pmatrix} \underline{U}_1 \\ \underline{U}_2 \end{pmatrix} \end{aligned} \tag{2.5.14}$$

Man kann sich die Regeln für die Parallelschaltung von Vierpolen einfach merken: Wie bei Widerständen addieren sich bei einer Parallelschaltung die Leitwerte.

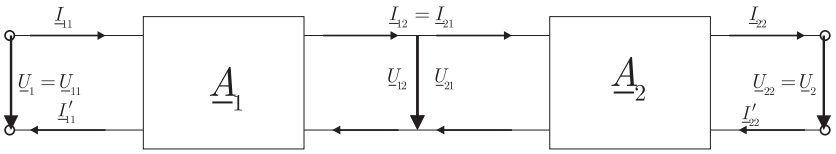



Abbildung 2.38.: Kettenschaltung zweier Vierpole

Bei der Kettenschaltung gilt:

$$\begin{aligned} \underline{U}_1 &= \underline{U}_{11} \\ \underline{U}_{12} &= \underline{U}_{21} \\ \underline{U}_{22} &= \underline{U}_2 \\ \underline{I}_1 &= \underline{I}_{11} \\ \underline{I}_{12} &= \underline{I}_{12} \\ \underline{I}_{22} &= \underline{I}_2 \end{aligned} \tag{2.5.15}$$

Unter Verwendung der Gleichungen (2.5.10)für die Kettenform erhält man

$$\begin{pmatrix} \underline{U}_1 \\ \underline{I}_1 \end{pmatrix} = \begin{pmatrix} \underline{a}_{111}\underline{a}_{211} + \underline{a}_{112}\underline{a}_{221} & \underline{a}_{111}\underline{a}_{212} + \underline{a}_{112}\underline{a}_{222} \\ \underline{a}_{121}\underline{a}_{211} + \underline{a}_{122}\underline{a}_{221} & \underline{a}_{121}\underline{a}_{212} + \underline{a}_{122}\underline{a}_{222} \end{pmatrix} \begin{pmatrix} \underline{U}_2 \\ \underline{I}_2 \end{pmatrix}$$
$$\begin{pmatrix} \underline{U}_1 \\ \underline{I}_1 \end{pmatrix} = \underline{A}_1 \cdot \underline{A}_2 \begin{pmatrix} \underline{U}_2 \\ \underline{I}_2 \end{pmatrix} \tag{2.5.16}$$

Wie bei jeder Matrixmultiplikation ist die Kettenschaltung von der Reihenfolge abhängig. Physikalisch kann man sich das wie folgt klar machen: Der Eingang des zweiten Vierpols belastet den Ausgang des ersten, während sein Ausgang unbelastet ist. Ebenso wir der Eingang des ersten von einer idealen Quelle angesteuert. Wechselt man nun die Reihenfolge, so sind die jeweiligen Ein- und Ausgänge nicht mehr gleich belastet. Entsprechend muss aus physikalischer Sicht das Resultat von der Reihenfolge der Vierpole abhängen.

## 2.5.2. Übertragungsfunktion eines Vierpols

Vielfach möchte man die Spannungs- oder Stromverstärkung eines mit der Lastimpedanz $\underline{Z}_L$ belasteten Vierpols wissen (Abbildung 2.39 . Die Lastimpedanz kann komplex sein, wir behandeln so auch die Frage nach kapazitiv belasteten Ausgängen.

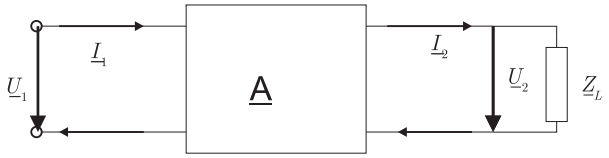



Abbildung 2.39.: Übertragungsfunktion eines Vierpols

Ausgangsstrom $\underline{I}_2$ und Ausgangsspannung $\underline{U}_2$ hängen dann wie folgt zusammen:

$$\underline{U}_2 = \underline{Z}_L \underline{I}_2 \tag{2.5.17}$$

Mit der Kettengleichung (2.5.10)wird

$$\begin{aligned} \underline{U}_1 &= \left(\underline{a}_{11} + \frac{\underline{a}_{12}}{\underline{Z}_L}\right) \underline{U}_2 \\ \underline{I}_1 &= \left(\underline{a}_{21}\underline{Z}_L + \underline{a}_{22}\right) \underline{I}_2 \end{aligned} \tag{2.5.18}$$

Damit ergibt sich für die Übertragungsfunktion der Spannung

$$\frac{\underline{U}_2}{\underline{U}_1} = \underline{g}_U = \frac{1}{\underline{a}_{11} + \frac{\underline{a}_{12}}{\underline{Z}_L}} \tag{2.5.19}$$

und des Stromes

$$\frac{\underline{I}_2}{\underline{I}_1} = \underline{g}_I = \frac{1}{\underline{a}_{21}\underline{Z}_L + \underline{a}_{12}} \tag{2.5.20}$$

Der Leistungsübertragungsfaktor ist

$$\underline{g}_P = \underline{g}_U \cdot \underline{g}_I \tag{2.5.21}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Die Eingangsimpedanz ist

$$\underline{Z}_I = \frac{\underline{U}_1}{\underline{I}_1} = \underline{Z}_L \cdot \frac{\underline{g}_I}{\underline{g}_U} = \frac{\underline{a}_{11}\underline{Z}_L + \underline{a}_{12}}{\underline{a}_{21}\underline{Z}_L + \underline{a}_{22}} \tag{2.5.22}$$

Weiter sind die Übertragungsimpedanz

$$\frac{\underline{U}_2}{\underline{I}_1} = \frac{\underline{Z}_L}{\underline{a}_{21}\underline{Z}_L + \underline{a}_{22}} \tag{2.5.23}$$

und die Übertragungsadmittanz

$$\frac{\underline{I}_2}{\underline{U}_1} = \frac{1}{\underline{a}_{11}\underline{Z}_L + \underline{a}_{12}} \tag{2.5.24}$$

Die Eingangsimpedanz $\underline{Z}_I$ hängt nach Gleichung (2.5.22) von der Ausgangsimpedanz $\underline{Z}_L$ ab. Sie kann Werte zwischen

$$\underline{Z}_I|_{\underline{Z}_L=\infty} = \frac{\underline{a}_{11}}{\underline{a}_{21}} \quad \text{Leerlaufeingangsimpedanz} \tag{2.5.25}$$

$$\underline{Z}_I|_{\underline{Z}_L=0} = \frac{\underline{a}_{12}}{\underline{a}_{22}} \quad \text{Kurzschlusseingangsimpedanz} \tag{2.5.26}$$

Analog erhält man für die Ausgangsimpedanz $\underline{Z}_A$ abhängig von der Quellimpedanz $\underline{Z}_Q$

$$\underline{Z}_A|_{\underline{Z}_Q=\infty} = \frac{\underline{a}_{22}}{\underline{a}_{21}} \quad \text{Leerlaufausgangsimpedanz} \tag{2.5.27}$$

$$\underline{Z}_A|_{\underline{Z}_Q=0} = \frac{\underline{a}_{12}}{\underline{a}_{11}} \quad \text{Kurzschlussausgangsimpedanz} \tag{2.5.28}$$

Der Wellenwiderstand des Eingangs $\underline{Z}_{01}$ oder Ausgangs $\underline{Z}_{02}$ ist das geometrische Mittel aus den entsprechenden Kurzschluss- und Leerlaufimpedanzen.

$$\underline{Z}_{01} = \sqrt{\underline{Z}_I|_{\underline{Z}_L=\infty}\,\underline{Z}_I|_{\underline{Z}_L=0}} = \sqrt{\frac{\underline{a}_{12}\underline{a}_{11}}{\underline{a}_{21}\underline{a}_{22}}} \quad \text{Eingangswellenwiderstand} \tag{2.5.29}$$

$$\underline{Z}_{02} = \sqrt{\underline{Z}_A|_{\underline{Z}_Q=\infty}\,\underline{Z}_A|_{\underline{Z}_Q=0}} = \sqrt{\frac{\underline{a}_{12}\underline{a}_{22}}{\underline{a}_{21}\underline{a}_{11}}} \quad \text{Ausgangswellenwiderstand} \tag{2.5.30}$$

Der Wellenwiderstand ist gerade der Abschlusswiderstand, für den der Vierpol angepasst ist. Ein mit $\underline{Z}_{02}$ am Ausgang abgeschlossener Vierpol hat gerade die Eingangsimpedanz $\underline{Z}_{01}$. Im Anpassungsfall, d.h. wenn die Impedanz der Quelle $\underline{Z}_Q = \underline{Z}_{01}$ ist und wenn der Lastwiderstand $\underline{Z}_L = \underline{Z}_{02}$ ist, hat man **Leistungsanpassung**
Die Wellenwiderstände lassen sich durch die Messung von Kurzschluss- und Leerlaufimpedanzen bestimmen. Diese Eigenschaft wird verwendet, um mit Netzwerkanalysatoren komplexe Hochfrequenzleiter oder Bauelemente auszumessen.
Besonders einfach ist die Bestimmung der Wellenwiderstände bei symmetrischen Vierpolen mit $\underline{a}_{11} = \underline{a}_{22}$. Dann ist

$$\underline{Z}_{01} = \underline{Z}_{02} = \sqrt{\frac{\underline{a}_{12}}{\underline{a}_{21}}} \tag{2.5.31}$$

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

### 2.5.3. Ersatzstrukturen für Vierpole

Für passive Vierpole ($\delta\underline{a} = \underline{a}_{11}\underline{a}_{22} - \underline{a}_{12}\underline{a}_{21} = 1$) können die Kettenparameter $\underline{a}_{ij}$ durch die Ein- und Ausgangsimpedanzen bestimmt werden (**Messrezept**).

$$\begin{aligned}
\underline{a}_{11} &= \sqrt{\frac{\underline{Z}_I|_{\underline{Z}_L=\infty}}{\underline{Z}_A|_{\underline{Z}_Q=\infty} - \underline{Z}_A|_{\underline{Z}_Q=0}}} \\
\underline{a}_{22} &= \sqrt{\frac{\underline{Z}_A|_{\underline{Z}_Q=\infty}}{\underline{Z}_I|_{\underline{Z}_L=\infty} - \underline{Z}_I|_{\underline{Z}_L=0}}} \\
\underline{a}_{21} &= \sqrt{\frac{1}{\underline{Z}_I|_{\underline{Z}_L=\infty}\left(\underline{Z}_A|_{\underline{Z}_Q=\infty} - \underline{Z}_A|_{\underline{Z}_Q=0}\right)}} = \sqrt{\frac{1}{\underline{Z}_A|_{\underline{Z}_Q=\infty}\left(\underline{Z}_I|_{\underline{Z}_L=\infty} - \underline{Z}_I|_{\underline{Z}_L=0}\right)}} \\
\underline{a}_{12} &= \underline{Z}_I|_{\underline{Z}_L=0}\sqrt{\frac{\underline{Z}_A|_{\underline{Z}_Q=\infty}}{\underline{Z}_I|_{\underline{Z}_L=\infty} - \underline{Z}_I|_{\underline{Z}_L=0}}} = \underline{Z}_A|_{\underline{Z}_Q=0}\sqrt{\frac{\underline{Z}_I|_{\underline{Z}_L=\infty}}{\underline{Z}_A|_{\underline{Z}_Q=\infty} - \underline{Z}_A|_{\underline{Z}_Q=0}}}
\end{aligned} \tag{2.5.32}$$

Das Übertragungsverhalten eines Vierpols lässt sich nun mit Ersatzschaltungen modellieren.

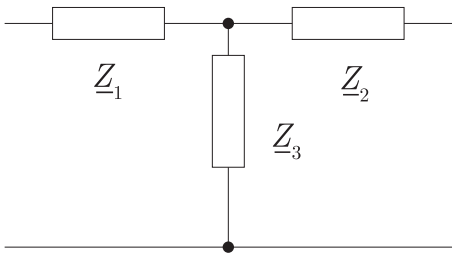



Abbildung 2.40.: Ersatzschaltung eines Vierpols: T- Glied (Sternschaltung)

Man erhält zum Beispiel für die Sternschaltung in Abbildung 2.40 folgende Beziehungen

$$\underline{Z}_I|_{\underline{Z}_L=\infty} = \tilde{\underline{Z}}_1 + \tilde{\underline{Z}}_3 \tag{2.5.33}$$

$$\underline{Z}_I|_{\underline{Z}_L=0} = \tilde{\underline{Z}}_1 + \frac{\tilde{\underline{Z}}_2\tilde{\underline{Z}}_3}{\tilde{\underline{Z}}_2 + \tilde{\underline{Z}}_3} \tag{2.5.34}$$

$$\underline{Z}_A|_{\underline{Z}_Q=\infty} = \tilde{\underline{Z}}_2 + \tilde{\underline{Z}}_3 \tag{2.5.35}$$

$$\underline{Z}_A|_{\underline{Z}_Q=0} = \tilde{\underline{Z}}_2 + \frac{\tilde{\underline{Z}}_1\tilde{\underline{Z}}_3}{\tilde{\underline{Z}}_1 + \tilde{\underline{Z}}_3} \tag{2.5.36}$$

Weitere mögliche Ersatzschaltbilder sind in den Abbildungen 2.41 und 2.42 dargestellt.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

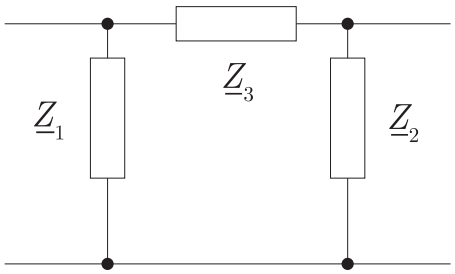



Abbildung 2.41.: Ersatzschaltung eines Vierpols: $\pi$- Glied (Dreiecksschaltung)

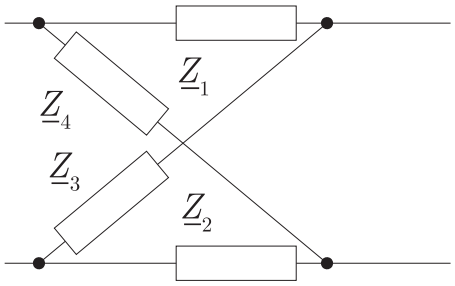



Abbildung 2.42.: Ersatzschaltung eines Vierpols: Kreuzglied

## 2.6. Filter

Filterschaltungen sind Baugruppen, die das Frequenzspektrum eines Signals verändern[**?**]. Sie werden verwendet, um zum Beispiel Rauschen zu entfernen, bei drahtloser Übertragung die Trägerfrequenz zu unterdrücken oder um gewisse Komponenten eines Signals zu bevorzugen. Die Filtereigenschaften lassen sich am einfachsten mit analogen Filtern erklären. Darauf aufbauend werden digitale Implementierungen von Filtern besprochen. Die dort erarbeiteten Konzepte können auch bei der Bearbeitung von Datensätzen im Computer verwendet werden.

### 2.6.1. Analogfilter

Als einfachstes Beispiel eines Filters betrachten wir einen RC Tiefpass. Die Filtercharakteristik geschrieben mit Frequenzen ist

$$\underline{A}(j\omega) = \frac{\underline{U}_a}{\underline{U}_e} = \frac{1}{1 + j\omega RC} \tag{2.6.1}$$

Es ist üblich, die Gleichung (2.6.1) so umzurechnen, dass die Frequenz $\Omega$, bei der $\frac{\underline{U}_a}{\underline{U}_e} = 2^{-1/2}$ ist, gleich 1 gesetzt wird. Mit $\Omega = \frac{\omega}{\omega_0}$ ($\omega_0 = \frac{1}{RC}$ ist die natürlich Grenzfrequenz unseres Tiefpasses. Gleichung (2.6.1) wird dann

$$\underline{A}(j\Omega) = \frac{1}{1 + j\Omega} \tag{2.6.2}$$

Nun haben wir im Abschnitt 2.4.3 über die Laplace-Transformation gesehen, dass diese Kausalität erzwingt. Indem $\omega$ mit $p$ identifiziert wird (man kann auch sagen, dass die allgemein komplexe Variable $p$ nur entlang der imaginären Achse betrachtet wird), erhält man die Übertragungsfunktion

$$\underline{A}(p) = \frac{L(\underline{U}_a)}{L(\underline{U}_e)} = \frac{1}{1 + pRC} \tag{2.6.3}$$

und daraus mit $P = \frac{p}{\omega_0}$ die normierte Darstellung der Übertragungsfunktion

$$\underline{A}(P) = \frac{1}{1 + P} \tag{2.6.4}$$

Die Darstellung in Gleichung (2.6.4) ist die einfachste denkbare Darstellung eines Tiefpassfilters. Die Grenzfrequenz ist normiert, alle Details der Realisierung werden mit der Normierung kaschiert. Das Vorgehen ist in gewissem Sinne analog zu dem der theoretischen Physiker, wenn sie $\hbar = 1$ und $c = 1$ setzen.
Für Sinuswellen ist das Verhältnis zwischen Ausgangs- und Eingangssignal ist

$$|\underline{A}(j\Omega)|^2 = \frac{1}{1 + \Omega^2} \tag{2.6.5}$$

Für grosse Frequenzen $\Omega \gg 1$ verhält sich die Ausgangsamplitude $|\underline{A}| = 1/\Omega$. Dies entspricht einem Verstärkungsabfall von 6 dB pro Oktave (Faktor 2) oder 20 dB pro Dekade (Faktor 10). Der Verstärkungsabfall pro Dekade ist charakteristisch für die Filterordnung. Pro Ordnung erhält man 20 dB pro Dekade.
Für einen steileren Abfall der Verstärkung kann man n Filter hintereinander schalten. Wenn man annimmt, dass jeder Teilfilter vom vorhergehenden entkoppelt ist

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

(keine Rückwirkung) und wenn man weiter annimmt, dass jeder Teilfilter eine andere Grenzfrequenz haben kann, charakterisiert durch den Faktor $\alpha_i$ dann ist die Übertragungsfunktion

$$A(P) = \frac{1}{(1+\alpha_1 P)\,(1+\alpha_2 P)\ldots(1+\alpha_n P)} \tag{2.6.6}$$

Die Koeffizienten $\alpha_i$ sind reell und positiv.
Für grosse Frequenzen gilt $|\underline{A}(j\Omega)| \propto \Omega^{-n}$. Der Abfall ist also **n** mal 20 dB pro Dekade.
Bei n gleichen, entkoppelten Tiefpässen ist die 3-dB Grenzfrequenz $\Omega = 1$, wenn gilt:

$$\alpha_i = \alpha = \sqrt{\sqrt[n]{2} - 1} \quad (i = 1, 2, \ldots, n) \tag{2.6.7}$$

Die Grenzfrequenz eines einzelnen Tiefpasses ist um $1/\alpha$ höher als die Grenzfrequenz der Gesamtschaltung. Diese Eigenheit ist bei allen zusammengesetzten Teifpässen zu bemerken. Die oben eingeführten Tiefpässe haben nur reelle Pole. Sie heissen kritische Tiefpässe. Tabelle F.1 im Anhang gibt eine Übersicht über die Filterkoeffizienten. Die Koeffizienten sind in Gruppen zu zwei geordnet, da man jedes Polynom mit reellen Koeffizienten in ein Produkt von Polynomen 2. Grades aufspalten kann.
Allgemein ist eine Filterfunktion gegeben durch

$$A(P) = \frac{A_0}{1 + c_1 P + c_2 P^2 + \ldots + c_n P^n} \tag{2.6.8}$$

Hier ist $A_0$ die Verstärkung bei $\Omega = 0$. Für beliebige reelle Koeffizienten $c_i$ kann das Nennerpolynom in Gleichung (2.6.8) in $n/2$ ($(n-1)/2$ bei ungeradem n) Polynome 2. Grades (und ein Polynom ersten Grades bei ungeradem n) aufgespalten werden. Diese Polynome zweiten Grades haben entweder zwei reelle Nullstellen, oder aber zwei konjugiert komplexe. Wir können also schreiben:

$$A(P) = \frac{1}{(1 + a_1 P + b_1 P^2)\,(1 + a_2 P + b_2 P^2)\ldots} \tag{2.6.9}$$

Wir vereinbaren, dass bei ungeradem n $b_1 = 0$ sein soll. Für unser kritisch gedämpftes Filter gilt nun:

$$a_i = 2\alpha \qquad b_i = \alpha^2 \tag{2.6.10}$$

Diese Koeffizienten sind in Tabelle F.1 aufgelistet.
Konjugiert komplexe Pole, wie sie in einem Filter höherer Ordnung auftreten können, sind nicht mit einfachen RC-Filtern realisierbar. Entweder man verwendet auch Spulen, also RLC-kreise, oder man benötigt aktive Schaltungen, wie sie im Kapitel 3 besprochen werden.
Es gibt nun verschiedene Optimierungsstrategien für die Filterkoeffizienten. Sie werden in den folgenden Abschnitten nun besprochen.

#### 2.6.1.1. Butterworth-Tiefpässe

Das Betragsquadrat der Verstärkung eines allgemeinen Tiefpasses hat die Form

$$|\underline{A}|^2 = \frac{A_0^2}{1 + d_2\Omega^2 + \ldots + d_{2n}\Omega^{2n}} \tag{2.6.11}$$

Wir fordern nun, dass die Verstärkung möglichst lange gleich $A_0$ sein soll. Das bedeutet, dass der Nenner möglichst lange in der Nähe von 1 sein muss. Dies heisst, dass die verschiedenen Potenzen von $\Omega$ so lange wie möglich klein gegen 1 sein müssen. Im Intervall $[0 \ldots 1]$ ist dies am besten für die höchste Potenz erfüllt. wir setzen also

$$|\underline{A}|^2 = \frac{A_0^2}{1 + d_{2n}\Omega^{2n}} \tag{2.6.12}$$

Den Koeffizienten $d_{2n}$ kann man aus der Normierungsbedingung

$$\frac{A_0^2}{2} = \frac{A_0^2}{1 + d_{2n}} \tag{2.6.13}$$

bestimmen. Wir erhalten so dass $d_{2i} = 0; \quad (i = 1, 2, \ldots, n-1)$ und $d_{2n} = 1$ ist. Der Nenner der Bestimmungsgleichung ist nun $\sqrt{1 + P^{2n}}$. Die daraus resultierenden Butterworth-Polynome sind in der Tabelle 2.14 zusammengefasst.

| Ordnung n | Polynom |
|---|---|
| 1 | $1 + P$ |
| 2 | $1 + \sqrt{2}P + P^2$ |
| 3 | $1 + 2P + 2P^2 + P^3 = (1 + P)(1 + P + P^2)$ |
| 4 | $1 + 2.613P + 3.414P^2 + 2.613P^3 + P^4 = (1 + 1.848P + P^2)(1 + 0.765P + P^2)$ |

Tabelle 2.14.: Butterworth-Polynome

### 2.6.1.2. Tschebyscheff-Tiefpässe

Der Übergang zwischen dem Durchlassbereich und dem Sperrbereich eines Tiefpasses kann man erhöhen, wenn man im Durchlassbereich eine gewisse Welligkeit zulässt. Übliche Polynome haben im Intervall $[0 \ldots 1]$ eine variable Welligkeit. Tschbyscheff-Polynome haben eine konstante Welligkeit, was man sehr leicht anhand der Definitionsgleichung ersehen kann.

$$T_n(x) = \begin{cases} \cos(n \arccos x), & \text{für } 0 \le x \le 1; \\ \cosh(n\, Arcosh\, x), & \text{für } x > 1. \end{cases} \tag{2.6.14}$$

Die ersten der resultierenden Polynome sind in Tabelle 2.15 angegeben.

| Ordnung n | Polynom |
|---|---|
| 1 | $T_1(x) = x$ |
| 2 | $T_2(x) = 2x^2 - 1$ |
| 3 | $T_3(x) = 4x^3 - 3x$ |
| 4 | $T_4(x) = 8x^4 - 8x^2 + 1$ |

Tabelle 2.15.: Tschebyscheff-Polynome

Die Tiefpassgleichung ist nun

$$|\underline{A}|^2 = \frac{dA_0^2}{1 + \epsilon^2 T_n^2(P)} \tag{2.6.15}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Die Normierungskonstanten $d$ und $\epsilon$ werden so gewählt, dass für $P = 0$ die Verstärkung $|A|^2 = A_0^2$ ist. Ein Vergleich mit Tabelle 2.15 zeigt, dass für ungerades $n$ $d = 1$ ist und für gerades $n$ $d = 1 + \epsilon^2$ ist. Dabei ist $\epsilon$ ein Mass für die Welligkeit im Durchlassbereich. Es gelten nun für die Welligkeit, die minimale und maximale Amplitude:

$$\begin{aligned} \frac{A_{max}}{A_{min}} &= \sqrt{1+\epsilon^2} \\ \left.\begin{aligned} A_{max} &= A_0\sqrt{1+\epsilon^2} \\ A_{min} &= A_0 \end{aligned}\right\} &\text{ bei gerader Ordnung} \\ \left.\begin{aligned} A_{max} &= A_0 \\ A_{min} &= \frac{A_0}{\sqrt{1+\epsilon^2}} \end{aligned}\right\} &\text{bei ungerader Ordnung} \end{aligned} \tag{2.6.16}$$

Die die Koeffizienten $d$ und $\epsilon$ werden in der Tabelle 2.16 für verschiedene Welligkeiten verglichen.

| Welligkeit | 0.5dB | 1dB | 2dB | 3dB |
|---|---|---|---|---|
| $A_{max}/A_{min}$ | 1.059 | 1.122 | 1.259 | 1.413 |
| $d$ | 1.112 | 1.259 | 1.585 | 1.995 |
| $\epsilon$ | 0.349 | 0.509 | 0.765 | 0.998 |

Tabelle 2.16.: Koeffizienten der Tschebyscheff-Polynome für verschiedene Welligkeiten

Die Koeffizienten der Tschebyscheff-Filter können nach Weinberg[**?**] mit den folgenden Gleichungen berechnet werden:

$$\left.\begin{aligned} cb_i' &= \frac{1}{\cosh^2\gamma - \cos^2\frac{(2i-1)\Pi}{2n}} \\ a_i' &= 2b_i' \sinh\gamma \cos\frac{(2i-1)\Pi}{2n} \end{aligned}\right\} \quad \text{für } \left(i = 1 \ldots \frac{n}{2}\right)\ n \text{ gerade} \tag{2.6.17}$$

$$\begin{aligned} b_1' &= 0 \\ a_1' &= \frac{1}{\sinh\gamma} \\ \left.\begin{aligned} cb_i' &= \frac{1}{\cosh^2\gamma - \cos^2\frac{(i-1)\Pi}{n}} \\ a_i' &= 2b_i' \sinh\gamma \cos\frac{(i-1)\Pi}{n} \end{aligned}\right\} &\quad \text{für } \left(i = 2 \ldots \frac{n+1}{2}\right)\ n \text{ ungerade} \end{aligned} \tag{2.6.18}$$

Dabei ist $\gamma = \frac{1}{n}\text{Arsinh}\frac{1}{\epsilon}$. Das so erhaltene Filter hat alle gewünschten Eigenschaften, bis auf die Normierung. Wir verschieben die Frequenzachse um den Faktor $\alpha$ so, dass $|A(j1)| = 1/\sqrt{2}$ ist. Die wahren Koeffizienten sind dann $a_i = \alpha a_i'$ und $b_i = \alpha^2 b_i'$. Die Tabellen F.4 bis F.7 zeigen die Filterkoeffizienten. Abbildungen E.7 bis E.10 zeigen die Frequenzgänge, die en sowie das Phasenbild.

### 2.6.1.3. Besseltiefpässe

Insbesondere für die Verarbeitung von steilen Impulssignalen ist es wünschenswert, wenn die Durchlaufzeit durch den Filter für alle Frequenzen konstant ist. Diese

| $n$ | |
|---|---|
| 1 | $1+P$ |
| 2 | $1+P+\frac{1}{3}P^2$ |
| 3 | $1+P+\frac{2}{5}P^2+\frac{1}{15}P^3$ |
| 4 | $1+P+\frac{3}{7}P^2+\frac{2}{21}P^3+\frac{1}{105}P^4$ |

Tabelle 2.17.: Tabelle der Besselpolynome

Durchlaufzeit wird im Alölgemeinen die Gruppenlaufzeit genannt. Die *Gruppenlaufzeit* ist eine Funktion der Phasenverschiebung (analog zur Gruppengeschwindigkeit bei Wellen).

$$t_{gr} = -\frac{d\varphi}{d\omega} \tag{2.6.19}$$

Die Phase $\varphi$ kann aus der Übertragungsfunktion $A(j\Omega)$ wie folgt berechnet werden:

$$\varphi = -\arctan\frac{\Im\left(A\left(j\Omega\right)\right)}{\Re\left(A\left(j\Omega\right)\right)} \tag{2.6.20}$$

Auf die *Gruppenlaufzeit* eines allgemeinen Filters n-ter Ordnung wollen wir das Butterworth-Konzept anwenden, um eine möglichst konstante Phase zu bekommen. Dazu verwenden wir die normierte Gruppenlaufzeit

$$T_{gr} = \frac{t_{gr}}{T_0} = t_{gr}f_0 = \frac{1}{2\pi}t_{gr}\omega_0 \tag{2.6.21}$$

Dabei ist $f_0$ die Grenzfrequenz des Filters und $T_0 = 1/f_0$ die dazugehörige Zeitkonstante. Aus Gleichung (2.6.21) erhält man

$$T_{gr} = -\frac{\omega_0}{2\pi}\frac{d\varphi}{d\omega} = -\frac{1}{2\pi}\frac{d\varphi}{d\Omega} \tag{2.6.22}$$

Man rechnet nun für ein zu optimierendes Filter die *Gruppenlaufzeit* aus. Für kleine Frequenzen $\Omega \ll 1$ sind nur die niedrigsten Potenzen von Bedeutung. Da in jedem Teilprodukt im Nenner und Zähler jeweils Potenzen von $\Omega^2$ und $\Omega^4$ auftreten, werden die Vorfaktoren von $\Omega^4$ nicht berücksichtigt und diejenigen von $\Omega^2$ gleichgesetzt. Williams[**?**] gibt eine Rekursionsformel für die Koeffizienten $c_i'$ der Besselpolynome an.

$$\begin{aligned} c_1' &= 1 \\ c_i' &= \frac{2\left(n-i+1\right)}{i\left(2n-i+1\right)}c_{i-1}' \end{aligned} \tag{2.6.23}$$

Die daraus resultierenden Besselpolynome sind in der Tabelle 2.17 angegeben. In der Tabelle F.2 sind die Koeffizienten wie üblich auf die 3-dB Grenzfrequenz umgerechnet. Die Abbildungen im Anhang E zeigen drastisch die unterschiedlichen *Gruppenlaufzeit*en (und damit die Impulsverzerrungen) der besprochenen Filtertypen.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

#### 2.6.1.4. Tiefpass-Hochpass-Transformation

Hochpässe können aus den Tiefpassfilterfunktionen abgeleitet werden, indem man die sogenannte Tiefpass-Hochpass-Transformation anwendet.

$$P \to \frac{1}{P} \tag{2.6.24}$$

Aus der Übertragungsfunktion

$$A(P) = \frac{A_0}{\prod_i (1 + a_i P + b_i P^2)} \tag{2.6.25}$$

wird

$$A(P) = \frac{A_\infty}{\prod_i (1 + a_i P^{-1} + b_i P^{-2})} \tag{2.6.26}$$

Damit können alle oben besprochenen Filtertypen als Hochpässe realisiert werden. Der Abschnitt E.2 im Anhang gibt eine Übersicht über die Übertragungsfunktionen, die Phasenbilder und die *Gruppenlaufzeit*en.

#### 2.6.1.5. Tiefpass-Bandpass-Transformation

Bandpässe können aus den Tiefpassfilterfunktionen abgeleitet werden, indem man die sogenannte Tiefpass-Bandpass-Transformation anwendet.

$$P \to \frac{1}{\Delta\Omega}\left(P + \frac{1}{P}\right) = \frac{1}{\Delta\Omega}\frac{P^2+1}{P} \tag{2.6.27}$$

$\Delta\Omega$ ist die Breite des Durchlassbereiches auf dem -3dB-Niveau. $\Delta\Omega$ hängt mit der Bandbreite $B$ und der Güte $Q$ wie folgt zusammen

$$\Delta\Omega = \Omega_{\max} - \Omega_{\min} = \frac{f_{\max} - f_{\min}}{f_0} = \frac{B}{f_0} = \frac{1}{Q} \tag{2.6.28}$$

Aus der Übertragungsfunktion

$$A(P) = \frac{A_0}{\prod_i (1 + a_i P + b_i P^2)} \tag{2.6.29}$$

wird

$$A(P) = \frac{A_\infty}{\prod_i \left(1 + a_i \frac{P^2+1}{\Delta\Omega P} + b_i \left(\frac{P^2+1}{\Delta\Omega P}\right)^2\right)} \tag{2.6.30}$$

Damit können alle oben besprochenen Filtertypen als Bandpässe realisiert werden. Zu beachten ist, dass es für Bandpässe nur gerade Ordnungen gibt. Aus einem Tiefpass 3. Ordnung wird so ein Bandpass 6. Ordnung. Ausser bei einem Bandpass 2. Ordnung kann durch die Wahl der Filterfunktion und der Güte $Q$ die Breite, die Flachheit im Durchlassbereich und die Steilheit getrennt gewählt werden. Der Abschnitt E.3 im Anhang gibt eine Übersicht über die Übertragungsfunktionen, die Phasenbilder und die *Gruppenlaufzeit*en.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

#### 2.6.1.6. Tiefpass-Bandsperren-Transformation

Bandsperren können aus den Tiefpassfilterfunktionen abgeleitet werden, indem man die sogenannte Tiefpass-Bandsperren-Transformation anwendet.

$$P \to \Delta\Omega \frac{1}{P + \frac{1}{P}} = \Delta\Omega \frac{P}{P^2 + 1} \tag{2.6.31}$$

$\Delta\Omega$ ist die Breite des Sperrbereiches auf dem -3dB-Niveau. $\Delta\Omega$ hängt mit der Bandbreite $B$ und der Güte $Q$ analog zum Bandpass zusammen (Gleichung (2.6.28)). Aus der Übertragungsfunktion

$$A(P) = \frac{A_0}{\prod_i (1 + a_i P + b_i P^2)} \tag{2.6.32}$$

wird

$$A(P) = \frac{A_\infty}{\prod_i \left(1 + a_i \Delta\Omega \frac{P}{P^2+1} + b_i \left(\Delta\Omega \frac{P}{P^2+1}\right)^2\right)} \tag{2.6.33}$$

Damit können alle oben besprochenen Filtertypen als Bandsperren realisiert werden. Zu beachten ist, dass es für Bandsperren nur gerade Ordnungen gibt. Aus einem Tiefpass 3. Ordnung wird so eine Bandsperre 6. Ordnung. Ausser bei einer Bandsperre 2. Ordnung kann durch die Wahl der Filterfunktion und der Güte $Q$ die Breite, die Flachheit im Durchlassbereich (ausserhalb des Sperrbereiches) und die Steilheit getrennt gewählt werden. Der Abschnitt E.4 im Anhang gibt eine Übersicht über die Übertragungsfunktionen, die Phasenbilder und die *Gruppenlaufzeit*en.

#### 2.6.1.7. Allpässe

Wenn gewünscht ist, dass ein *Signal* zwar verzögert, nicht aber in seiner Amplitude oder Form geändert wird, dann können Allpassfilter eingesetzt werden. Allpassfilter dienen auch als Phasenschieber. Die den Allpassfiltern zugrundeliegende Idee ist die folgende:

- Der Betrag eines Koeffizienten aus einer Funktion und ihrer konjugiert komplexen Funktion ist konstant und gleich eins.

Wir setzen also für die Übertragungsfunktion an

$$\begin{aligned} A(P) &= \frac{\bar{f}(j\Omega)}{f(j\Omega)} \\ &= \frac{\prod_i (1 - a_i P + b_i P^2)}{\prod_i (1 + a_i P + b_i P^2)} \\ &= \frac{\prod_i \left(\sqrt{(1 - b_i\Omega^2)^2 + a_i^2\Omega^2}\, e^{-j\alpha}\right)}{\prod_i \left(\sqrt{(1 - b_i\Omega^2)^2 + a_i^2\Omega^2}\, e^{j\alpha}\right)} \\ &= e^{-2j\alpha} = e^{j\varphi} \end{aligned} \tag{2.6.34}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

mit

$$\varphi = -2\alpha = -2\sum_i \arctan \frac{a_i \Omega}{1 - b_i \Omega} \tag{2.6.35}$$

Die Koeffizienten werden mit dem Butterworth-Ansatz so berechnet, das die *Gruppenlaufzeit* über einen möglichst grossen Frequenzbereich konstant bleibt. Im Anhang finden Sie im Abschnitt E.5 die Übertragungsfunktion, das Phasenbild und die *Gruppenlaufzeit* dieser Filter. Die Filterkoeffizienten sind in Tabelle F.8 angegeben.

## 2.6.2. Digitalfilter

Mit der Verfügbarkeit moderner Hochleistungsrechner werden immer öfter Filter auf digitaler Hardware implementiert. Digitale Filter können einerseits analoge Filter nachbilden, sind andererseits auch in der Lage wesentlich kompliziertere Filterfunktionen nachzubilden. Insbesondere sind Bauteiltoleranzen bei digitalen Filtern nicht problematisch. Rundungsfehler können mit mathematischen Methoden abgeschätzt werden. Sie sind dann für alle Implementationen gleich.

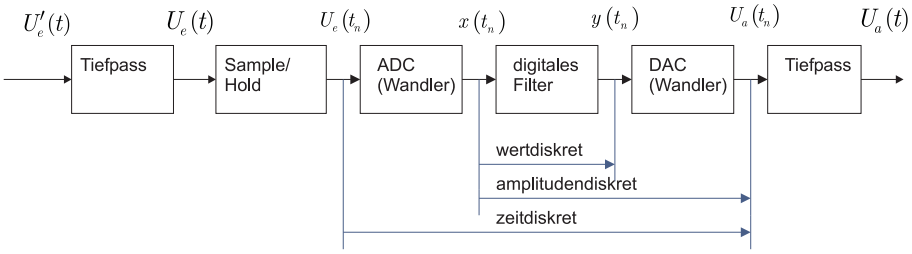



Abbildung 2.43.: Systemumgebung eines digitalen Filters

Ein digitales Filter wird üblicherweise in einer Umgebung wie in Abbildung 2.43 eingesetzt. Dabei wird das analoge *Signal* zuerst in ein digitales *Signal* umgewandelt. Nach der Verarbeitung des so entstandenen Zahlenstromes im digitalen Filter wird mit Digital-Analogwandlern wieder ein analoges *Signal* erzeugt. Filter am Eingang und am Ausgang sorgen dafür, dass keine unerwünschten Frequenzkomponenten vorhanden sind.

### 2.6.2.1. Abtastung

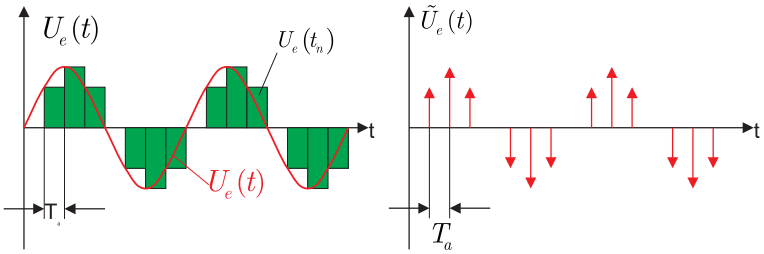



Abbildung 2.44.: Analoge Eingangsfunktion und die entsprechende Treppenfunktion (links) und Dirac-Pulsfolge (rechts).

Das analoge *Signal* muss zuerst in ein digitales *Signal* umgewandelt werden. Typische Auflösungen und Frequenzen bei dieser Umwandlung sind 44,1 kHz und 16 Bit bei Audiosignalen, 8 kHz und 12 Bit bei der Telephonie und 13,3 MHz und 8 Bit bei Fernsehsignalen. Abbildung 2.44 zeigt wie aus einem analogen *Signal* mit einem *Abtast-Halteglied* eine Treppenfunktion entsteht. Es ist jedoch leichter, anstelle einer Treppe äquivalente Dirac-Impulse zu verwenden. Auch dieses *Signal* ist in Abbildung 2.44 gezeigt. Mathematisch wird die Folge von Dirac-Impulsen (wir sind immer noch auf der Analogseite) wie folgt beschrieben:

$$\tilde{U}_e(t) = \sum_{n=0}^{\infty} U_e(t_n)\, T_a \delta(t - t_n) \tag{2.6.36}$$

Wenn Gleichung (2.6.36) fouriertransformiert wird, erhält man das Spektrum der Dirac-Pulsfolge.

$$\tilde{X}(jf) = T_a \sum_{n=0}^{\infty} U_e(nT_a)\, e^{-2\pi} e^{-2\pi j n \frac{f}{f_a}} \tag{2.6.37}$$

Hier ist $f_a = 1/T_a$ die Abtastfrequenz. Das Spektrum ist periodisch mit der Abtastfrequenz. Aus Abbildung 2.45 ist ersichtlich, dass ein Eingangsspektrum nur dann getreu wiedergegeben werden kann, wenn seine Bandbreite geringer als

$$f_a \geq 2 f_{max} \tag{2.6.38}$$

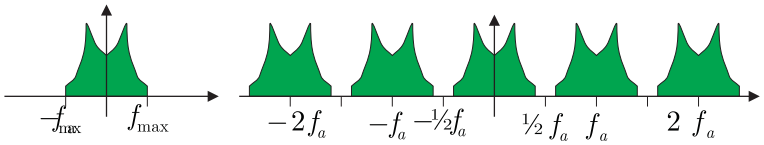



Abbildung 2.45.: Spektrum der Eingangsspannung vor und nach dem Abtasten

ist. Diese Bedingung wird Abtasttheorem genannt. $f_a/2$ heisst auch die Nyquist-Frequenz. Das Eingangsfilter in Abbildung 2.43 dient zur Verringerung der Bandbreite des Eingangssignals, so dass Gleichung (2.6.2) erfüllt ist. Das Eingangsfilter kann entweder explizit verwendet werden, oder aber man muss sicherstellen, dass das vorhergehende System keine Frequenzkomponenten über der Nyquistfrequenz abgibt. **Um die Konstruktion des Eingangsfilters einfach zu halten, wird meistens anstelle von Gleichung (2.6.38) die Bedingung $f_a \geq 5 f_{max}$ verwendet.** Wenn man bewusst ein schmalbandiges *Signal* um eine der Vielfachen der Abtastfrequenz digitalisiert und dafür sorgt, dass im eigentlichen Frequenzintervall um 0 keine Signalanteile vorhanden sind, dann ist es möglich ein schnelleres *Signal* als von der Abtastfrequenz gegeben, zu digitalisieren (Abbildung 2.46). Das Verfahren, Oversampling genannt, wird häufig in Höchstfrequenz-Oszilloskopen eingesetzt.
Die Rückgewinnung des analogen Signals erfordert ebenfalls eine Beschneidung der Bandbreite. Jeder *digital-analog-Wandler* erzeugt neben dem ursprünglichen *Signal* auch die um die Vielfachen der Abtastfrequenz gespiegelten Komponenten. Diese müssen mit einem Tiefpassfilter wie in Abbildung 2.43 herausgefiltert werden, es sei denn, man stellt sicher, dass das nachfolgende System wie zum Beispiel bei einem STM diese Funktion übernimmt.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

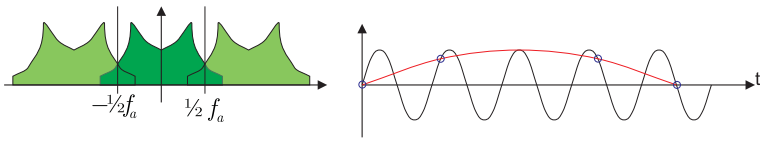



Abbildung 2.46.: Oversampling oder Schwebung beim Abtasten

In der Praxis ist es nicht möglich, Dirac-Impulse zu erzeugen. Wenn die Dirac-Pulse durch Pulse der Breite $\varepsilon T_a$ ersetzt werden, kann das Eingangssignal als

$$\tilde{U}'_e(t) = \sum_{n=0}^{\infty} U_e(nT_a)\, r_\varepsilon(t - nT_a) \tag{2.6.39}$$

geschrieben werden. $r_\varepsilon$ ist der Rechteckpuls. Als *Fouriertransformation* erhält man

$$\tilde{X}'(jf) = \frac{\sin \pi\varepsilon T_a f}{\pi\varepsilon T_a f} \tilde{X}(jf) \tag{2.6.40}$$

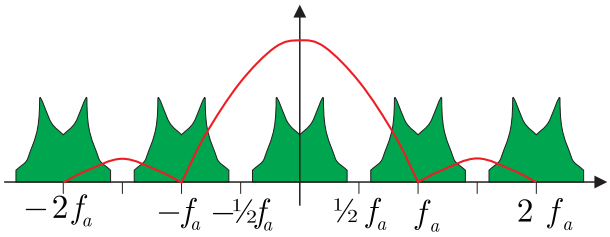



Abbildung 2.47.: Fensterfunktion beim Abtasten mit Rechteckpulsen

Abbildung 2.47 zeigt, dass bei einer geschickten Wahl von $\varepsilon$ die erste Nullstelle der Funktion $sinx/x$ gerade auf die Abtastfrequenz fällt. Dieser Effekt tritt genau dann auf, wenn $\varepsilon = 1$ ist, wenn man also eine Treppenfunktion hat. Die Fensterfunktion bewirkt andererseits, dass bei der halben Abtastfrequenz das *Signal* um den **Faktor 0,64** abgeschwächt ist. Man kann also nicht einfach ein Digitalsignal zurückwandeln, ohne den Einfluss der Fensterfunktion zu berücksichtigen. Nach Abbildung 2.48 muss das Tiefpassfilter am Ausgang zur Frequenzgangkorrektur verwendet werden. Alternativ, wenn $f_a \geq 5 f_{max}$ verwendet wird, ist das Filterproblem vielfach zu vernachlässigen. Das bei CD-Spielern als Errungenschaft verkaufte Oversampling dient in erster Linie dazu, Geld bei analogen Ausgangsfiltern zu sparen!

### 2.6.2.2. Bausteine für digitale Filter

Der Grundbaustein für ein digitales Filter ist die Verzögerungsstrecke um das Abtastintervall $T_a$. Diese Verzögerungsstrecke tritt auch als Zwischenspeicherung auf, wenn man im Rechner in einer Schleife Signale berechnet. Die Folge $\{x(t_n)\}$ wird in die Folge $\{y(t_n)\}$ übergeführt mit

$$\{y(t_n)\} = \{x(t_{n-1})\} \tag{2.6.41}$$

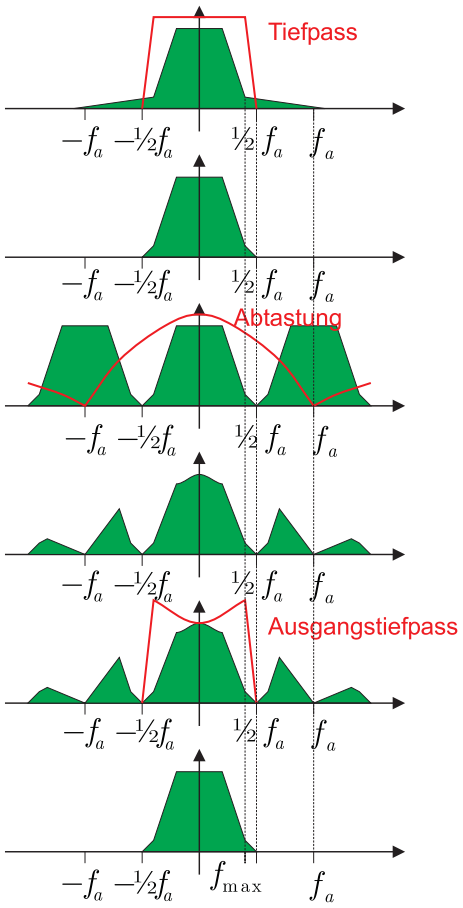



Abbildung 2.48.: Darstellung der Änderungen des Frequenzganges bei einer analog-digital-analog-Signalkette

Für eine harmonische Folge $x(t_n) = \hat{x}e^{j\omega t_n}$ gilt $y(t_n) = \hat{x}e^{j\omega t_n}e^{-j\omega T_a}$. Die Übertragungsfunktion eines Verzögerungsgliedes (auch Totzeitglied genannt) ist also

$$A\,(p) = \frac{y(t_n)}{x(t_n)} = e^{-pT_a} \tag{2.6.42}$$

Der Frequenzgang eines Verzögerungsgliedes ist also eine periodische Funktion. Das heisst, dass bei Verzögerungsstrecken immer beliebig grosse Phasenverschiebungen auftauchen. Die Phasenverschiebung ist also, im Gegensatz zu der eines Tiefpasses, nicht kompensierbar.
Mit der im Abschnitt 2.4.4.2 eingeführten z-Transformation kann unter Verwendung der Regel

$$z^{-1} = e^{-pT_a} \tag{2.6.43}$$

erhält man für die digitale Übertragungsfunktion

$$\tilde{A}(z) = z^{-1} \tag{2.6.44}$$

Durch Rückeinsetzen erhält man

$$\underline{A}(j\omega) = z^{-1} = e^{-j\omega T_a} = \cos\omega T_a - j\sin\omega T_a \tag{2.6.45}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Daraus bekommt man für die Amplitude, Phase und Gruppenlaufzeit

$$\begin{aligned} |\underline{A}(j\omega)| &= \sqrt{\cos^2 \omega T_a + \sin^2 \omega T_a} = 1 \\ \varphi &= \arctan \frac{-\sin \omega T_a}{\cos \omega T_a} = \arctan\left(-\tan \omega T_a\right) = -\omega T_a \\ t_{gr} &= -\frac{d\varphi}{d\omega} = T_a \end{aligned} \tag{2.6.46}$$

also das erwartete Resultat.
Der zweite Baustein für ein digitales Filter ist ein Summationsknoten, der dritte die Multiplikation mit einem konstanten Faktor.

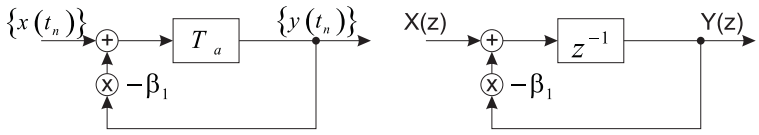



Abbildung 2.49.: Ein Beispieltiefpass

**2.6.2.2.1. Beispiel** Mit den oben genannten Bauteilen kann ein Tiefpass (Abbildung 2.49) aufgebaut werden. Wir setzen für die Folgen:

$$y\left(t_{n+1}\right) = x\left(t_n\right) - \beta_1 y\left(t_n\right) \tag{2.6.47}$$

Die entsprechende Funktion im z-Raum ist

$$Y(z) = \frac{z^{-1}}{1 + \beta_1 z^{-1}} X(z) \tag{2.6.48}$$

Daraus erhält man die Übertragungsfunktion

$$\tilde{A}(z) = \frac{Y(z)}{X(z)} = \frac{z^{-1}}{1 + \beta_1 z^{-1}} \tag{2.6.49}$$

Der Frequenzgang kann berechnet werden, indem man

$$z^{-1} = e^{-j\omega T_a} = \cos \omega T_a - j \sin \omega T_a \tag{2.6.50}$$

setzt. Die Übertragungsfunktion in der Zeit wird demnach

$$\underline{A}(j\omega) = \frac{1}{\beta_1 + e^{j\omega T_a}} = \frac{1}{\beta_1 + \cos \omega T_a + j \sin \omega T_a} \tag{2.6.51}$$

Amplitude und Phase sind

$$\begin{aligned} |\underline{A}(j\omega)| &= \frac{1}{\sqrt{\left(\beta_1 + \cos \omega T_a\right)^2 + \left(\sin \omega T_a\right)^2}} \\ \arg \underline{A}(j\omega) &= \frac{-\sin \omega T_a}{\beta_1 + \cos \omega T_a} \end{aligned} \tag{2.6.52}$$

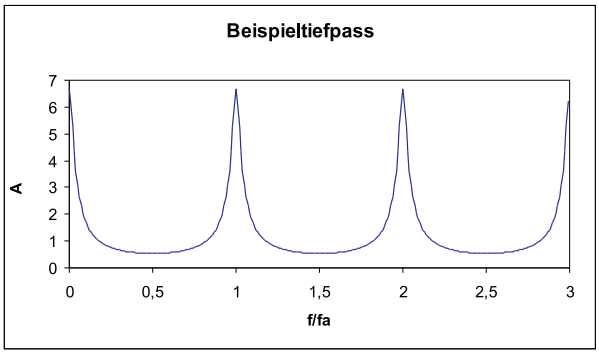



Abbildung 2.50.: Frequenzgang eines digitalen Tiefpasses mit $\beta_1 = -0,85$

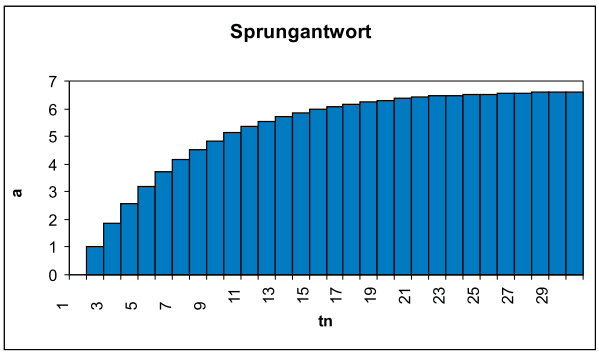



Abbildung 2.51.: Sprungantwort eines digitalen Tiefpasses mit $\beta_1 = -0,85$

Abbildung 2.50 zeigt den Frequenzgang eines Tiefpasses mit $\beta_1 = -0,85$. Abbildung 2.51 zeigt die dazugehörige Sprungantwort. Der Frequenzgang ist mit $1/T_a$ periodisch. Das Abtasttheorem sagt, dass der Tiefpass nur bis zur Frequenz $1/2T_a$ verwendet werden kann.

Ist $\beta_1$ positiv, erhält man eine Hochpasscharakteristik. Abbildung 2.52 zeigt den Frequenzgang für $\beta_1 = 0,85$, Abbildung 2.53 ist die dazugehörige Sprungantwort.

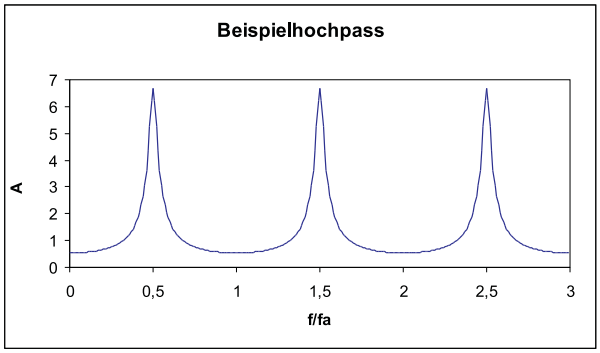



Abbildung 2.52.: Frequenzgang eines digitalen Hochpasses mit $\beta_1 = 0,85$

Setzt man $\beta_1 = -1$ so wird aus dem Tiefpass ein Integrator, wie man unschwer am Frequenzgang in Abbildung 2.50 ersehen kann. Dann gilt

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

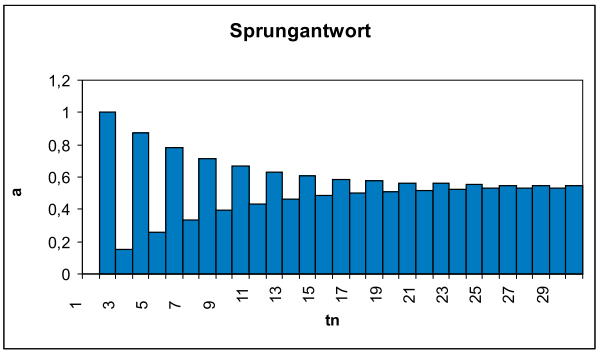



Abbildung 2.53.: Sprungantwort eines digitalen Hochpasses mit $\beta_1 = 0,85$

$$\begin{aligned}(\cos\alpha - 1)^2 + \sin^2\alpha &= \cos^2\alpha + \sin^2\alpha - 2\cos\alpha + 1 \\ &= 2\,(1 - \cos\alpha) \\ &= 2\left(1 - \left[\cos^2\frac{\alpha}{2} - \sin^2\frac{\alpha}{2}\right]\right) \\ &= 4\sin^2\frac{\alpha}{2} \end{aligned} \tag{2.6.53}$$

Eingesetzt in den Frequenzgang erhält man

$$|\underline{A}\,(j\omega)| = \frac{1}{2\sin\frac{\omega T_a}{2}} \sim \frac{1}{\omega} \tag{2.6.54}$$

was in der Tat der Frequenzgang eines Integrators ist.

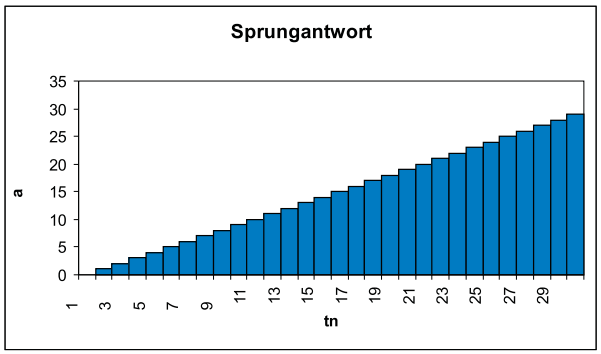



Abbildung 2.54.: Der digitale Tiefpass aus Abbildung 2.49 als Integrator. Gezeigt ist die Sprungantwort

### 2.6.2.3. Grundstrukturen digitaler Filter

Zentral für ein digitales Filter ist die Verzögerungskette. Für ihre Einbettung in die Filterstruktur gibt es drei mögliche Konfigurationen:

1. Abbildung 2.55 zeigt ein Filter mit verteilten Summierern. Da jeweils zu jeder Summenbildung auch eine Multiplikation gehört, ist diese Topologie

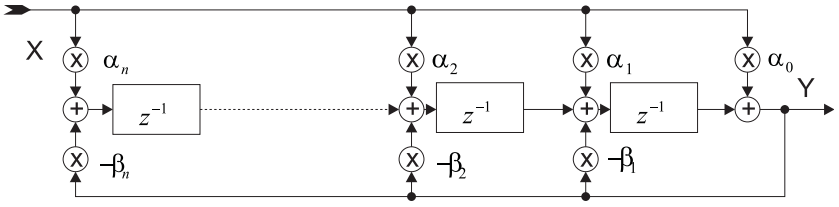



Abbildung 2.55.: Digitales Filter mit verteilten Summierern

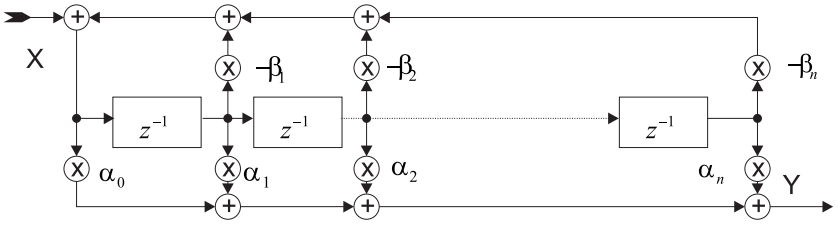



Abbildung 2.56.: Digitales Filter mit globalen Summierern an Eingang und Ausgang

ideal geeignet zur Implementation in einem digitalen Signalprozessor. Bei der seriellen Abarbeitung der Befehle in diesen Prozessoren ist es sogar von Vorteil, wenn die Summation nicht auf einen Summierer konzentriert ist. Jede Summation/Multiplikation ist jeweils um einen Taktschritt verzögert.

2. Bei Abbildung 2.56 ist der Eingang der Verzögerungskette das gewichtete Mittel aus allen Zwischenwerten sowie aus dem Eingangssignal. Der Ausgang andererseits ist das (anders gewichtete) Mittel aller Zwischenwerte.

3. Die Schaltung in Abbildung 2.57 hat nur einen einzelnen Summierer am Ausgang. Im Gegensatz zu den obigen Filtern wird hier eine zweite Verzögerungskette benötig.

Die Ordnung $n$ des Filters bestimmt auch, dass $n$ Verzögerungsstufen benötigt werden. Hier soll nur die gebräuchlichste Topologie analysiert werden, nämlich die aus Abbildung 2.55. Die Differenzengleichung dafür lautet:

$$y\left(t_n\right) = \sum_{k=0}^{n} \alpha_k x_{n-k} - \sum_{k=1}^{n} \beta_k y_{n-k} \tag{2.6.55}$$

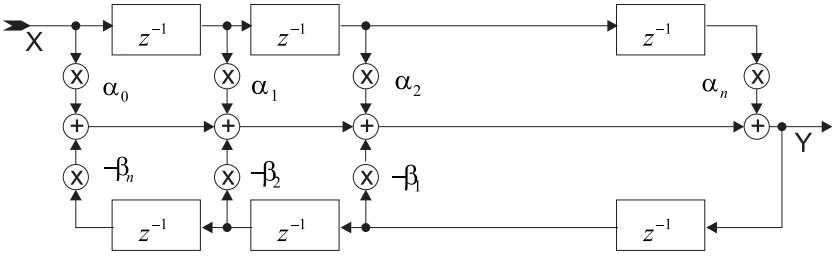



Abbildung 2.57.: Digitales Filter mit einem globalen Summierer am Ausgang

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Entsprechend berechnet sich aus

$$Y(z) = \sum_{k=0}^{n} \alpha_k z^{-k} X(z) - \sum_{k=1}^{n} \beta_k z^{-k} Y(z) \tag{2.6.56}$$

ist die Übertragungsfunktion

$$\begin{aligned} A(z) &= \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^{n} \alpha_k z^{-k} X(z)}{1 + \sum_{k=1}^{n} \beta_k z^{-k} Y(z)} \\ A(z) &= \frac{\alpha_0 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \ldots + \alpha_n z^{-n}}{1 + \beta_1 z^{-1} + \beta_2 z^{-2} + \ldots + \beta_n z^{-n}} \end{aligned} \tag{2.6.57}$$

Die komplexe Übertragungsfunktion kann mit der Identität (2.6.43) berechnet werden. Weiter normiert man die Frequenz auf die Abtastfrequenz

$$\begin{aligned} F &= \frac{f}{f_a} \\ \omega T_a &= 2\pi F \end{aligned} \tag{2.6.58}$$

Das Abtasttheorem verlangt, dass das Eingangssignal nur Frequenzen

$$\begin{aligned} 0 \leq \quad f \quad &\leq \frac{1}{2} f_a \\ 0 \leq \quad F \quad &\leq \frac{1}{2} \end{aligned} \tag{2.6.59}$$

Indem man formal für den in der Filterkette nicht vorkommenden Koeffizienten $\beta_0 = 1$ setzt, erhält man für die Übertragungsfunktion:

$$|\underline{A}(j\omega)| = \sqrt{\frac{\left[\sum_{k=0}^{n} \alpha_k \cos 2\pi k F\right]^2 + \left[\sum_{k=0}^{n} \alpha_k \sin 2\pi k F\right]^2}{\left[\sum_{k=0}^{n} \beta_k \cos 2\pi k F\right]^2 + \left[\sum_{k=0}^{n} \beta_k \sin 2\pi k F\right]^2}} \tag{2.6.60}$$

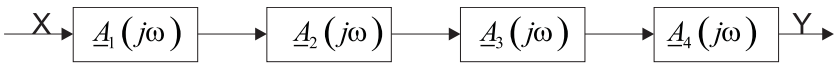



Abbildung 2.58.: Kaskadierung von digitalen Filtern

Filter können entweder als einen Block berechnet werden, oder aber wie in Abbildung 2.58 kaskadiert werden. Im Falle der Kaskadierung gilt

$$\begin{aligned} \underline{A}ges &= \underline{A}_1 \cdot \underline{A}_2 \cdot \underline{A}_3 \cdot \underline{A}_4 \\ \left|\underline{A}_{ges}\right| &= |\underline{A}_1| \cdot |\underline{A}_2| \cdot |\underline{A}_3| \cdot |\underline{A}_4| \\ N_{ges} &= N_1 + N_2 + N_3 + N_4 \end{aligned} \tag{2.6.61}$$

Der Frequenzgang ist also das Produkt der einzelnen Frequenzgänge, die Ordnung die Summe der einzelnen Ordnungen. Die Kaskadierung vereinfacht die Berechnung der Filterkoeffizienten und die Verifizierung von Filtern Insbesondere bei Filtern mit Rückkopplung ist dies ein wichtiger Punkt.
Man unterscheidet zwei Typen von Filtern

**FIR-Filter** **F**inite **I**mpulse **R**esponse Filter haben keine Rückkopplung. Sie sind deshalb unter allen Betriebszuständen stabil und vorhersagbar. Die genauere Untersuchung wird aber zeigen, dass man sehr hohe Filterordnungen benötigt. Andererseits ist Die Antwort auf einen Impuls von endlicher Länge.

**IIR-Filter** **I**nfinite **I**mpulse **R**esponse Filter besitzen einen Rückkopplungszweig. Man benötigt deshalb weniger Filterstufen für eine ähnlich steile Filterwirkung wie bei FIR-Filtern. Die Verzögerungszeiten sind deshalb auch geringer. Die Impulsantwort dieser Filter dauert unendlich lange.

#### 2.6.2.4. FIR-Filter

Bei FIR-Filtern sind alle Koeffizienten $\beta_i = 0$. Die Differenzengleichung lautet also:

$$\begin{aligned} y(t_m) &= \alpha_0 x(t_m) + \alpha_1 x(t_{m-1}) + \ldots + \alpha_{n-1} x(t_{m-n+1}) + \alpha_n x(t_{m-n}) \\ &= \sum_{k=0}^{n} \alpha_k x(t_{m-k}) \end{aligned} \tag{2.6.62}$$

Daraus folgt für den z-Raum

$$Y(z) = \left[\alpha_0 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \ldots + \alpha_n z^{-n}\right] X(z) \tag{2.6.63}$$

und für die Übertragungsfunktion

$$\tilde{A}(z) = \frac{Y(z)}{X(z)} = \sum_{k=0}^{n} \alpha_k z^{-k} \tag{2.6.64}$$

Mit Gleichung (2.6.43) bekommt man den komplexen Frequenzgang

$$\underline{A}(j\omega) = \sum_{k=0}^{n} \alpha_k e^{-j2\pi kF} \tag{2.6.65}$$

Wenn die Filterkoeffizienten Symmetriebedingungen genügen lassen sich besonders einfach zu berechnende Filterstrukturen realisieren. Bei gerader Symmetrie $\alpha_{n-k} = \alpha_n$ ist der komplexe Frequenzgang

$$\underline{A}(j\omega) = e^{-j2\pi nF} \sum_{k=0}^{n} \alpha_k \cos\left[\pi (n-2k) F\right] \tag{2.6.66}$$

Bei ungerader Symmetrie $\alpha_{n-k} = -\alpha_n$ erhält man entsprechend

$$\underline{A}(j\omega) = j e^{-j2\pi nF} \sum_{k=0}^{n} \alpha_k \sin\left[\pi (n-2k) F\right] \tag{2.6.67}$$

Bei ungerader Symmetrie in gerader Ordnung muss $\alpha_{\frac{1}{2}n} = 0$ sein. Also können Amplitude und Phase explizit angegeben werden:

$$\underline{A}(j\omega) = \begin{cases} B(\omega) & \text{gerade Symmetrie} \\ B(\omega) & \text{ungerade Symmetrie} \end{cases} \tag{2.6.68}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Der Betrag der Übertragungsfunktion lässt sich einfach aus den Summen in den Gleichungen (2.6.66) und (2.6.66) berechnen. Die Phase ist

$$\varphi = \begin{cases} -\pi n F & \text{gerade Symmetrie} \\ -\pi n F + \frac{\pi}{2} & \text{ungerade Symmetrie} \end{cases} \tag{2.6.69}$$

In beiden Fällen hängt die Phase linear von der Frequenz ab. Entsprechend ist die Gruppenlaufzeit

$$t_{gr} = -\frac{d\varphi}{d\omega} = -\frac{d\varphi}{dF}\frac{dF}{d\omega} = -\frac{T_a}{2\pi}\frac{d\varphi}{dF} = \frac{1}{2} n T_a \tag{2.6.70}$$

**Die *Gruppenlaufzeit* ist für alle FIR-Filter mit symmetrischen Koeffizienten frequenzunabhängig**. Symmetrische FIR Filter eingesetzt in Audiogeräten erzeugen keine Laufzeitverzerrungen! Ausser in Ausnahmefällen verwendet man nur FIR-Filter mit linearer Phase.

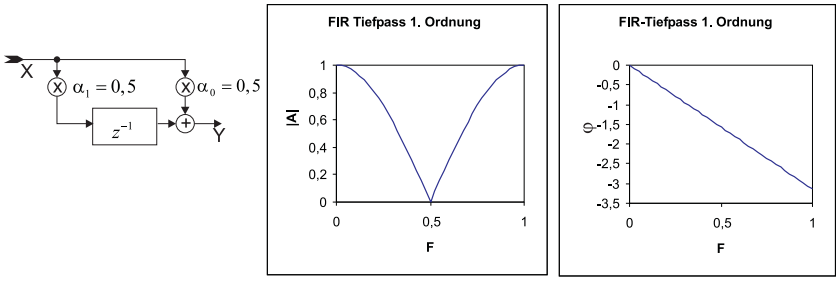



Abbildung 2.59.: Ein FIR Tiefpass 1. Ordnung. Links: Schaltschema. Mitte: Frequenzgang. Rechts: Phasengang

**2.6.2.4.1. FIR-Filter 1. Ordnung** Das in Abbildung 2.59 gezeigte Filter ist ein Tiefpass (Berechnet mit Excel-Tabelle[1]). Für eine Einheitsfolge $\{x_\mu\} = \{1, 1, 1, \ldots\}$ erhält man die Ausgangsfolge $\{y_\mu\} = \{1, 1, 1, \ldots\}$. Die Verstärkung ist hier also 1. Man berechnet dass die z-Übertragungsfunktion, die komplexe Übertragungsfunktion, ihr Betrag, die normierte Grenzfrequenz, die Phase und die Gruppenlaufzeit

$$\begin{aligned} \tilde{A}(z) &= 0.5\left(1 + z^{-1}\right) \\ \underline{A}(j\omega) &= 0.5\left(1 + \cos 2\pi F - j \sin 2\pi F\right) \\ |\underline{A}(j\omega)| &= |\cos \pi F| \\ F_g &= 0.25 \\ \varphi &= -\pi F \\ t_{gr} &= 0.5 T_a \end{aligned} \tag{2.6.71}$$

sind. Eine Analyse der allgemeinen FIR-Übertragungsfunktion zeigt, dass

> **die Gleichspannungsverstärkung eines FIR-Filters gleich der Summe aller Filterkoeffizienten**

[1] http://wwwex.physik.uni-ulm.de/lehre/PhysikalischeElektronik/Materialien/digifilter.xls

$$|\underline{A}(0)| = \sum_{k=0}^{n} \alpha_k \tag{2.6.72}$$

ist. Bei der halben Abtastfrequenz, der höchsten nach dem Nyquist-Theorem zulässigen Frequenz, ist die Eingangsfolge $\{x_\mu\} = \{1, -1, 1, -1, \ldots\}$. Das Ausgangssignal ist, wie man leicht nachprüft, die Nullfolge $\{y_\mu\} = \{0, 0, 0, 0, \ldots\}$. Auch diese Eigenschaft ist allgemeingültig.

**Die Verstärkung eines FIR-Filters bei der halben Abtastfrequenz ist gleich der Summe der im Wechsel mit +1 und -1 gewichteten Koeffizienten**

$$\left|\underline{A}\left(j\frac{F}{2}\right)\right| = \sum_{k=0}^{n} (-1)^k \alpha_k \tag{2.6.73}$$

Weiter stellt mann fest, dass

**wenn man in einem FIR-Filter alle Koeffizienten mit dem gleichen Faktor multipliziert, die Verstärkung um diesen Faktor geändert wird, ohne dass sich die Filtercharakteristik ändert.**

Speist man in ein FIR-Filter die Folge $\{x_\mu\} = \{1, 0, 0, 0, \ldots\}$. ein, so erhält man als Impulsantwort die Folge $\{y_\mu\} = \{\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_n, 0, \ldots\}$, also gerade die Folge der Filterkoeffizienten.

**Die Antwort eines FIR-Filters auf einen Einheitsimpuls ist immer die Folge seiner Filterkoeffizienten. Diese Folge ist bei einem Filter n-ter Ordnung $n+1$ Werte lang von null verschieden.**

Die Grenzfrequenz des Filters erhält man, indem man $|\underline{A}(j2\pi F_g)| = \frac{1}{\sqrt{2}} = \cos \pi F_g$ setzt.

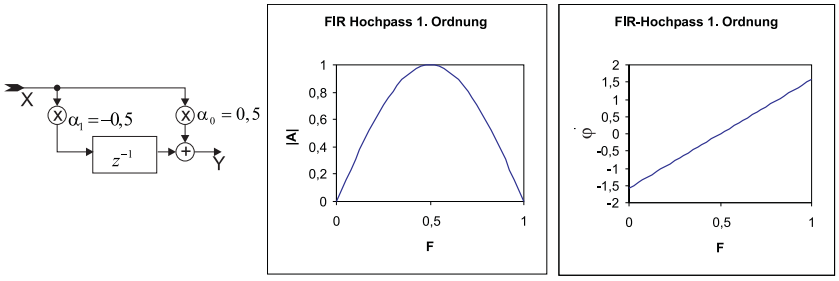



Abbildung 2.60.: Ein FIR Hochpass 1. Ordnung. Links: Schaltschema. Mitte: Frequenzgang. Rechts: Phasengang

Abbildung 2.60 zeigt einen FIR-Hochpass 1. Ordnung (Berechnet mit Excel-Tabelle[2]). Im Unterschied zum Tiefpass ist $\alpha_1$ mit $-1$ multipliziert. Unsere Regeln sagen, dass

[2] http://wwwex.physik.uni-ulm.de/lehre/PhysikalischeElektronik/Materialien/Digifilter.xls

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

die Verstärkung bei der Frequenz null auch null ist, und dass sie bei der halben Abtastfrequenz maximal ist. Im einzelnen sind die Kenngrössen:

$$\begin{aligned} \tilde{A}(z) &= 0,5\left(1-z^{-1}\right) \\ \underline{A}(j\omega) &= 0,5\left(1-\cos 2\pi F - j\sin 2\pi F\right) \\ |\underline{A}(j\omega)| &= |\sin \pi F| \\ F_g &= 0,25 \\ \varphi &= \pi\left(0,5-F\right) \\ t_{gr} &= 0,5T_a \end{aligned} \tag{2.6.74}$$

Bei tiefen Frequenzen ist die Verstärkung proportional zu $F$, die Schaltung arbeitet also (was man aus der Gleichung hätte erraten können!) als Differenzierer. Bandpässe und Bandsperren kann man mit einem Filter erster Ordnung, wie bei analogen Filtern, nicht realisieren.

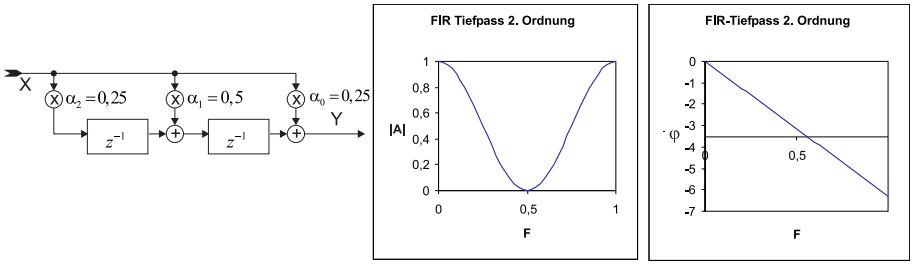



Abbildung 2.61.: Ein FIR Tiefpass 2. Ordnung. Links: Schaltschema. Mitte: Frequenzgang. Rechts: Phasengang

**2.6.2.4.2. FIR-Filter 2. Ordnung** Die Summe der Koeffizienten in der Abbildung 2.61 (Berechnet mit Excel-Tabelle[3]) ist eins, also ist die Verstärkung bei bei der Frequenz null eins. Die Übertragungsfunktion sowie die weiteren Kenndaten sind:

$$\begin{aligned} \tilde{A}(z) &= 0,25 + 0,5z^{-1} + 0,25z^{-2} \\ |\underline{A}(j\omega)| &= 0,5 + 0,5\cos 2\pi F \\ F_g &= \frac{1}{2\pi}\arccos\left(\sqrt{2}-1\right) = 0,182 \\ \varphi &= -2 * \pi \\ t_{gr} &= T_a \end{aligned} \tag{2.6.75}$$

Abbildung 2.62 zeigt einen FIR-Hochpass 2. Ordnung (Berechnet mit Excel-Tabelle[4]). Im Vergleich zum Tiefpass zweiter Ordnung ist das Vorzeichen von $\alpha_1$ gewechselt

[3] http://wwwex.physik.uni-ulm.de/lehre/PhysikalischeElektronik/Materialien/Digifilter.xls
[4] http://wwwex.physik.uni-ulm.de/lehre/PhysikalischeElektronik/Materialien/Digifilter.xls

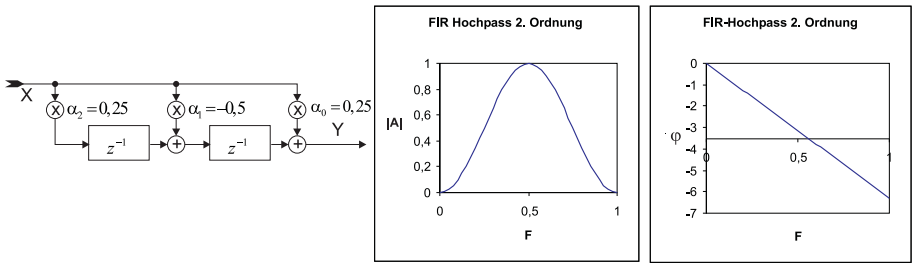



Abbildung 2.62.: Ein FIR Hochpass 2. Ordnung. Links: Schaltschema. Mitte: Frequenzgang. Rechts: Phasengang

worden. Die Übertragungsfunktion sowie die weiteren Kenndaten sind:

$$\begin{aligned} \tilde{A}(z) &= 0,25 - 0,5z^{-1} + 0,25z^{-2} \\ |\underline{A}(j\omega)| &= 0,5 - 0,5\cos 2\pi F \\ F_g &= \frac{1}{2\pi}\arccos\left(1-\sqrt{2}\right) = 0,318 \\ \varphi &= -2 * \pi \\ t_{gr} &= T_a \end{aligned} \tag{2.6.76}$$

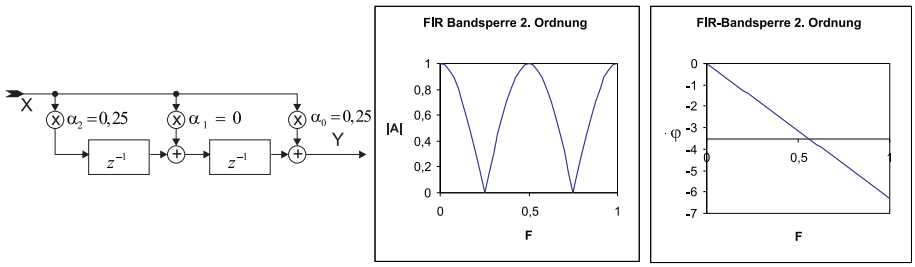



Abbildung 2.63.: Eine FIR Bandsperre 2. Ordnung. Links: Schaltschema. Mitte: Frequenzgang. Rechts: Phasengang

Setzt man $\alpha_1 = 0$ so erhält man aus dem Tiefpass oder Hochpass die in Abbildung 2.63 gezeigte FIR-Bandsperre 2. Ordnung (Berechnet mit Excel-Tabelle[5]. Die Übertragungsfunktion sowie die weiteren Kenndaten sind:

$$\begin{aligned} \tilde{A}(z) &= 0,25 + 0,25z^{-2} \\ |\underline{A}(j\omega)| &= |\cos 2\pi F| \\ F_r &= 0,25 \\ \varphi &= -2 * \pi \\ t_{gr} &= T_a \\ Q &= \frac{F_r}{B} = 1 \end{aligned} \tag{2.6.77}$$

Negiert man in der Bandsperre in Abbildung 2.63 $\alpha_2$ so erhält man aus der Bandsperre einen Bandpass (Abbildung 2.64) (Berechnet mit Excel-Tabelle[6]. Die Über-

[5] http://wwwex.physik.uni-ulm.de/lehre/PhysikalischeElektronik/Materialien/Digifilter.xls
[6] http://wwwex.physik.uni-ulm.de/lehre/PhysikalischeElektronik/Materialien/Digifilter.xls

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

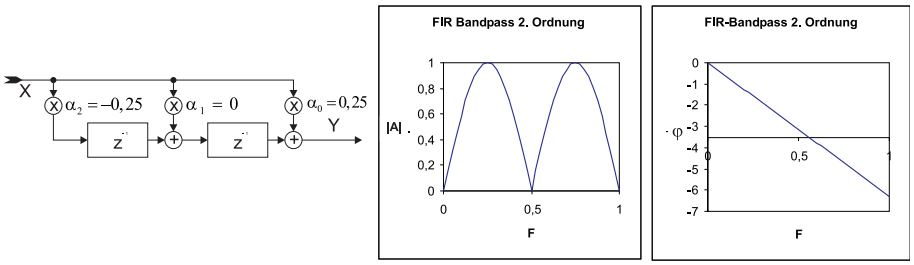



Abbildung 2.64.: Ein FIR Bandpass 2. Ordnung. Links: Schaltschema. Mitte: Frequenzgang. Rechts: Phasengang

tragungsfunktion sowie die weiteren Kenndaten sind:

$$\begin{aligned} \tilde{A}(z) &= 0,25 - 0,25z^{-2} \\ |\underline{A}(j\omega)| &= |\sin 2\pi F| \\ F_r &= 0,25 \\ \varphi &= -2 * \pi \\ t_{gr} &= T_a \\ Q &= \frac{F_r}{B} = 1 \end{aligned} \tag{2.6.78}$$

**2.6.2.4.3. Berechnung der Filterkoeffizienten für FIR-Filter** Eine Einführung in die Berechnung der FIR-Filterkoeffizienten kann in Tietze-Schenk[?] gefunden werden. Es gibt zwei bevorzugte Verfahren zu dieser Berechnung: die Fenster-Methode und den Remez Exchange Algorithmus. Die zweite Methode ist ein auf *Tschebyscheff-Polynom*en beruhender Optimierungsalgorythmus, der eine minimale Anzahl von Filterkoeffizienten liefert.
Wie in der Gleichung (2.6.62) gezeigt wurde, ist die Antwort eines FIR-Filters auf eine Impulsanregung

$$\{x(kT_a)\} = \begin{cases} 1 & \text{für} \quad k = 0 \\ 0 & \text{sonst} \end{cases} \tag{2.6.79}$$

Aus (2.6.79) folgt als Ausgangsfolge die Folge der Koeffizienten.

$$\{y(kT_a)\} = \alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_N = \{\alpha_k\} \tag{2.6.80}$$

Nun sind bei einem linearen System die Impulsantwort und der Frequenzgang durch eine (inverse) Fouriertransformation ineinander überführbar. Für zeitdiskrete Systeme mit $f_a = 1/T_a$ sind die Werte der Impulsantwort für $t = kT_a$ durch

$$y(kT_a) = \int_{-\frac{f_a}{2}}^{\frac{f_a}{2}} A_w(jf)\, e^{j2\pi f k T_a} df \tag{2.6.81}$$

aus dem Frequenzgang $A_w(jf)$ gegeben. man erhält die gewünschten Koeffizienten, indem man (2.6.80) und (2.6.81) gleichsetzt.
Das resultierende Filter muss nach Tietze und Schenk[?] normalisieret und auf die gewünschte Grenzfrequenz umgerechnet werden.

### 2.6.2.5. IIR-Filter

Die zweite Klasse digitaler Filter stellen die rekursiven Filter oder **I**nfinite **I**mpulse **R**esponse-Filter dar. Da Ihre Impulsantwort wie die der analogen Filter unendlich lange dauert, kann man die analogen Filter auf die digitalen abbilden.
Zur Abbildung eines analogen Filters auf ein digitales Filter verwendet man die bilineare Transformation[?][?]. Um den Freqenzbereich $[0 \ldots \infty]$ des analogen Filters auf den Frequenzbereich $[0 \ldots \frac{f_a}{2}]$ des digitalen Filters abgebildet.

$$f = \frac{f_a}{\pi} \tan \frac{\pi f'}{f_a} \tag{2.6.82}$$

Wie gewünscht strebt für $f \to \infty$ die digitale Frequenz $f' \to \frac{f_a}{2}$. Dabei ist die Verzerrung der Frequenzachse umso geringer, je geringer die Frequenz im Vergleich zur Taktfrequenz ist. Wir rechnen wieder mit normierten Frequenzen

$$\begin{aligned} F &= \frac{f}{f_a} \\ F_g &= \frac{f_g}{f_a} \end{aligned} \tag{2.6.83}$$

Die Transformationsgleichung für die normierten Frequenzen ist dann

$$F = \frac{1}{\pi} \tan \pi F' \tag{2.6.84}$$

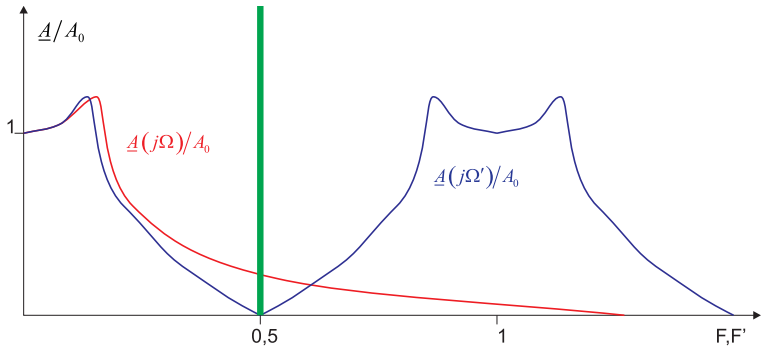



Abbildung 2.65.: Transformation eines Tschebyscheff-Tiefpasses auf einen periodischen Frequenzgang

Bei der Transformation eines Tschebyschefftiefpasses (Abbildung 2.65) bleibt das charakteristische Überschwingen erhalten. Damit die Grenzfrequenz des ursprünglichen Tiefpasses und des neuen Tiefpasses gleich bleibt, ersetzen wir in Gleichuing (2.6.84) den Vorfaktor $\frac{1}{\pi}$.

$$F = \frac{F_g}{\tan \pi F_g} \tan \pi F' = F_g \ell \tan \pi F' \tag{2.6.85}$$

Man erkennt in Abbildung 2.65, dass nun der Frequenzgang des digitalen Filters erst über der Grenzfrequenz $F_g$ abweicht. Die Variable $P$ in der normierten Übertragungsfunktion wird

$$P = \ell j \tan \pi F \tag{2.6.86}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Unter Verwendung der Umformung

$$j \tan x = -\tanh(-jx) = \frac{1 - e^{-2jx}}{1 + e^{2jx}} \tag{2.6.87}$$

sowie mit der Definition von $z$ ist

$$P = \ell \frac{1 - e^{-2jx}}{1 + e^{2jx}} = \ell \frac{1 - z^{-1}}{1 + z^{-1}} \tag{2.6.88}$$

Die Form der obigen Gleichung erklärt den Namen ¨Bilineare Transformation¨. Sie bildet den Arbeitsbereich eines analogen Filters auf den eines digitalen ab. Abweichungen sind desto geringer, je grösser die Abtastfrequenz $f_a$ im Vergleich zu den interessierenden Frequenzen ist. Aus

$$A(P) = \frac{d_0 + d_1 P + d_2 P^2 + \ldots}{c_0 + c_1 P + c_2 P^2 + \ldots} = \frac{\sum_{k=0}^{n} d_k P^k}{\sum_{k=0}^{n} c_k P^k} \tag{2.6.89}$$

Wenn man die bilineare Transformation oben einsetzt, erhält man

$$A(P) = \frac{\sum_{k=0}^{n} d_k \left(\ell \frac{1-z^{-1}}{1+z^{-1}}\right)^k}{\sum_{k=0}^{n} c_k \left(\ell \frac{1-z^{-1}}{1+z^{-1}}\right)^k} \tag{2.6.90}$$

Durch Koeffizientenvergleich erhält man schliesslich die Koeffizienten der Digitalfilter

$$A(z) = \frac{\alpha_0 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \ldots}{\beta_0 + \beta_1 z^{-1} + \beta_2 z^{-2} + \ldots} = \frac{\sum_{k=0}^{n} \alpha_k z^{-k}}{\sum_{k=0}^{n} \alpha_k z^{-k}} \tag{2.6.91}$$

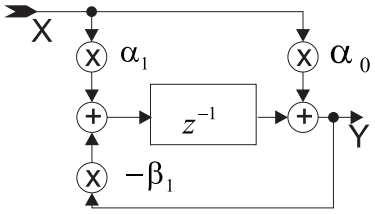



Abbildung 2.66.: Struktur eines IIR-Filters erster Ordnung

**2.6.2.5.1. IIR-Filter erster Ordnung** Für das in der Abbildung 2.66 dargestellte IIR-Filter erster Ordnung soll der analoge Frequenzgang (schliesst **alle** Filterarten ein)

$$A(P) = \frac{d_0 + d_1 P}{c_0 + c_1 P} \tag{2.6.92}$$

auf das IIR-Filter

$$\tilde{A}(z) = \frac{Y}{X} = \frac{\alpha_0 + \alpha_1 z^{-1}}{1 + \beta_1 z^{-1}} \tag{2.6.93}$$

Der Koeffizientenvergleich ergibt

$$\begin{aligned} \alpha_0 &= \frac{d_0 + d_1\ell}{c_0 + c_1\ell} \\ \alpha_1 &= \frac{d_0 - d_1\ell}{c_0 + c_1\ell} \\ \beta_1 &= \frac{c_0 - c_1\ell}{c_0 + c_1\ell} \end{aligned} \tag{2.6.94}$$

Die obigen Gleichungen angewandt auf einen allgemeinen Tiefpass $A(P) = \frac{A_0}{1+a_1P}$ ergeben

$$\begin{aligned} \alpha_0 &= \frac{A_0 l}{1 + a_1\ell} \\ \alpha_1 &= \frac{A_0 l}{1 + a_1\ell} \\ \beta_1 &= \frac{1 - a_1\ell}{1 + a_1\ell} \end{aligned} \tag{2.6.95}$$

Der Hochpass $A(P) = \frac{A_\infty P}{1+a_1P}$ wird in der digitalen Implementation

$$\begin{aligned} \alpha_0 &= \frac{A_\infty \ell}{a_1 + \ell} \\ \alpha_1 &= -\frac{A_\infty \ell}{a_1 + \ell} \\ \beta_1 &= \frac{a_1 - \ell}{1_1 + \ell} \end{aligned} \tag{2.6.96}$$

Bei einem digitalen Filter wird die Grenzfrequenz durch die Abtastfrequenz festgelegt. Dies bietet eine einfache Möglichkeit der Abstimmung des Filters.

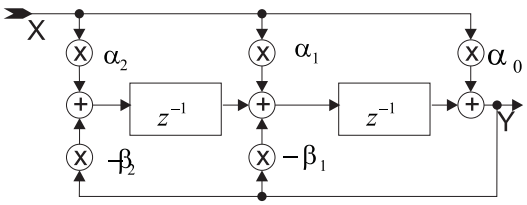



Abbildung 2.67.: Allgemeines IIR-Filter zweiter Ordnung

**2.6.2.5.2. IIR-Filter zweiter Ordnung** Die allgemeine Form eines analogen Filters zweiter Ordnung

$$A\left(P\right) = \frac{d_0 + d_1P + d_2P^2}{c_0 + c_1P + c_2Pr} \tag{2.6.97}$$

wird durch die bilineare Transformation zu

$$\tilde{A}\left(z\right) = \frac{\alpha_0 + \alpha_1 z^{-1} + \alpha_2 z^{-2}}{1 + \beta_1 z^{-1} + \beta_2 z^{-2}} \tag{2.6.98}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

der Übertragungsfunktion des digitalen IIR-Filters aus Abbildung 2.67. Die Koeffizienten sind:

$$\begin{aligned}
\alpha_0 &= \frac{d_0 + d_1\ell + d_2\ell^2}{c_0 + c_1\ell + c_2\ell^2} \\
\alpha_1 &= \frac{2\,(d_0 - d_2\ell^2)}{c_0 + c_1\ell + c_2\ell^2} \\
\alpha_2 &= \frac{d_0 - d_1\ell + d_2\ell^2}{c_0 + c_1\ell + c_2\ell^2} \\
\beta_1 &= \frac{2\,(c_0 - c_2\ell^2)}{c_0 + c_1\ell + c_2\ell^2} \\
\beta_2 &= \frac{c_0 - c_1\ell + c_2\ell^2}{c_0 + c_1\ell + c_2\ell^2}
\end{aligned} \tag{2.6.99}$$

Aus dem Tiefpass $A(P) = \frac{A_0}{1+a_1P+b_1P^2}$ erhält man das äquivalente IIR Filter mit den Koeffizienten

$$\begin{aligned}
\alpha_0 &= \frac{A_0}{1 + a_1\ell + b_1\ell^2} \\
\alpha_1 &= \frac{2A_0}{1 + a_1\ell + b_1\ell^2} = 2\alpha_0 \\
\alpha_2 &= \frac{A_0}{1 + a_1\ell + b_1\ell^2} = \alpha_0 \\
\beta_1 &= \frac{2\,(1 - b_1\ell^2)}{1 + a_1\ell + b_1\ell^2} \\
\beta_2 &= \frac{1 - a_1\ell + b_1\ell^2}{1 + a_1\ell + b_1\ell^2}
\end{aligned} \tag{2.6.100}$$

Der Hochpass $A(P) = \frac{A_\infty P^2}{1+a_1P+b_1P^2}$ wird zu

$$\begin{aligned}
\alpha_0 &= \frac{A_\infty b_1\ell^2}{1 + a_1\ell + b_1\ell^2} \\
\alpha_1 &= -\frac{2A_\infty b_1\ell^2}{1 + a_1\ell + b_1\ell^2} = -2\alpha_0 \\
\alpha_2 &= \frac{A_\infty b_1\ell^2}{1 + a_1\ell + b_1\ell^2} = \alpha_0 \\
\beta_1 &= \frac{2\,(1 - b_1\ell^2)}{1 + a_1\ell + b_1\ell^2} \\
\beta_2 &= \frac{1 - a_1\ell + b_1\ell^2}{1 + a_1\ell + b_1\ell^2}
\end{aligned} \tag{2.6.101}$$

Weiter wird der Bandpass $A(P) = \frac{A_r\frac{P}{Q}}{1+\frac{P}{Q}+P^2}$ transformiert in

$$\begin{aligned}
\alpha_0 &= \frac{\ell\frac{A_r}{Q}}{1+\frac{\ell}{Q}+\ell^2} \\
\alpha_1 &= 0 \\
\alpha_2 &= -\frac{\ell\frac{A_r}{Q}}{1+\frac{\ell}{Q}+\ell^2} = -\alpha_0 \\
\beta_1 &= \frac{2\left(1-\ell^2\right)}{1+\frac{\ell}{Q}+\ell^2} \\
\beta_2 &= \frac{1-\frac{\ell}{Q}+\ell^2}{1+\frac{\ell}{Q}+\ell^2}
\end{aligned} \tag{2.6.102}$$

Die Bandsperre $A(P) = \frac{A_0\left(1+P^2\right)}{1+\frac{P}{Q}+P^2}$ transformiert sich in

$$\begin{aligned}
\alpha_0 &= \frac{A_0\left(1+\ell^2\right)}{1+\frac{\ell}{Q}+\ell^2} \\
\alpha_1 &= \frac{2A_0\left(1-\ell^2\right)}{1+\frac{\ell}{Q}+\ell^2} \\
\alpha_2 &= \frac{2\left(1+\ell^2\right)}{1+\frac{\ell}{Q}+\ell^2} = \alpha_0 \\
\beta_1 &= \frac{A_0\left(1-\ell^2\right)}{1+\frac{\ell}{Q}+\ell^2} \\
\beta_2 &= \frac{1-\frac{\ell}{Q}+\ell^2}{1+\frac{\ell}{Q}+\ell^2}
\end{aligned} \tag{2.6.103}$$

Endlich wird aus dem Allpass $A(P) = \frac{1+a_1P+b_1P^2}{1+a_1P+b_1P^2}$ das folgende digitale IIR-Filter

$$\begin{aligned}
\alpha_0 &= \frac{1+a_1\ell+b_1\ell^2}{1-a_1\ell+b_1\ell^2} \\
\alpha_1 &= \frac{2\left(1-b_1\ell^2\right)}{1-a_1\ell+b_1\ell^2} \\
\alpha_2 &= \frac{1-a_1\ell+b_1\ell^2}{1-a_1\ell+b_1\ell^2} = 1 \\
\beta_1 &= \frac{2\left(1-b_1\ell^2\right)}{1-a_1\ell+b_1\ell^2} = \alpha_1 \\
\beta_2 &= \frac{1+a_1\ell+b_1\ell^2}{1-a_1\ell+b_1\ell^2} = \alpha_0
\end{aligned} \tag{2.6.104}$$

## 2.7. Modulationstheorie

Um Signale über grössere Distanzen transportieren zu können, wird das Nutzsignal oft auf einen Träger aufmoduliert. Grob gesagt, kann man so ein *Signal* in

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

einem Bereich mit schlechten Ausbreitungseigenschaften in einen Bereich mit längerreichweitiger Übertragungsmöglichkeit transferieren. Oder es können in einem Kabel mehrere bis sehr viele Telefongespräche gleichzeitig geführt werden. Modulation tritt auf, wenn zwei oder mehrere Signale mit unterschiedlichen oder gleichen Frequenzen durch ein nichtlineares Bauteil laufen. Liegen an einer Strecke mit der Strom-Spannungscharakteristik

$$I(t) = aU(t) + bU^2(t) \tag{2.7.1}$$

die zwei Signale $U_1 = \hat{U}_1 \cos \omega_1 t$ und $U_2 = \hat{U}_2 \cos \omega_2 t$ an, dann treten Terme bei den Summen- und Differenzfrequenzen auf. **Man beachte, dass mit reellen Grössen gerechnet werden muss. Werden die bequemen komplexen Darstellungen verwendet, muss immer auch das konjugiert-komplexe mitgenommen werden, also effektiv auch eine reelle Zahl.** Wir erhalten also

$$\begin{aligned} I(t) &= a\left(\hat{U}_1 \cos \omega_2 t + \hat{U}_2 \cos \omega_2 t\right) + b\left(\hat{U}_1 \cos \omega_2 t + \hat{U}_2 \cos \omega_2 t\right)^2 \\ &= a\hat{U}_1 \cos \omega_2 t + a\hat{U}_2 \cos \omega_2 t + \\ &\quad b\hat{U}_1^2 \cos^2 \omega_1 t + b\hat{U}_2^2 \cos^2 \omega_2 t + 2b\hat{U}_1\hat{U}_2 \cos \omega_1 t \cos \omega_2 t \\ &= \frac{b}{2}\left(\hat{U}_1^2 + \hat{U}_2^2\right) + \\ &\quad a\hat{U}_1 \cos \omega_2 t + a\hat{U}_2 \cos \omega_2 t + \\ &\quad \frac{b}{2}\left(\hat{U}_1^2 \cos 2\omega_1 t + \hat{U}_2^2 \cos 2\omega_2 t\right) + \\ &\quad b\hat{U}_1\hat{U}_2 \left[\cos\left(\omega_1 - \omega_2\right) t + \cos\left(\omega_1 + \omega_2\right)\right] \end{aligned} \tag{2.7.2}$$

neben den beiden ursprünglichen Signalen treten auch ein Gleichstromanteil sowie die doppelten, die Summen- und die Differenzfrequenzen auf. Eine Nichtlinearität von höherer Ordnung würde entsprechend den Produkt- und Summenregeln für Winkelfunktionen noch mehr Frequenzen ergeben.
Die oben gezeigte Rechnung illustriert die Amplitudenmodulation. Allgemein kann man schreiben:

$$I_{AM} = \hat{I}(s) \cos \omega_t t \tag{2.7.3}$$

$\omega_t$ ist die Trägerfrequenz. $\hat{I}(s)$ hängt nun vom aufmodulierten *Signal* ab:

$$\hat{I}(s) = \hat{I}_t + \hat{I}_s \cos \omega_s t \tag{2.7.4}$$

$\hat{I}_t$ ist die Intensität des Trägers, $\hat{I}_s$ diejenige des aufmodulierten Signals mit der Frequenz $\omega_t$. Typischerweise ist $\omega_s \ll \omega_t$. Mit der Abkürzung $m = \frac{\hat{I}_s}{\hat{I}_t}$ ergibt sich

$$I_{AM} = \hat{I}_t \left(1 + \frac{\hat{I}_s}{\hat{I}_t} \cos \omega_s t\right) \cos \omega_t t = \hat{I}_t \left(1 + m \cos \omega - st\right) \cos \omega_t t \tag{2.7.5}$$

$m$ heisst auch der Modulationsgrad. Anwendung der Additionssätze für Winkelfunktionen ergibt:

$$I_{AM} = \hat{I}_t \cos \omega_t t + \frac{\hat{I}_t m}{2} \left[\cos\left(\omega_t - \omega_s\right) t + \cos\left(\omega_t + \omega_s\right) t\right] \tag{2.7.6}$$

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

Das resultierende *Signal* enthält also sowohl die Summen- wie auch die Differenzfrequenzen. man ersieht aus der Herleitung, dass man zwar für eine Amplitudenmodulation eine quadratisch-nichtlineare Kennlinie nehmen könnte, dass es aber geschickter ist, einen Multiplizierer zu verwenden. Weiter ersieht man, dass um ein *Signal* der Frequenz $\omega_s$ zu übertragen, man eine Bandbreite

$$B = 2\frac{\omega_s}{2\pi} \tag{2.7.7}$$

benötigt wird. Gleichung (2.7.6) zeigt weiter, dass bei einem Modulationsgrad von 1 die Hälfte der Sendeenergie im Träger steckt und dass je ein Viertel in den beiden Seitenbändern vorhanden ist. Für eine Radioübertragung ist aber nur der Energiegehalt in den Seitenbändern wichtig. Deshalb gibt es Sendeverfahren, bei denen der Träger unterdrückt oder sogar mit dem Träger das eine Seitenband nicht gesendet wird. Die Stereoinformation bei einer Stereo-UKW-Sendung wird mit diesem Verfahren bandbreitensparend übertragen. Zur Wiederherstellung des Signals benötigt man den träger. Sind beide Seitenbänder vorhanden, kann man einen Oszillator im Mittel mit dem *Signal* mitlaufen lassen (das *Signal* hat, gemittelt, gerade die Trägerfrequenz). Bei Einseitenbandmodulation (SSB für Single Sideband Modulation) verwendet man entweder einen Pilotton (UKW-Stereo) oder man ist darauf angewiesen, das der Sender und der Empföänger sehr stabil laufen. Als Kuriosum sei erwähnt, dass eine Möglichkeit Sprache zu verschlüsseln, darin besteht, das Sprachsignal in Bänder aufzuteilen und mit SSB in der Frequenz zu verschieben. Die Frequenzverschiebung kann auch dynamisch sein.
Die Amplitudenmodulation ist eher störanfällig. Deshalb werden qualitativ hochwertigere Dienste mit Frequenzmodulation ausgestrahlt. Die Trägerfrequenz $\omega_t$ wird mit der Signalfrequenz $\omega_s$ moduliert. Das heisst, die Phase des Signals bewegt sich nicht mehr mit konstanter Geschwindigkeit. Also schreibt man anstelle von $\omega t = \varphi$, der phase, das Integral

$$\int_0^t (\omega_t + \Delta\omega_t \sin\omega_s t)\, dt = \omega_t t + \frac{\Delta\omega_t}{\omega_s}\cos\omega_s t \tag{2.7.8}$$

$\Delta\omega_t$ ist der Frequenzhub der Modulation. Das frequenzmodulierte *Signal* ist also:

$$\begin{aligned} I_{FM} &= \hat{I}_t \sin\left(\omega_t t + \frac{\Delta\omega_t}{\omega_s}\sin\omega_s t\right) \\ &= \hat{I}_t \sum_{n=-\infty}^{\infty} \hat{I}_n\,(m_F)\sin\left(\omega_t + n\omega_s\right)t \end{aligned} \tag{2.7.9}$$

Die Amplituden $\hat{I}_n\,(m_F)$ der einzelnen Teilfrequenzen $\omega_t + n\omega_s$ sind Bessel-Funktionen $n$-ter Ordnung. das Argument $m_F = \frac{\Delta\omega_t}{\omega_s}$. Man benötigt für die *FM-Übertragung* eine Bandbreite von

$$B \geq \frac{1}{\pi}\left(\Delta\omega_t + \Delta\omega_s\right) = \frac{\Delta\omega_s}{\pi}\left(1 + m_F\right) \tag{2.7.10}$$

Der Frequenzhub bei der FM hängt nicht von der Signalfrequenz, sondern nur von der Amplitude ab. Anstelle der Frequenz kann man auch die Phase modulieren. Das *Signal* sieht dann folgendermassen aus:

$$I_{PM} = \hat{I}_t \sin\left(\omega_t t + \Delta\varphi \sin\omega_s t\right) \tag{2.7.11}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Hier hängt der Phasenhub von der Amplitude ab, der resultierende Frequenzhub ist aber durch die Signalfrequenz bestimmt.

## 2.8. Rauschen

Mit Rauschen bezeichnet man die durch stochastische Prozesse bedingte Schwankung einer Grösse. Rauschen tritt nicht nur in der Elektronik auf, sondern in allen Vielteilchensystemen. So ist, zum Beispiel, die Brownsche Bewegung ein Rauschprozess. In der Elektronik betrachtet man den Transport von Strom, also einen Fluss einzelner Elektronen. Da diese, wie alle anderen Vielteilchensysteme den Gesetzen der statistischen Physik gehorchen müssen, tritt Rauschen auf. Man unterscheidet viele verschiedene Arten von Rauschen wie das Widerstandsrauschen oder das Schrotrauschen. In diesem Abschnitt soll zuerst das Widerstandsrauschen nach dem Buch von Reif[**?**] abgeleitet werden, es folgen dann andere relevante Rauschprozesse.

### 2.8.1. Widerstandsrauschen

Wir betrachten einen elektrischen Widerstand $R$, der an den Eingang eines idealen Verstärkers angeschlossen sei, der zwischen $\omega_1$ und $\omega_2$ eine konstante Verstärkung habe. Ausserhalb dieses Durchlassbandes sei die Verstärkung 0.
Die Elektronen im Widerstand ändern ihre Position und Geschwindigkeit durch zufällige thermische Fluktuationen. Also existiert ein fluktuierender Strom $I(t)$ und als Konsequenz auch eine fluktuierende *EMF* $U(t)$. Wenn wir die *Fouriertransformation* von $V(t)$ kennen, können wir nach Gleichung (2.4.17) die Varianz der *EMF* schreiben (Dar Mittelwert der *EMF* ist null, da wir keine treibende Potentialdifferenz annehmen)

$$\left\langle U^2(t)\right\rangle = \int_0^\infty J_+(\omega)\, d\omega \tag{2.8.1}$$

Wie im Abschnitt 2.4.2.0.1 gezeigt, ist $J_+(\omega)$ das Leistungsspektrum der EMF $U(t)$. Reif [**?**] zeigt im Abschnitt 15.8 mit der Langevin-Funktion, dass an einem Widerstand (analog zur Brownschen Bewegung) gilt

$$R = \frac{1}{2kT}\int_{-\infty}^{\infty} \langle V(0)\, V(s)\rangle_0\, ds \tag{2.8.2}$$

Diese Gleichung kann mit den Gleichungen (2.4.19) und (2.4.17) umgeschrieben werden (die Korrelationsfunktion ist unabhängig von t, und $e^{j\omega s} = 1$ wenn $\omega = 0$)

$$R = \frac{1}{2kT}\left[2\pi J(0)\right] \tag{2.8.3}$$

Die *Korrelationszeit* $\tau^*$ der *Fluktuationen* ist von der Grössenordnung der Zeit zwischen zwei Stössen eines Elektrons mit dem Gitter. Also ist $K(s) = \langle V(0)V(s)\rangle_0 = 0$ für alle $|s| \gg \tau^*$. Die *Korrelationsfunktion* ist in der Nähe von 0 konzentriert. Also ist im Bereich der nichtverschwindenden *Korrelationsfunktion* $e^{j\omega s} = 1$ für alle $\omega\tau^* \ll 1$. Für alle diese $\omega$-Werte hat das Integral den gleichen Wert

$$J(\omega) = J(0) \tag{2.8.4}$$

Das *Leistungsspektrum* der fluktuierenden *EMF* $U(t)$ ist also konstant. Man kann ziemlich allgemeingültig festhalten:

> **Kurze Korrelationszeiten bedingen breite Spektren, und umgekehrt.**

Diese Aussage ist übrigens analog zum *Heisenbergschen Unschärfeprinzip*.
Wir erhalten damit das *Leistungsspektrum* der Rausch-EMF an einem *Widerstand*

$$J_+(\omega) = \frac{2}{\pi} kTR \qquad für\ \omega \ll \frac{1}{\tau^*} \tag{2.8.5}$$

Die obige Gleichung ist das *Nyquist-Theorem*. Es ist ein Spezialfall der viel allgemeineren Verbindung zwischen Fluktuation und Dissipation. Zum Beispiel gibt es bei der *Brownschen Bewegung* eine ähnliche Beziehung zwischen dem Spektrum der fluktuierenden Kraft $F$ und dem Reibungskoeffizienten $\alpha$. Interessierte Leser mögen Literatur über das ¨*Fluktuations-Dissipations-Theorem*¨ lesen. Rauschen mit einer von der Frequenz unabhängigen spektralen Dichte heisst *weisses Rauschen*.

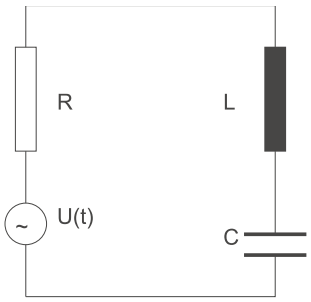



Abbildung 2.68.: Der RLC-Kreis für Rauschuntersuchungen

Wir wollen nun untersuchen, ob das Nyquist-Theorem konsistent mit der Thermodynamik von Gleichgewichtszuständen ist. Wir verwenden dazu die Schaltung aus Abbildung 2.68. Das System sei im thermischen Gleichgewicht bei der Temperatur $T$. Stromfluktuationen $I(t)$ haben ihre Ursache in der stochastischen EMF im Widerstand $R$. Wir betrachten die Fourierkomponente $U_0(\omega)\, e^{j\omega t}$ der *EMF* und verwenden die Differentialgleichung

$$L\frac{dI}{dt} + RI + \frac{1}{C}\int I dt = U(t) \tag{2.8.6}$$

Für die Frequenz $\omega$ erhält man

$$I_0(\omega) = \frac{U_0(\omega)}{Z(\omega)} \quad \text{wobei } Z(\omega) = R + j\left(\omega L + \frac{1}{\omega C}\right) \tag{2.8.7}$$

Zur Berechnung der in der *Induktivität* gespeicherten Energie betrachtet man den Strom $I$ als einen Parameter eines thermodynamischen Systems mit der freien Energie $\Delta F$ (Siehe auch das Buch von Reif[**?**]). Die Wahrscheinlichkeit $P(I)\, dI$ dass der Strom in einer Schaltung im Gleichgewicht mit der Temperatur $T$ einen Wert zwischen $I$ und $I + dI$ annimmt, ist proportional zu $e^{\Delta F/kT}$. $\Delta F$ ist die freie Energie im Zustand $I = 0$. Die Bewegung von Elektronen, ohne dass Ladung auf- oder abgebaut wird, ändert die *Entropie* $S$ nicht. Also ist $\Delta S = 0$ und damit

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

$\Delta F = \Delta E - T\Delta S = \Delta E$. Da Energie durch einen Strom nur in einer *Induktivität* gespeichert werden kann, gilt:

$$P(I)\,dI \propto e^{\Delta E/kT} dI = e^{-LI^2/2kT} dI \tag{2.8.8}$$

Daraus leitet man unter Verwendung elementarer Gesetze der Statistik ab

$$\left\langle \frac{1}{2} LI^2 \right\rangle = \frac{\int_{-\infty}^{\infty} P(I)\, I^2 dI}{\int_{-\infty}^{\infty} P(I)\, dI} = \frac{1}{2} kT \tag{2.8.9}$$

Analog gilt auch für einen *Kondensator*, dass im Mittel im thermischen Gleichgewicht gilt

$$\left\langle \frac{1}{2} CU^2 \right\rangle = \frac{1}{2} kT \tag{2.8.10}$$

Der Ausdruck für die mittlere in der Induktivität $L$ gespeicherte Energie kann in eine Fourierreihe entwickelt werden:

$$\begin{aligned}
\left\langle \frac{1}{2} LI^2 \right\rangle &= \frac{1}{2} L \frac{\pi}{\theta} \int_{-\infty}^{\infty} |I_0(\omega)|^2 \, d\omega \\
&= \frac{1}{2} L \frac{\pi}{\theta} \int_{-\infty}^{\infty} \frac{|U_0(\omega)|^2}{|Z(\omega)|^2} d\omega \\
&= \frac{1}{2} L \int_{-\infty}^{\infty} \frac{J(\omega)}{|Z(\omega)|^2} d\omega \\
&= \frac{1}{2} L \int_{0}^{\infty} \frac{J_+(\omega)}{|Z(\omega)|^2} d\omega
\end{aligned} \tag{2.8.11}$$

Dabei haben wir die Definition des *Leistungsspektrums* nach Gleichung (2.4.29) verwendet.

$$J_+(\omega) \equiv 2J(\omega) \equiv \frac{2\pi}{\theta} |V(\omega)|^2 \tag{2.8.12}$$

Die Gleichgewichtsbedingung (2.8.9) verlangt, dass unter Verwendung von (2.8.7) gilt:

$$\left\langle I^2 \right\rangle = \int_0^{\infty} \frac{J_+(\omega)\, d\omega}{R^2 + \left(\omega L + \frac{1}{\omega C}\right)^2} = \frac{kT}{L} \tag{2.8.13}$$

Umgeformt mit $\omega_0^2 = 1/LC$ bekommt man

$$\frac{1}{R^2} \int_0^{\infty} \frac{J_+(\omega)\, d\omega}{1 + \left(\frac{L}{\omega R}\right)^2 (\omega^2 + \omega_0^2)^2} = \frac{kT}{L} \tag{2.8.14}$$

Wir nehmen an, dass $L$ sehr gross ist, dass also der Schwingkreis eine sehr grosse Güte hat. Dann kann man $J_+(\omega) = J_+(\omega_0)$ aus dem Integral herausziehen. Ebenso

ist dann $\omega^2 - \omega_0^2 = (\omega + \omega_0)(\omega - \omega_0) \approx 2\omega_0(\omega - \omega_0)$ und $\frac{L}{\omega R} \approx \frac{L}{\omega_0 R}$. Mit der Abkürzung $\eta = \omega - \omega_0$ erhalten wir

$$\begin{aligned}\frac{kT}{L} &= \frac{J_+(\omega_0)}{R^2}\int_0^\infty \frac{d\omega}{1+\left(\frac{2L}{R}\right)^2(\omega-\omega_0)^2} \\ &= \frac{J_+(\omega_0)}{R^2}\int_{-\infty}^\infty \frac{d\eta}{1+\left(\frac{2L}{R}\right)^2\eta^2} \\ &= \frac{J_+(\omega_0)}{R^2}\left[\pi\frac{R}{2L}\right] \end{aligned} \tag{2.8.15}$$

Wir erhalten daraus das Resultat

$$J_+(\omega_0) = \frac{2}{\pi}kTR \tag{2.8.16}$$

Da $C$ in der Gleichung nicht explizit vorkommt, kann man damit jede Frequenz erreichen, wir haben also, auf eine andere Art und Weise das *Nyquist-Theorem* abgeleitet.

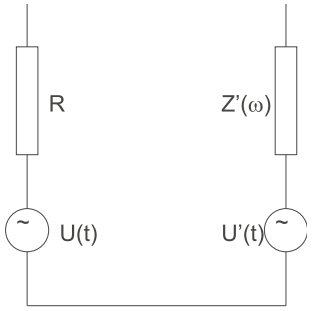



Abbildung 2.69.: Widerstand und allgemeine Impedanz im Gleichgewicht

In einem weiteren Beispiel wird, wie in Abbildung 2.69 im thermischen Gleichgewicht ein Widerstand $R$ mit einer allgemeinen Impedanz $Z'(\omega) = R'(\omega) + jX'(\omega)$ zusammengeschaltet. Sowohl der Widerstand $R'$ wie auch der reaktive Teil $X'$ sollen frequenzabhängig sein. Wir bezeichnen mit $U(t)$ die Spannung an $R$ und mit $U'(t)$ die Spannung an der Impedanz $Z'$. Im Gleichgewicht muss die mittlere Leistung $P'$, die durch die Impedanz $Z'$ wegen der Rauschspannung $U(t)$ am Widerstand $R$ absorbiert wird, gleich sein der Leistung $P$, die durch den Widerstand $R$ wegen der Rauschspannung $U'(t)$ der Impedanz $Z'$ absorbiert wird. Wir betrachten weiter ein enges Frequenzband zwischen $\omega$ und $\omega + d\omega$. Das Leistungsgleichgewicht muss für alle Frequenzbänder gelten. Nun erzeugt die Frequenzkomponente $U_0(\omega)$ in der Schaltung den Strom $I_0 = \frac{U_0}{R+Z'}$ Die Leistung $P'$ absorbiert durch $Z'$ ist dann

$$P' \propto |I_0|^2 R' = \left|\frac{U_0}{R+Z'}\right|^2 R' \tag{2.8.17}$$

$R'$ ist der (frequenzabhängige) Realteil der Impedanz $Z'$. Analog zur obigen Glei-

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

chung gilt auch

$$P \propto |I_0'|^2 R = \left| \frac{U_0'}{R + Z'} \right|^2 R \tag{2.8.18}$$

Also ist

$$|U_0|^2 R' = |U_0'|^2 R \tag{2.8.19}$$

oder

$$J_+(\omega) R'(\omega) = J_+'(\omega) R(\omega) \tag{2.8.20}$$

$J_+$ ist das Leistungsspektrum von $U(t)$ und $J_+'$ dasjenige von $U'(t)$. Also ist

$$J_+(\omega) = \frac{R'(\omega)}{R} \left( \frac{2}{\pi} kTR \right) = \frac{2}{\pi} kTR' \tag{2.8.21}$$

Das *Rauschspektrum* einer jeglichen Impedanz ist also immer mit dem Widerstandsteil dieser Impedanz verbunden. Ideale Spulen und Kondensatoren rauschen nicht. Deshalb kann man mit Hilfe parametrisch veränderter Kapazitäten sehr rauscharme Verstärker bauen.

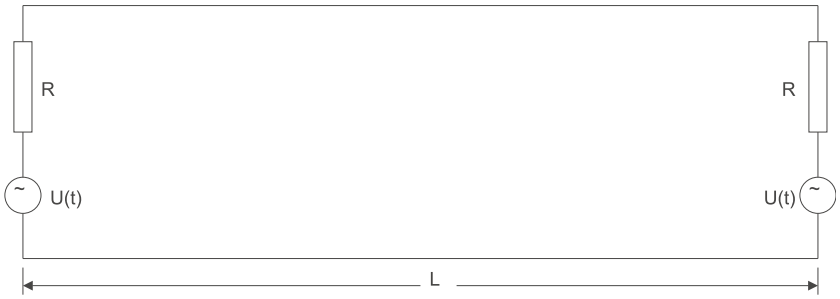



Abbildung 2.70.: Nyquists Anordnung zur Berechnung der Rauschspannung an einem Widerstand

*Nyquist* hatte sein Theorem aus Analogien der *Schwarzkörperstrahlung* abgeleitet. Abbildung 2.70 zeigt die von ihm verwendete Anordnung. Zwei Widerstände der Grösse $R$ sind über eine verlustfreie Leitung der Länge $L$ mit der Impedanz $R$ verbunden. Die ganze Anordnung ist im Gleichgewicht mit einem Wärmebad der Temperatur $T$. Durch die Impedanzanpassung wird jede vom linken Widerstand ausgehende Welle vom rechten vollständig absorbiert und umgekehrt. Die beiden Abschlusswiderstände sind also ein Analogon zum Schwarzen Körper. Eine Spannungswelle $U = U_0 e^{j(kr - \omega t)}$ breitet sich mit der Geschwindigkeit $c' = \frac{\omega}{k}$ aus. Zum Abzählen der möglichen Moden setzt man als Randbedingung $U(0) = U(L)$. Dann ist $kL = 2\pi n$ für jede ganze Zahl $n$. Die Anzahl Moden pro Einheitslänge im Frequenzintervall $\omega$ bis $\omega + d\omega$ ist dann

$$\Delta n = \frac{1}{2\pi} dk \tag{2.8.22}$$

Jede Mode hat eine Energie

$$\varepsilon(\omega) = \frac{\hbar\omega}{e^{\frac{\hbar\omega}{kT}} - 1} \to kT \quad \text{für } \hbar\omega \ll kT \tag{2.8.23}$$

Wir verwenden nun, dass im Gleichgewicht in jedem Frequenzintervall $\omega$ bis $\omega+d\omega$ die emittierte und die absorbierte Leistung eines Widerstandes gleich sein muss. Da es $\frac{1}{2\pi}\frac{d\omega}{c'}$ propagierende Moden pro Einheitslänge gibt, ist die auf einen Widerstand eintreffende Leistung

$$P_i = c'\left(\frac{1}{2\pi}\frac{d\omega}{c'}\right)\varepsilon(\omega) = \frac{1}{2\pi}\varepsilon(\omega)\,d\omega \tag{2.8.24}$$

Die emittierte Leistung ist gleich. Diese Leistung wird in Form einer Rauschspannung $U$ erzeugt. Deshalb muss auch ein Strom $I = \frac{U}{2R}$ fliessen. Wir müssen $2R$ verwenden, da der Generator der Rauschspannung die Serieschaltung **zweier** Widerstände $R$ sieht. Also ist

$$R\left\langle I^2\right\rangle = R\left\langle\frac{V^2}{4R^2}\right\rangle = \frac{1}{4R}\left\langle V^2\right\rangle = \frac{1}{4R}\int_0^\infty J_+(\omega)\,d\omega \tag{2.8.25}$$

Die Leistung im Frequenzintervall $\omega$ bis $\omega+d\omega$ ist dann

$$P_i' = \frac{J_+(\omega)}{4R}d\omega \tag{2.8.26}$$

Wenn man nun $P_i' = P_i$ setzt, dann bekommt man

$$\begin{aligned}\frac{1}{4R}J_+(\omega)\,d\omega &= \frac{1}{2\pi}\varepsilon(\omega)\,d\omega\\ J_+(\omega) &= \frac{2}{\pi}\frac{\hbar\omega}{e^{\frac{\hbar\omega}{kT}}-1}R\end{aligned} \tag{2.8.27}$$

Dies ist die Gleichung für das Rauschen unter Einbezug der quantenmechanischen Korrekturen. Für die üblichen Frequenzen, auch im Mikrowellenbereich, gilt $\hbar\omega\ kT$. Also wird aus Nyquists quantenmechanisch korrekter Formel

$$J_+(\omega) = \frac{2}{\pi}kTR \tag{2.8.28}$$

Betrachtet man nun das Rauschen in einem bestimmten Frequenzband bei genügend tiefen Frequenzen, dann ist

$$\begin{aligned}\left\langle V^2\right\rangle\Big|_{\omega_{\min}}^{\omega_{\max}} &= \int_{\omega_{\min}}^{\omega_{\max}} J_+(\omega)\,d\omega\\ &= \int_{\omega_{\min}}^{\omega_{\max}} J_+(0)\,d\omega\\ &= J_+(0)\int_{\omega_{\min}}^{\omega_{\max}} d\omega\\ &= J_+(0)\,(\omega_{\max}-\omega_{\min})\\ &= \frac{2\,(\omega_{\max}-\omega_{\min})\,kTR}{\pi}\\ &= 4BkTR\end{aligned} \tag{2.8.29}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

wobei $B = \frac{\omega_{\max} - \omega_{\min}}{2\pi}$ die Bandbreite der Detektion ist.
Das *Widerstandsrauschen* ist besonders bei breitbandigen Schaltungen der limitierende Teil.

## 2.8.2. Weitere Rauschquellen

### 2.8.2.1. Schroteffekt

In Elektronenröhren ist die statistische Fluktuation des Emissionszeitpunktes von Elektronen aus der Kathode (Dies würde übrigens auch für Feldemissionskathoden gelten) Ursache eines Rauschens. Schottky hat für dieses Rauschen die folgende Formel angegeben:

$$I_{schr}^2 = 2eI_a\Delta\nu \tag{2.8.30}$$

$I_a$ ist hier der *Anodenstrom*. Wie beim thermischen Rauschen ist die Ursache die Quantisierung der Ladung. Der *Schroteffekt* erlaubt, über eine Messung des *Rauschstrom*es die *Elementarladung* zu bestimmen.

### 2.8.2.2. Generations-Rekombinationsrauschen

Dieser Typ Rauschen tritt in Halbleitern auf. Der Rauschstrom, der sich aus der diskreten Natur von Elektronen und Löchern ergibt ist:

$$i_R^2 = A\left(\nu, T\right) E^2\Delta\nu \tag{2.8.31}$$

$E$ ist die elektrische Feldstärke im Kristall. $A$ ist ein frequenz- und temperaturabhängiger Faktor.

### 2.8.2.3. Flickerrauschen und 1/f-Rauschen

Man beobachtet überall da wo Ladung transportiert wird ein *Rauschspektrum*, dessen Dichte sich wie

$$I_R^2 \sim \frac{1}{f} \tag{2.8.32}$$

verhält. Dieses Rauschen ist besonders bei langsamen Messungen sehr störend. Lock-In- Verstärker sind eine Möglichkeit, eine Messung aus dem mHz-Bereich, wo $1/f$-Rauschen dominant sein kann in den kHz-Bereich zu verschieben.

## 2.8.3. Einfluss von Filtern auf das Rauschen

Filter verändern die Rauschspannung anders als die Signalspannung (Siehe Dostal [**?**]). Um den Effekt zu verstehen legen wir eine Rauschspannung mit dem Spektrum $J_+\left(\omega\right)$ an den Eingang eines Filters mit der Übertragungsfunktion $\underline{A}\left(\omega\right)$. Das Signalspektrum $J_s\left(\omega\right)$ wird durch das Filter wie folgt modifiziert:

$$J_{s,a}\left(\omega\right) = \underline{A}\left(\omega\right) J_s\left(\omega\right) \tag{2.8.33}$$

betreibt man das obige Filter mit einer Rauschspannung, dann ist

$$\left\langle U_a^2\left(t\right)\right\rangle = \int_0^\infty \left|\underline{A}\left(\omega\right)\right|^2 J_+\left(\omega\right) d\omega \tag{2.8.34}$$

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

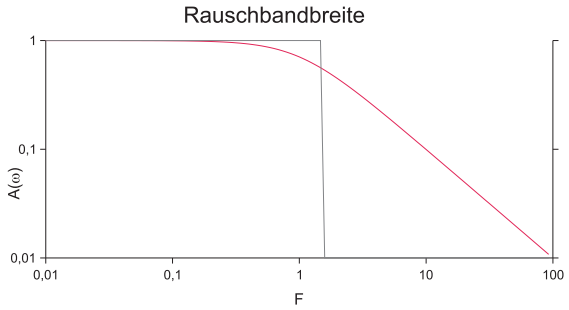



Abbildung 2.71.: *Signal-* und Rauschbandbreite an einem Tiefpassfilter erster Ordnung. Rot eingezeichnet ist der Frequenzgang des Filters. Die 3dB Bandbreite liegt hier bei 1. Die Rauschbandbreite ist mit schwarz eingezeichnet

Bei einer konstanten Verstärkung $A_0$ erhöht sich die Rauschspannung um den gleichen Wert $A_0$. Bei einem Filter, das mit weissem Rauschen gespiesen ist, kann man eine Rauschbandbreite definieren. Man nimmt (siehe Abbildung 2.71) an, man hätte ein Rechteckfilter mit der Verstärkung $A_m$ und der Bandbreite $\omega_m$ bis $\omega_m + \Delta\omega_m$. Nach dem Durchlauf dieses Filters soll die gemessene Rauschspannung gleich wie nach dem Passieren des zu untersuchenden Tiefpassfilters sein. Wir erhalten:

$$\left\langle U_{a,m}^2(t) \right\rangle = \int_{\omega_m}^{\omega_m+\Delta\omega_m} A_m^2 J_+(\omega)\, d\omega = A_m^2 \left\langle U^2(t) \right\rangle \Delta\omega_m \tag{2.8.35}$$

Für weisses Rauschen findet man nun

$$\Delta\omega_m = \frac{1}{A_m^2} \int_0^\infty |\underline{A}(\omega)|^2\, d\omega \tag{2.8.36}$$

Beim Tiefpass in Abbildung 2.71 ist die Übertragungsfunktion

$$|\underline{A}(\omega)| = \frac{1}{\sqrt{1+\omega^2 (RC)^2}} \tag{2.8.37}$$

Da die Verstärkung dieses Tiefpassfilters bei der Frequenz null den Wert eins hat, setzen wir $A_m = 1$ und erhalten

$$\Delta\omega_m = \int_0^\infty \frac{d\omega}{\sqrt{1+\omega^2 (RC)^2}} = \frac{\pi}{2}\frac{1}{RC} \tag{2.8.38}$$

Die Signalbandbreite des Filters ist definiert durch den 3dB-Abfall und gleich $\Delta\omega_s = \frac{1}{RC}$. Die Rauschbandbreite ist also um den Faktor $\frac{\pi}{2}$ grösser.

$$\Delta\omega_m = \frac{2}{\pi}\Delta\omega_s \tag{2.8.39}$$

Bei einem kritischen Tiefpass höherer Ordnung nähert sich die Signalbandbreite immer mehr der Rauschbandbreite. Ein Tiefpass $m$-ter Ordnung hat eine Über-

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

| Filterordnung m | $\frac{\delta\omega_m}{\Delta\omega_s}$ Kritischer Tiefpass | $\frac{\delta\omega_m}{\Delta\omega_s}$ Butterworth-Tiefpass |
|---|---|---|
| 1 | 1,571 | 1,571 |
| 2 | 1,220 | 1,111 |
| 3 | 1,155 | 1,047 |
| 4 | 1,129 | 1,026 |
| 5 | 1,114 | 1,017 |
| 6 | 1,105 | 1,012 |
| 7 | 1,098 | 1,008 |
| 8 | 1,094 | 1,004 |

Tabelle 2.18.: Verhältnis der Rausch- zur Signalbandbreite für kritische Filter und Butterworth-Filter

tragungsfunktion wie $|\underline{A}(\omega)| = \frac{1}{1(+x^2)^m}$. Das Integral zur Bestimmung der Rauschbandbreite hat die Rekursionslösung

$$\begin{aligned} \Delta\omega_m &= \Delta\omega_s\sqrt{2^{1/m}-1}J_m \\ J_1 &= \frac{\pi}{2} \\ J_m &= \frac{2m-1}{2\,(m-1)}J_{m-1} \end{aligned} \tag{2.8.40}$$

Das Verhältnis von Rausch- zur Signalbandbreite ist in Tabelle 2.18 aufgelistet. Zusätzlich sind auch die Werte für ein Butterworthfilter aufgelistet. Der Leser kann sich anhand des oben beschriebenen Verfahrens die Rauschbandbreiten für andere Filter leicht selber berechnen.

## 2.9. Digitale Signalprozessoren (DSP)

Digitale Signalprozessoren sind eine Klasse von Mikroprozessoren, die für die Berechnung digitaler Filter optimiert worden sind.

### 2.9.1. Klassische Rechner

Die Struktur eines klassischen Rechners nach John von Neumann[?] ist in der Abbildung 2.72 gezeigt. Bei dieser Computerarchitektur werden Programm und Daten im gleichen Speicher abgelegt. Daten werden über einen Datenbus vom Speicher zu dem Steuerwerk und dem Rechenwerk (oder umgekehrt) verschoben. Im Steuerwerk werden die Befehle aus dem Datenstrom dekodiert und in Steuersignale für das Rechenwerk oder den Speicher umgewandelt. Das Rechenwerk generiert, auch auf Grund von Rechenoperationen, die nächste Adresse.
Der detailierte Ablauf einer Berechnung wird in Abbildung 2.73 gezeigt.

1. Adresse des Befehls an den Speicher ausgeben.
2. Befehl ins Steuerwerk laden
3. Wenn der Befehl im Steuerwerk die Adresse der zu verarbeitenden Zahl enthält, diese laden

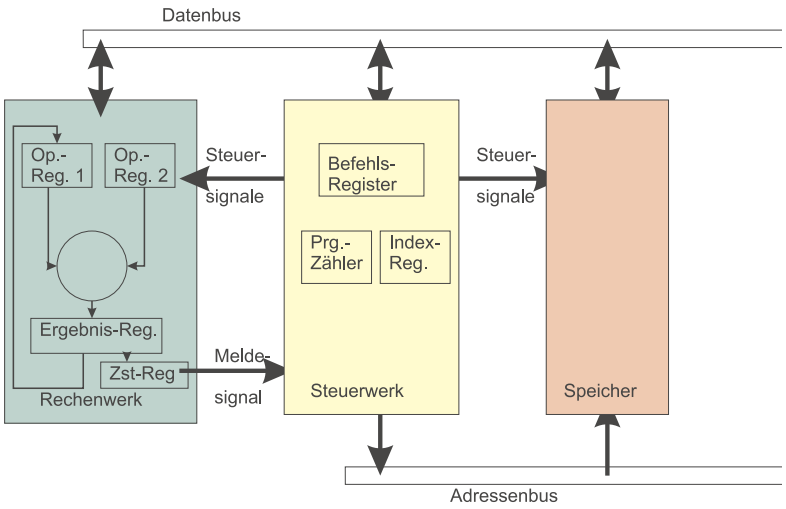



Abbildung 2.72.: von Neumannsche Struktur eines Computers. Abkürzungen: Op.-Reg.: Operationsregister; Ergebnis-Reg.: Ergebnisregister; Zst.-Reg.: Zustandsregister; Prg.-Zähler: Programmzähler; Index-Reg.: Indexregister

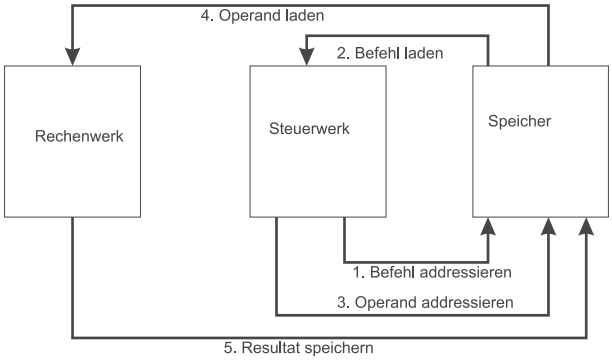



Abbildung 2.73.: Ablauf einer Operation in einem Rechner mit der von Neumannschen Struktur

4. evtl. Der Operand wird aus dem Speicher geladen.

5. evtl. Das Resultat wird in den Speicher zurückgeschrieben.

Abbildung 2.74 zeigt die Phasen der Ausführung eines Mikro-Maschinenbefehls in einem Mikroprozessor[**?**]. Die Mikro-Maschinenbefehle werden in vier Phasen verarbeitet.

1. Mikrobefehladdressierungsphase. Hier wird die Adresse des nächsten Befehls generiert

2. Mikrobefehlsholphase: Hier wird der Befehl geholt.

3. Mikrobefehlsdekodierphase: Hier können unter Umständen aus einem Befehl viele generiert werden (z.B. Die Berechnuing des sin-Wertes erfordert ein Mikroprogramm)

4. Die Mikrobefehlsausführphase

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

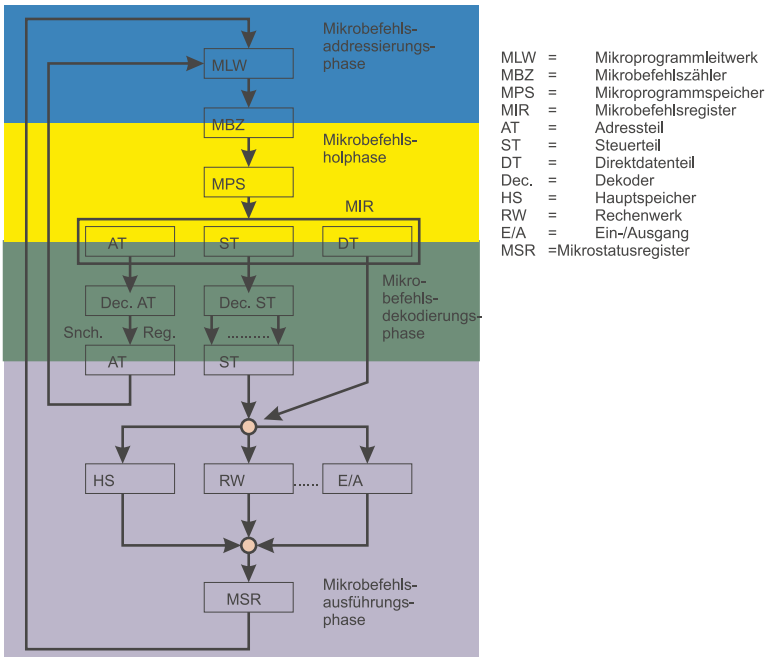



Abbildung 2.74.: Abarbeitung eines Maschinenbefehls in einem Mikrocomputer

Moderne Prozessoren holen und dekodieren Adressen auf Vorrat. Während zeitintensive Befehle abgearbeitet werden, wir spekuliert, wie die Berechnung weitergehen könnte. Die modernen Mikroprozessoren haben einen beachtlichen Teil ihrer Leistung dieser Strategie zu verdanken.
Digitale Signalprozessoren unterscheiden Sich von klassischen Mikroprozessoren dadurch, dass sie mehrere Rechenwerke besitzen und dass diese so ausgelegt sind, dass die Operationen in einem Taktzyklus beendet werden. Ist dies nicht möglich, so sollte die Ausführungszeit eines Befehls vorherbestimmt sein.

## 2.9.2. Digitale Signalprozessoren

Digitale Signalprozessoren werden für die folgenden Verfahren eingesetzt:

**Digitale Filter**
- FIR-Filter (Finite Impulse Response)
- IIR-Filter (Infinite Impulse Response)
- Korrelatoren
- Hilberttransformationen
- Adaptive Filter

**Signalverarbeitung**
- Sprachkompression
- Mittelung
- Energieberechnung

**Datenverarbeitung**
- Verschlüsselung
- Verschleierung
- Kodierung und Dekodierung

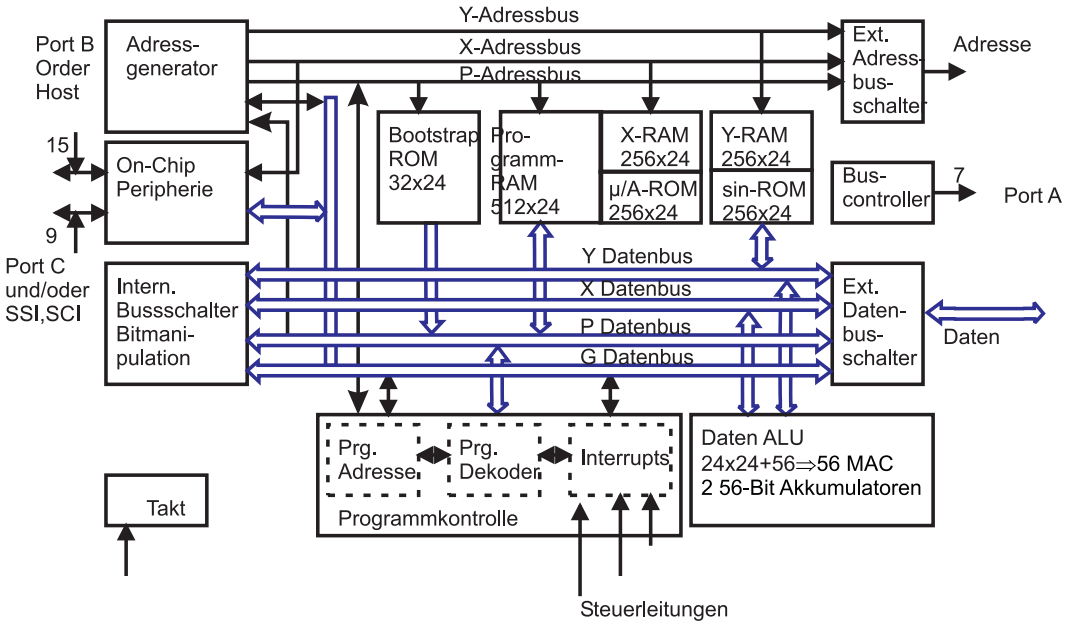



Abbildung 2.75.: Blockschaltbild des Motorola Signalprozessors DSP56001[**?**]

**Numerik**
- Skalar-, Vektor- und Matrixarithmetik
- Transzendentale Funktionen (Funktionsgenerator)
- Pseudo-Zufallszahlen, deterministisches Rauschen

**Modulation**
- Amplitude
- Frequenz
- Phase
- *Modems*

**Spektralanalyse**
- FFT Fast Fourier Transform
- DFT Diskrete Fourier-Transformation
- Sinus/Kosinus-Transformationen

Anwendung finden die Signalprozessoren unter anderem in den folgenden Geräten:

**Telekommunikation**
- Tongeneratoren
- Telefonie (Mehrfrequenz-Wahl)
- Lautsprech-Telefonie
- ISDN
- Rausch-Unterdrückung (Anti-Rauschen)
- Adaptive Differential Pulse Code Modulation (ADPCM) En- und Dekoder

**Datenkommunikation**
- Hochgeschwindigkeitsmodems (56k-Modem)
- Faxgeräte

**Funkkommunikation**
- Mobiltelefonie

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

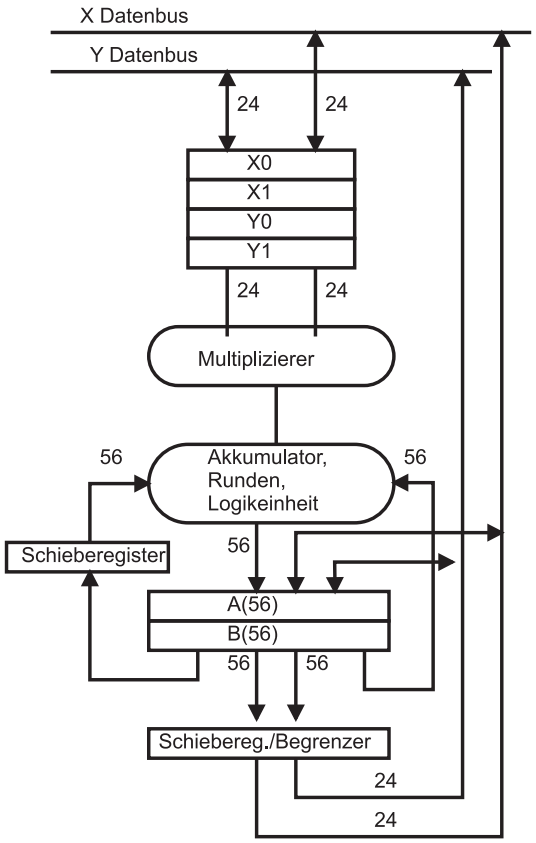



Abbildung 2.76.: Blockschaltbild des Ablaufs einer Rechenoperation im Motorola Signalprozessor DSP56001[**?**]

- Abhörsichere Funkverbindungen
- Radiosender

**Computer**
- Array-Prozessoren
- Graphik-Beschleuniger

**Bildverarbeitung**
- Mustererkennung
- OCR (Schrifterkennung)
- Bildwiederherstellung
- Bildkompression
- Bildverbesserung
- Bilderkennung und -verarbeitung für Roboter

**Instrumente**
- Spektralanalyse
- Funktionsgeneratoren
- Datenerfassung

**Audio-Signalverarbeitung**
- Digitale AM/FM-Radiosender/Empfänger
- Digitale Hi-Fi Vorverstärker
- Musiksynthesizer
- Equalizer
- Virtuelle Dolby-Surround-Prozessoren (mit zwei anstelle von 5 Lautsprechern)

**Hochgeschwindigkeitsregelungen**
- Laserdrucker
- Festplatten
- Roboter
- Motoren

**Vibrationsanalyse**
- Hochleistungselektromotoren
- Düsentriebwerke
- Turbinen

**Medizin**
- Cat-Scanner
- Elektrokardiogramme
- NMR
- Röntgengeräte mit Minimaldosen

Im Gegensatz zu einem klassischen Mikroprozessor hat ein DSP mehrere Daten- und Adressbusse (Abbildung 2.75 Diese Busse erlauben, in einem Taktzyklus mehrere Rechenoperationen gleichzeitig durchzuführen. Der Ablauf, der zum Beispiel in den Datenbüchern von Motorola[**?**] sehr schön beschrieben ist, ist in Abbildung 2.76 gezeigt. Je zwei Register, X0 und X1, beziehungsweise Y0 und Y1, arbeiten auf einen Multiplizierer. Sie sind mit den jeweiligen datenbussen (X und Y)

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

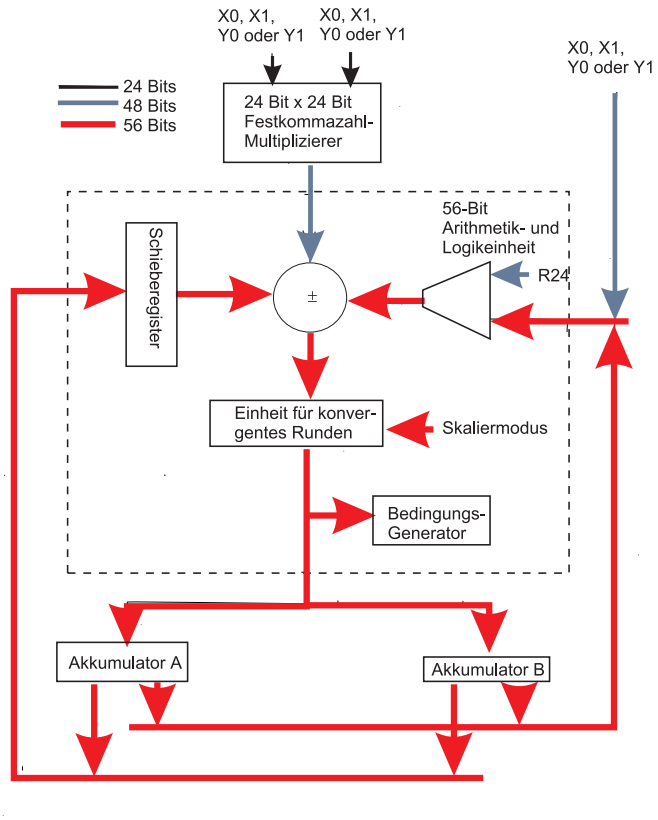



Abbildung 2.77.: Blockschaltbild der MAC-Einheit des Motorola Signalprozessors DSP56001[?]

verbunden. Auf den Multiplizierer folgt der Addierer, so dass in den Ausgangsregistern zum Beispiel in einem Zyklus $A = X0 * Y0 + A$ steht. Schieberegister und Rundungseinheiten komplettieren die ALU (Arithmetic Logic Unit)
Die Multiplikation/Addition wird im Kern von einer MAC-Einheit durchgeführt (Abbildung 2.77). Die datenpfade haben eine unterschiedliche Breite. ein Vergleich mit den Gleichungen für IIR-Filter oder FIR-Filtern (Abschnitt 2.6.2 zeigt, dass die MAC-Einheit in einem Zyklus einen Knoten dieser Filtertypen berechnen kann. Tabelle H.1 zeigt eine Zusammenfassung von Benchmark-Ergebnissen für den DSP 56001 (Taktfrequenz 27 MHz) von Motorola. Ein Vergleich mit Mikroprozessoren wie der Intel-Familie zeigt, dass diese etwa die fünf- bis zehnfache Taktfrequenz brauchen bis sie eine FFT ebenso schnell wie ein DSP berechnen können.

# 3. Bauelemente und Schaltungstechnik

## 3.1. Halbleiter–Grundlagen

### 3.1.1. Grundlagen

Die heutige Elektronik ist im Wesentlichen eine *Festkörperelektronik*. Von zentraler Bedeutung sind dabei *einkristalline Halbleiter*. So sind über 95 % aller kommerzieller Chips aus einkristallinem *Si*. Polykristalline und *amorphe Halbleiter* werden selten eingesetzt. Oxide, Polymere und Metalle sind von sekundärer Bedeutung; allerdings wirft beispielsweise die Herstellung eines isolierenden *Gateoxids*, von Photolacken oder verlustarmer *ohmscher Kontakte* und elektrischer Zuleitungen hochinteressante physikalische, chemische und technologische Fragen auf.

Der Begriff *Halbleiter* bezieht sich auf die elektrische Leitfähigkeit bzw. den spezifischen Widerstand reiner Materialien. Bei 300 K zeigen *Isolatoren* spezifische Widerstände von $> 10^8$ Ωcm, ein guter Isolator $> 10^{15}$ Ωcm; Metalle dagegen $< 10^{-4}$ Ωcm, Halbmetalle von $10^2 - 10^4$ Ωcm. Reine *Halbleiter* können durch gezielte Verunreinigungen (Dotierung) die Lücke zwischen Isolator und Metall ausfüllen, vgl. Bild 3.1. Das Temperaturverhalten der Leitfähigkeit von Metallen und Halbleitern unterscheidet sich aber wesentlich.

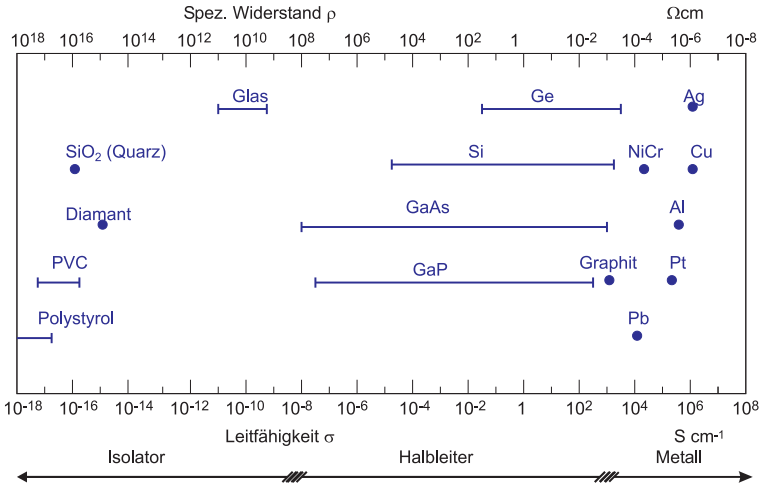



Abbildung 3.1.: Leitfähigkeit und spezifischer Widerstand von Metallen, Halbleitern und *Isolatoren* bei Zimmertemperatur.

Ein Blick auf das Energie–Termschema bzw. genauer das Energie–Bandschema in Abbildung 3.2 soll nochmals an die physikalischen Grundlagen erinnern.

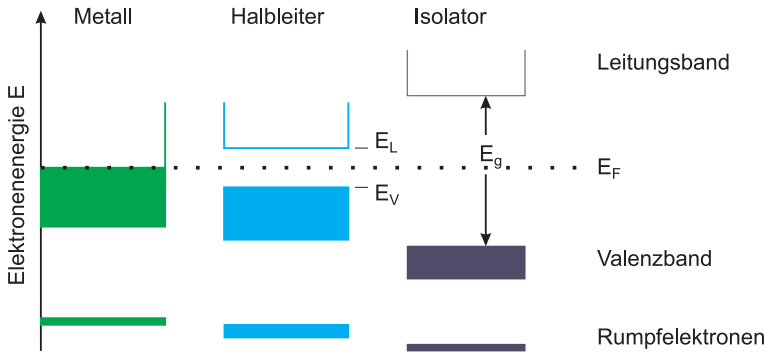



Abbildung 3.2.: Energieschema für Metall, *Halbleiter* und Isolator; schraffiert: besetzte Zustände. $E_F$: Fermi–Niveau, $E_G$: Bandlücke, $E_L$: Leitungsbandunterkante (Unterkante des niedrigsten leeren Bandes), $E_V$: Valenzbandoberkante (Oberkante des höchsten gefüllten Bandes), Skizze nach [**?**]

Metalle haben bei $T = 0$ K ein teilweise besetztes Band; bei Halbleitern und *Isolatoren* ist das vollständig besetzte Valenzband vom vollständig entleerten Leitungsband durch eine Bandlücke getrennt. Bei ideal reinen Materialien liegt die *Fermi-Energie* in der Bandlückenmitte. Da nur partiell gefüllte elektronische Bänder elektrischen Strom tragen können, sind bei $T = 0$ K auch *Halbleiter* isolierend. Bei erhöhten Temperaturen, z. B. Raumtemperatur, und nicht zu grossen Energielücken, z. B. 1, 5 eV, werden genügend Elektronen aus dem Valenzband ins Leitungsband angehoben, um eine merkliche elektrische Leitfähigkeit zu erhalten. (Wir werden noch sehen, dass ausser den Elektronen im Leitungsband auch die 'Löcher' im Valenzband zur Leitfähigkeit des Halbleiters beitragen.) Dagegen ist die Bandlücke bei *Isolatoren* so gross, dass auch bei einigen 100° C keine technisch relevante Leitfähigkeit beobachtet wird.

Halbleitende Materialien können aus Elementen, Verbindungen und Legierungen bestehen. Die Elementhalbleiter stehen in der IV. Hauptgruppe des Periodensystems: C (Diamant, 5, 47 eV), *Si* (engl. silicon, 1, 10 eV), Ge (0, 67 eV) und $\alpha$–Sn (0, 08 eV); mit zunehmender Ordnungszahl nimmt — typischerweise — die Energie der *Energielücke* (bei 300 K) ab. Binäre Verbindungshalbleiter realisiert die Natur auf verschiedene Weisen: Aus IV–IV–Elementen wie SiC (3, 26 eV), aus III-V-Elementen wie *GaAs* und GaN, aus II–VI–Elementen wie CdS und, nicht unmittelbar einzusehen, aus IV–VI–Elementen wie PbS.

Von zunehmend technologischer Bedeutung sind schliesslich $Ge_xSi_{1-x}$–Schichten auf *Si–Substraten*. Ternäre Verbindungshalbleiter wie $Al_xGa_{1-x}As$ oder $Ga_xIn_{1-x}As$ sind die wesentlichen Baustoffe der modernen Kommunikationstechnologie; der erste blaugrüne cw–HL–Laser bestand aus einem Schichtsystem aus ZnCdSe / ZnSSe / ZnMgSSe (Schicht 3: *quaternärer Halbleiter*).

Die beeindruckende Vielfalt schlägt sich auch in der Vielzahl der realisierten *Kristallstrukturen* fort. So finden wir beispielsweise das Diamantgitter bei C, Ge, *Si*, das Zinkblendegitter bei ZnS, *GaAs*, GaP, etc., das Wurzitgitter bei CdS, ZnS (beides möglich), etc., das Kochsalzgitter bei PbS, PbTe, etc. und schliesslich weisen Dichalcogenide wie $WSe_2$, $WS_2$, $MoS_2$, $MoTe_2$ (nicht aber $WTe_2$) eine dem Graphit ähnliche Schichtstruktur auf.

Eine Vorhersage aus 'first principles', ob ein Stoff bei Raumtemperatur halblei-

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

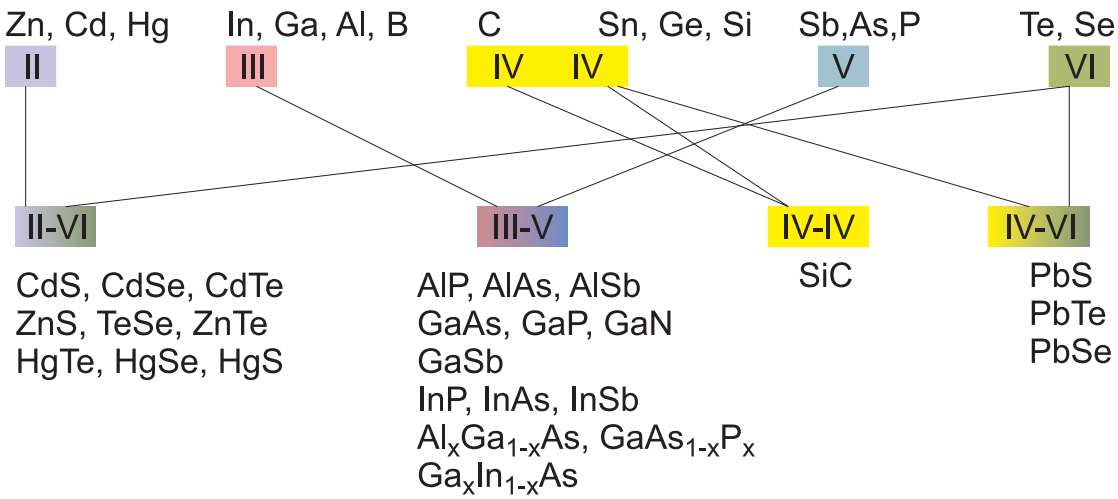



Abbildung 3.3.: Element-, binäre und ternäre Verbindungshalbleiter.

tend sein wird oder nicht fällt schwer; am besten gelingt dies noch bei Ge und *Si*. Ansonsten hilft nur die enge Kombination von Experiment und Bandstrukturrechnung.

Bei den Elementhalbleitern C, *Si* und Ge liegt die *kovalente Bindung* in ihrer reinsten Form vor. Zur Erinnerung: Bei der quantenmechanischen Behandlung des $H_2^+$–Molelüls lernten Sie das Überlappen der Einzelatomzustände erstmals kennen. Näherungsweise wird dort die neue Wellenfunktion durch eine Linearkombination der Einzelwellenfunktionen beschrieben. $\Psi_+ = \Psi_a + \Psi_b$ bildet den bindenden, energetisch (im Vergleich zum Ausgangszustand) tiefer liegenden Grundzustand aus; d. h. die Elektronendichte zwischen den Atomkernen wird erhöht und so ihre Coulomb–Abstossung reduziert. Der bindende Grundzustand ist (Spinentartung, Pauli–Prinzip) mit zwei Elektronen besetzbar. Der antibindende Zustand $\Psi_- = \Psi_a - \Psi_b$ liegt energetisch höher als die Ausgangszustände. Beim analog zu behandelnden $H_2$–Molekül bezeichnen wir diese Art von Bindung als Elektronenpaarbindung. Je stärker der räumliche Überlapp der Wellenfunktionen, desto stärker die kovalente Bindung. (Weiteres Beispiel: Zwei sich annähernde, unendlich hohe Rechteck–Potentialtöpfe mit je einem Elektron, eindimensional.)

Die kovalente Bindung braucht zu ihrer Realisierung unvollständig besetzte Einzelatomorbitale. Diamant zum Beispiel besitzt die Konfiguration $1\,s^2, 2\,s^2, 2\,p^2$; es stehen also nur die beiden p–Orbitale der kovalenten Bindung zur Verfügung. Man findet aber vier kovalente Bindungen. Die Erklärung liefert die sog. Hybridisierung (zur Erinnerung: $CH_4$–Molekül). Unter dem Einfluss der Nachbarn wird der kleine energetische Unterschied zwischen $2\,s^2$– und $2\,p^2$–Orbitalen wieder aufgehoben. Aus den Wellenfunktionen $2\,s$, $2\,p_x$, $2\,p_y$ und $2\,p_z$ werden vier neue Linearkombinationen gebildet: Die $sp^3$–Hybrid–Orbitale bilden im Raum einen perfekten Tetraeder aus ($109,5°$ – Winkel). Ordnen sich im periodischen Gitter der Festkörper die nächsten Nachbarn jeweils gerade auf den Tetraederpositionen an, so kommt es zur Ausbildung von 4 kovalenten Bindungen pro Atom. D. h. die Nachbaratome teilen sich die verfügbaren Elektronen gerade so, dass nur die bindenden Zustände besetzt sind. Abbildung 3.4 gibt das wieder: die s– und p–Orbitale spalten bei Annäherung der Atome auf, dann aber bilden sich die — ebenfalls aufgespaltenen — Hybridorbitale aus.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

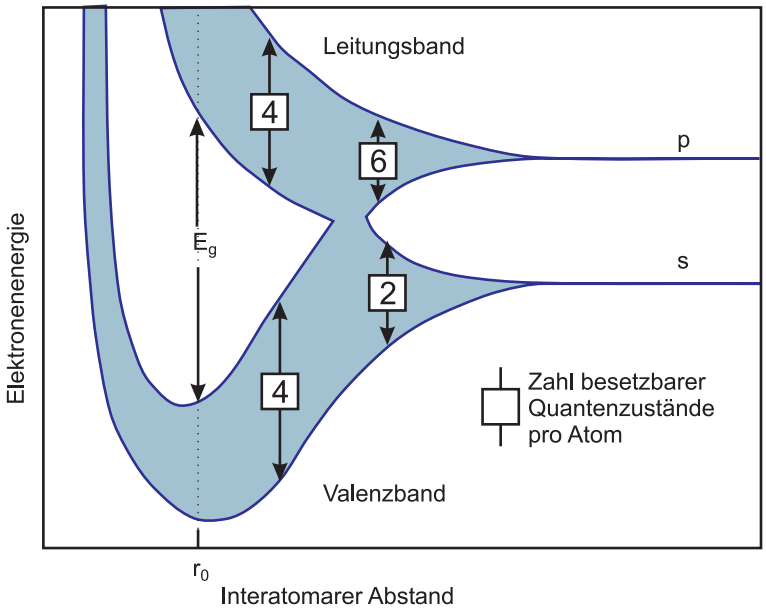



Abbildung 3.4.: Schematischer Verlauf der Bandaufspaltung als Funktion des interatomaren Abstandes für die tetraedrisch gebundenen *Halbleiter* Diamant (C), *Si* und Ge, gezeichnet nach [**?**].

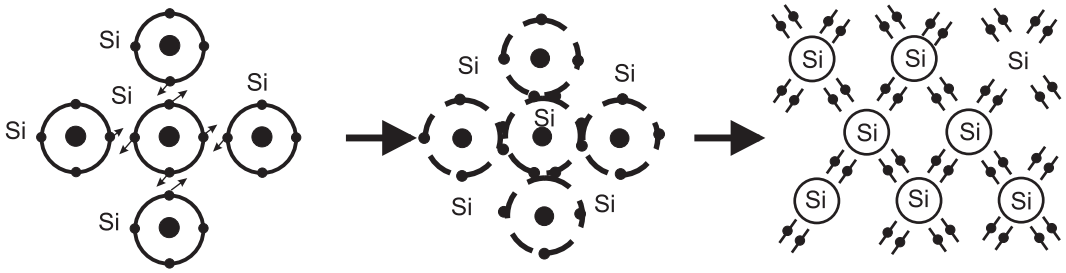



Abbildung 3.5.: Bindungsstruktur (kovalente Bindung) des Eigenhalbleiters in zweidimensionaler Darstellung mit zunehmender zeichnerischer Abstraktion.

Mit kleiner werdendem Atomabstand entsteht eine verbotene Zone, d. h. zwischen bindenden und antibindenden $sp^3$–Teilbändern tut sich ein 'Gap' auf. Alle vier pro Atom zur Verfügung stehenden Elektronen haben im tiefer liegenden, bindenden Band, dem sog. Valenzband, Platz: es liegt bei $T = 0$ K ein Isolator vor. Abbildung 3.4 gilt für kristalline und amorphe Elementhalbleiter gleichermassen, solange die tetraedrische Bindungsanordnung vollständig gegeben ist; allerdings führen die im Amorphen etwas variierenden Abstände und das Vorhandensein von unabgesättigten Bindungen zu Abweichungen: es gibt Zustände, die die *Energielücke* verkleinern. Ergänzend sei bemerkt, dass mit dem temperaturabhängigen Atomabstand auch die Energiebreiten der Bandlücke temperaturabhängig sein muss: bei $T = 0$ K ist sie am grössten.

Die oben besprochene räumliche Struktur der kovalenten Bindung wird gerne in eine zweidimensionale, abstrakte Darstellung ('Bindungsmodell') überführt, um z. B. die elektrische Leitung zu veranschaulichen.

Die $sp^3$–Hybridisierung findet man bei den III–V–Halbleitern wieder. Es liegt ei-

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

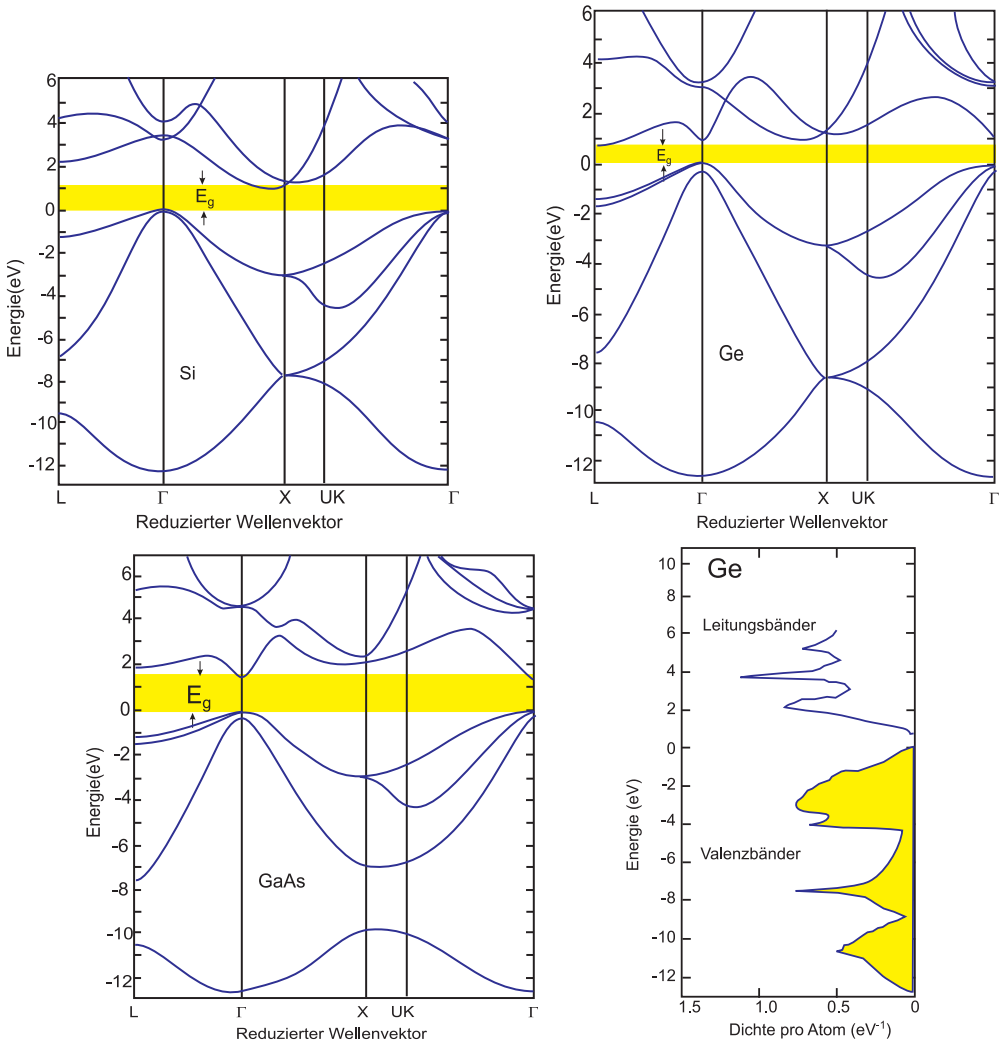



Abbildung 3.6.: Verschiedene Bandstrukturen gezeichnet nach[**?**].

ne Mischbindung aus ionischer Bindung (Ladungstransfer vom V er– zum III er–Material) und kovalenter Bindung vor; letztere überwiegt. Auch die *II–VI–Halbleiter* zeigen diese Mischbindung, mit grösserem ionischen Anteil als die III–V- er.

Bisher betonten wir die physikalischen Gemeinsamkeiten. Die unterschiedlichen atomaren Eigenschaften spiegeln sich in unterschiedlichen Bandstrukturen wieder. Im Abbildung 3.6 sind die $E(\boldsymbol{k})$–Darstellungen der elektronischen Bänder aus (an Experimente angepasste) Rechnungen für *Si*, Ge und *GaAs* angegeben. (Weiterführende Arbeiten zeigen noch kompliziertere Band–Feinstrukturen, z. B. führt die Berücksichtigung der Spin–Bahn–Aufspaltung bei *Si* und Ge zur Aufspaltung der Oberkante des Valenzbandes, es gibt dann leichte, schwere und 'split–off'–Löcher.)

*Si* und Ge sind sog. *indirekte Halbleiter*. Das Maximum der Valenzband–Oberkante liegt beidesmal beim Γ–Punkt ($\boldsymbol{k} = (0, 0, 0)$), aber das Minimum der Leitungsband–Unterkante liegt bei *Si* am X–Punkt (ΓX=[100]–Richtung) und bei Ge am L–Punkt (ΓL=[111]–Richtung). Auch GaP und AlSb haben eine indirekte Bandlücke. Aber die wichtigsten *III–V–Halbleiter* (*GaAs*, GaSb, InSb, InAs, InP) haben eine direkte Bandlücke, d. h. Valenzbandmaximum und Leitungsbandminimum liegen beide bei Γ; gleiches gilt für die *II–VI–Halbleiter* ZnO, ZnS, CdS, CdSe und CdTe. Man spricht von *direkten Halbleitern*. Auf Halbleitern mit direkter Bandlücke basieren

die optoelektrischen Bauelemente.
Der Vollständigkeit halber ist für Ge die theoretisch ermittelte elektronische Zustandsdichte $D(E)$ wiedergegeben, dabei sind die besetzten Zustände der Valenzbänder schraffiert worden. Einige kritische Punkte lassen sich auf Minima oder Maxima in der Bandstruktur zurückführen. Von den komplizierten Verläufen darf man sich nicht abschrecken lassen, für die elektrische Leitfähigkeit genügt es i. allg. den Verlauf des Valenz– und des Leitungsbands rund um $\Gamma$ zu kennen.

## 3.1.2. Intrinsischer Halbleiter

Bei $T = 0$ K zeigen *Halbleiter* keine Leitfähigkeit. Bei endlichen Temperaturen aber kommt es zu einer 'thermischen Anregung' von Elektronen über die Bandlücke hinweg. Sie hinterlassen im Valenzband jeweils eine positiv geladene Lücke, ein sog. Loch. In einem äusseren elektrischen Feld $\boldsymbol{E}$ können nicht nur die Elektronen im Leitungsband (wie bei den Metallen) Energie aufnehmen, sondern auch die im Valenzband. Vereinfachend wird dies beschrieben durch die Energieaufnahme der Löcher.
Für die Stromdichte gilt (im stationären Fall) bei Metallen:

$$\boldsymbol{j} = \sigma \boldsymbol{E} = e \cdot n \cdot \mu_{\text{Metall}} \cdot \boldsymbol{E} \tag{3.1.1}$$

und für die elektrische Leitfähigkeit

$$\sigma = e \cdot n \cdot \mu_{\text{Metall}} \tag{3.1.2}$$

mit $\mu$ als Beweglichkeit und $n$ als Anzahldichte der Elektronen.
Für die Leitfähigkeit bzw. für die Beweglichkeit bei Halbleitern gilt analog:

$$\boldsymbol{j} = -e \cdot n \cdot \mu_{\text{n}} \cdot \boldsymbol{E} + e \cdot p \cdot \mu_{\text{p}} \cdot \boldsymbol{E} \tag{3.1.3}$$

$$\text{und} \qquad \sigma = |e| \cdot (n\mu_{\text{n}} + p|\mu_{\text{p}}|) \ . \tag{3.1.4}$$

Im allg. gilt für die Beweglichkeit der Elektronen $\mu_{\text{n}}$ und der Löcher $\mu_{\text{p}}$ die Relation $\mu_{\text{n}} > \mu_{\text{p}} > \mu_{\text{Metall}}$; n und p sind die Volumenanzahldichten der Elektronen bzw. Löcher. Letztere zeigen, im Gegensatz zu $\mu_{\text{Metall}}$, eine starke Temperaturabhängigkeit. Ein *intrinsischer Halbleiter* ist gekennzeichnet durch das blosse Vorhandensein des oben beschriebenen thermischen Anregungsmechanismus. Man spricht auch vom *idealen Halbleiter*, weil Störstellen–freien Halbleiter. Wir haben also zwei Ladungsträgertypen im *Halbleiter*! Im thermodynamischen Gleichgewicht werden sie ständig generiert und rekombinieren — nach einer Lebensdauer $\tau$ — wieder ins Valenzband.
Die Beweglichkeiten $\mu_{\text{n}}$ und $\mu_{\text{p}}$ sind — was in der obigen Gleichung nicht enthalten ist — streng genommen Impuls– bzw. Energie–abhängige Grössen. Häufig genügt es jedoch völlig, die Ladungsträger in den Valenzbandmaxima und Leitungsbandminima zu berücksichtigen, d. h. nicht zu grosse $T$ und $\boldsymbol{E}$ zuzulassen. Dann gilt für die betrachteten Bänder die sog. *parabolische Näherung* (auch Näherung der Standardbänder):

$$E(\boldsymbol{k}) = E_0 \pm \hbar^2 \left( \frac{k_x^2}{2\, m_x^*} + \frac{k_y^2}{2\, m_y^*} + \frac{k_z^2}{2\, m_z^*} \right) , \tag{3.1.5}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

mit $m^*$ als konstante, d. h. Impuls– bzw. Energie–unabhängige *effektive Masse* ('effektive Massennäherung').
Für die tensorielle effektive Masse $m^*_{ij}$ gilt:

$$\left(\frac{1}{m^*}\right)_{ij} = \frac{1}{\hbar^2}\frac{\partial^2 E\,(\boldsymbol{k})}{\partial k_i\,\partial k_j} = \text{Krümmung von } E\,(\boldsymbol{k})\ , \tag{3.1.6}$$

ein kleines $m^*$ beschreibt also eine starke Bandkrümmung, ein grosses $m^*$ eine schwache.
Für die Dichten der Ladungsträger in Leitungs– und Valenzband gilt ganz allgemein:

$$n = \int_{E_\mathrm{L}}^{\infty} D_\mathrm{L}(E)\, f(E,T)\,\mathrm{d}E \tag{3.1.7}$$

$$\text{und} \qquad p = \int_{-\infty}^{E_\mathrm{V}} D_\mathrm{V}(E)\ [1-f(E,T)]\ \mathrm{d}E\,, \tag{3.1.8}$$

mit $E_\mathrm{L}$ bzw. $E_\mathrm{V}$ als Energien der Leitungsbandunterkante bzw. Leitungsbandoberkante, $D_\mathrm{L}$ und $D_\mathrm{V}$ als Zustandsdichten der Elektronen und Löcher und mit $f$ als Verteilungsfunktion gemäss der Fermistatistik mit $E_\mathrm{F}$, der Fermieenergie als chemischem Potential, also

$$f(E,T) = \frac{1}{\mathrm{e}^{\frac{E-E_\mathrm{F}}{k_B T}}+1}\ . \tag{3.1.9}$$

Für die Zustandsdichten gilt in parabolischer Näherung:

$$D_\mathrm{L}(E) = \frac{(2\,m_n^*)^{3/2}}{2\,\pi^2\hbar^3}\sqrt{E-E_\mathrm{L}} \quad , \quad (E > E_\mathrm{L}) \tag{3.1.10}$$

$$D_\mathrm{V}(E) = \frac{\left(2\,m_p^*\right)^{3/2}}{2\,\pi^2\hbar^3}\sqrt{E_\mathrm{V}-E} \quad , \quad (E < E_\mathrm{V}) \tag{3.1.11}$$

$$D(E) = 0 \quad , \quad (E_\mathrm{V} < E < E_\mathrm{L})\ . \tag{3.1.12}$$

Die sog. *Neutralisationsbedingung des Idealhalbleiters* ergibt sich aus der thermischen Anregung, die Generation eines Elektrons ins Leitungsband erzeugt ein Loch im Valenzband:

$$n = p = \sqrt{np} = n_i\,; \tag{3.1.13}$$

$n_i$ steht für Inversionsdichte, auch Eigenleitungskonzentration genannt; analog dazu bezeichnet man $\sigma_i$ als Eigenleitung.
Wenn die effektiven Massen $m^*_\mathrm{n}$ und $m^*_\mathrm{p}$ gleich sind, also auch die Zustandsdichten gleich sind, muss das Fermi–Niveau $E_\mathrm{F}$ in der Mitte der Bandlücke liegen. Bei ungleichen Massen wandert das Fermi–Niveau aus der Mitte, seine Lage ist dann schwach temperaturabhängig. (Konsequenzen für elektronische Bauelemente!)
Man sieht im linken Teilbild von 3.7, dass nur die 'Ausläufer' der Fermifunktion bei der Berechnung der Ladungsträgerkonzentrationen (zum Produkt für $D$) beitragen. Die 'Aufweichungszone' der Fermifunktion ($\approx 2\,k_B T$) ist bei Raumtemperatur klein: $k_B T \approx 25$ meV $= \frac{1}{40}$ eV. Die *Energielücke* ist — bis auf ein paar wenige Ausnahmen ($\alpha$–Sn, InSb) 10–100 mal grösser. Man darf deshalb die Fermifunktion durch die Boltzmann–Besetzungswahrscheinlichkeit annähern:

$$\frac{1}{\mathrm{e}^{\frac{E-E_\mathrm{F}}{k_B T}}+1} \approx \mathrm{e}^{-\frac{E-E_\mathrm{F}}{k_B T}} \ll 1 \quad \text{für} \quad E - E_\mathrm{F} \gg 2\,k_B T\ . \tag{3.1.14}$$

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

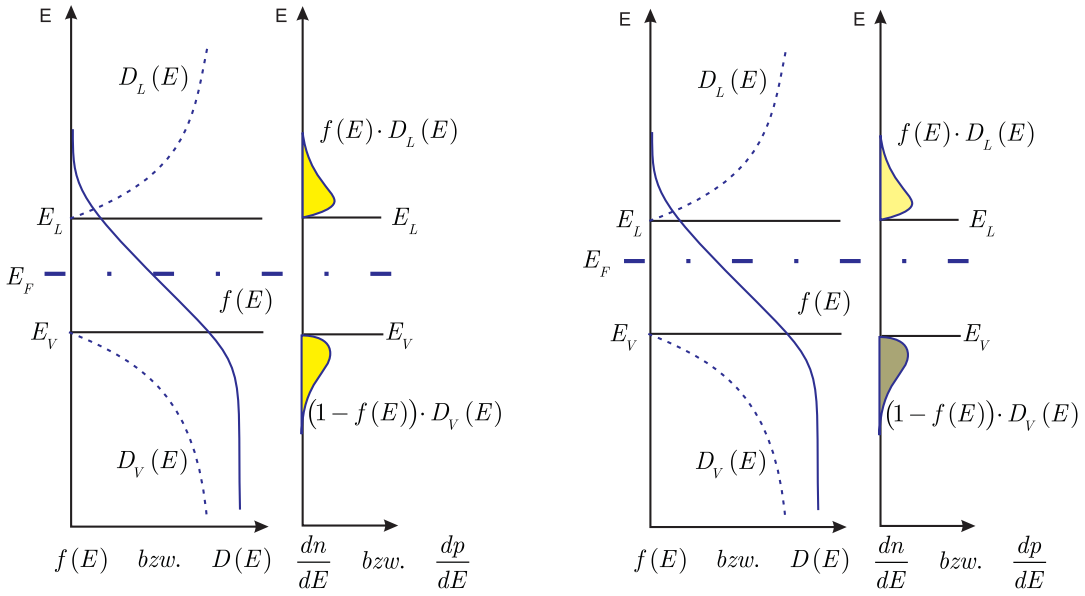



Abbildung 3.7.: Fermi–Funktion $f(E)$ und Zustandsdichte $D(E)$ sowie Elektronen– (n) bzw. Löcherkonzentration (p) in Leitungs– und Valenzband. Links: Gleiche Zustandsdichten in Leitungs– und Valenzband; rechts: verschiedene Zustandsdichten in Leitungs– und Valenzband (gezeichnet nach [**?**]).

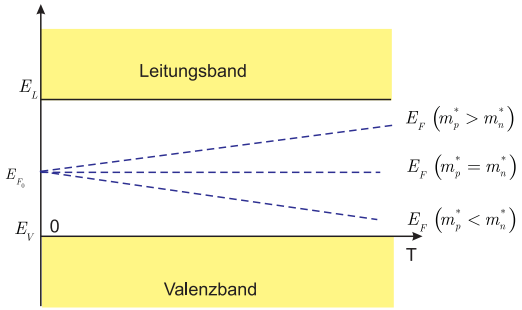



Abbildung 3.8.: Abhängigkeit der Fermi–Energie $E_F$ von der Temperatur.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

| | $n_i\ [cm^{-3}]$ |
|---|---|
| C | $6,7 \cdot 10^{-28}$ |
| Si | $1,5 \cdot 10^{10}$ |
| Ge | $2,4 \cdot 10^{13}$ |
| Ga As | $5 \cdot 10^{7}$ |

Tabelle 3.1.: Inversionsdichten $n_i$ einiger Materialien bei Raumtemperatur.

Diese Näherung nennt man die Näherung der Nichtentartung, sie ist gut für kleine Ladungsträgerkonzentrationen. Für diese liefert dann die Rechnung:

$$n = N_{\text{eff}}^{\text{L}} \cdot \mathrm{e}^{-\frac{E_{\text{L}}-E_{\text{F}}}{k_B T}} \quad , \quad N_{\text{eff}}^{\text{L}} = 2\left(\frac{2\pi m_{\text{n}}^* k_B T}{h^2}\right)^{3/2} \tag{3.1.15}$$

$$\text{und} \qquad p = N_{\text{eff}}^{\text{V}} \cdot \mathrm{e}^{-\frac{E_{\text{V}}-E_{\text{F}}}{k_B T}} \quad , \quad N_{\text{eff}}^{\text{V}} = 2\left(\frac{2\pi m_{\text{p}}^* k_B T}{h^2}\right)^{3/2} . \tag{3.1.16}$$

Die sog. effektiven Zustandsdichten (auch Entartungskonzentrationen) $N_{\text{eff}}^{L,V}$ gelten also formal für ein einziges Energieniveau, nämlich die Bandkante L, V.
Damit lässt sich die Neutralisationsbedingung in der Form eines Massenwirkungsgesetzes schreiben:

$$n_i = p_i = \sqrt{N_{\text{eff}}^{\text{L}}\, N_{\text{eff}}^{\text{V}}}\ \mathrm{e}^{-\frac{E_{\text{g}}}{2k_B T}} = 2\left(\frac{k_B T}{2\pi\hbar^2}\right)^{3/2} \left(m_{\text{n}}^*\, m_{\text{p}}^*\right)^{3/4} \mathrm{e}^{-\frac{E_{\text{g}}}{2k_B T}} . \tag{3.1.17}$$

Aus diesem Grund nennt man die *Halbleiter* 'Heissleiter' (und Metalle im Vergleich hierzu Kaltleiter). Weiter gilt:

$$\frac{N_{\text{eff}}^{\text{V}}}{N_{\text{eff}}^{\text{L}}} = \frac{\mathrm{e}^{\frac{2E_{\text{F}}}{k_B T}}}{\mathrm{e}^{\frac{E_{\text{V}}+E_{\text{L}}}{k_B T}}} \tag{3.1.18}$$

$$\text{und} \qquad E_{\text{F}} = \frac{E_{\text{V}}+E_{\text{L}}}{2} + \frac{k_B T}{2}\ \ln\left(\frac{N_{\text{eff}}^{\text{V}}}{N_{\text{eff}}^{\text{L}}}\right)$$

$$= \frac{E_{\text{L}}+E_{\text{V}}}{2} + \frac{3}{4}k_B T\ \ln\left(\frac{m_{\text{p}}^*}{m_{\text{n}}^*}\right) . \tag{3.1.19}$$

In Tabelle 3.1 sind zum Abschluss einige Zahlenwerte für die Inversionsdichte $n_i$ bei 300 K angegeben. Diese Werte sind in Relation zu sehen mit typischen Atomdichten von $> 2 - < 5 \cdot 10^{22}\ \frac{\text{Atome}}{cm^3}$.
Die experimentelle Beobachtbarkeit der intrinsischen Leitung setzt extrem sauberes *Halbleiter–Material* voraus. Die niedrigsten erreichbaren Verunreinigungskonzentrationen bei Halbleitereinkristallen wie Ge und *Si* liegen bei etwa $10^{12}\ cm^{-3}$. (Vergleiche tiegelfreies Zonenziehen und Zonenreinigen von Siliziumstäben; *Si* ist der am reinsten darstellbare Stoff überhaupt.) Reinstes *GaAs* dagegen hat heute Ladungsträgerdichten von $10^{16}\ cm^{-3}$.
Extrem reines Material ist bei Raumtemperatur sehr hochohmig (siehe Beispiel *Si*); der Transport von elektrischem Strom ist also sehr verlustreich. Deshalb werden gezielt elektrisch aktive Störstellen in den *Halbleiter* eingebaut. Erst die Möglichkeit, definiert räumliche Konzentrationsprofile von freien Elektronen und freien Löchern auf sub–$\mu$m–Skala vorgeben zu können, ermöglicht die moderne Festkörperelektronik.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

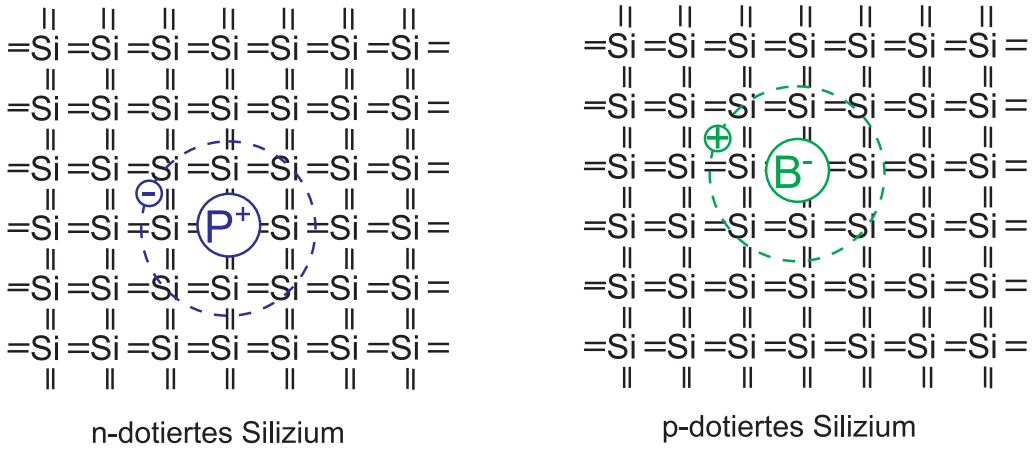



Abbildung 3.9.: Schematische Darstellung der Wirkung eines Donators (links) bzw. eines Akzeptors (rechts) in einem Silizium–Gitter (gezeichnet nach [**?**]).

### 3.1.3. Dotierung von Halbleitern

Verunreinigt man *Si* (oder Ge) gezielt mit fünfwertigen Atomen wie P, As oder Sb, so beobachtet man bei endlichen Temperaturen eine erhöhte *Ladungsträgerdichte* im Leitungsband. Diese Störstellen heissen dann *Donatoren*, der so dotierte Halbleiter heisst *n–Halbleiter*.
Baut man in vierwertige Halbleiter–Materialien dreiwertige Fremdatome wie B, Al, Ge oder In ein, so findet man bei $T > 0$ K eine erhöhte *Ladungsträgerdichte im Valenzband*. Solche Störstellen werden als *Akzeptoren* bezeichnet; in Analogie spricht man von *p–Halbleitern*.
Abbildung 3.9 zeigt schematisch den Einbau eines Donator– bzw. Akzeptoratoms auf einem Gitterplatz im *Si–Einkristall*. Im Falle des Donators nehmen vier Valenzelektronen an den kovalenten Bindungen zu den benachbarten *Si–Atomen* teil, das fünfte Elektron ist nur schwach an das Phosphoratom gebunden und kann schon bei kleinen Temperaturen angeregt bzw. ionisiert ($T \geq 10$ K), also ins Leitungsband angehoben werden. Analog gilt für ein Akzeptoratom, dass schon bei kleinen Temperaturen ein Elektron aus dem Valenzband die kovalente Bindung komplettieren kann und so ein schwach gebundenes Loch bzw. durch Ionisation ein zusätzliches freies Loch im Valenzband erzeugt wird.
Der Radius der Störstellenbahn beträgt ca. 10 Gitterabstände bzw. das schwach gebundene Elektron bzw. Loch ist über ca. $10^3$ *Si–Gitteratome* 'verschmiert'. Aus FIR–Absorptionsspektroskopie–Experimenten bei tiefen Temperaturen kennt man die energetischen Abstände $E_\mathrm{D}$ der Donatorniveaus bzw. $E_\mathrm{A}$ der Akzeptorniveaus vom Leitungsband bzw. Valenzband. Dies geben qualitativ die beiden in Abbildung 3.10 gezeigten Bändermodelle (Valenzbandmaximum und Leitungsbandminimum über Ortskoordinate, ergänzt um die Störstellen–Grundniveaus) wieder.
Die aus den genannten Absorptionsspektren bekannten angeregten Niveaus sind nicht eingezeichnet. Zur Abschätzung der Anregungs– und Ionisationsenergien, sowie der Ausdehnung von Störstellen kann ein Wasserstoffatom–Modell herangezogen werden. ($m_\mathrm{e}$ wird ersetzt durch $m^*$, $\epsilon_0$ durch $\epsilon_0 \cdot \epsilon_\mathrm{Si}$, $\epsilon_\mathrm{Si} = 11,7$ (Abschirmung der Coulomb–Anziehung zwischen $P^+$ und $e^-$ bzw. $B^-$ und positiv geladenem Loch)). Im Vergleich zu den Bandlücken sind die Störstellenabstände i. allg. klein

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

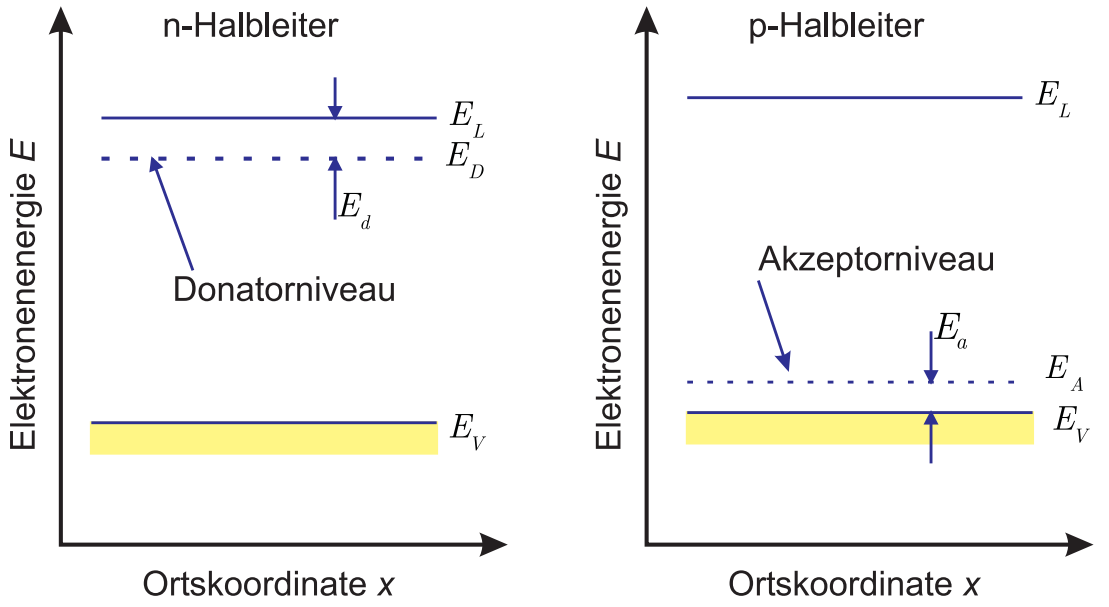



Abbildung 3.10.: Qualitative Lage der Grundzustandniveaus von *Donatoren* und *Akzeptoren* in Bezug auf die Unterkante des Leitungsbandes $E_\mathrm{L}$ bzw. die Oberkante des Valenzbandes $E_\mathrm{V}$ (gezeichnet nach [**?**]).

('flache Störstellen'), tiefer sitzende Störstellen sind schwerer bzw. praktisch gar nicht thermisch zu ionisieren und erhöhen die Ladungsträgerdichten nicht.
Als Beispiel sind in den Tabellen 3.2 und 3.3 einige wichtige Messwerte angegeben.

| | P [meV] | As [meV] | Sb [meV] |
|---|---|---|---|
| Si | 45 | 54 | 43 |
| Ge | 13 | 14 | 10 |

Tabelle 3.2.: Energetischer Abstand $E_\mathrm{D}$ einiger Donatorenniveaus vom Leitungsband für Silizium und Germanium[**?**].

| | B [meV] | Al [meV] | Ga [meV] | In [meV] |
|---|---|---|---|---|
| Si | 45 | 67 | 74 | 153 |
| Ge | 11 | 11 | 11 | 12 |

Tabelle 3.3.: Energetischer Abstand $E_\mathrm{A}$ einiger Akzeptorniveaus vom Valenzband für Silizium und Germanium[**?**].

Der Eigenhalbleiter hat im thermodynamischen Gleichgewicht immer die gleiche Konzentration an Elektronen und an Löchern. Im gezielt dotierten Material ist dies anders. Im Falle der n–Dotierung befinden sich mehr Elektronen im Leitungsband als Löcher im Valenzband. Die Elektronen sind also die sog. *Majoritätsladungsträger* und die Löcher die sog. *Minoritätsladungsträger*. Es gilt:

$$n(T) = n_i(T) + N_\mathrm{D}^+ \; . \tag{3.1.20}$$

$N_\mathrm{D}^+$ ($N_\mathrm{A}^-$) ist die Anzahldichte der ionisierten *Donatoren* (*Akzeptoren*). Für den

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

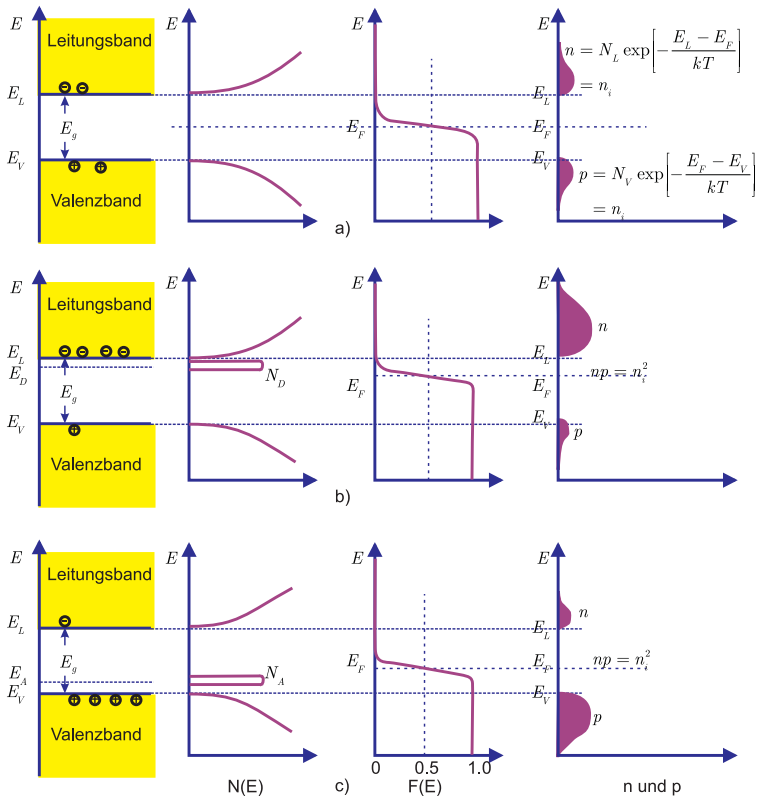



Abbildung 3.11.: Schematisches Bändermodell, Zustandsdichte und Fermi–Verteilung sowie Ladungsträgerdichten für a) intrinsische, b) n– und c) p–Halbleiter.

Fall der p–Dotierung gilt für die Majoritätsladungsträger analog:

$$p(T) = p_i(T) + N_{\mathrm{A}}^- \ . \tag{3.1.21}$$

Auch im *dotierten Halbleiter* gilt (im thermodynamischen Gleichgewicht) die grundlegende Beziehung:

$$n(T) \cdot p(T) = n_i^2(T) \ , \tag{3.1.22}$$

d. h. eine Erhöhung von $n(T)$ bewirkt eine Erniedrigung von $p(T)$ um denselben Faktor! I. allg. sind die Minoritätsladungsträger–Anzahldichten sehr klein im Vergleich zu denen der Majoritätsladungsträger; im *homogenen Halbleiter* sind sie praktisch vernachlässigbar, in Bauelementen mit ihren inhomogenen Dotierungen, Grenz– und Randschichten aber keinesfalls.

## 3.1.4. Ladungsträgerdichten im dotierten Halbleiter

Solange die Besetzung im Leitungsband bzw. im Valenzband in guter Näherung mit Hilfe der Boltzmann–Verteilung beschrieben werden kann (Fall des nicht entarteten Halbleiters), gilt auch für dotierte *Halbleiter* das Massenwirkungsgesetz:

$$n \cdot p = N_{\mathrm{eff}}^{\mathrm{L}} \cdot N_{\mathrm{eff}}^{\mathrm{V}} \, \mathrm{e}^{-\frac{-E_{\mathrm{g}}}{k_B T}} \ . \tag{3.1.23}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Eine etwas kompliziertere Neutralitätsbedingung regelt wieder die Lage des Fermi–Niveaus $E_F$ im homogen dotierten *Halbleiter*; die negative *Ladungsträgerdichte* muss gleich der positiven *Ladungsträgerdichte* sein:

$$n + N_A^- = p + N_D^+ \ , \tag{3.1.24}$$

wobei für die Störstellendichte gilt:

$$\begin{array}{lll} & N_D = N_D^0 + N_D^+ & N_D = N_D^+ \\ \text{und} & \text{bzw. bei RT:} & \\ & N_A = N_A^0 + N_A^- & N_A = N_A^- \ . \end{array} \tag{3.1.25}$$

$N_{D,A}^0$ bezeichnt dabei die Anzahldichte der nicht ionisierten *Donatoren* bzw. *Akzeptoren*. Für Störstellenkonzentrationen von $\geq 10^{17}\ cm^{-3}$, wie sie für p– bzw. n–Dotierung üblich sind, nicht aber für 'hohe Dotierungen' $p^+$ oder $n^+$ ($10^{18}\ cm^{-3}$), gilt in guter Näherung:

$$N_D = N_D^0 \cdot \left[1 + e^{\frac{E_D - E_F}{k_B T}}\right] \tag{3.1.26}$$

$$\text{und analog} \qquad N_A = N_A^0 \cdot \left[1 + e^{\frac{E_F - E_A}{k_B T}}\right] \ . \tag{3.1.27}$$

Der allgemeine Fall, wo gleichzeitig p– und n–Dotierung vorliegt, ist nur numerisch lösbar, reine n– oder p–Dotierung kann (mit den oben angegebenen Formeln) diskutiert werden. Für die n–Dotierung lautet die Lösung:

$$n \approx 2\, N_D \left(1 + \sqrt{1 + 4 \frac{N_D}{N_{eff}^L}\, e^{\frac{E_L - E_D}{k_B T}}}\right)^{-1} \ , \tag{3.1.28}$$

sie beschreibt für kleine Temperaturen das Regime der Störstellenreserve, dann den Erschöpfungszustand (der *Donatoren*) und für hohe Temperaturen den Bereich der intrinsischen Trägerkonzentration. Die Lage der *Fermi-Energie* verhält sich entsprechend: für $T = 0$ K liegt sie in der Mitte zwischen $E_D$ und der Leitungsbandunterkante $E_L$, reicht im mittleren Temperaturbereich von $E_L$ weg und endet im intrinsischen Bereich in der Mitte zwischen $E_D$ und $E_V$, also auf $E_i$.
Die experimentelle Bestimmung der Ladungsträgerdichten in Abhängigkeit von der Temperatur geschieht unter Benutzung des Hall–Effekts.
Bei Dotierungskonzentrationen z. B. von $\geq 10^{17}\ cm^{-3}$ bei *Si* ($n^+$ bzw. $p^+$) erreicht bzw. überschreitet man die sog. kritische Konzentration: die *Donatoren* bzw. *Akzeptoren* 'sehen' einander. Angeregte Störstellen–Zustände liegen unter $E_L$ oder über $E_V$ und die *Energielücke* des Halbleiters wird um einige 10 meV kleiner, gleichzeitig werden weniger Störatome ionisiert, als es bei der entsprechenden Temperatur zu erwarten wäre.

## 3.1.5. Leitfähigkeit in Abhängigkeit von Dotierkonzentration und Temperatur

Im Gegensatz zu den Metallen tragen bei den Halbleitern nicht nur Elektronen an der Fermi–Kante zur elektrischen Leitfähigkeit bei, sondern es müssen die von Elektronen bzw. Löcher besetzten Zustände im unteren Leitungsband bzw. oberen

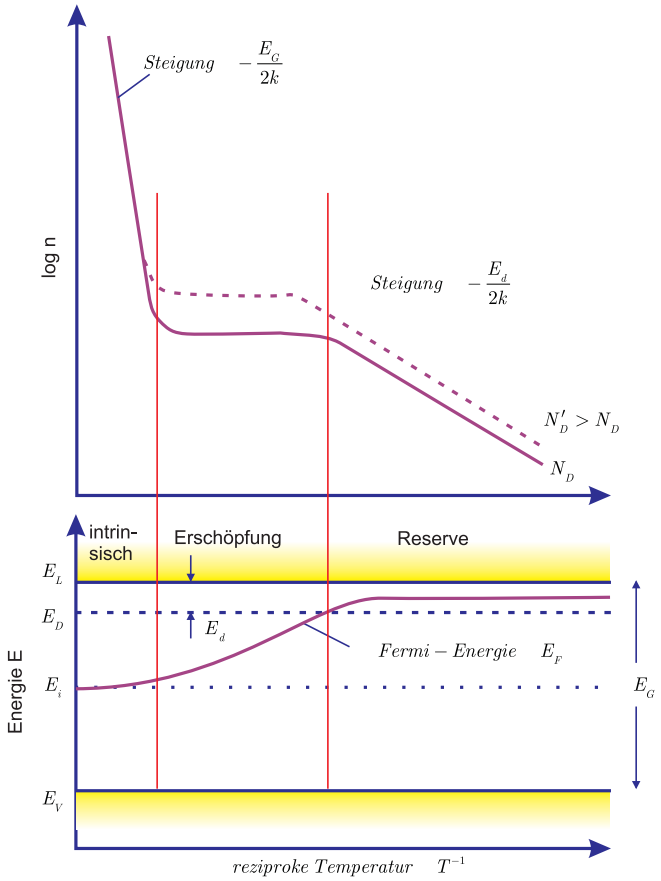



Abbildung 3.12.: oben: Qualitative Abhängigkeit der Elektronenkonzentration $n$ im Leitungsband eines n–Halbleiters von der Temperatur für zwei verschiedene Donatorkonzentrationen $N'_\mathrm{D} > N_\mathrm{D}$. unten: Qualitative Lage der Fermi–Energie $E_\mathrm{F}(T)$ bei demselben *Halbleiter* in Abhängigkeit von der Temperatur (gezeichnet nach [**?**]).

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

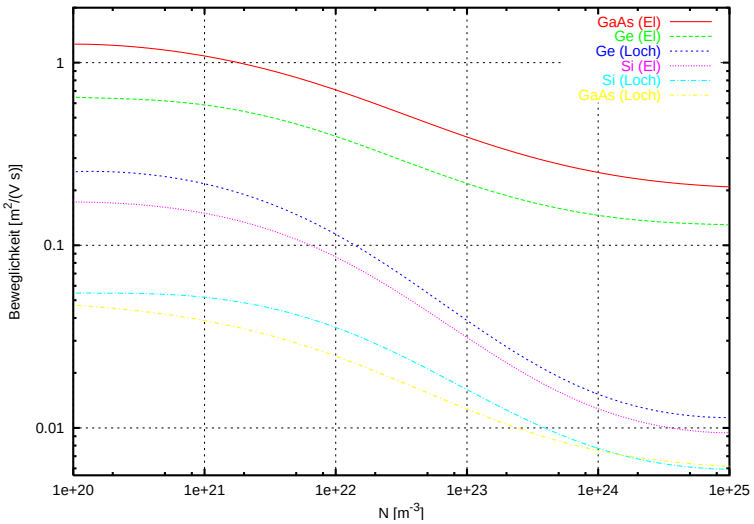



Abbildung 3.13.: Beweglichkeit als Funktion der Elektronen–Konzentration in Ge, *Si*, *GaAs* bei Raumtemperatur[**?**].

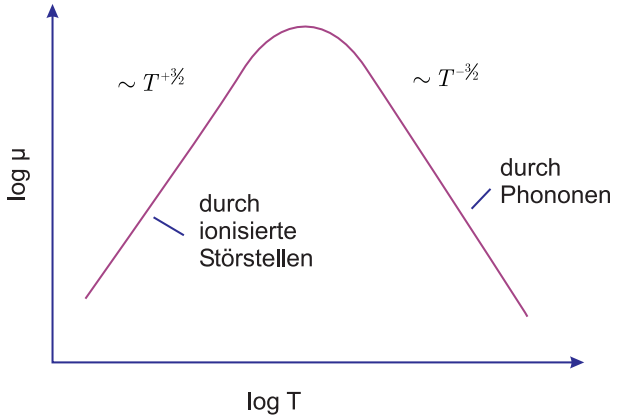



Abbildung 3.14.: Schematische Abhängigkeit der Beweglichkeit $\mu$ in einem *Halbleiter* von der Temperatur bei Streuung an Phononen und an geladenen Störstellen (gezeichnet nach [**?**]).

Valenzband berücksichtigt werden. Deshalb sind Grössen wie die Beweglichkeiten $\mu_\mathrm{n}$ und $\mu_\mathrm{p}$ immer als Mittelwerte aufzufassen, die auch vom elektrischen Feld $\boldsymbol{E}$ abhängen können; die folgenden Aussagen gelten für relativ kleine Feldstärken.
Ohne Diskussion von Details bleibt festzuhalten, dass die Ladungsträger zum einen hauptsächlich an akustischen Phononen und andererseits an gebundenen Störstellen (ionisierte *Donatoren* und *Akzeptoren*) gestreut werden. Bei niedrigen Dotierkonzentrationen beobachtet man den temperaturabhängigen Einfluss der Phononen, bei hohen Dotierkonzentrationen ist die Temperaturabhängigkeit sehr klein und die Beweglichkeit ist um 1–2 Grössenordnungen verringert.
Die Diskussion der Temperaturabhängigkeit der Leitfähigkeits–Messkurven ist noch etwas schwieriger, denn zur $T$–Abhängigkeit der Beweglichkeiten ist die der Trägerkonzentrationen zusätzlich zu bedenken.
Sehr viel einfacher dagegen sind die Widerstands–Konzentrationskurven, sie spiegeln einen eindeutigen Zusammenhang wieder. Wer eine ordentliche 4–Spitzen–Messung des Widerstands durchführt, kann bei bekanntem Dotierungstyp auf die

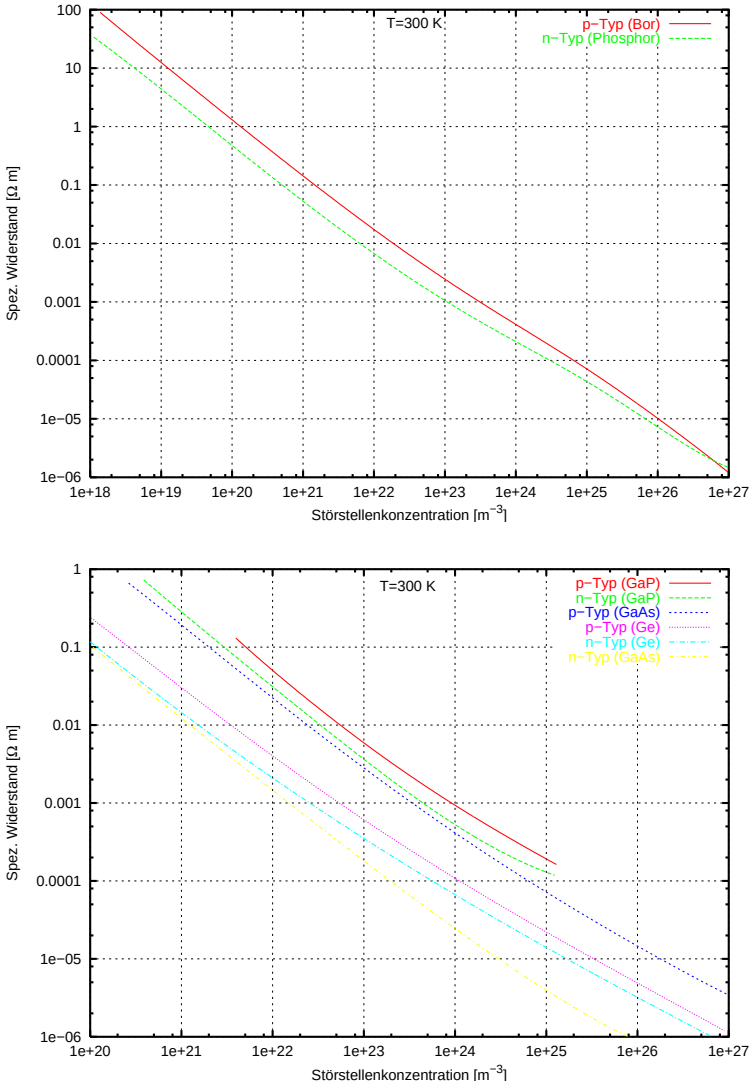



Abbildung 3.15.: Spezifischer Widerstand in Abhängigkeit von der Konzentration für verschiedene *Halbleiter*.

Dotierungskonzentration rückschliessen.

Das ohmsche Verhalten der Leitfähigkeit von Halbleitern gilt bis zu Feldstärken von typischerweise $10^3$ – $10^4$ V/cm (materialabhängig). In den aktuellen *Halbleiter–Bauelementen* mit Submikrometer–grossen Inhomogenitäten im Aufbau können Feldstärke–Werte von $10^5$ – $10^6$ V/cm auftreten. Die Driftgeschwindigkeit erreicht bei Silizium (Löcher und Elektronen) einen Sättigungswert von $10^7$ cm/s, wobei vor allem die Wechselwirkung der Ladungsträger mit den optischen Phononen hierfür verantwortlich ist. (Diese Sättigungswerte sind höher als die des *GaAs*; allerdings zeigen GaN, *GaAs* und InP bei kleineren Feldstärken ein deutlich höher liegendes Maximum in $v_\mathrm{D}(E)$.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

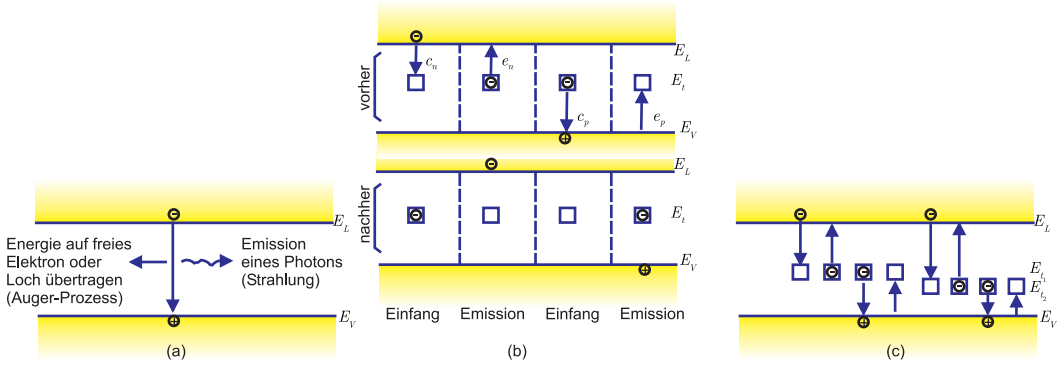



Abbildung 3.16.: Rekombinationsprozesse (gezeichnet nach Sze[**?**]).

### 3.1.6. Rekombinationsprozesse und Ladungsträgertransport: Grundgleichungen zur Funktion von Halbleiter–Bauelementen

Wenn in einem physikalischen System die Bedingung des thermischen Gleichgewichts verletzt ist, gibt es stets Prozesse, die das System wieder ins Gleichgewicht zurück bringen. Wird beispielsweise (in einem beliebig dotierten *Halbleiter*) durch optische Anregung lokal die *Ladungsträgerdichte* erhöht, so dass $p \cdot n \neq n_\mathrm{i}^2$ gilt, so relaxiert sie am Ende wieder zu $p \cdot n = n_\mathrm{i}^2$. Im *Halbleiter* geschieht dies, anders als im Metall, wesentlich durch die sog. *Rekombinationsprozesse*. Abbildung 3.16 gibt die grundlegenden *Rekombinationsprozesse* der *Halbleiter* wieder.

In Abbildung 3.16 (a) ist die sog. Elektron–Loch–Rekombination gezeigt: das Elektron macht eine Band–Band–Rekombination, die Übergangsenergie wird an ein Photon ('strahlender Rekombinationsprozess'), wichtigster Prozess bei direkten Halbleitern, oder an ein freies Elektron im Leitungsband bzw. an ein freies Loch im Valenzband ('nichtstrahlender Rekombinationsprozess') abgegeben.

Im weiteren werden die für *indirekte Halbleiter* wie *Si* so wichtigen Störstellen–Rekombinationsprozesse aufgezeigt. Eine tiefe Störstelle in (b), vereinfacht mit einem einzigen Energieniveau angenommen, kann Elektronen bzw. Löcher 'trappen' und wieder freisetzen; verschiedene Störstellen mit mehreren Energieniveaus besitzen noch mehr Rekombinationsmöglichkeiten (c).

Besonders effektiv wirken Störstellen in der Mitte der Bandlücke. (Deshalb sind Au– oder Cu–Verunreinigungen im *Si* i. allg. gefürchtet.) Gezielt eindiffundierte tiefe Störstellenatome, hochenergetische elektromagnetische Strahlung und energiereiche Partikelstrahlung ermöglichen es, die Rekombinationsraten lokal kontrolliert zu erhöhen. Meist muss man aber tiefe Traps unbedingt vermeiden.

Die Umkehr der Rekombinationsprozesse von Teilbild (a), nämlich der direkte optische Übergang (bei *direkten Halbleiter* wie *GaAs*) bzw. die Stossionisation (von Elektronen im Valenzband) bei hohen elektrischen Feldern geben zwei Wege zur Erzeugung zusätzlicher Ladungsträger im thermischen Nichtgleichgewicht an. Die lokal erhöhte Elektronen– bzw. Löcherkonzentration zerfällt räumlich durch Diffusion (aufgrund der zufälligen thermischen Bewegung der Ladungsträger) und zeitlich durch die oben eingeführten Rekombinationsprozesse.

Der Strom von Elektronen, der in der Abbildung 3.17 von links den Ort $x$ erreicht, wird aufgrund der vorhandenen Rekombinationsprozesse im Wegintervall d$x$ um

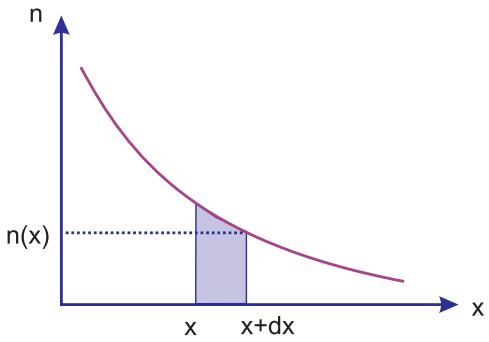



Abbildung 3.17.: Elektronendiffusion mit Rekombinationsprozessen (gezeichnet nach [**?**]).

den Betrag $\mathrm{d}j_\mathrm{n}$ geschwächt. Immer wenn in Halbleiter–Bauelementen Distanzen vergleichbar oder grösser als die sog. *Diffusionslängen* der Elektronen oder der der Löcher ($L_\mathrm{n} = \sqrt{D_\mathrm{n} \cdot \tau_\mathrm{n}}$ bzw. $L_\mathrm{p} = \sqrt{D_\mathrm{p} \cdot \tau_\mathrm{p}}$) sind, verändert die Rekombination entlang des Wegs die Stromdiche der Elektronen oder Löcher. Die entsprechenden Diffusionslängen in *Si* und Ge betragen ca. 10 mm. (Die in *GaAs* ca. $0,1$ mm.)
Zur Beschreibung von *Si–Bauelementen* darf die für den *intrinsischen Halbleiter* in Kapitel 3.1.2 angegebene Driftstromdichte (drift current) um eine Diffusionskomponente erweitert werden. Probleme mit Ladungsträger–Konzentrationsgradienten werden so behandelbar.
Die *Volumenstromdichtengleichungen* lauten:

$$\boldsymbol{j}_\mathrm{n} = |e\mu_\mathrm{n}| \cdot n \cdot \boldsymbol{E} + |e| \cdot D_\mathrm{n} \cdot \mathrm{grad}\, n \,, \tag{3.1.29}$$

‘conduction current’ + ‘diffusion current’

$$\boldsymbol{j}_\mathrm{p} = |e\mu_\mathrm{p}| \cdot p \cdot \boldsymbol{E} + |e| \cdot D_\mathrm{p} \cdot \mathrm{grad}\, p \,, \tag{3.1.30}$$

$$\text{und} \qquad \boldsymbol{j} = \boldsymbol{j}_\mathrm{n} + \boldsymbol{j}_\mathrm{p} \,, \tag{3.1.31}$$

mit $D_\mathrm{n}$ und $D_\mathrm{p}$ als *Diffusionskonstanten* (’diffusion coefficient, diffusion constant’). Damit im Halbleitermaterial im Gleichgewichtszustand *Ladungsneutralität* herrscht, muss $\boldsymbol{j} = 0$ sein. Dies ist bei *schwach dotierten Halbleitern* die einzige Bedingung. Bei *stark dotierten Halbleitern* jedoch muss für die einzelnen Ladungsträgerarten $\boldsymbol{j}_n = 0$ und $\boldsymbol{j}_p = 0$ gelten.

*Nebenbemerkung*:
Im Falle der nichtentarteten *Halbleiter* gilt die Einsteinrelation $D_\mathrm{n} = \left(\frac{k_B T}{e}\right) \cdot \mu_\mathrm{n}$ und $D_\mathrm{p} = \left(\frac{k_B T}{e}\right) \cdot \mu_\mathrm{p}$, die die Tatsache wiedergeben, dass die Diffusion der Ladungsträger von ihrer Beweglichkeit abhängt. Die angegeben Gleichungen enthalten noch keine Magnetfeld–Effekte.
Bei kleineren $E$–Feldstärken (in V/cm) gilt für den (für Bauelemente besonders interessanten) eindimensionalen Fall:

$$j_\mathrm{n} = |e\mu_\mathrm{n}|\, n E_x + |e|\, D_\mathrm{n} \frac{\mathrm{d}n}{\mathrm{d}x} = |e|\, \mu_\mathrm{n} \left( n E_x + \frac{k_B T}{e} \frac{\mathrm{d}n}{\mathrm{d}x} \right) \,. \tag{3.1.32}$$

Für $j_\mathrm{p}$ gilt die analoge Gleichung.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Unter äusserem Einfluss (optische Anregung, hohe elektrische Felder) können im *Halbleiter–Volumen* also lokal Elektronen und Löcher generiert werden: 'excess concentration of carriers'. Die zugehörigen *Generationsraten* bezeichnen wir mit $G_\mathrm{n}$ und $G_\mathrm{p}$ (in $\mathrm{cm}^3/\mathrm{s}$). Analog führen wir *Rekombinationsraten* $R_\mathrm{n}$ und $R_\mathrm{p}$ ein. Die sog. *Kontinuitätsgleichung* lauten damit:

$$\frac{\partial n}{\partial t} = G_\mathrm{n} - R_\mathrm{n} + \frac{1}{|e|}\,\mathbf{div}\,\boldsymbol{j}_\mathrm{n}\ , \tag{3.1.33}$$

und

$$\frac{\partial p}{\partial t} = G_\mathrm{p} - R_\mathrm{p} + \frac{1}{|e|}\,\mathbf{div}\,\boldsymbol{j}_\mathrm{p}\ . \tag{3.1.34}$$

Im eindimensionalen Fall und unter der Bedingung, dass die injizierte *ladungsträgerdichte* sehr viel kleiner als die Majoritätsladungsträgerdichte ist ('low injection condition') gilt:

$$R_\mathrm{n} = \frac{\mathrm{d}\Delta n}{\mathrm{d}t} \approx \frac{\Delta n}{\tau_\mathrm{n}}\ , \tag{3.1.35}$$

wobei $\Delta n$ die Abweichung der Minoritätsladungsträgerdichte vom thermodynamischen Gleichgewicht angibt; $\tau_\mathrm{n}$ steht für die Lebensdauer der (Minoritäts–) Elektronendichte. In elektrisch neutralen Raumteilen gilt $\Delta n = \Delta p$.
Im eindimensionalen Fall gilt weiter:

$$\frac{\partial n}{\partial t} = G_\mathrm{n} - \frac{\Delta n}{\tau_\mathrm{n}} + n\mu_\mathrm{n}\frac{\partial E}{\partial x} + \mu_\mathrm{n}E_x\frac{\partial n}{\partial x} + D_\mathrm{n}\frac{\partial^2 n}{\partial x^2}\ , \tag{3.1.36}$$

und analog

$$\frac{\partial p}{\partial t} = G_\mathrm{p} - \frac{\Delta p}{\tau_\mathrm{p}} + p\mu_\mathrm{p}\frac{\partial E}{\partial x} + \mu_\mathrm{p}E_x\frac{\partial p}{\partial x} + D_\mathrm{p}\frac{\partial^2 p}{\partial x^2}\ . \tag{3.1.37}$$

Einfache Beispiele für die Anwendbarkeit dieser Gleichungen sind:

1. *n–Halbleiter* unter Beleuchtung (Bestimmung der Minoritätsladungsdauer nach Stevenson und Keyes).
   Mit $\Delta n \sim \mathrm{e}^{-t/\tau}$ zerfällt nach dem Abschalten die Nichtgleichgewichtskonzentration. Bei intrinsischen Halbleitern sind die Lebensdauern typischerweise $> \mu$s. Bei dotierten Halbleitern sind $\tau_\mathrm{n}$ bzw. $\tau_\mathrm{p}$ stark von der Dotierkonzentration abhängig, die Werte reichen von 1 ns (bei $10^{21}\ cm^{-3}$) bis 10 $\mu$s (bei $10^{15}\ cm^{-3}$) bei *Si*.

2. Überschussladungsträger–Injektion von einer Seite (Bestimmung der Diffusionslängen).

3. Punktförmige optische Anregung mit und ohne elektrisches Feld (Diffusionsexperiment nach Hayes und Shockely).

4. Oberflächen–Rekombination.

Die Tatsache, dass ein Halbleitereinkristall ein Ende hat, bedeutet, dass er immer an dieser Oberfläche lokalisierte Störstellen besitzt. Deren Dichte kann sehr gross sein (ca. $10^{15}\ \mathrm{cm}^{-2}$). Energetisch liegen sie gerade zwischen $E_\mathrm{V}$ und $E_\mathrm{L}$ und damit bilden sie tiefe *Oberflächenstörstellen* ('surface trapping centers'). (Für vollkommen saubere Oberflächen nennt man diese Oberfächenzustände nach ihrem Entdecker Tamm–Zustände. Sie sind bedingt durch die freien, unabgesättigten Valenzen

der in ihrer Lage leicht verschobenen Oberflächenatome. Gebundene Fremdatome bewirken ebenfalls Störstellen mit allerdings deutlich anderen Eigenschaften. Man trachtet immer danach, diese zu vermeiden.) Die obigen Kontinuitätsgleichungen sind entsprechend zu ergänzen, für Elektronen lautet sie:

$$\boldsymbol{j}_{\mathrm{n}} = |e|\, S_{\mathrm{n}}\, \Delta n\, \boldsymbol{e}_A, \tag{3.1.38}$$

dabei wird $\boldsymbol{e}_A$ als Einheitsnormalenvektor an der Oberfläche des Halbleiters und $S_{\mathrm{n}}$ als *Oberflächenrekombinations–Geschwindigkeit* ('surface recombination velocity') eingeführt. Die *Oberflächenrekombinations–Geschwindigkeit* charakterisiert jedes Interface, nur bei einer Dotierungsgrenzfläche (p–n oder n–$n^+$) ist sie vernachlässigbar. Abschliessend einige Zahlen:

| | |
|---|---|
| für *Metall–Halbleiter–Kontakte*: | $S_{\mathrm{n}} \approx 10^6$ cm/s , |
| für Si–$SiO_2$–Grenzflächen: | $S_{\mathrm{n}} \approx 1$ cm/s . |

Den wichtigen Anwendungsfall der Injektion durch vorwärtsgespannte p–n–Übergänge behandeln wir später. (In den Gleichungen Gleichung (3.1.36) und Gleichung (3.1.37) sind im Falle der p–n–Diode die Grössen $\Delta n$ durch $\Delta n_{\mathrm{p}}$ und $p$ durch $p_{\mathrm{n}}$ zu ersetzen, sprich Elektronen (Minoritätsladungsträger) im p–Gebiet, Löcher (ebenfalls Minoritätsladungsträger) im n–Gebiet.)

## 3.2. Phänomene elektrischer Kontakte

### 3.2.1. Grundlagen

Ein elektrischer Kontakt zwischen zwei Medien besteht, wenn durch ihre Grenzfläche ein Ladungsträgertransport möglich ist. Im Folgenden betrachten wir einige, für die *Halbleiter–Bauelemente* relevante Beispiele:

1. Die abrupte Grenzfläche verschiedener Dotierungen in einem *Halbleiter–Einkristall*: p–n–Übergänge ('homo–junctions').

2. Die Grenzfläche zwischen verschiedenen, epitaktisch gewachsenen *Halbleiter–Schichten*: 'Hetero–Übergänge' ('hetero–junctions') (Bsp. $GaAs - AlGaAs$).

3. Grenzflächen zwischen zwei Materialien: (Metall–Vakuum,) *Metall–Halbleiter*: Schottky–Dioden und ohmsche Kontakte.

4. Grenzflächen zwischen drei Materialien: Metall–Isolator–Halbleiter: MOS–Übergänge (metall–oxide–semiconductor).

Die Wirkungsweise der elektrischen Bauelemente beruht meist auf dem Ladungstransport durch Kontakte. Dieser ist praktisch ausschliesslich durch den ortsabhängigen Verlauf der potentiellen Energie der Ladungsträger in der unmittelbaren Umgebung der Grenzfläche festgelegt. Den sog. Potentialverlauf $V(x)$ erhält man durch Divison der potentiellen Energie für Elektronen (bzw. für Löcher) mit $-e$ bzw. $(+e)$.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Ursache für diese Ortsabhängigkeiten sind *Raumladungen*, im Gegensatz zum Metall–Metall–Kontakt. Den formalen Zusammenhang zwischen dem elektrischen Potential, auch Makropotential genannt, und der gesamten elektrischen *Ladungsträgerdichte* $\rho(x)$ liefert (für statische oder niederfrequente Betrachtungen) die *Poisson–Gleichung* der Elektrostatik:

$$\frac{\partial^2 V(x)}{\partial x^2} = -\frac{1}{\epsilon\epsilon_0}\rho(x) \; . \tag{3.2.1}$$

In dieser eindimensionalen Form ist sie Fixpunkt der folgenden quasiklassischen Beschreibung der allmählichen Bandverbiegung, deren auslösende Ursache ideale, auch abrupte Dotierwechsel sein können.
Einen Ausgangspunkt zur Erklärung eines Kontaktphänomens bietet das *Bänderschema* rund um die ebene Grenzfläche im *thermodynamischen Gleichgewicht*. Über den Kontakt hinweg ist die *Fermi-Energie* im ganzen Bauelement konstant; im Bandschema liegt sie also stets waagrecht. Der resultierende Strom von Ladungsträgern einer bestimmten Energie ist in jeder Raumrichtung gleich Null. (Erinnerung: Das chemische Potential, d. h. die Änderung der freien Energie mit der Teilchenzahl bei konstantem Volumen und konstanter Temperatur, ist gleich der Fermi–Grenzenergie.)

## 3.2.2. p–n–Übergänge

Denken wir uns einen *Si–Kristall*, der inhomogen dotiert wurde, in dem in der linken Hälfte *Akzeptoren* (B, Al, Ga) und in der rechten *Donatoren* (P, As, Sb) eingebracht wurden (siehe Dotierverfahren, später); zur Vereinfachung nehmen wir einen abrupten Wechsel der Störatomsorten an. Die Abbildung 3.18 gibt den Gang der Argumentation wieder.
Zunächst denkt man sich die beiden Kristallhälften getrennt; die *Fermi–Niveaus* liegen in beiden Gebieten verschieden hoch bezogen auf dieselbe Energieskala. Für $T \approx 300$ K sind die Störstellen nahezu vollständig erschöpft, $E_\mathrm{F}$ liegt entsprechend oberhalb von $E_\mathrm{A}$ bzw. unterhalb von $E_\mathrm{D}$ (b). Im nächsten Schritt setzt man die beiden Teile zusammen: im thermischen Gleichgewicht muss das Fermi–Niveau als elektrochemisches Potential ($E_\mathrm{F} = E'_\mathrm{F} + eU$) gleich hoch und waagrecht sein. Im Übergangsbereich kommt es zu einer sog. Bandverbiegung, während ausserhalb die Ausgangslagen ungeändert erhalten bleiben (c). Die bereits eingeführte Poissongleichung Gleichung (3.2.1) verknüpft die Krümmung des Makropotentials mit einer *Raumladung* $\rho(x)$, die über den sog. Raumladungsbereich oder die sog. *Raumladungszone* ausgedehnt ist (d). Im Falle der Störstellenerschöpfung sind links (fast) alle *Akzeptoren* geladen ($N_\mathrm{A}^-$) und rechts ebenso (fast) alle *Donatoren* ($N_\mathrm{D}^+$). Ausserhalb der *Raumladungszone* herrscht Ladungsträgerneutralität; links sorgen die positiven Löcher als Majoritätsladungsträger ($\mathrm{p_p}$) hierfür, rechts die negativen Elektronen als Majoritätsladungsträger ($\mathrm{n_n}$, sprich Elektronen (n) im n–Gebiet) (e). Die frei beweglichen Löcher und Elektronen diffundieren aufgrund des Konzentrationsgradienten ins angrenzende n– bzw. p–Gebiet. Dort sind sie Minoritätsladungsträger und ihre Konzentrationen werden als $\mathrm{n_p}$ (Elektronen im p–Gebiet) bzw. $\mathrm{p_n}$ (Löcher im n–Gebiet) geschrieben; sie rekombinieren kräftig mit den vorhandenen Majoritätsladungsträgern. Folglich bleiben die Ladungen der ionisierten unbeweglichen Störstellen im Bereich der *Raumladungszone* unkompensiert; im p–Gebiet bleibt also eine negative, im n–Gebiet eine positive, ortsfeste

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

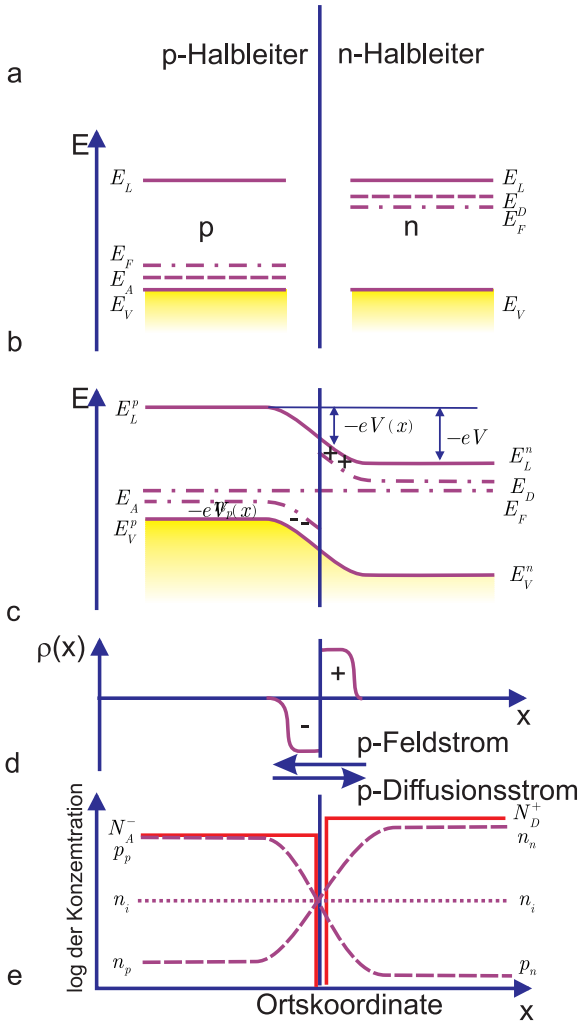



Abbildung 3.18.: Qualitatives Schema eines p–n–Übergangs im thermischen Gleichgewicht bei Raumtemperatur (gezeichnet nach Ibach-Lüth[?]). Beachten Sie bitte die logarithmische Auftragung!

*Raumladung* übrig. Dieses elektrische Feld der 'inneren Diode' ist Ursache für Feldströme, die den Diffusionsströmen der Löcher und Elektronen im thermischen Gleichgewicht entgegenfliessen und ihnen die Waage halten. (Aufgrund der endlichen Temperatur werden ständig und überall Elektron–Loch–Paare generiert und diese rekombinieren natürlich wieder.) Im thermischen Gleichgewicht gilt auch im Raumladungsbereich $n \cdot p = n_\mathrm{i}^2$.
Die sich einstellende Diffusionsspannung $V_\mathrm{D}$ (maximale Differenz des elektrischen Potentials $V(x)$) hängt natürlich von den Dotierkonzentrationen ab:

$$eV_\mathrm{D}(T) = k_B T \ln \frac{\mathrm{p_p}(T)\mathrm{n_n}(T)}{n_\mathrm{i}^2} < E_\mathrm{g} \; . \tag{3.2.2}$$

Die Leitungsbandkante ist dann $E_\mathrm{L}^\mathrm{p} - eV(x)$ und die ortsabhängige Elektronenanzahldichte lässt sich (im Falle des nichtentarteten Halbleiters) angeben mit:

$$n(x) = N_\mathrm{eff}^\mathrm{L} \exp\left(-\frac{E_\mathrm{L}^\mathrm{p} - eV(x) - E_\mathrm{F}}{k_B T}\right) \; . \tag{3.2.3}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

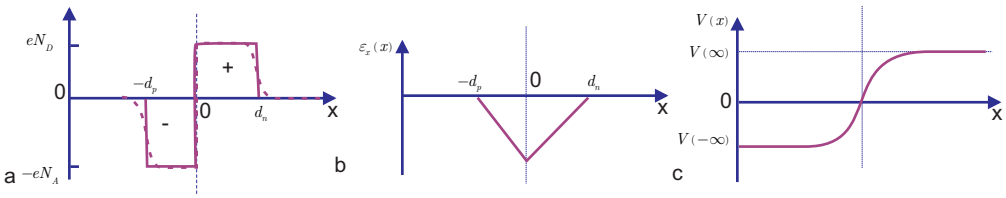



Abbildung 3.19.: Das Schottky–Modell der *Raumladungszone* eines p–n–Übergangs (bei $x = 0$) (Zeichnung nach Ibach-Lüth[**?**]).

Die Berechnung des Verlaufs der Raumladungsdichte $\rho(x)$ ist schwierig. Für den abrupten p–n–Übergang greift man daher zum 'Schottky–Modell' der *Raumladungszone*. In Abbildung 3.19 ist dies wiedergegeben.
Der unbekannte $\rho(x)$–Verlauf wird durch Kastenfunktionen angenähert. Die zugehörige Poissongleichung ist leicht lösbar, man erhält:

$$E_x(x) = -\frac{e}{\epsilon\epsilon_0} N_\mathrm{D}\,(d_\mathrm{n} - x)\;, \tag{3.2.4}$$

$$V(x) = V_\mathrm{n}(\infty) - \frac{e}{2\epsilon\epsilon_0} N_\mathrm{D}\,(d_\mathrm{n} - x)^2\;, \tag{3.2.5}$$

und damit die Bandverläufe. Die Ausdehnungen der *Raumladungszone* $(\mathrm{d_n} + \mathrm{d_p})$ kann man bestimmen zu

$$d_\mathrm{n} = \left(\frac{2\epsilon\epsilon_0 V_\mathrm{D}}{e}\,\frac{N_\mathrm{A}/N_\mathrm{D}}{N_\mathrm{A} + N_\mathrm{D}}\right)^{1/2}\;, \tag{3.2.6}$$

$$\text{und} \qquad d_\mathrm{p} = \left(\frac{2\epsilon\epsilon_0 V_\mathrm{D}}{e}\,\frac{N_\mathrm{D}/N_\mathrm{A}}{N_\mathrm{A} + N_\mathrm{D}}\right)^{1/2}\;. \tag{3.2.7}$$

Bei *Si* und bei typischen Störstellenkonzentrationen von $10^{14} - 10^{18}\ cm^{-3}$ beträgt $V_\mathrm{D} \approx 0,5 - 0,8$ V $< E_\mathrm{g}$ und es sind Ausdehnungen von $\mathrm{d_n}$ bzw. $\mathrm{d_p}$ von 1000 – 10 nm zu erwarten. Die Feldstärken in den Raumladungszonen liegen zwischen $10^4$ und $10^6$ V/cm, die Feldströme bzw. die kompensierten Diffusionsströme bei einigen kA/mm$^2$. Das sind beträchtliche Werte!

**Dotiertechniken**
Ein zentraler Schritt bei der *Herstellung von p–n–Dioden* ist die *Dotierung*. Es gibt hierzu mehrere Fabrikationswege. Die einfachste Art ist in Teilbild a) von Abbildung 3.20 gezeigt: durch Erwärmen wird der feste, im direkten Kontakt stehende Dotierstoff aufgeschmolzen und einlegiert. Definierter sind die nächsten beiden Ofen–Diffusionsprozesse. Im Teilbild b) wird nach der grossflächigen Diffusion (bei $800° - 1200°$ C) der Kontakt lateral durch einen Ätzprozess festgelegt; es entsteht eine 'Mesastruktur'. Im Teil c) dagegen wird, wie heute allgemein üblich, durch eine Öffnung in einer Oxidmaske eindiffundiert. Die seitliche Diffusion unter die Oxidmaske ist allerdings unvermeidlich und begrenzt die erzielbare laterale Auflösung. Im modernen MOS–Prozess schliesslich kommt das letztgezeigte Verfahren zum Einsatz. Mit Hilfe eines 'Ionenimplanters' werden bei Beschleunigungsspannungen zwischen 50 und 300 keV ionisierte Störatome implantiert. Sie sitzen auf Leerstellen und in der Hauptsache auf Zwischengitterplätzen und müssen in einem weiteren Ofenprozess bei ca. 900° C erst noch auf Gitterplätze gebracht werden ('elektrische Aktivierung').

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

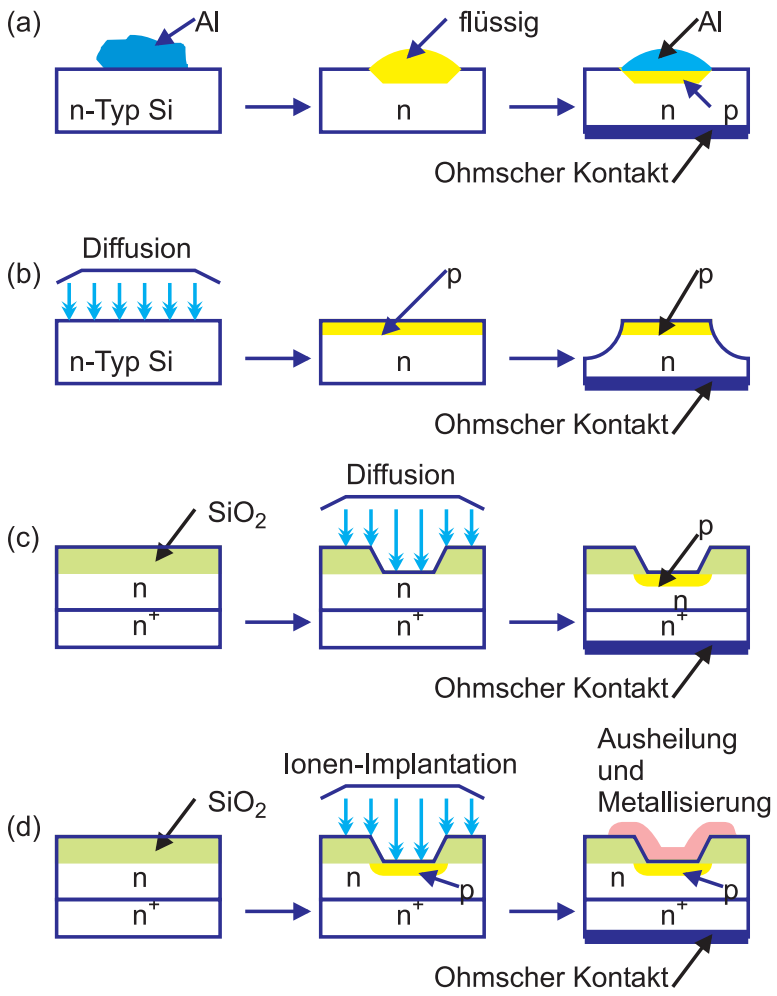



Abbildung 3.20.: Einige Herstellungsmethoden für p–n–Dioden (gezeichnet nach Sze[?]).

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

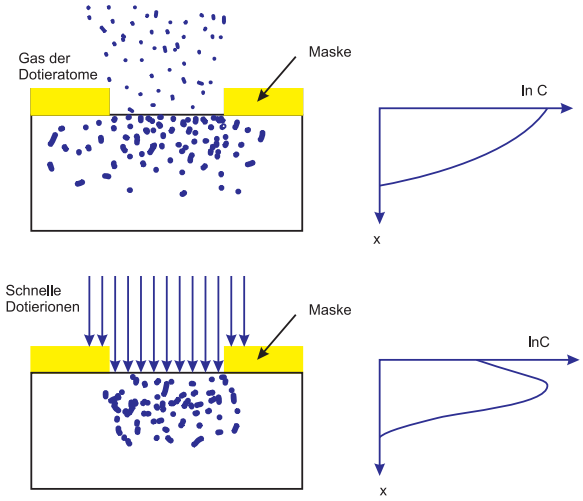



Abbildung 3.21.: Vergleich der Dotierprofile, die mit Hilfe eines Ofenprozesses (oben) und der Ionenimplantation hergestellt wurden, schematisch[**?**].

Bei der Diffusion helfen einem mehrere glückliche Umstände. Erstens ist die Löslichkeit von As, B und P in *Si* grösser als die von z. B. Cu, Au oder O. Zum zweiten ist der Diffusionskoeffizient der gängigen Dotierstoffe gerade klein, d. h. ionenimplantierte Dotierprofile unter der Oberfläche überstehen den Aktivierungsprozess und nachfolgende, mit hohen Temperaturen verbundene Prozessschritte relativ unbeschadet.
Die Ofen–gestützen Dotierprozesse und die Ionenstrahl–gestützten Dotiermethoden erzeugen unterschiedliche Dotierprofile unter der Siliziumoberfläche, siehe Bild 3.21 Mehrere Ionenimplantier–Prozessschritte mit unterschiedlichen Beschleunigungsenergien ermöglichen die Anlage von in der Tiefe gestaffelten, ausgedehnten oder modulierten Dotierprofilen (vergl. CMOS–Herstellung).

## 3.2.3. Vorgespannte p–n–Übergänge — gleichrichtende Dioden

Legt man eine zeitlich konstante, äussere Spannung $U$ an einen p–n–Übergang, so erhält man einen stationären Zustand nahe am thermischen Gleichgewicht. Die Fermi–Energie als elektrochemisches Potential neigt sich jedoch nicht gleichmässig zur positiv vorgespannten Seite, sondern es fällt fast der gesamte Betrag von $U$ über der *Raumladungszone* ab. Diese hat einen wesentlich höheren elektrischen Widerstand als die nebenliegenden Halbleiter–Gebiete. Der Grund liegt in der oben beschriebenen Rekombination von freien Löchern und freien Elektronen; die *Raumladungszone* ist gekennzeichnet durch eine Verarmung an freien Ladungsträgern (sog. Verarmungszone).
Über der *Raumladungszone* falle also jetzt statt der Diffusionsspannung der Wert

$$V_{\mathrm{D}} - U = V_{\mathrm{n}}(\infty) - V_{\mathrm{p}}(-\infty) \tag{3.2.8}$$

ab. Entsprechend dem Vorzeichen unterscheidet man zwei Fälle:

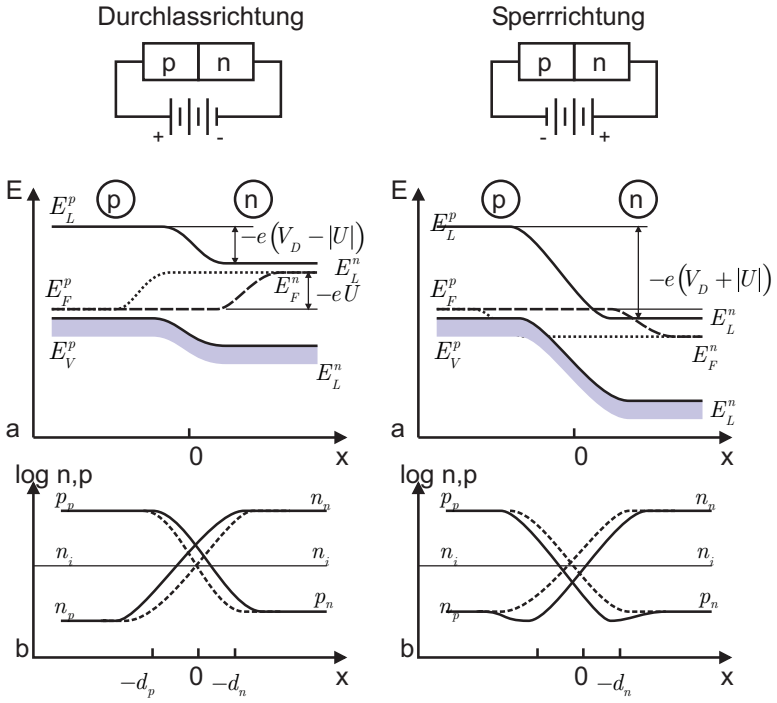
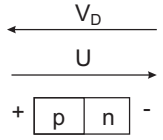
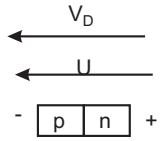



Abbildung 3.22.: Schema eines p–n–Übergangs, der in Durchlass– bzw. Sperrrichtung gepolt ist (Nichtgleichgewichtszustände) (gezeichnet nach [**?**]).

1. 


(p–Teil positiv gegen n–Teil):
'Flussrichtung oder Durchlassrichtung'
Potentialhöhe: $eV_\mathrm{D} - eU$,
die Diffusionsspannung wird erniedrigt.

2. 


(p–Teil negativ gegen n–Teil):
'Sperrrichtung"
Potentialhöhe: $eV_\mathrm{D} + eU$,
die Diffusionsspannung wird erhöht.

Die Ausdehnung der *Raumladungszone* ändert sich: in Flussrichtungs–Polung wird sie schmaler, in Sperrrichtungs–Polung wird sie breiter. Aufgrund der Verarmung an freien Ladungsträgern hat dies direkte Folgen für die Leitfähigkeit der gesamten Anordnung, wir haben eine Diode.

Abbildung 3.22 gibt die Verhältnisse für den Fall des bereits eingeführten symmetrischen, abrupten p–n–Übergangs wieder. In Durchlassrichtung erniedrigt sich die Potentialhöhe. Um näherungsweise eine Beschreibung der Besetzung der Bänder mit Hilfe der Boltzmann–Statistik zu ermöglichen, werden satt des Fermi–Niveaus formal zwei sog. Quasi–Fermi–Niveaus für die Elektronen und für die Löcher eingeführt, obwohl strenggenommen kein thermisches Gleichgewicht mehr herrscht; das Massenwirkungsgesetz gilt in der *Raumladungszone* nicht mehr. (Das Fermi–Niveau liegt in der Mitte zwischen den Quasi–Fermi–Niveaus.)

Die Ladungsträgerdichten sind erhöht, ihre Ausläufer reichen diffusionsbedingt in die neutralen Gebiete. Die Elektronen laufen von rechts gegen die Potentialschwelle

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

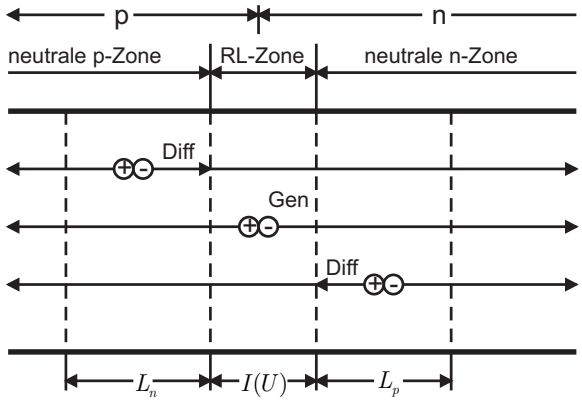



Abbildung 3.23.: Schematische Darstellung des Zustandekommens des Sperrstromes einer p–n–Diode mit Angabe des jeweils begrenzenden Mechanismuses(gezeichnet nach [**?**]).

und ein Bruchteil überwindet diese gemäss dem Boltzmannfaktor

$$\sim \quad \exp\left[\frac{-e\,(V_{\mathrm{D}} - U)}{k_B T}\right] \ .$$

Der Strom in Durchlassrichtung hängt also stark ab von der angelegten Spannung. Weil in Flussrichtung die Trägerdichten grösser sind als die Gleichgewichsdichten, überwiegt die Rekombination: der Ladungstransport erfolgt durch eindiffundieren von Elektronen und Löchern in die neutrale Zone und ihre dortige Rekombination: 'Rekombinationsströme'.
In Sperrpolung ist es anders: die Trägerdichten in der *Raumladungszone* und in den naheliegenden neutralen Zonen sind unter den Gleichgewichtswerten; das System antwortet mit einer thermischen Netto–Generation von Ladungsträgern in diesem Bereich. Im elektrischen Feld der *Raumladung* werden Elektronen aus dem p–Gebiet ins n–Gebiet und Löcher aus dem n–Gebiet ins p–Gebiet ('Feldströme der Minoritätsladungstäger') gezogen. Diese 'Generationsströme' sind praktisch unabhängig von der Potentialhöhe und also unabhängig von der angelegten Spannung; sie sind proportional zu den entsprechenden Trägerdichten. Sie liefern den Sperrstrom der Diode.
Bildet man die Summe der Rekombinations– und Generationsströme von Löchern und Elektronen, so erhält man für die Kennlinie eines p–n–Übergangs (nur für die 'innere Diode', ohne Bahnwiderstände):

$$I(U) = \left(\underbrace{I^{\mathrm{gen}}(n_{\mathrm{p}}) + I^{\mathrm{gen}}(p_{\mathrm{n}})}_{I_{\mathrm{Sättigung}}}\right)\left[\exp\left(\frac{\pm eU}{mk_B T}\right) - 1\right] \ . \qquad (3.2.9)$$

Diese Gleichung beschreibt das Verhalten der Minoritätsladungsträger. Diese werden laufend am Rande der Verarmungszone gebildet. Der Vorfaktor ist materialabhängig.
Dies ist die bekannte, extrem asymmetrische Gleichrichter–Kennlinie der Diode, wie sie im Bild unten gezeigt wird. $m$ ist ein Korrekturfaktor ('Idealitätsfaktor'). Massgebend für diesen Kurvenverlauf ist also der Potentialverlauf am p–n–Übergang und der Rekombinations–/Generationsmechanismus.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

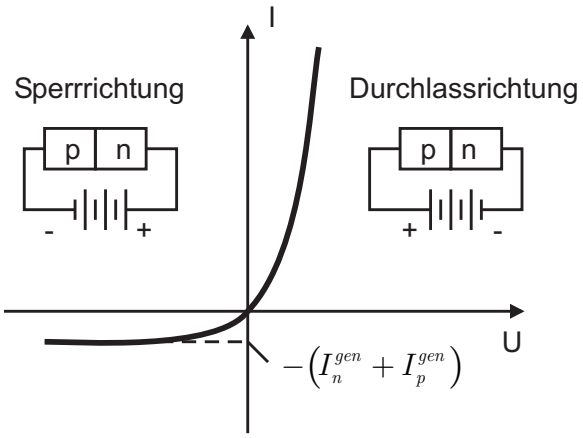



Abbildung 3.24.: Schema der Strom–Spannungs ($I$–$U$)–Kennlinie eines p–n–Übergangs(gezeichnet nach [**?**]).

*Nebenbemerkung*:
Die Berechnung von $I_S$ und $m$ ist für Spezialfälle von Shockley und Mitarbeitern durchgeführt worden. Für den Fall der sog. Diffusionsstrom–Näherung bei geringer Ladungsträgerinjektion lautet die Lösung:

$$j(U) = \left(\frac{eD_p}{L_p}p_n + \frac{eD_n}{L_n}n_p\right)\left[\exp\left(\frac{eU}{mk_BT}\right) - 1\right] \tag{3.2.10}$$

und beschreibt einen Kennlinienabschnitt im Durchlassbereich, vergleiche Bild 3.25. Diese Gleichung verknüpft alle in Kapitel 3.1.6 eingeführten Grössen in einfacher Weise.
Nicht mehr gültig ist die Kennlinienformel oberhalb der Durchbruchspannung; die Ursache hierzu kann eine im Feld der *Raumladungszone* ausgelöste Stossionisation sein ($e^-$ aus Valenzband ins Leitungsband), die sich lawinenartig vermehrt ('Lawineneffekt', 'Lawinendurchbruch', 'Ladungsträgermultiplikation') oder aber der sog. Zenereffekt, ein Tunneleffekt vom Valenzband im p–Teil zum Leitungsband im n–Teil.
*Nebenbemerkung*:
Esaki–Dioden (sehr hoch und steil dotierte Dioden) zeigen Tunneleffekte für beide Polungen.
*Nebenbemerkung*:
Symmetrische abrupte p–n–Übergänge sind selten, meist findet man asymmetrische Übergänge mit fertigungsbedingten ausgeschmiertem Dotierprofil. Reale Kennlinien weichen deshalb von der oben gezeigten Formel häufig ab.
Abschliessend soll noch auf die kapazitive Wirkung der in der *Raumladungszone* gespeicherten Ladung hingewiesen werden. Mit dem Anlegen einer äusseren Spannung ändert sich die Zone, d. h. man transportiert Majoritäts– und Minoritätsladungsträger; im ersten Fall entspricht dies der sog. Sperrschichtkapazität, im zweiten Fall der sog. Diffusionskapazität.

## 3.2.4. Hetero–Übergänge

Mit der Verfügbarkeit moderner Epitaxieverfahren wie der Molekularstrahlepitaxie und der Gasphasenepitaxie (MBE, molecular beam epitaxy, oder MOVCD,

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

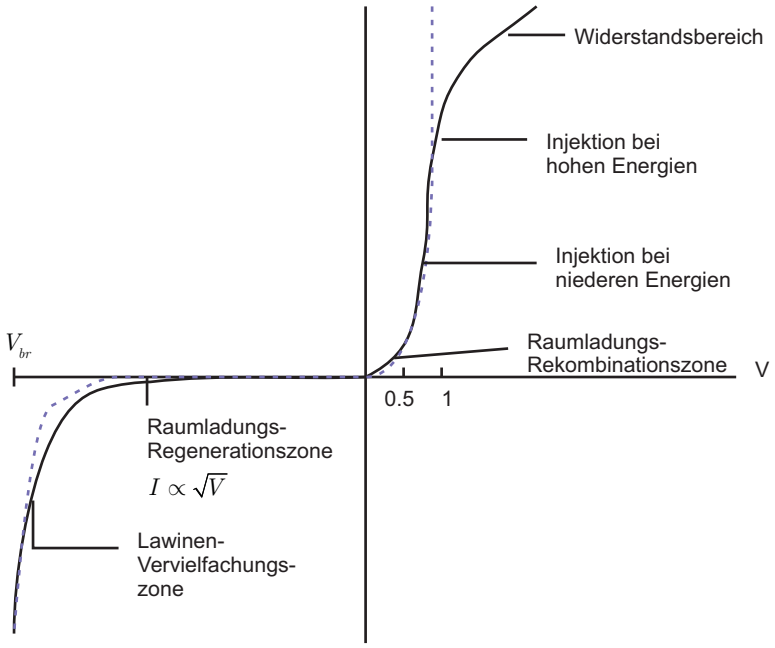



Abbildung 3.25.: $I$–$U$–Kennlinie einer p–n–Diode mit den Geltungsbereichen verschiedener Theorien bzw. Effekte (gezeichnet nach [**?**]).

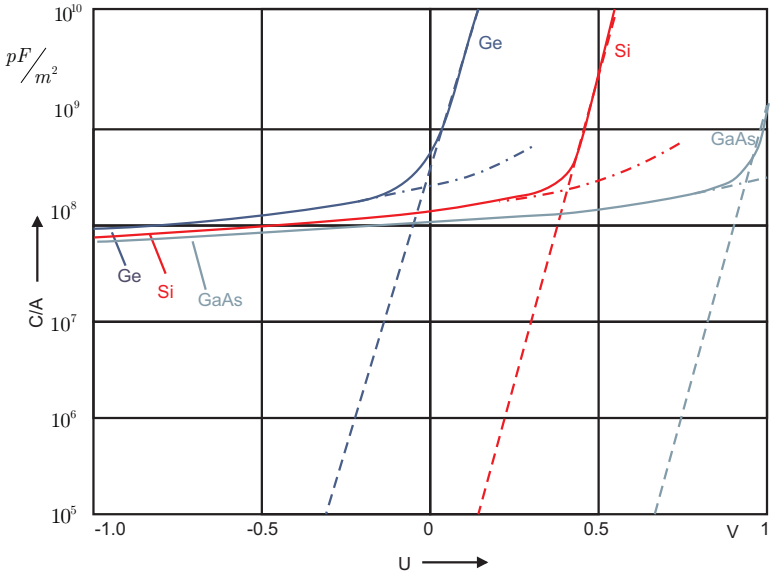



Abbildung 3.26.: Sperrschichtkapazität (strichpunktiert), Diffusionskapazität (gestrichelt) und Gesamtkapazität (voll ausgezogen) je Flächeneinheit einer $p^+n$–Diode als Funktion der Diodenspannung $U$ für Zimmertemperatur, berechnet für
Ge: $N_D = 10^{21}\ m^{-3}$, $N_A = 10^{24}\ m^{-3}$, $\tau_p = 10^{-3}$ s;
Si: $N_D = 10^{21}\ m^{-3}$, $N_A = 10^{24}\ m^{-3}$, $\tau_p = 10^{-3}$ s;
GaAs: $N_D = 10^{21}\ m^{-3}$, $N_A = 10^{24}\ m^{-3}$, $\tau_p = 10^{-3}$ s.
(gezeichnet nach [**?**])

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

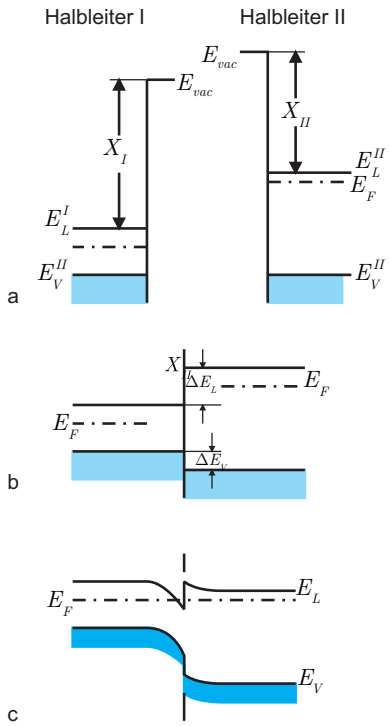



Abbildung 3.27.: *Bänderschemata*, die sich bei der Bildung einer Heterostruktur aus den Halbleitern I und II ergeben(gezeichnet nach [**?**]).

metal organic chemical vapour deposition)ist es möglich geworden, zwei verschiedene *Halbleiter* kristallin aufeinander aufzuwachsen. Zum Beispiel *GaAs* und AlAs oder, mit tolerablen Vorspannungen, SiGe auf *Si*. Mit der Variation der Zusammensetzung der Legierung (z. B. aus GaAlAs) kann man die *Energielücke* (z. B. zwischen $1,4 - 2,2$ eV) gezielt einstellen. Die Übergänge können extrem scharf, d. h. auf atomarer Skala von Lage zu Lage wechselnd, hergestellt werden. (Schichtverfahren, siehe späteres Kapitel).
Was geschieht nun, wenn man die beiden *Halbleiter* in Kontakt bringt? Es gibt zwei unterschiedliche Gesichtspunkte: erstens muss es aufgrund der unterschiedlichen Bandlückenbreiten sog. *Banddiskontinuitäten* $\Delta E_\mathrm{V}$ und $\Delta E_\mathrm{L}$ geben und zweitens wird es — genau wie beim p–n–Übergang in *Si* — eine *Raumladungszone* und also auch *Bandverbiegungen* geben, nur dass es hierfür mehr Möglichkeiten gibt (n–n, p–p, p–n, etc.).
Die Anpassung der beiden Bandstrukturen aneinander erfolgt von *Atomlage* zu *Atomlage*, die elektrischen Felder sind also in der Grössenordnung atomarer Felder ($\gtrsim 10^8$ V/cm), sie sollten also die Banddiskontinuitäten festlegen. Das von Anderson stammende, in vielen Büchern zitierte Modell sieht dies auch so vor: die *Bänderschemata* der beiden *Halbleiter* sind so zusammenzusetzen, dass die Vakuum–Energieniveaus $E_\mathrm{vac}$ (oder $E_\infty$) gleich sind. Für die Diskontinuität des Leitungsbandes $\Delta E_\mathrm{L}$ kann man aus Abbildung 3.27 direkt ablesen:

$$\Delta E_\mathrm{L} = \chi_\mathrm{I} - \chi_\mathrm{II} = \Delta\chi \ , \tag{3.2.11}$$

mit den Elektronenaffinitäten $\chi_i$, den Differenzen zwischen $E_\mathrm{vac}$ und den unteren Leitungsbandkanten.
(Erinnerung: Das sog. Vakuumpotential ist die potentielle Energie des Elektrons

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

im Vakuum in unendlicher Entfernung vom Kristallvolumen. Es gilt: $\chi = E_L - E_\infty$. Man könnte auch die Austrittsarbeiten $\Phi = E_F - E_\infty$ zur Diskussion heranziehen.) Das Anderson–Modell liefert aber nicht die richtigen Werte, weil die bereits früher angesprochenen elektronischen Grenzflächenzustände lokal als *Donatoren* oder *Akzeptoren* zusätzlich wirksam sind. Wir folgen Ibach und Lüths[**?**] phänomenologischer Vorgehensweise und benützen für $\Delta E_L$ bzw. $\Delta E_V$ experimentell gefundene Werte.
Die Bandverbiegungen erhalten wir auf dieselbe Weise wie bei der *Si–Diode*. Im thermischen Gleichgewicht sind die durch eventuelle Dotierungen festgelegten Fermi–Niveaus gleich; das bedingt einen Ladungstransfer, Rekombination und damit wieder die Ausbildung einer *Raumladungszone*. Wir erwarten wieder ausgedehnte Bandverbiegungen (typisch einige 10 nm) und Raumladungsstärken in der Grössenordnung $10^5$ V/cm. Die einfachste theoretische Beschreibung erfolgt wieder mit dem Schottky–Modell, die verschiedenen Dielektrizitätskonstanten bewirken aber zwei verschiedene Diffusionsspannungen bzw. äussere Spannungen. In Bild 3.28 sind die Verhältnisse an einem sog. p–N–Heteroübergang wiedergegeben. Die *negative Raumladung* im p–Teil senkt die Bänder wieder zur Grenzschicht hin ab, die *positive Raumladung* im n–Teil rechts erhöht wieder die Potentiale. Die Diskontinuität $\Delta E_V$ erscheint als senkrechtes Verbindungsstück in der Grenzfläche.

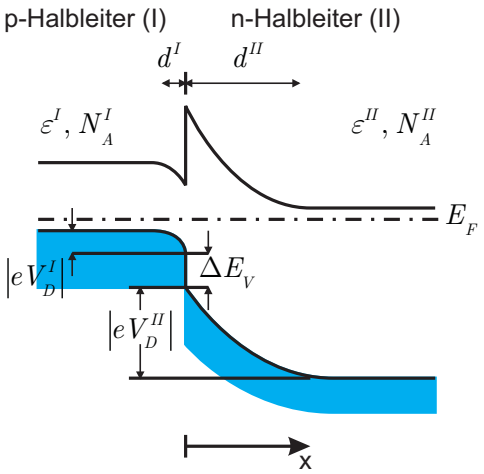



Abbildung 3.28.: *Bänderschema* eines *Halbleiter–Heteroübergangs* (gezeichnet nach [**?**]).

Hochinteressante physikalische Eigenschaften erhält man bei sog. *isotypen Heteroübergängen*; diese sind aus verschiedenen Halbleitern mit gleicher Dotierung zusammengesetzt, also z. B. ein n–N–Übergang. Im n–Gebiet der *Raumladungszone* entsteht eine Potentialtopf–artige *Anreicherungs–Raumladungszone* mit lokal sehr stark angehobener Elektronenkonzentration; dabei kann die Dotierung im n–Gebiet fast intrinsisch sein, eine kräftige Dotierung im N–Gebiet liefert ja die Ladungsträger. Die zurückbleibenden ionisierten Donatoratome sind aufgrund der räumlichen Trennung nicht mehr in der Lage, den Elektronentransport in der Grenzfläche durch Streuung zu schwächen. (Hohe Dotierung bedeutet normalerweise hohe Streuung, hier nicht!) Diese einseitige Dotierung wird '*Modulationsdotierung*' genannt. Bringt man noch eine ca. 10 nm dicke undotierte Schicht zwischen n– und N–Gebiet, so vermeidet man die Streuprozesse von lokalen Grenzflächenstörstellen zusätzlich. Auf diese Weise werden extrem hohe Beweglichkeiten

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

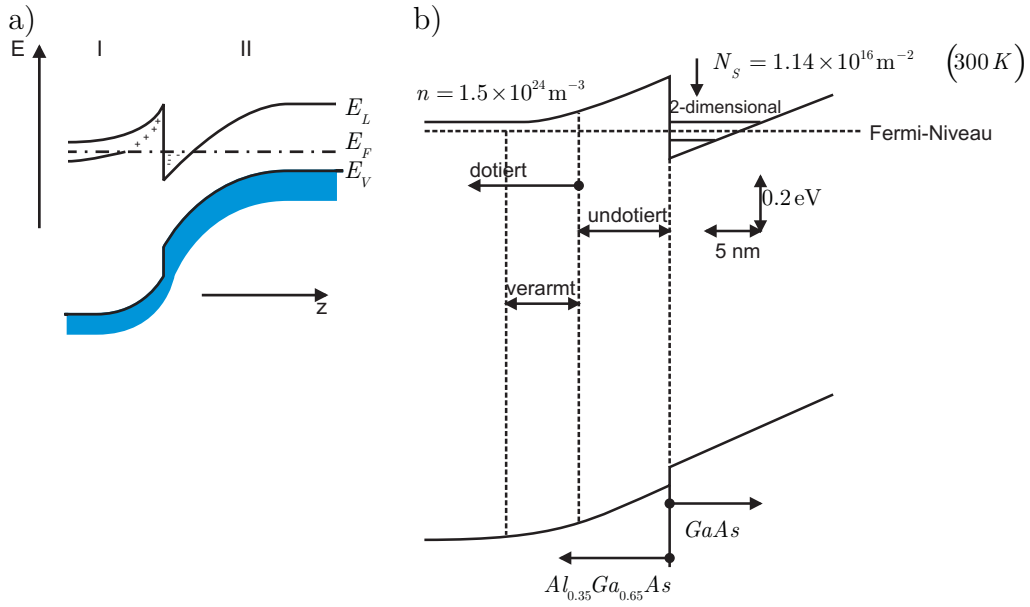



Abbildung 3.29.: *Modulationsdotierter Heteroübergang* (gezeichnet nach [**?**, **?**]).

($> 10^6\ \frac{\mathrm{cm}^2}{\mathrm{Vs}}$) bei tiefen Temperaturen erreicht.
Wächst man eine ganze Serie von solchen Schichtpaketen so auf, dass ein sog. *Kompositionsgitter* entsteht, so hat man eine ganze Reihe von Potentialtöpfen (sog. Quantentöpfen) zum parallelen Stromtransport zur Verfügung (siehe Abbildung 3.30). Zur Kontrolle des Wachstums verwendet man RHEED (Siehe Abschnitt 4.10.16.4).
Wenn die Schichtdicken genügend klein sind und damit die Quantentöpfe genügend eng ($< 10$ nm), so treten in der Richtung senkrecht zur Schicht neue Quantisierungseffekte auf. Anschaulich: die Ladungsträger sind in einer Richtung eingesperrt, senkrecht dazu in der Kontaktebene sind sie frei beweglich.
Die theoretische Untersuchung löst i. allg. die zeitunabhängige Schrödingergleichung; im einfachsten Fall nimmt man in der relevanten Richtung ein unendlich hohes Rechteckpotential an, den Fall des isotropen Heteroübergangs löst man näherungsweise mit einem Dreieckspotential. Abbildung 3.31 zeigt die neuen Energieeigenwerte, die zugehörigen Subbänder und die konstante (!) Zustandsdichten der Subbänder. Für einen einzelnen Quantentopf erhält man scharfe Energiezustände. In Übergittern können aber die Abstände der Potentialtöpfe so klein sein ($< 10$ nm), dass sich die Wellenfunktionen teilweise überlappen und man erhält wieder Bandaufspaltungen; je kleiner die Abstände, um so mehr 'Subband–Aufspaltung', siehe Photolumineszensspektroskopie.

## 3.2.5. Metall–Halbleiter–Kontakte (Ohmsche Kontakte, Schottky–Dioden)

Elektronische Bauelemente benötigen elektrische Zuleitungen mit geringen Verlusten. Also nimmt man ein Metall (meist Al bzw. genauer eine AlSiCu–Legierung, neuerdings Cu) und dampft oder sputtert es auf den *Halbleiter*. Unter bestimmten Bedingungen erhält man so einen ohmschen Kontakt, oder aber das Gegenteil, nämlich eine Diode. Die Physik dazu hat vieles mit der bereits bei der p–n–Diode

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

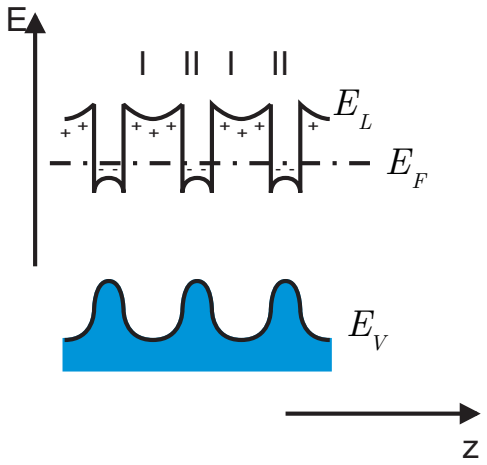



Abbildung 3.30.: *Bänderschema* eines modulationsdotierten Kompositions–Übergitters (gezeichnet nach [**?**]).

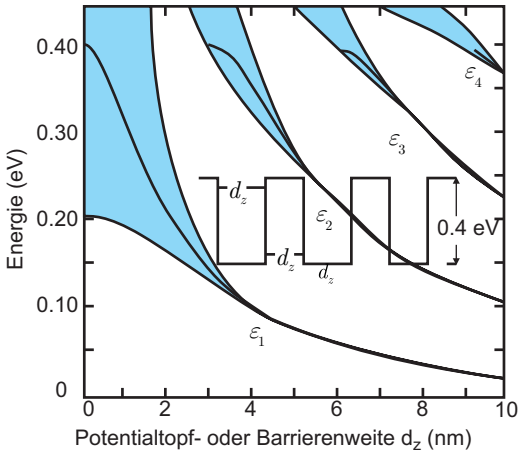



Abbildung 3.31.: Energiezustände von Elektronen, die in rechteckigen Potentialtöpfen des Leitungsbandes eines Kompositionsgitters ‘eingesperrt’ sind (gezeichnet nach [**?**]).

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

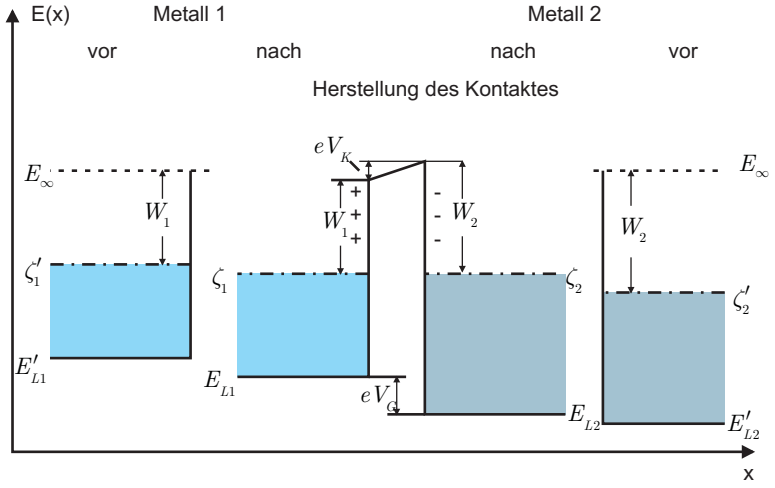



Abbildung 3.32.: *Bänderschema* eines Metall–Metall–Kontakts.

eingeführten gemeinsam.
Zu Beginn sei an die einfachen Verhältnisse des Metall–Metall–Kontakts erinnert (siehe Abbildung 3.32).
Das Bezugsniveau vor dem Kontakt sei wieder $E_{\mathrm{vac}}$ (oder $E_\infty$); die Fermi–Energie $E_{\mathrm{F}_i}$ (oder $\zeta_i'$), die Austrittsarbeiten $\Phi_i$ (oder $W_i$) und die Elektronenaffinitäten $\chi_i$ der beiden Metalle seien ungleich. Im Kontakt müssen die beiden Fermi–Niveaus gleich sein: es diffundieren vom Metall mit der kleineren *Austrittsarbeit* mehr Elektronen zu dem mit der grösseren *Austrittsarbeit* als umgekehrt (Thermoemission), da die Ströme $\sim \exp\left(-\Phi/k_BT\right)$ sind. In der Grenzschicht bildet sich eine elektrische Doppelschicht aus, Grenzflächenzustände spielen keine Rolle. Das Potential ändert sich in der Grenzschicht, (aufgrund der vergleichsweise sehr hohen Ladungsdichte der Metalle) nicht aber innerhalb der Metalle. Es bildet sich also keine *Raumladungszone* aus, Rekombination kann nicht auftreten. Man beobachtet das sog. *Kontaktpotential* $V_{\mathrm{K}} = \frac{1}{e}\left|\Phi_1 - \Phi_2\right|$, eine direkt messbare Potentialdifferenz (Messung mittels Kelvin–Methode); der Sprung der Leitungsbandunterkanten $eV_{\mathrm{G}}$ ist nicht messbar (sog. Volta–Spannung). Die Ladungsträgerdichten bleiben trotz des Ladungstransfers praktisch konstant. Es gilt:

$$eV_{\mathrm{K}} = E_{\mathrm{F}_1} + E_{\mathrm{F}_2}\ . \tag{3.2.12}$$

Komplizierter sind die Verhältnisse am *Metall–Halbleiter–Kontakt*. Betrachten wir zuerst den Kontakt zwischen *Metall* und *n–Halbleiter*.
Beschränkt man sich auf den Fall $W_{\mathrm{M}} > W_{\mathrm{n}}$ mit $W_{\mathrm{M}} - W_{\mathrm{n}} \ll E_{\mathrm{gap}}(\Delta E_i)$, braucht man nur das Leitungsband im *n–Halbleiter* mitzunehmen (sog. Einband–Modell). Wieder sorgt die Thermoemission für die negative Oberflächen–Ladung im Metall. Die positive Gegenladung im *Halbleiter* ist aber aufgrund der endlichen Anzahldichte ionisierbarer *Donatoren* ähnlich ausgedehnt wie bei der n–Zone des p–n–Übergangs: *positive Raumladung* durch ortsfeste (ionisierte) *Donatoren*.
Die Raumladungsdichte kann ganz analog zur theoretischen Beschreibung des p–n–Übergangs mit Hilfe der Poissongleichung benützt werden, um die sich einstellende Bandkrümmung auszurechnen. Die potentielle Zustandsenergie ist gegenüber $E_{\mathrm{F}}$ erhöht, das Band krümmt sich in der Ladungszone nach oben, die Randschicht verarmt an Majoritätsladungsträgern.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

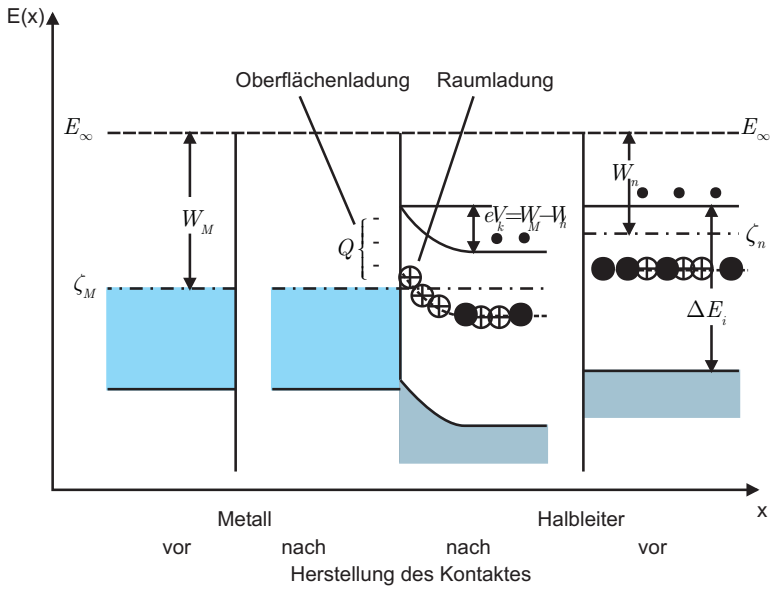



Abbildung 3.33.: *Bänderschema* eines Kontakts zwischen Metall und n–Halbleiter.

Beim Kontakt zwischen *Metall* und *p–Halbleiter* mit $W_\mathrm{M} < W_\mathrm{p}$ und $W_\mathrm{p} - W_\mathrm{M} \ll E_\mathrm{gap}$ diffundieren Elektronen in den *Halbleiter*, rekombinieren mit Löchern und übrig bleibt die negative Ladung der ortsfesten, ionisierten *Akzeptoren*. Er zeigt eine Verarmungsschicht für Löcher, den Majoritätsladungsträgern im p–Halbleiter. Im Gegensatz zum p–n–Kontakt spielen hier die Minoritätsladungsträger beim Ladungstransport keine Rolle.

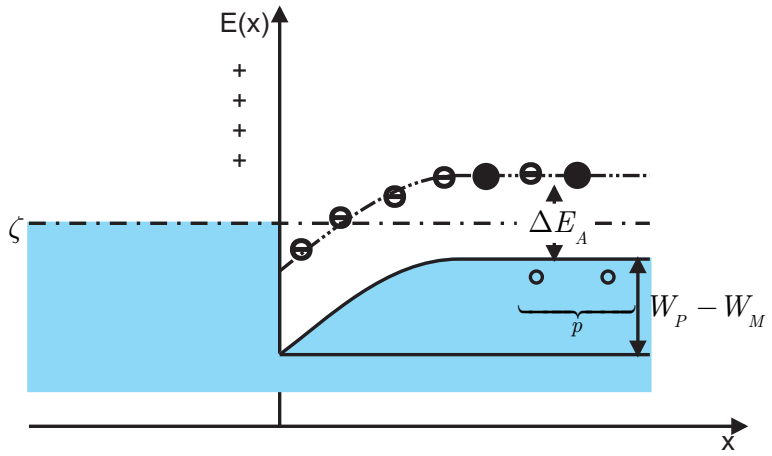



Abbildung 3.34.: Verarmungsrandschicht für Löcher.

Die Verarmungszonen–Breite ist wieder abhängig von einer aussen anlegbaren Spannung. Wir haben keinen ohmschen Kontakt, sondern eine (dem p–n–Übergang sehr ähnliche) Diode. Die thermische Beschreibung erfolgt wieder mit einem kastenförmigen Raumladungsprofil, im *Schottky–Modell*. Dies liefert einen parabelformigen Potentialverlauf, eine Raumladungstiefe $d \sim N_\mathrm{D}^{-1/2}$ (Grössenordnung 10 nm und mehr) und eine spannungsabhängige Kapazität $C \sim \sqrt{\epsilon \cdot N_\mathrm{D}}$ , die in der *Halbleiter–Messtechnik* zur Bestimmung der Donatorkonzentration herangezogen wird.
Einen weiteren interessanten Fall erhalten wir, wenn $|W_\mathrm{M} - W_\mathrm{H}| \approx \Delta E_i = E_{gap}$ ist,

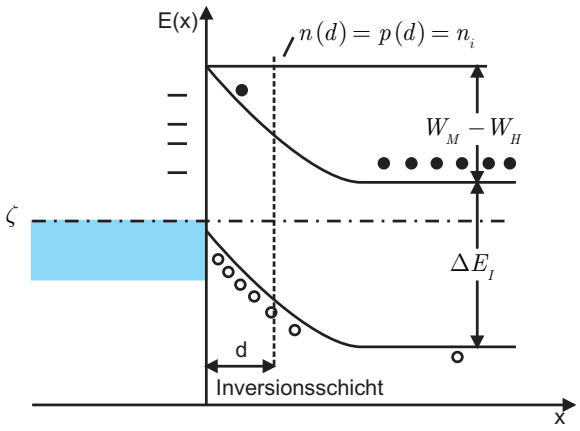



Abbildung 3.35.: Löcherinjektion im Zwei–Band–Modell.

wir also beide Bänder des Halbleiters und beide Ladungsträgerarten betrachten müssen (Zwei–Band–Modell). Die Elektronen, die ins Metall übergehen, können jetzt aus beiden Bändern stammen.

In diesem Fall erhalten wir im Leitungsband des Halbleiters eine Verarmung an Elektronen, aber im Valenzband eine Anreicherung an Defektelektronen (Löchern). Deshalb heisst dieser Randschichtbereich *Inversionsschicht*. Auf einem *n–Halbleiter* eignet sie sich zur *Injektion von Löchern*; auf einem *p–Halbleiter* (Anreicherung von Elektronen) zur *Injektion von Elektronen* (Fall hier nicht bildlich gezeigt). *Injizierende Kontakte* spielen für viele Bauelemente eine bedeutende Rolle.

Ein weiteres Beispiel für ein Zwei–Bänder–Modell ist der Kontakt von Metall und *undotiertem Halbleiter*. Aufgrund der geringen *Ladungsträgerdichte* wird die *Raumladungszone* extrem breit, z. B. bei Ge $> \mu$m.

Soweit stimmen die gemachten Ausführungen mit einem Grossteil der Buchliteratur überein. Allerdings zeigen sich die Gleichrichterwirkung oder die Kapazität der Metall–Halbleiter–Kontakte im Experiment weitgehend unabhängig von einem gewählten Metall und seiner speziellen *Austrittsarbeit*. Ähnlich wie die in Kapitel 3.1.6 erwähnten Oberflächenzustände sind beim Metall–Halbleiter–Kontakt *Grenzflächenzustände* entscheidend für das Kontaktverhalten. Ist die von diesen 'Zwischenzuständen' verursachte Bandverbiegung grösser als die Differenz der Austrittsarbeiten, so ist der Einfluss des Metalls praktisch vernachlässigbar.

Diese Oberflächenzustände wirken entweder als *Donatoren* oder *Akzeptoren* und sind entsprechend der Temperatur besetzt oder nicht, d. h. ihre Ladung wird durch die Lage des Fermi–Niveaus bestimmt; ihre Anzahldichte ist sehr gross (ca. $10^{14}$ cm$^{-2}$). Die Ladung der Grenzflächenzustände wird zur Herstellung der Ladungsneutralität durch die Ladung der ortsfesten ionisierten *Donatoren* kompensiert. Wir haben eine Verarmungszone, die zwar indirekt durch das die Grenzflächenzustände mitbestimmende Metall verursacht wird, die aber praktisch wenig mit der *Austrittsarbeit* des Metalls zu tun hat. Auch die Variation der Dotierung im *Halbleiter* ändert fast nichts an der *Raumladung*, ebensowenig das Anlegen einer äusseren Spannung. Die sog. *Schottky–Barriere* $e\Phi_{SB}$ ist deshalb eine für den speziellen Metall–Halbleiterübergang charakteristische Grösse, sie hängt allerdings etwas von präparativen Details ab.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

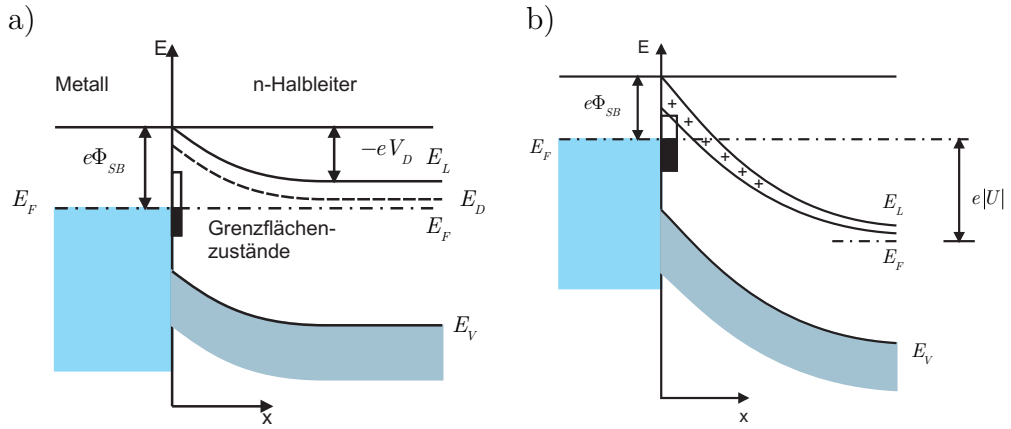



Abbildung 3.36.: Elektronisches *Bänderschema* eines Metall – n–Halbleiter–Übergangs mit Schottky–Barriere a) im thermischen Gleichgewicht, b) mit angelegter äusserer Spannung (gezeichnet nach [**?**]).

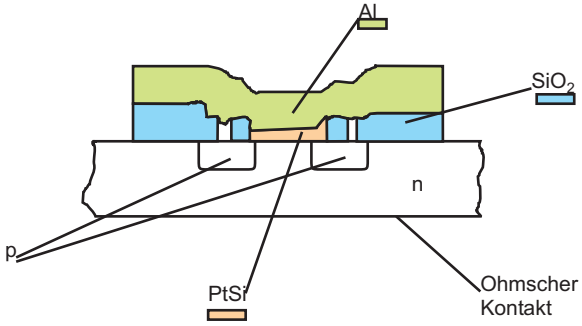



Abbildung 3.37.: Aufbau einer PtSi–n–Si–Diode(gezeichnet nach [**?**]).

Der *reale Kontakt* mit Verarmungsrandschicht zeigt eine Kennlinie der Form

$$j(U) = j_S \cdot \left[ e^{\frac{\pm eU}{\beta k_B T}} - 1 \right] , \tag{3.2.13}$$

wobei $\beta > 1$ ist. In Durchlassrichtung ist die Verarmungsrandschicht verengt und die Potentialbarriere verringert; der Thermoemissionsstrom beträgt

$$j_{H\to M} = C(T) \cdot e^{-\frac{E_L + |eV_D| - E_F}{k_B T}} \cdot e^{\pm \frac{eU}{k_B T}} . \tag{3.2.14}$$

Der Sättigungsstrom ist fast unabhängig von der äusseren Spannung und lautet

$$j_{H\to M} = j_s = C(T) \cdot e^{-\frac{\Phi_{SB}}{k_B T}} . \tag{3.2.15}$$

Zur hochohmigen Randschicht kommen noch der niederohmige Bahnbeitrag des 'Halbleiter–bulks' hinzu.
In Abbildung 3.37 ist eine PtSi–n–Si–Diode im Aufbau gezeigt.
*Ohmsche Kontakte* können auf mehrere Arten realisiert werden. Genannt wurde schon der Fall der Anreicherungsschicht (Inversionsschicht) beim injizierenden Kontakt; die Schicht ist niederohmiger als das *Halbleiter–Volumen*, die Bahnwiderstände dominieren; wir haben einen sog. sperrfreien Kontakt. Zwei weitere Möglichkeiten sind dadurch gegeben, dass man die Barriere durch Wahl eines besonders kleinen $e\Phi_{SB}$ sehr niedrig macht oder dadurch, dass man durch eine hochdotierte

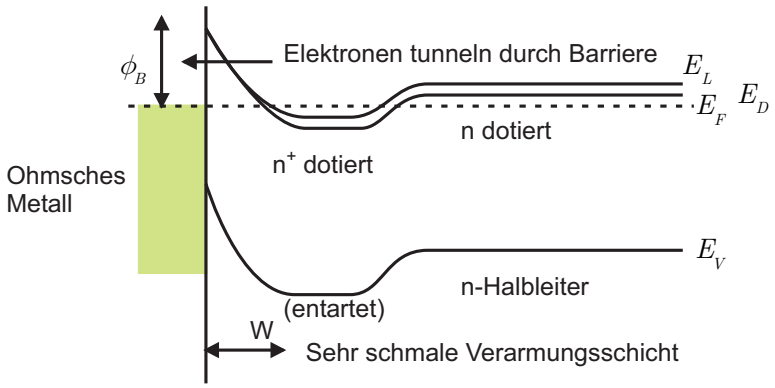



Abbildung 3.38.: Ohmscher Metall – n–Halbleiter–Kontakt (gezeichnet nach [**?**]).

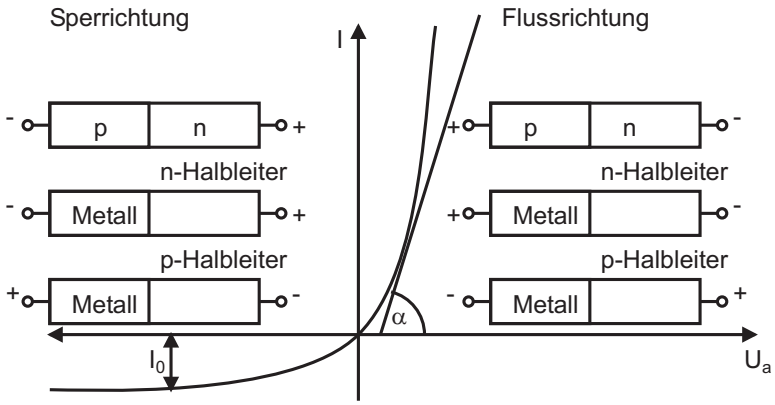



Abbildung 3.39.: Kennlinie eines gleichrichtenden Metall–Halbleiter–Kontaktes.

Schicht die Verarmungsschicht sehr dünn macht, so dass Elektronentunneln massiv möglich wird.
*Reale ohmsche Kontakte* stellt man auf *p–Si* her, so dass man Al direkt aufbringt und durch Ein–Legieren eine gute Haftung und eine $p^+$–dotierte Zwischenschicht, also sehr gute Leitfähigkeit erhält. Dies funktioniert aber deshalb so gut, weil Al ein gutes Akzeptormaterial ist. Auf *n–Si* würde man so nur eine Umdotierung erreichen. Man scheidet deshalb ein ganzes Schichtpaket ab: Al–TiN–Ti–$TiSi_2$–n–Si, wobei Ti als Diffusionsbarriere für das Al eingesetzt wird; der eigentliche Kontakt bildet wieder die Silizid–n–Si–Anordnung. Die Silizide (Ti, W, Mo, Pt, Ni) werden heute fast immer eingesetzt, da sie niederohmigere Kontakte als die elementaren Metalle erlauben.
Zum Abschluss des Kapitels werden die Eigenschaften von p–n– und Schottky–Dioden bildlich zusammengestellt.

## 3.2.6. Metall–Isolator–Halbleiterkontakte (MIS– und MOS–Dioden)

MIS (metal-isolator-semiconductor) –Kontakte sind ein extrem leistungsfähiger Grundbaustein in der Festkörperelektronik. Zum einen lassen sich damit Halbleiteroberflächen physikalisch untersuchen, zum anderen ist diese Anordnung in einer Vielzahl von Bauelementen enthalten, nämlich als Diode, als spannungsvariable Kapazität, als 'charge–coupled device' (CDD), als *Transistor* und schliesslich als

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

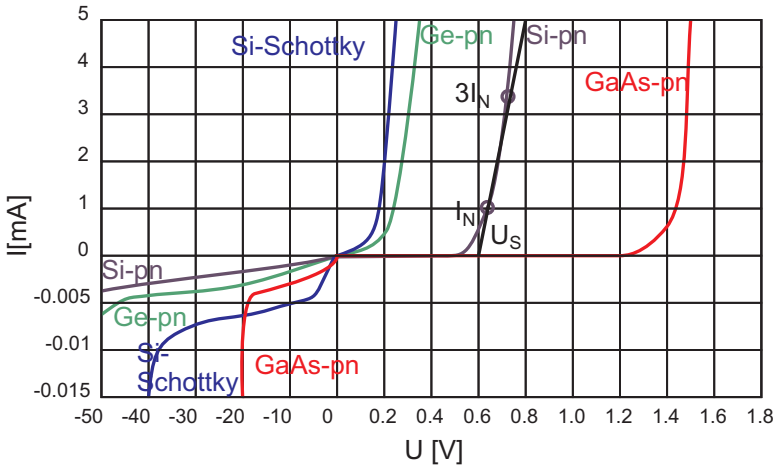



Abbildung 3.40.: Zum Vergleich: Beispiele gemessener Diodenkennlinien (gezeichnet nach [**?**]).

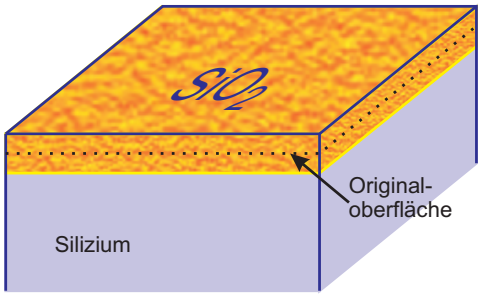



Abbildung 3.41.: Aufwachsen von $SiO_2$ (gezeichnet nach [**?**]).

das meistverbreiteste Element der Höchstintegration. Diese grosse technische Bedeutung erlangte der MIS–Kontakt in einer einzigen speziellen Zusammensetzung: MOS auf *Si*, das Oxid des Siliziums hat nämlich aufgrund seiner Herstellung durch *thermische Oxidation* ganz hervorragende Eigenschaften.

Man bringt eine sehr gut gereinigte *Si–Oberfläche* in einen Rohrofen und leitet $O_2$–Gas ein. Der Sauerstoff reagiert an der Waferoberfläche mit dem Silizium und es bildet sich *amorphes* $SiO_2$. Weiter angebotener Sauerstoff diffundiert durch die Oxidschicht und reagiert an der Grenzschicht. Durch die Massenzunahme wird die Waferoberfläche angehoben, knapp 45 % der Oxiddicke liegen unter der ursprünglichen Oberfläche.

Bei der sog. *trockenen Oxidation* leitet man nur reinen Sauerstoff bei 1000° C - 1200° C ein; man erzeugt das ‘Gate–Oxid’ für MOS–Transistoren ($<$ 100 nm), die noch dünneren Tunneloxide stellt man bei ca. 800° C her. Die Gateoxid–Präparation ist der kritischste Schritt bei der Herstellung z. B. eines dynamischen RAM–Speichers. Bei der *nassen Oxidation* erhält man höhere Aufwachsraten, aber geringere Durchbruchfestigkeiten; so gewachsene Oxide werden als Feldoxide bezeichnet und häufig bei der Fertigung als Maskenmaterial eingesetzt. Ein spezielles Nassverfahren ist die sog. $H_2O_2$–Verbrennung; sie liefert sehr reine, allerdings weniger dichte Oxide, die bevorzugt als Kondensatordielektrikum in DRAM–Speichern eingesetzt werden. Die thermische Oxidation ist ein sehr langsamer Prozess mit Aufheiz– und Abkühlraten von ca. 1° C pro Minute, dabei können steile bzw.

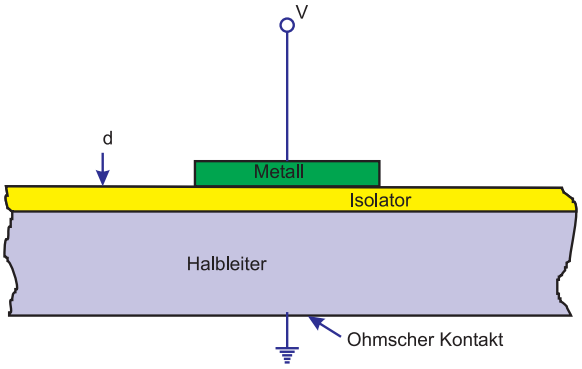



Abbildung 3.42.: Schematischer Aufbau einer MIS–Diode(gezeichnet nach [**?**]).

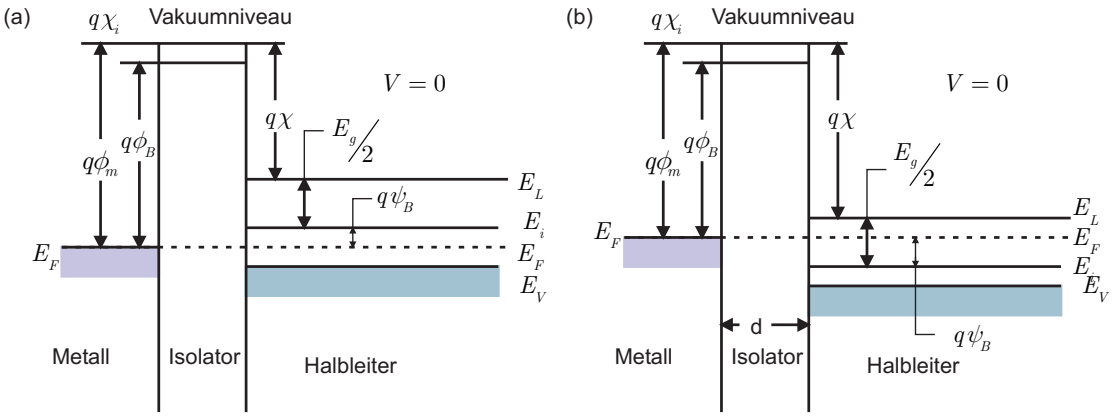



Abbildung 3.43.: *Bänderschema* einer idealen MIS–Diode: (a) mit *p–Halbleiter*; (b) mit *n–Halbleiter* (gezeichnet nach [**?**]).

schmale Dotierprofile nicht stabil bleiben. Deshalb setzt man neuerdings RTP (rapid thermal processing) – Öfen für Einzelwafer mit Raten von $10 - 100°$ C pro Minute ein. So lassen sich sehr dünne, dichte und glatte Oxide wachsen.

Den schematischen Aufbau einer MIS–Diode zeigt Abbildung 3.42. Eine äussere Spannung $V$ sei positiv, wenn das Metall positiver als der ohmsche Kontakt des Halbleiters ist.

Das Bändermodell einer *idealen MIS–Diode* bei $V = 0$ zeigt im thermodynamischen Gleichgewicht eine durchgängig waagrechte *Fermi-Energie* ('flat-band condition'); die Differenz der Austrittsarbeiten von Metall und *Halbleiter* ist Null. Durch die Isolatorbarriere gibt es keinen Ladungstransport beim Anlegen einer Gleichspannung $V$. Die einzigen Ladungen, die unter $V$ auftreten, sind dann im Metall an der Oberfläche und gleich gross mit entgegengesetztem Vorzeichen im *Halbleiter*. Im *Halbleiter* treten Phänomene auf, wie wir sie schon bei der idealen Schottky–Diode kennengelernt haben. Für negative Spannungen, also $V < 0$, erhalten wir im Falle des p–Halbleiters eine Anreicherungsschicht für die Majoritätsladungsträger, für positive Spannungen, also $V > 0$, erhält man eine Verarmungsschicht und für noch grössere positive $V$ schliesslich einer Inversionsschicht: die Minoritätsladungsträgerdichte in der *Raumladungszone* ist grösser als die Majoritätsladungsträgerdichte. Die rechte Spalte in Abbildung 3.44 zeigt für den *n–Halbleiter* dieselben Effekte bei umgekehrter Spannungspolung. Die räumliche Ausdehnung der Bandverbiegung ist von derselben Grössenordnung wie bei den Beispielen im

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

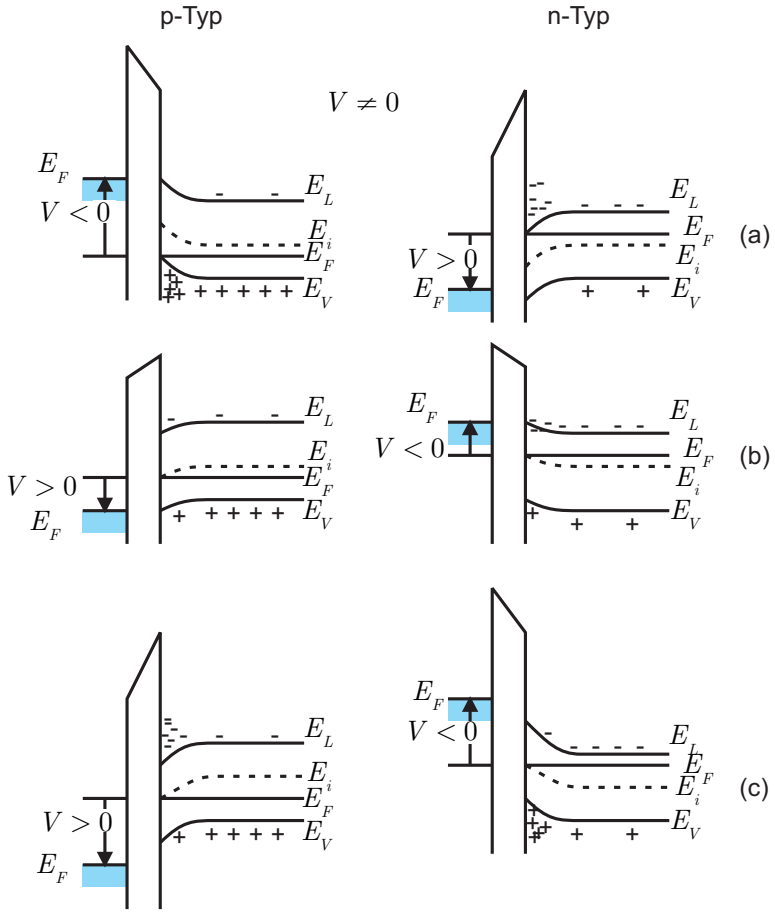



Abbildung 3.44.: *Bänderschema* einer idealen MIS–Dioden mit äusserer Spannung $V$ (gezeichnet nach [**?**]).

vorigen Kapitel (typisch 100 nm).

Unter der Annahme, dass die Ladungsträgerdichten im Volumen und an der Oberfläche jeweils exponentiell vom Potential ($\sim \exp(q\psi/k_BT)$) abhängen, lässt sich mit Hilfe der Poissongleichung die Raumladungsdichte für eine bestimmte Dotierkonzentration als Funktion des Oberflächenpotentials (bezogen auf $E_{\text{intrinsisch}}$) berechnen.

Die angelegte Spannung fällt linear über der Isolatorschicht und exponentiell in der *Raumladungszone* ab: $V = V_{\text{Isolator}} + \psi_S$, $\psi_S$=Oberflächenpotential. Analog ist die Gesamtkapazität der MIS–Diode eine Reihenschaltung der Isolator–Kapazität und der Verarmungsschicht-Kapazität im *Halbleiter*: $C = \frac{C_{\text{Isolator}}\,C_D}{C_{\text{Isolator}}+C_D}$. $C_{\text{Isolator}}$ ist konstant und entspricht dem Maximalwert des Systems, $C_D$ ist spannungs– und frequenzabhängig.

Die realen Verhältnisse in der Si–$SiO_2$–MOS–Diode sind deutlich komplizierter als die ideale MIS–Diode es beschreibt. Abbildung 3.45 zeigt zum einen bewegliche Ladungen ($Na^+$, $K^+$) und getrappte Ladungen im $SiO_2$ und eine wenige nm tiefe, unvollständig oxidierte Zwischenschicht. Dazu ist die Tatsache zu bedenken, dass im amorphen Material immer unabgesättigte Bindungen vorhanden sind und dass es an den Kontaktflächen Metall–Isolator und Isolator–Halbleiter immer ionisierbare Grenzflächenzustände geben muss (‘interface trapped charge’), wobei letztere die *Raumladung* deutlich mitbestimmen (Erinnerung: Schottky–Diode).

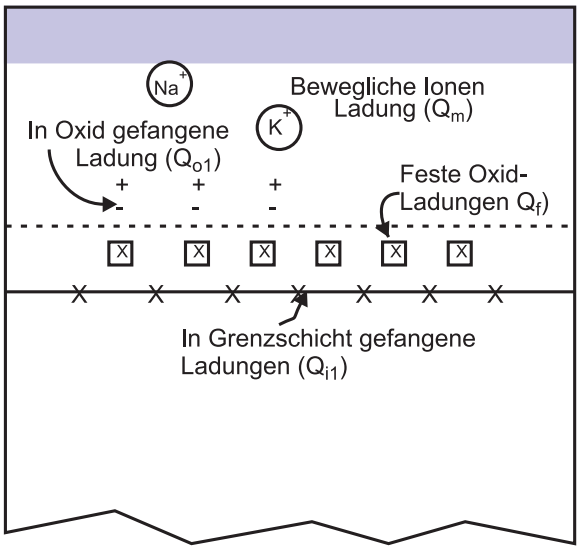



Abbildung 3.45.: Ladungsverhältnisse in einer realen MOS–Diode (gezeichnet nach [**?**]).

*Nebenbemerkung*:
Durch einen Tempergang in $H_2$–Atmosphäre bei ca. 450° C kann dieser Einfluss stark reduziert werden. Durch Aufwachsen des Oxids unter Zugabe von $Cl^-$–Ionen (HCl–Dampf oder Trichlorethangas) können die beweglichen Ionen ortsfest gebunden werden.
Die verbleibenden ortsfesten und getrappten Oxidladungen verändern das Bild von der *Raumladung der idealen MIS–Diode*, es shiften beispielsweise die flat–band–Bedingung, die $C$–$V$–Kurven und natürlich das Oberflächenpotential $\psi_S$. Die Differenz der Austrittsarbeiten $\Phi_{MS}$ im thermodynamischen Gleichgewicht ist nicht mehr Null (Beispiel Al–$SiO_2$–Si–Diode). In der Praxis gibt es aber für jede Dotierkonzentration im *Si* Materialien, die es im Falle $V = 0$ erlauben, eine Anreicherungs– oder Verarmungs–/Inversionsschicht gezielt zu realisieren.
Für den Fall des p–Typ–Substrats seien im obigen Fall die wichtigsten Fälle vereinfacht im Bändermodell und in den Raumladungen zusammengefasst (siehe Abbildung 3.46). Für negative Spannungen erhält man den Fall der Anreicherung mit der *Raumladungszone* $Q_h$ der Löcher. Aufgrund der positiven Oberflächenladungen $Q_{Ox}$ benötigt die flat–band–Bedingung ebenfalls eine negative Vorspannung; ohne Vorspannung erhält man aus demselben Grund im thermodynamischen Gleichgewicht bereits eine Verarmung, d. h. die ausgedehnten Raumladungen der ionisierten *Akzeptoren* und schliesslich unten, wenn das Leitungsband so stark nach unten gekrümmt ist, dass es unter $E_F$ sinkt, kommt noch eine dünne Lage aus Elektronen (Minoritätsladungsträger, $Q_E$) im Falle der Invesion hinzu.
Äussere Einflüsse wie Temperatur, Licht, ionisierende Strahlung oder die Injektion heisser Ladungsträger beeinflussen das MOS—Dioden–Verhalten erheblich. Dem interessanten Fall, nämlich dem der Beleuchtung werden wir ein eigenes Kapitel widmen, nämlich der CCD im Kapitel 3.3.1.
Zum Abschluss dieses Kapitels sei noch auf das Problem der Segregation bei der thermischen Oxidation von *dotiertem Si* hingewiesen. Beim Tempern diffundieren die Störatome und wenn ihre Löslichkeit im $SiO_2$ grösser ist als im *Si* selber, kommt es zu einer Absenkung der Dotierkonzentration im *Si* nahe dem Interface

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

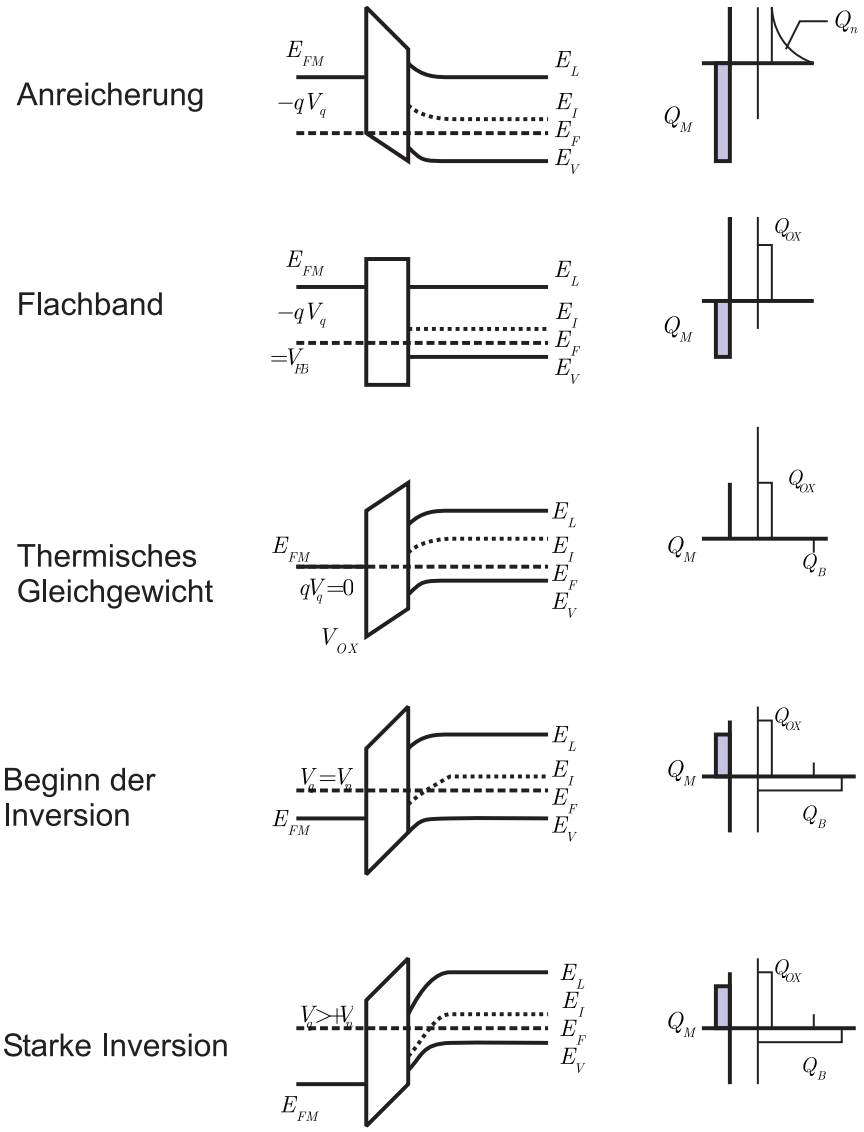



Abbildung 3.46.: *Bänderschema* und Raumladungen eines realen MOS–Kontaktes für verschiedene äussere Spannungen (p–Typ–Substrat), (vereinfacht gezeichnet nach [**?**]).

('pike–down'). Dies ist für den gebräuchlichsten Akzeptor Bor gerade der Fall. Der umgekehrte Fall ('pike–up') spielt in der Praxis keine grosse Rolle.

## 3.3. Wichtige Halbleiter–Bauelemente (Aufbau, Funktion, Technologie)

Mit den besprochenen Kontaktphänomenen haben wir die Grundbausteine kennengelernt, aus denen alle Halbleiter–Bauelemente bestehen. Zuerst waren dies die bipolaren p–n–Dioden, danach die unipolaren Metall–Halbleiter– und Metall–Oxid–Halbleiter–Dioden. Letztere finden wir in den 'ladungsgekoppelten' CCDs, in den MOS–Transistoren und in den hoch– und höchstintegrierten Bausteinen (IC 'integrated circuit') für die Datenverarbeitung (Logik, Mikroprozessor, etc.) und die Speicherung (DRAM, etc.) wieder; allesamt auf Siliziumsubstraten. Bei den bipolaren Bauelementen sind herauszustellen die Transistoren und die grosse Vielfalt der optoelektrischen Bauelemente auf Basis der *direkten Halbleiter*, insbesonders auf der Grundlage von *GaAs*. Im folgenden also eine Auswahl aus dem oben genannten Themenbereich.

### 3.3.1. Ladungsgekoppelte Bauelemente (CCD charge coupled devices)

Das CCD in seiner einfachsten Form ist eine lineare Anordnung ('linear array') von dichtbenachbarten MOS–Dioden, deren Vorspannung so festgesetzt ist, dass an der Oberfläche eine ausgeprägte Verarmung an Majoritätsladungsträgern besteht ('biased in deep surface depletion').
An dem bereits vom vorigen Kapitel gewohnten Beispiel des positiv vorgespannten Metall–p–Typ–Halbleiters sieht man im Bereich der Verarmungszone, dass sich im Leitungsband ein leerer Oberflächen–Leitungskanal ('surface channel') zwangsläufig ausgebildet hat. Bringt man in diesen zweidimensionalen Potentialtopf Elektronen, so sind diese zunächst in der Oberflächenschicht frei beweglich. Im realen MOS–Diodenbauelement ist die Metallelektrode natürlich nur endlich ausgedehnt; d. h. wir haben auch lateral in der Oberflächenschicht einen endlich ausgedehnten Potentialsee.
Legt man an die MOS–Kondensatoren–Kette mit Hilfe eines Taktgenerators ('clock') eine bestimmte Impulsfolge an, so kann elektrische Ladung in Form von Paketen kontrolliert an der Oberfläche eines *Si–Substrats* transportiert werden. Wir haben also die Grundstruktur eines *dynamischen Analog–Schieberegisters* vorliegen, wobei die Information in der Ladung $Q_{\text{Signal}}$ steckt (und nicht, wie üblich, im Strom oder der Spannung). Neben seiner Anwendung als Schieberegister kann das CCD als Datenspeicher (mit Regenerationsstufen), als Verzögerungsglied, als Filter oder für logische Operationen eingesetzt werden.
Die verbreiteste Anwendung der CCD ist heute als *Bildsensor* (*CCD–Kamera*). In einem typischen Abstand von z. B. 7, 13 oder 27 $\mu$m sitzen MOS–Dioden als optische Detektoren mit Grössen von z. B. $1,7$ $\mu$m. In zweidimensionaler Anordnung erhält man so einen Bildsensor.
Beim optischen Detektor ist das Metall bzw. das Polysilizium meist transparent. Photonen, deren Energie grösser als die *Energielücke des Si* sind, generieren im

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

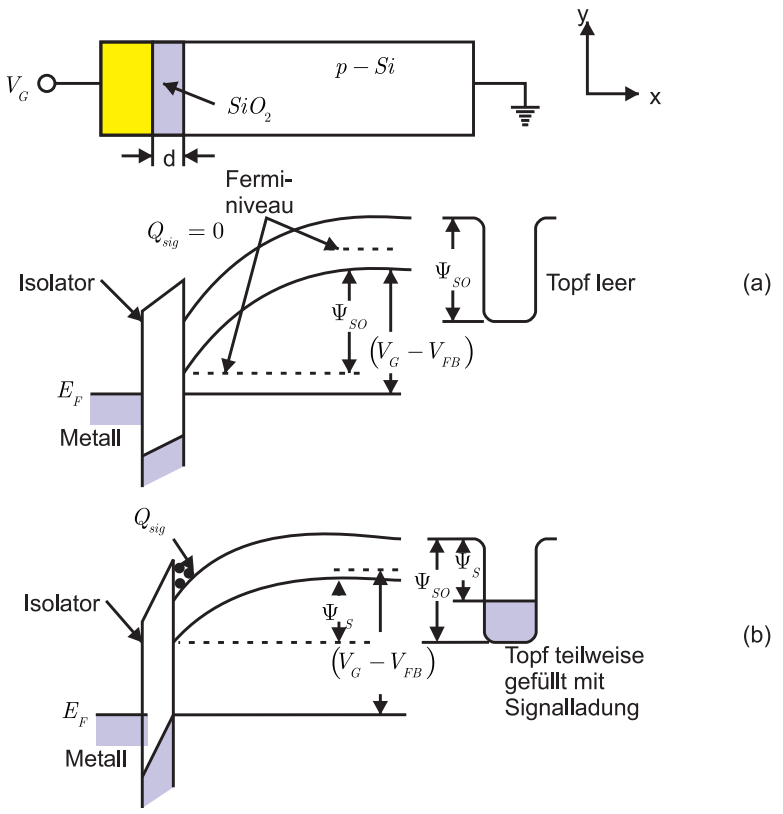



Abbildung 3.47.: *Bänderschema* einer MOS–Diode mit leerem Oberflächen–Leitungskanal (gezeichnet nach [**?**]).

*Si* ein Elektron–Loch–Paar. Das Loch diffundiert im Valenzband in die Tiefe des *Si–Substrats*, die Elektronen sammeln sich im n–Kanal an der Oberfläche. Das lichtempfindliche Bauelement kann im Prinzip aber auch eine Schottky–Diode, eine p–i–n–Fotodiode oder ein Fototransistor sein. Das CCD-Chip enthält aber in jedem Fall ein oder zwei MOS–Schieberegister zum Auslesen des bzw. der optischen Detektoren.

Den Aufbau der Grundeinheit eines Ladungsverschiebelements gibt Abbildung 3.49 wieder. Drei benachbarte MOS–Dioden bilden, abhängig von der Grösse der positiven, angelegten Spannung Potentialtöpfe aus; die Ladung bleibt annähernd im tiefsten Topf erhalten oder fliesst zum tiefsten Topf hin. Das gezeigte Bild gibt eine Drei–Phasen–Verschiebeeinheit wieder.

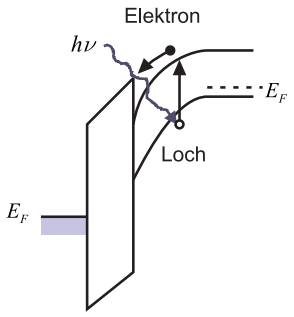



Abbildung 3.48.: *Bänderschema* einer Si–$SiO_2$–MOS–Diode unter Beleuchtung (gezeichnet nach [**?**]).

Setzt man mehrere solche Baugruppen hintereinander (siehe Bild 3.50) und versieht Anfang und Ende der Kette mit einer Eingangs– und Ausgangsschaltung (n–p–Diode und MOS–Gate) ist das Schieberegister komplett.
Legt man in geeigneter Weise Spannungen an den Eingang und die MOS–Gates und schliesslich an den Ausgang, funktioniert der Ladungstransport (Abbildung 3.51).
Es gibt verschiedene Elektronenanordnungen und zugehörige Takttechniken ('push–clock–Systeme'): 1–4 Phasensysteme mit Taktfrequenzen von einigen kHz - 100 MHz. 2–Phasensysteme benötigen 'unsymmetrische' Gate–Elektroden, 1–Phasensysteme zusätzliche Dotierung.
Die bisher präsentierte CCD–Bauform bezeichnet man als *SCCD* (*surface channel CCD*). Ihr Hauptnachteil liegt im Vorhandensein von Trapping–Zentren am Kanalrand. Dies führt zu zusätzlichem Rauschen.
Eine verbesserte Ladungstransfer–Effektivität erhält man mit einem weiteren Grundtyp: *BCCD* (*burried channel CCD*). Die 'vergrabene' n–Kanalschicht im p–Substrat erhält man durch eine *ionenimplantierte n–Schicht* auf dem p–Substrat. Streng genommen liegt ein völlig neues Bauelement vor, nämlich ein MO–Kontakt über einem n–p–Kontakt; es handelt sich aber wieder nur um ein unipolares Bauelement. Im zugehörigen Bändermodell zeigt sich, dass der entstehende n–Kanal nicht mehr direkt an der kritischen Oxid–Halbleiter–Grenzfläche liegt. Das Rauschverhalten dieser Bauelemente ist deutlich besser (Anwendung: hochempfindliche Bildsensoren).
Ein– oder zweidimensionale *Bildaufnahmeeinheiten* bestehen aus Zeilen oder Arrays von Detektoren, die entweder integrierende Speicher besitzen oder ihre Ladungen ständig an eine Speicherzeile weitergeben müssen. Häufig sind diese Speicher durch eine lichtundurchlässige Schicht geschützt ('Anti–Blooming').
Beim eindimensionalen *Interline–Konzept* wird neben jede Sensorzeile eine CCD-Transportregister–Zeile gesetzt bzw. genauer zwischen zwei Sensorzeilen wird eine Registerzeile gesetzt und die Halbbilder werden abwechselnd ausgelesen in ein sog. CCD–Ausleseregister.
Beim dem in der Abbildung 3.55 dargestellten *Frame–Transfer–Konzept* wird der gesamte Inhalt eines ganzen Bildbereichs in einem lichtdichten Speicherbereich zeilenweise parallel verschoben. Das Auslesen erfolgt zeilenweise. Zur Erzeugung analoger Videosignale benötigt man einen Videoverstärker, für digitale Kameras geeignete A–D–Wandler und Speicher.

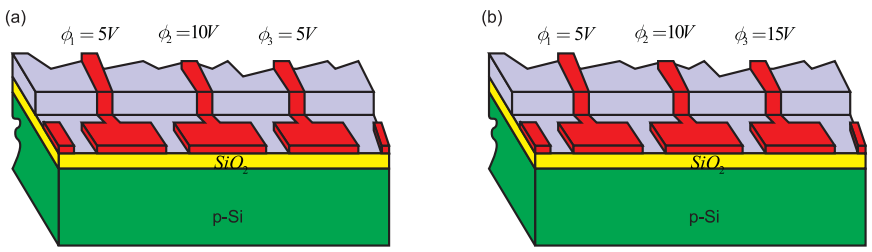



Abbildung 3.49.: Schnitt durch eine Dreiphasen–CCD–Verschiebeeinheit (gezeichnet nach [**?**]).

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

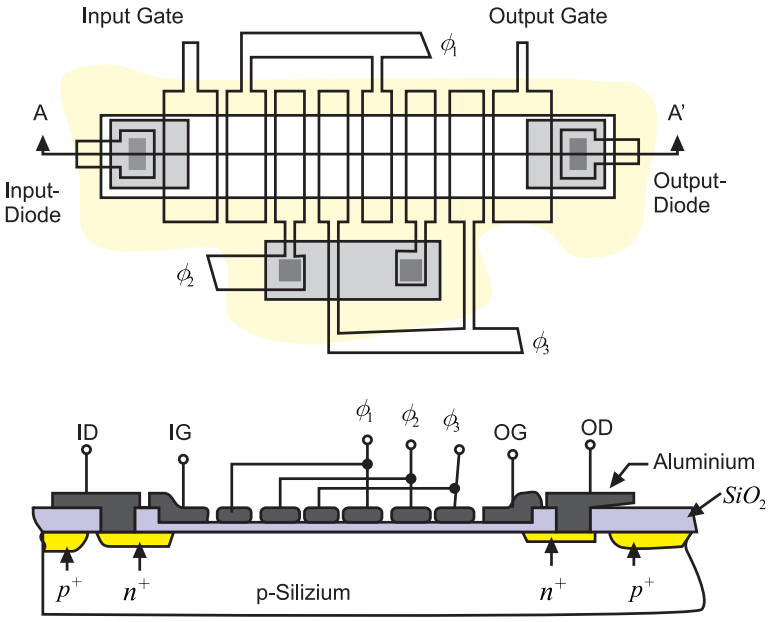



Abbildung 3.50.: Aufbau eines n–Kanal–CCD (gezeichnet nach [**?**]).

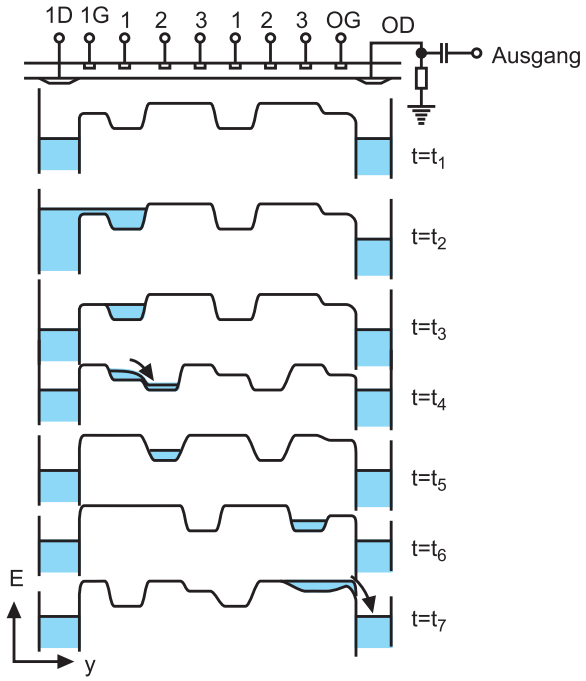



Abbildung 3.51.: Potentialverlauf und Ladungsverteilung im oben abgebildeten CCD–Chip(gezeichnet nach [**?**]).

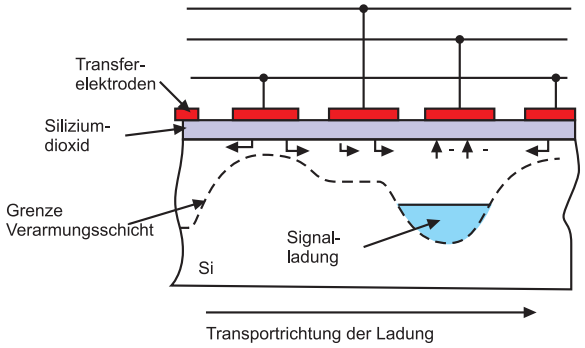



Abbildung 3.52.: Trapping (gezeichnet nach [**?**]).

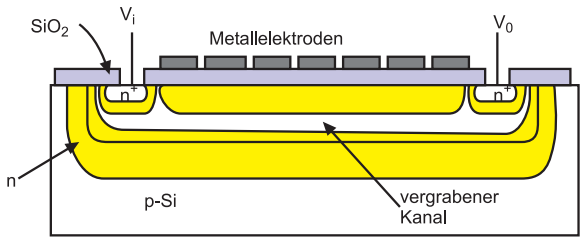



Abbildung 3.53.: Aufbau eines n–Kanal–BCCD (gezeichnet nach [**?**]).

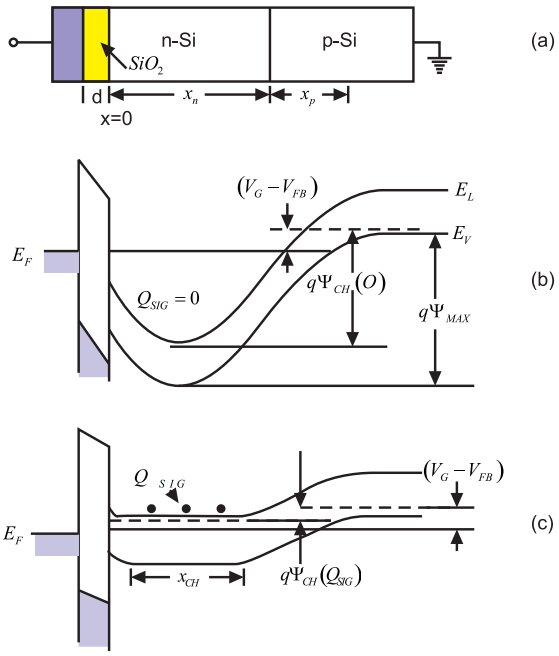



Abbildung 3.54.: Aufbau eines BCCD a). Bändermodell ohne b) bzw. mit c) Signal (gezeichnet nach [**?**])

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

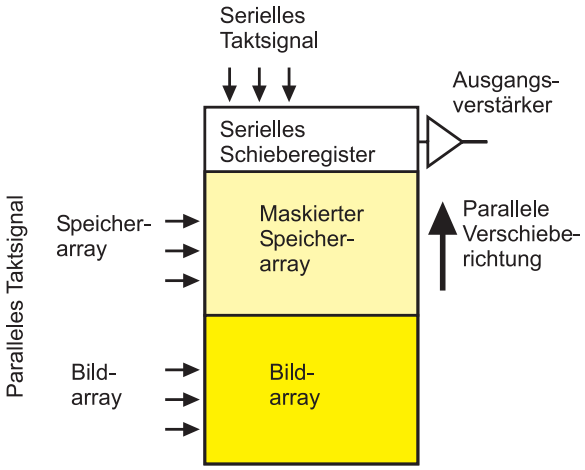



Abbildung 3.55.: Frame–Transfer–Konzept.

## 3.3.2. Feldeffekt–Transistoren (Unipolare Transistoren)

Für informationstechnische Schaltkreise (logische Schaltungen, Speicher) und Leistungsverstärker sind die sog. *Transistoren* die wichtigsten aktiven Bauelemente. Während die bislang besprochenen Dioden sog. Zweitor–Bauelemente mit zwei äusseren Kontakten sind, kennzeichnet die Transistoren das Vorhandensein von drei Kontakten: sog. *Dreitor–Bauelemente*. Ströme oder Spannungen zwischen zwei Kontakten werden durch einen dritten Kontakt gesteuert oder ein– und ausgeschaltet.
Es gibt zusätzliche Möglichkeiten zur Klassifizierung von Halbleiter–Bauelementen. Ist für die Funktion des Bauelements nur eine Ladungsträgersorte wesentlich, so spricht man von *'unipolaren' Bauelementen*; sind dagegen beide Ladungsträgertypen wesentlich beteiligt, redet man von 'bipolaren' Bauelementen. So auch bei den Transistoren. Abweichend von der Mehrzahl der Autoren stellen wir die Behandlung der Bipolartransistoren noch zurück (schwieriger zu verstehen!). Stattdessen nutzen wir die in Kapitel 3.2.5 und 3.2.6 zuletzt gelegten Grundlagen für das Verständnis der *Feldeffekt–Transitoren* (FET 'field effect transistor') aus.
Im Prinzip sind FETs Widerstände mit zwei Kontakten, die durch eine äussere Spannung an einem dritten Kontakt gesteuert werden. Die Art des dritten Kontakts kann verschieden sein: *MESFET* und *MISFET* (Metall–Semiconductor FET und Metall–Isolator–Semiconductor FET). Der Steuerkontakt ist im ersten Fall also ein Schottky–Kontakt, im zweiten Fall ein MIS– oder (in den meisten Fällen in der Praxis) ein MOS–Kontakt.
Der *gesteuerte Widerstand* besteht aus einem an beiden Enden kontaktierten Stromkanal für die eine, für die Funktion wesentliche Ladungsträgersorte. Es gibt Bauelemente mit einem n–Kanal und hierzu komplementäre mit einem p–Kanal; die letztgenannten Bauelemente sind aufgrund der geringeren Beweglichkeit der Löcher immer langsamer (aber bisweilen einfacher zu fertigen).
Der Kontak am Eingang des leitenden Kanals ist die sog. *Source* (Quelle), der am Kanalausgang der sog. *Drain* (wörtlich Abfluss). Der Steuerkontakt trägt die Bezeichnung *Gate* (Tor). In den FET–Grundformen gibt es ein oder zwei zusammengeschaltete Gates. (*Nebenbemerkung*: Beim sog. Double Gate FET werden die beiden Gates nicht miteinander verbunden, sondern voneinander unabhängig zur Steuerung des Majoritätsladungsträgerstroms insbesondere für Misch– und Regelschaltungen herangezogen. Nicht zu verwechseln mit dem Thyristor, einem bipolaren Bauelement mit zwei Steuerelektroden.) Kennzeichnend für alle FET–Typen ist die hochohmige Trennung von Steuer–Kontakt und Kanal über eine Verarmungsschicht oder einen Isolator/ein Oxid. Die Eingangswiderstände sind vergleichsweise sehr hoch; die Steuerung erfolgt über die Spannung, die Gateströme sind extrem klein. Die Steuerelektrode (control electrode) ist kapazitiv an die aktive Kanalregion gekoppelt, ein elektrisches Feld kontrolliert die Ladungen in der aktiven Zone und verändert diese, daher der Name *'Feldeffekt'–Transistor*.
Das Prinzip des FETs besteht also in der Steuerung des *Leitwerts* eines leitenden Kanals:

$$G = e \cdot n \cdot \mu \cdot \frac{A}{l} = \sigma \frac{A}{l} \ . \tag{3.3.1}$$

Man erkennt leicht, dass (bei fester Kanallänge $l$) entweder der Kanalquerschnitt $A$ oder die *Ladungsträgerdichte* $n$ zur Steuerung variiert werden muss.
Die Querschnitt–Steuerung nutzen zwei bekannte Transistorklassen. Erstens die

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

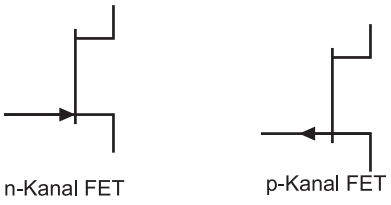



Abbildung 3.56.: Schaltsymbole für den FET.

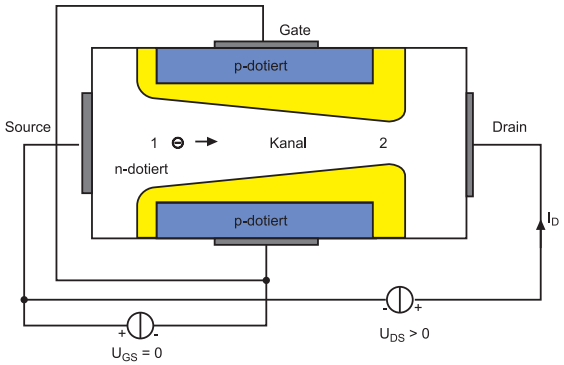



Abbildung 3.57.: Prinzip des Sperrschicht–FET.

sog. *Sperrschicht–FET* (JFET, Junction FET) und, später hinzugenommen, zweitens die MESFETs. Erstere werden meist in *Si*, letztere meist in *GaAs* realisiert. Die Steuerung der Trägerdichte ist kennzeichnend für die MISFETs und MOSFETs, sowie für die moderne Weiterentwicklung des MESFET, den HEMT (high electron mobility).
Historisch gesehen war der MISFET der erste *Transistor*, der erdacht worden ist (1939). Im Gegensatz zum Bipolartransistor kommt es bei unipolaren Transistoren nur auf die Majoritätsladungsträger an, d. h. die Anforderungen an das *Halbleiter–Material* (Reinheit, Kristallinität) sind geringer. Tatsächlich sind unipolare Transistoren in einer grossen Vielzahl von halbleitenden Materialien realisiert worden. Der grosse Nachteil des Vorschlags war aber die technisch äusserst schwierig zu beherrschende Grenzfläche zwischen Isolator und *Halbleiter*. Der Sperrschicht–FET umgeht dieses Problem: die Steuerung wird weg von der Halbleiter–Oberfläche ins Innere des Halbleiters gelegt. Seine technische Realisierung gelang als erste.

**Sperrschicht–FET (JFET)**
Die konventionelle Bauform eines JFET kann man sich als ein Stäbchen aus z. B. mässig n–dotiertem Material vorstellen. Die beiden Stirnseiten werden mit ohmschen Kontakten versehen und bilden Source und Drain. Von der Oberseite und von der Unterseite wird stark p–dotiert. Darüber werden die Gatekontakte aufgebracht; diese werden miteinander fest verbunden.
Werden alle Kontakte miteinander verbunden, dann bilden sich jeweils asymmetrische Sperrzonen aus, die wegen der schwächeren Dotierkonzentration hauptsächlich im n–Gebiet liegen. Legt man an die Gates eine negative Spannung an, werden die Sperrzonen breiter und tiefer ins n–Gebiet reichen und sich, bei geeigneter Substratdichte, berühren. Umgekehrt wird eine positive Gate–Spannung den n–Kanal (n–channel) verbreitern. Die *Gate–Steuerspannung* legt also wie gewünscht die *Kanalbreite* fest.

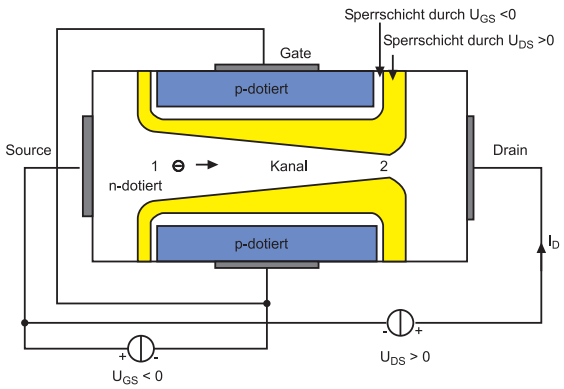



Abbildung 3.58.: JFET mit negativer Gate–Source–Spannung.

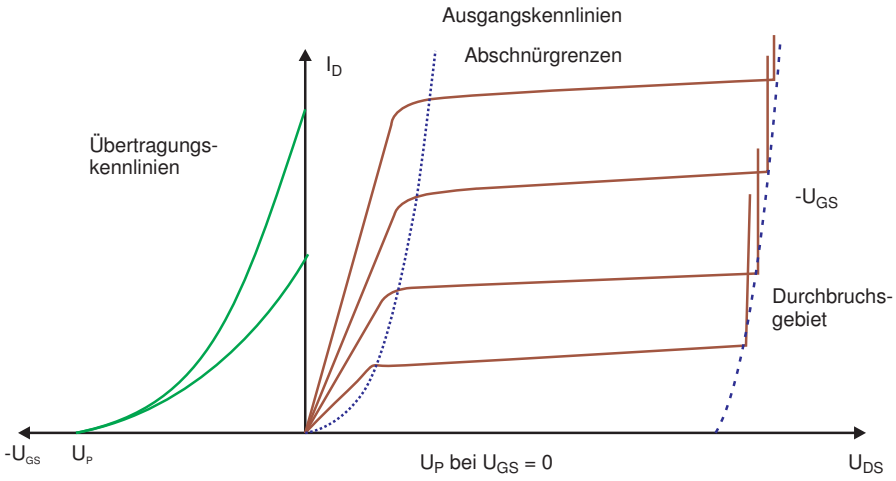



Abbildung 3.59.: Kennlinien des JFET.

Gehen wir nochmals zurück zur Ausgangslage mit verbundenen Kontakten. Wird jetzt einseitig an Drain eine positive Spannung angelegt, beginnt ein Elektronenstrom vom Source–Kontakt zum Drain–Kontakt zu fliessen. Gleichzeitig werden zum positiven Potential hin die Sperrzonen breiter; bei einer bestimmten Drain–Source–Spannung $U_{DS}$ werden sich die Sperrzonen an ihrem Drain–Ende erstmals berühren: die Abschnürspannung $U_p$ (pinch–off) ist erreicht. Für $U_{DS} < U_p$ werden wir eine lineare Abhängigkeit des Drainstroms $I_D \sim U_{DS}$ erwarten, darüber einen fast konstanten Sättigungstrom.
Mit einer negativen Gatespannung $U_{GS} < 0$ lassen sich die Sperrzonen verbreitern und die Abschnürspannung wird bereits bei kleineren Drainströmen erreicht. Entsprechende Übertragungs– und Ausgangskennlinien sind in Abbildung 3.59 wiedergegeben. Wegen des symmetrischen Aufbaus erhält man die gezeigten Ausgangs–Kennlinien für positive $I_D$ bei positiven $U_{DS}$, und gespiegelt am Ursprung solche für negative $I_D$ bei negativen $U_D$; im linearen Kennlinienbereich kann man den JFET als regelbaren Widerstand für Wechselspannungen (symmetrisch zu Null) benutzen. Bei grossen $U_{DS}$ brechen die Drain–Gate–Strecken durch.
Die wichtigsten JFET–Eigenschaften sind:

- sehr hoher Eingangswiderstand ($10^8 - 10^{10}\ \Omega$)
- dynamischer Ausgangswiderstand $r_{DS} = \frac{dU_{DS}}{dI_D} \approx 10^4 - 10^6\ \Omega$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

- Steilheit $g$ (transconductance) $= \left.\frac{\mathrm{d}I_{\mathrm{DS}}}{\mathrm{d}U_{\mathrm{GS}}}\right|_{u_{\mathrm{DS}}} \approx \ldots 10$ mA/V
- maximale Spannungsverstärkung $\approx 50 - 300$ (kleiner als beim Bipolar–Transitor).

Den JFET zeichnet eine vergleichsweise sehr lineare, aber kleine Spannungsverstärkung aus, deshalb wird er, wenn auch selten, als linearer Verstärker mit hohem Eingangswiderstand, kleiner Verstärkung und kleinen Rauschzahlen eingesetzt. Wie bei allen FET–Typen sorgt ein negativer Temperaturgradient ($I_{\mathrm{D}}$ sinkt, wenn die Temperatur $T$ steigt) für einen homogenen Stromtransport und damit für eine stabile Funktion.
Die statischen Strom–Spannungs–Kennlinien lassen sich nach Shockely (gradual channel approximation model) berechnen, siehe z. B. Roulston[?]. Das Bemerkenswerte daran ist, dass diese Theorie nicht nur auf JFET anwendbar ist, sondern auch auf MESFETs. Beide Male liegt eine Verarmungszone vor, aber im einen Fall ist es die eines p–n–Übergangs, im anderen die eines Schottky–Kontakts.

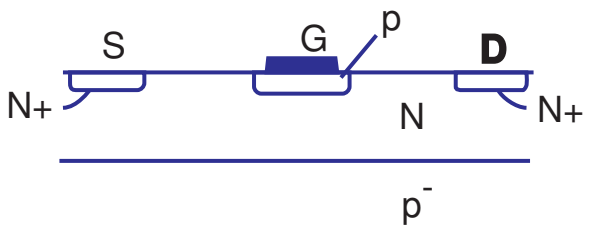



Abbildung 3.60.: Schematischer Aufbau eines planaren JFET (gezeichnet nach Roulston[?]).

Es gibt auch eine *planare Aufbau*variante des JFETs (siehe Abbildung 3.60). Auf einem schwach *p–dotierten Si–Substrat* wird eine n–dotierte Epitaxieschicht aufgewachsen. Für die Kontakte werden stark dotierte $n^+$–Wannen hergestellt, der Steuerkontakt besteht wieder aus einem stark überdotierten p–Bereich und der Metallisierung darüber. Man erhält von oben eine asymmetrische *Raumladungszone* in den p–Kanal hinein. Von unten wird dieser durch eine asymmetrische Verarmungszone begrenzt, die weit ins p–Substrat reicht. Je negativer die angelegte Gatespannung ist, desto schmaler der Kanal. Legt man wieder eine positive $U_{\mathrm{DS}}$ an, so erhält man wieder analog eine asymmetrische Verarmungszone und schliesslich berühren sich die Verarmungszonen wieder, mit den bekannten Folgen. Eine Diskussion eines p–Kanal–JFET liefe analog.
Die technische Verwirklichung des bereits angesprochenen *Double–Gate JFET* zeigt gegenüber dem besprochenen Planaraufbau folgende Modifikationen: das Substrat ist kontaktiert und bildet das Gate 2; bei negativer Vorspannung wandert die Sperrzone weiter in die Epischicht; bei angelegter $U_{\mathrm{DS}}$ wird sie asymmetrisch.

***GaAs–MESFET***

Auf den ersten Blick gleichen sich die Verhältnisse beim *GaAs MESFET* sehr. Auf einem intrinsischen oder semi–isolierenden (Cr–dotierten) Material wird zunächst eine undotierte 'Bufferlayer' gewachsen, dann folgt die n–dotierte Epischicht; die weiteren Schritte folgen analog. Die Herstellung ist ziemlich einfach.
Der Vorteil besteht einerseits in der hohen Beweglichkeit der Elektronen in *GaAs* und andererseits sind die Kapazitäten $C_{\mathrm{GS}}$ und $C_{\mathrm{GD}}$ deutlich kleiner als beim JFET

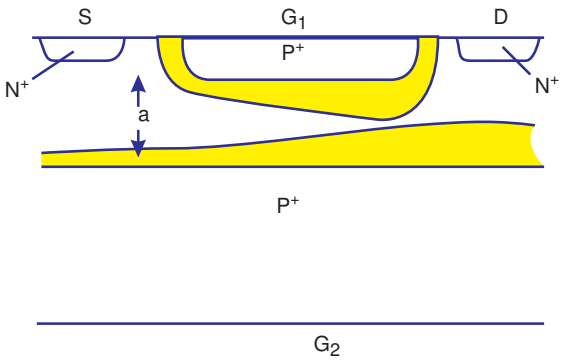



Abbildung 3.61.: Aufbau eines Double–Gate JFET (gezeichnet nach Roulston[?]).

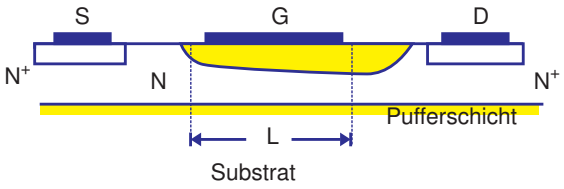



Abbildung 3.62.: Aufbau eines *GaAs MESFET* (Gezeichnet nach Roulston[?]).

(Seitenwandbeiträge entfallen). Für Mikrowellenanwendungen liegen die kritischen Längen heute unter 200 nm. Die maximale Frequenz steigt wegen $f_{\max} = g/C_{\text{Gate}}$ an.

Die Bauelemente zeigen exzellente DC– und HF–Eigenschaften. Sie werden als Subnanosekundenschalter oder als sehr rauscharmer Mikrowellenverstärker eingesetzt. Es gibt sogar integrierte Schaltkreise (MMICs, Monolithic Microwave ICs in Radarsystemen). Da für die *III–V–Halbleiter* kein brauchbares Oxid zur Verfügung steht, gibt es zum MESFET keine Alternative.

In der Praxis werden die einfach erscheinenden Verhältnisse des idealen Schottky–Kontakts durch die hohe Zahl der Grenzflächenzustände am Metall–Halbleiter–Interface des Gates doch recht kompliziert. Aber wie in Kapitel 3.2.5 schon gezeigt wurde, ändert dies nur etwas am Verständnis, nicht aber an der Funktion der Bauelemente.

Das nebenstehende Bild zeigt neben dem Grundaufbau die zugehörigen elektronischen *Bänderschemata*. Im *intrinsischen GaAs–Substrat* liegt $E_F$ (nahe) der Energielückenmitte. Im dotierten n–Kanal liegt $E_F$ nahe unter der Leitungsbandunterkante. Zum Metall–Halbleiter–Interface hin haben wir die Verarmungszone des Schottky–Kontakts, die positiven Ladungen der ortsfesten *Donatoren* verursachen

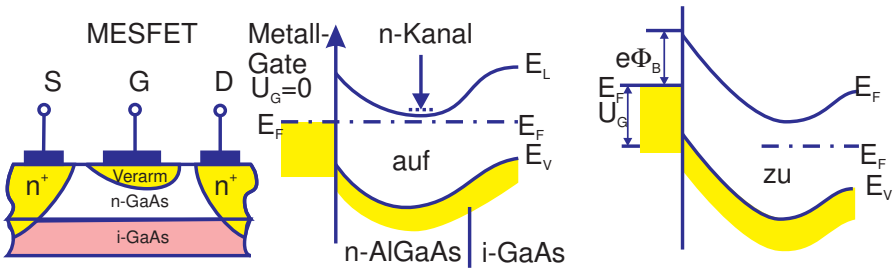



Abbildung 3.63.: MESFET (gezeichnet nach Ibach/Lüth[?]).

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

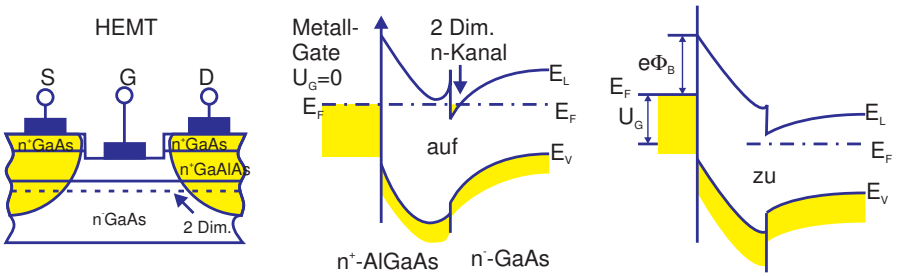



Abbildung 3.64.: HEMT (gezeichnet nach Ibach/Lüth[**?**]).

Bänderkrümmungen nach oben. Das Fermi–Niveau an der Grenzfläche ist durch die hohe Anzahldichte der Grenzflächenzustände nahe der Mitte des verbotenen Bandes 'gepinnt'. Der metallene Steuerkontakt ist also hochohmig vom n–Kanal getrennt.
Ohne Vorspannung können auf der niederohmigen Source–Drain–Strecke Elektronen injiziert werden, die im Potentialtopf des n–Kanals, abgesehen von einer leichten Rekombination, sicher transportiert werden können.
Durch Anlegen einer negativen Gatespannung $U_G$ wird der Kanal abgeschnürt, indem das Fermi–Niveau des Metalls gegenüber dem des *i–GaAs* angehoben wird. Da $E_F$ an der Grenzfläche gepinnt bleibt, werden die angrenzenden Bänder angehoben, die Verarmungszone wächst nach rechts, der Kanal verarmt an Ladungsträgern und wird schliesslich sperrend.
Die beiden Einträge 'Auf' und 'Zu' zeigen die Schalterfunktion des Transistors. Neben dem beschriebenen 'Normal–Auf'–MESFET (normally on), dessen n–Epischicht dicker als die Verarmungsschicht (ca. 50 nm) des Schottky–Kontakts ist, gibt es auch einen 'Normal–Zu'–MESFET (normally off). Dieser hat eine dünnere n–Epischicht (vergleichbar mit der Schottky–Kontakt–Verarmung) und sperrt deshalb bei $U_G = 0$; für $U_G > 0$ schaltet er auf.

**High Electron Mobility Transistor (HEMT)** Eine Weiterentwicklung des *GaAs–MESFET* ist der HEMT. Möglich wurde das durch die Fortschritte in der MBE (Molekularstrahlepitaxie) der *III–V–Halbleiter* bei der Herstellung von Heterostruktur–MESFETs und durch den gemischten Einsatz von optischen Lithographieverfahren und der *Elektronenstrahllithographie* zur Bearbeitung kleinster Abmessungen.
Resultat ist ein extrem rauscharmer *Transistor* für Höchstfrequenzanwendungen. Bei Gatelängen von 50 nm wurden Stromverstärkungs–Abschneidefrequenzen $f_t$ von weit über 300 GHz gemessen. Die Anwendung liegt bei Radar– und Satellitenkommunikation.
Der HEMT ist ein Heterostruktur–FET, dessen n–Kanal durch ein *2–dimensionales Elektronengas* (2 DEG) an der Grenzfläche zwischen einer AlGaAs/GaAs–Heterostruktur gebildet wird (vergleiche Kapitel 3.2.4). Im 'Auf'–Zustand (negative Gate–Spannung) können die Potentialtopf–Zustände gefüllt werden, im 'Zu'–Schaltzustand liegt das *Fermi-Energie*–Niveau zu tief, der 2 DEG–n–Kanal hat keine Ladungsträger.

**MOSFET (Metal Oxid Semiconductor FET)**
Wiederholt wurde schon darauf hingewiesen, dass MISFETs (oder auch IGFETs, Isolated Gate FETs genannt) praktisch nur auf *Si* als MOSFET realisiert wer-

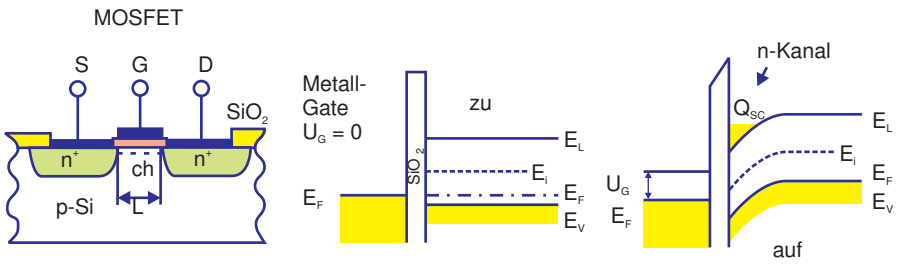



Abbildung 3.65.: MOSFET (gezeichnet nach [**?**]).

den. Wie schon in Kapitel 3.2.6, kann man am MIS–Kontakt, abhängig von der Substrat–Dotierung und der angelegten Spannung an der Grenzfläche Anreicherung oder Verarmung der Majoritätsladungsträger erreichen (oder gar Inversion der Minoritätsladungsträger; letztere spielte beim CCD–Element die entscheidende Rolle). Beim JFET war nur der Verarmungsbetrieb möglich, beim MISFET kommt die Möglichkeit des Anreicherungsbetriebs hinzu. Prinzipiell sind n– und p–Kanäle denkbar. Aber da Grenzschichtladungen meist positiv sind, entspricht dies einer Vorspannung am Gate und es gibt deshalb eine Tendenz zum n–Kanal (Verarmung im p–Material, Anreicherung im n–Material).

Beim MOSFET kontrolliert das elektrische Feld (Gatespannung gegen Kanalpotential) über das Oxid die *Ladungsträgerdichte* und damit den Kanalwiderstand. Der Eingangswiderstand ist extrem hoch ($10^{12} - 10^{16}\ \Omega$). Der abgebildete MOSFET–Aufbau gibt einen *n–Kanal–MOSFET* wieder: Ausgangsmaterial ist ein mässig *p–dotierter Si–Wafer*. Die beiden Source– und Draingebiete sind durch Ionenimplantation hoch n–dotiert worden; darauf befindet sich hoch n–dotiertes Poly–Silizium für die Kontaktierung.

Bei Gatespannung $U_G = 0$ befindet sich (die Grenzflächenladungen seien hier vernachlässigt) *p–Si* unterm Gate. Liegt eine Spannung $U_{SD}$ an, ist ein p–n–Übergang gesperrt, der Strom $I_D$ ist nahezu Null; der *Transistor* ist im 'Zu'–Zustand.

Machen wir jetzt das Gate positiver als die Source ($U_{GS} > 0$), so erhät man durch Inversion ($E_i < E_F$) eine starke n–Anreicherungsschicht nahe der Grenzschicht. Dieser n–Kanal verbindet die Source– und Gate–Wannengebiete leitend. Eine positive Gatespannung macht den n–Kanal–MOSFET 'auf'.

Bei invertierter Dotierung erhält man einen *p–Kanal–MOSFET* in einem *n–Si–Substrat*. Bei Gatespannung Null ist dieser Typ im 'Zu'–Zustand, aber bei negativer Gatespannung (invertierter Spannung) ist auch hier der 'Auf'–Zustand erreicht.

In Abbildung 3.66 sind die *Ausgangskennlinien* zweier n–Kanal–MOSFET–Typen zum Vergleich wiedergegeben. Auch hier zeigt die kleinere Kanallänge die besseren Eigenschaften, nämlich hier den höheren Ausgangsstrom. Aber aufgrund der kleineren Gatekapazität ist auch die Grenzfrequenz höher, das Schaltverhalten schneller.

Die Anwendung der MOSFETs nutzt die sehr guten dynamischen Eigenschaften dieser Bauelemente in HF–Verstärkern und zu Schaltzwecken. (Über die Verwendung in der Verstärkerintegration wird noch zu reden sein.) Grösster Nachteil der MOSFETs ist die maximal verträgliche Gatespannung von 50 - 100 V; elektrostatische Ladungen liegen im Alltag häufig 100 mal höher.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

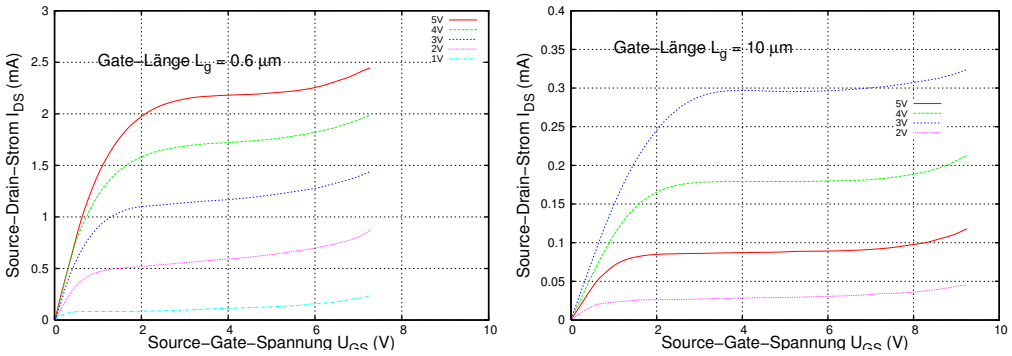



Abbildung 3.66.: Kennlinien von Kurz– und Langkanal–MOSFETs im Vergleich (nach Ibach/Lüth[**?**]).

### 3.3.3. CMOS–Technologie und Halbleiter–Speicher

MOSFETs eignen sich zur Herstellung hochintegrierter Digitalschaltungen besser als Bipolartransistoren; sie benötigen weniger Fläche auf dem *Si–Wafer* und sind deshalb preisgünstiger. Bis Mitte der achziger Jahre wurden die Schaltungen in NMOS–Technologie ausgeführt. Mit dem 1 Mb–DRAM–Speichern (1 Megabit Dynamic Random Access Memory) wechselten die Hersteller zur deutlich komplexeren CMOS–Technologie; die kleinsten Strukturgrössen betrugen MFS $= 1,2\ \mu$m (Minimal feasable size).

Durch die Herstellung von n–Kanal und p–Kanal Anreicherungs–MOSFETS nebeneinander auf einem Substrat ist deren Verschaltung zu einem Inverter mit besonderen Eigenschaften möglich: er schaltet schnell; seine Übertragungskennlinie zeigt einen steilen Wechsel, am Ausgang liegt entweder die volle Versorgungsspannung $V_{DD}$ oder die Null an. In den stationären Zuständen fliesst kein Strom; der Ruhestrom ist Null, nur beim Schalten wird Leistung verbraucht. Bei hohen Taktraten und häufigem Schalten relativiert sich zwar dieser Vorteil, aber die CMOS–Technik hat noch weitere: vereinfachtes logisches Design, kleine Anzahl von Transistoren in den peripheren Hilfs–Schaltkreisen (support circuits) (der sog. Speicherwirkungsgrad wird $> 50\,\%$), kleinere Rauschempfindlichkeit.

Heute beträgt der Marktanteil der CMOS–Technologie über 75 %; ca. 10 % dieser Bauelemente sind analoge Schaltungen. Das Bild unten gibt als technologisches Beispiel das Schaltbild und den schematischen Aufbau eines CMOS–Inverters auf n–Substrat wieder. Dabei sind n–MOSFET und p–MOSFET zueinander 'komplementär': CMOS (Complementary MOS).

Aus zwei Invertierern lässt sich ein einfaches *Flip–Flop* aufbauen, eine Schaltung mit zwei stabilen, eindeutig unterscheidbaren Schaltzuständen. Fügt man noch zwei weitere MOS–Transistoren hinzu, dann erhält man die Zelle eines SRAM–Speichers (Static RAM), eines sehr wichtigen Halbleiterspeichers also, siehe später.

Am Beispiel der technologischen Entwicklung des Inverters lässt sich die historische Entwicklung der *Halbleiter–Technologie* demonstrieren: Den einfachsten Invertierer kann man aus einem Ladewiderstand und einem *PMOS–Transistor* herstellen. Integrierte Planarwiderstände benötigen aber eine grosse Fläche. Viel platzsparender ist es, Source und Gate zu verbinden und so einen Widerstand zu realisieren. Die entsprechede Technologie heisst '*PMOS Aluminium–Gate–Prozess*'. Es handelt sich um eine sog. Einkanal MOS–Technik auf einem n–Silizium–Substrat, bei

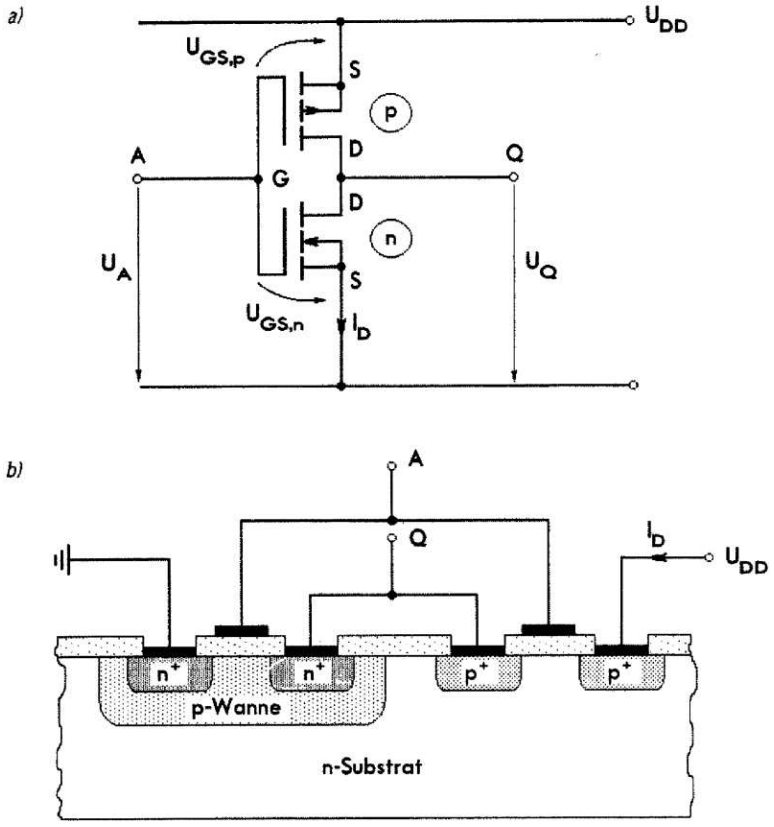



Abbildung 3.67.: CMOS–Technik: Inverter im a) Schaltbild und b) im technologischen Aufbau.

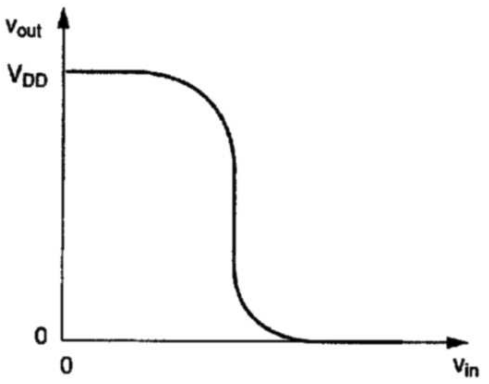



Abbildung 3.68.: CMOS–Inverter: Übertragungskennlinie[**?**].

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

der es nur selbstsperrende p–Kanal–MOSFETs (enhancement type E) gibt. Diese Technik benötigt nur vier verschiedene Masken und eine Metallisierungs–Schicht.

Eine Voraussetzung für diese Technik ist das Vorhandensein ausgeklügelter Lithographietechniken. Bis heute verwendet man in der Chipfertigung *optische Lithographietechniken*, um in aufgeschleuderten, photoempfindlichen Polymerlacken laterale Strukturen lokal zu öffnen, durch die additiv oder subtraktiv die Waferoberfläche oder oberflächennahe *Si–Schichten* gezielt verändert werden können (Oxidieren, Ätzen, Dotieren). Schritt für Schritt können so hochkomplexe Planar–Schaltungen aufgebaut werden. Die kleinste Strukturgrösse betrug anfangs $4 - 10$ $\mu$m. Schneller wurden die Inverter durch die Einführung des 'NMOS Aluminium–Gate–Prozesses' auf p–Substrat.

Eine merkliche Verbesserung erzielte man aber erst mit der Einführung der 'n–Kanal Aluminium–Gate MOS–Technik'. Die Ladewiderstände werden hier durch selbstleitende NMOSFETs gebildet (depletion type D), die Schalttransistoren sind wieder vom E–Typ: sog. E/D–Inverter. Der Preis für die Verbesserung sind zwei zusätzliche Masken– und ein zusätzlicher As–Implantations–Dotierschritt zur Einstellung der Schwellenspannung des D–Typ–FETs.

Die ursprüngliche Technik erfuhr im Laufe weniger Jahre zahlreiche Verbesserungen. Die wichtigste war wohl der Übergang zur 'NMOS Silizium–Gate Technologie'. Dabei wird Aluminium als Material zur Gate–Kontaktierung ersetzt durch polykristallines Silizium, das durch nachträgliche Dotierung leitfähiger gemacht werden kann ('Polysilizium', durch LPCVD–Abscheidung). Dies geschieht gleichzeitig mit der Source/Drain–Implantation, d. h. man benötigt hierfür keine spezielle Maske mehr, dieser Implantationsschritt ist 'selbstjustierend'. Zusätzlich verkleinern sich die störenden Überlappungskapazitäten ('Miller capacities') zwischen Gate und p–dotierten Source/Drain–Kanälen.

Weitere Schritte sind vergrabene $n^+$–dotierte *Si–Leiterbahnen* und eine weitere selbstjustierende Oxidationstechnik, die LOCOS–Technik (Local Oxidation of Silicon), die zusätzlich Abscheide– und Ätztechniken für das hierbei benötigte $Si_3N_4$ erfordert. Überhaupt werden die eingesetzten Materialien immer zahlreicher und raffinierter; etwa Silizide zur Kontaktierung, Nitride als Diffusionsbarrieren, Mischoxide, Gläser und AlSiCu–Legierung bzw. seit neuestem Cu (als Leiterbahnenmaterial in bis zu vier Metallisierungsebenen).

Zunächst aber erzwang die in den hochintegrierten Bauelementen mit der Verkleinerung der Strukturgrössen und der Steigerung der zu einer Schaltung gehörenden Transistoranzahl über die Jahre angestiegene Verlustleistung den Wechsel zur CMOS–Technologie. Der zugehörige Inverter wurde bereits eingangs besprochen.

Das Endprodukt des '*n–Wannen Silizium–Gate CMOS–Prozesses*' gibt die Abbildung 3.69 wieder. Auf einem mässig p–dotierten (100)–Substrat wurde rechts der NMOS–FET direkt hergestellt, links dagegen wurde eine mehrere $\mu$m tiefe 'Wanne' (engl. well) mit Phosphor n–dotiert, um darauf durch $p^+$–Implantation den PMOSFET zu bauen. Die Schaltungsdesigner bemühen sich, möglichst nicht nur einzelne, sondern immer mehrere p–Kanal–Transferelemente in eine grosse n–Wanne zu plazieren, was die Funktionssicherheit im späteren Betrieb erhöht.

Reale CMOS–ICs herzustellen erfordert heute etwa 25 Masken. Enthalten die Schaltungen (wie bei den DRAMs) noch aufwendige MOS–Kondensatoren, so steigt die Maskenanzahl auf 30 bis 35 an. (Zum Vergleich: für eine moderne Laserdiode benötigt man nur 5 verschiedene Masken.)

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

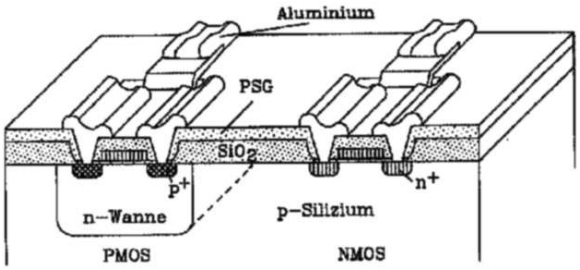


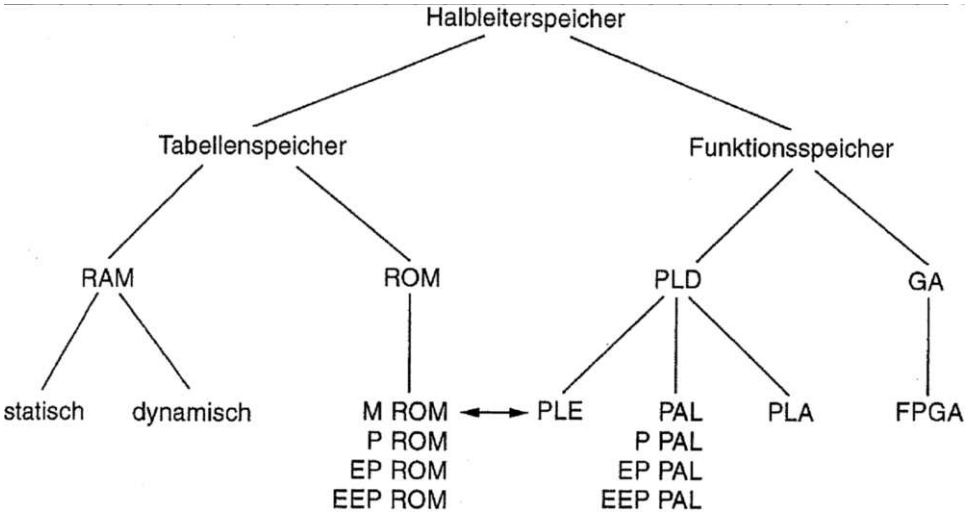

Abbildung 3.69.: CMOS: p– und n–Kanal–E–Type MOSFETs, schematisch[**?**].

Die produzierten Schaltungen enthalten immer mehr Transistoren:

| | | |
|---|---|---|
| $10^5 - 10^7$ | VLSI | Very Large Scale Integration Technique, |
| $10^7 - 10^9$ | ULSI | Ultra Large Scale Integration Technique, |
| $> 10^9$ | SLSI | Super Large Scale Integration Technique. |

Der im Jahr 2000 gerade in Serienproduktion gegangene 256 Mb–DRAM beispielsweise enthält rund 270 Milliarden Transistoren, das 1 Gb–DRAM wird das erste SLSI–Bauelement sein.
Hierzu werden die Schaltungslayouts nicht nur von Speichergeneration zu Generation 'geshrinkt', sondern dreidimensionaler. Die MOSFET–Zelle wird dabei mit dem Skalierungsfaktor $\alpha$ verkleinert, z. B. $\alpha = 0,7$. Dann erhöht sich die Schaltgeschwindigkeit mit $1/\alpha$, die Schaltungsdichte $\sim 1/\alpha^2$, der Leistungsverbrauch pro Inverter sinkt $\sim \alpha^2$, die Verlustleistungsdichte bleibt annähernd konstant.
Die aktuelle Leiterbahnbreite liegt bei $110 - 130$ nm in der Spitzentechnologie. Die nächsten Stützpunkte (nodes) in der technischen Entwicklung werden, siehe den Skalierungsfaktor $\alpha$, sein: 100 nm (im Jahre 2005/6), 70 nm, 50 nm (im Jahre 2011/12) und 35 nm (2014 oder früher). Ein Ende jeglicher CMOS–Technologie wird bei 20-30 nm erwartet. Die Generationswechsel erfolgen gemäss dem 'Mooreschen Gesetz' etwa alle $\lesssim 3$ Jahre, siehe NTRS (National Roadmap for Semiconductors) der SIA (Semiconductor Industry Association) der USA bzw. seit 1999 die ITRS (International Technology Industry Association) aller Industriestaaten. Diese 'road maps' wagen jeweils eine 15 Jahre–Prognose für die künftige technologische Entwicklung in der Halbleiterindustrie, an der sich die Produzenten und Zulieferer orientieren können.
Eine der spannendsten Fragen des Jahres 2001 wird sein, mit welchem Lithographieverfahren die Industrie die 100 nm–Technologie angehen wird. Excimer–Laser gestützte 'optische' Lithographieverfahren haben überraschenderweise auch hier sehr grosse Chancen. Allerdings wird das Problem der immer kleiner werdenden Tiefenschärfe immer gravierender; die Topographie einer CMOS–Schaltung muss deshalb auch nach vielen Prozessschritten noch sehr plan bleiben, will heissen $< 1/2$ $\mu$m. Aus diesem Grund werden bereits heute planarisierende Polierschritte (CMP Chemomechanical Polishing) in die Prozessfolge eingefügt, was exzessive Reinigungsschritte mit sich zieht. Für die nachfolgenden Technologiegenerationen stehen Projektionsverfahren mit Elektronen und Ionen optional zur Verfügung, ebenso Photonen des EUV (Extreme Ultraviolet bei 13,5 nm Wellenlänge). Dabei ist die Wechselwirkung zwischen Ionen und Resist die für die Auflösung günstigste. Dagegen wird die Röntgenstrahl–Lithographie als Schattenwurfverfahren nur ein wichtiges Spezialverfahren in der Mikromechanik zur Belichtung sehr dicker

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*




| | | | | | |
|---|---|---|---|---|---|
| RAM | = | Random Access Memory | PLD | = | Programmable Logic Device |
| ROM | = | Read Only Memory | PLA | = | Programmable Logic Array |
| M | = | Masken-programmiert | PAL | = | Programmable Array Logic |
| P | = | Programmierbar | PLE | = | Programmable Logic Element |
| EP | = | Löschbar und programmierbar | GA | = | Gate Array |
| EEP | = | Elektrisch löschbar und programmierbar | FPGA | = | Field Programmable GA |

Abbildung 3.70.: Übersicht über die verbreitesten Halbleiter–Speichertypen[?].

Resists (mehrere mm) mit $\leq$ $\mu$m Auflösung bleiben.
Das wichtigste Marktsegment neben den Mikroprozessoren ist das der digitalen Halbleiter–Speicher. Die Abbildung 3.70 gibt einen Überblick über die gängigen Bauelement–Typen.
Im Folgenden sollen die Schaltungszellen der wichtigsten Tabellenspeicher jeweils kurz diskutiert werden.
In Tabellenspeichern legt man 'wortweise' den Inhalt von beliebigen Tabellen (z. B. Computerprogramm, Messwerte) ab. Jedes Wort hat seine eigene Adresse; bei RAMs (Random Access Memory) und ROMs (Read Only Memory) kann, im Gegensatz zu Schieberegistern, jederzeit auf jede Adresse zugegriffen werden ('wahlfreier Zugriff'). RAMs werden im Normalbetrieb beschrieben und gelesen; beim Abschalten der Betriebsspannung verlieren sie ihren Speicherinhalt (flüchtiger Speicher, volatile memory). ROMs sind im Normalbetrieb nur auslesbare Festwertspeicher. Je nach Typ sind sie ein– oder wenige Male beschreibbar; sie behalten aber auch ohne Versorgungsspannung sehr lange ihren Inhalt (nicht flüchtiger Speicher, non-volatile memory).
*Statische RAMs* (*SRAMs*) sind wie DRAMs bit– und wortweise organisiert, benötigen also Hilfsschaltkreise wie Adressdecoder (Spalten– und Zeilen–Decoder), Ein– und Ausleseverstärker, Zwischenspeicher (adress–latch), etc. Die Speicher–Matrix besteht aus Flip–Flop–Speichern. In Abbildung 3.71 sieht man eine Flip–Flop–Speicherzelle in CMOS–Technik, die sog. statische Sechs–Transistor–Zelle. Es handelt sich um ein kreuzgekoppeltes Flip–Flop in Komplementärkanal–Technik ($T_1$, $T_3$ und $T_2$, $T_4$). Zum Ändern der stabilen Zustände werden die n–Kanal–E–Typ–MOSFETs $T_5$ und $T_6$ benötigt. Zum Schreiben werden kurzzeitig die Schalter $S_w$ und $S_s$ oder $\overline{S_s}$ geschlossen, zum Lesen nur $S_w$. Nach dem Einschalten befin-

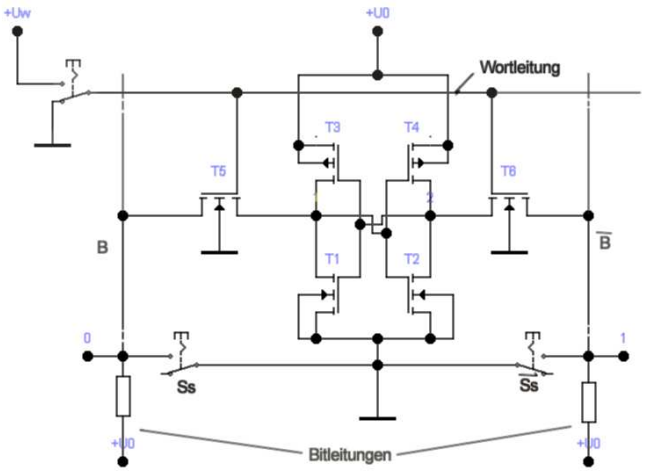



Abbildung 3.71.: CMOS–Speicherzelle eines SRAMs[**?**].

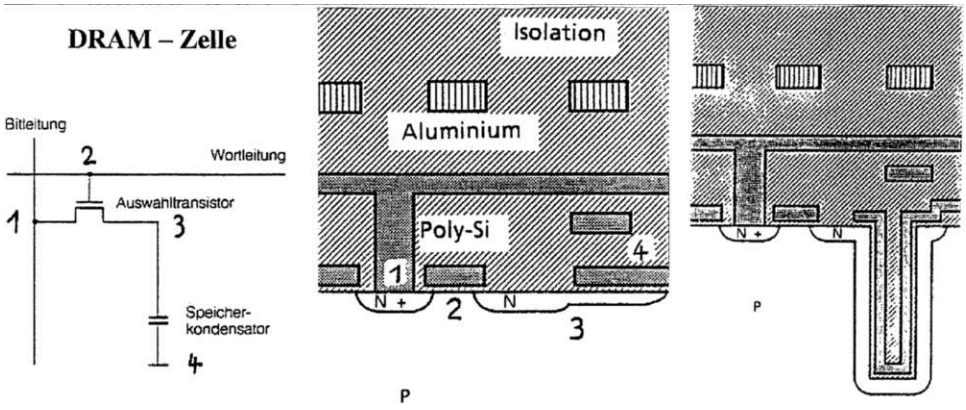



Abbildung 3.72.: DRAM–Zelle, von links: Schaltbild, schematischer Aufbau der 1 Mb– und einer 4 Mb–Zelle mit Grabenkondensator[**?**].

den sich die Flip–Flops statisch in den Zuständen 0 oder 1 und können sofort beschrieben werden.

*Dynamische RAMs* (DRAMs) haben ungefähr die vierfache Speicherkapazität vergleichbarer SRAMs. Ihre Speicherzelle besteht nur aus einem Auswahltransistor (Schalter) und einem MOS–Kondensator (Speicherkapazität $C_S$). Die Source des selbstsperrenden NMOS–Transistors ist mit der Bitleitung (Datenleitung mit der Leitungskapazität $C_b$) verbunden, das Gate mit der Wortleitung. Schliesst man — über eine positive Spannung am Gate — den Schalter, fliessen Ladungen von oder zur Bitleitung; der Kondensator wird ent– oder aufgeladen. Den Speicherkondensator bilden die flächig erweiterte Drain und, isoliert über ein hochwertiges Oxid ($d < 150$ nm), die Gegenelektrode aus Metall oder heute üblicherweise aus Polysilizium. (Im Bild: 1 = Bitleitung über der Source, 2 = Wortleitungsende als Gate über dem n–Kanal, 3 = Drain mit planarer Erweiterung (storage node), 4 = flächige Topelektrode (cell plate), darüber im isolierenden Oxid eine Wortleitung einer nebenliegenden Speicherzelle.)

Beim Lesen wird die Kondensatorspannung registriert und gleichzeitig die Ladung verändert (destruktives Lesen). Das Vorhandensein von Ladung entspricht der logischen ‘1’ und diese wird beim Lesen verkleinert, da die Datenleitung auf

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

niedrigerem Potential liegt, umgekehrt liegt die Datenleitung aber auf höherem Potential als der leere Kondensator, der die logische '0' repräsentiert. Das beim Ladungstransfer erzeugte *Signal* $\Delta V_\mathrm{b}$ ist sehr klein, typisch 100 - 200 mV, denn die Kapazität der Datenleitung und unvermeidbare Streukapazitäten sind mehr als zehn Mal grösser als die Speicherkapazität, die man ja der Fläche wegen klein halten will. Näherungsweise gilt:

$$\Delta V_\mathrm{b} = \frac{V_\mathrm{DD}}{2} \frac{1}{1 + \frac{C_\mathrm{b}}{C_\mathrm{S}}} \; . \tag{3.3.2}$$

Weil bei jedem Technologiegenerationswechsel sowohl die interne Versorgungsspannung $V_\mathrm{DD}$ als auch $C_\mathrm{b}$ um den selben Faktor $d$ verkleinert werden, muss $C_\mathrm{S} \gtrsim 25 - 50$ fF pro Zelle konstant bleiben. Die empfindlichen Schreib–Lese–Verstärker sind praktisch nicht mehr verbesserbar. Jedes (wortweise) Lesen erfordert als ein (wortweises) Neuschreiben der Speicherinhalte (Refresh).
Unvermeidbar sind die Source–Drain–Leckströme des Schalttransistors, d. h. die Kondensatorladung ändert sich immer. Deshalb müssen alle Speicherzellen periodisch 'aufgefrischt' werden (Refresh Zyklen). Die Haltezeit einer Zelle beträgt typischerweise 1 – einige 100 ms; bei DRAMs mit 60 ns Zugriffszeiten erfolgt das Auffrischen in 16 $\mu$s alle 8 ms. Die Refresh–Schaltungen beanspruchen einen erheblichen Flächenanteil.
Die Kapazität des planaren Speicherkondensators ist mit $C_\mathrm{S} = \frac{Q}{U} = \epsilon_0 \cdot \epsilon_\mathrm{r} \cdot \frac{A}{d}$ gegeben. Bei einer internen Spannung von 2,5 V und bei Verwendung des gewohnten $SiO_2$ macht es die Durchschlagfestigkeit des Oxids (7 MV/cm, eigentlich 5 MV/cm) nötig, dass ein 30 fF–Kondensator eine Fläche von rund 3 $\mu$m$^2$ benötigt. Für ein 4 Mb–DRAM war dieser Flächenbedarf bereits zu gross. Die Auswege sind:

1. Die Verwendung von Oxiden mit grösserem $\epsilon_\mathrm{r}$ und/oder grösserer Durchschlagsfestigkeit (Beispiel: ONO = oxidized nitride–oxide sandwich isolator).

2. Die Vergrösserung der Fläche durch Übergang zu dreidimensionalen Zell–Kondensatoren (Beispiele: Grabenkondensator (trench), Stapelkondensator (stack).

Gerade der Übergang zum vergrabenen Kondensator hat die Integrationsdichte in den letzten Jahren stärker erhöht, als es nach den 'Roadmaps' zu erwarten war. Diese Technik stellt grösste Anforderungen sowohl an die isotrope Nassätztechniken als auch an die anisotropen Plasmaätztechniken. Das Ätzen eines merere $\mu$m tiefen, verrundeten Zylinderlochs, die Abscheidung eines geschlossenen, hochwertigen Oxids und das Einfüllen der Deckelelektrode sind äusserst komplexe Verfahrensschritte. Die selben Techniken werden in STI (Shallow Trench Isolation) –Prozess zur elektrischen Trennung von vergrabenen Leitungen verwendet. Die Kombination von Oxid– und Trench–Isolation kennzeichnet heute die 'Advanced CMOS–Techniques'. Obwohl die Stapelkondensatoren weit grössere Möglichkeiten in der Wahl des Isolators zulassen, führt die mit ihrer Herstellung verbundene Aufrauhung der Chipoberfläche in der gegenwärtigen Lithographietechnik zu grossen Problemen.
Abschliessend einige Bemerkungen zu den nichtflüchtigen Festwertspeichern. Sie basieren ebenfalls auf MOS–Strukturen. Die früher üblichen *PROM*–Bausteine (Programmable ROM) mit ihren Schmelzsicherungen für jedes Bit sind heute nicht

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

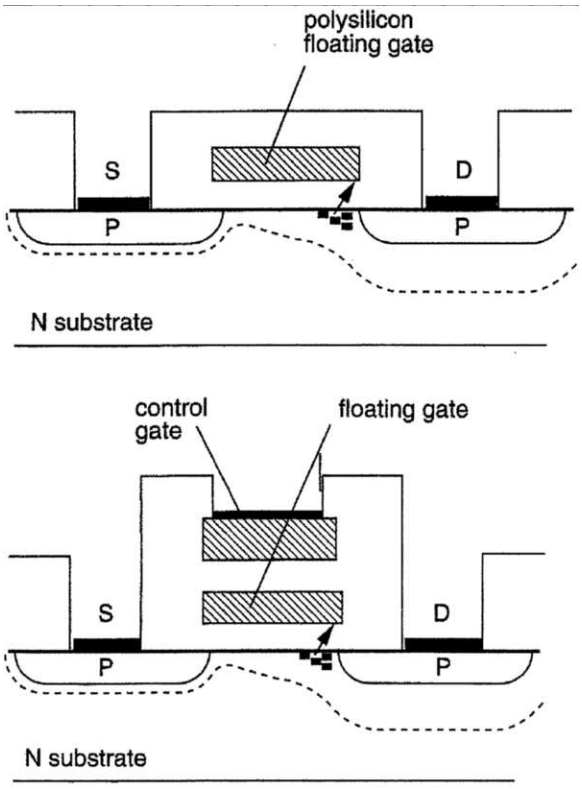



Abbildung 3.73.: FAMOS–Struktur des EPROMs und SAMOS–Struktur des EEPROMs[**?**].

mehr üblich. An ihre Stelle sind die EPROMs (Erasable PROM) getreten. Die ursprüngliche MOS–Struktur verwendet ein 'Floating Gate', das von einem etwa 100 nm dicken Oxid vom Silizium isoliert ist und so praktisch keine Leckströme zulässt. Legt man beim Schreiben zwischen Source und Drain eine ca. 20 V hohe Spannung an, so gelangen via 'avalanche injection' heisse Elektronen auf die Gateelektrode, wo sie wenigstens 10 Jahre als Ladung erhalten bleiben. Vorhandene Ladungen schalten die Source–Drain–Strecke bleibend durch, nicht vorhandene Gateladung sperrt. Mit einer ca. zwanzigminütigen UV–Lichtbestrahlung ist der EPROM–Speicher löschbar.
Eine verbesserte Struktur ist die SAMOS (stacked gate), die eine zweite, floatende Gate–Elektrode als Auswahl–Elektrode vorsieht. Zum Schreiben wird hier zusätzlich auf das Auswahlgate eine positive Spannung gelegt.
Eine SAMOS–Struktur verwendet auch das EEPROM (Electrically Erasable PROM). Zum Löschen wird eine positive Spannung ans Kontrollgate gelegt, alle anderen Kontakte auf Null; die Elektronen tunneln dann von der Floating–Gate–Speicherelektrode zum Kontrollgate. Modifizierte SAMOS–Strukturen haben z. B. verringerte Abstände zu einem erweiterten Draingebiet und Löschen durch Elektronentunneln vom Floating–Gate zum Drainkontakt. Die Anzahl der Schreib/Löschzyklen ist auf ca. $10^4$ begrenzt.
*Nebenbemerkung:* Eine modifizierte FAMOS–Struktur verwenden die sog. Flash–Speicher, die dadurch ebenfalls elektrisch löschbar sind. Ihr Oxid ist lokal dünner und erfordert ein extremes Mass an Perfektion für ca. $10^6$ Schreib/Löschzyklen.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

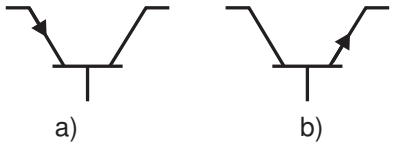



Abbildung 3.74.: Schaltsymbole eines a) pnp– und b) npn–Transistors (gezeichnet nach [**?**]).

### 3.3.4. Bipolare Transistoren (BJT Biplor Junction Transistor, Injektionstransistoren)

Zwei gegeneinander geschaltete p–n–Übergänge bilden einen Bipolartransistor. Es gibt zwei Reihungsmöglichkeiten: *npn* und *pnp*. Im gebräuchlichen Fall des npn–Transistors tragen Elektronen den für das Bauelement bestimmenden Strom, im Falle des pnp–Transistors sind es Löcher. Die drei Dotierbereiche sind jeweils ohmsch kontaktiert; die Anschlüsse werden immer *'Emitter'*, *'Basis'* und *'Kollektor'* genannt.

Die Bezeichnung bipolar rührt von dem Fakt her, dass in jedem Dotiergebiet Majoritäts– und Minoritätsladungsträger für die Bauelementefunktion wesentlich sein können; der Begriff 'Injektion' im Zweitnamen weist darauf hin, dass vom Emittergebiet Majoritätsladungsträger in das Basisgebiet injiziert werden, wo sie Minoritätsladungsträger weit ab vom thermodynamischen Gleichgewicht sind. Der Emitter–Basis–Übergang wird im Durchlass–, der Basis–Kollektor–Übergang stets in Sperrrichtung betrieben. Die relevanten Ausdehnungen aller Dotiergebiete sind (100 - 1000 mal) kleiner als die Diffusionslängen der Ladungsträger, d. h. die Kontaktzonen beeinflussen einander stark.

Verglichen mit den unipolaren Transistoren, die i. allg. schnell, rauscharm und temperaturstabil gebaut werden können, sind bipolare Transistoren in geeigneten Beschaltungen vor allem hervorragende Leistungsverstärker. Man begegnet ihnen häufig als Einzelelement; in der Höchstintegration spielen sie zahlenmässig eine untergeordnete Rolle.

#### n–p–n–Bipolartransistoren

Abbildung 3.75 zeigt den prinzipiellen eindimensionalen Aufbau dieses Bauelements in sog. *Basisschaltung*, d. h. der Basiskontakt liegt auf Masse. Der EB–Kontakt ist mit $U_{\mathrm{BE}} \approx -0,7$ V leitend gesteuert, $U_{\mathrm{BC}}$ wählt man so hoch (z. B. $= \pm 5$ V), dass auch bei grösserem Lastwiderstand der BC–Kontakt kräftig gesperrt wird. Die Dotierkonzentrationen sind sehr ungleich und die Übergänge abrupt. Das Emittergebiet ist stark $n^+$ dotiert, die Basis ist beim klassischen *Transistor* homogen und deutlich niedriger p dotiert, der Kollektor nochmals deutlich niedriger n dotiert. Die Basisschicht ist sehr dünn, das Kollektorgebiet vergleichsweise weit. Die Raumladungszonen sind asymmetrisch; die der Injektionsdiode sehr schmal und niederohmig, die der gesperrten Diode sehr breit und tief im Kollektorgebiet, ihre zugehörige Kapazität sehr klein, ihr Widerstand hoch.

Das *elektronische Bänderschema* (Abbildung 3.76) gibt — gestrichelt — die Verhältnisse ohne Vorspannungen im thermodynamischen Gleichgewicht wieder und — durchgezogen — die des Normalbetriebs. Von links diffundieren Elektronen $I_{\mathrm{nE}}$ durch die schmale Sperrzone; je grösser $N_{\mathrm{E}}$, desto mehr. Gegenüber dem thermi-

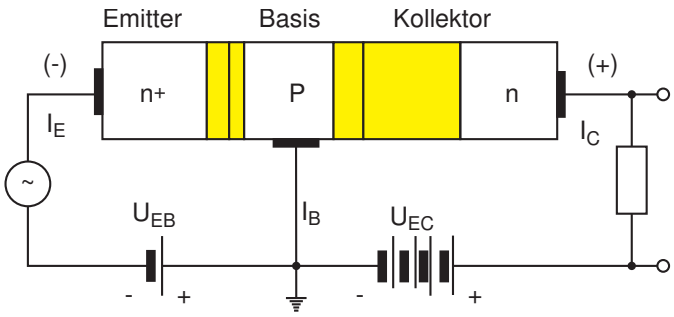



Abbildung 3.75.: Basisschaltung eines npn–Transistors (gezeichnet nach [**?**]).

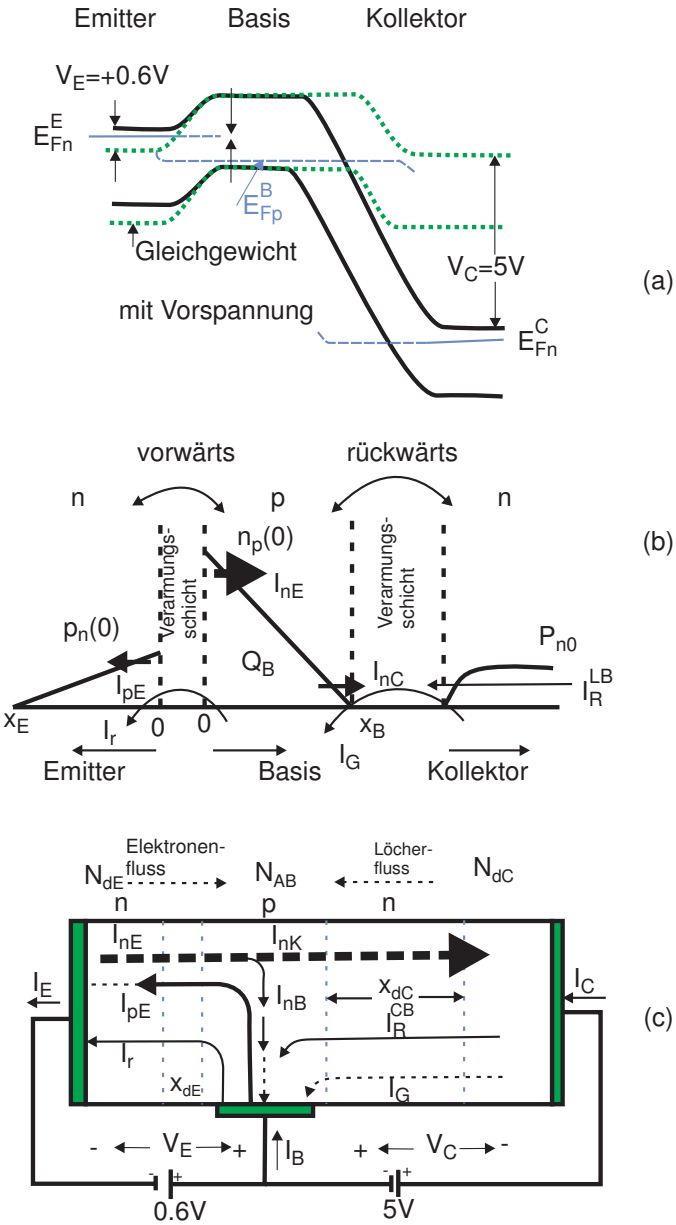



Abbildung 3.76.: *npn–Transistor* im Gleichgewicht und Nichtgleichgewicht (gezeichnet nach [**?**]).

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

schen Gleichgewicht $n_{\mathrm{B}_0}$ ist am emitterseitigen Basisrand die Minoritätsladungsrägerdichte $n_{\mathrm{B}}$ erhöht gemäss

$$n_{\mathrm{B}}(0) = n_{\mathrm{B}_0} \exp\left[U_{\mathrm{BE}}/k_B T\right] \ . \tag{3.3.3}$$

Der Emitterstrom teilt sich in der Basis auf. Der Grossteil der Elektronen ($\geq 99\,\%$) wird vom elektrischen Feld des gesperrten B–C–Kontakts in den Kollektor abgesaugt: $I_{\mathrm{nC}}$. Der kleine Rest ($\leq 1\,\%$) rekombiniert mit den an der Basis injizierten Löchern. (Diese Stromverteilung ist eine Ursache des vergleichsweise vergrösserten Rauschens.) Näherungsweise gilt also $I_{\mathrm{nE}} \approx I_{\mathrm{nC}}$, solange die Basislänge deutlich kleiner als die Diffusionslänge $L_{\mathrm{n}}$ ist.
*Nebenbemerkung*:
Das Eindiffundieren der Elektronen in die Basis limitiert das zeitliche Verhalten des klassischen Bipolartransistors durch Laufzeiteffekte. Moderne Varianten sind durch ein inhomogenes Basisdotierprofil gekennzeichnet; das damit verbundene elektrische Feld beschleunigt die Ladungsträger zum Kollektor hin: 'Drifttransistoren' statt 'Diffusionstransistoren'. (Diese erreichen Transitfrequenzen von einigen GHz.)
Die Verteilung der Minoritätsladungsträger in der Basis ist gegenüber der einer blossen Diode ($\exp\left[-x/L_{\mathrm{n}}\right]$) stark verändert; die Anzahldichte fällt aufgrund der Sogwirkung des gesperrten BC–Kontakts ($\sim \exp\left[-x/x_{\mathrm{B}}\right]$) praktisch am kollektorseitigen Rand der Basis (bei $x_{\mathrm{B}}$: abhängig von $U_{\mathrm{BC}}$) auf Null. Oberhalb des thermischen Gleichgewichts, zum Emitter hin, überwiegt die bereits oben eingeführte Rekombination, zum rechten Rand der Basis hin ist $n_{\mathrm{B}} < n_{\mathrm{B}_0}$, d. h. hier liegt eine schwache Generation von Minoritätsladugsträgern vor. Für den Elektronenstrom gilt:

$$I_{\mathrm{nE}} \approx eAD_{\mathrm{nB}}\mathrm{grad}\ n_{\mathrm{B}} \ , \tag{3.3.4}$$

$$\text{mit} \qquad \mathrm{grad}\ n_{\mathrm{B}} = \frac{\partial n_{\mathrm{B}}}{\partial x} = -\frac{n_{\mathrm{B}}(0)}{X_{\mathrm{B}}} \ . \tag{3.3.5}$$

Daraus folgt:

$$I_{\mathrm{nE}} \approx \frac{eAD_{\mathrm{nB}} n_{\mathrm{B}_0}}{x_{\mathrm{B}}} \exp\left[eU_{\mathrm{BE}}/k_B T\right] \approx I_{\mathrm{nC}} \ . \tag{3.3.6}$$

Die Minoritätsladungsträgerdichte im Kollektor ist zum Basisrand hin aufgrund des Löchergenerationsstroms $I_{\mathrm{G}}$ ebenfalls auf nahezu Null reduziert. Ein zweiter Leckstrom besitzt noch mehr Relevanz. Die in die Basis injizierten Löcher können in der Basis, im EB–Sperrgebiet oder aber erst im Emittergebiet rekombinieren: $I_{\mathrm{R}}$. Dieser Löcherstrom wächst mit der Majoritätsladunsgträgeranzahldichte, also mit der p–Dotierkonzentration der Basis. Die Basis kann deshalb nicht beliebig niederohmig gemacht werden, die eindiffundierenden Löcher und die damit verbundene Kapazität limitieren das Schalt–/Zeitverhalten aller Bipolartransistorvarianten. (Eine kleine Löcherdiffusionslänge $L_{\mathrm{pE}}$ wäre hilfreich.)
Ein Mass für die Leistungsfähigkeit des Transistors ist das Verhältnis zwischen dem Majoritätsladungsträgerstrom, der die Emittergrenze erreicht und dem Anteil, der im Emitter durch Rekombination verloren geht (Emitterwirkungsgrad). Häufig angegeben wird die sog. *Emittereffizienz* $\gamma$:

$$\gamma = \frac{I_{\mathrm{nE}}}{I_{\mathrm{E}}} = \frac{I_{\mathrm{nE}}}{I_{\mathrm{nE}} + I_{\mathrm{pE}} + I_{\mathrm{R}}} < 1 \ , \tag{3.3.7}$$

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

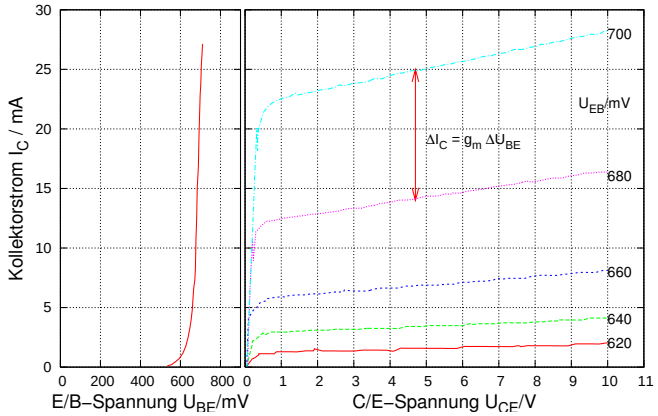



Abbildung 3.77.: Übertragungs– und Ausgangskennlinie nach [**?**].

da $I_{\mathrm{pE}}$ und $I_{\mathrm{R}}$ entgegengesetztes Vorzeichen besitzen. Die klassische Bipolartransistoren–Theorie liefert die Aussage, dass für eine hohe Effizienz das Produkt $D_{\mathrm{nB}} \cdot n_{\mathrm{B}_0}$ möglichst gross sein soll und deshalb wird das Emittergebiet sehr hoch dotiert.
Für die technischen Ströme gilt der Zusammenhang:

$$I_{\mathrm{E}} \;=\; I_{\mathrm{B}} + I_{\mathrm{C}} \; . \tag{3.3.8}$$

Im Detail gilt:

$$I_{\mathrm{E}} \;=\; I_{\mathrm{nE}} + I_{\mathrm{pE}} \qquad \text{und} \tag{3.3.9}$$

$$I_{\mathrm{C}} \;=\; I_{\mathrm{nC}} + I_{\mathrm{pC}} \; . \tag{3.3.10}$$

Speziell für den Basisstrom gilt:

$$I_{\mathrm{B}} \;=\; I_{\mathrm{nB}} + I_{\mathrm{pE}} - I_{\mathrm{pC}} \; , \tag{3.3.11}$$

er setzt sich also aus dem $\approx 1\,\%$–Anteil von $I_{\mathrm{nE}}$ und den beiden Leckströmen zusammen. Im Gegensatz zu den Feldeffekt–Transistoren geschieht die *Steuerung* der bipolaren Transistoren nicht leistungslos, der Basisstrom belastet stets den Steuergenerator.
Die erwähnte ‘klassische’ Theorie liefert schliesslich für die Übertragungskennlinie $I_C(U_{\mathrm{B}E})$ näherungsweise den Zusammenhang:

$$I_{\mathrm{E}} \approx \frac{eAD_{\mathrm{nB}}n_{\mathrm{B}_0}}{L_{\mathrm{n}}} \exp\left[eU_{\mathrm{BE}}/k_BT - 1\right] \approx I_{\mathrm{C}} \; , \tag{3.3.12}$$

d. h. wir haben eine typische Diodenkennlinie vorliegen, deren ‘Vorwärtsstrom’ $I_{\mathrm{C}}$ wieder bei $U_{\mathrm{BE}} \gtrsim 0,6$ V merklich einsetzt.
Die sog. Steilheit (small–signal transconductance) $g_{\mathrm{m}} = \frac{\partial I_{\mathrm{C}}}{\partial U_{\mathrm{BE}}}$ ist eine weitere Grösse zur Charakterisierung, siehe Abbildung 3.77 (Einfluss einer Steuerspannungsänderung auf den Ausgangsstrom). Die Ausgangskennlinien zeigen die Sperrströme der BC–Diode, die Höhe des Sperrstromes wird durch $U_{\mathrm{BE}}$ bzw. $I_{\mathrm{B}}$ gesteuert.
Es gibt drei verschiedene Grundschaltungen; der Kontakt, der auf Erdpotential gelegt wird, gibt der Schaltung ihren Namen: Basis–, Emitter– und Kollektorschaltung.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

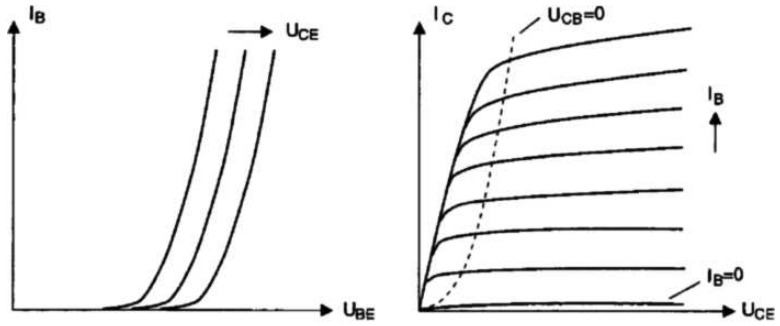



Abbildung 3.78.: Emitterschaltung: Eingangs– und Ausgangskennlinien[?].

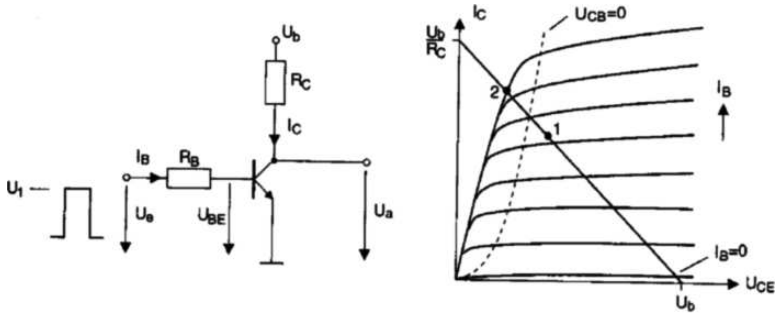



Abbildung 3.79.: Emitterschaltung: Bipolartransistor als Schalter[?].

Die *Basisschaltung* ist gekennzeichnet durch einen niederohmigen Eingang, einen hochohmigen Ausgang; ihre Stromverstärkung ist $\lesssim 1$, eine Spannungsverstärkung und damit Leistungsverstärkung ist möglich.
Die gebräuchlichste Grundschaltung ist die *Emitterschaltung*. Der Emitteranschlusss liegt auf Masse, die Spannungsversorgung des Transistors erfolgt über einen Arbeitswiderstand $R_\mathrm{A}$, der in der Praxis den Ausgangswiderstand der Schaltung bestimmt. Der ausgangsseitige Kollektorstrom $I_\mathrm{C}$ wird vom vergleichsweise kleineren Basisstrom $I_\mathrm{B}$ gesteuert.
Die Eingangskennlinie gibt das Verhalten der Injektionsdiode wieder; die Ausgangskennlinien zeigen das Sperrstromverhalten der BC–Diode bei $I_\mathrm{B} = 0$ bzw. stufenweise angehobenen Kennlinien mit wachsendem $I_\mathrm{B}$ bzw. $U_\mathrm{BE}$. Wird das Kennlinienfeld nach rechts fortgesetzt, folgt der Lawinendurchbruch in der BC–*Raumladungszone*. Weitere Eigenschaften sind: Ausgangsimpendanz von $10 - 100$ kΩ, nichtlineare Spannungsverstärkung von 20 - 100, sehr lineare Stromverstärkung $B$ von 50 - 500, typisch 100, grosse Steilheit, Durchbruch 2. Art (posiviter Temperaturkoeffizient bedeutet inhomogene Stromverteilungen, die zu lokaler Erhitzung und schliesslich zur lokalen Zerstörung insbesonders von Leistungs– und Hochspannungstransistoren führen).
Ein *Schalter* benötigt zwei klar unterscheidbare, verlustarme Zustände. Der Normalzustand (in Abbilding 3.79 mit 1 bezeichnet) zeigt erhebliche Verluste und wird bei der Verwendung des Transistors als Schalter nur als Zwischenzustand durchlaufen. Der Zustand ‘Zu’ oder ‘Aus’ wird erreicht, indem $I_B \approx 0$ gemacht wird. Dann sinkt $U_\mathrm{BE}$ unter die Durchlassspannung ($U_\mathrm{BE} < U_\mathrm{S} \approx 0,6$ V) und $I_\mathrm{C}$ sinkt auf einen minimalen Restbetrag ab. Der Widerstand des völlig gesperrten Transistors ist maximal und gross gegen den vorgeschalteten Kollektorwiderstand $R_\mathrm{C}$, d. h. die Versorgungsspannung $U_\mathrm{B}$ fällt praktisch am *Transistor* ab: $U_\mathrm{A} \approx U_\mathrm{B}$.

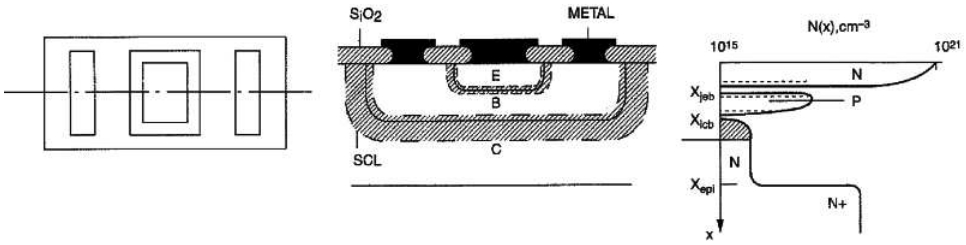



Abbildung 3.80.: Querschnitt und Dotierprofil des vertikalen npn–Bipolartransistors, schematisch[**?**].

Der Zustand ‘Auf’ oder ‘Ein’ wird durch sprungartiges Anlegen einer Spannung an den vorgeschalteten Basiswiderstand $R_B$ erreicht, so dass $I_B$ so gross wird, dass der Sättigungszustand (Punkt 2 in der Abbildung) erreicht wird. Dabei ist $U_{CE,Sättigung} \approx 0,2 - 0,5$ V, also kleiner als $U_{BE}$: D. h. $U_{CB}$ wird umgepolt und die (bislang immer gesperrte) B–C–Diode wird (erstmals) leitend.
Das zeitliche Schaltverhalten wird natürlich durch das Umladen der Sperrschichtkapazitäten bedingt. Es treten beim Ein– und Ausschalten Verzögerungszeiten auf. (Setzt man eine Schottkydiode so zwischen Basis und Kollektor, dass ihre Anode mit der Basis verbunden ist, lassen sich diese Verzögerungszeiten deutlich verringern: Schottky–TTL–Technologie).
Die Herstellung der Bipolartransistoren erfolgt wieder in Planartechnik.
Der *vertikale npn–Bipolartransistor* ist in Abbildung 3.80 gezeigt. Im Querschnitt ist mittig der Emitterkontakt zu sehen, daran schliessen beidseitig zwei Basiskontakte an. Die Breiten der Raumladungszonen stehen i. allg. im Verhältnis von etwa 1:2; die *BC–Raumladungszone* beträgt heute bei High–speed–Transistoren etwa 100 nm, bei Hochspannungstransistoren ca. 10 $\mu$m. Der n–dotierte Kollektor wurde auf ein $n^+$–dotiertes Substrat aufgewachsen. Für die Fertigung integrierter vertikaler npn–Transistoren muss der Kollektor ebenfalls an die Oberfläche geführt werden. (Der Emitter sitzt in einer p–dotierten Basiswanne, die im n–dotierten Kollektorgebiet eingelassen ist, das Substrat ist $p^-$–dotiert, die ohmschen Kontakte sind aus Polysilizium). Die Herstellung eines pnp–Transistors erfolgt analog, die Anordnung ist etwas einfacher: *lateraler pnp–Transistor*. Die Flächengrössen eines Transistors hängen vom Verwendungszweck ab. Für kleine Spannungen benötigt man heute $< 1$ $\mu$m, für einen 600 V–Schalttransistor beispielsweise $\varnothing > 10$ mm. Der Abstand der Raumladungszonen ist kritisch, eine weitere Verkleinerung der $I^2L$, der integrated injection logic, scheint technisch kaum möglich.

**HBT Hetero–Bipolartransistor (Heterjunction Bipolar Transistor)**
Beim konventionellen Bipolartransistor konnte die Leitfähigkeit der Basis zur Erzielung einer höheren Diffusionsgeschwindigkeit durch weitere Dotierung nicht gesteigert werden, weil der Rückstrom von Löchern von der Basis und dem Emitter dadurch zunahm und der Emitter–Injektionswirkungsgrad (durch Rekombination) absank.
Für HF– und Hochgeschwindigkeits–Schalt–Anwendungen kann man diese Grenze durch zwei Schritte hinausschieben.
Erstens kann man das n–dotierte Emittermaterial durch ein ebenfalls *n–dotiertes Halbleiter–Material* grösserer Bandlücke ersetzen. Die Valenzband–Diskontinuität am Heterokontakt vermindet das Eindringen der Löcher in das Emittergebiet sehr

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

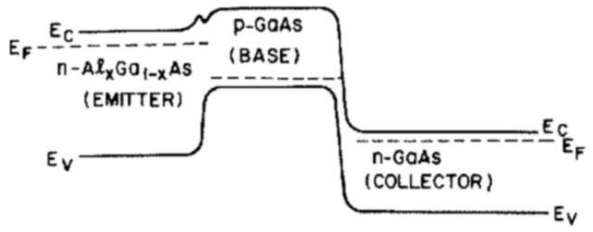



Abbildung 3.81.: AlGaAs–GaAs–Hetero–Bipolartransistor: Elektronisches Bandschema[?].

effektiv. Die Basis kann dadurch höher dotiert werden als der Emitter.
In einem weiteren, auf *Si–Wafern* aufgewachsenen System kann eine ähnliche Physik realisiert werden: *Si/SiGe–Bipolartransistor*. Hier hat die Basis eine kleinere *Energielücke* als das Emitter– und Kollektormaterial.
In diesem System ist auch der zweite Verbesserungsschritt realisiert worden, nämlich die '*Graded base band gap technique GBT*'. Hier besitzt die Basis zum einen wie in den Drifttransistoren eine inhomogene Dotierung und zusätzlich (neu) eine ortsabhängige Bandlücke, die ein hohes Driftfeld (15 kV/cm) hervorruft. Dies wird möglich, da mit der Variation des zulegierten Ge–Anteils in einem gewissen Rahmen die kontinuierliche Variation der *Energielücke* möglich ist. Die Elektronen brauchen zum Durchqueren der z. B. 50 nm breiten Basis noch weniger Zeit. Abschneidefrequenzen von rund 100 GHz sind so erreicht worden und die künftige Verbindung mit der ULSI (Ultra Large Scale Integration)–Technik eröffnet neue Horizonte.

## 3.3.5. Einige Optoelektronische Bauelemente

Bestrahlt man *Halbleiter–Material* mit Licht geeigneter Wellenlänge, so wird seine Leitfähigkeit (Dunkelleitfähigkeit) erhöht um die sog. *Fotoleitfähigkeit*. Dabei werden i. allg. die Ladungsträgerdichten erhöht, während die Ladungsträgerbeweglichkeiten praktisch unverändert bleiben.
Licht kann also in Fotostrom umgewandelt werden: *innerer Fotoeffekt* oder innerer lichtelektrischer Effekt. Auch die Umkehr ist möglich und dient der Erzeugung von inkohärentem oder kohärentem Licht. Die Anwendungsfelder sind riesig; sie umfassen die Sensorik, die Energietechnik, die kommerzielle Elektronik und die Telekommunikation etc. Glasfasernetze mit ultraniedrigen Verlusten verbinden die Kontinente; längst gibt es die integrierte Optoelektronik. Wir wollen uns hier aber auf die Grundlagen beschränken.
Aus den Absorptionsmessungen am intrinsischen *Halbleiter–Material* weiss man, dass die Photonen eine Mindestenergie benötigen, um ein Elektron–Loch–Paar über die Bandlücke anzuregen. (Im *Halbleiter* ist die optische Aktivierungsenergie gleich der thermischen.) Für die Grenzwellenlänge gilt:

$$\lambda_{\mathrm{Gr}}[\mu\mathrm{m}] = \frac{1,24}{E_{\mathrm{gap}}[\mathrm{eV}]} \; . \tag{3.3.13}$$

Dabei ist die Wahrscheinlichkeit für einen sog. indirekten Übergang (mit einer Änderung des Quasiimpulses des Elektrons durch ein Phonon) auch bei Raumtempe-

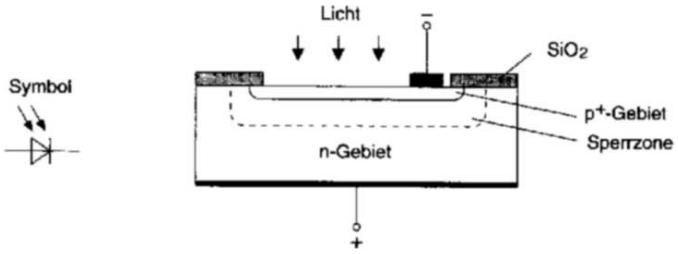



Abbildung 3.82.: Schema der $P^+NN^+$–Fotodiode[**?**].

ratur um Grössenordnungen kleiner als für einen direkten. Es liegt die *'Grundgitter–Fotoleitung'* vor.
Bei dotierten Halbleiter–Materialien werden die vorhandenen *Donatoren* oder *Akzeptoren* ionisiert; die sog. Ausläuferabsorption setzt also schon bei grösseren Wellenlängen ein und führt zur sog. '*Störstellen–Fotoleitung*'. Dieser Fotoleitungstyp ist unipolar, während die Grundgitter–Fotoleitung im Prinzip bipolar ist. (Ungleiche Beweglichkeiten und Lebensdauern können jedoch zu einer Unipolarität führen.)
Die aus dem thermischen Gleichgewicht durch Bestrahlung zusätzlich erzeugten Ladungsträger haben eine mittlere Lebensdauer, nach der sie wieder rekombinieren. Die Rekombination legt letzlich fest, ob zwischen Fotoleitfähigkeit und Beleuchtungsstärke ein linearer Zusammenhang besteht (oder etwa $\sim \sqrt{I}$).

#### 3.3.5.1. Halbleiter–Fotodetektoren

Der einfachste Fotodetektor ist ein beidseitig kontaktierter Halbleiter–Dickfilmstreifen; unter Beleuchtung erniedrigt sich sein Widerstand. Im sichtbaren Bereich langjährig bewährte Fotowiderstände sind CdS und CdSe. Sie sind hochohmig, sehr empfindlich und sehr langsam.
Aufwendiger sind *Fotodioden*. In das Absorptionsgebiet dicht unter der Oberfläche wird eine möglichst ausgedehnte *Raumladungszone* gelegt. Bei der klassischen *pn–Fotodiode* ist dies ein stark asymmetrisch dotierter pn–Übergang, bei der *pin–Fotodiode* ein ausgedehnter intrinsischer Zwischenbereich. Die Übergänge werden natürlich in Sperrrichtung betrieben, um die Raumladungszonen möglichst zu verbreitern, die Stärke des Driftfeldes ist von sekundärer Bedeutung. Für das Zeitverhalten sind aber neben den internen Laufzeiten (dotierkonzentrationsabhängig) die Umladezeiten der Sperrschichtkapazität massgebend. Wegen $C \sim 1/\sqrt{U}$ ist es sinnvoll, bei der maximalen Sperrspannung (ca. ... 10 V) zu arbeiten; die Schaltzeiten liegen dann bei ca. 10 ns. PIN–Dioden sind hier wegen ihrer kleineren Kapazität grundsätzlich im Vorteil, Schaltzeiten bis hinab zu 100 ps sind möglich. Die häufigst verwendeten Materialien sind *Si*, Ge und *GaAs*.
Prinzipiell sind alle Halbleiter–Kontakte mit ihren jeweiligen Raumladungszonen zur Detektion von Licht (oder härterer Strahlung) geeignet: die Schottkydiode, die Metall–I–N–Diode, die MOS–Diode. Die Heteroepitaxiemethoden ermöglichen neuartige Bauelemente, in denen Absorptionsgebiet und Speichergebiete räumlich getrennt sind.
Mit einer von aussen angelegten Sperrspannung detektiert man einen Fotostrom. Ohne Vorspannung fliesst bei Kurzschluss ein 'Kurzschluss–Strom' beziehungsweise erzeugt die Fotodiode bei offenen Kontakten eine Spannung. Abhängig von der

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

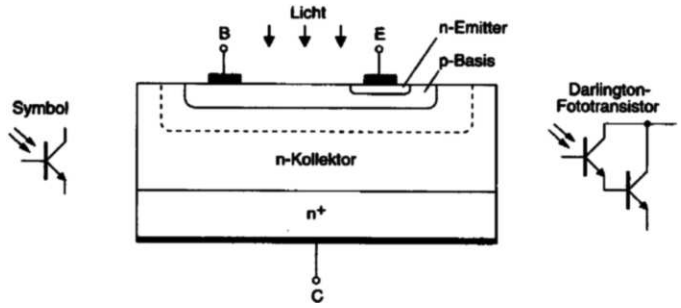



Abbildung 3.83.: Prinzip des Fototransistors[**?**].

Bauform spricht man von *Fotoelement* oder von der *Solarzelle*.
Die bisher genannten Bauelemente funktionieren ohne Verstärkung. Innere Verstärkung bis Faktoren von $10^4$ kann man in *Avalanche Fotodioden* (Lawinen–Fotodioden) erreichen. Durch geeignete Dotierung und hohe Sperrspannungen werden in diesem Element Ladungsträger bewusst durch lokale Stossionisation und weniger durch Beleuchtung erzeugt, im Driftfeld beschleunigt und so ein Lawinendurchbruch gezündet. Meist weist die Diode eine $P^+P^-PN^+$–Struktur auf, es kann aber auch ein Metall-Halbleiter–Kontakt oder eine PIN–Struktur, etc. sein. Die Anstiegszeiten können unter 100 ps liegen.
Eine sehr effektive, lineare Möglichkeit zur Verstärkung bieten die *Fototransistoren*. Beim *bipolaren Transistor* ist die gesperrte Basis–Kollektor–Diode das Absorptionsgebiet. Die Löcher erhöhen die Basis–Emitter–Spannung, so dass der Emitter mehr Elektronen in die Basis und damit in den Kollektor injiziert. Die Bipolar–Fototransistoren sind allerdings langsam (ca. 100 kHz). Beim Sperrschicht–Feldeffekt–Fototransistor und beim MOSFET–Fototransistor werden bevorzugt die breiten Sperrzonenbereiche beleuchtet. Hervorzuheben ist beim ersten Typ die Rauscharmut, beim zweiten die kurzen Ansprechzeiten.
*Solarzellen* sind besondere Bauelemente; kein kommerzielles Detektorelement wurde so sehr auf Effizienz, Grossflächigkeit und Preiswertheit optimiert. Sie gibt es in einer Vielzahl von Bauformen, meist aus *Si*. Einkristalline Solarzellen haben Wirkungsgrade von über 26 % erreicht.

### 3.3.5.2. Halbleiter–Strahlungsquellen

pn–Dioden aus Halbleitern mit direkter Bandlücke emittieren Strahlung, wenn sie in Durchlassrichtung betrieben werden: LED (Light emitting diode). Die Ursache ist die sog. strahlende direkte Rekombination über die Bandlücke hinweg. *GaAs* selbst emittiert im Infratoren, im Sichtbaren werden $GaAs_{1-x}P_x$–Materialien und GaP:N–Materialien eingesetzt. Neuerdings spielt GaN eine bedeutende Rolle. Das Emissionsspektrum ist i. allg. sehr breit und temperaturabhängig. Die Richtcharakteristik ist ausgesprochen breit und wird in der Praxis z. B. durch Kunststofflinsen in Vorwärtsrichtung verbessert. Die Schaltzeiten können 1 $\mu$s deutlich unterschreiten.
Die strahlende Rekombination kann auch über einen Zwischenzustand (Lumineszenz–Zentrum) erfolgen. Bekanntes Beispiel sind die blauen SiC–LEDs (Siemens). Die neuen blauen LEDs bestehen aus AlGaN/InGaN–Doppelheterostrukturen. Werden sie mit YAG (Yttrium Aluminium Garnet) und Phosphoren direkt beschich-

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

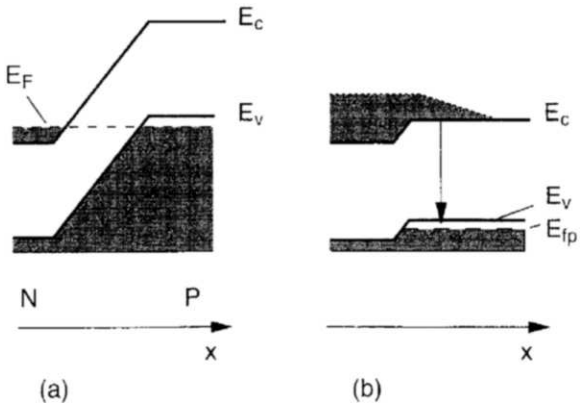



Abbildung 3.84.: $P^+N^+$–Laserdiode a) ohne Spannung und b) mit angelegter Spannung und dadurch hervorgerufener Besetzungsinversion.[?]

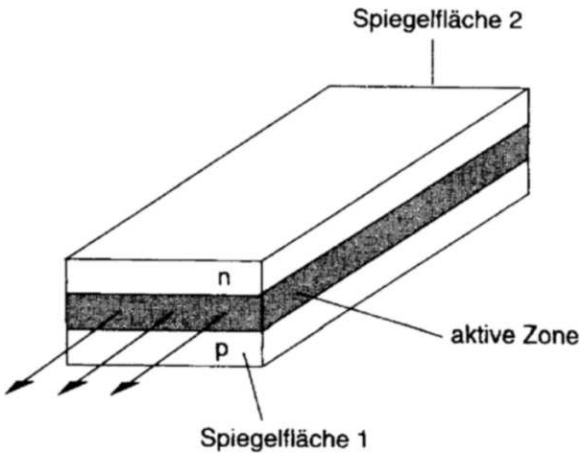



Abbildung 3.85.: Prinzip des Laserresonators[?]. Eine ausführliche Darstellung findet sich im Abschnitt 4.5.1

tet, entsteht eine *weisse LED*. Ein weiterer Trend geht zu grossflächigen LEDs, einzelne Emitterflächen reichen an 1 $mm^2$.
Das LED–Prinzip lässt sich zum *Halbleiter–Laser* weiterentwickeln. Dazu sind zwei Dinge notwendig. Erstens muss die induzierte Emission die bei der LED ausschliesslich vorhandene spontane Emission deulich übertreffen. Hierzu ist in der sog. aktiven Zone eine ausreichende Besetzungsinversion notwendig. In einer beidseitig sehr hoch dotierten ($> 10^{19}\ cm^{-3}$) entarteten pn–Diode, die in Durchlassrichtung betrieben wird, ist die Ladungsträgerinjektion tatsächlich ausreichend gross, um gepulsten Laserbetrieb zu erhalten. Vorausgesetzt, die zweite Bedingung ist erfüllt: die gesamten Verluste der Strahlungsmode müssen kleiner sein als ihr Gewinn. Erreicht wird dies durch einen länglichen (ca. 1 mm) Resonator. Man erhält ihn durch Brechen entlang einer niederinduzierten Kristallebene ((110) in *GaAs*), die Seitenflächen werden aufgeraut. Dieser sog. Kantenstrahler emittiert an beiden Enden.
Die aktive Zone ist mehrere $\mu$m hoch und seitlich noch unbegrenzt; auch die Strompfade sind noch undefiniert. Deshalb ist die sog. Schwellstromdichte noch sehr hoch, die Verlustwärme zerstört die Laserdiode rasch.
Viel besser wäre es, wenn der optische Resonator durch einen Wellenleiter seitlich

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

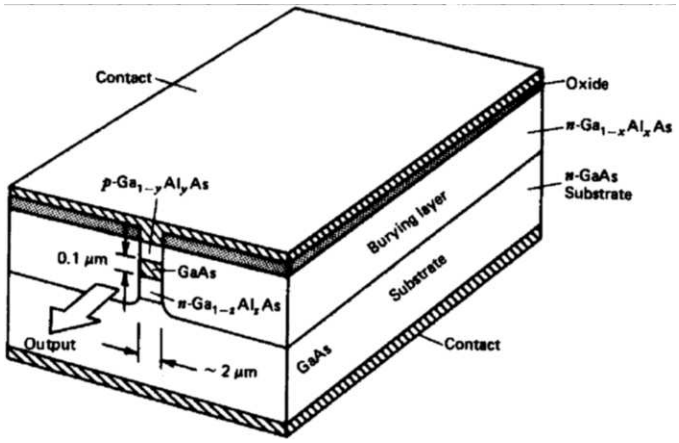



Abbildung 3.86.: Quantum Well Lasers.

auf seine Grundmode eingeschränkt würde; das *Halbleiter–Material* der aktiven Zone müsste also einen deutlich höheren Brechungsindex haben als das sie umgebende Material: 'optisches Confinement'. Auch der Diodenstrom müsste nur durch die aktive Zone und auf den Zuleitungswegen möglichst niederohmiges Material durchfliessen müssen: 'Elektrisches Confinement'. Die mässige spektrale Bandbreite, bedingt durch die energetische Breite der besetzten Zustände ($h\nu > E_{\text{gap}}$ !) und die gebrochenen planen Endflächen sollten durch einen wellenlängenselektiven hochreflektierenden Spiegel (z. B. DFB Distributed feedback) ersetzt werden.

Technologisch haben die *Halbleiter* diesen weiten Weg über viele Jahre mit schrittweisen Verbesserungen zurückgelegt. Hier können nur wenige genannt werden: der Heterojunction Laser, der Doppel–Heterojunction Laser mit dem optischen Confinement in der Senkrechten und einer senkrechten Ausdehnung der aktiven Zone von ca. 200 nm durch ein elektrisches Confinement mit Hilfe der Banddiskontinuitäten.

Abbildung 3.86 zeigt eine weitere verbesserte Variante dieses Typs. Der Wellenleiter ist zusätzlich seitlich begrenzt; der Zuleitungskontakt ist durch eine Oxidmaske streifenförmig definiert. Mit solchen Lasern sind ca. 10 mW im Dauerbetrieb bei Raumtemperatur möglich.

Noch niedrigere Schwellströme erreicht man, wenn die aktive Zone durch einen *Single Quantum Well* oder gar ein *Multiquantum Well* gebildet wird. Die hochdefinierte Art und die Höhe der beteiligten Elektronen– und Löcherzustandsdichten ermöglicht unbekannte Verstärkungen (pro aktive Zonenlänge). Mit diesem Konzept konnte z. B. im ZnSe–Materialsystem der erste blaue Dauerstrichlaser bei Raumtemperaur realisiert werden (Nakayama, 1993).

Durch Verwendung mehrerer Quantum Wells und Spiegeln aus Braggreflektoren gelang es schliesslich, VCSELs (Vertical–cavity surface–emitting lasers) zu realisieren. Zum Abschluss sei ein Beispiel der Universität Ulm (Arbeitsgruppe Ebeling) hierzu gezeigt. Vier 6 nm dicke GaAs Quantentöpfe sind zwischen $Al_{0,3}Ga_{0,7}$–Barrieren eingebettet. Der Laser emittiert einen zylindersymmetrischen Strahl nach oben (top emitting).

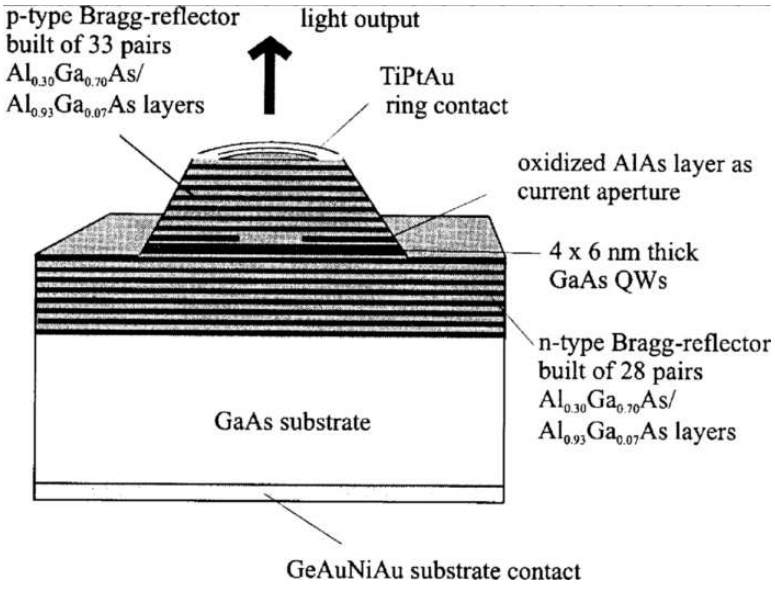



Abbildung 3.87.: *GaAs–VCSEL* für 750 nm, Arbeitsgruppe Ebeling, Universität Ulm.

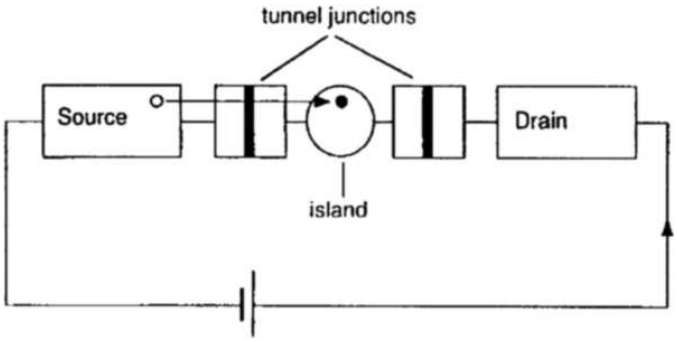



Abbildung 3.88.: Coulomb–Blockade und Single electron electronics.

## 3.3.6. Ausblick

Bedingt durch die Dominanz der Silizium–CMOS–Technologie bei den integrierten Schaltungen muss ein Resümee mit einem Blick auf den *MOS–Transistor* beginnen. Mit jedem Technologieschritt werden seine Abmessungen kleiner und sein Gateoxid immer dünner. Irgendwann zwischen 2010 und 2020, so sagt dies die 'technology roadmap' der SIA (Semiconductor Industries Association/USA) voraus, wird die CMOS–Technologie aus physikalischen Gründen nicht mehr anwendbar sein. (Ein Schaltvorgang wird aus Dimensionsgründen nur noch von ca. 10 Elektronen getragen werden. Die Wellenfunktionen des Halbleiters und besonders des Metalls lecken in das Oxid hinaus, so dass mindestens 4 – 5 Oxidmonolagen nötig sind, damit sich Isolatorbandzustände überhaupt ausbilden können.)
Eine ca. 1 nm dicke Gate–Oxidschicht kann von einzelnen Elektronen durchtunnelt werden: Tunneloxid. Darauf beruht die 'Single electron electronics', die auf einer Grössenskala $< 10$ nm arbeitet.
In der Grundanordnung befindet sich zwischen 2 Elektroden, jeweils durch Tunnelstrecken getrennt, eine Nanoinsel aus leitendem Material, z. B. ein metallener 'nanodot'. Die Inselkapazität soll möglichst klein sein (ca. $10^{-19} - 10^{-18}$ F), d. h.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

der Partikel darf nur einen Durchmesser von wenigen Nanometer haben.
Bringt man durch einen Spannungsimpuls ein Elektron über die Source–seitige Tunnelstrecke auf die Nanoinsel, so verändert die Ladung des Elektrons die Tunnelverhältnisse für weitere Elektronen; diese sehen eine zusätzliche 'Barriere' und ein weiteres Elektron muss zusätzlich die Energie $e^2/2C$ mitbringen, um ebenfalls auf die Nanoinsel tunneln zu können: *Coulomb–Barriere*. Für Spannungen $|V_C| > e^2/2C$ können Elektronen in beide Richtungen transportiert werden, was man durch Messen der $I(U)$–Kennlinie jeweils überprüfen kann.
Nimmt man eine dritte Elektrode (Gate) hinzu, so kommt man zum *SET* (*Single electron transistor*). Es gibt erste Beispiele, die bei Raumtemperatur funktionieren ($e^2/2C \gg k_B T$) (Siehe auch Abschnitt 4.10.10) .
Zum Abschluss sei noch auf eine Merkwürdigkeit hingewiesen. Bei allen elektronischen Bauelementen spielte der Elektronenspin überhaupt keine Rolle. Dies soll sich mit der 'Spin–Elektronik' ändern. Allerdings stehen die Konzepte noch auf dem Papier (z. B. eine neue Schaltungslogik), die erfolgreiche experimentelle Realisierung steht noch aus. Packen wir's an!

## 3.4. Grundschaltungen

### 3.4.1. Lineare passive Bauelemente

Passive Bauelemente besitzen keine eingebaute Leistungsquelle; ihre Ausgangsleistung kann also niemals grösser als ihre Eingangsleistung sein. (Wohl aber kann die Ausgangsspannung grösser als die Eingangsspannung sein, siehe Transformator.)
Passive Bauelemente sind stets zweipolig. Lineare Zweipole verknüpfen ein Eingangssignal linear mit dem zugehörigen Ausgangssignal. Dem wichtigsten nichtlinearen Zweipol — der Diode — widmen wir einen eigenen Abschnitt im Anschluss.
In einem Gleichstromkreis gibt es neben Spannungs- und Stromquellen nur Widerstände ($> 0\ \Omega$ bis $< \infty\ \Omega$); Kondensatoren und Spulen sind erst in Wechselstromkreisen wirksam. Im erweiterten Sinn sind auch Schalter, Relais, Verbinder, etc. sowie *Messgeräte* Zweipole.
Zur Erinnerung: Zum Repertoire zur Beschreibung von Zweipolschaltungen gehören das ohmsche Gesetz, die Begriffe Spannung, Strom, Leistung, die Knoten– und Maschenregel (Kirchhoffschen Gesetze), die Formeln für Serien– und Parallelschaltungen von Widerständen und für den Spannungsteiler.
Die Untersuchung eines Netzwerkes aus Widerständen und Spannungsquellen kann sehr mühselig sein. Aber sie ist beispielsweise notwendig, wenn es um die 'Belastung' einer Schaltung geht. Ein einfaches Beispiel hierfür ist der belastete Spannungsteiler, siehe Abbildung 3.90.

#### 3.4.1.1. Theoreme von Thévenin und Norton

Zwei Theoreme helfen dabei:
Das Theorem nach Thévenin sagt aus, dass jedes Netzwerk aus Zweipolen, z. B. aus Spannungsquellen und Widerständen, äquivalent durch eine Ersatzschaltung aus einer (idealen) *Spannungsquelle* mit $U_{\mathrm{Th}}$ und einem (inneren) Widerstand $R_{\mathrm{Th}}$ in Reihe beschreibbar ist.

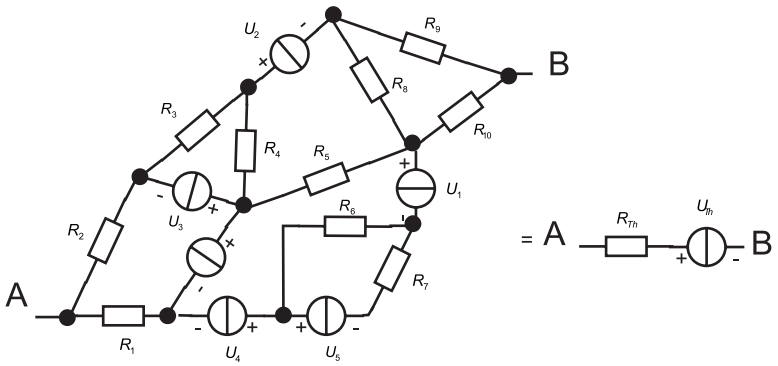



Abbildung 3.89.: Théveninsche Ersatzschaltung (gezeichnet nach [**?**]).

Zur Berechnung der Théveninschen Ersatzschaltung geht man folgendermassen vor:

1. Eine eventuelle Last zwischen $A$ und $B$ wird entfernt.
2. Man misst die Spannung $U_{Th}$ zwischen $A$ und $B$ mit einem idealen Spannungsmesser ($R_i = \infty$).
3. Man ersetzt alle Spannungsquellen durch Kurzschlüsse, behält aber deren Innenwiderstände (die ja in Serie zur Spannungsquelle sind).
4. Man ersetzt alle Stromquellen durch Unterbrüche, behält aber die Innenwiderstände (die ja parallel zur Stromquelle sind).
5. Man misst oder berechnet den Widerstand $R_{Th}$ zwischen $A$ und $B$.

Das Theorem von Norton besagt das gleiche für eine Ersatzschaltung aus einer (idealen) Stromquelle und einem parallelen Widerstand. Hier geht man wie folgt vor:

1. Eine eventuelle Last zwischen $A$ und $B$ wird entfernt.
2. Man misst den Kurzschlussstrom $I_N$ zwischen $A$ und $B$ mit einem idealen Strommesser ($R_i = 0$).
3. Man ersetzt alle Spannungsquellen durch Kurzschlüsse, behält aber deren Innenwiderstände (die ja in Serie zur Spannungsquelle sind).
4. Man ersetzt alle Stromquellen durch Unterbrüche, behält aber die Innenwiderstände (die ja parallel zur Stromquelle sind).
5. Man misst oder berechnet den Widerstand $R_N$ zwischen $A$ und $B$.

$U_{\mathrm{Th}}$ ist die Leerlaufspannung sowohl der Ersatzschaltung als auch des äquivalenten Schaltnetzwerks und kann berechnet oder gemessen werden. Der Kurzschlussstrom im geschlossenen Ersatzkreis ist gleich dem Kurzschlussstrom im äquivalenten

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

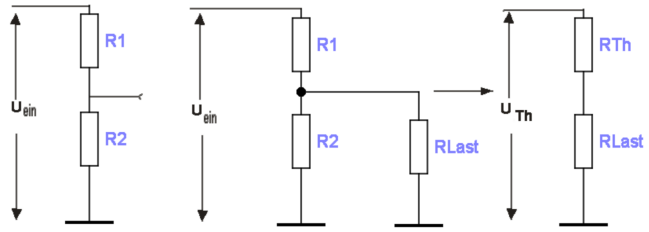



Abbildung 3.90.: Unbelasteter und belasteter Spannungsteiler nebst seiner Théveninschen Ersatzschaltung[**?**].

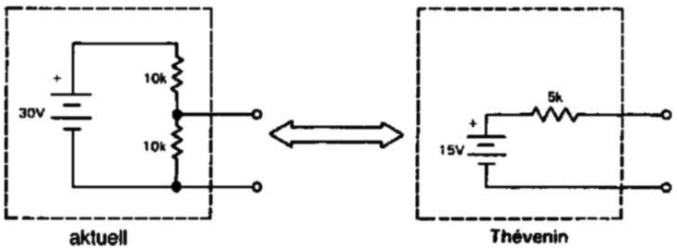



Abbildung 3.91.: Beispiel für die Ersatzschaltung eines Spannungsteilers[**?**].

Netzwerk, also

$$I = \frac{U_{\mathrm{Th}}}{R_{\mathrm{Th}}} \qquad \text{‘Kurzschlussstrom’,}$$

$$U_{\mathrm{Th}} = U(\text{offener Kreis}) \qquad \text{‘Leerlaufspannung’ am Ausgang,}$$

$$R_{\mathrm{Th}} = \frac{U(\text{offener Kreis})}{I(\text{geschlossener Kreis})} \qquad \text{‘Innenwiderstand’,}$$

$$U_{\mathrm{L}} = U_{Th}\frac{R_L}{R_{Th}+R_L} \qquad \text{‘Spannung am Lastwiderstand’.}$$

Übungsaufgabe: Bestimmung von $U_{\mathrm{Th}}$ und $R_{\mathrm{Th}}$ für den oben gezeigten Spannungsteiler. Die belastete Ersatzschaltung ist wieder ein Spannungsteiler!
Das Bild von einem äquivalenten Innenwiderstand ist nicht nur auf die *Spannungsquelle* ‘Spannungsteiler’ beschränkt, sondern wird ebenso auf Batterien, Oszillatoren, Verstärker und Sensoren angewand. Die Ausdrücke Quellenwiderstand, Innenwiderstand, Ausgangswiderstand und Théveninscher Ersatzwiderstand meinen alle dasselbe.
Müssen Kondensatoren und Spulen berücksichtigt werden, so gilt das Thévenin–Theorem in verallgemeinerter Form: Jedes Zweipol–Netzwerk aus Widerständen, Kondensatoren, Spulen und Signalquellen ist äquivalent einer einzigen komplexen Impedanz in Serie mit einer einzigen Signalquelle. Die Ersatzspannungsquelle und die Ausgangsimpedanz werden wie oben ermittelt.
Übungsaufgabe: Berechnen Sie $U_{\mathrm{Th}}$ und $R_{\mathrm{Th}}$ für den Fall, dass die Widerstände 100 kΩ, 1 MΩ, 100 MΩ betragen.

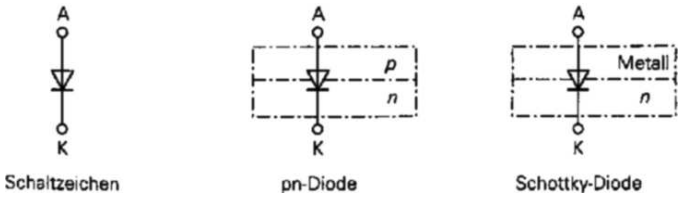



Abbildung 3.92.: Diode: Schaltzeichen und Aufbau[**?**].

Eine wichtige Anwendung des Théveninschen Modells ist die Vorhersage der Ausgangsspannung einer *Spannungsquelle* (z. B. der genannte Spannungsteiler) oder einer ganzen Schaltungsgruppe unter Last. Noch allgemeiner: Wenn ein Schaltkreis A einen Schaltkreis B ansteuert, — also mit ihm belastet wird — in welchem Verhältnis muss der Ausgangswiderstand von A zum Eingangswiderstand von B stehen, damit der Spannungseinbruch von A einen gewünschten Prozentsatz nicht übersteigt? Die Betrachtung folgt immer dem selben Weg: erst ermittelt man die beiden Elemente $U_{\mathrm{Th}}$ und $R_{\mathrm{Th}}$ der Ersatzschaltung für Schaltkreis A, dann betrachtet man den Spannungsteiler unter Berücksichtigung des Eingangswiderstands des Schaltkreises B als Last. Sukzessive können so weitere Lasten angehängt werden.
Für die Praxis gilt folgende Merkregel: Um den Spannungseinbruch einer Quelle klein zu halten, sollte der Lastwiderstand möglichst gross gehalten werden. Es sollte gelten:

$$R_{\mathrm{Last}} \gtrsim 10 \cdot R_{\mathrm{Innen}} \ . \tag{3.4.1}$$

Bei Anwendungen in der Hochfrequenztechnik ('reflexionsfreie Anpassung') oder zur optimalen Leistungsanpassung gilt aber stets:

$$R_{\mathrm{Last}} \stackrel{!}{=} R_{\mathrm{Innen}} \ . \tag{3.4.2}$$

## 3.4.2. Dioden

Die physikalischen Grundlagen der Dioden wurden bereits in Kapitel 3.2 ausführlich behandelt. Es folgen hier einige praxisbezogene Bemerkungen.
Diskrete Dioden haben stets zwei Anschlüsse: Anode (A) und Kathode (K); integrierte Dioden besitzen noch einen dritten Anschluss (Substratanschluss S). Die Gehäuseformen von Einzeldioden ähneln denen von Widerständen, der Farbring kennzeichnet die Kathode. Weitere Formen, insbesondere Leistungsdioden.
Die in Abbildung 3.94 gezeigten Kennlinien geben nochmals die drei Betriebsbereiche wieder: Durchlass–, Sperr– und Durchbruchsbereich. Im Durchlass beträgt die sog. Flussspannung $U_{\mathrm{F}}$ (forward voltage) bei typischen Betriebsströmen bei Ge– und Schottkydioden ca. $0,3 - 0,4$ V, bei *Si–Dioden* ca. $0,6 - 0,7$ V. Dieser 'Durchlass–Spannungsabfall' ist annähernd konstant; ein Fakt, der die Diskussion von Schaltungen mit Dioden ganz wesentlich vereinfacht. Für eine erste Betrachtung kann er häufig ganz vernachlässigt werden. Für hohe Diodenströme $I_{\mathrm{D}}$ sind die Durchlasswiderstände sehr klein: $R_{\mathrm{D}} \approx 0,01 \ - \ 10\ \Omega$.
Umgekehrt sind die Sperrwiderstände $R_{\mathrm{S}}$ sehr hoch, bei *Si* i. allg. $> 10\ \mathrm{M}\Omega$. Die zugehörigen Sperrströme $I_{\mathrm{R}}$ (reverse current) sind extrem klein, meist $< 10^{-7}$ A. Schliesslich wird die Sperr–Durchbruchspannung $U_{\mathrm{BR}}$ (peak inverse voltage) erreicht, die meist bei einigen 10 V liegt, aber auch 1 kV bei Stromgleichrichterdioden betragen kann. Ausser bei Zenerdioden sollte diese Sperrspannung keinesfalls

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

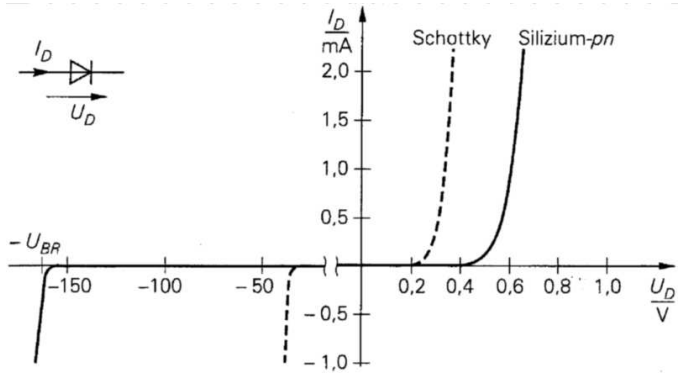



Abbildung 3.93.: Kleinsignal–Diode: Strom–Spannungs–Kennlinien[?].

erreicht werden. In einer ersten Schaltungsbetrachtung haben die Dioden in Sperrrichtung einen $\infty$–grossen Widerstand.

Das *dynamische Verhalten* von Dioden wurde bereits in Kapitel 3.2 prinzipiell diskutiert. In der Praxis spielt das Schaltverhalten bei ohmscher bzw. ohmsch–induktiver Last und bei höheren Frequenzen eine entscheidende Rolle. In Durchlassrichtung stellen sich konstante Betriebsbedingungen meist in wenigen Nanosekunden ein, in Sperrrichtung werden ähnliche Werte für die Abfallzeiten nur bei Schottkydioden mit kleineren Kapazitäten erreicht, Siliziumdioden wie die 1 N 4148 Kleinsignaldiode benötigen ca. 100 $\mu$s, bei Stromgleichrichterdioden werden mehrere $\mu$s benötigt. Weitere Informationen.

CAD–Programme zur Schaltungssimulation wie z. B. PSPICE verwenden nichtlineare *Diodenmodelle* mit den Grössen für den Diffusions–, Rekombinations– und Durchbruchstrom, für den Bahnwiderstand und für die Sperrschicht– und Diffusionskapazitäten. Zahlreiche Parameter charakterisieren das statische, dynamische und thermische Verhalten. Muss man auf die Hilfe von Computern verzichten, helfen linearisierte Kleinsignalmodelle, die im Wesentlichen den differentiellen Diodenwiderstand und die beiden Kapazitäten in einem Arbeitspunkt berücksichtigen.

Es gibt eine Vielzahl von Dioden–*Anwendungen*. Einer der wichtigsten ist die *Gleichrichtung* von Wechselspannungen, d. h. der periodische Wechsel von Durchlass– und Sperrbetrieb ('Gleichrichterdiode'). Am einfachsten geschieht dies in Spannungsteilern; Abbildung 3.94 zeigt neben dem bekannten Einweg– oder Halbwellengleichrichter zwei verwandte Schaltungen, nämlich sog. Klemmen.

Am Eingang wird jeweils eine (ausreichend niederfrequente) Wechselspannung von $V_{\mathrm{P-P}} = 20$ V angelegt. Am Ausgang des Einweggleichrichters beobachtet man die von der Diode durchgelassene, positive Halbwelle, reduziert um $U_{\mathrm{F}} \approx 0,6$ V. Bei der linken Klemme sperrt die Diode bei der positiven Halbwelle, d. h. weil der Diodenwiderstand $R_{\mathrm{S}} \gg R$ ist, fällt fast die gesamte Spannung am Ausgang an. In der negativen Halbwelle ist die Diode duchlässig, jetzt ist $R \gg R_{\mathrm{D}}$; am Ausgang liegt praktisch $U_{\mathrm{F}} \approx 0,6$ V an. Die Klemme 'klemmt' die negative Ausgangsspannung auf $-0,6$ V. Die Klemme rechts ist erweitert worden um eine vorgespannte zweite Diode. Solange die positive Halbwelle Werte von $U < 5$ V annimmt, sperrt die Diode 2. Für $U > 5,6$ V ist die Diode 2 sehr niederohmig, die Diode 1 aber extrem hochohmig; die zusätzliche Gleichspannung klemmt den maximalen positiven Spannungswert am Ausgang auf 5,6 V.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

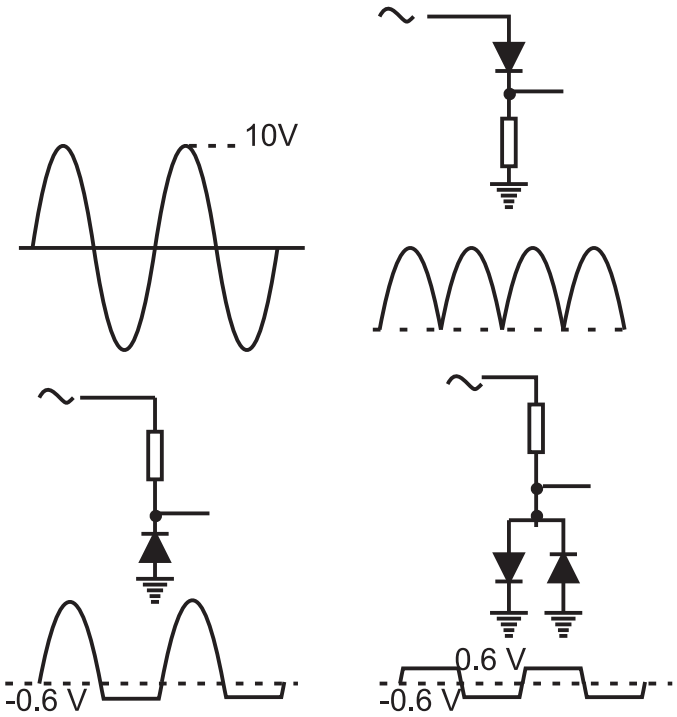



Abbildung 3.94.: Einweggleichrichter oben und zwei Klemmen unten (gezeichnet nach [**?**]).

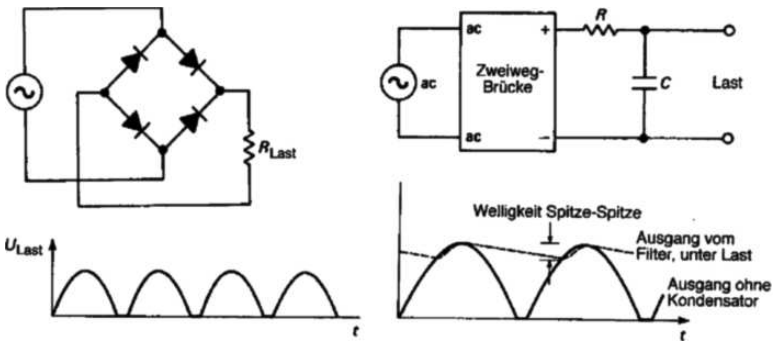



Abbildung 3.95.: Zweiweg–Brückengleichrichter links ohne und rechts mit Glättungs–Kondensator[**?**].

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

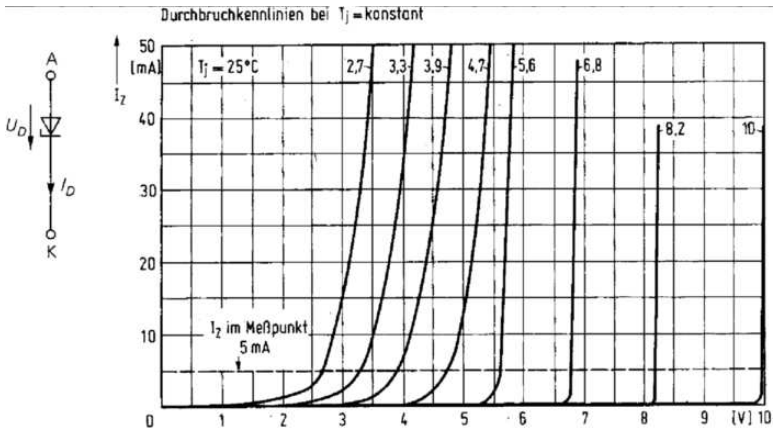



Abbildung 3.96.: Schaltsymbol und Durchbruchkennlinien[?].

Vollwellen– oder Zweiweg–Brückengleichrichter nützen beide Halbwellen aus, siehe Bild 3.95. Allerdings fällt pro Halbwelle an zwei Dioden $U_F$ ab. Ein erster Schritt zum Einsatz als Netzgerät ist die Hinzunahme eines Glättungskondensator mit $R_{\text{Last}}C \gg 1/f$, wobei $f$ die Brummfrequenz, also die doppelte Netzfrequenz meint. Die Brücke ist heutzutage ein kleines, integriertes Bauteil. Damit die Restwelligkeit kleiner und der Kondensator nicht zu gross und teuer werden muss, lässt man der obigen Anordnung ein weiteres integriertes Bauteil, einen sog. Spannungsregler folgen. Diese benutzen eine aktive Rückkopplungsschaltung; ein Prinzip, das wir bei den Operationsverstärkern erstmals kennenlernen werden.
Übungsaufgabe: Die Stromkennlinie $I_{\text{aus}}(I_{\text{ein}})$ zeigt einen linearen, symmetrischen Verlauf. Tut dies die Spannungskennlinie $U_{\text{aus}}(U_{\text{ein}})$ auch?
Die bisher besprochenen Anwendungen betrafen Kleinsignaldioden bzw. zumeist Netzgleichrichterdioden. Es gibt noch eine ganze Reihe spezieller Dioden, die für spezielle Anwendungen optimiert werden.
*Zenerdioden* werden im Sperrbereich betrieben; genauer, man nützt ihr Verhalten im Durchbruch. Abbildung 3.96 mit typischen Durchbruchkennlinien zeigt, dass Dioden mit kleineren Durchbruchspannungen noch nicht ganz ideal schalten. Die Anwendung im Spannungsteiler ist die *Spannungsbegrenzung* bzw. bei belastetem Spannungsteiler die *Spannungsstabilisierung*.
*Kapazitätsdioden* (Abstimmdioden, varicap) werden ebenfalls in Sperrrichtung betrieben. Man nützt die Spannungsabhängigkeit der Sperrschichtkapazität aus. Durch sog. hyperabrupte Dotierung (inhomogene, zur Grenzschicht hin ansteigende Konzentration) erreicht man Kapazitätskoeffizienten bis 1 und Kapazitätsvariationen bis Faktor 10. Die Anwendung liegt insbesondere in der Frequenzabstimmung von LC–Kreisen (z. B. im Radio) in Wobbelsendern oder zur Frequenzmodulation; hierzu sind meist 2 oder 3 Abstimmdioden in einem Gehäuse untergebracht.
Auf den negativen Widerstand von *Esaki–Tunnel–Dioden* wurde bereits hingewiesen. Sie finden in GHz–Oszillatoren und schnellen Schaltern ihre Verwendung.
*Schaltdioden* sind durch kleine Sperrkapazitäten und sehr steile Durchlasskennlinien ausgezeichnet. Sie werden als elektronische Schalter zum Ersatz von mechanischen Schaltern eingesetzt.
*Schottky–Dioden* (Mikrowellendioden, hot carrier diodes) haben kleine $U_F \approx 0,4$ V und können als Mikrowellendiode sehr kleine Kapazitäten aufweisen. Sie werden für Frequenzen $> 15$ GHz als Mischer und Detektoren eingesetzt, sowie als sehr

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

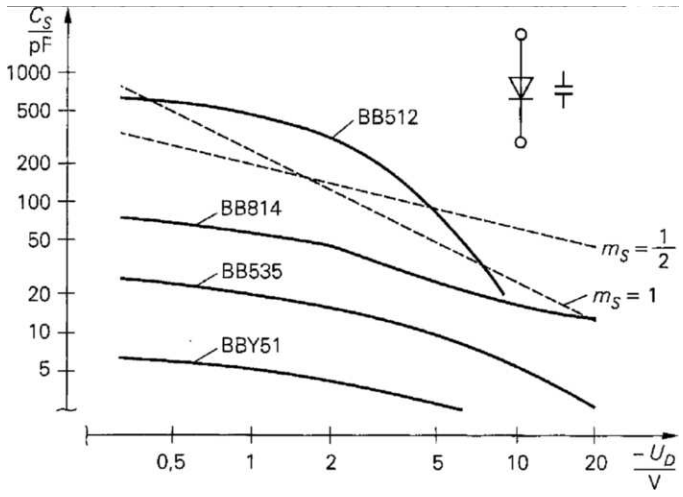



Abbildung 3.97.: Schaltzeichen und spannungsabhängige Kapazitäten an typischen Kapazitätsdioden[**?**].

schneller Schalter. Vergleiche Datenblätter und Praktikum (Ringmischer).

Als *PIN–Dioden* (current–controlled RF–resistor) bezeichnet man Dioden mit geringer Störstellendichte und deshalb hoher Lebensdauer der Ladungsträger in der intrinsischen Schicht. Solche Dioden sperren zuverlässig nur bei relativ kleinen Frequenzen. Bei ca. $f > 10$ MHz kann man die Dioden als gleichstromgesteuerten Wechselspannungswiderstand einsetzen, z. B. in HF–Dämpfungsgliedern.

## 3.4.3. Bipolartransistoren

Transistoren sind die wichtigsten aktiven Bauelemente. Ihre Ausgangsleistung kann grösser sein als ihre Eingangsleistung: Transistoren verstärken und können deshalb (durch Rückkopplung) der aktive Teil eines Oszillators sein. Die zusätzliche Leistung liefert eine 'externe' Quelle.

Transistoren sind praktisch in jeder elektronischen Schaltung enthalten; in IC's findet man z. B. ca. 20 Transistoren in einer einfachen Operationsverstärkerschaltung oder in höchstintegrierten DRAM's, Mikroprozessoren oder Logigschaltungen heute weit über 100 Millionen. Entsprechend ihrer Beschaltung werden, wie in den vorigen Abschnitten gezeigt, eine Vielzahl von Bauformen entwickelt.

Im Folgenden betrachten wir Einzeltransistoren und ihre Schaltungen.

Allen Bipolartransistoren sind einige für die praktische Anwendung relevante Eigenschaften gemeinsam. Meist werden sie im Normalbetrieb (forward region) eingesetzt. Das heisst, die Basis–Emitter–Diode wird in Durchlassrichtung und die Basis–Kollektor–Diode in Sperrrichtung betrieben. Dies gilt für npn– und pnp–Transistoren gleichermassen, mit vertauschten Vorzeichen für Ströme und Spannungen.

Das Verhalten des jeweiligen Transistortyps beschreiben die sog. Kennlinienfelder, die in den zugehörigen Datenblättern gezeigt werden. Die Kennlinien sind immer

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

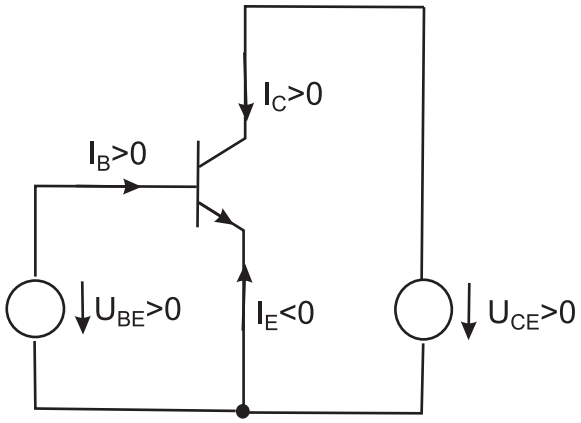



Abbildung 3.98.: Spannungen und Ströme eines npn–Transistors in Normalbetrieb[**?**].

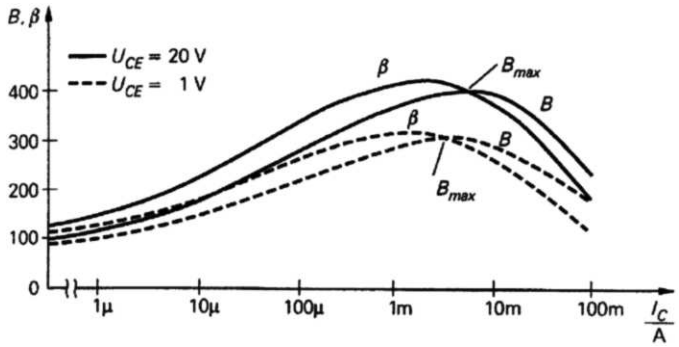



Abbildung 3.99.: Gross– und Kleinsignalverstärkung $B$ und $\beta$ eines Kleinleistungstransistors im Normalbetrieb[**?**].

nichtlinear oder nur abschnittsweise linear. Es gelten die Gleichungen

$$I_\mathrm{C} = I_\mathrm{S} \cdot \mathrm{e}^{\frac{U_\mathrm{BE}}{U_\mathrm{T}}} \left(1 + \frac{U_\mathrm{CE}}{U_\mathrm{A}}\right) \quad \text{und} \tag{3.4.3}$$

$$I_\mathrm{B} = \frac{1}{B} I_\mathrm{C} \quad \text{mit } B = B(U_\mathrm{BE}, U_\mathrm{CE}) \tag{3.4.4}$$

$I_\mathrm{S} \approx 10^{-16} - 10^{-12}$ A ist der Sättigungssperrstrom, $U_\mathrm{T} \approx 26$ mV die Temperaturspannung bei Raumtemperatur, $U_\mathrm{A}$ die im Ausgangskennlinienfeld ($I_\mathrm{C}(U_\mathrm{CE})$) extrapolierte Early–Spannung und $B$ die Stromverstärkung. Je grösser $U_\mathrm{A}$ bzw. je flacher die Ausgangskennlinien, desto übersichtlicher sind die Kennlinienfelder.
Die praktische Verwendung wird häufig bestimmt durch die Nichtlinearität/Stromabhängigkeit der Stromverstärkung $B$ bzw. der differentiellen Stromverstärkung $\beta = \frac{\mathrm{d}I_\mathrm{C}}{\mathrm{d}I_\mathrm{B}}$, siehe Bild 3.99. Ebenso ist der Eingangswiderstand stromabhängig. Im sog. Grosssignalbetrieb wird das Transistorverhalten sehr komplex; insbesondere das dynamische Verhalten. (Bei hohen Strömen verändern sich die Schaltgeschwindigkeit, die Grenzfrequenz sinkt.) Stets erfordern die möglichen Durchbruchserscheinungen die strikte Beachtung der ausgewiesenen Grenzwerte (maximale Sperrspannung, maximale Emitterspannung, maximale Ströme, maximale Verlustleistung). Die Folgen möglicher Temperaturdrift müssen meist beim Schaltungsdesign berücksichtigt werden.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

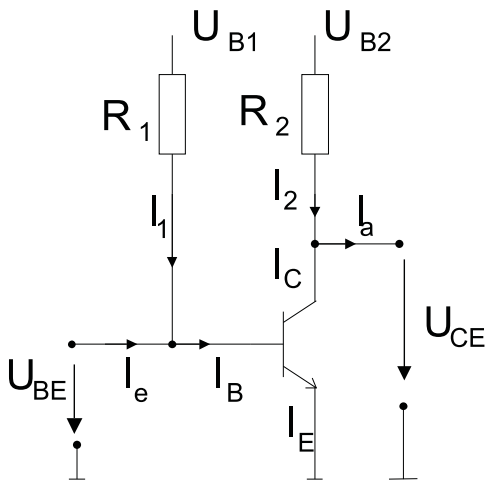



Abbildung 3.100.: Zur Bestimmung des Arbeitspunktes eines npn–Transistors(gezeichnet nach [?]).

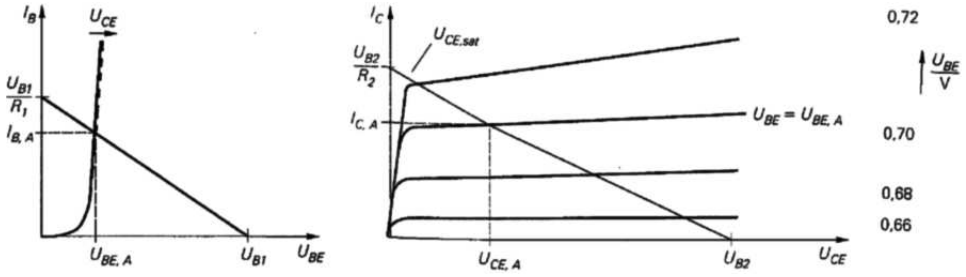



Abbildung 3.101.: Zur Bestimmung des Arbeitspunktes im Eingangs– und Ausgangskennlinienfeld[?].

Eine wichtige Anwendung des Transistors ist die lineare Verstärkung von Signalen im sog. Kleinsignalbetrieb. Hierzu wird der *Transistor* — durch äussere Beschaltung — in einem Arbeitspunkt festgehalten und mit kleinen Signalen um den Arbeitspunkt ausgesteuert. Der ausgesteuerte Kennlinienabschnitt wird näherungsweise beschrieben durch die Tangente im Arbeitspunkt; je besser die Übereinstimmung von Kurve und Tangente im Aussteuerbereich, desto besser ist das 'lineare Verhalten'.
Der Arbeitspunkt A wird durch die Spannungen $U_{\mathrm{CE,A}}$ und $U_{\mathrm{BE,A}}$ und die Ströme $I_{\mathrm{C,A}}$ und $I_{\mathrm{B,A}}$ bestimmt. Die Bestimmung des Arbeitspunktes bei bekannter Beschaltung soll anhand des Beispiels in Abbildung 3.100 kurz aufgerissen werden. Der *Transistor* wird im Maximum der Kleinsignalstromverstärkung betrieben, keinesfalls aber bei grösseren Strömen; den entsprechenden Wert findet man im Datenblatt: $I_{\mathrm{C,A}}$. Vernünftigerweise gelte immer $U_{\mathrm{CE,A}} > U_{\mathrm{CE,sat}}$. Man berechnet $I_{\mathrm{B,A}}$ und zeichnet im Eingangskennlinienfeld die Lastgerade $I_1 = \frac{1}{R_1}(U_{\mathrm{B1}} - U_{\mathrm{BE}})$ ein: Da die Kennlinie nur schwach von $U_{\mathrm{CE}}$ abhängt, ist $I_{\mathrm{B,A}}$ eindeutig festlegbar und die Lastgerade von $U_{\mathrm{B1}}$ durch $I_{\mathrm{B,A}}$ legbar. Der Schnittpunkt mit der Stromachse liefert den Wert für $R_1$. Im Ausgangskennlinienfeld ist die Vorgehensweise analog: $U_{\mathrm{B2}}$ ist vorgegeben, $U_{\mathrm{CE,A}}$ wird gewählt, so dass die mit $U_{\mathrm{BE,A}}$ festgelegte Kennlinien im Schnittpunkt wieder $I_{\mathrm{C,A}}$ liefert und die Extrapolation auf $U_{\mathrm{B2}}/R_2$, also auf $R_2$ erlaubt. Dieser graphischen Bestimmung des Arbeitspunktes ist die Lösung des Gleichungssystems aus den beiden Kennliniengleichungen des Transistors und

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

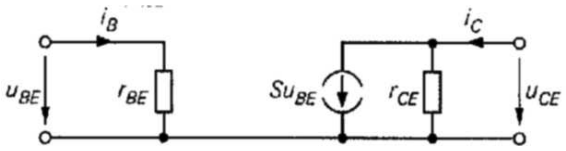



Abbildung 3.102.: Gleichstrom–Kleinsignalersatzschaltbild des Bipolartransistors[**?**].

den beiden Lastgeradengleichungen äquivalent.
Im Kleinsignalbetrieb werden die Abweichungen von den Arbeitspunktwerten als Kleinsignalspannungen $u_{\mathrm{BE}}$ und $u_{\mathrm{CE}}$ bzw. als Kleinsignalströme $i_{\mathrm{B}}$ und $i_{\mathrm{C}}$ bezeichnet. Die Verknüpfung zwischen den Strömen und den Spannungen liefern die sog. Kleinsignalgleichungen. Man kann sie auch in Matrizen–Form wiedergeben:

$$\begin{bmatrix} i_{\mathrm{B}} \\ i_{\mathrm{C}} \end{bmatrix} = Y_{\mathrm{e}} \begin{bmatrix} u_{\mathrm{BE}} \\ u_{\mathrm{CE}} \end{bmatrix} = \begin{bmatrix} Y_{11,\mathrm{e}} & Y_{12,\mathrm{e}} \\ Y_{21,\mathrm{e}} & Y_{22,\mathrm{e}} \end{bmatrix} \begin{bmatrix} u_{\mathrm{BE}} \\ u_{\mathrm{CE}} \end{bmatrix} . \tag{3.4.5}$$

Diese Darstellung ist äquivalent der Leitwert–Darstellung eines Vierpols. Der Index e der $Y$–2 x 2–Matrix steht für Emitterschaltung. Die Matrixkomponenten werden durch die sog. Kleinsignalparameter festgelegt, die alle aus den Kennlinienfeldern als Steigung der Tangente in den Arbeitspunkten ermittelt werden können:

$$Y_{11,\mathrm{e}} = \frac{1}{r_{\mathrm{BE}}} , \quad \text{mit dem Kleinsignalwiderstand } r_{\mathrm{BE}} = \left.\frac{\partial U_{\mathrm{BE}}}{\partial I_{\mathrm{B}}}\right|_{\mathrm{A}} \tag{3.4.6}$$

$$Y_{12,\mathrm{e}} = S_{\mathrm{r}} , \quad \text{mit der Rückwärtssteilheit } S_{\mathrm{r}} = \left.\frac{\partial I_{\mathrm{B}}}{\partial U_{\mathrm{CE}}}\right|_{\mathrm{A}} \approx 0 \tag{3.4.7}$$

$$Y_{21,\mathrm{e}} = S , \quad \text{mit der Steilheit } S = \left.\frac{\partial I_{\mathrm{C}}}{\partial U_{\mathrm{BE}}}\right|_{\mathrm{A}} \tag{3.4.8}$$

$$Y_{22,\mathrm{e}} = \frac{1}{r_{\mathrm{CE}}} , \quad \text{mit dem Kleinsignalausgangswiderstand } r_{\mathrm{CE}} = \left.\frac{\partial U_{\mathrm{CE}}}{\partial I_{\mathrm{C}}}\right|_{\mathrm{A}} \tag{3.4.9}$$

Den Kleinsignalgleichungen entspricht (mit der Näherung $S_{\mathrm{r}} = 0$) ein Ersatzschaltbild, das sog. Kleinsignalersatzschaltbild des Bipolartransistors. Neben dem Eingangs– und Ausgangswiderstand bestimmt die Stromquelle $S \cdot u_{\mathrm{BE}}$ das Niederfrequenzverhalten des Transistors.
Neben dieser Leitwertdarstellung gibt es noch eine zweite, übliche Variante: die Hybriddarstellung mit der $H$–Matrix:

$$\begin{bmatrix} u_{\mathrm{BE}} \\ i_{\mathrm{C}} \end{bmatrix} = H_{\mathrm{e}} \begin{bmatrix} i_{\mathrm{B}} \\ u_{\mathrm{CE}} \end{bmatrix} . \tag{3.4.10}$$

Beide Darstellungen sind ineinander überführbar. Mit Hilfe dieser Vierpoldarstellungen können Transistorschaltungen unter Benutzung der Matrizenarithmetik berechnet werden. Die Vierpolparameter sind allerdings vom gewählten Arbeitspunkt und der Frequenz abhängig. Ihre konkrete Formulierung hängt jeweils von den gewählten Grundgleichungen und den daraus abgeleiteten Modellen ab. Sie sind Grundlage von Schaltungssimulationsprogrammen wie — siehe Praktikum — PSPICE (Firma MicroSim).

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

Bekannte *Modelle*, die das statische Verhalten eines Bipolartransistors beschreiben, sind das Ebers–Moll–Modell, das Transportmodell und das Gummel–Poon–Modell. Letzteres ist das vollständige, nichtlineare Modell aus dem durch Linearisierung am Arbeitspunkt statische und dynamische Kleinsignalmodelle abgeleitet werden. Dem Buch von P. Horrowitz/W. Hill folgend, kann man mit Hilfe zweier einfacher Modelle bereits ein elementares Verständnis für viele *Transistor–Grundschaltungen* entwickeln, ohne zum $H$–Parameter–Modell oder diversen Ersatzschaltungen greifen zu müssen.
Das einfachste Transistormodell ist das des 'Stromverstärkers'. Es lautet:

1. Der Kollektor muss stets positiver als der Emitter sein.
2. Die Basis–Emitter–Diode leitet, die Basis–Kollektor–Diode sperrt (Normalbetrieb), also $U_\mathrm{B} \approx U_\mathrm{E} + 0,6\ \mathrm{V}$.
3. Grenzwerte ($I_\mathrm{C}$, $I_\mathrm{B}$, $U_\mathrm{CE}$, $U_\mathrm{BE}$, $I_\mathrm{C} \cdot U_\mathrm{CE}$, $T$) sind zu beachten.
4. $I_\mathrm{C} = B \cdot I_\mathrm{B} = \beta I_\mathrm{B}$, mit $B \approx 100$.

Im Stromverstärker–Modell bestimmt $I_\mathrm{B}$ also den Wert von $I_\mathrm{C}$, man gibt den Basisstrom vor. Die letzte Aussage bedeutet auch, dass ein kleiner (Basis–) Strom einen einhundert mal grösseren zum Kollektor hin steuert. Leider ist die Stromverstärkung $B$ keine feste und stabile Grösse. Eine Schaltung sollte nie ein bestimmtes, konstantes $B$ voraussetzen.
Das verbesserte Modell nennt sich 'Transkonduktanzverstärker'–Modell. In diesem wird die vierte Modellannahme des Stromverstärker–Modells verbessert: $I_\mathrm{C}$ wird jetzt von $U_\mathrm{BE}/U_\mathrm{T}$ und $I_\mathrm{S}(T)$ bestimmt:

$$I_\mathrm{C} = I_\mathrm{S}\left[\mathrm{e}^{\frac{U_\mathrm{BE}}{U_\mathrm{T}}} - 1\right] \approx I_\mathrm{s} \cdot \mathrm{e}^{\frac{U_\mathrm{BE}}{U_\mathrm{T}}}\ , \quad \text{'Ebers–Modell–Gleichung'}\ . \tag{3.4.11}$$

Man gibt statt des Stromes $I_\mathrm{B}$ jetzt die Spannung $U_\mathrm{BE}$ in der Schaltung vor. In der Praxis führt das aber wegen des hohen negativen Temperaturkoeffizienten der Basis–Emitter-Spannung ($I_\mathrm{S}(T)$, -2,1 mV/K) nicht direkt zu erfolgreichen Schaltungsentwürfen. Temperaturstabile Schaltungen erhält man erst durch Spannungs– oder Stromgegenkopplung, siehe später.
*Nebenbemerkung*: Man nennt das 'Transkonduktanzverstärker'–Modell in der obigen Form auch 'Reduziertes Ebers–Moll–Modell für den Bipolartransistor–Normalbetrieb'.
*Grundschaltungen* mit einem Bipolartransistor sind die Emitterschaltung (common emitter configuration), die Kollektorschaltung (common collector configuration) und die Basisschaltung (common base configuration). Den Namen der Schaltung bestimmt der Transistoranschluss, der als gemeinsamer Bezugsknoten für den Eingang und den Ausgang der Schaltung dient. (Schwächeres Kriterium).
Die *Emitterschaltung* (ohne Gegenkopplung) ist in Abbildung 3.103 gezeigt. Eine *Spannungsquelle* $U_\mathrm{g}$ mit dem Innenwiderstand $R_\mathrm{g}$ liefert die Eingangsspannung $U_\mathrm{E}$. Der Emitteranschluss liegt — kennzeichnend für diese Schaltung — direkt an Masse, der Kollektor über den Kollektorwiderstand $R_\mathrm{C}$ (typ. 1 kΩ) an der Versorgungs–*Spannungsquelle* $U_\mathrm{B}$ (typ. 5 V).
Die gemessene Übertragungskennlinie $U_\mathrm{A}(U_\mathrm{g})$ ist in Abbildung 3.104 oben gezeigt. Für $0,5\ \mathrm{V} \le U_\mathrm{g} \le 0,7\ \mathrm{V}$ nimmt $U_a = U_\mathrm{b} - I_\mathrm{C} \cdot R_\mathrm{C}$ ab; in diesem Bereich des Normalbetriebs liegt der Arbeitspunkt (o).

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

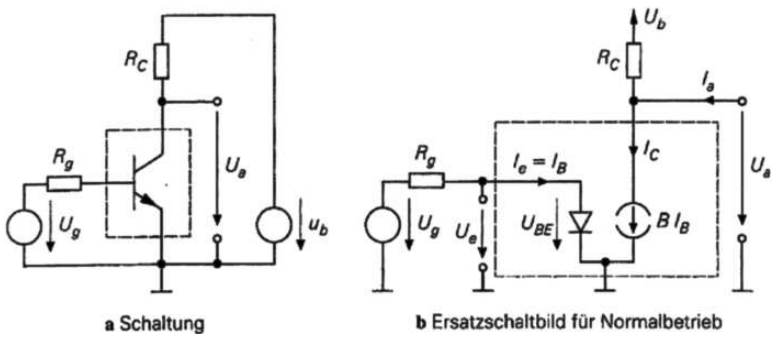



Abbildung 3.103.: Emitterschaltung: Schaltbild und Ersatzschaltbild für den Normalbetrieb[?].

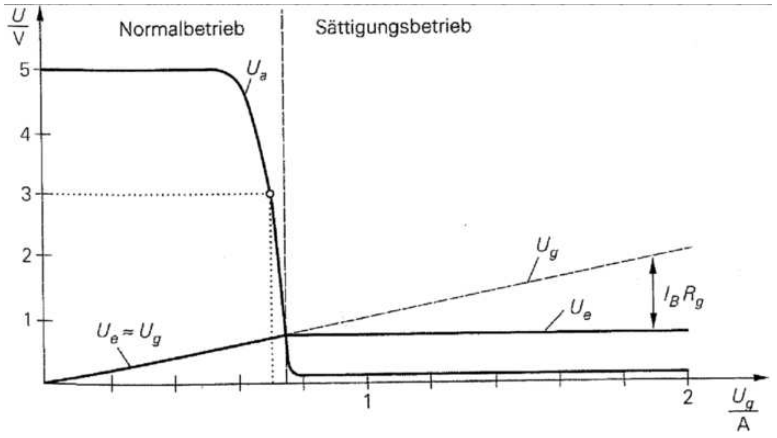



Abbildung 3.104.: Emitterschaltung: Übertragungskennlinie[?].

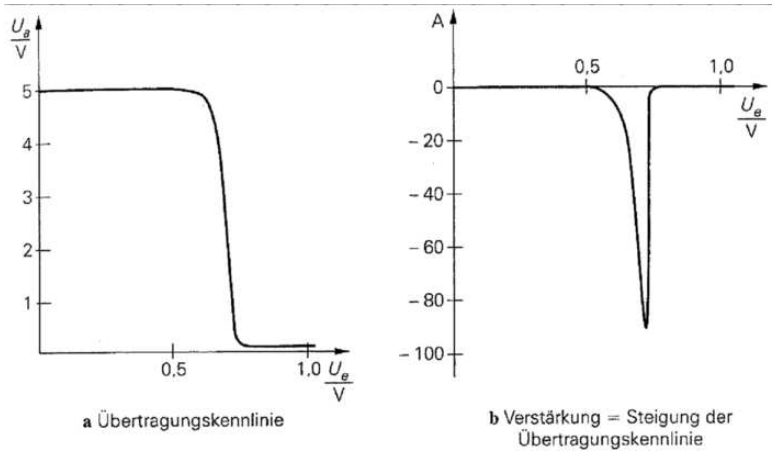



Abbildung 3.105.: Emitterschaltung: Übertragungskennlinie und Verstärkung[?].

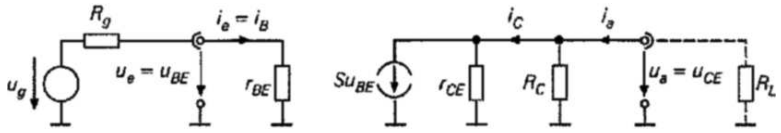



Abbildung 3.106.: Emitterschaltung: Kleinsignal–Ersatzschaltbild[**?**].

Das oben ebenfalls gezeigte Ersatzschaltbild für diesen Kennlinienbereich ist das sog. vereinfachte Transportmodell; dabei ist der Early–Effekt vernachlässigt und es gilt $B = \beta =$ konst. Man erhält:

$$I_\mathrm{C} = B \cdot I_\mathrm{B} = I_\mathrm{S} \cdot \mathrm{e}^{\frac{U_\mathrm{BE}}{U_\mathrm{T}}} \tag{3.4.12}$$

$$U_\mathrm{A} = U_\mathrm{CE} \overset{I_\mathrm{A}=0}{=} U_\mathrm{b} - I_\mathrm{C} R_\mathrm{C} \tag{3.4.13}$$

$$U_\mathrm{e} = U_\mathrm{BE} = U_\mathrm{g} - \frac{I_\mathrm{C} R_\mathrm{g}}{B} \approx U_\mathrm{g} \tag{3.4.14}$$

Typische Zahlenwerte am gewählten Arbeitspunkt sind mit $B = \ldots 100$ und $I_\mathrm{S} = \ldots$ fA: $I_\mathrm{C} \approx$ mA, $I_\mathrm{B} \approx \ldots$ $\mu$A, $U_\mathrm{e} < 1$ V und $U_\mathrm{g} < 1$ V.

Einen Hinweis auf das Kleinsignalverhalten erhält man aus der Kleinsignal–Spannungsverstärkung $A$; sie entspricht der Steigung der Übertragungskennlinie. Sie ist sehr stark vom Arbeitspunkt abhängig, d. h. dieser muss genau und temperaturstabil eingestellt werden. Schon bei kleinsten Aussteuerungen erhält man eine nichtlineare Verzerrung des Signals.

Aus dem Kleinsignal–Ersatzschaltbild erhält man (ohne Lastwiderstand $R_\mathrm{L}$):

$$A = \left.\frac{u_\mathrm{A}}{u_\mathrm{e}}\right|_{i_\mathrm{A}=0} = -S(R_\mathrm{C} || r_\mathrm{CE}) \overset{r_\mathrm{CE} \gg R_\mathrm{C}}{\approx} -SR_\mathrm{C} \,, \tag{3.4.15}$$

$$r_\mathrm{e} = \frac{u_\mathrm{e}}{i_\mathrm{e}} = r_\mathrm{BE} \,, \tag{3.4.16}$$

$$r_\mathrm{A} = \frac{u_\mathrm{A}}{i_\mathrm{A}} = R_\mathrm{C} || r_\mathrm{CE} \overset{r_\mathrm{CE} \gg R_\mathrm{C}}{\approx} R_\mathrm{C} \,. \tag{3.4.17}$$

Grösster Schwachpunkt der Emitterschaltung ist aber die Temperaturinstabilität. Wird die Temperatur um 1 Grad erhöht, muss man die Eingangsspannung um 2 mV erniedrigen, um den Kollektorstrom konstant zu halten. Andernfalls driftet die Ausgangsspannung um ca. $A \cdot 2$ mV. Eine Temperaturdrift verschiebt also den Arbeitspunkt und damit die Verstärkung etc. Praxistaugliche Schaltungen erhält man durch Gegenkopplung.

Die *Emitterschaltung mit Stromgegenkopplung*: Durch Einfügen eines Emitterwiderstand $R_\mathrm{E}$ (typ. 500 $\Omega$) wird die Übertragungskennlinie deutlich verändert, siehe Bild unten. Der Arbeitsbereich beträgt jetzt ein gutes Volt und die Kennlinie verläuft annähernd linear; der Arbeitspunkt wird mittig gewählt. Die Kleinsignal–Verstärkung spiegelt dies in einem annähernd konstanten Mittelteil wieder. Allerdings erkauft man sich diesen Stabilitätsgewinn durch einen Verlust an Verstärkung. Dies zwingt einen zum Bau mehrstufiger Verstärker.

Aus dem Ersatzschaltbild erhält man näherungsweise für $r_\mathrm{CE} \gg R_\mathrm{C}, R_\mathrm{E}$ und $\beta \gg$

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

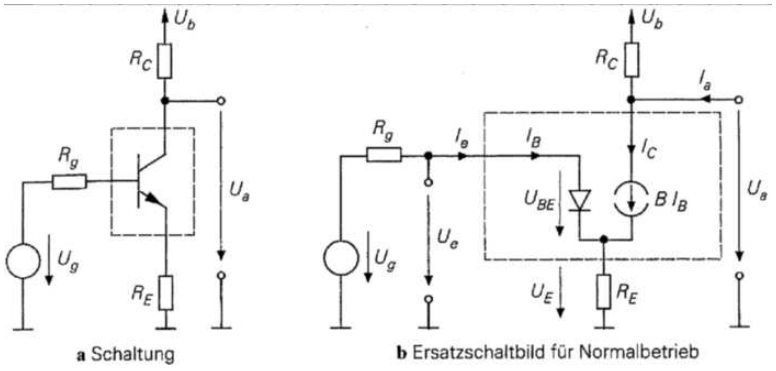



Abbildung 3.107.: Emitterschaltung mit Stromgegenkopplung: Schaltung und Ersatzschaltbild (Normalbetrieb)[**?**].

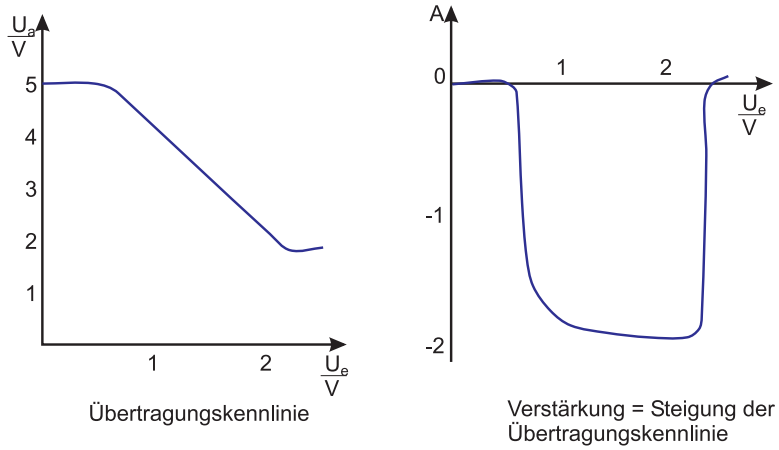



Abbildung 3.108.: Emitterschaltung mit Stromgegenkopplung: Übertragungskennlinie und Verstärkung[**?**].

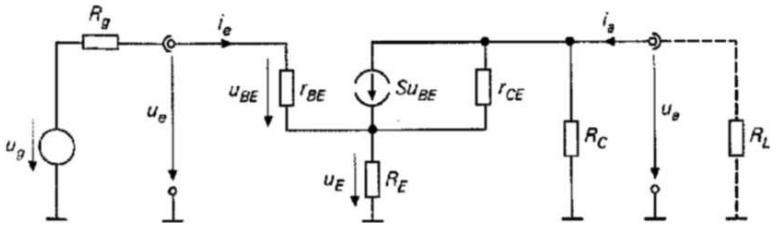



Abbildung 3.109.: Emitterschaltung mit Stromgegenkopplung: Kleinsignal–Ersatzschaltbild[**?**].

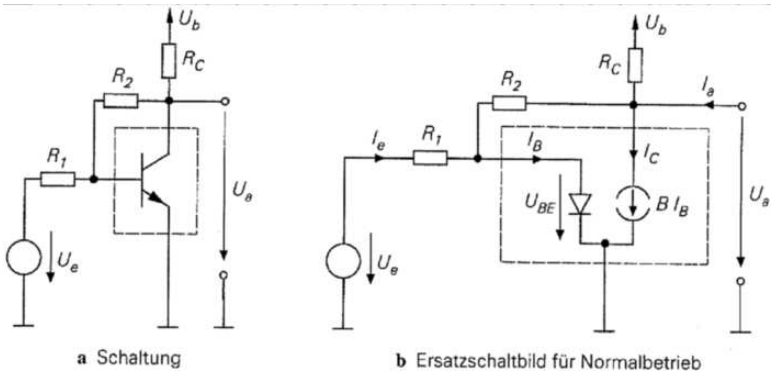



Abbildung 3.110.: Emitterschaltung mit Spannungsgegenkopplung: Schaltung und Ersatzschaltbild (Normalbetrieb)[**?**].

1, sowie $R_\mathrm{L} = 0$:

$$A = \left.\frac{u_\mathrm{A}}{u_\mathrm{e}}\right|_{i_\mathrm{A}=0} \approx -\frac{SR_\mathrm{C}}{1+S_\mathrm{E}} \overset{SR_\mathrm{E}\gg 1}{\approx} -\frac{R_\mathrm{C}}{R_\mathrm{E}} \tag{3.4.18}$$

$$r_\mathrm{e} = \frac{u_\mathrm{e}}{i_\mathrm{e}} \approx r_\mathrm{BE} + \beta R_\mathrm{E} = r_\mathrm{BE}(1+SR_\mathrm{E}) \tag{3.4.19}$$

$$r_\mathrm{A} = \frac{u_\mathrm{A}}{i_\mathrm{A}} \approx R_\mathrm{C}\ . \tag{3.4.20}$$

Die Verstärkung wird in erster Näherung nicht vom *Transistor*, sondern von den äusseren Beschaltungswiderständen festgelegt. $R_\mathrm{E}$ sorgt für Stabilität und Linearität, erhöht den Eingangswiderstand (!), belässt den Ausgangswiderstand unverändert, aber senkt die Verstärkung stark ab und darf deshalb auch nicht zu gross gewählt werden. Das Verstärkungs–Bandbreite–Produkt bleibt im übrigen fast unverändert.
Die *Emitterschaltung mit Spannungsgegenkopplung*: Hier fügt man zwischen Basis– und Kollektoranschluss einen Widerstand ein (typisch $R_1 = R_\mathrm{C} = 1$ kΩ, $R_2 = 2$ kΩ), wodurch ein Teil der Ausgangsspannung auf den Eingang zurückgekoppelt wird. Nochmals ist der Arbeitsbereich vergrössert, die Verstärkung ist von vergleichbarer Qualität und annähernd symmetrisch um $U_\mathrm{e} = 0$.
Aus dem Ersatzschaltbild erhält man näherungsweise:

$$A = \left.\frac{u_\mathrm{A}}{u_\mathrm{e}}\right|_{i_\mathrm{A}=0} \approx -\frac{R_2}{R_1 + \frac{R_1+R_2}{SR_\mathrm{C}}} \overset{SR_\mathrm{C}\gg 1+R_2/R_1}{\approx} -\frac{R_2}{R_1} \tag{3.4.21}$$

$$r_\mathrm{e} = \frac{u_\mathrm{e}}{i_\mathrm{e}} \approx R_1 \tag{3.4.22}$$

$$r_\mathrm{A} = \frac{u_\mathrm{A}}{i_\mathrm{A}} \approx R_\mathrm{C} || \left(\frac{1}{S}\left(1+\frac{R_2}{R_1}\right) + \frac{R_2}{\beta}\right)\ . \tag{3.4.23}$$

Die Verstärkung wird durch $R_2/R_1$ bestimmt, also wieder durch die Gegenkopplung. Der Eingangswiderstand ist relativ klein, der Ausgangswiderstand der niedrigste im Vergleich; dies ist bei niederohmigen oder kapazitiven Lasten von Vorteil.

Wählt man $R_1 = 0$ und gibt einen Ansteuerstrom $I_\mathrm{e}$ vor, so erhält man einen *Strom–Spannungs–Wandler* (Transimpedanzverstärker). Über einen typischen Bereich von etwa einem mA arbeitet dieser Verstärker mit guter Linearität.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

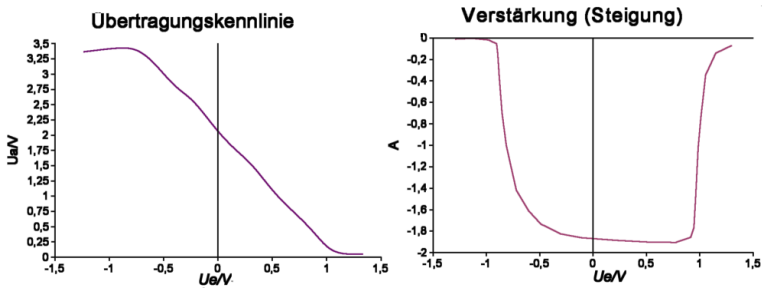



Abbildung 3.111.: Emitterschaltung mit Spannungsgegenkopplung: Übertragungskennlinie und Verstärkung[?].

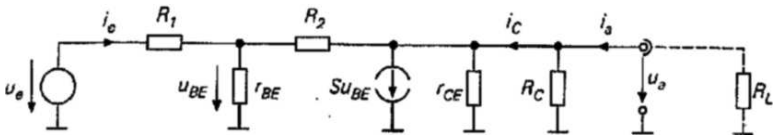



Abbildung 3.112.: Emitterschaltung mit Spannungsgegenkopplung: Kleinsignal–Ersatzschaltbild[?].

Zur Arbeitspunkteinstellung ist ganz allgemein festzustellen, dass der Arbeitspunkt möglichst wenig von den — streuenden — Transistorparametern abhängen sollte. Es gibt zwei prinzipiell unterschiedliche Verfahren, nämlich die Gleichspannungs–Kopplung und die Wechselspannungs–Kopplung. Im letzteren Fall werden Ein– und Ausgang des Verstärkers über Koppelkondensatoren mit dem *Signal* bzw. der Last verbunden.
Die *Kollektorschaltung* — siehe Bild 3.113 — besteht aus dem *Transistor* mit $R_\mathrm{E}$ und $U_\mathrm{B}$ (typisch 1 kΩ und +5 V), der Signalspannungsquelle $U_\mathrm{g}$ mit ihrem Innenwiderstand $R_\mathrm{g}$.
Aus dem Ersatzschaltbild und dem vereinfachten Transportmodell folgt:

$$U_\mathrm{A} = (I_\mathrm{C} + I_\mathrm{B} + I_\mathrm{A})R_\mathrm{E} \approx I_\mathrm{C} \cdot R_\mathrm{E} \quad \text{für } I_\mathrm{A} = 0 \text{ und} \tag{3.4.24}$$

$$U_\mathrm{e} \approx U_\mathrm{g} \,. \tag{3.4.25}$$

Die Kennlinien oben zeigen, dass die Ausgangsspannung $U_\mathrm{A}$ der Eingansspannung $U_e$ im Abstand $U_\mathrm{BE}$ folgt; man nennt die Kollektorschaltung deshalb auch *Emitterfolger*.

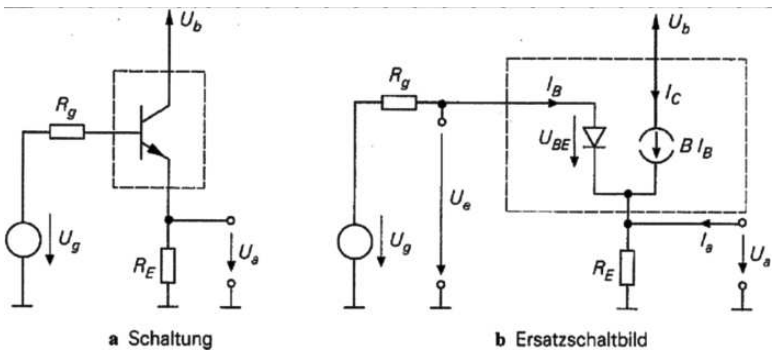



Abbildung 3.113.: Kollektorschaltung: Schaltbild und Ersatzschaltbild für den Normalbetrieb[?].

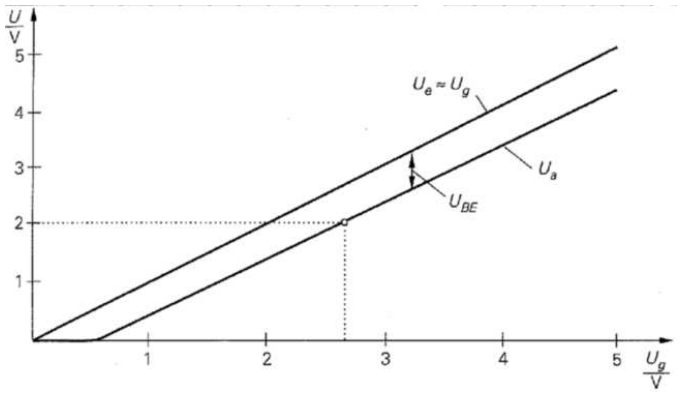



Abbildung 3.114.: Kollektorschaltung: Kennlinien[**?**].

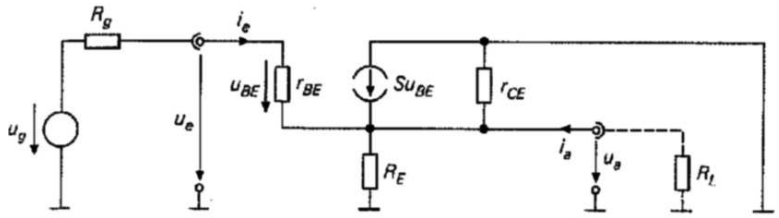



Abbildung 3.115.: Kollektorschaltung: Kleinsignal–Ersatzschaltbild[**?**].

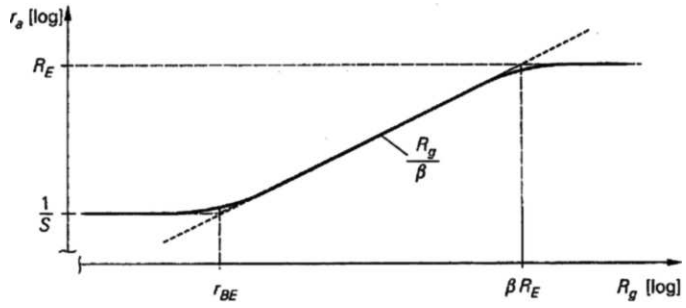



Abbildung 3.116.: Kollektorschaltung: Widerstandstransformation $r_{\mathrm{A}}(R_{\mathrm{g}})$[**?**].

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

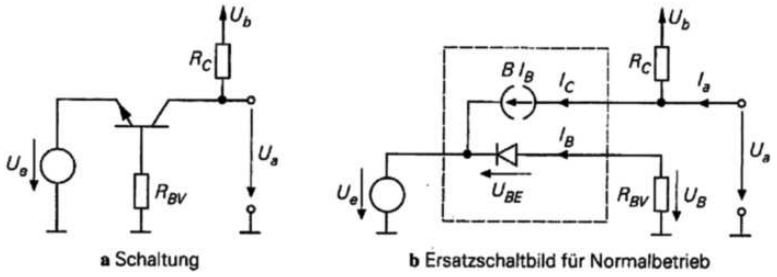



Abbildung 3.117.: Basisschaltung: Schaltbild und Ersatzschaltbild für den Normalbetrieb[**?**].

Das Kleinsignalverhalten entnimmt man wieder dem zugehörigen Ersatzschaltbild. Will man zusätzlich den Einfluss eines Lastwiderstandes $R_\mathrm{L}$ berücksichtigen, muss man statt $R_\mathrm{E}$ die Parallelschaltung von $R_\mathrm{E}$ und $R_\mathrm{L}$ berücksichtigen. Man erhält:

$$A = \left.\frac{u_\mathrm{A}}{u_\mathrm{e}}\right|_{i_\mathrm{A}=0} = \frac{SR_\mathrm{E}}{1+S_\mathrm{E}} \overset{SR_\mathrm{E}\gg 1}{\approx} 1 \tag{3.4.26}$$

$$r_\mathrm{e} = \frac{u_\mathrm{e}}{i_\mathrm{e}} \approx r_\mathrm{BE} + \beta R_\mathrm{E} \overset{SR_\mathrm{E}\gg 1}{\approx} \beta r_\mathrm{BE} \tag{3.4.27}$$

$$r_\mathrm{A} = \frac{u_\mathrm{A}}{i_\mathrm{A}} \approx R_\mathrm{E} || \left(\frac{R_\mathrm{g}}{\beta} + \frac{1}{S}\right) . \tag{3.4.28}$$

Von praktischer Bedeutung ist die Verwendung der Kollektorschaltung zur Impedanztransformation. Bei rein ohmschen Widerständen gibt die obenstehende Abbildung die Verhältnisse wieder. Der Eingangswiderstand ist in erster Näherung vom Lastwiderstand abhängig, der Ausgangswiderstand hängt vom Innenwiederstand $R_\mathrm{g}$ der Signalquelle ab. Die Widerstandstransformation lässt sich zur Impedanztransformation verallgemeinern.
Die *Basisschaltung* schliesslich ist im nächsten Bild gezeigt. Die Basis liegt auf konstantem Potential, der Widerstand $R_\mathrm{BV}$ (z. B. 1 kΩ) begrenzt den Basisstrom bei Übersteuerung und spielt im Normalbetrieb keine Rolle. Das Eingangssignal wird direkt an den Emitter gekoppelt. Kapazitives Übersprechen ist hierdurch minimiert, höchste Frequenzen werden noch verstärkt. Die Basischaltung wird deshalb gern in Hochfrequenzschaltungen eingesetzt. Weitere Werte $R_\mathrm{C} \approx 1$ kΩ, $U_\mathrm{B} = +5$ V typ.
Aus dem Ersatzschaltbild und dem vereinfachten Transportmodell erhält man:

$$U_\mathrm{A} = U_\mathrm{b} - I_\mathrm{C} R_\mathrm{C} \quad \text{für } I_\mathrm{A} = 0 \tag{3.4.29}$$

$$U_\mathrm{e} \approx -U_\mathrm{BE} . \tag{3.4.30}$$

Der Arbeitspunkt wird mittig in den abfallenden Bereich der Übertragungskennlinie gelegt. Man erhält näherungsweise und mit $R_\mathrm{L} = 0$:

$$A = \left.\frac{u_\mathrm{A}}{u_\mathrm{e}}\right|_{i_\mathrm{A}=0} = \frac{\beta R_\mathrm{C}}{r_\mathrm{BE} + R_\mathrm{BV}} \overset{r_\mathrm{BE}\gg R_\mathrm{BV}}{\approx} SR_\mathrm{C} \tag{3.4.31}$$

$$r_\mathrm{e} = \frac{u_\mathrm{e}}{i_\mathrm{e}} \approx \frac{1}{S} + \frac{R_\mathrm{BV}}{\beta} \overset{r_\mathrm{BE}\gg R_\mathrm{BV}}{\approx} \frac{1}{S} \tag{3.4.32}$$

$$r_\mathrm{A} = \frac{u_\mathrm{A}}{i_\mathrm{A}} \approx R_\mathrm{C} . \tag{3.4.33}$$

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

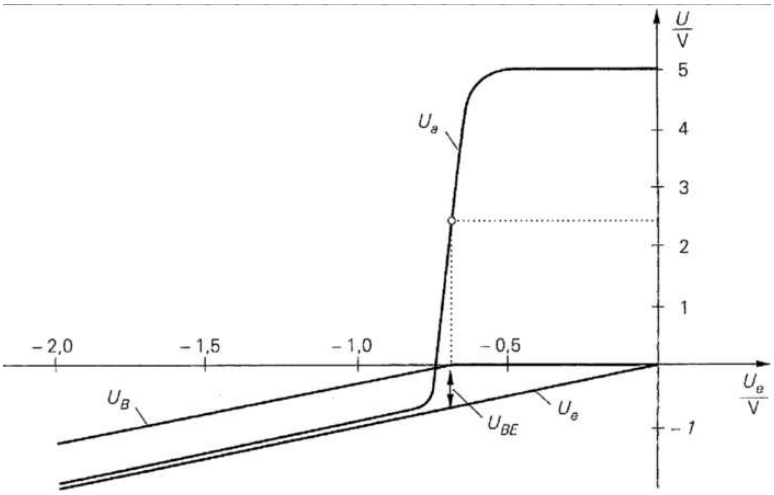



Abbildung 3.118.: Basisschaltung: Kennlinien[**?**].

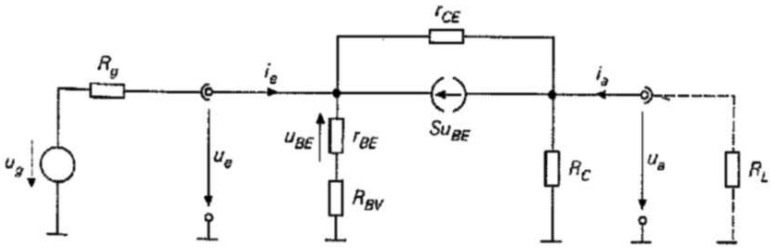



Abbildung 3.119.: Basisschaltung: Kleinsignal–Ersatzschaltbild[**?**].

Die Eingangsimpedanz ist also vergleichsweise niedrig und passt damit sehr gut an planare Wellenleiter (siehe integrierte Optoelektronik).
Steuert man die Basisschaltung mit einer Stromquelle an, so tritt anstelle der Verstärkung $A$ der Übertragungswiderstand $R_\mathrm{T}$, die sog. Transimpedanz $R_\mathrm{T} = \left.\frac{u_a}{i_a}\right|_{i_\mathrm{A}=0} \approx R_\mathrm{C}$. Der Klirrfaktor des Strom–Spannungs–Wandlers in der Basisschaltung ist besonders klein.
Reicht die Stromverstärkung eines einzelnen Transistors nicht aus, so greift man zur *Darlington–Schaltung*:
Der Emitterstrom des Eingangstransistors fliesst in die Basis des zweiten Transistors. In erster Näherung gilt für die Stromverstärkung:

$$B \approx B_1 \cdot B_2 \; . \tag{3.4.34}$$

Darlington–Transistoren kann man wie Einzeltransistoren (in einem Gehäuse) kaufen. Sie werden häufig als Schalter eingesetzt: Aufgrund der grossen Verstärkung können grosse Ströme geschaltet werden. Allerdings schaltet der 2. *Transistor* und damit der gesamte *Darlington–Transistor* langsam. Die in der Basis gespeicherte Ladung leitet man deshalb über den Widerstand $R$ ab; der *Darlington–Transistor* kann so schneller gesperrt werden, allerdings auf Kosten der Stromverstärkung.

## 3.4.4. Feldeffekttransistoren

Schaltungen, die Feldeffekttransistoren enthalten, verwirren viele. Zu unrecht, denn die Unterschiede zu den Bipolartransistoren sind im Grundsatz sekundär. Den

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

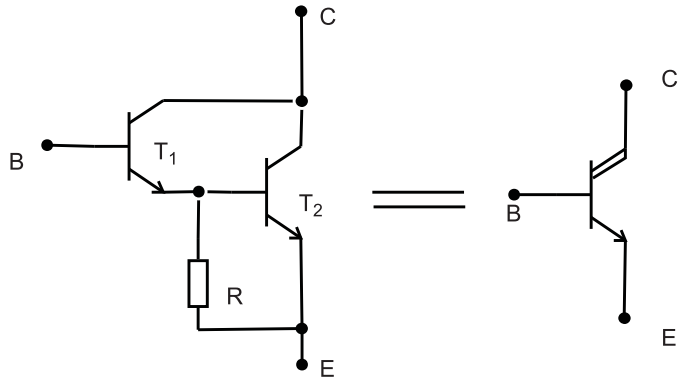



Abbildung 3.120.: *npn–Darlington–Transistor*: Schaltung und Schaltzeichen (gezeichnet nach [**?**]).

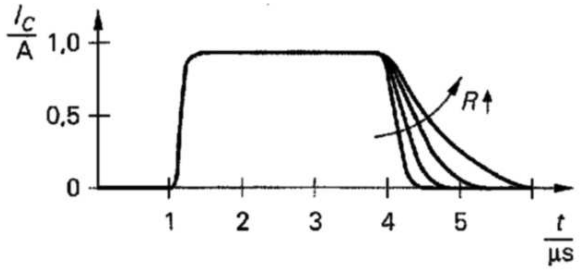



Abbildung 3.121.: *Darlington–Transistor*: zum Schaltverhalten[**?**].

Bipolartransistor–Anschlüssen Emitter (E), Basis (B), Kollektor (K) und Substrat (S) entsprechen die Feldeffekttransistor–Anschlüsse Source (S), Gate (G), Drain (D) und Bulk (B). Ersetzt man im vorigen Kapitel in den Gleichungen bei den Strömen und Spannungen die korrespondierenden Anschlussbezeichnungen, erhält man erstaunlich häufig die gültigen Bezeichnungen für die Grundschaltungen mit Feldeffekttransistoren und in nullter Näherung zeigen die entsprechenden Kennlinien den selben Verlauf. Auch die Grenzwerte ähneln einander stark.

Es gibt eine ganze Reihe von Feldeffekttransistoren: 4 Typen von MOSFETs, 2 Typen von JFETs und 2 von MESFETs. Zur Übersicht trägt aber bei, dass die Ausgangskennlinien aller Bauelemente mit n–Kanal sich ähneln (positive $J_\mathrm{D}$ für positive $U_\mathrm{DS}$) und die für alle p–Kanal–Bauelemente ebenfalls (negative $I_\mathrm{D}$ für negative $U_\mathrm{DS}$). Auch die Übertragungskennlinien der n–Kanal (bzw. gespiegelt am Ursprung die der p–Kanal–Bauelemente) zeigen einen Dioden–ähnlichen Verlauf; die der einzelnen Typen sind lediglich auf der $U_\mathrm{GS}$–Achse gegeneinander verschoben. (Die zugehörigen Schwellspannungen $U_\mathrm{GS,th}$ können positive und negative Werte annehmen.)

Das statische Verhalten eines FETs beschreiben wieder seine Kennlinien. Die Eingangskennlinien zeigen im normalen Arbeitsbereich nur Nulllinien: der Gatestrom ist vernachlässigbar klein, die Steuerung mit $U_\mathrm{GS}$ erfolgt leistungslos. Dies ist ein wesentlicher Unterschied zu den Bipolartransistoren, d. h. das entsprechende '$r_\mathrm{BE}$' ist $\infty$.

Eine empirische Beschreibung der Kennlinien geben die sog. Grosssignalgleichun-

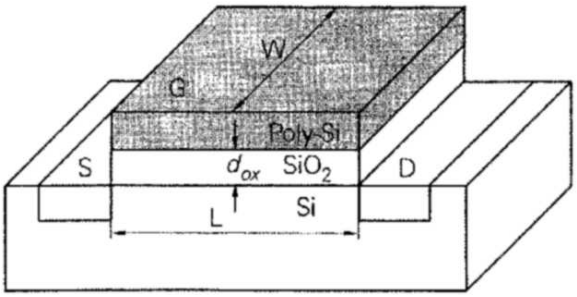



Abbildung 3.122.: Zum Steilheitskoeffizient[?].

gen (näherungsweise) wieder.

$$I_{\mathrm{D}} = \begin{cases} 0 & \mathrm{SB} \\ K U_{\mathrm{DS}} \left(U_{\mathrm{GS}} - U_{\mathrm{th}} - \frac{U_{\mathrm{GS}}}{2}\right)\left(1 + \frac{U_{\mathrm{DS}}}{U_{\mathrm{A}}}\right) & \mathrm{OB} \\ \frac{K}{2}\left(U_{\mathrm{GS}} - U_{\mathrm{th}}\right)^2 \left(1 + \frac{U_{\mathrm{D}} S}{U_{\mathrm{A}}}\right) & \mathrm{AB} \end{cases} \tag{3.4.35}$$

$$I_{\mathrm{G}} = \begin{cases} 0 & \mathrm{MOSFET} \\ I_{\mathrm{G,S}} \left(\mathrm{e}^{\frac{U_{\mathrm{GS}}}{U_{\mathrm{T}}} - 1}\right) & \mathrm{Sperrschicht\text{–}FET} \end{cases} \tag{3.4.36}$$

$$\left.\begin{array}{lcl} \mathrm{SB} & : & \mathrm{Sperrbereich} \\ \mathrm{OB} & : & \mathrm{ohmscher\ Bereich} \\ \mathrm{AB} & : & \mathrm{Abschnürbereich} \end{array}\right\} \Rightarrow \begin{cases} U_{\mathrm{GS}} < U_{\mathrm{th}} \\ U_{\mathrm{GS}} \geq U_{\mathrm{th}},\ 0 \leq U_{\mathrm{DS}} < U_{\mathrm{GS}} - U_{\mathrm{th}} \\ U_{\mathrm{GS}} \geq U_{\mathrm{th}},\ U_{\mathrm{DS}} \geq U_{\mathrm{GS}} - U_{\mathrm{th}} \end{cases}$$

Die Gleichungen enthalten die bereits eingeführte Gate–Source–Schwellspannung $U_{\mathrm{th}}$, ab der der Drainstrom $I_{\mathrm{D}}$ merklich einsetzt und die (in Analogie zu den Verhältnissen beim Bipolartransistor) sog. Early–Spannung $U_{\mathrm{A}}$, die die endliche Steigung der Ausgangskennlinien bei grossen Spannungen (im 'Abschnürbereich') berücksichtigt. Des weiteren den sog. Steilheitskoeffizient $K$:

$$K = K' \frac{W}{L} = \mu\, C_{\mathrm{Ox}} \cdot \frac{W}{L}\ . \tag{3.4.37}$$

Der Anstieg im annähernd linearen ohmschen Kennlinienbereich ist mit der Beweglichkeit der Ladungsträger im Kanal, dem Kapazitätsbelag des Gate–Oxids und der Geometrie des Gates verbunden. Diese Gleichungen gelten für die vergleichsweise gross dimensionierten Einzel–FETs, wobei Drain und Source i. allg. nicht vertauscht werden dürfen (unsymmerische Bauelemente). Bei den immer kleiner dimensionierten integrierten FETs wird die Beschreibung zunehmend schwieriger und die zugehörigen Modelle, insbesondere zur Beschreibung der dynamischen Eigenschaften, wie sie in Schaltungssimulationen benötigt werden, werden immer umfangreicher. Im Grundsatz gilt aber, dass $K \sim \frac{1}{d_{\mathrm{Ox}}} \cdot \frac{1}{L}$ ist, d. h. die Miniaturisierung die Steilheit vergrössert.
Im ohmschen Bereich kann der FET als *steuerbarer Widerstand* betrieben werden. In der Umgebung von $U_{\mathrm{DS}} = 0$ wirkt der FET bei Aussteuerung mit kleinen Amplituden als steuerbarer linearer Widerstand, bei grösseren Amplituden zunehmend nichtlinear, ebenso für $U_{\mathrm{DS}} > 0$. Mögliche linearisierte Schaltungen vermeiden letzteres auf Kosten der Grösse des Abstimmbereichs.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

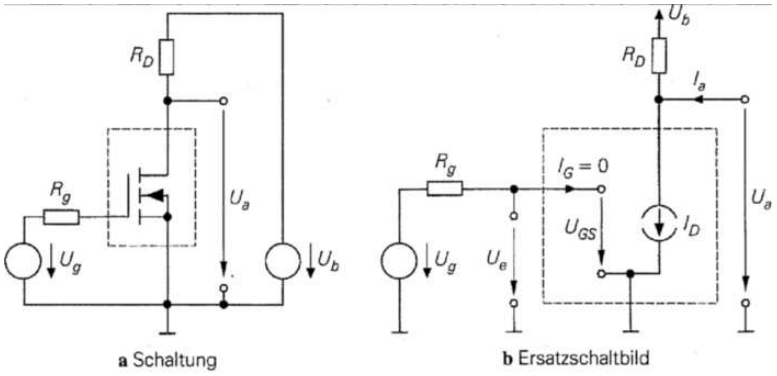



Abbildung 3.123.: Sourceschaltung: Schaltung und Ersatzschaltbild[**?**].

Obwohl das Grosssignalverhalten des FETs deutlich besser ist als das der Bipolartransistoren, ist eine wichtige FET–Anwendung die lineare Verstärkung von Signalen im Kleinsignalbetrieb. Der Arbeitspunkt liegt im Abschnürbereich. Das früher zu Kleinsignalgrössen und –Gleichungen gesagte gilt analog. Besonders einfach werden die Verhältnisse wieder beim Einzel–FET, bei dem Source und Bulk miteinander verbunden sind und die Source–Bulk–Diode wegen $U_{\mathrm{SB}} = 0$ im Kleinsignalersatzschaltbild nicht berücksichtigt werden muss (Die Stromquelle mit der Substrat–Steilheit $S_{\mathrm{B}}$ und die Bulk–Source–Kapazität entfallen.).
Die Unterschiede vom FET–Verhalten gegenüber dem der Bipolartransistoren kann man zusammenfassen:
FETs sind extrem ladungsempfindlich und erfordern besondere Schutzmassnahmen. Sie besitzen einen negativen Temperaturkoeffizient (JFETs) bzw. werden meist im Bereich mit negativen Temperaturkoeffizient betrieben (MOSFET); $\frac{\mathrm{d}U_{\mathrm{th}}}{\mathrm{d}T}$ liegt bei rund $-2\ \frac{\mathrm{m}V}{\mathrm{K}}$ und sie benötigen keinen zusätzlichen Stabilisierungsaufwand. Das Grosssignalverhalten ist günstiger. Bei gleichen Aussteueramplituden sind FET–Kleinsignalverstärker klirrärmer, bei hochohmigen Quellen sind FETs deutlich rauschärmer (Anwendung: Photodioden–Vorverstärker); JFETs zeigen dabei das geringere $1/f$–Rauschen (kleine Frequenzen!). Bei niederohmigen Quellen dagegen hat der Bipolartransistor das deutlich bessere Rauschverhalten.
*Grundschaltungen* mit einem FET sind die Source–Schaltung (common source configuration), die Drain–Schaltung (common drain configuration) und die Gate–Schaltung (common gate configuration). Der gemeinsame Bezugspunkt für Schaltungsein– und –Ausgang gibt wieder den Namen. (Schwächeres Kriterium)
Die *Sourceschaltung* entspricht der Emitterschaltung, siehe Abbildungen. Wegen des grösseren Steilheitskoeffizienten (höhere Ladungsträgerbeweglichkeit) werden meist n–Kanal–MOSFETs bevorzugt, selbstsperrende MOSFETs häufiger als selbstleitende.
Der MOSFET hat einen extrem hohen Eingangswiderstand, $I_{\mathrm{G}} = 0$ und die BE–Diode der korrespondierenden Emitterschaltung entfällt hier. Der Innenwiderstand $R_{\mathrm{g}}$ der Quelle hat keinen Einfluss auf die Kennlinie, wohl aber auf das dynamische Verhalten.
Das Kleinsignalverhalten der FETs ist gekennzeichnet durch die geringere Verstärkung bei vergleichbaren Strömen bzw. Arbeitspunkten aufgrund ihrer geringeren Steilheit im Vergleich zu Bipolartransistoren. Die korrespondierenden Kleinsignalersatzschaltbilder unterscheiden sich nur in $r_{\mathrm{e}} = \infty$ (gegenüber $r_{\mathrm{e}} = r_{\mathrm{BE}}$), ebenso

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

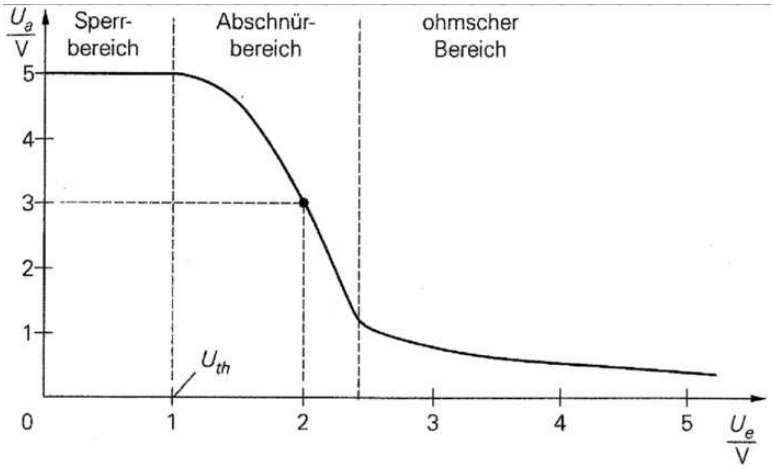



Abbildung 3.124.: Sourceschaltung: Übertragungskennlinie[**?**].

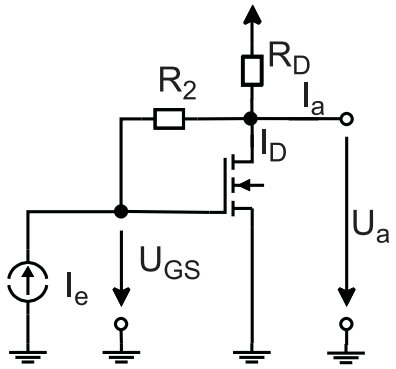



Abbildung 3.125.: Transimpedanzverstärker (gezeichnet nach [**?**]).

die Ausdrücke für die Spannungsverstärkung $A$, sowie die Ein– und Ausgangswiderstände $r_\mathrm{e}$ und $r_\mathrm{A}$.
Aufgrund der günstigeren Klirrfaktoren wird die Sourceschaltung gerne in schmalbandigen HF–Verstärkern eingesetzt.
Die vorhandene Nichtlinearität und Temperaturabhängigkeit kann wieder durch Einfügen eines Widerstands $R_\mathrm{S}$ verbessert werden: *Sourceschaltung mit Stromgegenkopplung*. Auch hier gilt, bis auf $r_\mathrm{e} = \infty$ die Analogie zur Emitterschaltung mit Stromgegenkopplung.
Ähnliches gilt für die *Sourceschaltung mit Spannungsgegenkopplung*. Schaltung, Kennlinie und Kleinsignalersatzschaltbild entsprechen ganz, die Kleinsignal–Formeln sind modifiziert. Diese Schaltung wird sehr selten eingesetzt, da der hohe FET–Eingangswiderstand nicht genutzt wird. Entfernt man aber den eingangsseitigen Widerstand, so erhält man eine Strom–Spannungs–Wandler–Schaltung (Transimpedanzverstärker). Diese rauscharme Schaltung wird gerne bei Photodioden–Empfängern eingesetzt; die Diode wird im Sperrbereich betrieben und wirkt als Stromquelle mit hohem Innenwiderstand.
Die *Drainschaltung* entspricht der Kollektorschaltung. Analog wird sie auch Sourcefolger genannt. Sie ist hervorragend geeignet, hochohmige Quellen zu puffern. Die Eingangskapazität ist extrem klein. Da der Ausgangswiderstand $> 50\ \Omega$ ist, benützt man gerne eine Darlingtonschaltung aus FET und Bipolarverstärker mit kleinerem $r_\mathrm{A}$ und höheren Ausgangsströmen.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

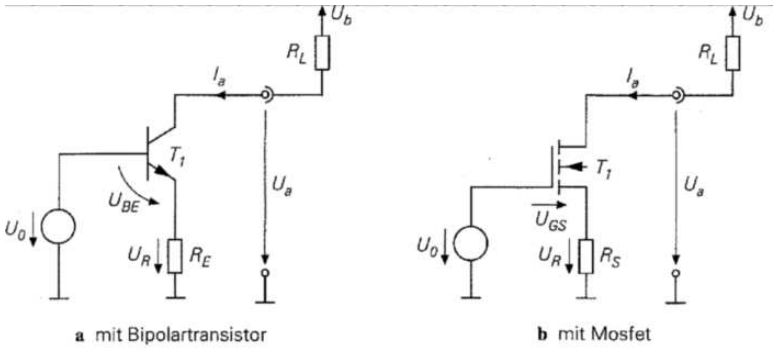



Abbildung 3.126.: Prinzip einer einfachen Stromquelle[?].

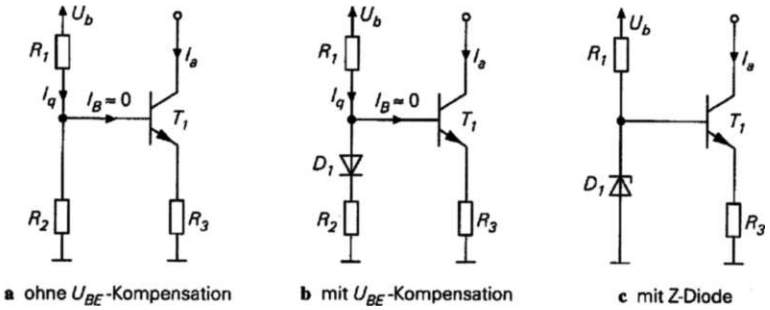



Abbildung 3.127.: Beispiele einfacher Stromquellen[?].

Bei der *Gateschaltung* schliesslich gilt wieder viel Analoges zur Basisschaltung einschliesslich des Kleinsignalverhaltens. Hervorzuheben ist hier die sehr übersichtliche Verwendung als Strom–Spannungs–Wandler mit einem $r_\mathrm{A} \approx R_\mathrm{D}$.

### 3.4.5. Einige Grundschaltungen mit Transistoren

Eine *Stromquelle* (current source) liefert an einem Verbraucher $R_\mathrm{L}$ einen konstanten Strom: Konstantstromquelle; idealerweise unabhängig vom momentanen Spannungsabfall an $R_\mathrm{L}$. Stromquellen sind spannungsgesteuert, d. h. eine kleine Steuerspannung bestimmt das Verhalten einer Schaltung mit grossem Ausgangswiderstand. Eingesetzt werden sie z. B. zur Einstellung von Arbeitspunkten, als wesentliche Bestandteile diverser Generatoren oder zur Versorgung von elektrochemischen Zellen und optoelektronischen Bauelementen (LEDs, Laserdioden) etc.

Das Prinzip einer einfachen Stromquelle beruht auf der Emitterschaltung bzw. (bei diskreten Schaltungen selten) auf der Source–Schaltung, siehe Abbildung 3.126; zur Stabilisierung meist mit Stromgegenkopplung. Dabei wird die Basis (bzw. das Gate) auf möglichst konstantem Potential gehalten. Verwendet man Transistoren (im Normalbetrieb) mit grosser Early–Spannung, dann ist im 'Arbeitsbereich' der Ausgangsstrom durch $R_\mathrm{L}$ praktisch konstant.
Einige einfache Stromquellen für diskrete Schaltungen zeigt Abbildung 3.127. Von links geschieht die Festlegung der konstanten Basisspannung durch a) einen Spannungsteiler, b) einen Spannungsteiler mit Diode zur Kompensation der Temperaturabhängigkeit von $U_\mathrm{BE}$ des Transistors und c) mit Zenerdiode (mit der geringen Temperaturabhängigkeit des Zenerdurchbruchs).

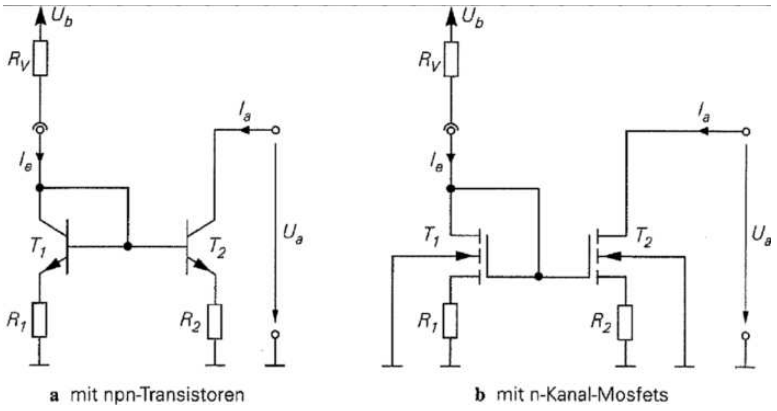



Abbildung 3.128.: Prinzip eines einfachen Stromspiegels[**?**].

Ein *Stromspiegel* (current mirror) ist eine stromgesteuerte Stromquelle, d. h. am Ausgang erhält man eine verstärkte oder identische oder abgeschwächte Kopie des Eingangsstroms. Bei konstantem Eingangsstrom funktioniert jeder Stromspiegel als Konstant–Stromquelle.
Der einfache Stromspiegel besteht aus zwei Transistoren (genauer aus *npn–Diode* und *npn–Transistor* bzw. aus n–Kanal–Diode und n–Kanal–MOSFET) und — optional — aus zwei Widerständen zur Stromgegenkopplung. Mit einem zusätzlichen, hinreichend grossen Vorwiderstand $R_\mathrm{V}$ kann man den Eingangsstrom als konstanten Refererenzstrom festlegen; man erhält so eine Konstant–Stromquelle.
Das sog. Übersetzungsverhältnis $K_\mathrm{i} = \frac{I_\mathrm{A}}{I_\mathrm{e}}$ hängt beim npn–Stromspiegel ohne Gegenkopplung (möglich bei integrierten Schaltungen) vom Verhältnis $\frac{I_{\mathrm{S}_2}}{I_{\mathrm{S}_1}}$ ab, also vom Grössenverhältnis der Transistoren. In diskreten Schaltungen ist nur der npn–Stromspiegel mit Gegenkopplung realisierbar; hier gilt $K_\mathrm{i} \approx \frac{R_1}{R_2}$. Ohne Gegenkopplung ist bei konstantem Übersetzungsverhältnis (z. B. $K_\mathrm{i} = 1$) die Übertragungskennlinie $I_\mathrm{A}(I_\mathrm{e})$ über mehrere Dekaden linear, mit Gegenkopplung etwas kleiner.
*Nebenbemerkung*: Der Arbeitsbereich von npn–Stromspiegeln ist immer grösser als der vergleichbarer n–Kanal–Stromspiegel.
Das limitierte Verstärkungs–Bandbreite–Produkt der Emitterschaltung begrenzt auch den einfachen Stromspiegel. Man greift deshalb gerne zur *Kaskodenschaltung*, meistens um eine grössere Verstärkung zu erreichen. Sie ist u. a. in vielen OP– und HF–Verstärkern enthalten.
Bei der Kaskodenschaltung werden eine Emitter– und eine Basisschaltung (bzw. eine Source– und eine Gateschaltung) in Reihe gesetzt, wodurch der Miller–Effekt der Emitter–Schaltung vermieden wird. Von den Lastschwankungen an $R_\mathrm{C}$ sieht der *Transistor* $\mathrm{T}_1$ praktisch nichts, sein $U_\mathrm{CE}$ bleibt unverändert und damit bleibt auch der Strom durch die Kaskode und durch $R_\mathrm{V}$ konstant.
Ersetzt man in der Prinzipschaltung den Emitterwiderstand $R_\mathrm{E}$ durch einen einfachen Stromspiegel, erhält man den ‘Stromspiegel mit Kaskode’, einen Stromspiegel mit 3 Transistoren. Sein Hauptvorteil ist der erhöhte Ausgangswiderstand.
Eine weitere Vergrösserung des Aussteuerbereichs erfordert komplexere Kaskoden–Stromspiegel mit 4, 5, 6 Transistoren plus Transistoren zur Arbeitspunkteinstellung.
Eine äusserst wichtige Grundschaltung in der integrierten Schaltungstechnik ist

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

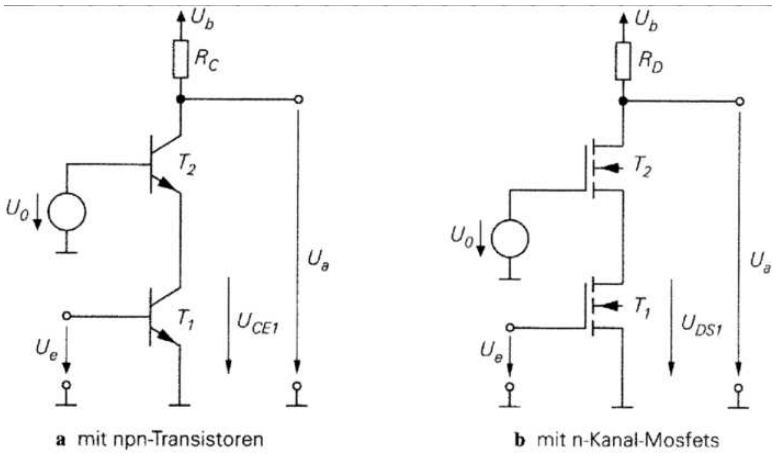



Abbildung 3.129.: Prinzip der Kaskodenschaltung[?].

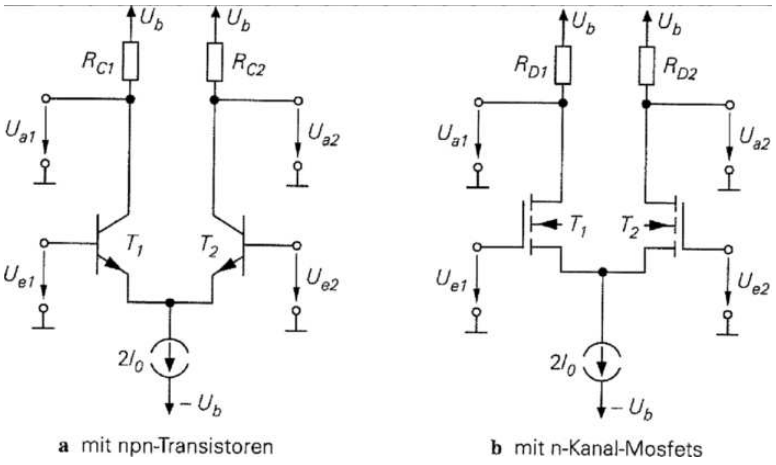



Abbildung 3.130.: Grundschaltung des Differenzverstärkers[?].

der *Differenzverstärker* (differential amplifier, long tailed pair); er bildet die Basis für den Operations–Verstärker.
Seine Grundschaltung ist symmetrisch aufgebaut mit zwei Eingängen und zwei Ausgängen. Zwei Emitterschaltungen (bzw. Source–Schaltungen) sind so angeordnet, dass die Emitter (bzw. Sources) gemeinsam mit einer Stromquelle verbunden sind, siehe Abbildung 3.130. Die Versorgungsspannung ist meist symmetrisch positiv und negativ: $\pm U_\mathrm{B}$.
Bei symmetrischem Aufbau fliesst durch jeden *Transistor* genau der halbe Strom. Erhöht die Spannung an einem Eingang den Strom durch den *Transistor*, so muss — wegen der Konstantstromquelle — der Strom am anderen *Transistor* entsprechend sinken. Die Spannungsänderungen an den 'Lastwiderständen' $R_{\mathrm{C}_1}$ und $R_{\mathrm{C}_2}$ sind also gegensinnig.
Legt man gleichsinnige Signale an die beiden Eingänge gleichzeitig an, so ändern sich zwar die Transistorwiderstände, aber die Konstantstromquelle hält die Ströme konstant und damit bleiben die Spannungsabfälle an $R_{\mathrm{C}_1}$, $R_{\mathrm{C}_2}$, also an den Ausgängen konstant. Die sog. Gleichtaktverstärkung (common mode gain) ist im Idealfall gleich Null; solange der Gleichspannungsanteil hinreichend klein ist, kann jede Signalquelle direkt an den Differenzverstärker angeschlossen werden.
Legt man dagegen eine schiefsymmetrische Differenzspannung $\pm U_\mathrm{D}/2$ an die Ein-

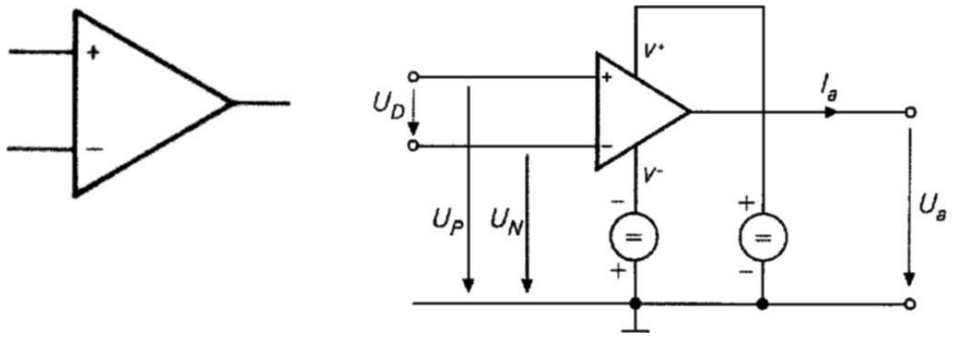



Abbildung 3.131.: Operationsverstärker: Schaltsymbol und Anschlüsse[**?**, **?**].

gänge, so tritt diese verstärkt an den Ausgängen auf (Differenzverstärkung, differential gain). Solange man im Aussteuerbereich bleibt, unterdrückt ein Differenzverstärker also die Gleichtaktsignale und verstärkt die Gegentaktsignale. Die MOSFET–Variante tut dies mit sehr hohem Eingangswiderstand.
Die Übertragungskennlinien zeigen eine relativ schlechte Linearität für die npn–Grundschaltung. Hier hilft in der Praxis wieder die Stromgegenkopplung, z. B. zwei gleiche Emitterwiderstände $R_{\mathrm{E}}$, auf Kosten der Differenzverstärkung.
Liegt an den Eingängen eine Spannungsdifferenz $U_{\mathrm{D}} = 0$ an, so ist beim idealen Differenzverstärker $U_{\mathrm{a}_1} = U_{\mathrm{a}_2}$. Beim realen Bauelement beobachtet man aber $U_{\mathrm{a}_1} \neq U_{\mathrm{a}_2}$, als ob eine endliche Differenzspannung am idealen Verstärker anliegen würde: Offsetspannung. Die Ursache sind unvermeidbare Unsymmetrien (Toleranzen) in der Schaltung. Zur Korrektur sind verschiedene Massnahmen üblich (Korrekturspannung auf einen Eingang, einen Kollektorwiderstand als Potentiometer, gemeinsamer Emitterwiderstand als Poteniometer, etc). Gute Differenzverstärker haben Offsetspannungen kleiner 1 mV und Temperaturdrifts kleiner 1 $\mu$V/K.
Der grosse Vorteil monolithischer Differenzverstärkerschaltungen ist ihre Stabilität bei erstaunlicher Einfachheit; die Güte der Stromquelle (z. B. Temperaturkonstanz) ist massgebend.
*Nebenbemerkung*: Verwendet man statt der Lastwiderstände Stromquellen — einfache Stromspiegel oder Kaskode–Stromspiegel sind üblich —, kann die Differenzverstärkung bzw. das Verstärkungs–Bandbreite–Produkt um etwa eine Grössenordnung gesteigert werden.

# 3.5. Operationsverstärker

## 3.5.1. Grundlagen, Grundtypen, Rückkopplung

Ein Operationsverstärker ist eine integrierte Analogschaltung in SSI– (Small Scale Integration) Technik. Es handelt sich um einen meist mehrstufigen Gleichspannungsverstärker sehr hoher Verstärkung (bis $10^6$); seine Eigenschaften können durch äussere Gegenkopplungs–Beschaltung festgelegt werden. Auf Grund der verfügbaren Vielfalt und zahlreichen Möglichkeiten sind heute in diskreten Schaltungen OPV, aber i. allg. sehr wenige Einzeltransistoren zu finden. Die Grosssignal–Bandbreiten der Standardtypen umfassen den NF–Bereich, die der anderen reichen bis zu einigen 100 MHz, vereinzelt bis ins GHz–Gebiet.
OP–Verstärker besitzen zwei Eingänge, den nichtinvertierenden (auch + – oder

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

P–Eingang) und den invertierenden (– – oder N–Eingang) und einen Ausgang. Die Versorgung erfolgt meist symmetrisch ($V^+$, $V^-$, meist $\pm$ 15 V); die Ruhepotentiale am Eingang und am Ausgang betragen dann 0 Volt. Es gibt keinen speziellen Masseanschluss. Der + –Eingang ist immer hochohmig.
Beim wohlvertrauten *Standard–OPV* (Voltage Feedback Operational Amplifier) ist auch der – –Eingang hochohmig, der Ausgang dagegen niederohmig. Die Eingänge sind spannungsgesteuert, der Ausgang verhält sich wie eine *Spannungsquelle*: *VV–OP* (Voltage–Voltage–OP). Im linearen Arbeitsbereich (Ausgangsaussteuerbarkeit) gilt:

$$U_\mathrm{A} = A_\mathrm{D} \cdot U_\mathrm{D} = A_\mathrm{D}\,(U_\mathrm{p} - U_\mathrm{n}) \quad \text{mit} \quad A_\mathrm{D} = \left.\frac{\mathrm{d}U_\mathrm{A}}{\mathrm{d}U_\mathrm{D}}\right|_{AP} . \tag{3.5.1}$$

Beim idealen OPV wird nur die an die Eingänge angelegte Spannungsdifferenz verstärkt und die Differenzverstärkung $A_\mathrm{D} = \infty$ angenommen; reale Werte liegen bei $10^4$ – $10^6$. Beispiele von Übertragungskennlinien.
Ersetzt man beim VV–OP den V–Ausgang durch eine hochohmige Stromquelle, so erhält man einen *VC–OP* (C=current), den sog. *Transkonduktanz–Verstärker* (Transkonduktanz=Übertragungssteilheit). Analog gilt:

$$I_\mathrm{A} = S_\mathrm{D} \cdot U_\mathrm{D} = S_\mathrm{D}\,(U_\mathrm{p} - U_\mathrm{n}) \quad \text{mit} \quad S_\mathrm{D} = \left.\frac{\mathrm{d}I_\mathrm{A}}{\mathrm{d}U_\mathrm{D}}\right|_{AP} , \tag{3.5.2}$$

der Differenzsteilheit. (Typische Werte von $10^2$.)
Es gibt heute auch OPV mit niederohmigenm stromgesteuertem N –Eingang, siehe Bild unten. Mit einer Stromquelle am Ausgang heisst er *Transimpedanz–Verstärker* (current feedback amplifier) oder auch CV–OP mit einem hochohmigen, stromgesteuerten Ausgang heisst er *Strom–Verstärker* oder CC–OP (Diamond Transistor, ursprünglich eine Firmenbezeichnung von Burr Brown, weil er sich wie ein idealer Transistor verhält). Der Stromübertragungsfaktor beträgt max. 10. Ein normaler VV–OPV wird praktisch nie ohne Gegenkopplung, als nie im 'open–loop' betrieben. Stattdessen koppelt man den Ausgang (grundsätzlich) auf den invertierenden Eingang zurück, sodass das ursprüngliche Eingangsignal verkleinert wird, die Verstärkung also verringert wird. Oder dynamisch gesehen: die Ausgangsspannungsänderung wirkt beim Einschwingvorgang der Eingangsspannungsänderung entgegen. Das Rückkopplungsnetzwerk kann linear oder nichtlinear, kann frequenzunabhängig oder –abhängig gewählt werden; es bestimmt im Wesentlichen die Eigenschaften des so beschalteten OPVs. Ein Rückkopplungsnetzwerk kann (bei hoher Ausgangsimpedanz) als Stromquelle und (bei kleiner Ausgangsimpedanz) als *Spannungsquelle* arbeiten. Allgemein gilt, dass die rückgekoppelte Grösse auch die durch die Rückkopplung verbesserte Grösse ist, z. B. hinsichtlich Stabilität, Linearität, Frequenzgang.
*Nebenbemerkung*: Man kann den gegengekoppelten OPV als Regelkreis auffassen, der OPV selbst arbeitet dabei als Regelstrecke; Führungsgrössenformer und Reglerfunktion werden durch die äussere OPV–Beschaltung gebildet. Die Subtraktion von Soll– und Istwert geschieht entweder durch den Differenzeingang des OPVs oder ebenfalls durch die äussere Beschaltung.
Im einfachsten Fall besteht die äussere Beschaltung aus einem ohmschen Spannungsteiler. Man erhält, siehe Bild oben, die bekannten nichtinvertierenden und

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

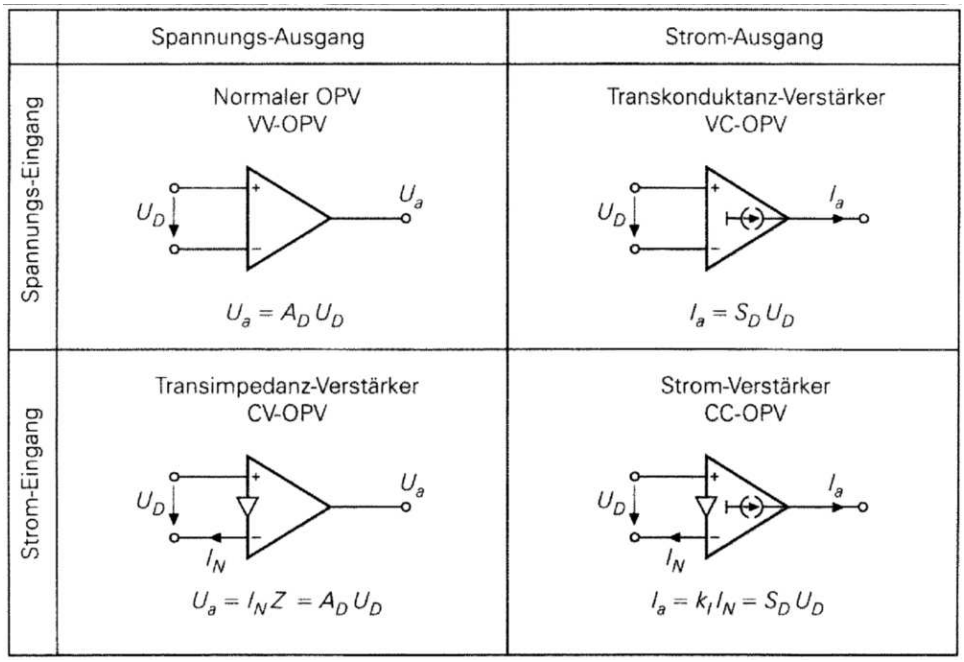



Abbildung 3.132.: OPV–Typen: Schaltsymbole und Übertragungsgleichungen[**?**].

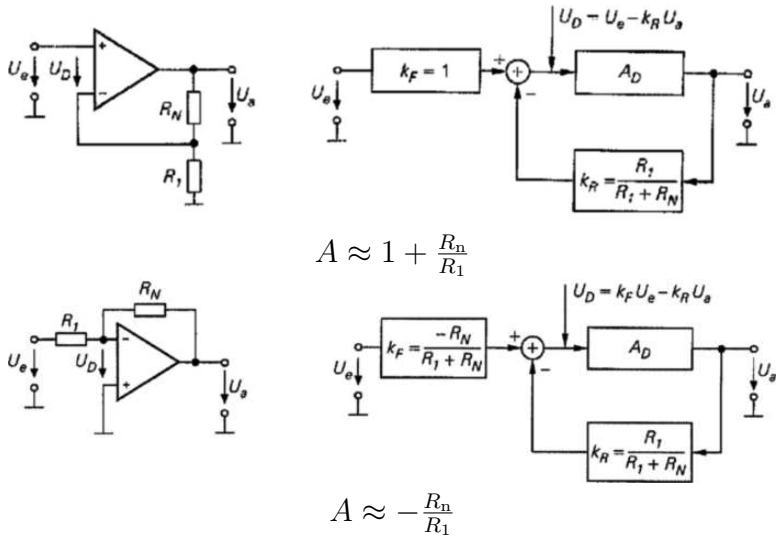



Abbildung 3.133.: VV-OPV mit einfacher ohmscher Gegenkopplung: nichtinvertierender und invertierender OPV[**?**, **?**].

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

invertierenden Standard–OP–Verstärkerschaltungen. Die Leerlaufverstärkung $A_D$ (Differenzverstärkung des nicht rückgekoppelten Verstärkers, open loop gain) wird durch die Gegenkopplung reduziert: $A$ (closed loop gain). Die sog. Schleifenverstärkung $g = A_D/A$ (loop gain) verknüpft beide, ebenso wie der Rückkoplungsfaktor $k_r$ mit $g = k_r \cdot A_D$ bzw. der Kehrwert $1/k_r$ (noise gain).
*Nebenbemerkung*: Vertauscht man beim nichtinvertierenden Verstärker die Eingänge, so erhält man statt der Gegenkopplung eine Mitkopplung. Die Schaltung funktioniert ganz anders, nämlich als invertierender Schmitt–Trigger! Weitere Beispiele gewollter Mitkopplung: Oszillatoren und Kippschaltungen.
Die wichtigsten Regeln ('Goldene Regel') zur Berechnung von OP–Verstärkerschaltungen lauten:

1. Die Ausgangsspannung eines Operationsverstärkers stellt sich so ein, dass die Eingangsspannungsdifferenz Null wird. D. h. bei ausreichend grosser Schleifenverstärkung (idealerweise $A_D = \infty$) liegt bei invertierenden OP–Verstärkerschaltungen der – –Eingang auf der sog. virtuellen Masse.

2. Die Eingänge ziehen keinen Strom.

*Nebenbemerkung*: Zur Berechnung der Verstärkungen der genannten Schaltungen sowie von Addieren und Subtrahierern, siehe Praktikumsanleitung OP–Versuch.
Wenn statt der Gegenkopplung RC–Netzwerke verwendet werden, so erhält man den Integrator, Differentiator oder aktive Filter; nichtlineare Bauelemente (Dioden, Transistoren) ermöglichen Exponierglieder (e–Funktionsgenerator) und Logarithmierer, siehe Praktikumsversuch, sowie komplexere Schaltungen wie Multiplizierer.

## 3.5.2. Standard–Operationsverstärker (VV–OPV)

Allen OPV sind einige Forderungen gemeinsam: Gleichspannungskopplung, Differenzeingang, Eingangs– und Ausgangsruhepotential Null und hohe Spannungsverstärkung. Der Praktiker wünscht sich dabei gute Nullpunktstabilität, einen hohen Eingangswiderstand, einen niedrigen Ausgangswiderstand und einen definierten Frequenzgang.
Diese Forderungen bestimmen den inneren Aufbau des OPVs, wobei sowohl Bipolar– oder/und Feldeffekt–Transistoren eingesetzt werden. Wir beschränken uns hier auf erstere.
Die erstgenannten Forderungen führen direkt zum im vorigen Kapitel eingeführten Differenzverstärker, genauer zu einem Bipolartransistor–Differenzverstärker mit unsymmetrischem Ausgang, dessen Ausgang mit einem weiteren Bipolartransistor als Emitterfolger verstärkt wird. Die Forderung der Gleichspannungskopplung bedingt, dass bei Verwendung eines npn– (pnp–) Transistors zur Verstärkung das Ausgangspotential positiv (negativ) gegenüber dem Eingangspotential verschoben ist. (Erinnerung: das Basisruhepotential einer einfachen Emitterschaltung beträg rund 0,6 V.) Man benötigt also ein weiteres Schaltungselement zur Rückverschiebung des Potentials; es gibt hierzu mehrere Möglichkeiten.

1. Spannungsteiler (schwächt das *Signal*),

2. Z–Dioden (üblich bei npn–Emitterfolgern in HF–Schaltungen),

3. Konstantstromkopplung (lange bevorzugte Bauart),

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

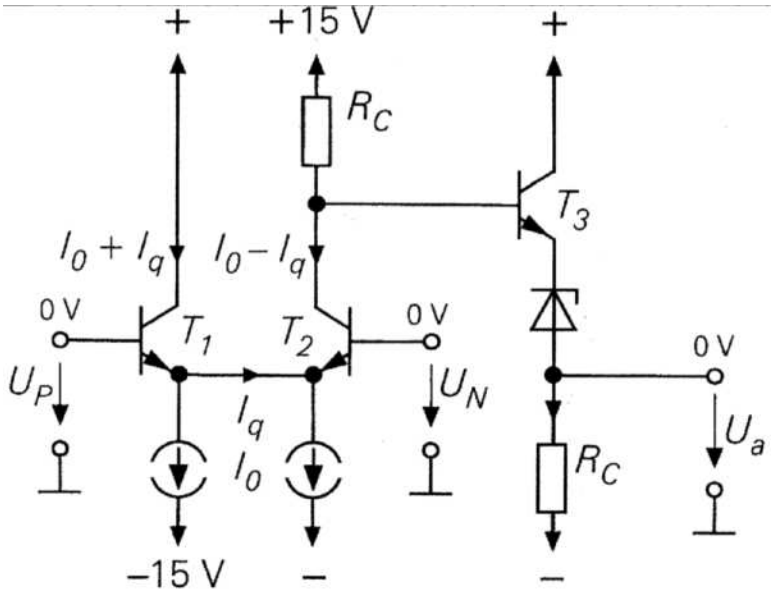



Abbildung 3.134.: Aufbau eines einfachen VV–OPV[**?**].

4. Komplementäre Transistoren (einfachste Möglichkeit, etwas teurer).

Die notwendige hohe Spannungsverstärkung ($A_D = 10^4 \ldots 10^6$) erzielt man meist durch mehrstufige Verstärkung; entsprechend mehrfach hat auch die Potentialrückverschiebung zu sein. (Ebenso hat eine Darlingtonstufe eine Rückverschiebung um 1,2 V nötig.) Diese abgebildete, einfache Schaltung aus npn–Transistoren hat weder die geforderte hohe Differenzverstärkung, noch eine befriedigende Aussteuerbarkeit (Gleichtakt– und Ausgangs–Aussteuerbarkeit).
Die sog. *Universalverstärker* erreichen die hohe Verstärkung durch Hinzufügen einer weiteren Verstärkerstufe, siehe Abbildung 3.135. Dann benötigt man aber in der Regel eine Frequenzgangkorrektur. Die Schaltung besteht aus einem Differenzverstärker mit Stromspiegel. Der Darlingtontransistor $T_5$ entkoppelt die 2. Stufe von $T_2$. Im Abbildung 3.136 sieht man, dass der bekannte Standard–OP 741 ein komplementäres Layout verwendet. Die Endstufe $T_6$, $T_7$ wird, wie bei integrierten OPV immer, als komplementärer Emitterfolger ausgeführt, um positive und negative Ausgangsströme zu ermöglichen, die gross gegen den Ruhestrom sind. Die beiden Dioden erzeugen eine Basisvorspannung, die leicht kleiner ist als die Spannung, bei der die beiden Ausgangstransistoren leitend werden (sog. AB–Betrieb (current on demand)). Der Kondensator $C_K$ wirkt als Miller–Kondensator und dient der Frequenzgangkorrektur. Die Betriebsspannungen sind die des Normalbetriebs.
*Nebenbemerkung*: Breitband–OPV erreichen die hohe Spannungsverstärkung mit nur einer Verstärkerstufe durch Verwendung der Kaskodeschaltung im Stromspiegel, der zur Potentialverschiebung eingesetzt wird.
Operationsverstärker der 741–Klasse sind also mehrstufige Verstärker. Jede Stufe verhält sich wie ein Tiefpass. Dies spiegelt sich im folgenden Bode–Diagramm wieder: Zur Erinnerung: Bei der Grenzfrequenz $f_g = \frac{1}{2\pi RC}$ eines Tiefpasses beginnt die 'Verstärkung' um 20 dB/Dekade oder 6 dB/Oktave abzufallen. Bereits früher setzt eine Phasennacheilung ein; bei der Grenzfrequenz beträgt sie $-45^0$ und wächst asymptotisch auf $-90^0$ an. Abbildung 3.138 gibt die drei wichtigsten

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

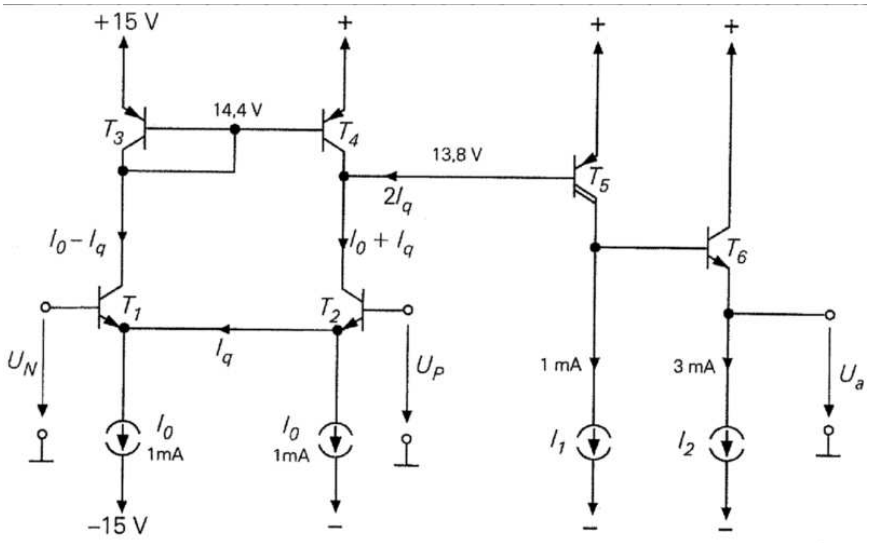



Abbildung 3.135.: Einfacher Universalverstärker: VV–OPV mit zwei Stufen zur Spannungsverstärkung[?].

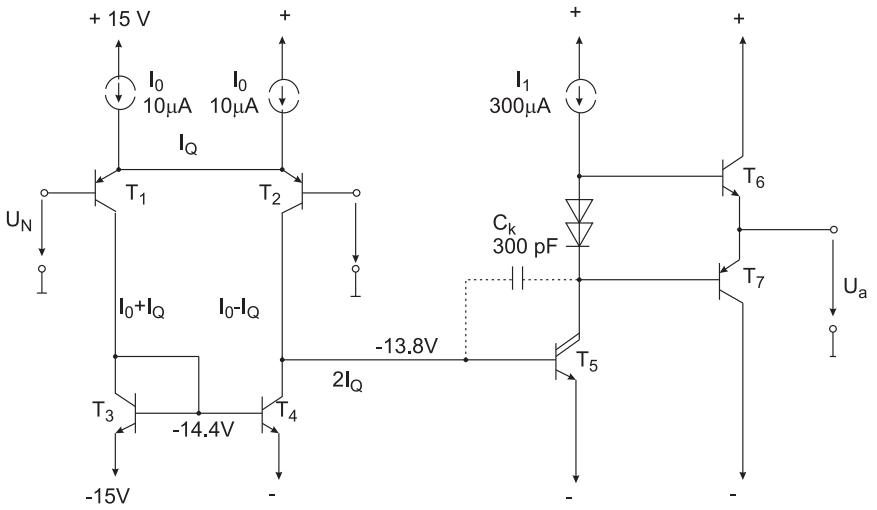



Abbildung 3.136.: Operationsverstärker der 741–Klasse[?].

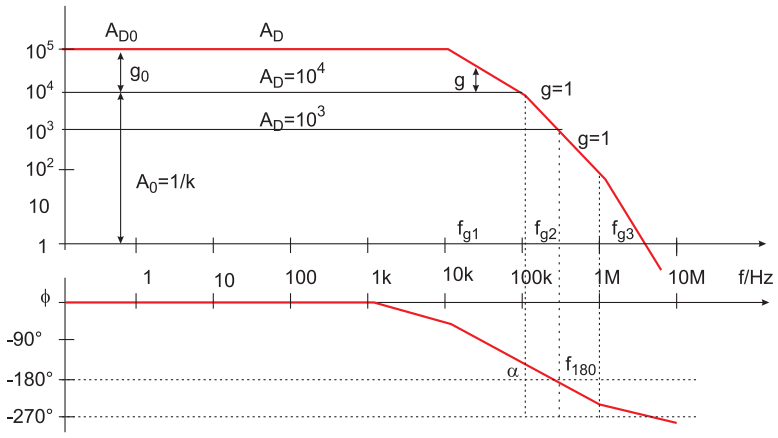



Abbildung 3.137.: Bode–Diagramm eines unkorrigierten OPVs der 741–Klasse[?].

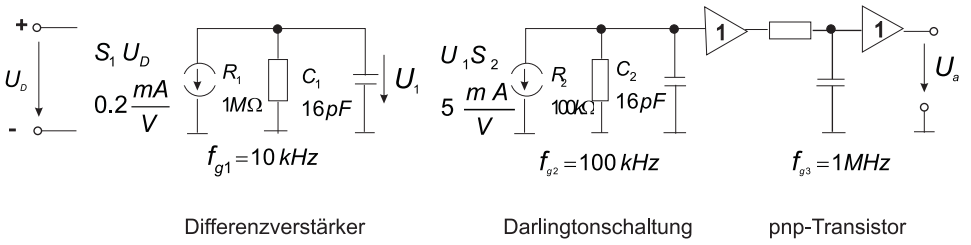



Abbildung 3.138.: Grenzfrequenzen der OPV der 741–Klasse (nach [**?**]).

Grenzfrequenzen unserer Beispiel–OPV–Klasse wieder; sie werden von der Differenzverstärkerstufe, von der Darlingtonverstärkerstufe und – bei preisgünstigen OPV – von minderwertigen pnp–Transistoren verursacht. Oberhalb von 10 kHz sieht man im Bode–Diagramm drei Stufen in der Verstärkungskurve und die Phasenverzögerung wächst ab ca. 1 kHz sukzessive auf $-270^0$.
Koppelt man hypothetisch den Ausgang zurück auf den invertierenden Eingang, dann haben wir für kleine Frequenzen bis ca. 1 kHz perfekte Gegenkopplung. Aus dem Bodediagramm entnimmt man, dass eine Phasenverzögerung von $-180^0$ bei $f_{180} \approx 300$ kHz erreicht wird. D. h. bei $f_{180}$ liegt vollständige Mitkopplung vor. Durch Rückkopplung sinkt auch die Verstärkung (von $A_\mathrm{D}$ auf $A$); die Schleifenverstärkung $g = k_\mathrm{r} \cdot A_\mathrm{D} = A_\mathrm{D}/A$ ist im Bodediagramm gerade der Abstand zwischen Leerlauf– und gegengekoppelter Verstärkung. Dieser verkleinert sich für Frequenzen $> f_{\mathrm{g}_1}$ und die Kurven schneiden sich ($\log g = 0$ oder $g = 1$) bei $f_{\mathrm{g}'}$.
Eine OPV–Schaltung wird instabil, wenn

$$\begin{array}{rrcll} & |k_\mathrm{r}| \cdot |A_\mathrm{D}| & = & 1 & \text{(Amplitudenbedingung)} \\ \text{und} & \phi(k_\mathrm{r} \cdot A_\mathrm{D}) & = & -180^\circ & \text{(Phasenbedingung) .} \end{array} \tag{3.5.3}$$

Bei $f_{\mathrm{g}'}$ sind beide Bedingungen erfüllt, man erhält eine Schwingung mit konstanter Amplitude. Ist bei erfüllter Phasenbedingung $|g| > 1$, so schwingt der Verstärker in die Übersteuerung. Nur wenn $|g| < 1$ bleibt, beobachtet man eine gedämpfte Schwingung. In der Praxis hat man für den Schaltungsentwurf die entsprechende Berechnung auszuführen.
Zur Stabilitätscharakterisierung führt man die sog. *Phasenreserve* $\alpha$ (auch Phasenspielraum, phase margin) ein. Dabei gibt man — bei erfüllter Amplitudenbedingung — den Abstand zur Phasenverschiebung $-180^0$ an:

$$\alpha = 180^\circ - \phi(f_\mathrm{k}) \; . \tag{3.5.4}$$

Die Phasenverschiebung darf noch um den Winkel $\alpha$ zunehmen, bis eine ungedämpfte Schwingung einsetzt. Bei der 'kritischen Frequenz' $f_\mathrm{k}$ ist jeweils die Amplitudenbedingung erfüllt. Im Bild unten sind Einschwingvorgänge für verschiedene Phasenreserven nebst den zugehörigen Frequenzgängen wiedergegeben. Kleine $\alpha$ sind jeweils durch starke Überschwinger gekennzeichnet, bei $\alpha = 60^\circ - 65^\circ$ hat man für die Praxis die günstigsten Werte, für $\alpha = 90^\circ$ liegt der aperiodische Grenzfall (mit seiner verlängerten Anstiegszeit) vor. Im obengezeigten Bodediagramm ist $\alpha \approx 45^\circ$ bei $A \approx 10^4$; ein grösseres $\alpha$ erfordert ein noch grösseres $A$. Umgekehrt gilt für diese unkorrigierten OPV: ein stark rückgekoppelter OPV (kleines $A$) schwingt.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

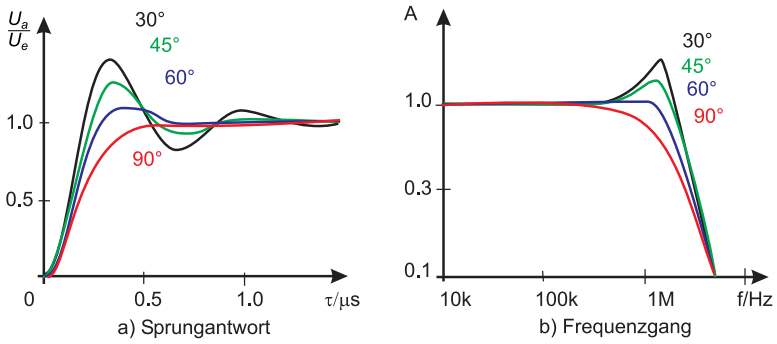



Abbildung 3.139.: Zur Phasenreserve $\alpha$ (nach [?]).

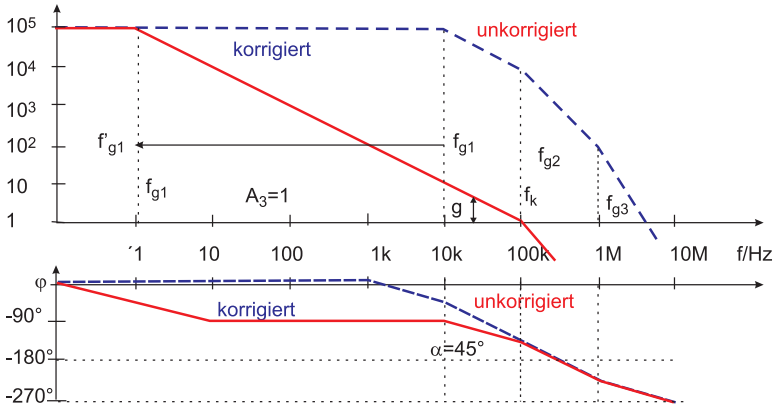



Abbildung 3.140.: Zur universellen Frequenzgangkorrektur der 741–Klasse (nach [?]).

Abhilfe schafft die sog. *Frequenzgang–Korrektur* (Frequenzgang–Kompensation). Hierzu haben einige OPV Extraanschlüsse, z. B. Typ $\mu$A 748, für individuelle Korrekturmassnahmen. Meist ist die Korrektur schon mit integriert. Bei unserem Beispieltyp $\mu$A 741 und vielen anderen Typen wird die sog. *universelle Frequenzgang–Kontrolle* realisiert. Wird bei maximaler Rückkopplung, also Verstärkung $A = 1$ eine Phasenreserve von 90° gefordert, so muss über den kompletten Frequenzgang ein RC–Tiefpassverhalten vorliegen. Hierzu muss die unterste Grenzfrequenz zu kleinen Frequenzen verschoben werden, so dass die zweite Grenzfrequenz gleich der Transitfrequenz und dort $a = 45°$ wird. Der Verstärker ist bei voller Gegenkopplung noch stabil, allerdings wird auch die Verstärkung merklich verändert.
Wird, zusätzlich zur Absenkung von $f_{g_1}$, die zweite Grenzfrequenz $f_g$ erhöht, um den stabilen Arbeitsbereich zu vergrössern, so spricht man von 'Pole Splitting'. Die Miller–Kapazität der 741–Klasse (ca. 30 pF) ist hierfür ein Beispiel.
Hinweis: Das Bodediagramm des kommerziell erhältlichen $\mu$A 741, also eines Frequenzgang–korrigierten Universal–OPVs wird im Praktikum gemessen.
Neben der Reduzierung der Bandbreite und der Verstärkung wirkt sich die Frequenzgang–Korrektur auch auf die maximale Anstiegsgeschwindigkeit der Ausgangsspannung, die sog. *Slew–Rate*, negativ aus. Da der Ausgangsstrom des Differenzverstärkers begrenzt ist, kann der Korrekturkondensator nur in endlicher Zeit umgeladen werden und die Ausgangsspannung kann sich typischerweise um $0,6$ V/$\mu$s ändern. Bei Verstärkern der 741–Klasse wird bei Frequenzen oberhalb ca. 10 kHz die volle Ausgangsspannung nicht mehr erreicht und Signalformen erscheinen verzerrt. Ein

Kompromiss ist die 'Teilkorrektur', die eine kleinere Korrekturkapazität verwendet.
*Kapazitive Lasten* bilden zusammen mit dem OPV–Ausgangswiderstand einen weiteren RC–Tiefpass. Dieser kann sich in relevanter Weise im Bodediagramm bemerkbar machen. Es gibt spezielle, intern Last–korrigierte OPV im Handel.

### 3.5.3. Transkonduktanz–Verstärker (VC–OPV)

Beim *Transkonduktanzverstärker* (Operational Transconductance Amplifier *OTA*) sind die beiden Eingänge — wie beim besprochenen Standard–OPV — spannungsgesteuert, der Ausgang aber ist hochohmig und verhält sich wie eine Stromquelle (spannungsgesteuerte Stromquelle).
VC–OPV eignen sich besonders zum Treiben von Koaxialleitungen, deren Wellenwiderstände klein gegen ihren Ausgangswiderstand sind. I. allg. zeigen sie bei kapazitiven Lasten keine Stabilitätsprobleme. Ihre Grosssignal–Bandbreiten sind beachtlich, z. B. 200 MHz.
Man kann jeden VV–OPV in einen VC–OPV umwandeln, indem man den Emitterfolger am Ausgang weglässt. Ein einfacher innerer Aufbau besteht also aus der Differenzverstärkerschaltung und vier Stromspiegeln (z. B. LM 13 700 von National Semiconductor, CA 3080 von Harris). Die Übertragungssteilheit (Transconductance) kann bei neueren Typen über den Versorgungsstrom ($I_{\mathrm{Bias}}$) am Emitterstromspiegel des Differenzverstärkers über mehrere Grössenordnungen von aussen eingestellt werden, d. h. der maximale Ausgangsstrom kann so festgelegt werden. Der hohe Ausgangswiderstand wird duch Transistoren in Emitterschaltung realisiert.
In den Standardbüchern findet man praktisch keine Anwendungsschaltungen; bei den Herstellern gibt es aber ausführliche Datenblätter, die auf den Webseiten leicht zu finden sind.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

# 4. Sensoren und Messverfahren

In diesem Kapitel werden die grundlegenden *Messverfahren* für Spannungen, Ströme, und viele weitere elektrische oder magnetische Signale besprochen. Insbesondere wird Wert auf die Diskussion aktueller elektronischer Schaltkreise gelegt. Die gezeigten Simulationen wurden mit dem *Elektronik-Design-Labor*[**?**] berechnet.

## 4.1. Basismessverfahren

Zu den Basismessverfahren zählen die Messung von Strom und Spannung, sowohl im Falle von Gleichstrom und -spannung, wie auch für Wechselstrom und -spannung.

### 4.1.1. Strom

Ströme können sowohl mit Spannungsmessern als auch mit Strommessern oder elektronischen Mitteln bestimmt werden. Für Gleichströme zeigt Abbildung 4.1 die Schaltung, unter Berücksichtigung des endlichen Ausgangswiderstandes $R_1$ der Stromquelle und des endlichen Innenwiderstandes $R_2$ des Strommessers. Wenn durch $R_j$ der Strom $I_j$ fliesst, so gilt für den Quellstrom $I_S = I_1 + I_2$. Weiter müssen die Spannungsabfälle an $R_1$ und $R_2$ gleich sein, da ja der Strommesser ideal sein soll.

$$I_2 = I_S - I_1 = \frac{U}{R_2} = \frac{R_1 R_2}{R_1 + R_2}\frac{1}{R_2} I_S = I_S \cdot \frac{R_1}{R_1 + R_2} = \frac{I_S}{1 + \frac{R_2}{R_1}} \tag{4.1.1}$$

Gleichung (4.1.1) zeigt, dass der Messfehler umso kleiner ist, je grösser der Ausgangswiderstand der Quelle und je kleiner der Widerstand des Messwerkes ist. Die beiden Anordnungen in Abbildung 4.1 unterscheiden sich im Innenwiderstand $R_2$ des Messwerkes. Rechts ist der Innenwiderstand kleiner, der Fehler also auch kleiner.
Die untere Zeile der Abbildung 4.1 zeigt den Aufbau eines analogen Drehspulenstrommessers. Die Spule bewegt sich in einem engen Spalt zwischen dem zylinderförmigen Südpol und den aussen liegenden Nordpolen. Im Spalt wird, ähnlich wie bei einem Plattenkondensator ein in guter Näherung homogenes Magnetfeld erzeugt. Der Strom, der durch die Spule (rot) fliesst bewirkt ein Drehmoment aufgrund der Lorentzkraft. Die Spiralfeder erzeugt ein rückstellendes Richtmoment, so dass das Drehmoment aufgrund des Stromes mit dem Drehmoment der Spiralfeder verglichen wird. In guter Näherung ist das rückstellende Drehmoment der Spiralfeder proportional zum Auslenkungswinkel.
In der Abbildung 4.2 wird der Strom $I_S$ mit einem Spannungsmesser bestimmt. Ein idealer Spannungsmesser hat den Innenwiderstand $\infty$. Der Widerstand $R_1$ ist wieder der Ausgangswiderstand der Stromquelle. $R_2$ ist der *Messwiderstand*: er

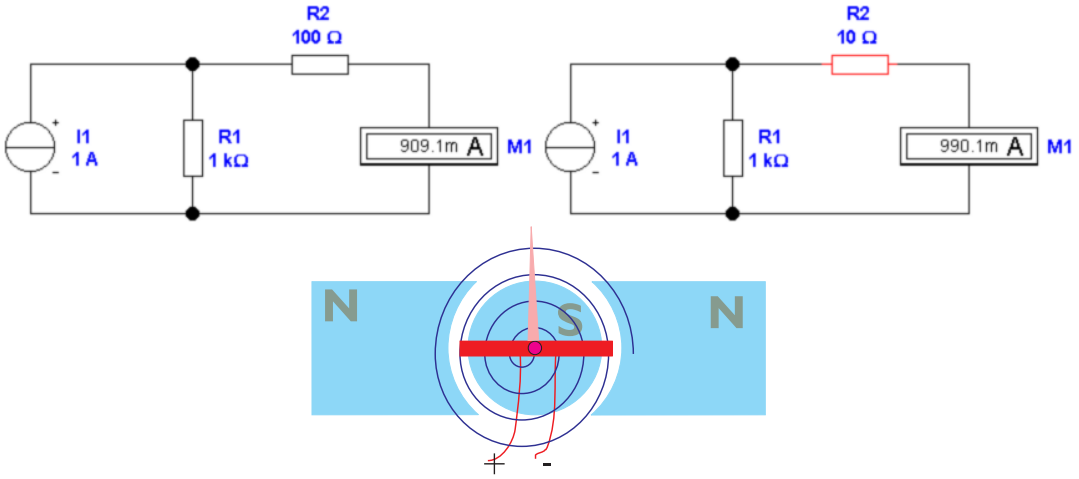



Abbildung 4.1.: Prinzipielle Schaltung für die Messung des Ausgangsstroms einer Stromquelle mit einem Strommesser. Die beiden Darstellungen unterscheiden sich durch den Innenwiderstand. Die untere Zeile zeigt das Prinzipbild eines analogen Strommessers.

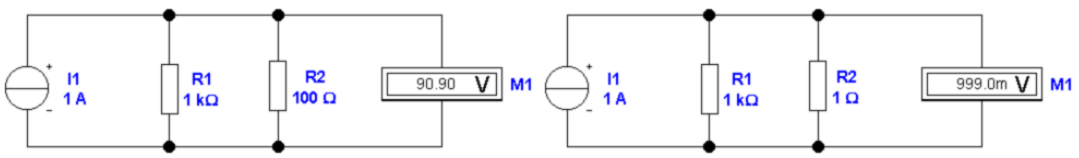



Abbildung 4.2.: Prinzipielle Schaltung für die Messung des Ausgangsstroms einer Stromquelle mit einem Spannungsmesser. Die beiden Darstellungen unterscheiden sich durch den *Messwiderstand* $R_2$.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

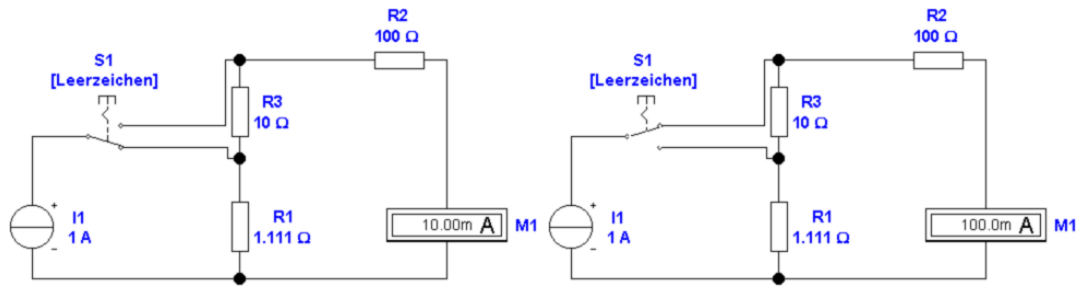



Abbildung 4.3.: Prinzipielle Schaltung für die Messung des Ausgangsstroms einer Stromquelle mit einem umschaltbaren Strommesser. Die beiden Darstellungen zeigen beispielhaft zwei Messbereiche.

enthält, implizit, den *Eingangswiderstand* des Spannungsmessers. Der Ausgangsstrom der Stromquelle verteilt sich wieder auf die beiden Widerstände $R_1$ und $R_2$ nach dem Gesetz $I_S = I_1 + I_2$. Gemessen wird

$$\begin{aligned} U_2 &= I_2 R_2 & &= (I_S - I_1)\, R_2 = I_S R_2 - U_2 \frac{R_2}{R_1} \\ U_2 &= I_S \frac{R_2}{1 + \frac{R_2}{R_1}} & &= I_S \frac{R_1 R_2}{R_1 + R_2} \end{aligned} \tag{4.1.2}$$

Wieder ist ersichtlich, dass der Fehler minimal wird, wenn der Ausgangswiderstand $R_1$ der Stromquelle gross gegen den *Messwiderstand* $R_2$ ist. Die beiden Darstellungen in Abb. 4.2 unterscheiden sich durch den Wert des Messwiderstandes. Aus $U_2$ wird mit $I_2 = \frac{U_2}{R_2}$ auf den Strom geschlossen.
In Abb. 4.3 ist dargestellt, wie mit einer Umschaltung Messbereiche eingestellt werden können. Hier ist $R_2$ der Innenwiderstand des Messwerkes. In der Abbildung 4.3 ist $R_3$ in Serie dazu geschaltet und $R_1$ ist der *Messwiderstand*. In der Abbildung rechts ist der *Messwiderstand* $R_1 + R_3$, der Innenwiderstand des Messwerkes wird allein durch $R_2$ gebildet. Allgemein gilt für den Messstrom $I_2$

$$I_2 = I_S \frac{\sum_{\text{Widerstände in Serie zu} R_1} R_i}{\sum_{\text{Alle Widerstände}} R_j} \tag{4.1.3}$$

Mit *„Alle Widerstände“* sind alle gemeint, einschliesslich des Innenwiderstands des Messwerkes. In Abb 4.3 sind die Werte so berechnet worden, dass links der Messstrom ein hundertstel und rechts ein Zehntel des Quellstromes ist.
In Abb. 4.4 wird die Spannung am *Messwiderstand* $R_1$ mit einem Spannungsmesser mit umschaltbaren Bereichen gemessen. Die Widerstände $R_3$ und $R_4$ werden in Serie zu $R_2$ geschaltet. Der Quellstrom $I_S = I_1 + I_2$ setzt sich aus dem Strom $I_1$ durch den *Messwiderstand* $R_1$ und dem Strom $I_2$ durch das Messwerk zusammen. Es gilt für $I_2$

$$I_2 = I_S - I_1 = I_S - \frac{U}{R_1} = I_S - \frac{1}{R_1} \frac{I_S}{\frac{1}{R_1} + \frac{1}{R_{3,4} + R_2}} = I_S \frac{1}{1 + \frac{R_2 + R_{3,4}}{R_1}} \tag{4.1.4}$$

Somit ist klar, dass mit $R_{3,4}$ die Empfindlichkeit umgeschaltet werden kann. Nachteilig ist, dass der *Messwiderstand* konstant bleibt, dass also die Verlustleistung $P = I_1 R_1^2$ extrem hoch werden kann.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

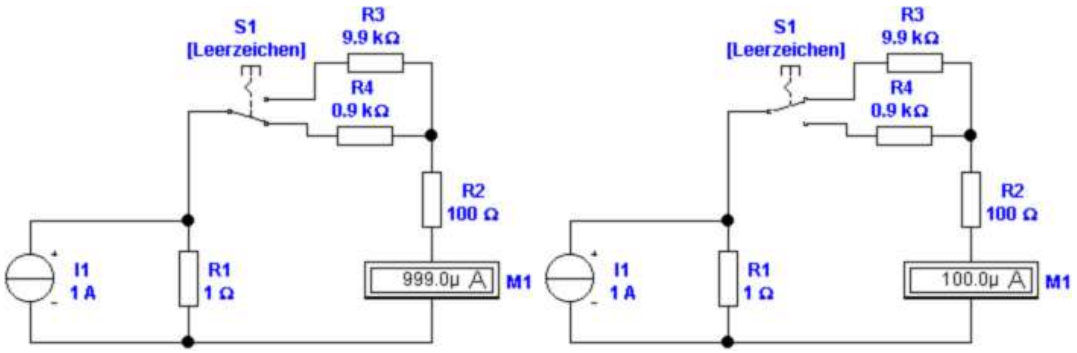



Abbildung 4.4.: Prinzipielle Schaltung für die Messung des Ausgangsstroms einer Stromquelle mit einem Strommesser mit umschaltbarem Innenwiderstand. Die beiden Darstellungen unterscheiden sich durch Messbereich.

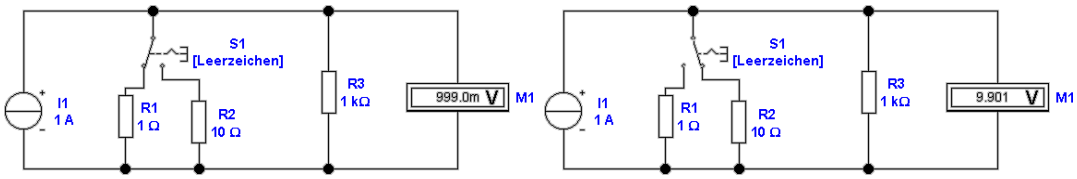



Abbildung 4.5.: Prinzipielle Schaltung für die Messung des Ausgangsstroms einer Stromquelle mit einem umschaltbaren *Messwiderstand* und einem Spannungsmesser. Die beiden Darstellungen unterscheiden sich durch den Innenwiderstand.

Die Schaltung in Abb. 4.5 ist die in der Messtechnik gebräuchliche. $R_3$ ist der Innenwiderstand des Spannungsmessers. Für die gemessene Spannung $U$ gilt:

$$U = I_S R_{eff} = I_S \frac{1}{\frac{1}{R_{1,2}} + \frac{1}{R_3}} = I_S \frac{R_{1,2} R_3}{R_{1,2} + R_3} \tag{4.1.5}$$

Da der Spannungsbereich fest ist, nimmt die Verlustleistung an den Messwiderständen $R_{1,2}$ nur linear zu, anders als in den Schaltungen der Abb. 4.4.
In Abb. 4.6 wird die Messung eines Stromes mit der Kompensationsmethode dargestellt (Äquivalent zur *Poggendorffschen Spannungskompensation*). Die linke Seite der Abbildung zeigt einen nicht-abgeglichenen Zustand, die rechte Seite ist abgeglichen.
Kleine Ströme werden heute meistens mit Strom-Spannungswandler-Schaltungen gemessen. Abb. 4.7 zeigt solche Schaltungen. Oben links ist die Schaltung mit

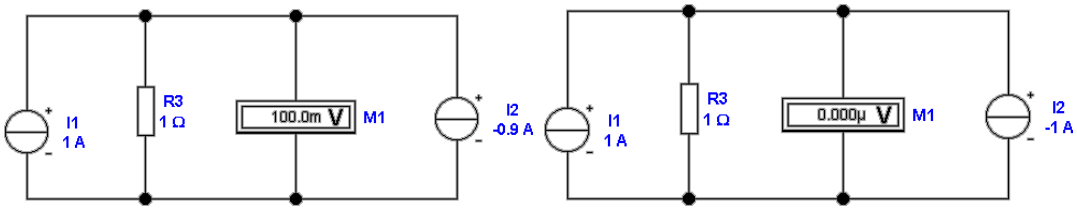



Abbildung 4.6.: Prinzipielle Schaltung für die Messung des Ausgangsstroms einer Stromquelle mit der Kompensationsmethode.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

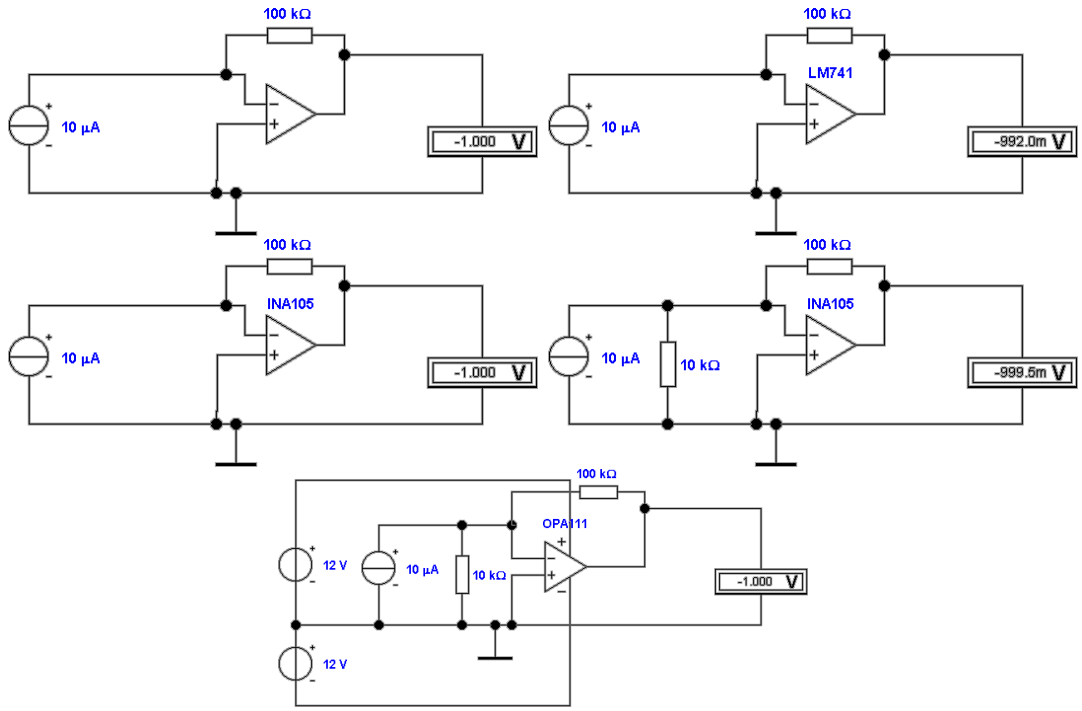



Abbildung 4.7.: Prinzipielle Schaltung für die Messung des Ausgangsstroms einer Stromquelle mit Operationsverstärkern.

einem idealen Operationsverstärker aufgebaut. Es gilt:

$$U_{aus} = -RI_{ein} \tag{4.1.6}$$

In der Abbildung ist $R = 100k\Omega$ und $I_{ein} = 10\mu A$. Entsprechend ist die Ausgangsspannung $U_{aus} = 1V$. Oben rechts ist die gleiche Schaltung mit einem Operationsverstärker *LM741*. Dieses Bauteil hat relativ grosse Bias-Ströme, die zu den Eingangsströmen dazu zu zählen sind. Der resultierende Fehler ist immerhin 0.8%. In der zweiten Reihe links ist die Schaltung mit dem Verstärker *INA105* aufgebaut. Dieser Typ hat sehr viel kleiner Eingangsströme. Ist der Ausgangswiderstand der Stromquelle jedoch relativ klein, wie in der zweiten Reihe rechts gezeigt, dann ist der Ausgang des *INA105* ebenfalls fehlerbehaftet. Hier hilft, wie unten in Abb. 4.7 gezeigt, ein Präzisionsverstärker mit Eingangsströmen im fA-Bereich, der *OPA111*. Natürlich könnte man, wie im vorangegangen Kapitel besprochen, die BIAS-Kompensationstechniken anwenden.
Abb. 4.8 zeigt einen erdfreien Strom-Spannungswandler. Die beiden Eingangsverstärker sind als Stromwandler geschaltet, wobei die Rückkopplung zum jeweils anderen Operationsverstärker geht. Die Widerstände R8 und R9 bilden Die Rückkopplung der beiden Verstärker $AR4$ und $AR5$. Deren invertierende Eingänge sind zusammengeschlossen. Deshalb sind ihre beiden Ausgänge dem Betrage nach auf der gleichen Differenzspannung, aber mit umgekehrtem Vorzeichen. Der Rückkopplungsstrom des einen Verstärkers wird vom Ausgang des anderen Verstärkers aufgebracht. Deshalb ist es richtig, dass die Strom-Spannungswandlerschaltung den nichtinvertierenden Eingang verwendet. Diese Schaltung hat eine Verstärkung von 1 für Gleichtaktsignale. Der abschliessende Differenzverstärker unterdrückt das Gleichtaktsignal. Hier sollte ein Typ eingesetzt werden, der eine gute Gleicht-

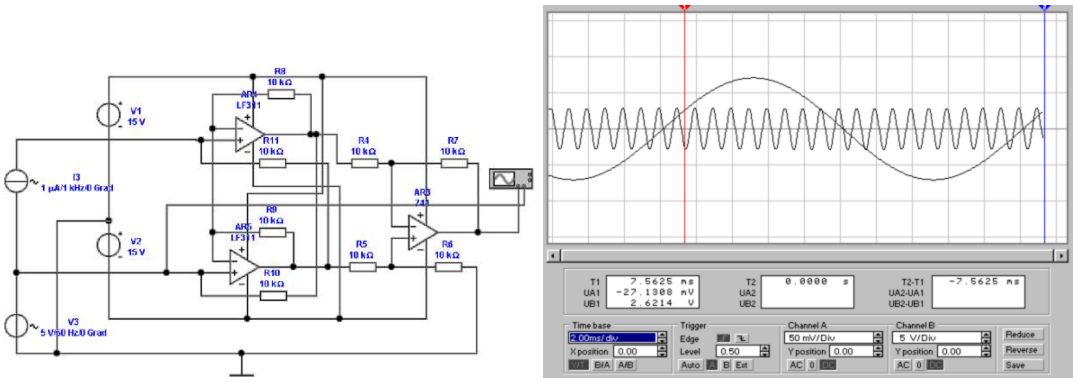


Abbildung 4.8.: Erdfreie Präzisionsmessung des Stromes mit einem Instrumentenverstärker-artigen Strom-Spannungs-Wandler. Links ist die Schaltung dargestellt, rechts zeigt ein Oszilloskopbild, dass Ströme von wenigen $\mu$A über Gleichtaktspannungen von einigen Volt messbar sind.

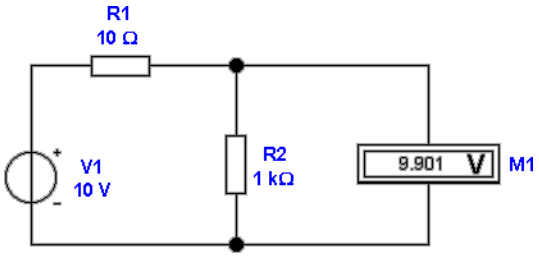



Abbildung 4.9.: Prinzipielle Schaltung für die Messung der Ausgangsspannung einer *Spannungsquelle* mit einem Spannungsmesser.

aktunterdrückung hat. Die Ausgangsspannung ist

$$U_a = 2RI \tag{4.1.7}$$

wenn die vier Widerstände der beiden Eingangsverstärker alle gleich sind.

## 4.1.2. Spannung

Abbildung 4.9 zeigt die Messung einer Spannung. Der Innenwiderstand der *Spannungsquelle* beträgt 10 Ohm. Die Spannung wird mit einem Innenwiderstand von einem Kiloohm gemessen. Durch die Belastung der *Spannungsquelle* mit dem *Messgerät* wird ein Fehler von etwa einem Prozent erzeugt. Die gemessene Spannung ist also:

$$U_{mess} = U_{Quelle} \frac{R_{mess}}{R_{mess} + R_{Quelle}} \tag{4.1.8}$$

Abbildung 4.10 zeigt wieder eine Spannungsmessung, diesmal aber mit einem Strommesser. Der Vorwiderstand des *Messgerätes* ist der gleiche wie in Abbildung 4.9.Da die *Spannungsquelle* den gleichen Innenwiderstand wie vorher hat, ist Messfehler auch gleich. Man könnte ihn verringern, indem man den vor Widerstand

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

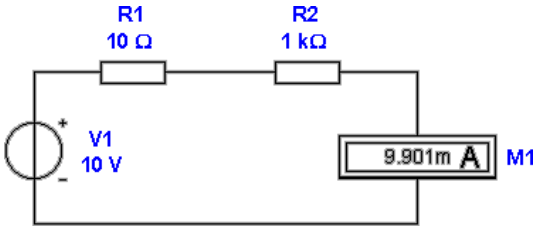



Abbildung 4.10.: Prinzipielle Schaltung für die Messung der Ausgangsspannung einer *Spannungsquelle* mit einem Strommesser.

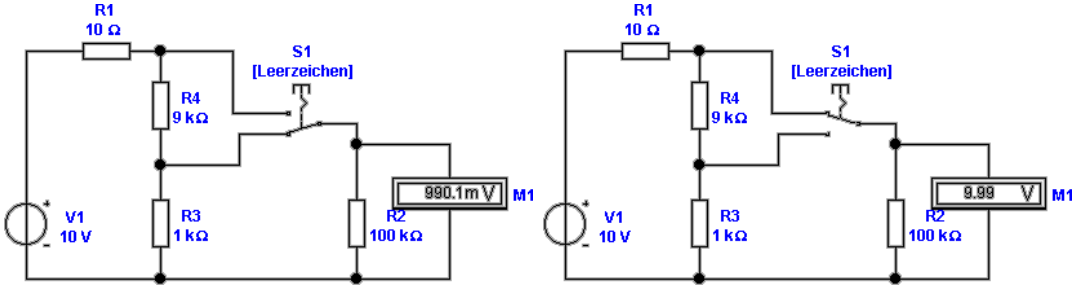



Abbildung 4.11.: Prinzipielle Schaltung für die Messung der Ausgangsspannung einer *Spannungsquelle* mit einem Spannungsmesser und ein Teilerkette.

vergrössert. Dadurch fliesst geringerer Strom durch das Messwerk.

$$I_m = \frac{U_{Quelle}}{R_{Quelle} + R_{Messwerk}} \tag{4.1.9}$$

Gerade bei den *analogen Messgeräten*, bei denen das Magnetfeld eine Kraft gegen eine Feder aufbringen muss, wird ein minimaler Strom benötigt. In der Praxis ist also eine Spannungsmessung eine Optimierungsaufgabe.
Abbildung 4.11 zeigt die Messung mit einem umschaltbaren Spannungsmesser. Der Innenwiderstand der Quelle beträgt wieder 10 Ohm. Eine Widerstandskette bestehend aus dem 9 Kiloohm Widerstand und dem 1 Kiloohm Widerstand fungiert als Spannungsteiler. In unserem Fall hat das *Messgerät* einen Innenwiderstand von 100 Kiloohm. Der Innenwiderstand des *Messgerätes* muss gross sein gegenüber dem Gesamtwiderstand der Spannungsteiler-Kette. Dann ist der Messfehler vernachlässigbar. In unserem Falle beträgt er etwa ein Promille. Die *Spannungsquelle* wird mit

$$R_{eff} = \sum_{Widerstände\,über\,dem\,Abgriff} R + \frac{1}{\frac{1}{R_i} + \frac{1}{\sum_{Widerstände\,unter\,dem\,Abgriff} R}} \tag{4.1.10}$$

belastet. $R_i$ ist der Innenwiderstand des Spannungsmessers.
Abbildung 4.12 zeigt die Spannungsmessung mit einem Strommesser und schaltbaren Vorwiderständen. Für den Messstrom gilt:

$$I_m = \frac{U}{R_i + \sum_{Eingeschaltete\,Widerstände} R} \tag{4.1.11}$$

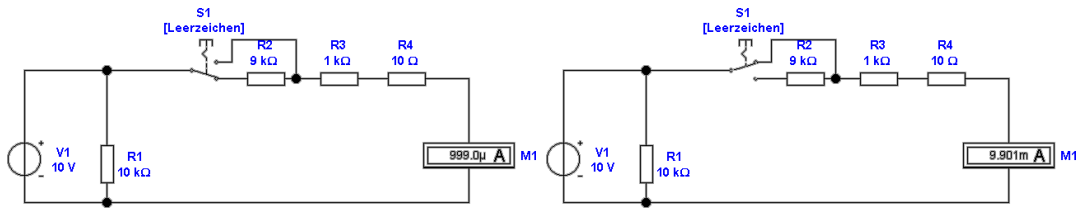



Abbildung 4.12.: Prinzipielle Schaltung für die Messung der Ausgangsspannung einer *Spannungsquelle* mit einem Strommesser und schaltbaren Vorwiderständen.

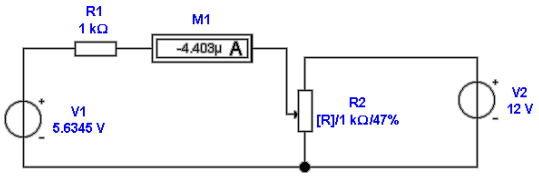



Abbildung 4.13.: Prinzipielle Schaltung für die Messung der Ausgangsspannung einer *Spannungsquelle* mit der Poggendorff'sche Kompensationsmethode.

Hier ist $R_i$ der Innenwiderstand des Strommessers und $U$ die angelegte Spannung. Die *Spannungsquelle* wird mit

$$\frac{1}{R_{eff}} = \frac{1}{R_{mess}} + \frac{1}{R_i + \sum\limits_{Eingeschaltete\,Widerstände} R} \tag{4.1.12}$$

belastet. $R_mess$ ist in Abb. 4.12 der 10 kΩ-Widerstand. Die Schaltung ist prinzipiell gleich wie in der Abbildung 4.11. Wieder gilt, dass der Strom durch das Messwerk genügend gross sein muss, damit die Mechanik vernünftig ansprechen kann.
Abbildung 4.13 zeigt die Poggendorff'sche Kompensationsmethode. Diese Methode, die auch im Praktikum angewandt wird, vergleicht die Spannung der Quelle mit einer Referenzquelle.Die Referenzquelle ist abstimmbar. Ihre Spannung wird so lange verändert, bis sie gleich der der zu messenden Quelle ist.

$$U_{Quelle} = \alpha R * U_{ref} \tag{4.1.13}$$

Hier ist $\alpha$ den Teiler, den man an $R$ einstellt. Im abgeglichenen Falle wird die Quelle nicht mit einem Strom belastet. Man ist also die unbelastete Ausgangsspannung.
Abbildung 4.14 zeigt, wie man Spannungen mit Operationsverstärkern messen kann. Für die Ausgangsspannung im idealen Falle gilt:

$$U_{aus} = U_{ein}\left(1 + \frac{R_2}{R_1}\right) = U_{ein}\left(1 + \frac{9k\Omega}{1k\Omega}\right) = 10U_{ein} \tag{4.1.14}$$

Oben links ist eine ideale Schaltung angegeben. Der Verstärker verstärkt das *Signal* um den Faktor 10. Die *Spannungsquelle* hat einen Innenwiderstand von 100 Kiloohm. Ihr Spannungswert von 10 mV wird durch den Verstärker auf 100 mV erhöht. oben rechts ist die gleiche Schaltung mit einem Verstärkern **LM741** aufgebaut. Dieser Verstärker hat relativ grosse Eingangsströme. Deshalb ist die gemessene Spannung am Ausgang um 80 Prozent falsch. Wenn der Strom $I_{Bias}$ in den positiven Eingang fliesst, dann gilt:

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

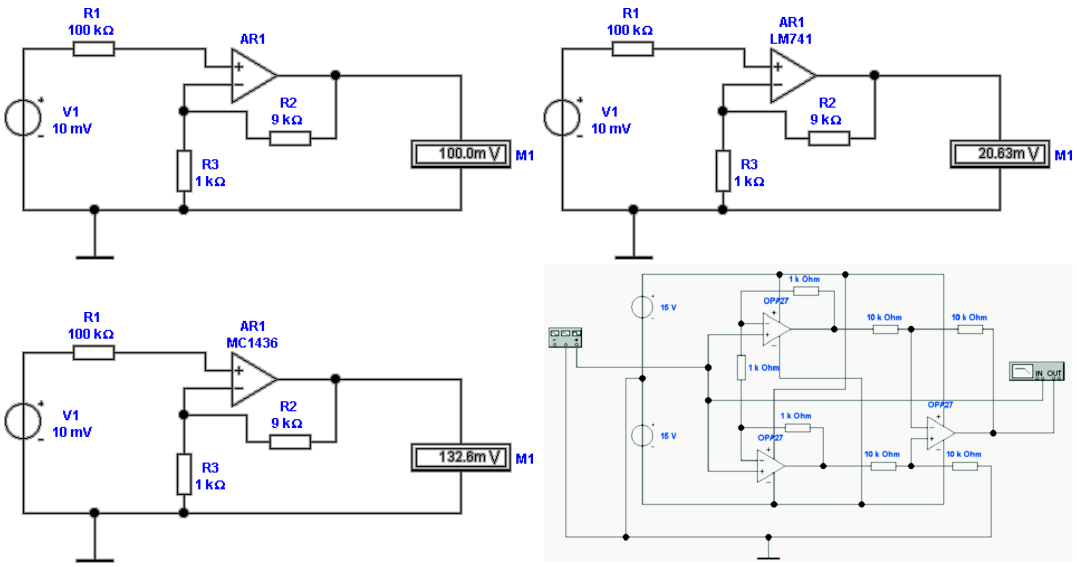



Abbildung 4.14.: Prinzipielle Schaltung für die Messung der Ausgangsspannung einer *Spannungsquelle* mit Operationsverstärkern.

$$\begin{aligned} U_{aus} &= (U_{ein} - I_{Bias} R_i) \left(1 + \frac{R_2}{R_1}\right) \\ &= (U_{ein} - I_{Bias} R_i) \left(1 + \frac{9k\Omega}{1k\Omega}\right) = 10\,(U_{ein} - I_{Bias} R_i) \end{aligned} \tag{4.1.15}$$

wobei $R_I$ der Innenwiderstand der Quelle ist. Hier kann man übrigens ausrechnen, dass $I_{Bias} = 0.8\mu A$ ist. Aus der Richtung schliesst man weiter, dass die Eingangstransistoren NPN-Transistoren sind. In der zweiten Zeile ist die Schaltung mit dem Schaltkreis **MC1436** aufgebaut. Bei diesem Schaltkreis fliessen die Eingangsströme in umgekehrter Richtung. Deshalb ist hier die gemessene Spannungen um 30 Prozent zu hoch. Hier berechnet man, dass $I_{Bias} = -0.3\mu A$ ist. Aus der Richtung schliesst man weiter, dass die Eingangstransistoren PNP-Transistoren sind. Unten rechts ist die Schaltung mit dem Verstärker **INA105** gezeigt. Dieser Präzisionsverstärker hat einen Fehler von 0,5 Prozent. Es zeigt sich, dass genaue Spannungsmessungen bei grossem Innenwiderstand der Quelle nur mit Präzisionsverstärkern möglich ist. Geld sparen lohnt sich hier nicht.
Abbildung 4.15 zeigt einen Instrumentenverstärker. Die beiden Operationsverstärker am Eingang (links) sind als nichtinvertierende Verstärker geschaltet. Da die Referenz nicht das Erdpotential ist, sondern der jeweils andere Verstärker, ist die Gleichtaktverstärkung 1. Der folgende Differenzverstärker unterdrückt weiter das Gleichtaktsignal. Dieser dritte Operationsverstärker muss eine gute Gleichtaktunterdrückung besitzen.

### 4.1.3. Wechselstrom und Wechselspannung

Die Messung von Wechselspannungen aus Quellen mit hohen Impedanzen ist eine schwierige Aufgabe. Solche Quellen können unter anderem Photodioden oder die Tunnelübergänge in einem STM (Scanning Tunneling Microscope) sein. Insbe-

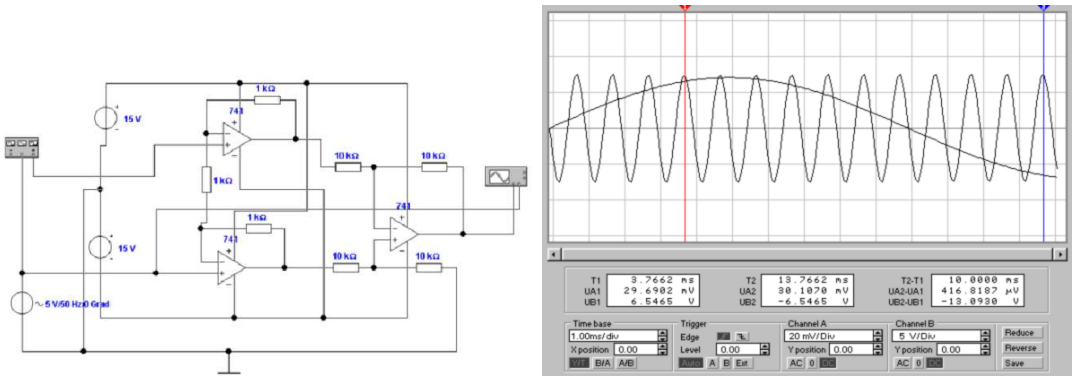



Abbildung 4.15.: Spannungsmessung mit einem Instrumentenverstärker. Im rechten Bild ist dargestellt, dass sogar Verstärker wie der LM741 bei 1 kHz ein *Signal* mit einer Amplitude von 10 mV von einem 6 V, 50 Hz Gleichtaktsignal trennen können.

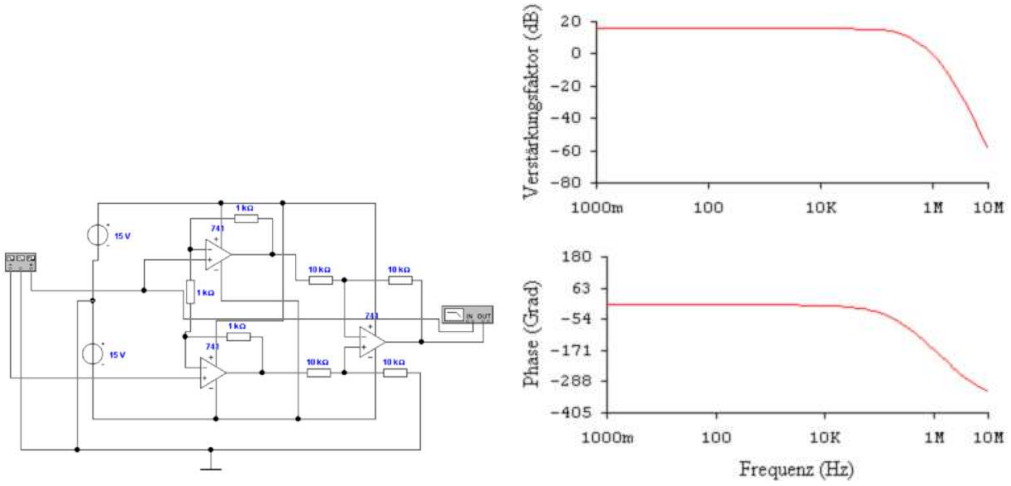



Abbildung 4.16.: Frequenzgang der Gegentaktverstärkung an einem Instrumentenverstärker. Links ist die Schaltung dargestellt, rechts der Bodeplot.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

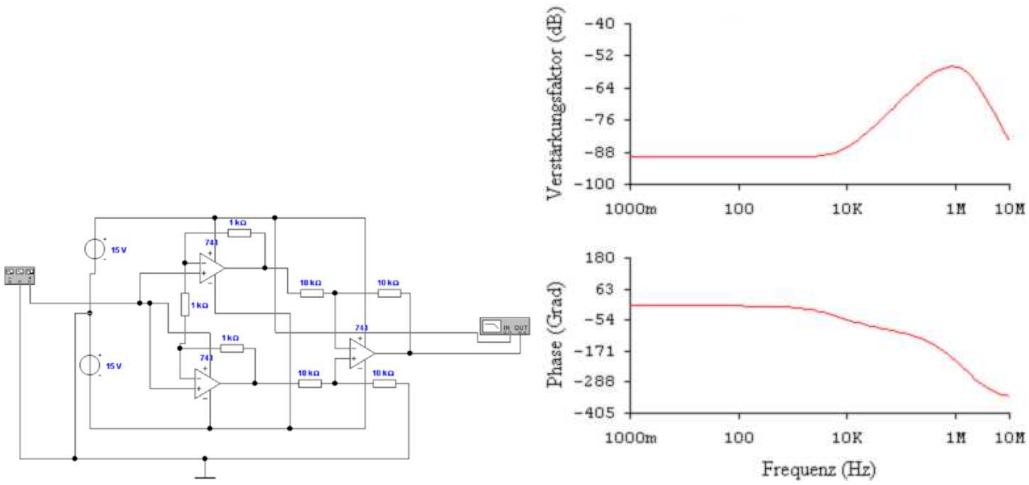



Abbildung 4.17.: Frequenzgang der Gleichtaktverstärkung an einem Instrumentenverstärker. Links ist die Schaltung dargestellt, rechts der Bodeplot.

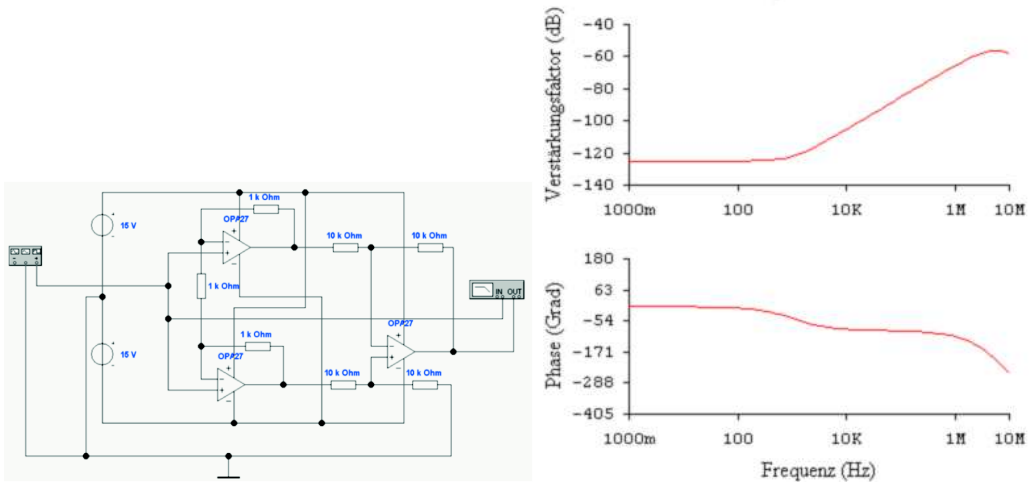



Abbildung 4.18.: Frequenzgang der Gleichtaktverstärkung an einem Instrumentenverstärker mit hochwertigen Operationsverstärkern vom Typ OP27 . Links ist die Schaltung dargestellt, rechts der Bodeplot.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

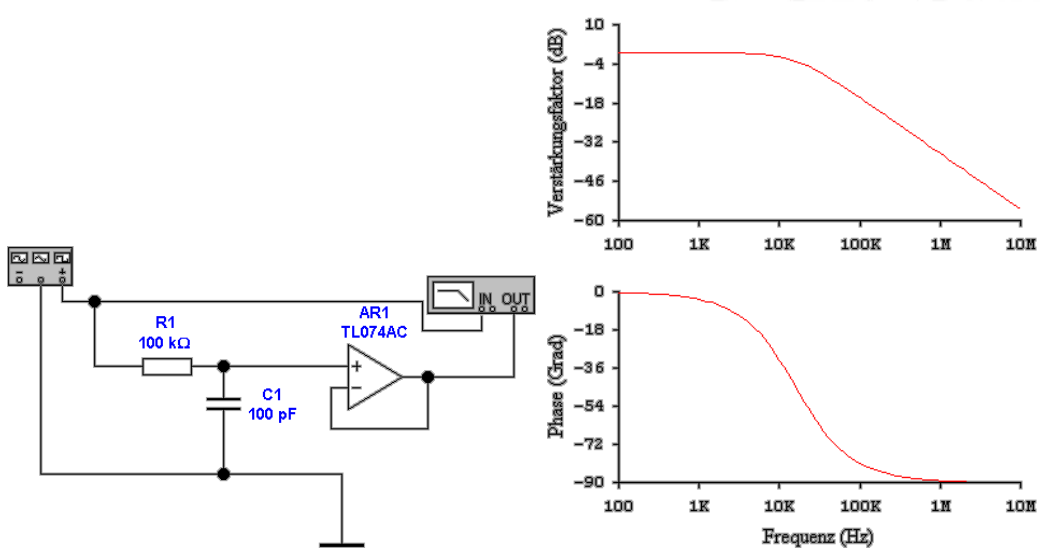



Abbildung 4.19.: Spannungsmessung über ein Koaxialkabel mit einem Operationsverstärker. Links ist die Schaltung, rechts das Bode-Diagramm.

sondere wenn die Quelle mit dem Messverstärker über ein Koaxialkabel verbunden wird, kann die Bandbreite der Messvorrichtung sehr eingeschränkt werden. Abbildung 4.19, links, zeigt ein Modell dieses Messsystems. Die *Spannungsquelle* wird über ihren Innenwiderstand von 100 kΩ an einen Spannungsverstärker mit der Verstärkung 1 angeschlossen. Das *Signal* der Quelle ist mit einem Koaxialkabel, Kapazität 100 pF, an den Verstärker angeschlossen. Die rechte Seite von Abb. 4.19 zeigt das entsprechende Bode-Diagramm. Der Tiefpass aus Innenwiderstand und Kabelkapazität bildet einen Tiefpass, mit einer Zeitkonstante von $100pF \times 100k\Omega = 10\mu s$. Dies entspricht einer Grenzfrequenz von 16 kHz, wie sie insbesondere aus dem Phasenbild ersichtlich ist.
In Abb. 4.20 wurde der Schirm des Koaxialkabel, wie von Tietze-Schenk[**?**] vorgeschlagen, an den Ausgang des Operationsverstärkers gelegt. Die Bandbreite wird, wie im in der Referenz angegebenen Buch beschrieben, breiter. Jedoch entsteht bei hohen Frequenzen eine Resonanz. Diese rührt von der Wechselwirkung der Phasenverschiebungen des Operationsverstärkers mit denen des Koaxialkabels her.
Diese Resonanz kann, wie in Abb. 4.21 gezeigt, gedämpft werden, indem der Schirm des Koaxialkabels nicht direkt, sondern über einen Widerstand von hier 1 kΩ angeschlossen wird. Die Grösse des Widerstandes hängt vom Operationsverstärker sowie von der Kabelkapazität ab.
Zur Detektion von Wechselspannungen ist es nötig, die Wechselspannung in eine Gleichspannung umzuwandeln. Abb. 4.22 zeigt eine entsprechende Schaltung. Das Wechselspannungssignal wird über die Diode gleichgerichtet und der Spitzenwert auf dem Kondensator akkumuliert. Die Schaltung oben ist mit einer Diode 1N4001GP aufgebaut. Bei hohem Frequenzen koppelt die beträchtliche Sperrschichtkapazität der Diode das Hochfrequenzsignal über. Besser ist es, eine Kleinsignaldiode wie die 1N4148 zu verwenden (siehe Abb. 4.23). HF-Dioden, deren Sperrschichtkapazität minimiert ist, sind eine ideale Wahl. In Frage kommt, u.A. der Typ 1N914. Man beachte, dass die detektierte Spannung um eine Diodendurchlassspannung geringer ist. Tabelle 4.1 gibt für die beiden Dioden 1N4001 und 1N 4148 die gemessenen Spannungen als Funktion der angelegten Spannung und der

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

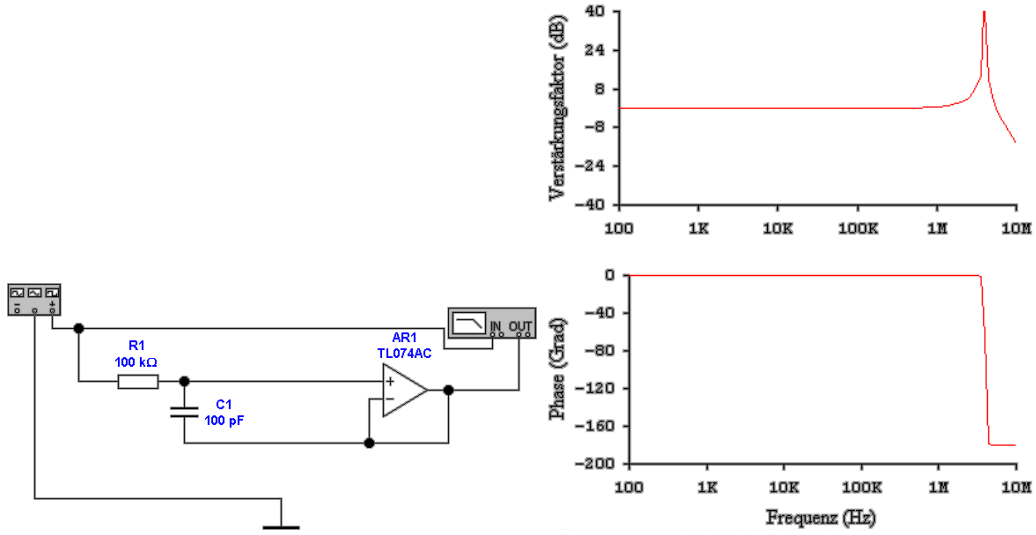



Abbildung 4.20.: Spannungsmessung über ein Koaxialkabel mit einem Operationsverstärker. Die Schirmung des Koaxialkabels wird mit dem Operationsverstärker getrieben. Links ist die Schaltung, rechts das Bode-Diagramm.

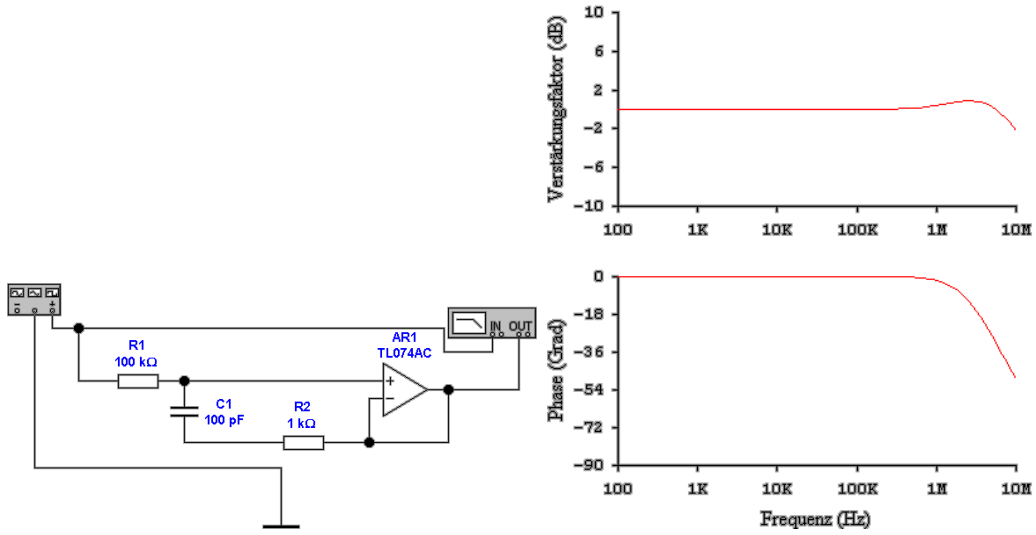



Abbildung 4.21.: Spannungsmessung über ein Koaxialkabel mit einem Operationsverstärker. Die Schirmung des Koaxialkabels wird mit dem Operationsverstärker getrieben und durch den 1-kΩ-Widerstand isoliert. Links ist die Schaltung, rechts das Bode-Diagramm.

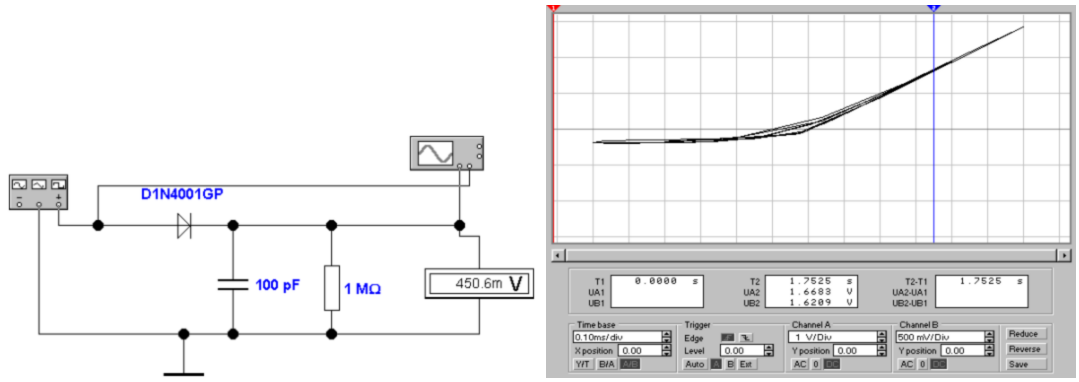



Abbildung 4.22.: Messung der Scheitelspannung einer Wechselspannung mit einer Diode und einem Kondensator. Das erste Bild zeigt die Schaltung. Die Anregungsspannung beträgt 3 V bei einem kHz. Das Zweite Bild zeigt ein Oszilloskopbild, wobei die Anregung horizontal und die Spannung am Detektor vertikal aufgetragen ist.

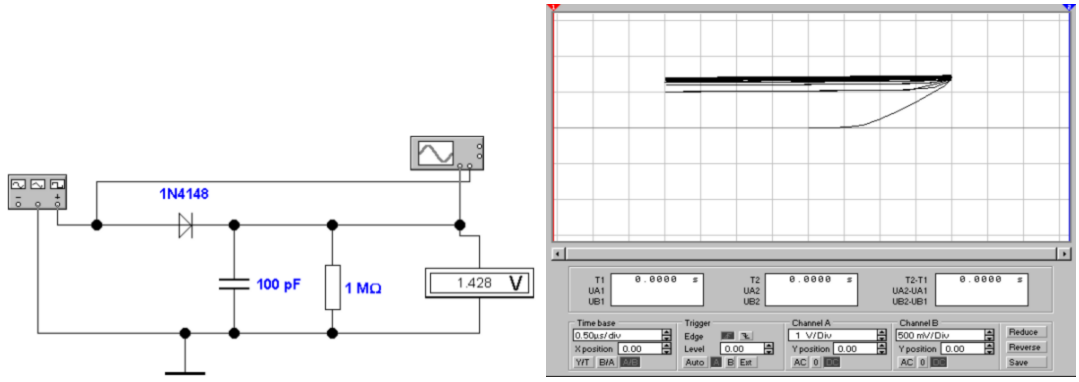



Abbildung 4.23.: Messung der Scheitelspannung einer Wechselspannung mit einer Kleinsignaldiode und einem Kondensator. Im Vergleich zu Abb. 4.22 ist die Scheitelspannungsdetektion gut zu erkennen.

Frequenz an. Aus dieser Tabelle wird schön ersichtlich, dass bei der Diode 1N4001 die Sperrschichtkapazität die die Brauchbarkeit limitierende Grösse ist. Generell ist das Messresultat bei kleinen Spannungen nicht aussagekräftig.
Abbildung 4.24 zeigt einen Halbwellengleichrichter aufgebaut mit einem Operationsverstärker und zwei Dioden. Die Schaltung funktioniert folgendermassen. Für die negative Halbwelle (Ausgangsspannung von AR1 positiv) wird vom Ausgang ein Strom geliefert, der durch D2, R1 und R2 fliesst. Da Die Verbindung zwischen R1 und R2 vom Operationsverstärker virtuell auf Erde gehalten wird, gilt für die Ausgangsspannung

$$U_A = -U_e * \frac{R_1}{R_2} \quad wenn\ U_e \leq 0 \tag{4.1.16}$$

Die Diode D3 ist in Sperrichtung gepolt. Für die positive Halbwelle ist D3 in Durchlassrichtung gepolt. Der Ausgang wird über R1 und R2 vom Eingang mit Strom versorgt. Die Diode D2 zieht nun den Ausgang auf eine Diodendurchlassspannung über der Ausgangsspannung des Operationsverstärkers. Dieser liegt aber, wegen der Durchlassspannung von D3 auf - einer Diodendurchlassspannung. Also

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

| Frequenz | Anregung AC | 1N4148 DC | 1N4148 AC | 1N4001 DC | 1N4001 AC |
|---|---|---|---|---|---|
| 1 kHz | 4 V | 1.065 V | 1.33 V | 1.098 V | 1.551 V |
| 10 kHz | 4 V | 1.685 V | 1.033 V | 1.626 V | 1.215 V |
| 100 kHz | 4 V | 3.176 V | 243.3 mV | 2.095 V | 799.4 mV |
| 1 MHz | 4 V | 3.286 V | 42.56 mV | 1.273 V | 1.339 V |
| 10 MHz | 4 V | 3.321 V | 32.98 mV | -330.5 mV | 2.803 V |
| 100 MHz | 4 V | 1.294 V | 660.9 mV | -389.0 mV | 2.817 V |
| 1 kHz | 2 V | 413.2 mV | 554.0 mV | 481.0 mV | 799.5 mV |
| 10 kHz | 2 V | 731.0 mV | 449.0 mV | 741.9 mV | 623.7 mV |
| 100 kHz | 2 V | 1.377 V | 106.1 V | 943.4 mV | 422.8 mV |
| 1 MHz | 2 V | 1.410 V | 11.03 mV | 807.2 mV | 499.2 mV |
| 10 MHz | 2 V | 1.414 V | 1.111 mV | -285.7 mV | 1.389 V |
| 100 MHz | 2 V | 1.397 V | 109.5 $\mu$V | -378.8 mV | 1.408 V |
| 1 GHz | 2 V | 1.398 V | 63.80 $\mu$V | -418.2 mV | 1.410 V |
| 1 kHz | 1 V | 121.8 mV | 192.4 mV | 185.3 mV | 428.9 mV |
| 10 kHz | 1 V | 232.0 mV | 149.0 mV | 313.4 mV | 330.7 mV |
| 100 kHz | 1 V | 436.8 mV | 33.51 mV | 417.4 mV | 218.8 mV |
| 1 MHz | 1 V | 446.6 mV | 3.438 mV | 379.9 mV | 235.9 mV |
| 10 MHz | 1 V | 446.0 mV | 341.4 $\mu$V | 117.8 mV | 399.0 mV |
| 100 MHz | 1 V | 445.1 mV | 59.73 $\mu$V | -174.6 mV | 640.3 mV |
| 1 GHz | 1 V | 425.6 mV | 67.36 $\mu$V | -361.3 mV | 702.7 mV |
| 1 kHz | 0.5 V | 10.07 mV | 19.4 mV | 51.86 mV | 19.40 mV |
| 10 kHz | 0.5 V | 20.98 mV | 14.18 mV | 114.8 mV | 184.5 mV |
| 100 kHz | 0.5 V | 26.66 mV | 1.948 mV | 167.2 mV | 117.5 mV |
| 1 MHz | 0.5 V | 26.55 mV | 194.1 $\mu$V | 161.5 mV | 118.7 mV |
| 10 MHz | 0.5 V | 23.87 mV | 110.2 $\mu$V | 77.47 mV | 159.1 mV |
| 100 MHz | 0.5 V | 11.98 mV | 60.89 $\mu$V | -86.42 mV | 262.3 mV |
| 1 GHz | 0.5 V | 2.507 mV | 11.3 $\mu$V | -157.8 mV | 309.1 mV |
| 10 MHz | 0.1 V | 18 nV | 0 nV | 8.762 mV | 29.04 mV |

Tabelle 4.1.: Gemessene Wechselspannungen mit den Schaltungen 4.22 und 4.23. Für 100 MHz wäre die ideale Gleichspannung nach dem Detektor $3.381V \approx 4V - 0.7V$ und die Wechselspannung wäre 355 $\mu$V.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

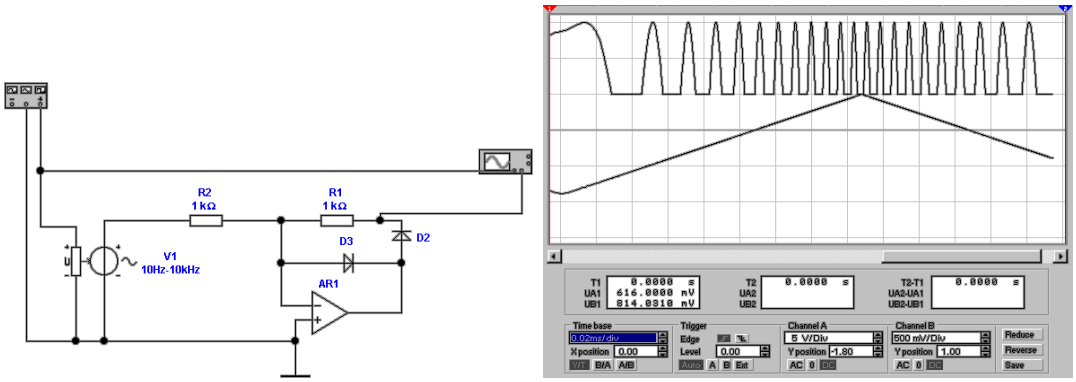



Abbildung 4.24.: Halbwellengleichrichter mit Operationsverständer. Links ist die Schaltung zu sehen, rechts das Ausgangssignal (10 Hz bis 10 kHz

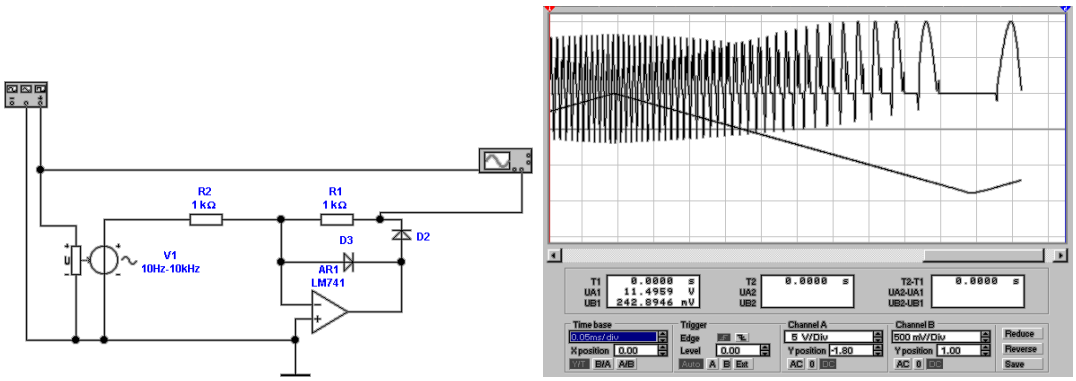



Abbildung 4.25.: Halbwellengleichrichter mit LM741. Links ist die Schaltung zu sehen, rechts das Ausgangssignal (10 Hz bis 10 kHz

ist der Ausgang auf 0 V. Die Schaltung 4.24 ist also ein Halbwellengleichrichter.

Während die Schaltung in Abbildung 4.24 mit einem idealen Operationsverstärker aufgebaut ist, ist die in Abb. 4.25 mit einem LM741 aufgebaut. Die Schaltung hat bei Frequenzen über einigen 100 Hz kein Gleichrichterverhalten mehr. Grund dafür ist die bescheidene Slew-Rate (Anstiegszeit) dieses Verstärkers. Immer wenn das Vorzeichen des Eingangssignales wechselt, muss eine Spannung von zwei Diodendurchlassspannungen übersprungen werden. Dies bedeutet beim LM741 eine Zeitverzögerung von etwa 2 $\mu$s. Diese, auf den ersten Blick unbedeutende Verzögerung ist jedoch für das schlechte Verhalten verantwortlich. Durch die Zeitverzögerung entsteht im Regelkreis des Operationsverstärkers eine Überkompensation.

**Für die Schaltung des Halbwellengleichrichters aus Abb. 4.25 werden Verstärker mit einer schnellen Anstiegsgeschwindigkeit des Ausgangssignals benötigt.** Abb. 4.26 zeigt die gleiche Schaltung aufgebaut mit einem TL074. Nun ist das Ausgangssignal bis zu 10 kHz von befriedigender Güte. Höhere Frequenzen können mit dieser Schaltung jedoch nur schwer vernünftig gleichgerichtet werden.

Abb. 4.27 zeigt, wie die Schaltung aus Abb. 4.26 zu einem Vollwellengleichrichter ausgebaut werden kann. Die Idee ist die folgende Dem Halbwellengleichrichter wird ein invertierender Verstärker parallelgeschaltet, der neben dem Eingangssignal auch den Ausgang des Halbwellengleichrichters mit doppeltem Gewicht ver-

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

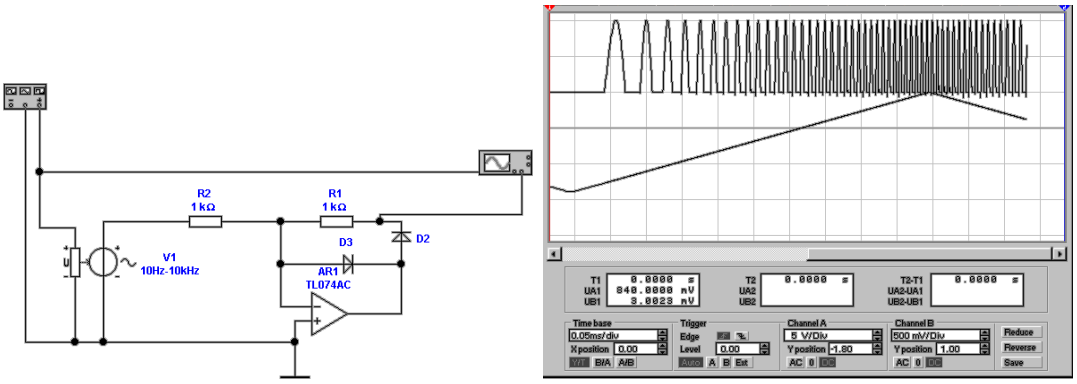



Abbildung 4.26.: Halbwellengleichrichter mit TL074. Links ist die Schaltung zu sehen, rechts das Ausgangssignal (10 Hz bis 10 kHz

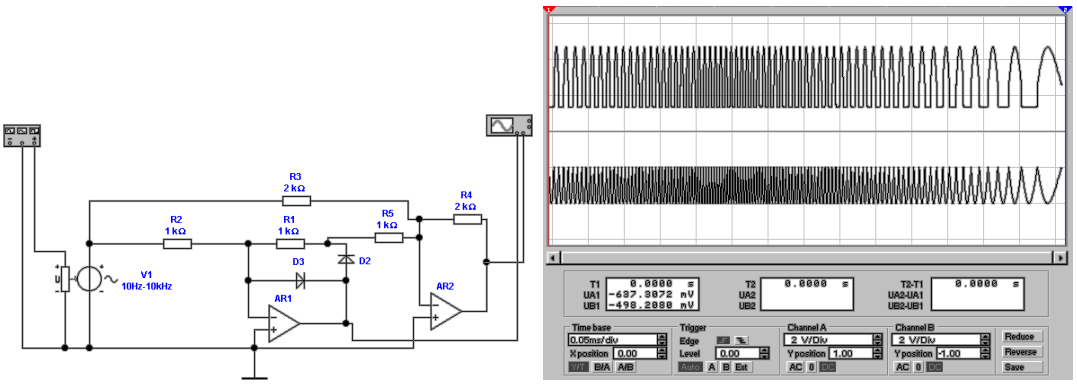



Abbildung 4.27.: Vollwellengleichrichter mit Operationsverstärker. Links ist die Schaltung zu sehen, rechts das Ausgangssignal (10 Hz bis 10 kHz

stärkt. Wir erhalten dann für die positive Halbwelle $U_e \geq 0$

$$U_A = -\frac{R_4}{R_3} U_e = -U_e \tag{4.1.17}$$

und für die negative Halbwelle $U_e \leq 0$

$$U_A = -\frac{R_4}{R_5} U_{A,Halbwelle} - \frac{R_4}{R_3} U_e = -2U_{A,Halbwelle} - U_e = -2\left(-U_e\right) - U_e = U_e \tag{4.1.18}$$

In der Summe ergibt sich also (für $U_e < 0$ ist $U_e = -\left|U_e\right|$!)

$$U_A = -\left|U_e\right| \tag{4.1.19}$$

Wieder ist die Schaltung aus Abb. 4.27 als ideale Schaltung nicht repräsentativ für die Güte realer Schaltungen. Abb. 4.28 zeigt wieder, dass Verstärker vom Typ LM741 nicht brauchbar für diese Anwendung sind, sie könnten höchstens bis einige 10 Hz verwendet werden.
Schliesslich ist aus Abb. 4.29 ersichtlich dass ein Vollwellen-Präzisionsgleichrichter mit dem TL074 bis 10 kHz arbeiten kann.
Wesentlich besser als die Operationsverstärkerschaltungen sind Transistorschaltungen als Gleichrichter für höchste Frequenzen geeignet. Abb. 4.30 zeigt eine

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

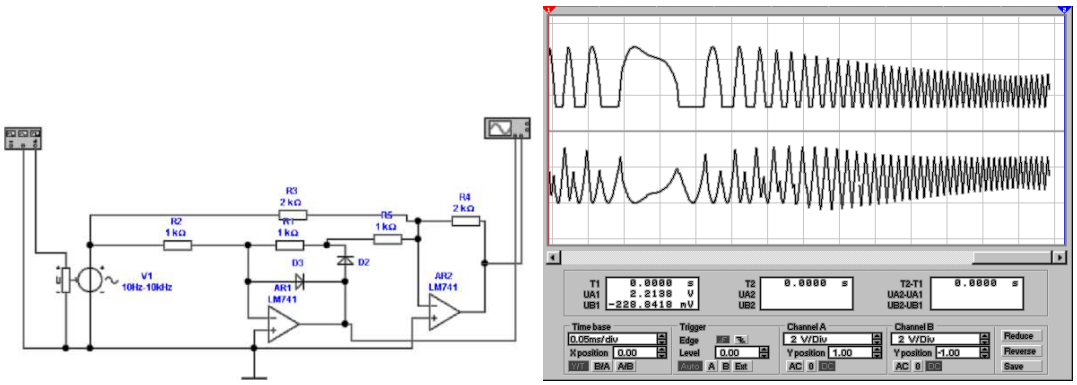



Abbildung 4.28.: Vollwellengleichrichter mit LM741. Links ist die Schaltung zu sehen, rechts das Ausgangssignal (10 Hz bis 10 kHz

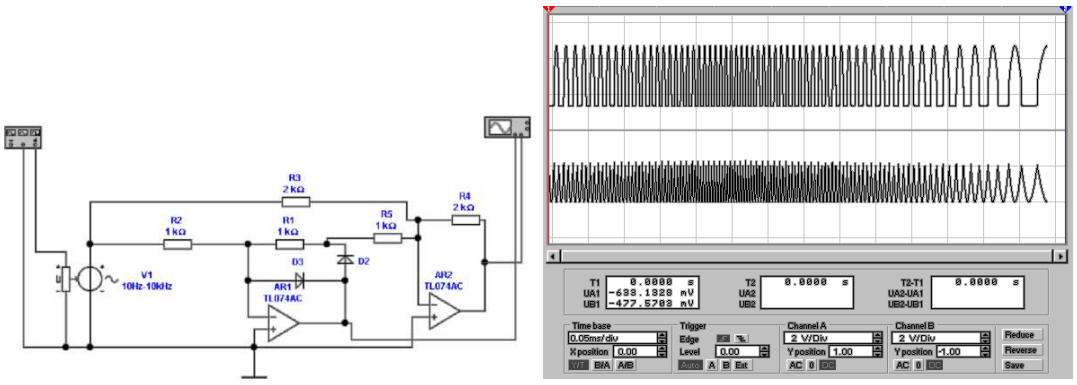



Abbildung 4.29.: Vollwellengleichrichter mit TL074. Links ist die Schaltung zu sehen, rechts das Ausgangssignal (10 Hz bis 10 kHz

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

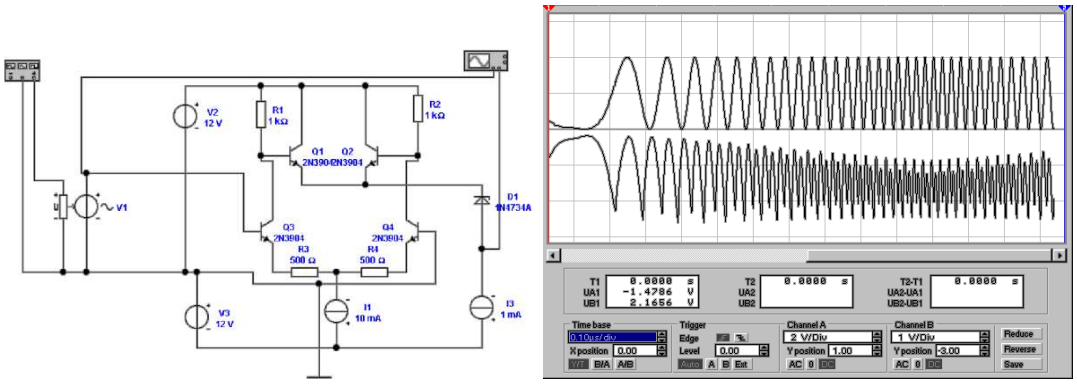



Abbildung 4.30.: Vollwellengleichrichter mit Transistoren. Links ist die Schaltung zu sehen, rechts das Ausgangssignal (3 kHz bis 3 MHz)

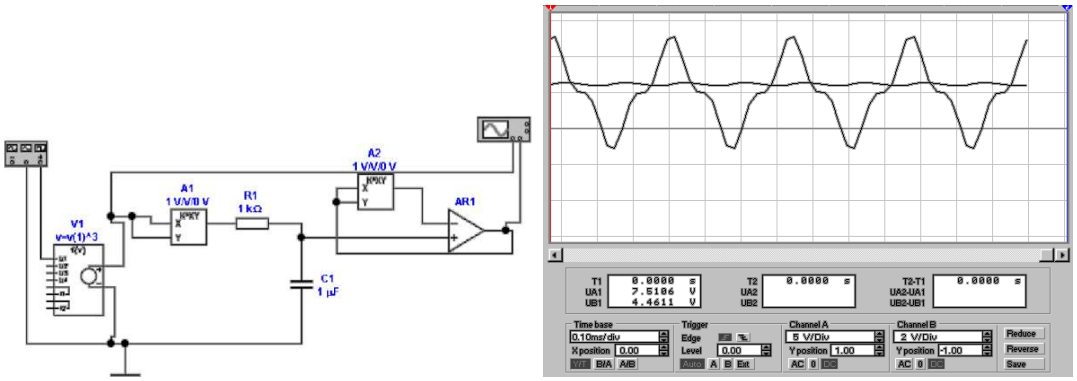



Abbildung 4.31.: True-RMS Wandler. Links ist die Schaltung zu sehen, rechts das Oszilloskopbild für ein $\sin^3$-*Signal*

mögliche Implementation eines Transistoren-Vollwellengleichrichters. Die Transistoren Q3 und Q4 bilden einen Differenzverstärker. Die Transistoren Q1 und Q2, deren Emitter und Kollektoren zusammengeschaltet sind, verstärken jeweils das positivere Potential. So kommt eine Vollwellengleichrichtung zustande. Die Schaltung nach Abb. 4.30 kann bei geeigneter Wahl der Transistoren mit bis zu 100 MHz betrieben werden.
Abbildung 4.31 zeigt einen True-RMS-Gleichrichter[1]. Die vorherigen Schaltungen haben den Betrag der Wechselspannung bestimmt. Nun ist der Betrag jedoch längstens nicht so interessant wie die dissipierte Leistung an einem Widerstand. Es ist bekannt, dass

$$P = \frac{U^2}{R} = I^2 R \tag{4.1.20}$$

Deshalb ist, wenn man Signalformen sucht, die die gleiche dissipierte Leistung an einem Widerstand $R$ erzeugen. Die Grösse

$$U_A(t) = \left( \frac{1}{T} \int_{t-T}^{t} U_e(\tau)^2 d\tau \right)^{\frac{1}{2}} \tag{4.1.21}$$

[1]RMS = Root Mean Square

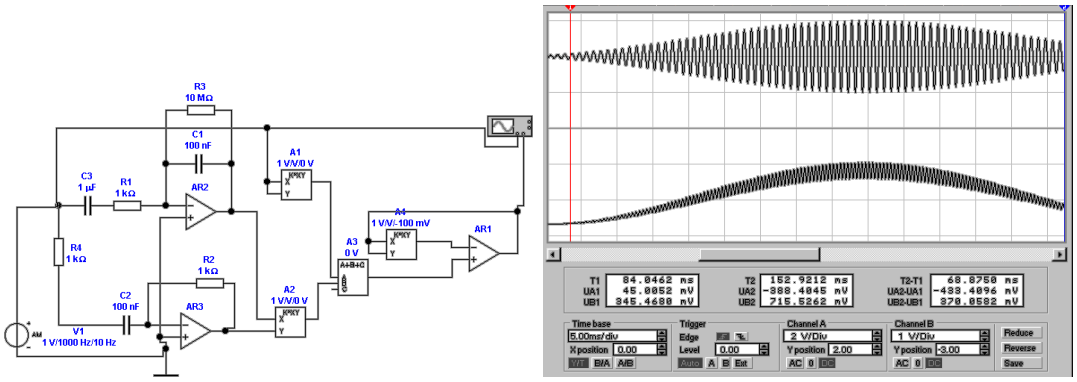



Abbildung 4.32.: Real-Time-Vektormesser. Links ist die Schaltung zu sehen, rechts Eingangs- und Ausgangssignal (Amplitudenmodulation)

sehr viel interessanter. Dies ist die Definition des RMS-Wertes. Die Schaltung in Abbildung 4.31 ist eine Implementation dieser Gleichung. Von links gesehen folgt zuerst ein Quadrierer, dann ein Integrator in Form eines Tiefpassfilters und zu letzt ein Radizierer. Dieser Schaltungsteil benützt die Tatsache, dass eine Impedanz im Rückkopplungszweig eines Operationsverstärkers sich zur dualen Funktion transformiert, das heisst vom Quadrat zur Wurzel, von der Exponentialfunktion zum Logarithmus.
Abb. 4.32 zeigt einen Vektormesser. Die Idee dahinter ist, wenn ich $A\cos\omega t$ und $A\sin\omega t$ habe, dann kann ich mit

$$A = \sqrt{(A\cos\omega t)^2 + (A\sin\omega t)^2} \tag{4.1.22}$$

rechnen. Angewandt auf die Schaltung von Abb. 4.32 führt dies zu

$$U_{A2,Y} = -R_2C_2\frac{dU_e}{dt} \quad = -\hat{U}_eR_2C_2\frac{d\sin\omega t}{dt} \quad = -\hat{U}_e\omega R_2C_2\cos\omega t \tag{4.1.23}$$

$$U_{A2,X} = -\frac{1}{R_1C_1}\int U_e dt \quad = -\frac{1}{R_1C_1}\int \hat{U}_e\sin\omega t dt \quad = \frac{\hat{U}_e}{R_1C_1\omega}\cos\omega t \tag{4.1.24}$$

$$U_{A3,B} = \frac{U_{A2,X}U_{A2,Y}}{1V} \quad = -\frac{\hat{U}_e{}^2}{1V}\cos^2\omega t \tag{4.1.25}$$

$$U_{A3,out} \quad = \frac{\hat{U}_e{}^2}{1V}\sin^2\omega t + \frac{\hat{U}_e{}^2}{1V}\cos^2\omega t \quad = \frac{\hat{U}_e{}^2}{1V} \tag{4.1.26}$$

Abb. 4.33 zeigt eine Brückenschaltung nach Graetz. Die Dioden oben rechts und unten links oder die Dioden oben links und unten rechts sind jeweils in Durchlassrichtung geschaltet. Im Diagonalzweig der Brücke steht die gleichgerichtete Spannung zur Verfügung. Die Schaltung hat zwischen dem Eingangsteil und dem Ausgangsteil kein gemeinsames Bezugspotential[2]. Die Graetz-Schaltung wird vorwiegend zur Gleichrichtung nach einem Trenntransformator verwendet.

$$U_= \max\left(|U_e| - 2U_{Diode}\,;\,0\right) = \max\left(|U_e| - 1.4V\,;\,0V\right) \tag{4.1.27}$$

[2]Das unvorsichtige Anschliessen eines Oszilloskopes kann Kurzschlüsse hervorrufen: warum?

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

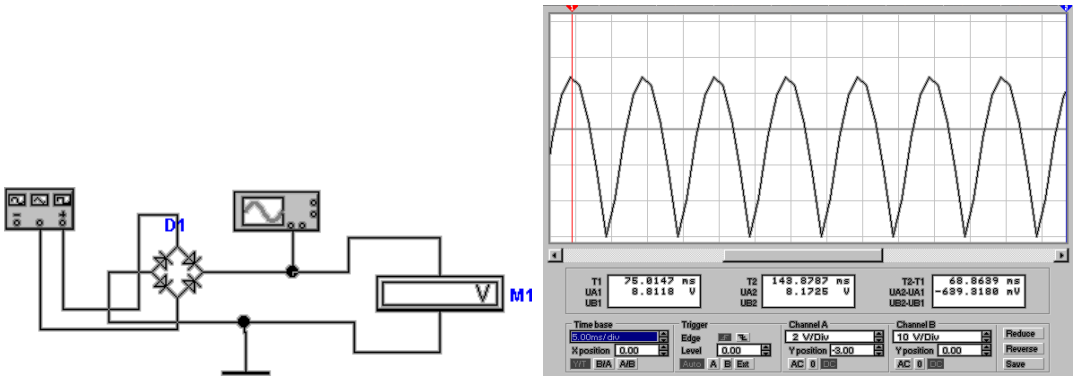



Abbildung 4.33.: Brückenschaltung (Graetz-Schaltung). Links ist die Schaltung zu sehen, rechts das Ausgangssignal für 50 Hz, 10 V

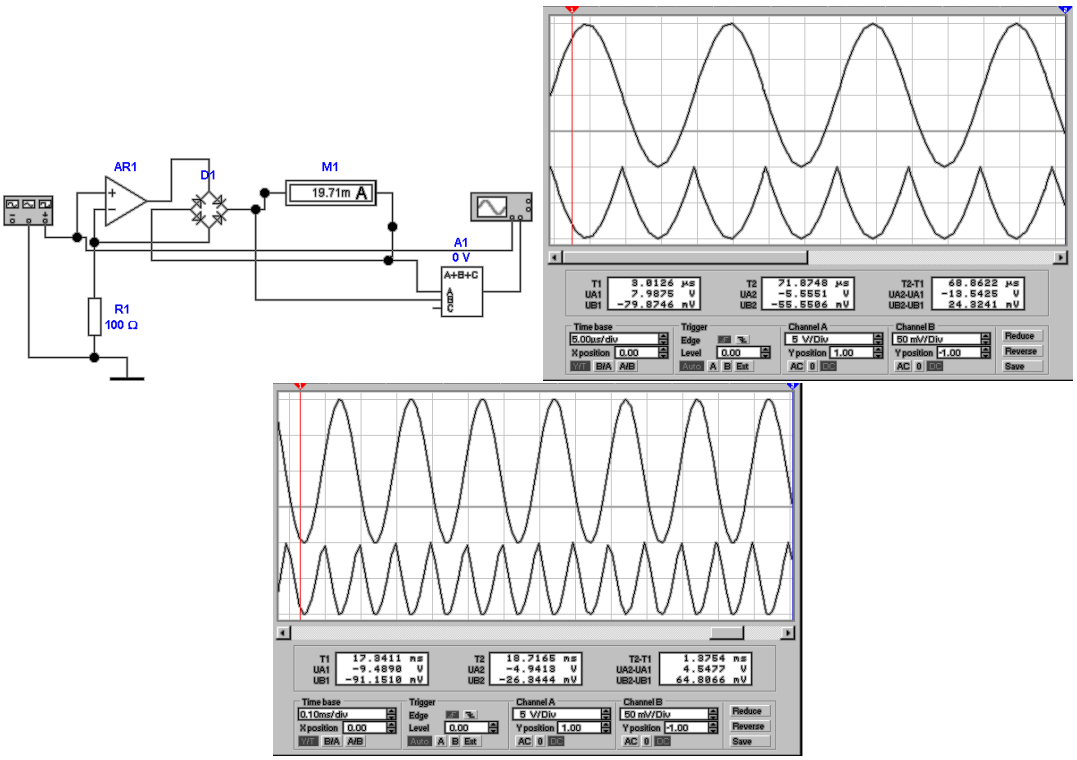



Abbildung 4.34.: Brückenschaltung (Graetz-Schaltung) mit Operationsverstärker. Links ist die Schaltung zu sehen, rechts das Ausgangssignal für 50 kHz, 10 V. Unten ist das Ausgangssignal der Schaltung gezeigt, wenn für den Operationsverstärker ein LM741 eingesetzt wird (Frequenz 5 kHz, Amplitude 10 V)

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

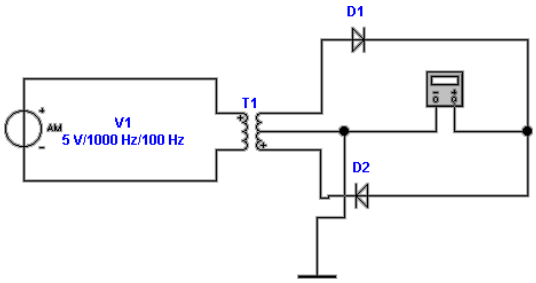



Abbildung 4.35.: Brückenschaltung mit Transformator als Wandler

Abb. 4.34 zeigt eine zu 4.33 analoge Schaltung. Dadurch, dass der Brückengleichrichter im Rückkoppelzweig ist, arbeitet die Schaltung mit eingeprägtem Strom. Die Spannungsabfälle über den Dioden werden so kompensiert.

$$U_{M1} = -\max\left(|U_e| - 2U_{Diode}\,;\,0\right) = -\max\left(|U_e| - 1.4V\,;\,0V\right) \tag{4.1.28}$$

Schliesslich zeigt Abb. 4.35 eine Brückenschaltung mit getrenntem Transformator. Die Schaltung ist besonders geeignet, wenn Transformatoren billig im Vergleich zu Dioden sind. Weiter ist der Spannungsabfall

$$U_{M1} = n\max\left(|U_e| - U_{Diode}\,;\,0\right) = n\max\left(|U_e| - 0.7V\,;\,0V\right) \tag{4.1.29}$$

nur eine Diodenspannung. In Gleichung (4.1.29) ist $n$ das Übertragungsverhältnis von $T1$.

### 4.1.4. Ladung

Die Messung von Ladung stellt eines der schwierigeren Messprobleme dar. Um Ladung messen zu können, muss Energie auf das Messwerk übertragen werden. Typische Energien sind jedoch

$$E_{1\,Elektron} = 1.6 \times 10^{-19}C \times 0.1V = 1.6 \times 10^{-20}J \tag{4.1.30}$$

also sehr klein.
Abbildung 4.36 zeigt eine Schaltung, mit der Ladungen gemessen werden können. Ziel ist, die Ladung auf dem Kondensator $C_1$ zu messen. Damit Die Ladung möglichst vollständig transferiert werden kann, muss die Kapazität $C_2$ sehr viel grösser als die von $C_1$ sein.
Vor dem Schliessen von $S_2$ gilt

$$Q_1 \neq 0$$
$$Q_2 = 0$$

Wird $S_2$ geschlossen gilt für die Ladungen $\tilde{Q}_1$ und $\tilde{Q}_2$ sowie das Spannungsgleichgewicht dass

$$Q_1 = \tilde{Q}_1 + \tilde{Q}_2$$
$$U_{C_1} = U_{C_2} \qquad \Rightarrow \frac{\tilde{Q}_1}{C_1} = \frac{\tilde{Q}_2}{C_2}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

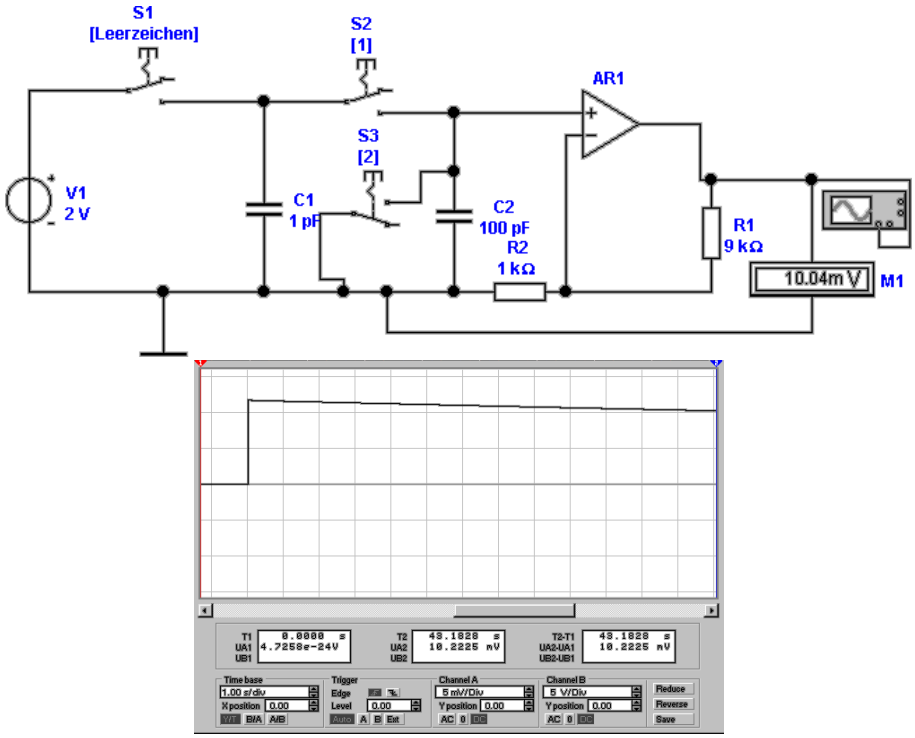



Abbildung 4.36.: Ladungsmessung über Ladungstransfer. Links die Schaltung und rechts das *Signal*, wenn $2 \times 10^{-12}C \approx 10^{7}e^{-}$ transferiert werden.

Daraus folgt:

$$\tilde{Q}_2 = Q_1 \frac{C_2}{C_1 + C_2} \tag{4.1.31}$$

gilt. Wenn $C_2 \gg C_1$ gilt, dann wird praktisch alle Ladung von $C_1$ auf $C_2$ übertragen und $\tilde{Q}_2 = Q_1$. $S_3$ dient zum Entladen des Messkondensators $C_2$. Aus der unteren Hälfte von 4.36 ersieht man, dass die Ausgangsspannung einen Sprung von der Grösse $U_{C_2} = \frac{\tilde{Q}_2}{C_2} = \frac{Q_1}{C_1+C_2}$ macht. Ebenso ist ersichtlich, dass Die Spannung sehr schnell wieder abnimmt, da Leckströme sogar im Modell eines idealen Verstärkers nicht zu vermeiden sind.
Zum Vergleich mit einem idealen Operationsverstärker in 4.36 zeigt Abb. 4.37 die gleiche Schaltung mit einem realen Verstärker. Die Widerstände $R_3$, $R_5$, $R_6$ und $R_7$ dienen zur Kompensation des Bias-Stromes im invertierenden Eingang von $AR_1$. Die Widerstände $R_2$, $R_8$, $R_9$ und $R_{10}$ dienen zur Kompensation der Offsetspannung. Nur mit diesen beiden Massnahmen ist es möglich, ein *Signal* wie in Abb. 4.37 zu bekommen. Wieder ist der Spannungssprung das Mass für die Ladung, der restliche Kurvenverlauf hängt von Störeinflüssen ab.
Abbildung 4.38 zeigt die Ladungsmessung mit einem Integrator. Die Verstärkung des Operationsverstärkers $AR_1$ bewirkt, dass der Kondensator $C_1$ **vollständig** entladen wird. Der resultierende Strom lädt wieder, ohne Verluste, $C-2$ auf. man erhält also

$$Q_1 = Q_2 \Rightarrow U_2 = -\frac{Q_2}{C_2} \tag{4.1.32}$$

Wie die rechte Seite von Abb. 4.38 zeigt, entsteht nach dem Anlegen von $C_1$ am Ausgang von $AR_1$ ein Spannungssprung. Anders als in der Schaltung von Abb.

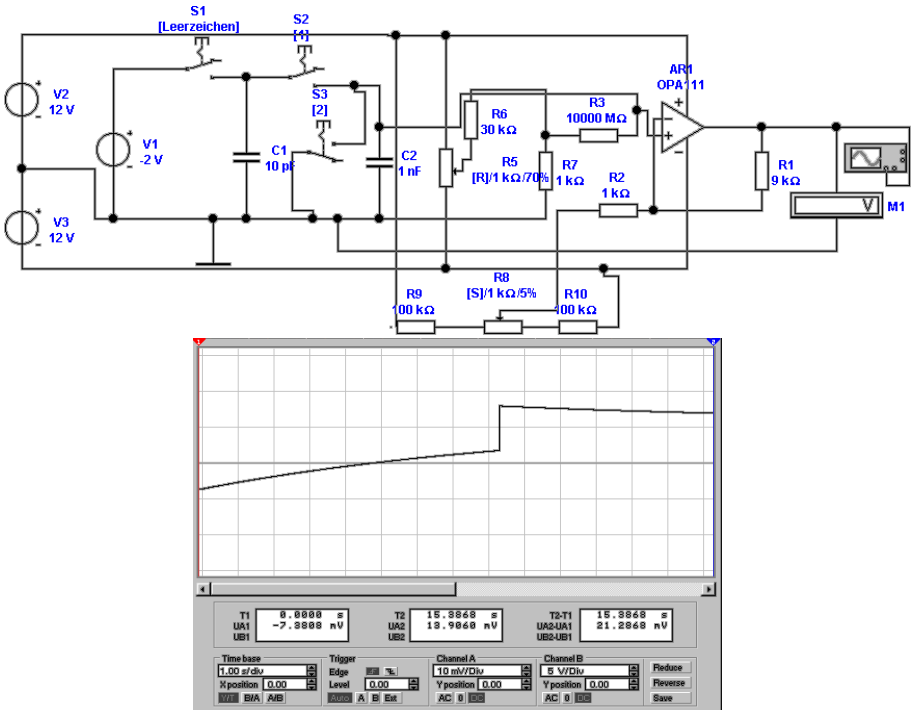



Abbildung 4.37.: Ladungsmessung über Ladungstransfer mit einem sehr guten, realen Operationsverstärker. Links die Schaltung und rechts das *Signal*, wenn $1.6 \times 10^{-12}C = 10^7 e^-$ transferiert werden.

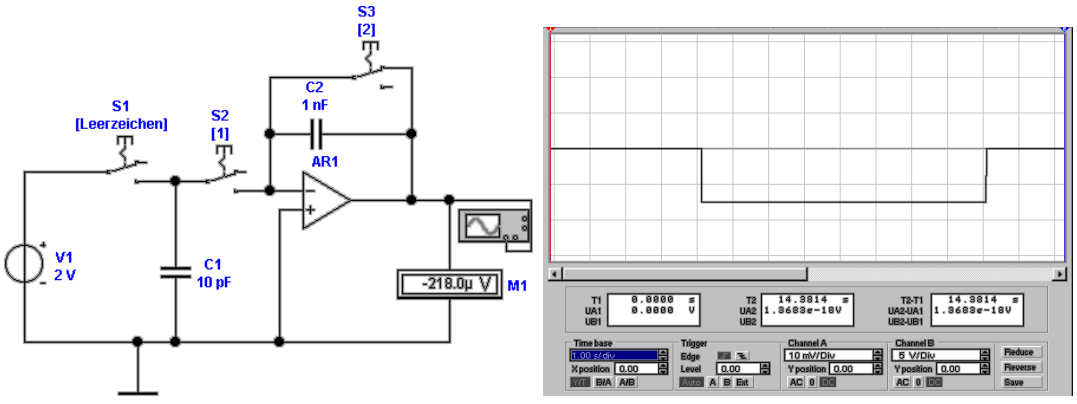



Abbildung 4.38.: Ladungsmessung mit Integrator. Links die Schaltung und rechts das *Signal*, wenn $2 \times 10^{-12}C \approx 10^7 e^-$ transferiert werden.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

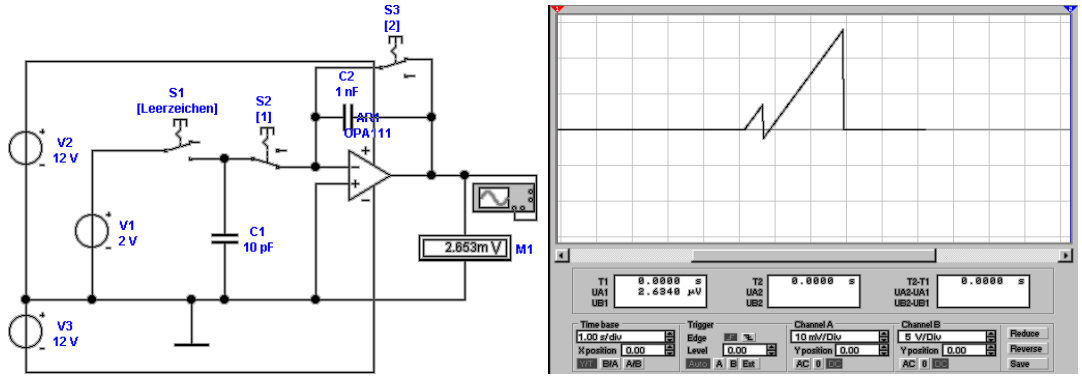



Abbildung 4.39.: Ladungsmessung mit Integrator mit einem sehr guten, realen Operationsverstärker. Links die Schaltung und rechts das *Signal*, wenn $2 \times 10^{-12}C \approx 10^7 e^-$ transferiert werden. Im Gegensatz zu Abb. 4.37 wurde keine Bias-Kompensation implementiert.

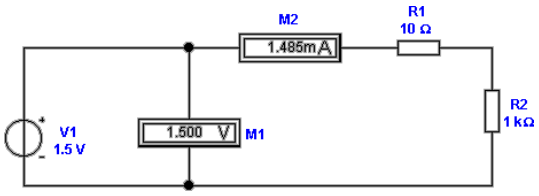



Abbildung 4.40.: Stromrichtige Widerstandsmessung.

4.36 ist das Ausgangssignal konstant. Hier zeigt sich der Vorteil, wenn man die Eingänge der Operationsverstärker in der Nähe des Spannungsnullpunktes hält. Der reale Operationsverstärker in Abb. 4.39 erzeugt genauso eine Rampe. Hier wurde keine Kompensationsschaltung verwendet: deshalb die doch steilen Spannungsverläufe. Wieder ist der Spannungssprung das Mass für die Ladung, und nicht der sonstige Spannungsverlauf.

### 4.1.5. Widerstand

In Abb. 4.40 wird die stromrichtige Messung eines Widerstandswertes gezeigt. Die Spannung $U$ wird über das Strommessgerät $M_2$ mit seinem Innenwiderstand $R_1$ an den zu messenden Widerstand $R_2$ gelegt. Die Messung heisst *stromrichtig*, da der Strom durch $R_2$ richtig gemessen wird, jedoch der Spannungsabfall an $R_1$ meistens nicht berücksichtigt wird. Ist $R_1$ bekannt, kann man mit

$$R_2 = \frac{U}{I} - R_1 \tag{4.1.33}$$

den genauen Wert von $R_2$ bestimmen. Eingesetzt ergibt sich:

$$R_2 = \frac{1.5V}{0.001485A} - 10\Omega = 1000.1\Omega$$

Abbildung 4.41 zeigt eine spannungsrichtige Widerstandsmessung. Hier ist das Spannungsmessgerät $M_1$ parallel zum zu messenden Widerstand angeschlossen. Die Spannung am Widerstand wird also **richtig** gemessen. Der Strom, den das

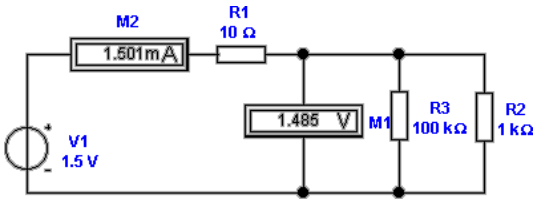



Abbildung 4.41.: Spannungsrichtige Widerstandsmessung.

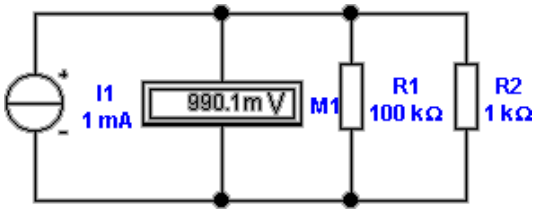



Abbildung 4.42.: Widerstandsmessung mit bekannter Stromquelle.

Ampèremeter $M_2$ misst, setzt sich aus dem Strom durch den zu prüfenden Widerstand $R_2$ sowie aus dem Strom durch den Innenwiderstand $R_3$ des Spannungsmessers zusammen. Is $R_3$ bekannt, so ergibt sich

$$\frac{1}{R_2} = \frac{I}{U} - \frac{1}{R_3} \tag{4.1.34}$$

Andernfalls muss sichergestellt werden, dass $R_2 \ll R_3$ ist. Setzt man die Werte aus Abbildung 4.41 in Gleichung (4.1.34) ein, erhält man

$$\begin{aligned} \frac{1}{R_2} &= \frac{0.001501A}{1.485V} - \frac{1}{100k\Omega} = 0.00100\overline{077441}\frac{1}{\Omega} \\ R_2 &= 999.2\Omega \end{aligned}$$

Die Widerstandsmessung kann vereinfacht werden, wenn man anstelle einer *Spannungsquelle* eine bekannte Stromquelle einsetzt (siehe Abb. 4.42). Wieder wird damit, spannungsrichtig, der Parallelwiderstand aus $R_2$ sowie dem Innenwiderstand des Spannungsmessers, $R_1$, gemessen.

$$\frac{1}{R_2} = \frac{I}{U} - \frac{1}{R_1} \tag{4.1.35}$$

Die Werte aus Abbildung 4.42 ergeben

$$\begin{aligned} \frac{1}{R_2} &= \frac{0.001A}{0.9891V} - \frac{1}{100k\Omega} = 0.00099999899\frac{1}{\Omega} \\ R_2 &= 1000.001\Omega \end{aligned}$$

Ist der Innenwiderstand $R_1$ des Spannungsmessers nicht genau bekannt, muss man die Annahme $R_1 \gg R_2$ machen.
Abb. 4.43 zeigt eine Widerstandsmessung mit einer *Spannungsquelle* sowie einem Voltmeter. Die gemessene Spannung entsteht am Spannungsteiler bestehend aus $R_3$, dem Innenwiderstand der *Spannungsquelle* und der Parallelschaltung $R_1 \| R_2$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

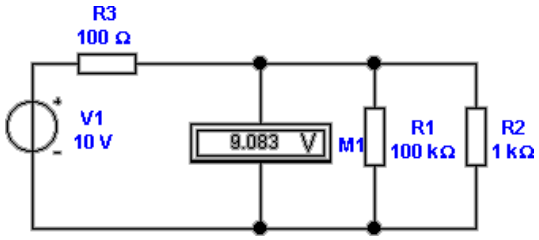



Abbildung 4.43.: Widerstandsmessung mit bekannter *Spannungsquelle* und Spannungsmesser.

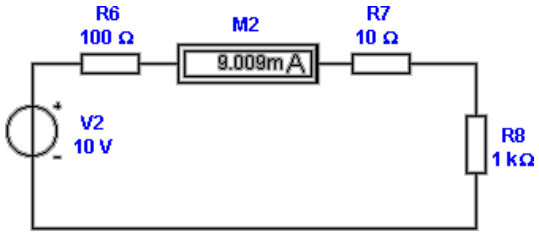



Abbildung 4.44.: Widerstandsmessung mit bekannter *Spannungsquelle* und Strommesser.

des Innenwiderstandes $R_1$ des Spannungsmessers $M_1$ sowie des zu prüfenden Widerstandes $R_2$

$$U_{mess} = U_{Quelle} \frac{R_1 \| R_2}{R_3 + R_1 \| R_2} \tag{4.1.36}$$

Aus Gleichung (4.1.35 folgt für die Parallelschaltung $R_1 \| R_2$ der Widerstand

$$R_1 \| R_2 = R_3 \frac{U_{mess}}{U_{Quelle} - U_{mess}} \tag{4.1.37}$$

Hier ist das Resultat eine Spannungsdifferenz einer nicht messbaren Quellspannung $U_{Quelle}$ und einem abgelesenen Wert $U_{mess}$: die Schaltung ist SEHR FEHLERBEHAFTET. Sie sollte nur eingesetzt werden, wenn es nicht anders geht. Das Resultat für Abb. 4.43 ist:

$$\begin{aligned} R_1 \| R_2 &= 990.5\Omega \\ R_2 &= 1000.4\Omega \end{aligned}$$

Vergleicht man den numerischen Wert dieses Resultates mit den vorherigen Ergebnissen fällt die doch deutlich schlechtere Genauigkeit auf.
Besser ist die Schaltung nach Abb. 4.44. Hier wird, bei bekannter *Spannungsquelle* der durch den Widerstand fliessende Strom gemessen. Es gilt

$$(R_6 + R_7 + R_8)\, I = U_{Quelle} \tag{4.1.38}$$

Damit wird der Wert des zu messenden Widerstandes $R_8$

$$R_8 = \frac{U_{Quelle}}{I} - (R_6 + R_7) \tag{4.1.39}$$

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

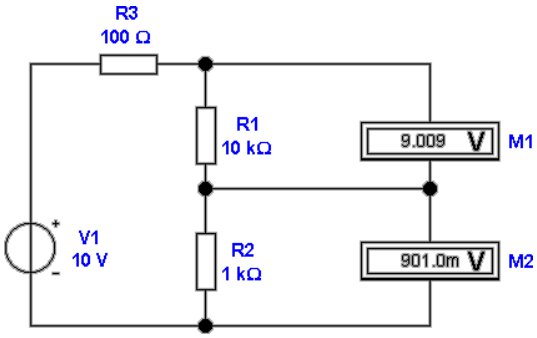



Abbildung 4.45.: Widerstandsmessung durch Vergleich mit Referenz.

Die Innenwiderstände des Strommessers, $R_7$ (gut bekannt), und der *Spannungsquelle*, $R_6$ (sehr schlecht bekannt), müssen vom aus Spannung und Strom berechneten Wert abgezogen werden. Das Resultat ist

$$R_8 = 1000.00\Omega$$

Diese Schaltung ist wesentlich genauer als die mit Spannungsmesser. Andererseits muss gefordert werden, dass der Innenwiderstand der *Spannungsquelle* $R_6$ sehr viel kleiner als der zu messende Widerstand $R_8$ ist.
Abbildung 4.45 zeigt eine ratiometrische Schaltung zur Widerstandsmessung, wie sie bei **digitalen Voltmetern** üblich ist. Unter der Annahme, dass der Innenwiderstand der Spannungsmesser $M_1$ und $M_2$ $R_i \gg R_{1,2}$ ist liefert diese Schaltung hervorragende, von der Betriebsspannung unabhängige Resultate. Trifft die Annahme nicht zu, so muss mit $R_{1,2} \| R_i$ gerechnet werden. Der Widerstandswert von $R_2$ wird so berechnet:

$$R_2 = R_1 \frac{U_2}{U_1} \tag{4.1.40}$$

Dabei ist der gemessene Wert von $R_2$ in sehr guter Näherung unabhängig von $R_3$ oder $U_{Quelle}$. Als Resultat erhält man mit den obigen Werten:

$$R_2 = 1000.1\Omega$$

Wie Abbildung 4.46 zeigt wird diese Schaltung in vielen Digitalvoltmetern verwendet.
Abbildung 4.47 zeigt die ratiometrische Widerstandsmessung mit Strömen. Unter der Annahme, dass der Innenwiderstand der Strommesser $M_3$ und $M_4$ $R_i \ll R_{7,8}$ sei werden direkt die richtigen Widerstandswerte, sonst wird $R_{7,8} + R_i$ gemessen. Es gilt

$$I_3 R_7 = I_4 R_8 \tag{4.1.41}$$

unabhängig von $R_3$ oder $U_{Quelle}$. Wir erhalten schliesslich

$$R_8 = R_7 \frac{I_3}{I_4} \tag{4.1.42}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

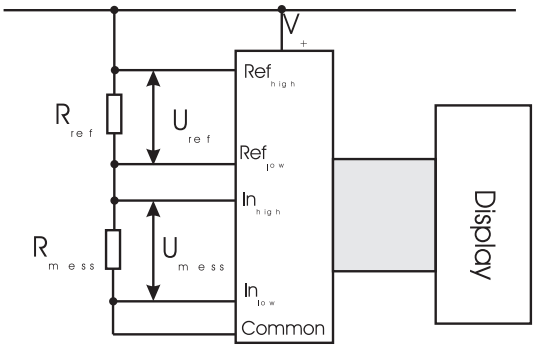



Abbildung 4.46.: Widerstandsmessung durch Vergleich mit Referenz. Die angegebene Schaltung stammt aus einer „Application Note“ für den Schaltkreis Teledyne ICL 7107.

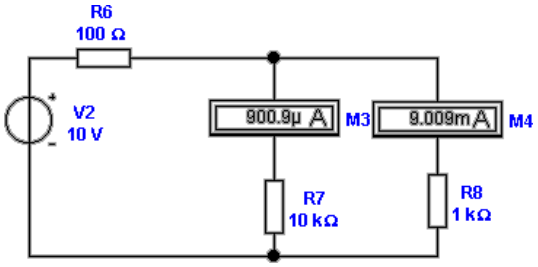



Abbildung 4.47.: Widerstandsmessung durch Vergleich mit Referenz. Hier werden Ströme verglichen.

und eingesetzt den Wert

$$R_8 = 1000\Omega$$

Die bisher vorgestellten Messmethoden sind für ganz kleine Widerstände nicht geeignet. Abbildung 4.48 zeigt die Vierdraht-Methode zur Widerstandsmessung. Eine *Spannungsquelle* wird über einen Referenzwiderstand $R_1$ und über Kabel mit den Widerständen $R_6$ und $R_9$ an den zu messenden Widerstand $R_2$ angelegt. Um den Spannungsabfall in $R-6$ und $R_9$ zu kompensieren, wird die Spannung am Widerstand direkt abgegriffen und über die Widerstände $R_7$ und $R_8$ zum Spannungsmesser gebracht. Wenn der Innenwiderstand von $M_1$ und $M_2$ sehr viel grösser als die anderen beteiligten Widerstände sind, dann hat man eine analoge Messmethode wie in Abb. 4.45. In diesem Falle sind die Kabelwiderstände und die Übergangswiderstände nicht relevant. Wir erhalten für $R_2$

$$R_2 = R_1 * \frac{U_2}{U_1} \tag{4.1.43}$$

Die Widerstände $R_4$ und $R_5$ in Abbildung 4.49 ermöglichen eine automatische Umschaltung von der 2-Draht- zur 4-Draht-Methode. Sie sind aber auch die Ursache für zusätzliche Fehler. Mit der Nebenbedingung:

$$R_{4,5} \gg R_{6,9} + R_{7,8} \tag{4.1.44}$$

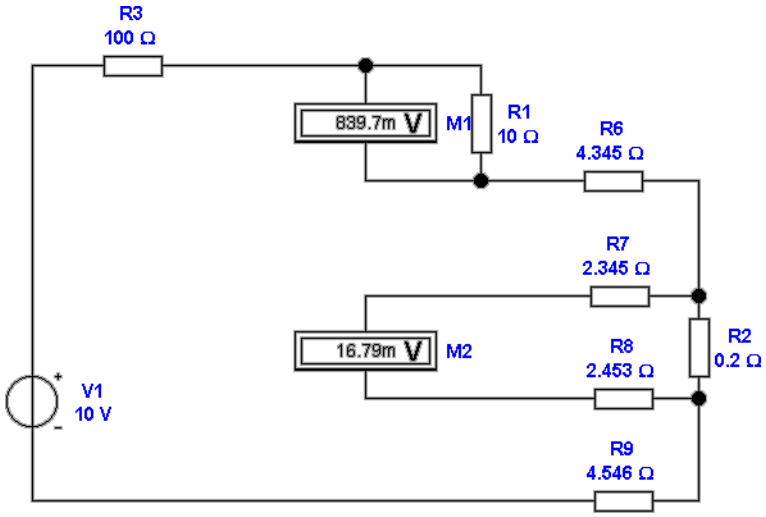



Abbildung 4.48.: Vierdraht-Widerstandsmessung für kleine Widerstände. Diese Schaltung kann nur im 4-Drahtverfahren betrieben werden.

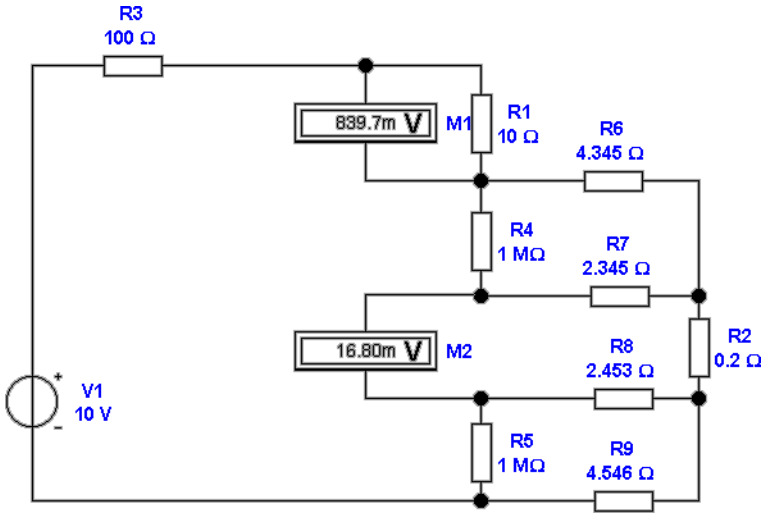



Abbildung 4.49.: Vierdraht-Widerstandsmessung für kleine Widerstände. Hier ist eine Schaltung angegeben, die automatisch von 2-Draht-Messung auf 4-Draht-Messung umschaltet.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

oder

$$R_{4,5} \gg \frac{R_{6,9}R_{7,8}}{R_2} - R_{6,9} - R_{7,8} \tag{4.1.45}$$

können deren Widerstandswerte vernachlässigt werden. Wir erhalten dann wie ohen $R_4$ und $R_5$

$$R_2 = R_1 * \frac{U_2}{U_1} \tag{4.1.46}$$

Ist man sicher, dass der Spannungsabfall in den Widerständen $R_6$ und $R_9$ nicht wesentlich über $0.1V$ ansteigt, könne man die Widerstände durch jeweils eine oder zwei in Durchlassrichtung geschaltete Dioden ersetzen. Die durch $R_4$ und $R_5$ hervorgerufenen Spannungsfehler sind:

$$I_m = \frac{U_1}{R_1} \tag{4.1.47}$$

$$U_{R_6} = I_m \left(R_6 \| \left(R_4 + R_7\right)\right) = \frac{I_m}{\frac{1}{R_6} + \frac{1}{(R_4+R_7)}} = \frac{I_m R_6}{1 + \frac{R_6}{(R_4+R_7)}} \tag{4.1.48}$$

$$U_{R_7} = U_{R_6} \frac{R_7}{R_4 + R_7} = \frac{R_7}{R_4 + R_7} \frac{I_m R_6}{1 + \frac{R_6}{(R_4+R_7)}} = \frac{I_m R_6 R_7}{R_4 + R_6 + R_7} \tag{4.1.49}$$

$$U_{R_7} = \frac{U_1 R_6 R_7}{R_1 \left(R_4 + R_6 + R_7\right)} \tag{4.1.50}$$

Die beiden oben gezeigten Schaltungen sind ideal zu Messung kleiner Widerstände geeignet. Sie haben jedoch einen gravierenden Nachteil: **je kleiner die Widerstände sind, desto höher werden sie thermisch belastet**. Dies liegt daran, dass zur Messung eines Spannungsabfalls ein minimaler Wert benötigt wird. Für $R \to \infty$ gilt

$$\lim_{R\to\infty} \frac{U^2}{R} = \infty \tag{4.1.51}$$

Abbildung 4.50 zeigt, wie mit einer gepulsten Messung das Problem der zu grossen thermischen Belastung umgangen werden kann. Wir nehmen hier an, dass die *Spannungsquelle* $V_2$ (6 mV) die durch thermische Belastung hervorgerufenen Spannungen zusammenfasst. Auf der linken Seite wird die Schaltung mit eingeschalteter Messspannungsquelle gezeigt, rechts ist die Messspannungsquelle ausgeschaltet. Indem die Messspannungsquelle nur den Bruchteil $\varepsilon$ der gesamten Messzeit eingeschaltet wird wird die Verlustleistung an $R_2$

$$P_{R_2} = \varepsilon \frac{U_{mess}^2}{R_2} \tag{4.1.52}$$

Zusätzlich kann man, wenn die Messpannungsquelle mit der Frequenz $\nu$ geschaltet wird, Lock-In-Verstärker verwenden und so die Empfindlichkeit steigern und die thermische Belastung senken. Im oberen Teil, bei eingeschalteter Messspannungsquelle gilt

$$U_{2,ein} = \frac{U_1}{R_1} + U_{Therm.} \tag{4.1.53}$$

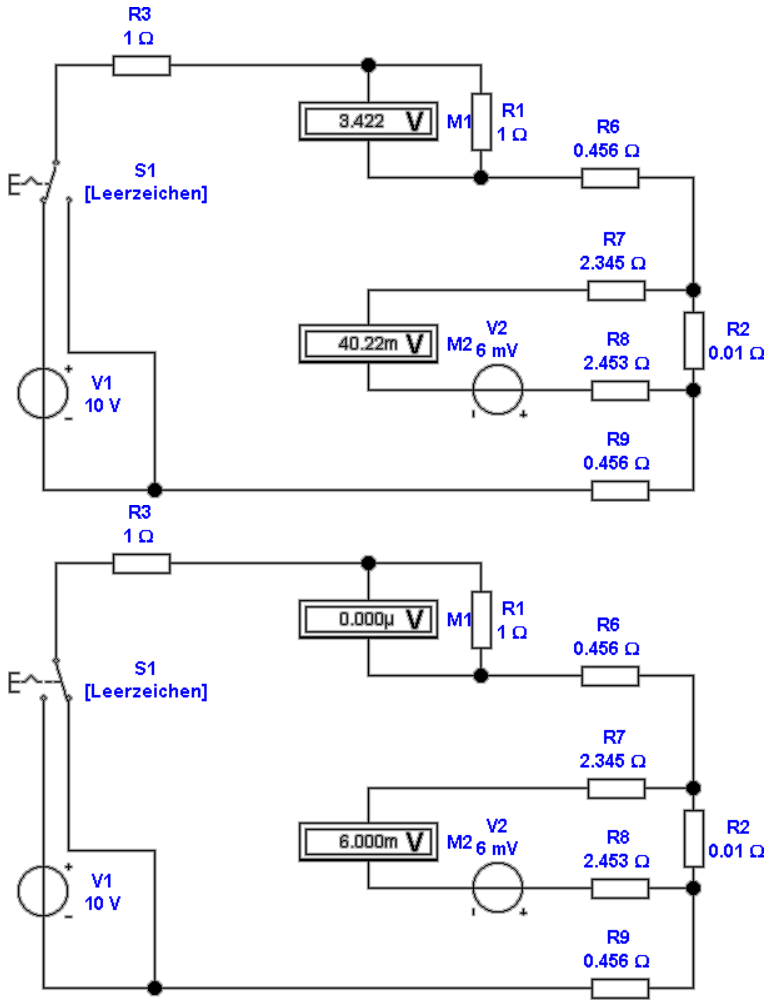



Abbildung 4.50.: Vierdraht-Widerstandsmessung für sehr kleine Widerstände im gepulsten Betrieb. Oben ist der Zustand mit eingeschalteter Messspannungsquelle gezeigt, unten mit ausgeschalteter Quelle.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

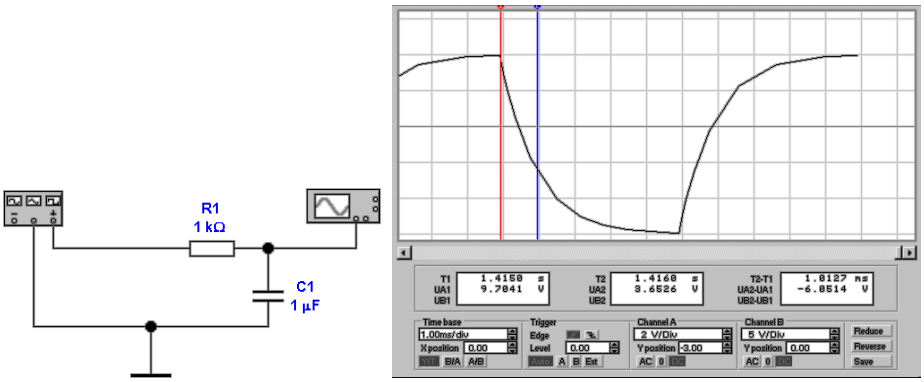



Abbildung 4.51.: Messung der Kapazität über die Anstiegs- und Abfallzeit

Im unteren Teil, ohne Messspannungsquelle hat man

$$U_{2,aus} = U_{Therm.} \tag{4.1.54}$$

Den gesuchten Widerstand $R_2$ findet man mit

$$R_2 = R_1 \frac{U_{2,ein} - U_{2,aus}}{U_1} \tag{4.1.55}$$

Achtung!

Sollte dem Widerstand $R_2$ jedoch grosse kapazitive oder, wahrscheinlicher, induktive Komponenten parallel geschaltet sein, ist die Messung nicht mehr zuverlässig.

## 4.1.6. Messung von L und C

Die Messungen von Kapazitäten und Induktivitäten kann auf verschiedene Arten erfolgen:

1. Messung der Zeitkonstanten bei Einschalt- oder Ausschaltvorgängen.
2. Messung von Resonanzfrequenzen in Schwingkreisen
3. Messung der komplexen Impedanzen.

In Abbildung 4.51 wird eine Kapazität $C_1$mit einer periodischen Wechselspannung über einen Widerstand $R_1$ geladen und entladen. Im Entladefall hat man, dass

$$U(t) = U_0 e^{-\frac{t}{R_1 C_1}} \tag{4.1.56}$$

Nun ist bei $t = \tau = R_1 C_1$ die Spannung gerade $U(\tau) = U_0 * e^{-1} = 0.3679 U_0$. Aus der Zeitkonstante für Entladung in Abb. 4.51, rechts, (Differenz zwischen blauer und roter Markierung) liest man ab, dass $\tau = 1.0127ms$ ist. Daraus folgt mit $R_1 = 1k\Omega$ dass $C_1 = \frac{\tau}{R_1} = \frac{1.0127ms}{1000\Omega} = 1.0127\mu F$ ist. Man ersieht aus der kurzen Rechnung, dass eine genau Messung Schwierigkeiten bieten dürfte.
Für die Anstiegszeit gilt

$$U(t) = U_0 \left(1 - e^{-\frac{t}{R_1 C_1}}\right) = U_0 \left(1 - e^{-\frac{t}{\tau}}\right) \tag{4.1.57}$$

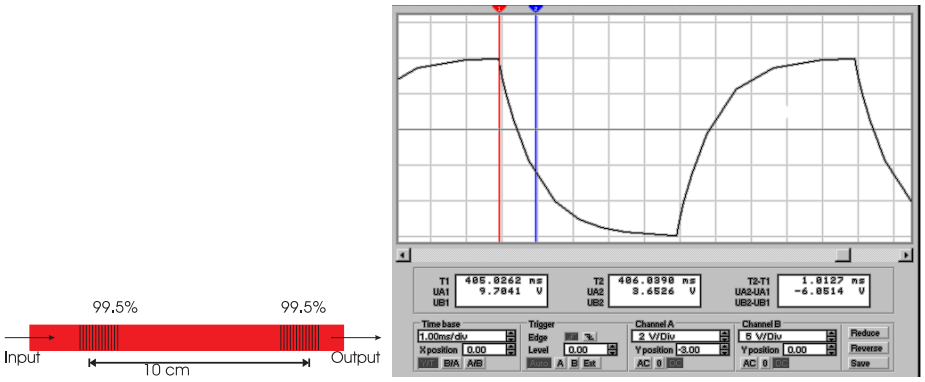



Abbildung 4.52.: Messung der Induktivität über die Anstiegs- und Abfallzeit

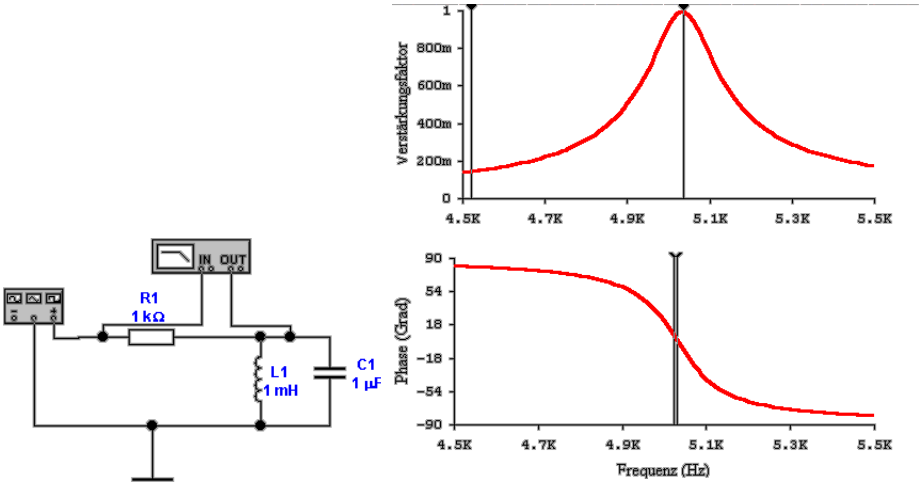



Abbildung 4.53.: Messung der Induktivität oder Kapazität mit einem Schwingkreis

d.h. man rechnet analog zum Entladefall. Hier ist angenommen worden, dass die Spannung $U_0$ zwischen 0V und ihrem (positiven) Maximalwert hin- und hergeschaltet wird. ist die untere Spannung nicht null, muss ihr Wert als Offset abgezogen werden.
In Abbildung 4.51 wird über eine Induktivität $L_1$ an den Widerstand $R_1$ eine periodischen Rechteckspannung angelegt. Wenn die Eingangsspannung von $U_0$ 0 wechselt hat man

$$U(t) = U_0 e^{-\frac{tR_1}{L_1}} \tag{4.1.58}$$

Nun ist bei $t = \tau = \frac{L_1}{R_1}$ die Spannung gerade $U(\tau) = U_0 * e^{-1} = 0.3679 U_0$. Aus der Zeitkonstante für Entladung in Abb. 4.51, rechts, (Differenz zwischen blauer und roter Markierung) liest man ab, dass $\tau = 1.0368ms$ ist. Daraus folgt mit $R_1 = 1\Omega$ dass $L_1 = \tau R_1 = 1.0368ms\, 1\Omega = 1.0368mH$ ist. Man ersieht auch aus der kurzen Rechnung, dass eine genau Messung Schwierigkeiten bieten dürfte.
Für die Anstiegszeit gilt analog

$$U(t) = U_0 \left(1 - e^{-\frac{tR_1}{L_1}}\right) = U_0 \left(1 - e^{-\frac{t}{\tau}}\right) \tag{4.1.59}$$

Abb. 4.53 zeigt, wie man mit einem Schwingkreis Kapazitäten oder Induktivitäten bestimmen kann. Für einen Schwingkreis gilt allgemein:

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

$$A(\omega) = A_0 \frac{\omega_0^2}{\sqrt{\left(\omega^2 - \omega_0^2\right)^2 + \frac{\omega^2\omega_0^2}{Q^2}}} \tag{4.1.60}$$

$$\alpha(\omega) = \arctan\left(\frac{1}{Q}\frac{\omega\omega_0}{\omega_0^2 - \omega^2}\right) \tag{4.1.61}$$

dabei ist $\omega_0$ die Resonanzfrequenz und $Q$ die Güte des Schwingkreises. In der Phase $\alpha(\omega)$ gilt in unserem Falle

$$\alpha\omega_0 = 0 \tag{4.1.62}$$

$$\left.\frac{d\alpha(\omega)}{d\omega}\right|_{\omega=\omega_0} = \frac{2Q}{\omega_0} \tag{4.1.63}$$

Damit ist die Resonanzfrequenz und die Güte einfach aus dem Phasenbild ablesbar.

Achtung!

Der Amplitudengang hat zwar prinzipiell die gleiche Aussagekraft wie der Phasengang, ist aber wesentlich ungenauer auszumessen. Eine fast immer gültige Regel besagt: **Resonanzfrequenz $\omega_0$ und Güte $Q$ bestimmt man aus der Phase.**

Die letzte Möglichkeit, die Werte von Kapazitäten und Induktivitäten zu bestimmen, ist mit ihren komplexen Impedanzen zu rechnen.

$$Z_L = i\omega L$$
$$Z_C = \frac{1}{i\omega C}$$

In den im Abschnitt 4.1.5 besprochenen Schaltungen wird die Gleichspannung durch eine Wechselspannung mit bekannter Amplitude $U$ und Frequenz $\omega$ ersetzt. Der so bestimmte Impedanzwert $Z$ kann dann umgerechnet werden nach

$$C = \frac{1}{|\omega Z|} \tag{4.1.64}$$

$$L = \left|\frac{Z}{\omega}\right| \tag{4.1.65}$$

## 4.1.7. Brückenschaltungen

Mit Brückenschaltungen kann man komplexe Impedanzen sehr schnell und genau vermessen. Abbildung 4.54 zeigt eine Widerstandsbrückenschaltung. Im Idealfall erhält man für das Widerstandsverhältnis im abgeglichenen Falle

$$R_1 : R_4 = R_2 : R_3$$
$$R_1 R_3 = R_2 R_4 \tag{4.1.66}$$

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

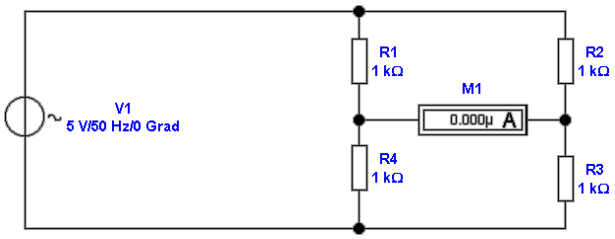



Abbildung 4.54.: Wheatstone-Brücke

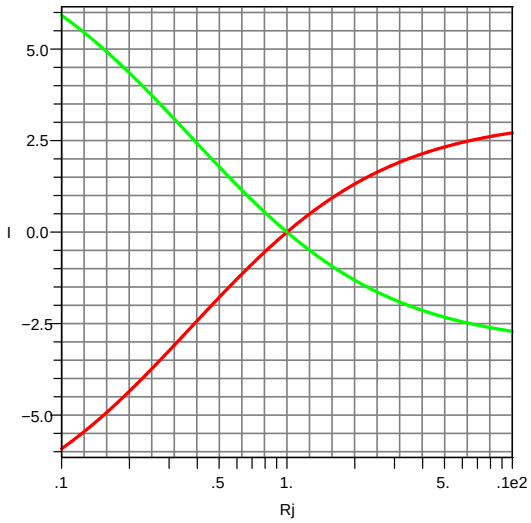



Abbildung 4.55.: Unabgeglichene Wheatstone-Brücke. Variiert werden $R_{1,2}$ (grün) und $R_{3,4}$ (rot). Die statischen Widerstände sind jeweils $1k\Omega$, der Innenwiderstand des Strommessers ist $R_i = 0.1k\Omega$. Die Brückenspannung ist $U = 10V$

Für die unabgeglichene Brücke erhält man:

$$I_i = U \frac{R_2 R_4 - R_1 R_3}{R_1 R_4 (R_2 + R_3) + R_2 R_3 (R_1 + R_4) + R_i (R_1 + R_4)(R_2 + R_3)} \tag{4.1.67}$$

Die Herleitung von Gleichung (4.1.67) finden Sie im Anhang B.
Abbildung 4.55 zeigt die Änderung des Querstromes in der Brücke als Funktion der Änderung der einzelnen Teilwiderstände. Sehr schön ist aus dieser Darstellung ersichtlich, dass die Ausgangsspannung der Brücke nichtlinear ist.
Die Grösse des Querstromes hängt nicht nur vom Ungleichgewicht der Brücke ab, sondern auch vom Innenwiderstand des Strommessers zum Nullabgleich. Abbildung 4.56 zeigt den Einfluss des Innenwiderstandes auf die Ausgangskurve, wenn $R_1$ variiert wird. Analog dazu zeigt Abbildung 4.57 den Einfluss des Innenwiderstandes auf die Ausgangskurve, wenn $R_4$ variiert wird.
Interessant ist der Fall, wo $R_1$ und $R_4$ gegengleich sich ändern, wo also $R_4 = 1/R_1$ gilt. Dieser Fall ist bei Sensoren wie Dehnungsmessstreifen oder piezoresistive AFM-Cantilever gegeben. Da variieren beide Widerstände in einem Brückenzweig.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

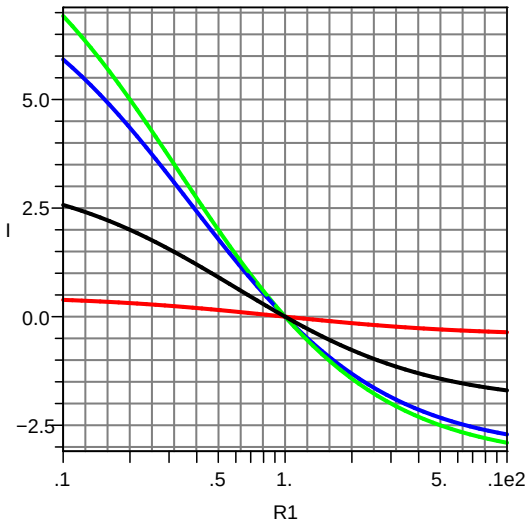



Abbildung 4.56.: Unabgeglichene Wheatstone-Brücke. Variiert wird $R_1$ mit $R_i = [0k\Omega, 0.1k\Omega, 1k\Omega, 10k\Omega]$ Innenwiderstand (Reihenfolge grün, blau, schwarz, rot). Die statischen Widerstände sind jeweils $1k\Omega$. Die Brückenspannung ist $U = 10V$

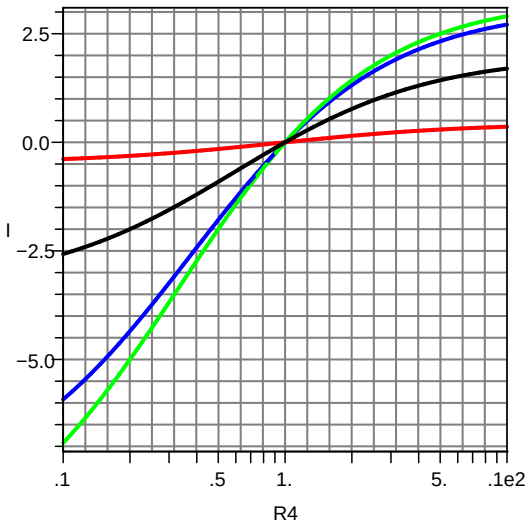



Abbildung 4.57.: Unabgeglichene Wheatstone-Brücke. Variiert wird $R_4$ mit $R_i = [0k\Omega, 0.1k\Omega, 1k\Omega, 10k\Omega]$ Innenwiderstand (Reihenfolge grün, blau, schwarz, rot). Die statischen Widerstände sind jeweils $1k\Omega$. Die Brückenspannung ist $U = 10V$

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

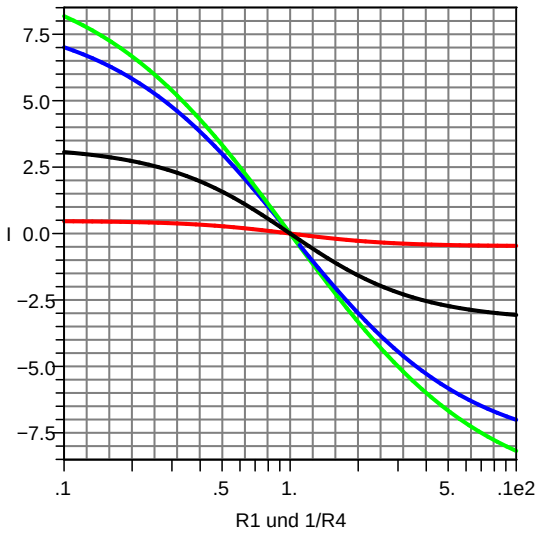



Abbildung 4.58.: Unabgeglichene Wheatstone-Brücke. Variiert werden $R_1$ und $R_4 = 1/R_1$ mit $R_i = [0k\Omega, 0.1k\Omega, 1k\Omega, 10k\Omega]$ Innenwiderstand (Reihenfolge grün, blau, schwarz, rot). Die statischen Widerstände sind jeweils $1k\Omega$. Die Brückenspannung ist $U = 10V$

Abbildung 4.58 zeigt die Ausgangskurven. Es ist bemerkenswert, um wieviel linearer das *Signal* ist.
Für den Fall dass der Innenwiderstand des Strommessers $R_i = 0$ ist, erhält man die vereinfachte Gleichung:

$$I_i = U\frac{R_2R_4 - R_1R_3}{R_1R_4\left(R_2 + R_3\right) + R_2R_3\left(R_1 + R_4\right)} \tag{4.1.68}$$

Misst man die Brückenspannung, so ergibt sich aus Gleichung 4.1.67

$$U_i = UR_i\frac{R_2R_4 - R_1R_3}{R_1R_4\left(R_2 + R_3\right) + R_2R_3\left(R_1 + R_4\right) + R_i\left(R_1 + R_4\right)\left(R_2 + R_3\right)} \tag{4.1.69}$$

Weiter sieht man, dass für $R_i \to \infty$

$$U_i = U\frac{R_2R_4 - R_1R_3}{\left(R_1 + R_4\right)\left(R_2 + R_3\right)} \tag{4.1.70}$$

ist. Abbildung 4.59 fasst den Einfluss des Innenwiderstandes nochmals zusammen. Die Empfindlichkeit auf Veränderungen von $R_1$ oder $R_4$ ergibt sich aus

$$\begin{aligned}\frac{\partial I_i}{dR_1} =& \frac{UR_3}{R_i\left(R_1 + R_4\right)\left(R_2 + R_3\right) + R_1R_4\left(R_2 + R_3\right) + R_2R_3\left(R_1 + R_4\right)} - \\ & \frac{U\left(R_2R_4 - R_3R_1\right)\left(\left(R_i + R_4\right)\left(R_2 + R_3\right) + R_2R_3\right)}{\left(R_i\left(R_1 + R_4\right)\left(R_2 + R_3\right) + R_1R_4\left(R_2 + R_3\right) + R_2R_3\left(R_1 + R_4\right)\right)^2}\end{aligned} \tag{4.1.71}$$

$$\begin{aligned}\frac{\partial I_i}{dR_4} =& \frac{UR_2}{R_i\left(R_1 + R_4\right)\left(R_2 + R_3\right) + R_1R_4\left(R_2 + R_3\right) + R_2R_3\left(R_1 + R_4\right)} - \\ & \frac{U\left(R_2R_4 - R_3R_1\right)\left(\left(R_i + R_1\right)\left(R_2 + R_3\right) + R_2R_3\right)}{\left(R_i\left(R_1 + R_4\right)\left(R_2 + R_3\right) + R_1R_4\left(R_2 + R_3\right) + R_2R_3\left(R_1 + R_4\right)\right)^2}\end{aligned} \tag{4.1.72}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

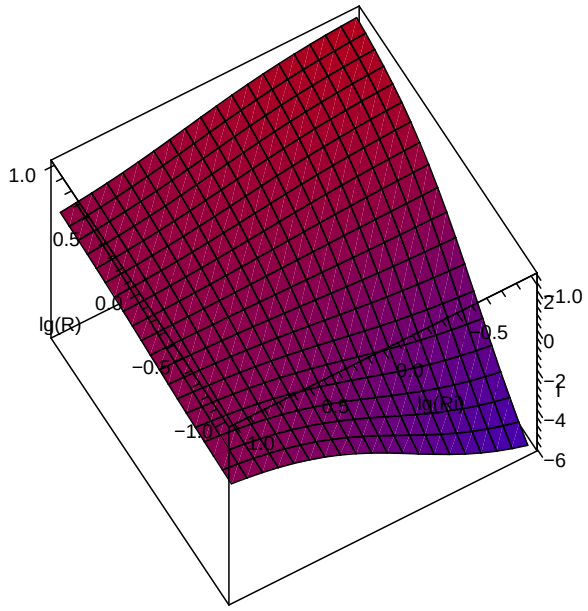



Abbildung 4.59.: Einfluss der Impedanz des Strommessers im Nullpunktszweig der Wheatstone-Brücke. R entspricht $R_4$ in Abb. 4.65, $R_i$ ist der Innenwiderstand des Strommessers.

Abbildung 4.60 vergleicht dabei die Variation von $R_1$ und $R_4$. Die Steigungen ändern sich extrem, das heisst, dass der lineare Bereich doch stark eingeschränkt ist.
Abbildung 4.61 zeigt den Einfluss des Innenwiderstandes $R_i$, wenn $R_1$ sich ändert. Die Empfindlichkeit der Brücke nimmt mit steigendem Innenwiderstand ab.
Abbildung 4.62 zeigt den Einfluss des Innenwiderstandes $R_i$, wenn $R_4$ sich ändert. Die Empfindlichkeit der Brücke nimmt mit steigendem Innenwiderstand ab.
Wenn sich $R_1$ und $R_4 = 1/R_1$ gegengleich ändern, dann variiert die Empfindlichkeit wie in Abbildung 4.63 angegeben.
Eine Detaildarstellung der normierten Empfindlichkeit in Abbildung 4.64 zeigt, dass für grosse Innenwiderstände $R_i$ die Empfindlichkeit am wenigsten variiert. Die Messkurve kann mit guter Näherung als linear mit einem kleinen paraboloiden Korrekturterm angesehen werden.
Für den Fall dass der Innenwiderstand des Strommessers $R_i = 0$ ist, erhält man die vereinfachte Gleichung:

$$\frac{\partial I_i}{dR_1} = \frac{UR_3}{R_1R_4\left(R_2+R_3\right)+R_2R_3\left(R_1+R_4\right)} - \frac{U\left(R_2R_4-R_3R_1\right)\left(R_4\left(R_2+R_3\right)+R_2R_3\right)}{\left(R_1R_4\left(R_2+R_3\right)+R_2R_3\left(R_1+R_4\right)\right)^2} \tag{4.1.73}$$

$$\frac{\partial I_i}{dR_4} = \frac{UR_2}{R_1R_4\left(R_2+R_3\right)+R_2R_3\left(R_1+R_4\right)} - \frac{U\left(R_2R_4-R_3R_1\right)\left(R_1\left(R_2+R_3\right)+R_2R_3\right)}{R_1R_4\left(R_2+R_3\right)+R_2R_3\left(R_1+R_4\right)^2} \tag{4.1.74}$$

Die Empfindlichkeit für Spannungsmessungen ist

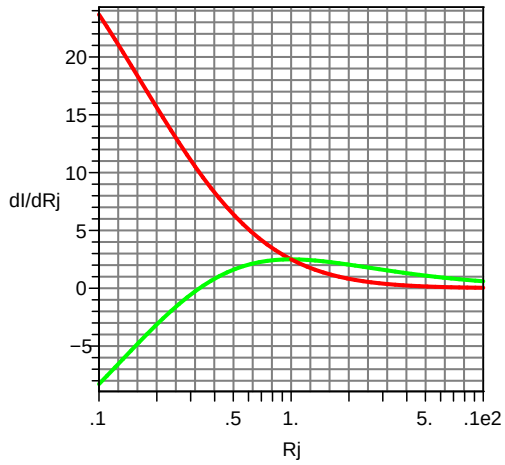



Abbildung 4.60.: Empfindlichkeit der unabgeglichenen Wheatstone-Brücke. Variiert werden $R_1$ (grün) und $R_4$ (rot) mit $R_i = 0k\Omega$ Innenwiderstand. Die statischen Widerstände sind jeweils $1k\Omega$. Die Brückenspannung ist $U = 10V$

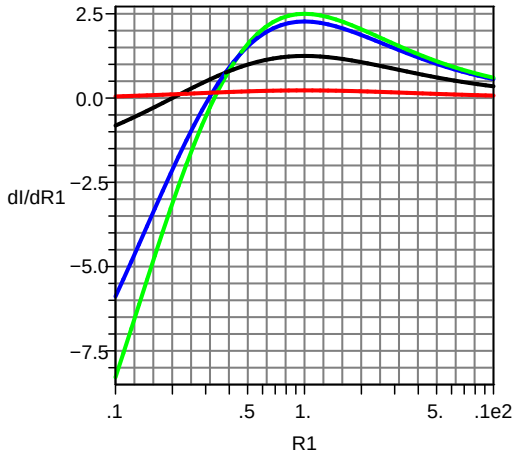



Abbildung 4.61.: Empfindlichkeit der unabgeglichenen Wheatstone-Brücke. Variiert wird $R_1$ mit $R_i = [0k\Omega, 0.1k\Omega, 1k\Omega, 10k\Omega]$ Innenwiderstand (Reihenfolge grün, blau, schwarz, rot). Die statischen Widerstände sind jeweils $1k\Omega$. Die Brückenspannung ist $U = 10V$

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

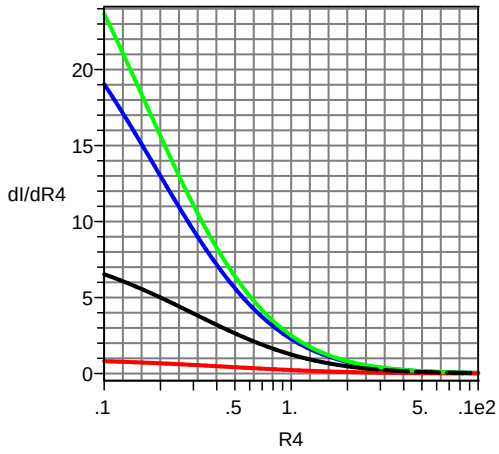



Abbildung 4.62.: Empfindlichkeit der unabgeglichenen Wheatstone-Brücke. Variiert wird $R_4$ mit $R_i = [0k\Omega, 0.1k\Omega, 1k\Omega, 10k\Omega]$ Innenwiderstand (Reihenfolge grün, blau, schwarz, rot). Die statischen Widerstände sind jeweils $1k\Omega$. Die Brückenspannung ist $U = 10V$

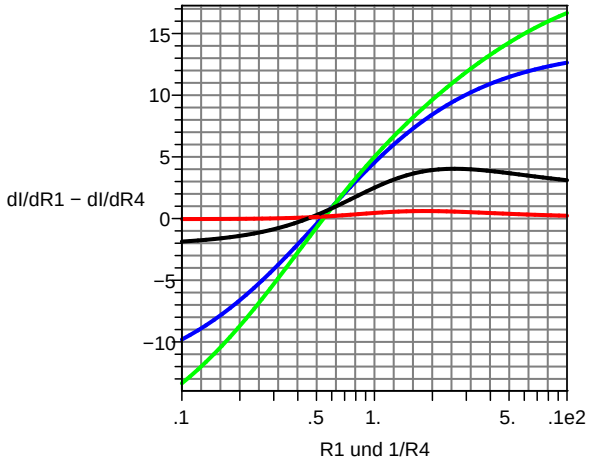



Abbildung 4.63.: Empfindlichkeit der unabgeglichenen Wheatstone-Brücke. Variiert werden $R_1$ und $R_4 = 1/R_1$ mit $R_i = [0k\Omega, 0.1k\Omega, 1k\Omega, 10k\Omega]$ Innenwiderstand (Reihenfolge grün, blau, schwarz, rot). Die statischen Widerstände sind jeweils $1k\Omega$. Die Brückenspannung ist $U = 10V$

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

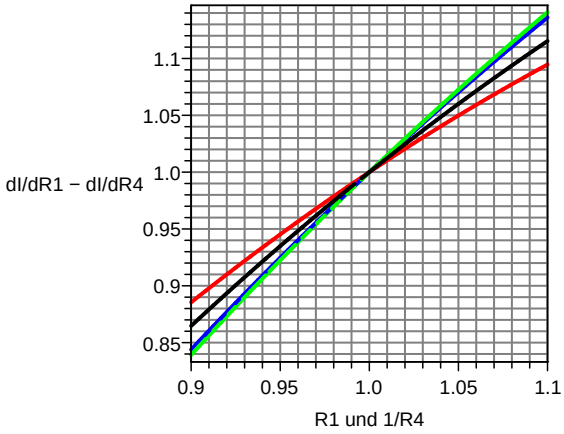



Abbildung 4.64.: Empfindlichkeit der unabgeglichenen Wheatstone-Brücke normiert auf den abgeglichenen Fall. Variiert werden $R_1$ und $R_4 = 1/R_1$ mit $R_i = [0k\Omega, 0.1k\Omega, 1k\Omega, 10k\Omega]$ Innenwiderstand (Reihenfolge grün, blau, schwarz, rot). Die statischen Widerstände sind jeweils $1k\Omega$. Die Brückenspannung ist $U = 10V$

$$\frac{\partial U_i}{dR_1} = \frac{UR_iR_3}{R_i\left(R_1+R_4\right)\left(R_2+R_3\right)+R_1R_4\left(R_2+R_3\right)+R_2R_3\left(R_1+R_4\right)} - \frac{UR_i\left(R_2R_4-R_3R_1\right)\left(\left(R_i+R_4\right)\left(R_2+R_3\right)+R_2R_3\right)}{\left(R_i\left(R_1+R_4\right)\left(R_2+R_3\right)+R_1R_4\left(R_2+R_3\right)+R_2R_3\left(R_1+R_4\right)\right)^2} \tag{4.1.75}$$

$$\frac{\partial U_i}{dR_4} = \frac{UR_iR_2}{R_i\left(R_1+R_4\right)\left(R_2+R_3\right)+R_1R_4\left(R_2+R_3\right)+R_2R_3\left(R_1+R_4\right)} - \frac{UR_i\left(R_2R_4-R_3R_1\right)\left(\left(R_i+R_1\right)\left(R_2+R_3\right)+R_2R_3\right)}{\left(R_i\left(R_1+R_4\right)\left(R_2+R_3\right)+R_1R_4\left(R_2+R_3\right)+R_2R_3\left(R_1+R_4\right)\right)^2} \tag{4.1.76}$$

Schliesslich erhält man für $R_i \to \infty$

$$\frac{\partial U_i}{dR_1} = \frac{UR_3}{\left(R_1+R_4\right)\left(R_2+R_3\right)} - \frac{U\left(R_2R_4-R_3R_1\right)\left(R_2+R_3\right)}{\left(\left(R_1+R_4\right)\left(R_2+R_3\right)\right)^2} \tag{4.1.77}$$

$$\frac{\partial U_i}{dR_4} = \frac{UR_2}{\left(R_1+R_4\right)\left(R_2+R_3\right)} - \frac{U\left(R_2R_4-R_3R_1\right)\left(R_2+R_3\right)}{\left(\left(R_1+R_4\right)\left(R_2+R_3\right)\right)^2} \tag{4.1.78}$$

Die Aussagen über die Empfindlichkeit für die Strommessung gelten auch für die Spannungsmessung. **Ein möglichst lineares Ausgangssignal benötigt hohe**

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

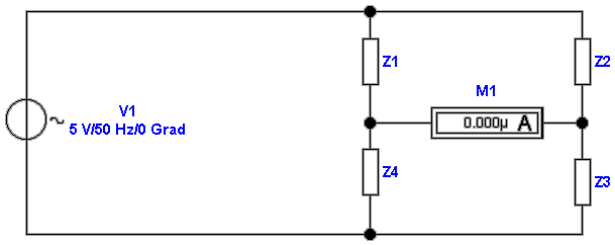



Abbildung 4.65.: Wheatstone-Brücke für allgemeine Impedanzen

**Querwiderstände $R_i$: Spannungsmessungen sind für nicht abgeglichene Brücken vorzuziehen.**
Die Gleichungen für die Wheatstone-Brücke für allgemeine Impedanzen (Abb. 4.65 können aus Gleichung (4.1.66) abgeleitet werden. Folgende Ersetzungen sind vorzunehmen:

$$\begin{aligned} R_1 &\rightarrow & |Z_1|e^{i\varphi_1} \\ R_2 &\rightarrow & |Z_2|e^{i\varphi_1} \\ R_3 &\rightarrow & |Z_3|e^{i\varphi_1} \\ R_4 &\rightarrow & |Z_4|e^{i\varphi_1} \end{aligned} \tag{4.1.79}$$

Gleichung wird dann zu

$$\begin{aligned} \frac{Z_1}{Z_4} &= \frac{Z_2}{Z_3} \\ \Rightarrow \quad \frac{|Z_1|e^{i\varphi_1}}{|Z_4|e^{i\varphi_4}} &= \frac{|Z_2|e^{i\varphi_2}}{|Z_3|e^{i\varphi_3}} \\ \Rightarrow \quad |Z_1|e^{i\varphi_1} * |Z_3|e^{i\varphi_3} &= |Z_2|e^{i\varphi_2} * |Z_4|e^{i\varphi_4} \\ \Rightarrow \quad |Z_1| * |Z_3|e^{i\varphi_1+i\varphi_3} &= |Z_2| * |Z_4|e^{i\varphi_2+\varphi_4} \end{aligned} \tag{4.1.80}$$

Daraus ist ersichtlich, dass eine Brücke nur abgleichbar ist, wenn sowohl die Beträge wie auch die Phasen abgeglichen sind. Diese Bedingungen sind:

$$|Z_1| : |Z_4| = |Z_2| : |Z_3|$$

$$|Z_1|\,|Z_3| = |Z_2|\,|Z_4| \tag{4.1.81}$$

$$\varphi_1 + \varphi_3 = \varphi_2 + \varphi_4 \tag{4.1.82}$$

## 4.1.8. Wandlerschaltungen

Wandlerschaltungen werden benötigt, um digitale mit analogen Schaltkreisen zu verbinden. Während digitale Darstellungen von Signalen prinzipiell mit beliebiger Genauigkeit machbar sind, limitiert das Rauschen von analogen Schaltkreisen (siehe auch Abschnitt 2.8.1). Da Analog-Digitalwandler auf Digital-Analog-Wandlern aufbauen werden zuerst diese beschrieben.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

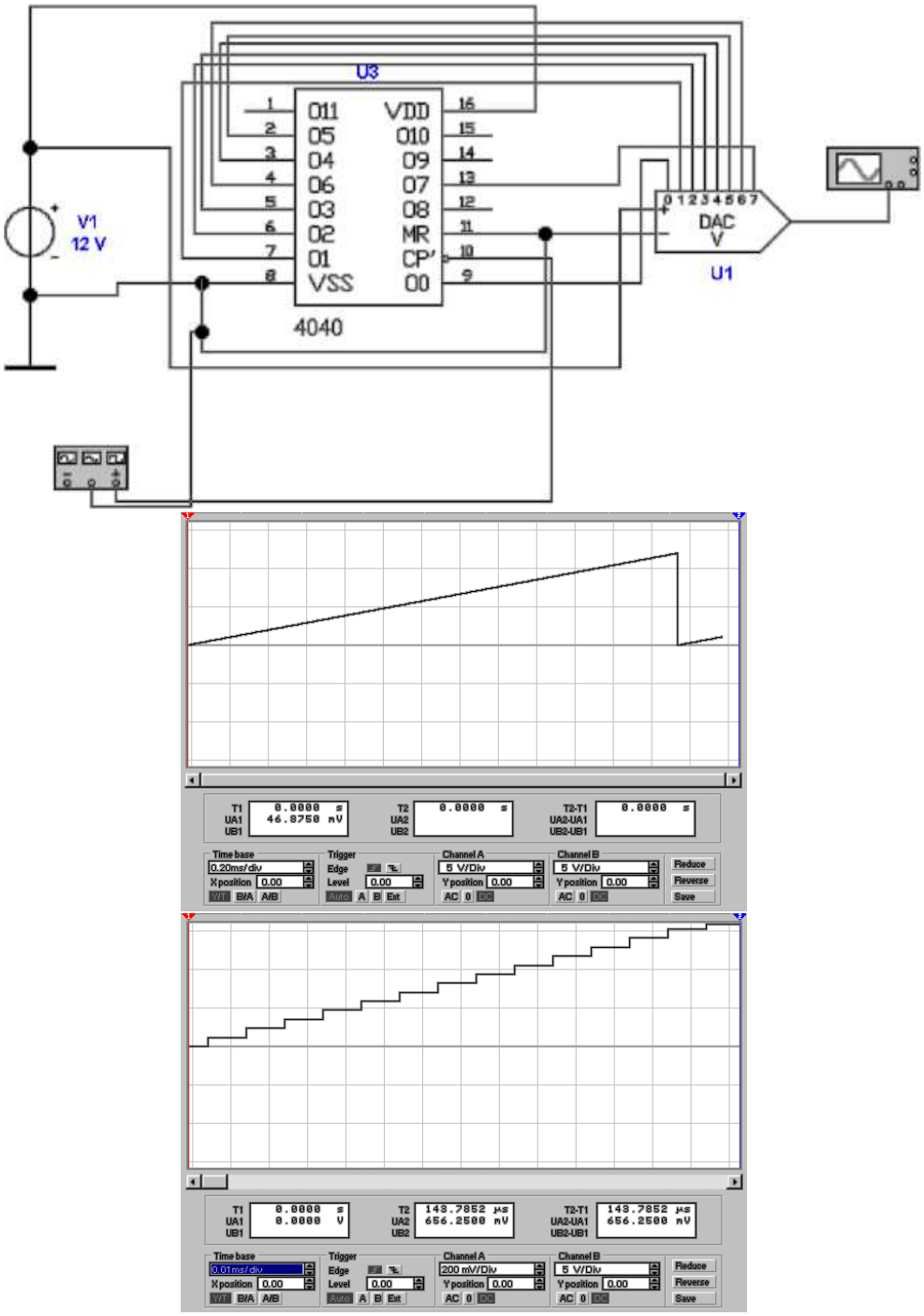



Abbildung 4.66.: Digital-Analog-Wandlerschaltung. Mitte Ausgangskurve, unten: vergrösserte Ausgangskurve

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

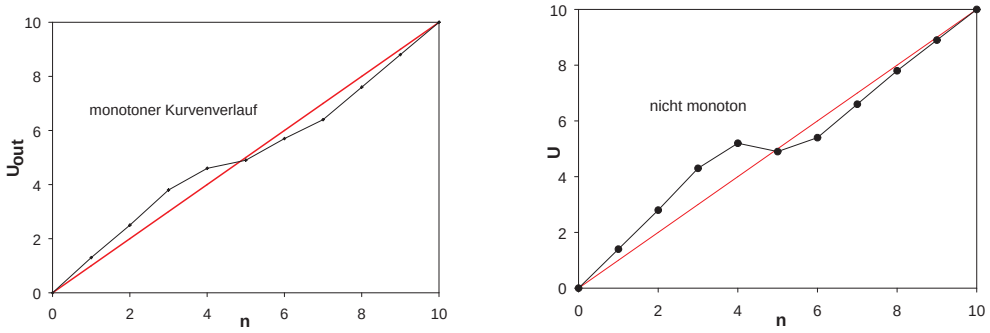



Abbildung 4.67.: Links eine monotone Wandlerkennlinie, rechts eine nicht monotone. Zum Vergleich ist die ideale Ausgangsgerade eingezeichnet.

#### 4.1.8.1. Digital/Analog-Wandler

Die Ausgangsspannung eines Digital/Analog-Wandlers ist prinzipiell wertdiskret. Bei sehr kleinen Diskretisierungsschritten kann das Rauschen von analogen Bauteilen diese Spannungsschritte maskieren. Abbildung 4.66 zeigt die prinzipielle Schaltung sowie die Ausgangskurven. Ein 4040 CMOS-Zähler U3 gespiesen vom Funktionsgenerator zählt von 000 nach FFF. Die untersten acht Bit werden in den generischen, idealen *Digital-Analog-Wandler* U1 gespiesen. Seine Ausgangsspannung wird in der Mitte und unten in Abb. 4.66 gezeigt. Die mittlere Abbildung zeigt die Ausgangsrampe. Um die Stufenhöhe auflösen zu können, ist ein Teil der Messkurve in der unteren Darstellung vergrösstert. Sehr schön sind die einzelnen Stufen im Ausgangssignal zu sehen. da dies ein idealer Wandler ist, sind die Stufen im gleichen Abstand.
Digital-Analogwandler haben die folgenden Fehler:

- Die Stufenhöhe ist nicht konstant. Beide Kurven in in Abb. 4.67 zeigen diesen Fehler.
- Die Ausgangsspannung ist keine monotone Funktion der Eingangsspannung. Die rechte Kurve in Abb 4.67 ist nicht monoton.

Die Grösse der Bitfehler kann auf zwei Arten bestimmt werden. Einerseits kann, wie in Abb. 4.68, linke Seite, eine Gerade durch die Endpunkte als Referenz verwendet werden. Die daraus resultierenden Fehler werden in Tabelle 4.2 in der zweiten und dritten Spalte aufgelistet. Andererseits kann eine durch Regression bestimmte Gerade als Referenz dienen (Abb. 4.68, rechts, und Tabelle 4.2, Spalten 4 und 5). Die so berechneten Fehler sind kleiner[3].

**4.1.8.1.1. Direkte D/A-Wandler** Direkte *Digital-Analog-Wandler* setzen jede Zahlenkombination am Eingang in einen mit einem diskreten Widerstand (hier R12, R14-R20) kodierten Wert um. Die Abbildung 4.69 zeigt eine mögliche Implementierung.

[3]Sie müssen in Datenblättern immer erst die Definition der Fehler nachschauen, um vergleichen zu können.

| Zahl Zahl | Ausgang DAC | Fehler durch Endpunkte | angepasste Gerade | Fehler angep. Gerade |
|---|---|---|---|---|
| 0 | 0 | 0 | 0,44 | 0,44 |
| 1 | 1,3 | -0,3 | 1,36 | 0,06 |
| 2 | 2,5 | -0,5 | 2,28 | -0,22 |
| 3 | 3,8 | -0,8 | 3,2 | -0,6 |
| 4 | 4,6 | -0,6 | 4,12 | -0,48 |
| 5 | 4,9 | 0,1 | 5,04 | 0,14 |
| 6 | 5,7 | 0,3 | 5,96 | 0,26 |
| 7 | 6,4 | 0,6 | 6,88 | 0,48 |
| 8 | 7,6 | 0,4 | 7,8 | 0,2 |
| 9 | 8,8 | 0,2 | 8,72 | -0,08 |
| 10 | 10 | 0 | 9,64 | -0,36 |

Tabelle 4.2.: Tabelle der Ausgangswerte von Digital-Analog-Wandlern. Die erste Spalte zeigt die Zahlenwerte. Die zweite Spalte die Ausgangswerte des Wandlers. In der dritten Spalte werden die Fehler angegeben, bezogen auf eine Gerade durch die Endpunkte. Die vierte Spalte zeigt die Ausgangsgerade $U = 0.44 + 0.92n$. Die letzte Spalte zeigt die Fehler dazu.

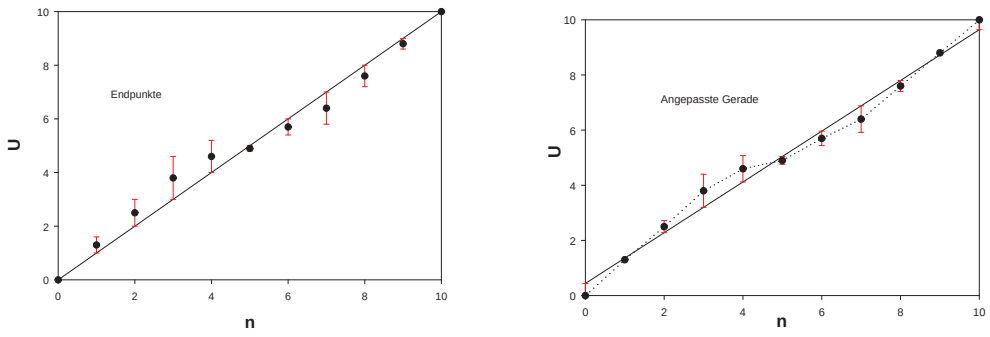



Abbildung 4.68.: Berechnung der Fehler: Links mit einer Kurve durch die Endpunkte, rechts eine mit angepasster Gerade

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

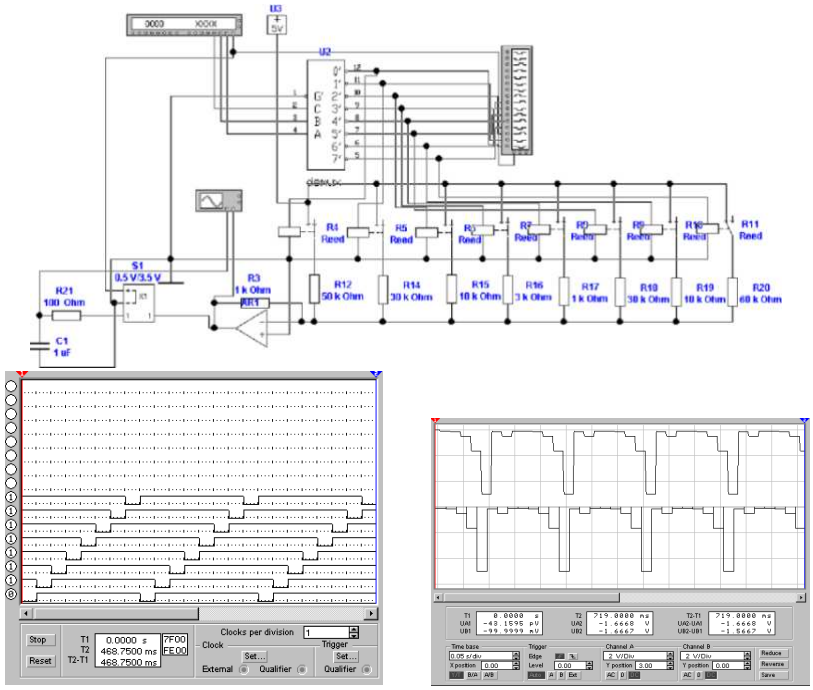


Abbildung 4.69.: Direkter Digital/Analog-Wandler. Oben die Schaltung, unten links der Logikanalysator und unten rechts das Oszilloskopbild.

Mit einem Bitmustergenerator werden die Zahlen von 0H bis FH generiert. Das niederwertigste Bit dient dabei als Taktgenerator.Der Demultiplexer U2 set eine 3-Bit-Zahl (0-7) in acht Ausgänge um, die je einzeln auf dem 0-Pegel liegen. Dies ist aus dem Logik-Analysator (Abb. 4.69, unten links) ersichtlich. Die einzelnen Ausgänge des Demultiplexers steuern einzelne Relais an, die die individuell programmierbaren Widerstände $R_{12}$, $R_{14} - R_{20}$ einzeln mit der Referenzspannung $U_3$ versorgen. Der Operationsverstärker $AR_1$ summiert die Stromwerte auf. Am Eingang Y2 des Oszilloskopes, (Abb. 4.69, unten rechts) sieht man, dass Spannungssprünge auftreten. Dies kann verhindert werden, indem eine sogenannte Deglitcher-Schaltung nachgeschaltet wird. Sie besteht hier aus einem Analog-Schalter $S_1$, dem mit $R_{21}$ und $C_1$ ein Tiefpassglied nachgeschaltet ist. Dieses Tiefpassglied dient als Analogspeicher und speichert während der Glitch-Phase das *Signal* zwischen. Die Schaltung, bestehend aus $S_1$, $R_{21}$ und $C_1$ wird auch Sample&Hold-Schaltung oder *Abtast-Halte-Glied* genannt.

Die direkte Wandlung von digitalen zu analogen Signalen kann sehr schnell sein, bedingt aber einen enormen Aufwand an Widerständen und vor allem, Schaltern.

**4.1.8.1.2. Stromwägeverfahren** Das Stromwägeverfahren, wie es in Abbildung 4.70 gezeigt wird, ist eine sehr viel effizientere Möglichkeit, digitale Zahlenwerte in Spannungen umzuwandeln. Anders als beim direkten Verfahren (Absatz 4.1.8.1.1) können jedoch die Ausgangswerte pro Bit nicht unabhängig gewählt werden.

In Abbildung 4.70 ist ein 4-Bit *Digital-Analog-Wandler* gezeigt. Die Analogschalter $S_1$ bis $S_4$ verbinden die Widerstände $R_2$, $R_4$ bis $R_6$ mit dem als Strom-Spannungsschalter geschalteten Operationsverstärker. Der Referenzwiderstand ist hier $R_1 = 1k\Omega$. Wenn wir den Referenzwiderstand mit $R_{ref}$ und den Widerstand

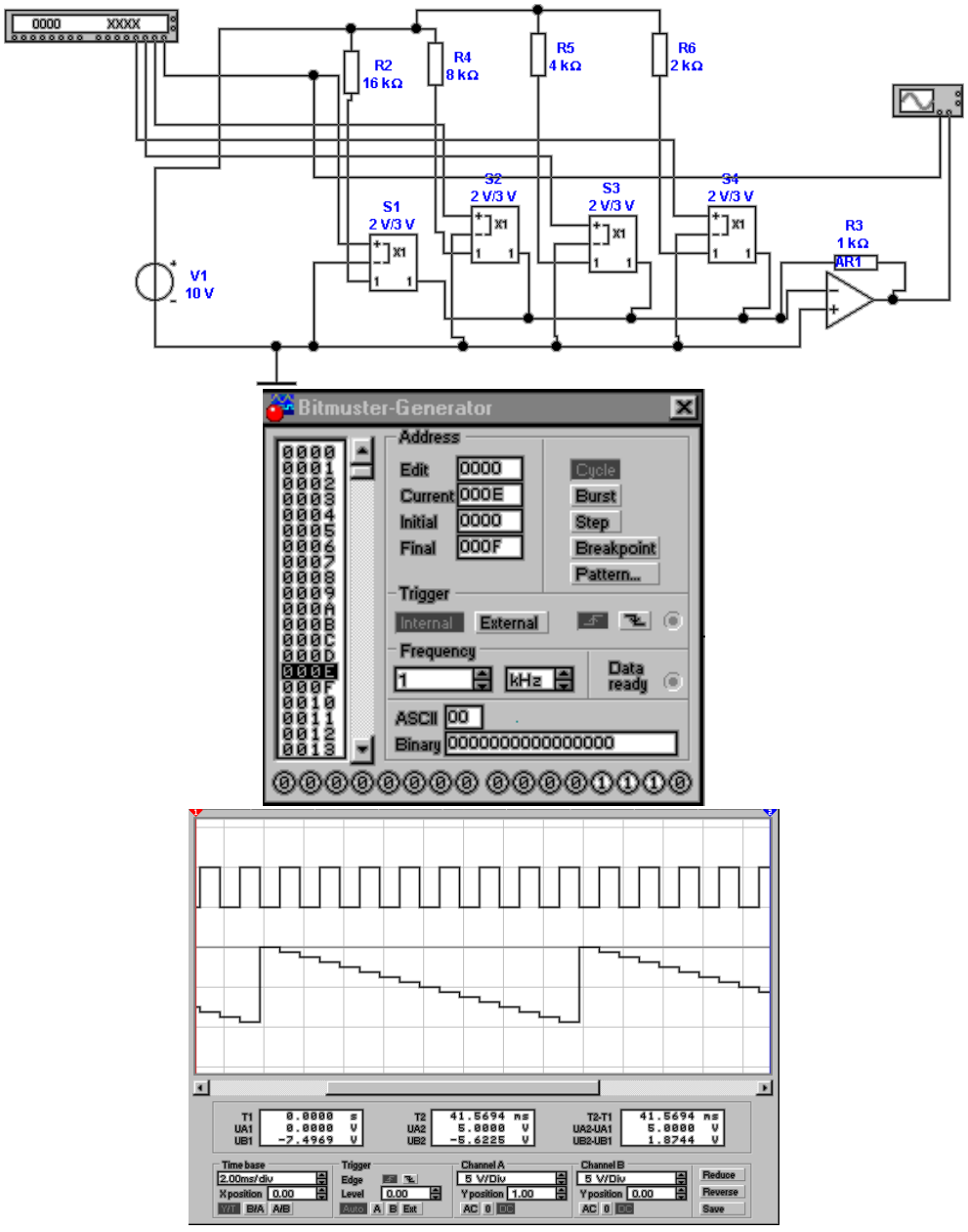



Abbildung 4.70.: Digital-Analog-Wandlerschaltung mit Analogschaltern. Rechts Bitmustergenerator, unten: Ausgangskurve

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

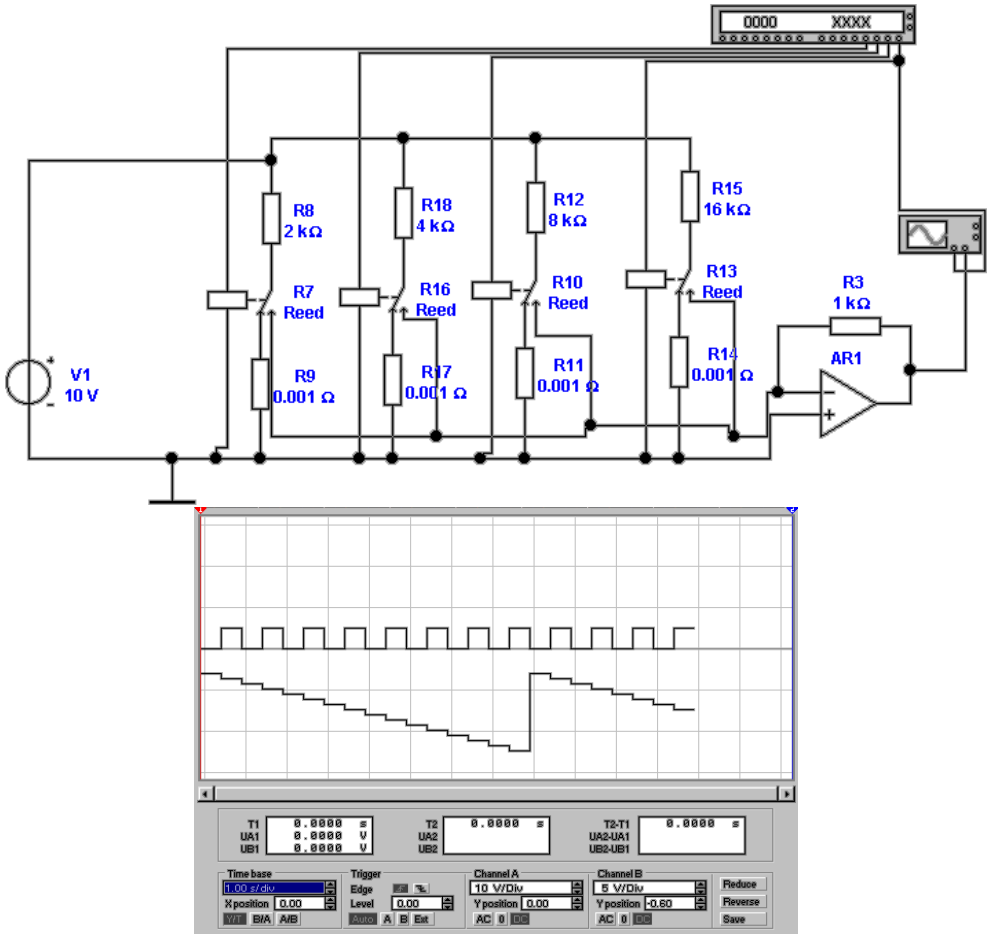



Abbildung 4.71.: Digital-Analog-Wandlerschaltung mit Stromwechselschaltern. Unten: Ausgangskurve

für das m-te Bit ($0 \leq m < n$ bei einem n-Bit-Wandler) dann gilt

$$R_m = R_{ref} * 2^{1+n-m} \quad m = 0 \ldots n-1 \tag{4.1.83}$$

Hier ist $n$ die Bitzahl des Wandlers, $n-1$ ist das höchstwertige Bit und 0 das niederwertigste Bit. Damit die Spannungsänderung von Bit 0 nicht in der mangelnden Genauigkeit von Bit $m$ untergeht, muss für die Widerstandstoleranz gelten:

$$\frac{\triangle R_m}{R_m} \leq \frac{1}{2^{m+1}} \tag{4.1.84}$$

Bei einem 8-Bit Wandler bedeutet dies eine Genauigkeit des kleinsten Widerstandes (grössten Leitwertes) von $\frac{1}{256} = 0.4\%$. Bei einem 12-Bit Wandler muss die Genauigkeit $\frac{1}{4096} = 0.024\%$ sein, bei einem 16-Bit Wandler[4] $\frac{1}{65536} = 0.0015\%$ sein. Dies sind illusorische Genauigkeiten, die nur mit verheerend grossen Kosten erreichbar wären.
In der Abbildung 4.70 wird in der Mitte die Einrichtung des Bitmustergenerators und unten das Ausgangssignal gezeigt. Die Stufigkeit dieses Signals kommt dabei sehr schön zum Ausdruck.

[4]Das entspricht der CD-Qualität

Digital/Analogwandler nach Abbildung 4.70 belastet die Stromquelle sehr ungleichmässig. Deshalb wird in der Strom durch die Widerstände nicht unterbrochen, sondern wie in Abbildung 4.71 nur zwischen dem invertierenden Eingang des Strom-Spannungswandlers und der Erde geschaltet. Gleichzeitig erreicht man, dass über dem geschlossenen Schalter keine Spannung abfällt, dass also Leckströme sehr effektiv unterdrückt werden. Für die Dimensionierung der Widerstände gilt Gleichung (4.1.83).
Der Digital-Analogwandler nach Abbildung 4.72 verwendet ein R-2R-Netzwerk, bei dem nur zwei Widerstandswerte vorkommen, nämlich R und 2R. Dieses Netzwerk kann wie folgt verstanden werden. Wir betrachte das rechte Ende der Kette, beginnend mit dem Knoten **1** zwischen $R_{20}$, $R_{21}$ und $R_{24}$. der Knoten **2** liegt zwischen $R_{21}$, $R_{22}$ und $R_{23}$. Die Widerstände $R_{23}$ und $R_{24}$ speisen den Strom in den Digital/Analog-Wandler. Dabei soll der Strom durch $R_{24}$, $I_{24}$ gleich dem doppelten des Stromes $I_{23}$ durch $R_{23}$ sein. Damit gilt: $U_1 = 2U_2$. Somit können wir das folgende Gleichungssystem aufstellen:

$$\begin{aligned} I_{21} &= I_{22} + I_{23} \\ I_{21} &= \frac{U_1 - U_2}{R_{21}} \\ I_{22} &= \frac{U_2}{R_{22}} \\ I_{23} &= \frac{U_2}{R_{23}} \\ U_1 &= 2U_2 \end{aligned}$$

Eingesetzt ergibt sich

$$\frac{U_1 - \frac{U_1}{2}}{R_{21}} = \frac{U_1}{2R_{22}} + \frac{U_1}{2R_{23}} \tag{4.1.85}$$

Diese Gleichung ist unabhängig von $U_1$. Man bekommt

$$R_{21} = \frac{R_{22}R_{23}}{R_{22} + R_{23}} \tag{4.1.86}$$

Wird in Gleichung (4.1.86) $R_{23} = R_{22}$ gesetzt, so ist $R_{21} = \frac{R_{22}}{2}$. Am Knoten **1** hat die Kombination aus $R_{21} \ldots R_{23}$ die Impedanz $R_{21} + (R_{22} \| R_{23}) = R_{21} + (2R_{21} \| 2R_{21}) = R_{21} + R_{21} = 2R_{21} = R_{22}$. Damit kann der Knoten **1** wie der Knoten **2** behandelt werden: der Strom verdoppelt sich bei diesem Netzwerk, wenn man nach links geht, und er halbiert sich, wenn man nach rechts geht.
Das R-2R-Netzwerk hat den Vorteil, dass nur Widerstände, die in der Grösse um den Faktor 2 variieren, auf der Chipfläche hergestellt werden müssen. Damit ist diese Struktur kompatibel zu der Halbleiterfertigung. Die grössten Fehlern sind

- Widerstand der Schalter im eingeschalteten Zustand
- Offsetspannungen
- Bei grossen Bitzahlen werden die Ströme so klein, dass thermische Ströme oder Rauschströme grösser werden können.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

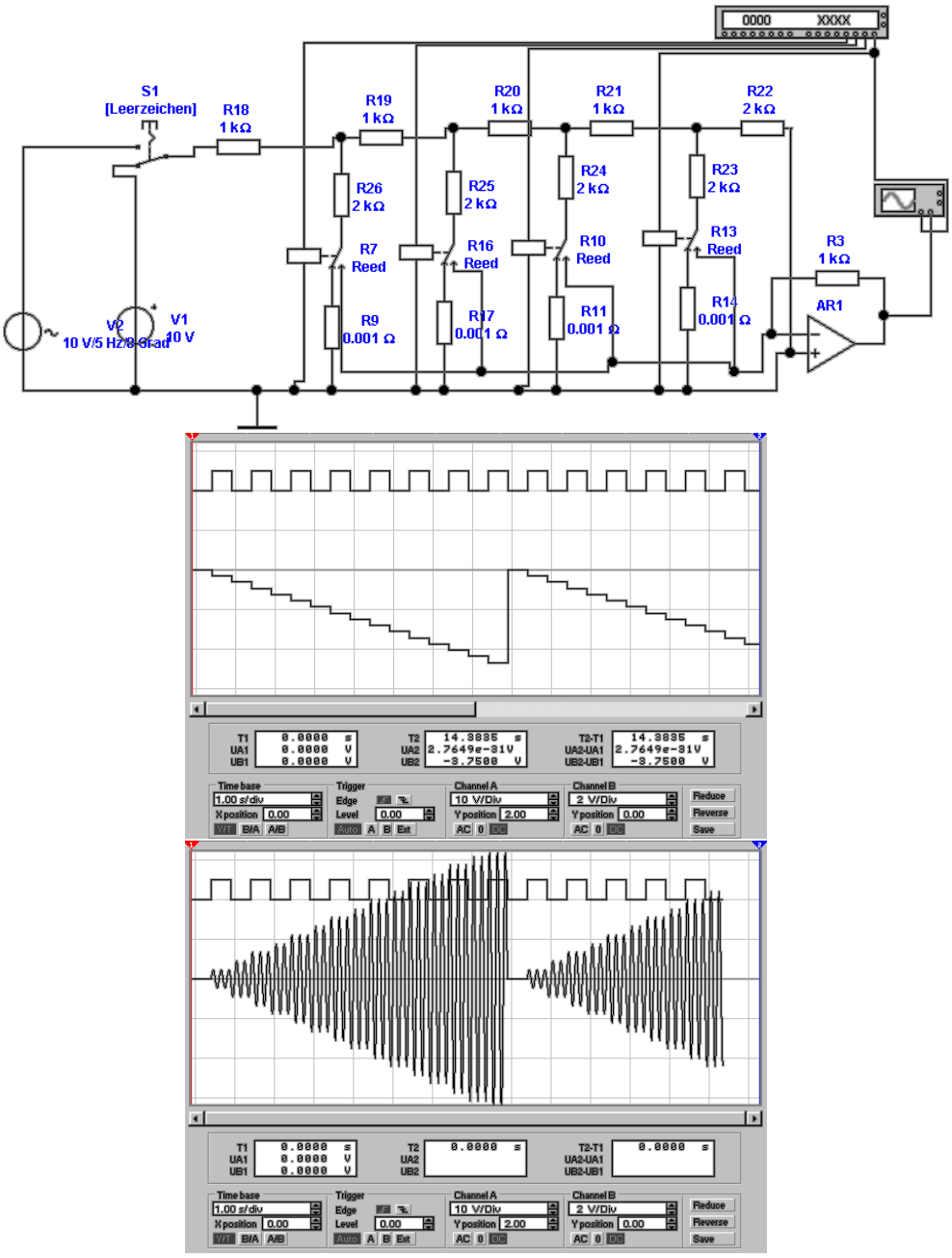



Abbildung 4.72.: Digital-Analog-Wandlerschaltung mit Stromwechselschaltern und R-2R-Netzwerk. Mitte: Ausgang bei Ansteuerung mit einer Gleichspannung. Unten: Ausgang bei Ansteuerung mit einer Wechselspannung.

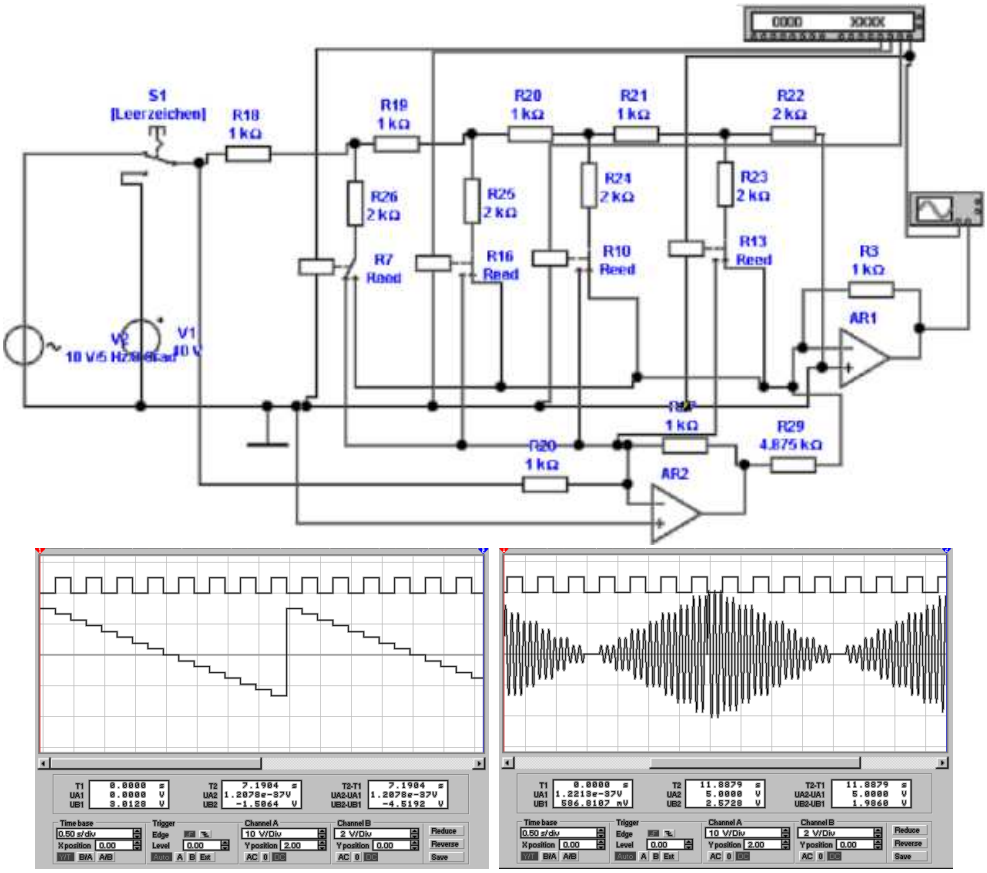



Abbildung 4.73.: Digital-Analog-Wandlerschaltung mit Stromwechselschaltern, R-2R-Netzwerk und bipolarem Ausgang. Mitte: Ausgang bei Ansteuerung mit einer Gleichspannung. Unten: Ausgang bei Ansteuerung mit einer Wechselspannung.

Die *Spannungsquelle* wird mit dem Widerstand $R_{21}$ belastet. $R_{18}$ kann auch weggelassen werden. Dann ist die Belastung der Referenzquelle $R_{22}$.
Schliesslich ist es möglich, wie in Abb. 4.72 gezeigt, das Netzwerk auch mit veränderlichen Spannungen zu betreiben. Die Digitalzahl wirkt dabei wie ein Multiplikator.
Die Schaltung nach Abbildung 4.73 erweitert das R-2R-Netzwerk mit einem bipolaren Ausgang.

**4.1.8.1.3. Pulslängenmodulation** Bei den oben besprochenen Digital-Analog-Wandlern gibt es zwei fundamentale Probleme:

- Beim Umschalten des MSB können grosse Störsignale entstehen. je mehr Bit *Auflösung* ein Wandler hat, desto schwieriger wird es, dieses Problem in den Griff zu bekommen.
- Die Linearität kann bei sehr hochauflösenden Wandlern nicht mehr garantiert werden.
- Und nicht zuletzt, die Kosten für einen Wandler steigen überproportional mit seiner Bitzahl.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

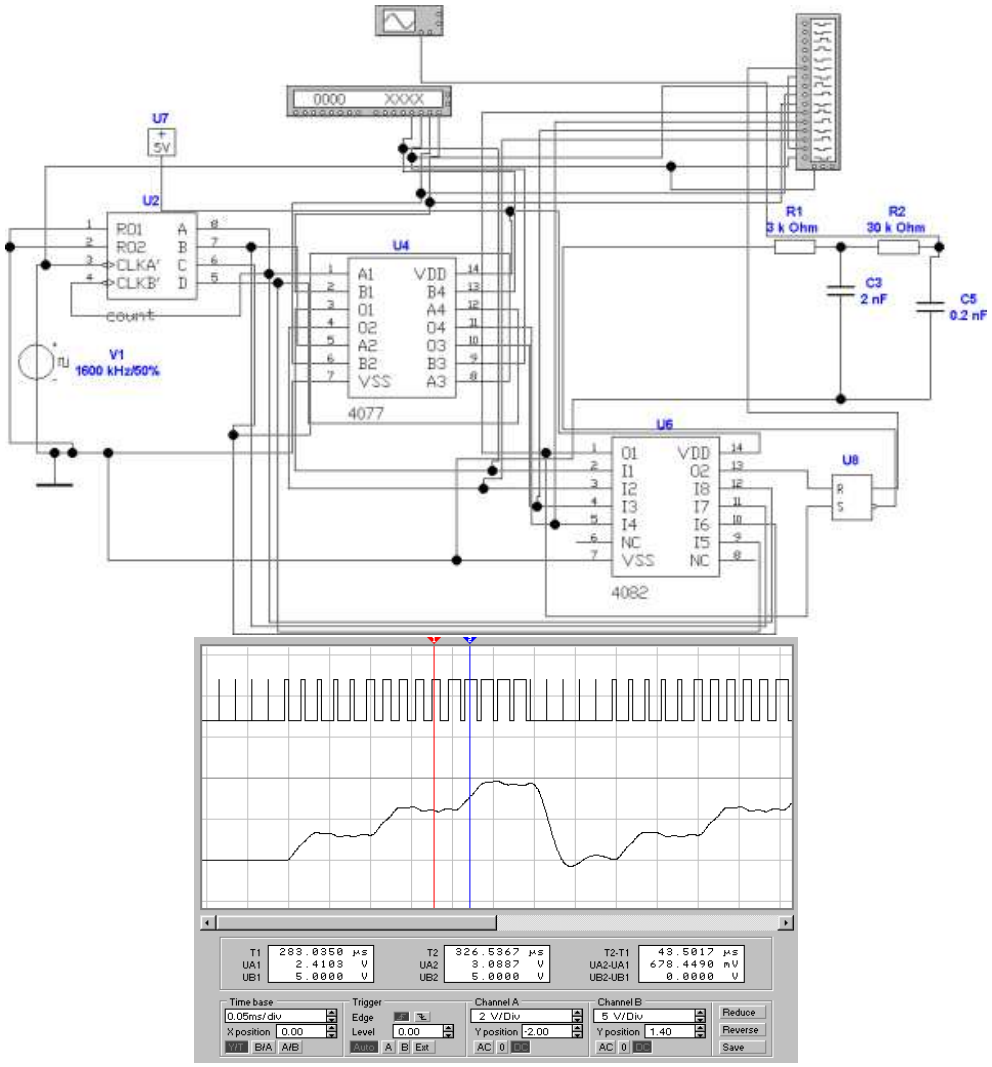



Abbildung 4.74.: Digital-Analog-Wandlerschaltung mit Pulslängenmodulation. Unten: Ausgang mit Pulsweitensignal und tiefpassgefiltertem *Signal*.

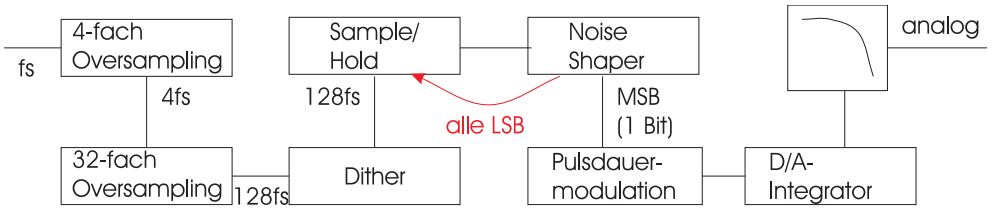



Abbildung 4.75.: 1-Bit Wandler.

Dieses Problem kann umgangen werden, indem man die Ausgangsspannung wie in Abb. 4.74 zwischen zwei Spannungswerten hin- und herschaltet und dabei die Pulslänge moduliert.

Das Eingangssignal stammt von dem Bitmustergenerator. Um das Prinzip klarzumachen, verwenden wir nur zwei Bits. Mit dem Taktgenerator $V_1$ wird der Binärzähler $U_2$ angesteuert. Das XOR-Gatter $U_4$ vergleicht jeweils ein Bit des Zählers mit dem entsprechenden Bit des Bitmustergenerators. Das AND-Gatter $U_{13}$ detektiert, wenn beide Bit-Werte vom Zähler und vom Bitmustergenerator gleich sind. Dann wird das RS-Flip-Flop $U_8$ zurückgesetzt. Mit der fallenden Flanke von Bit B aus dem Zähler $U_1$ wird der Monoflop $U_6$ getriggert. Sein Ausgangspuls setzt das RS-Flip-Flop $U_8$. Damit ist das Ausgangssignal von $U_8$ eins, solange der Zähler eine kleinere Zahl als der Bitmuster-Generator hat. Das Eingangssignal könnte auch von einem Computer kommen. Das Ausgangssignal wird Null für den Rest der Periode. Damit ist diese Schaltung ein digitaler Pulsweiten-Modulator. Schliesslich wird das Ausgangssignal in der Schaltung $A_1$ Tiefpassgefiltert (Die Parameter ergeben ein Tschebyscheff-Tiefpassfilter dritter Ordnung mit 0.5 dB Welligkeit und einer Grenzfrequenz von 10kHz).

Die Abb. 4.74, unten zeigt in der oberen Hälfte des Oszilloskpbildes das Pulsweitensignal und unten das gefilterte Ausgangssignal. Die Bitstufen im 8 kHz-Takt sind klar getrennt. Die Taktfrequenz ist dabei 200 kHz.

**4.1.8.1.4. 1-Bit Wandler** 1-Bit-Wandler kombinieren das Pulslängen-Modulationsprinzip mit zusätzlicher digitaler Logik. Abbildung 4.75 zeigt eine mögliche Schaltung aus einem CD-Spieler[**?**]. Das abgetastete *Signal* wird zuerst 4-fach interpoliert (oversampled) und dann 32-fach interpoliert. Aus der CD-Abtastfrequenz von 44.1 kHz wurde nun eine Abtastfrequenz von 5.6 MHz. Das *Signal* wird mit einem 352 kHz-*Signal* so digital moduliert, dass die Ausgangsspannung um das wenigstwertige Bit (LSB) schwankt. damit muss der Wandler konstant das Ausgangssignal ändern: es können keine niederfrequenten Störsignale entstehen. In einer digitalen Sample/Hold-Stufe werden die Datenworte verdoppelt. Die Abtastfrequenz ist nun 11.2 MHz. Aus diesem *Signal* wird das MSB-Bit im Noise-Shaper abgetrennt und dem Pulsweiten-Modulator (1-Bit!) zugeführt. Die nicht-verwendeten Signalbits werden zum nächsten Datenbit im Sample/Hold dazu gezählt und gehen so nicht verloren. Die Wandlung wird nun von einer 1-Bit Puls-Dichte-Modulatorstufe durchgeführt, deren niedrigste Frequenz durch die 352kHz-Modulationsspannung gegeben ist. Das Ausgangssignal wird nun durch einen Integrator tiefpassgefiltert.

Die Bitzahl m, die man gewinnen kann, hängt von der Oversamplingrate r wie

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

| Oversampling | Erhöhung der Bit-Zahl |
|---|---|
| 1 | 0 |
| 2 | 1 |
| 4 | 2 |
| 8 | 3 |
| 16 | 4 |
| 32 | 5 |
| 64 | 6 |
| 128 | 7 |
| 256 | 8 |

Tabelle 4.3.: Erhöhung der Bitzahl durch oversampling

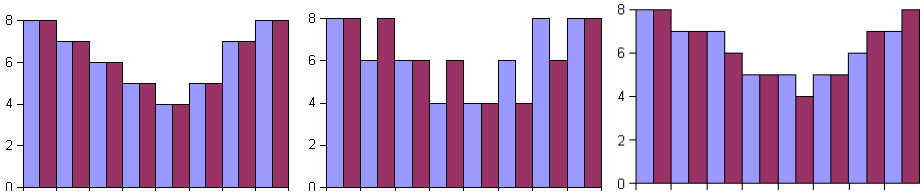


Abbildung 4.76.: Oversampling. Links ist der Output, wie ihn ein 3 Bit- Wandler erzeugen würde. In der Mitte ist das Ausgangssignal eines 2-Bit-Wandlers. Dabei schwankt das letztwertige Bit, wenn links eine ungerade Zahl herausgegeben wurde. Rechts ist der gemittelte (Tiefpass-gefilterte) Ausgang. Die weinroten Balken haben den gleichen Wert links und rechts.

folgt ab:

$$m = \log_2(r) \tag{4.1.87}$$

Tabelle 4.3 zeigt einige charakteristische Werte. Warum funktioniert die 1-Bit-Wandlung, obwohl man im obigen Beispiel ausrechnen kann, dass die Bitzahl $1 + 8 = 9$ Bit ist. das menschliche Ohr hat seine maximale Empfindlichkeit bei 1 kHz, so dass für diese Frequenz nochmals eine etwa 32-fache Überabtastung resultiert. Damit ist die effektive Bitzahl $9 + 5 = 14$. Der Noise-Shaper ermöglicht eine zusätzliche, digitale Erhöhung der Quantisierung.

**4.1.8.1.5. Oversampling** Die Funktionsweise des Oversampling wird in Abb. 4.76 gezeigt. Links wird das Ausgangssignal, wie es ein 3-Bit-Wandler erzeugen würde, gezeigt. Das mittlere Bild stellt den Ausgang eines 2-Bit-Wandlers dar. Wenn dabei das ursprüngliche *Signal* zwischen zwei möglichen Ausgangswerten liegt, wird das Ausgangssignal zwischen den beiden, dem ursprünglichen *Signal* benachbarten werten, hin-und hergeschaltet. Der mittlere Teil von Abb. 4.76 zeigt das entsprechende *Signal*. Der rechte Teil von Abb. 4.76 zeigt das mit einem gleitenden Mittelwert gefilterte Ausgangssignal. Betrachtet man nur die weinroten Balken im linken und im rechten Diagramm, stellt man fest, dass sie identisch sind.
Abb. 4.77 zeigt ein vierfach-Oversampling. Links ist das ursprüngliche *Signal* mit 4-Bit *Auflösung*. Das mittlere Bild zeigt das Interpolationsresultat wen je-

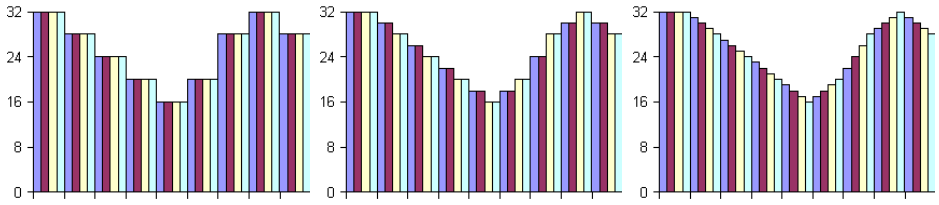



Abbildung 4.77.: 4-fach Oversampling. Links ist der Output, wie ihn ein 4 Bit-Wandler erzeugen würde. In der Mitte ist in einer ersten Stufe die Ausgangsfrequenz verdoppelt und die Werte gemittelt. Rechts ist der doppelt gemittelte (Tiefpass-gefilterte) Ausgang.

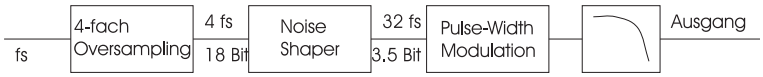



Abbildung 4.78.: MASH (Multi-Stage Noise Shaping)-Wandlung

weils über 2 und 2 Ausgangsbalken gemittelt wird. Rechts ist das voll interpolierte 4-fach Oversampled-*Signal*. bei CD-Plattenspielern wird bei 16-Bit Digital-Analogwandlern maximal 16-fach überabgetastet. mehr macht nicht Sinn, da 16-Bit Wandler damit an ihre Geschwindigkeitsgrenze kommen[**?**]. Das verfahren ist bekannt unter dem Namen High-Bit-Verfahren.

Wie das Beispiel mit der Pulsweiten-Modulation gezeigt hat, kann es sinnvoll sein, die Taktfrequenz sehr hoch zu setzen und die *Auflösung* durch Filteroperationen im digitalen Bereich zu erhalten. Werden weniger Bits und preiswerte analoge Filter verwendet, nennt man das Verfahren Bitstream-Verfahren.

**4.1.8.1.6. MASH-Verfahren** Eine Weiterentwicklung des Bitstream-Verfahrens ist die MASH-Technik. MASH heisst **M**ulti-**S**tage noise **SH**aping. Dabei wird, wie in Abb. 4.78 gezeigt, der Puls-Weiten-Modulator mit mehr als einem Bit angesteuert. Die Pulsweite kann dabei in $2^k$ Schritten eingestellt werden. Damit hat man eine höhere *Auflösung* im Wandler. Da die Bitzahl klein bleibt, ist die notwendige Präzision gewährleistet. Die Netto-Auflösung ist dann

$$n_{Auflösung} = k_{Wandler} + m_{Oversampling} \tag{4.1.88}$$

Bei einem Pulsweiten-Modulator mit 4-Bit *Auflösung*, der mit 45,1 MHZ betrieben wird (1024 Oversampling) stehen bei 44.1 kHz 14 Bit zur Verfügung. Bei dem für den Hörer wichtigen Frequenzbereich von unter 5 kHz stehen nun 17 Bit zur Verfügung.

### 4.1.8.2. Analog/Digital-Wandler

*Analog-Digital-Wandler* bereiten analoge Eingangssignale in digitale Signale auf. Da es bei der Analog-Digital-Wandlung keine so einfachen Konzepte wie das Successive Approximation Verfahren gibt, ist die Umsetzung von analogen in digitalen Signale meistens mit einem grösseren Aufwand verbunden.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

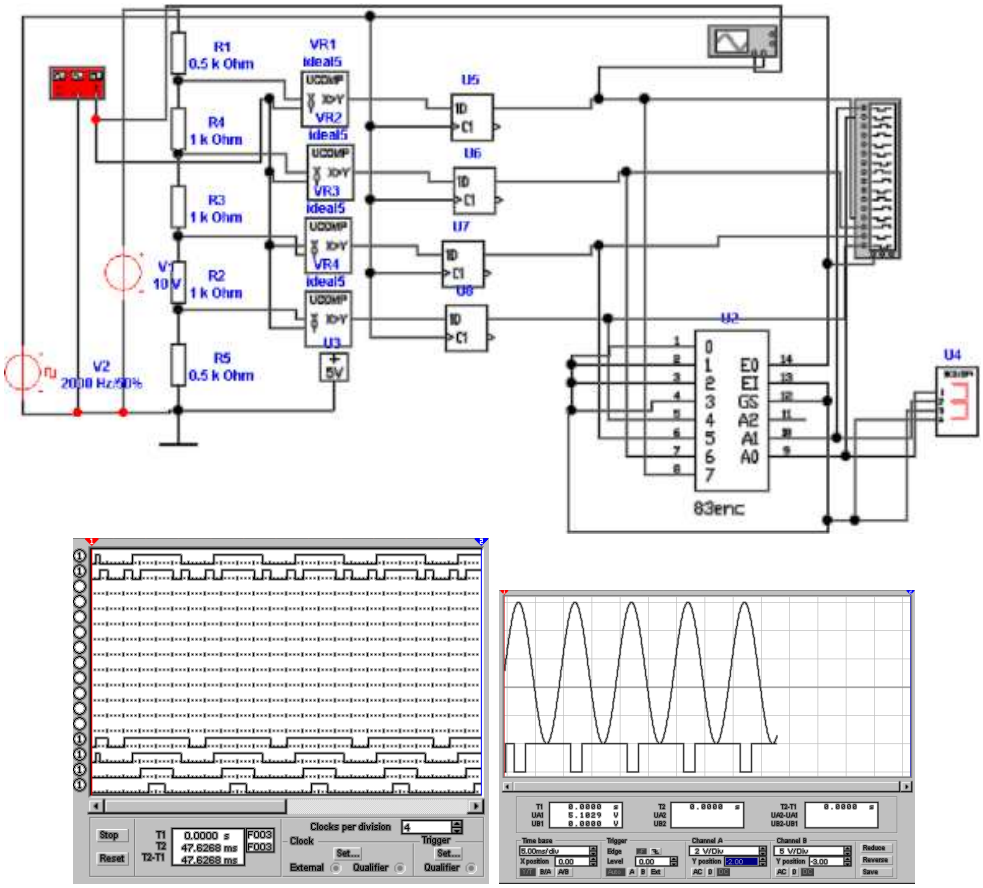



Abbildung 4.79.: Direkter *Analog-Digital-Wandler*. Unten links: Logikdiagramm. Unten rechts: Komparator

**4.1.8.2.1. Direkte Verfahren** Das am einfachsten zu begreifende Verfahren für die Analog-Digital-Wandlung ist das Direktverfahren[5], wie es in Abb. 4.79 gezeigt wird. Als Beispiel wird eine 2-Bit Wandlung gezeigt. Die Referenzspannung wird durch eine Teilerkette $R_1 \ldots R_5$ in gleichabständige Spannungswerte gewandelt. dabei sind die Spannungswerte jeweils um ein halbes LSB verschoben, um eine korrekte Wandlung zu erreichen. Die Komparatoren $VR_1 \ldots VR_4$ vergleichen die Eingangsspannung mit den Referenzwerten. Die D-Flipflops $U_5 \ldots U_8$ transferieren zu einem genau festgelegten Zeitpunkt die Komparatorsignale an den Ausgang. Diese sind im Logikanalysator (Abb. 4.79, unten links) zu sehen. der Priority-Encoder $U_2$ gibt nun ein 2-Bit-Ausgangssignal, das von der Adresse des höchstwertigen Komparators bestimmt ist. Die LED-anzeige stellt den Wert dar. Das Oszilloskopbild in Abb. 4.79, unten rechts, zeigt das Verhältnis des obersten Komparators zum Eingangssignal.
*Analog-Digital-Wandler* nach diesem Prinzip arbeiten bis in den GHz-Frequenzbereich. Damit dienen sie zur Wandlung von Videosignalen und werden in digitalen Höchstfrequenz-Oszilloskopen eingesetzt. Sie werden typischerweise mit 8-Bit *Auflösung* hergestellt.

**4.1.8.2.2. Nachlaufverfahren** Das Nachlaufverfahren nach Abb. 4.80 verwendet einen *Digital-Analog-Wandler* und eine Nachlaufregelung um ein Analogsignal zu wandeln. Das Eingangssignal wird im Subtrahierer $A_1$ vom Ausgangssignal des Digital-Analog-Wandlers $U_7$ abgezogen. Der Komparator $VR_1$ vergleicht die Differenz mit 0 und steuert so die Zählrichtung des Up/Down-Zählers $U_{11}$. Der Ausgang dieses Zählers steuert den *Digital-Analog-Wandler* $U_7$. Das *Signal* wird auch in der 7-Segment-Anzeige und im Logikanalysator angezeigt.
Die Abb. 4.80 unten links zeigt das Bild des Logikanalysators. Unten rechts wird schliesslich das Ausgangssignal des Digital-Analog-Wandlers mit dem Eingangssignal verglichen.
Nachlaufende *Analog-Digital-Wandler* benötigen *Digital-Analog-Wandler* mit monotoner Ausgangskennlinie. Sie sind sehr schnell bei kleinen Änderungen, benötigen aber bis zu $2^n T_{takt}$ zum wandeln eines Spannungssprunges um n Bits.

**4.1.8.2.3. Wägeverfahren** Beim Wögeverfahren (auf englisch: Successive Approximation) wird der gesuchte Zahlenwert schrittweise ermittelt. Abb. 4.81 zeigt das *Blockschema* eines Wandlers nach dem Wägeverfahren. Die Eingangsspannung wird in einem Sample/Hold-Glied zwischengespeichert[6]. Die Wandlung läuft nun folgendermassen ab:

- Die Steuerlogik setzt das höchstwertige Bit.
- Der Komparator vergleicht das Ausgangssignals des Digital-Analog-Wandlers (jetzt die halbe Referenzspannung $U_{ref}$) mit dem Eingangssignal. Ist das Eingangssignal grösser, bleibt das Bit gesetzt, sonst wird es zurückgesetzt.
- Nun wird das nächste Bit gesetzt. Die Spannung am *Digital-Analog-Wandler* ist nun $U_{ref}/4$ oder $3U_{ref}/4$, je nach Ausgang des ersten Schrittes.

[5] Flash-Converter
[6] Anders als bei den vorherigen Verfahren muss bei diesem Wandlerprinzip das Eingangssignal einen konstanten Wert haben

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

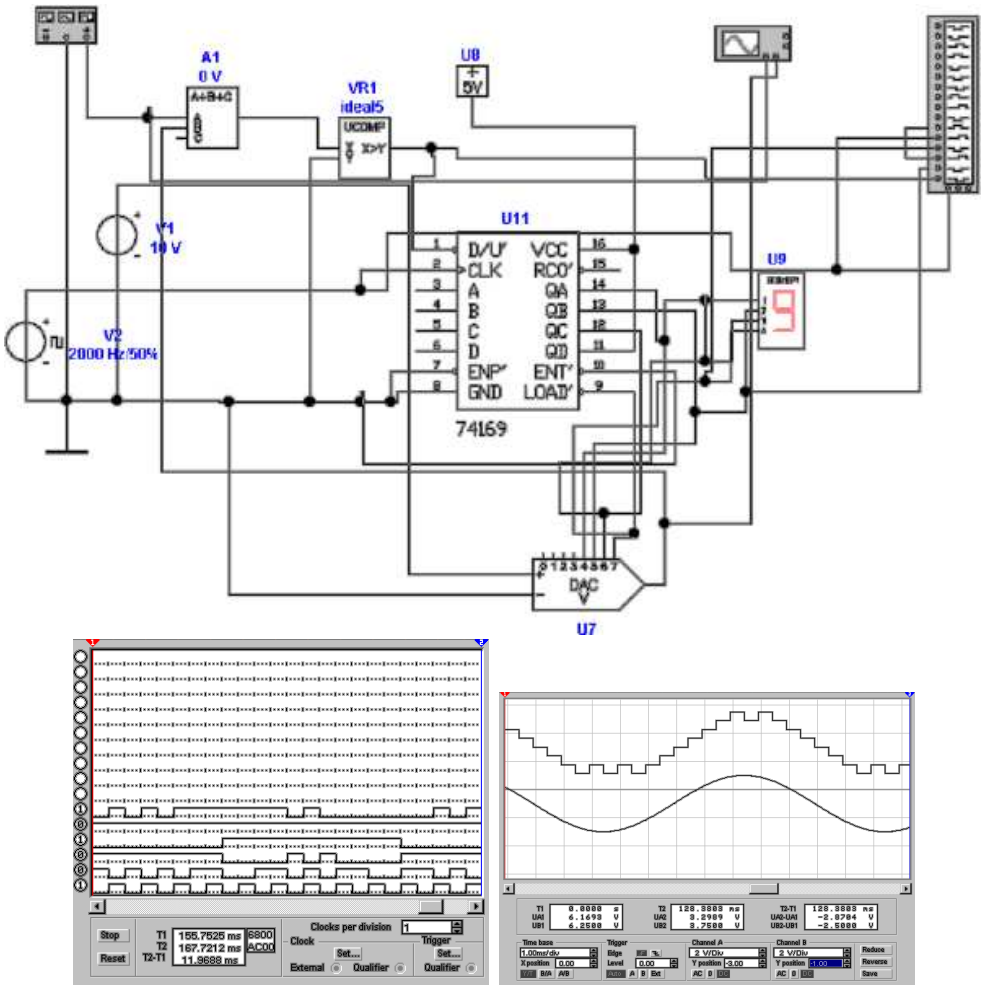



Abbildung 4.80.: *Analog-Digital-Wandler* nach dem Nachlaufverfahren. Unten links: Logikdiagramm. Unten rechts: Komparator

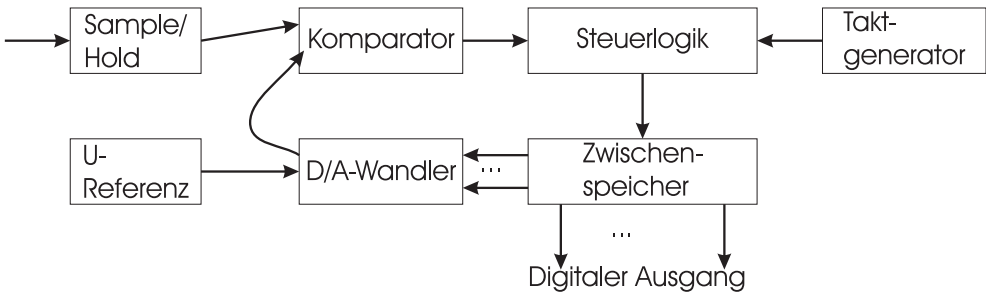



Abbildung 4.81.: *Analog-Digital-Wandler* nach dem Wägeverfahren

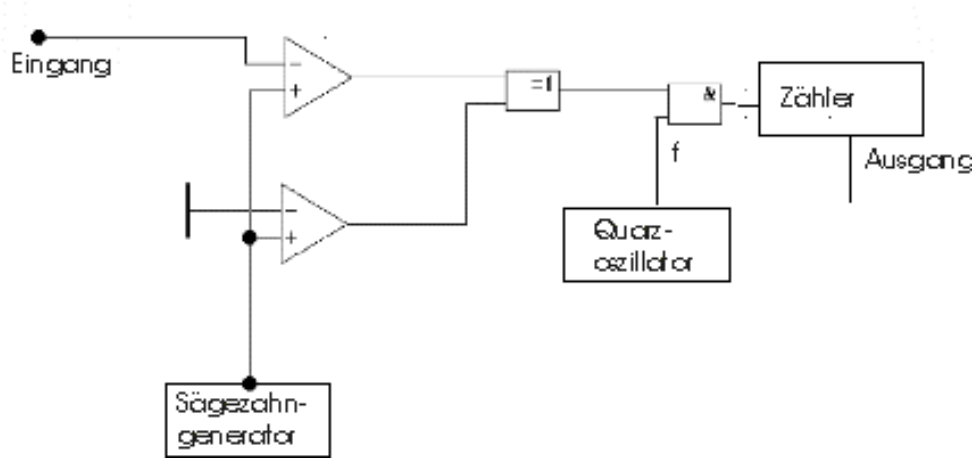



Abbildung 4.82.: *Analog-Digital-Wandler* nach dem Sägezahnverfahren

- Das zweite Bit wird nun gelöscht, wenn die Eingangsspannung kleiner als die Ausgangsspannung des Digital-Analog-Wandlers ist.
- Die obige Prozedur wird für jedes Bit wiederholt.
- Bei einem n-Bit-Wandler steht das Resultat nach n Schritten zur Verfügung.

Wandler nach dem Wägeprinzip benötigen *Digital-Analog-Wandler* die über den ganzen Spannungsbereich monoton sind. Die Wandler haben einen mittleren Geschwindigkeitsbereich, bis einige 10 MHz Taktrate. Dies heisst Wandelzeiten um die 1 $\mu$s.

**4.1.8.2.4. Integrierende Verfahren** Integrierende *Analog-Digital-Wandler* können sehr einfach aufgebaut werden. Abbildung 4.82 zeigt einen Wandler nach dem Sägezahnverfahren. Zwei Komparatoren vergleichen die Sägezahnspannung mit Null und mit der Eingangsspannung. Während die Sägezahnspannung zwischen Null und der Eingangsspannung ist, wird der Quarzoszillator auf den Zähler geschaltet. Die Sägezahnspannung hat den folgenden Funktionsverlauf:

$$V_S = \frac{U_{ref}}{\tau} - V_0 \tag{4.1.89}$$

Die Zeit, während der der Zähler angesteuert wird, ist:

$$\Delta t = \frac{\tau}{U_{ref}} U_e \tag{4.1.90}$$

In der Zeit werden die Schwingungsperioden $T$ des Quarzoszillators gezählt. der Zählerstand ist am Ende der Wandlung:

$$Z = \frac{\Delta t}{T} = \frac{\tau f}{U_{ref}} U_e \tag{4.1.91}$$

Das Sägezahnverfahren funktioniert theoretisch hervorragend. In der Praxis gibt es damit aber fast unüberwindliche Probleme.

- Die Frequenzunsicherheit (Jitter) des Sägezahnoszillators begrenzt die Genauigkeit.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

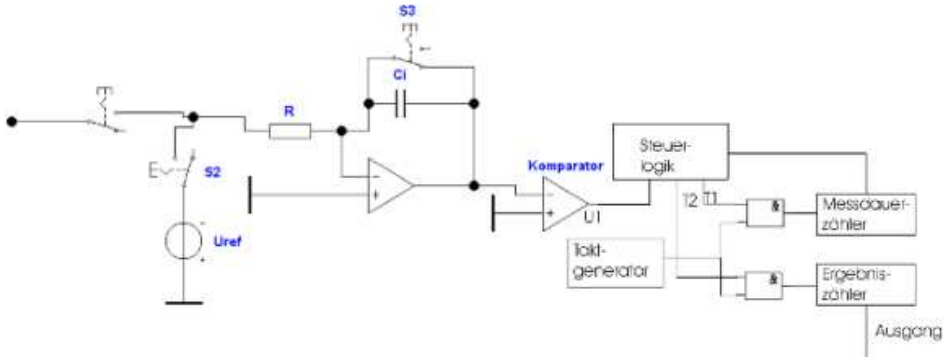



Abbildung 4.83.: *Analog-Digital-Wandler* nach dem Dual-Slope-Verfahren

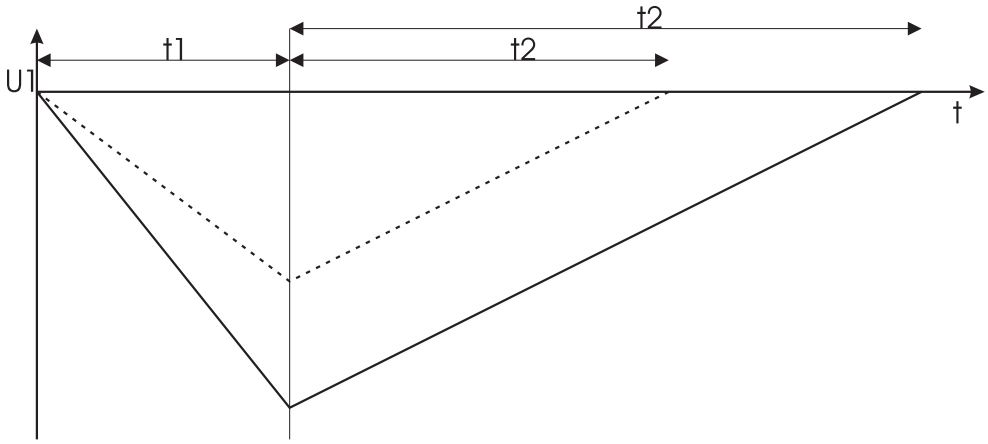



Abbildung 4.84.: Spannungsverlauf beim *Analog-Digital-Wandler* nach dem Dual-Slope-Verfahren. Es ist die Messung einer grossen Spannung (untere Kurve) und einer kleineren Spannung (obere Kurve) angegeben

- Drift und der Einfluss der Temperatur verändern die Schaltschwellen und beeinflussen damit die Genauigkeit.
- Kondensatoren sind schwer mit genügender Genauigkeit zu bekommen.
- Durch den Quarzoszillator und die weiteren Komponenten ist die Schaltung relativ teuer.

Die Abbildung 4.83 zeigt einen Wandler nach dem Dual-Slope-Prinzip. Zuerst wird der Kondensator $C_i$ mit dem Schalter $S_3$ entladen. Dann wird, gesteuert durch die Steuerlogik, das Eingangssignal während einer **festen** Zeit $t_1$ integriert. Dann wird das Eingangssignal vom Integrator getrennt und und die Referenzspannung $U_{ref}$ integriert, bis die Ausgangsspannung des Integrators wieder Null ist. Der Spannungsverlauf ist in Abb. 4.84 für zwei verschiedene Eingangsspannungen gezeigt. Die beiden Zeiten werden vom Messdauerzähler und vom Ereigniszähler bestimmt. Der Ablauf nochmals in Kürze:

1. Integration der Eingangsspannung über eine vorgegebene Zeit $t_1$

2. Integration der fixen Referenzspannung (mit umgekehrter Polarität wie die Eingangsspannung) bis der Kondensator Entladen ist. Diese Zeit $t_2$ wird gemessen.

Die Ausgangsspannung am Integrator nach der Zeit $t_1$ ist

$$U_1(t) = -\frac{1}{\tau}\int_0^{t_1} U - e dt = -\frac{\bar{U_e} n_1 T}{\tau} \tag{4.1.92}$$

Dabei ist $n_1$ die Anzahl Zählimpulse für die Zeit $t_1$. $T$ ist die Periodendauer des Taktoszillators, $\tau$ ist die Integrationskonstante des Integrators. Die Zeit $t_2$ für das Zurückintegrieren ist

$$t_2 = n_2 T = \frac{\tau}{U_{ref}} |U_1(t_1)| \tag{4.1.93}$$

Daraus erhält man für den Zählerstand im Ergebniszähler:

$$Z = n_2 = \frac{\bar{U_e}}{U_{ref}} n_1 \tag{4.1.94}$$

Die Eigenschaften des Dual-Slope-Verfahrens sind:

- Das Ergebnis hängt nicht von der Taktfrequenz ab, da alle Zeiten von ihr abgeleitet werden.
- Der Absolutwert des R-C-Gliedes beeinflusst das Ergebnis nicht. Durch die zweimalige Integration sind die Integrationszeiten und $U_{ref}$ wichtig.
- Das Verfahren ist wenig Anfällig gegen Störspannungen. Alle Frequenzen, die ein Vielfaches von $1/t_1$ sind werden unterdrückt.
- Die Referenzspannungsquelle muss die geforderte Präzision haben.
- Der Integrationskondensator sollte eine möglichst geringe Spannungshysterese haben, also zum Beispiel Polystyrol als Dielektrikum haben.
- Dieser Wandler ist sehr billig herzustellen.

**4.1.8.2.5. Sigma-Delta-Verfahren** Abbildung 4.85 zeigt einen Sigma-Delta-Wandler, der neuerdings die bevorzugte Bauart für höchstauflösende *Analog-Digital-Wandler* ist[7]. Der Wandler besteht aus einem Subtrahierer am Eingang, $A_1$, gefolgt von einem Integrierer, $A_2$, und einem Hystereseschalter bestehend aus $A_3$ und $S_1$. Die Eingangsspannung muss zwischen den beiden Ausgangswerten des Schalters $S_1$ liegen. Hier sind das 1V und -1V. Für Spannungswerte in diesem Bereich funktioniert die Schaltung. Eine gute Beschreibung dieses Funktionsprinzips gibt Jim Thompsons Website[8][**?**].
Für eine Eingangsspannung von 0V ergibt sich folgendes:

[7] Siehe ADS1252 24-Bit, 40 kHz *Analog-Digital-Wandler* von Burr-Brown
[8] http://www.ee.washington.edu/conselec/CE/kuhn/onebit/primer.htm

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

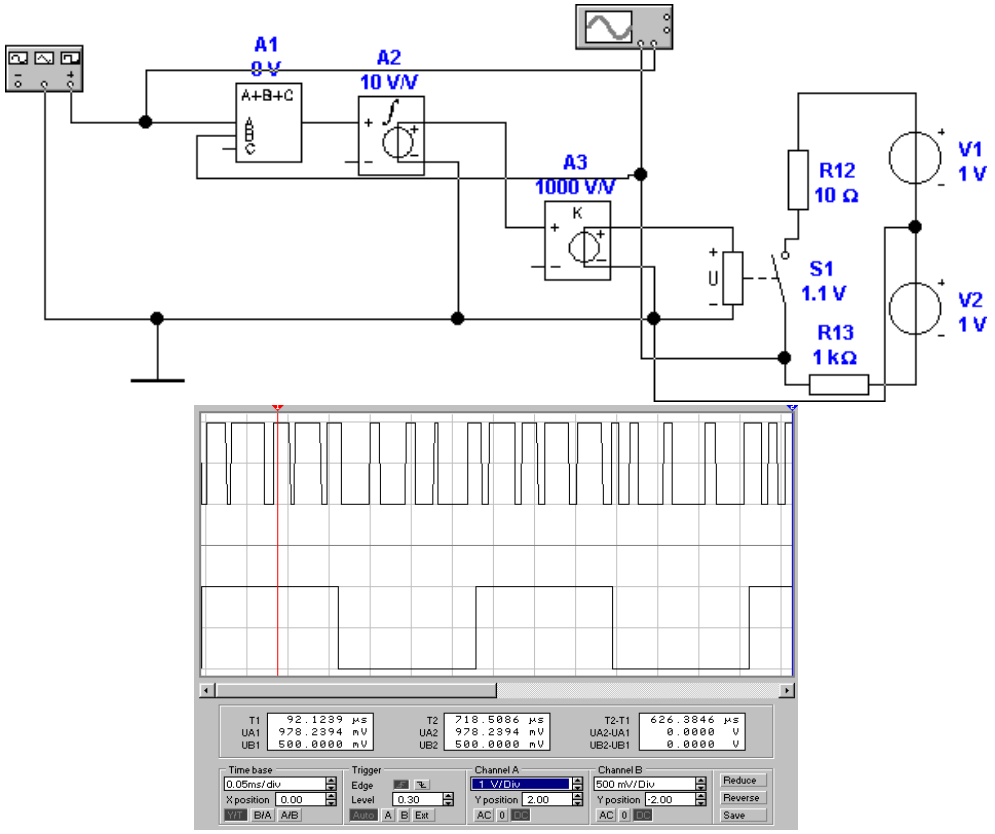



Abbildung 4.85.: Sigma-Delta-Wandler. Oben die Schaltung, unten die Signalformen für zwei Eingangsspannungen.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

- Der Ausgang des Integrators $A_2$ sei auf +1mV. Dann ist $S_1$ eingeschaltet und am Eingang B von $A_1$ liegt -1V. Die Ausgangsspannung von $A_1$ ist dann -1V, der Integrierer integriert mit einer Verstärkung von 10V/Vs gegen -1mV, der unteren Umschaltschwelle von $S_1$.

- Die Integrationsrate ist $|-1V|10V/Vs = 10V/s$. Daraus ergibt sich die Integrationszeit zu $t_{int} = 2mV/(10V/s) = 0.2ms$.

- Dann schaltet $S_1$ auf +1V. Der Integrator $A_2$ integriert nun von -1mv auf 1mV, wieder in 0.2 ms. Das Ausgangssignal ist also ein 2.5kHz Rechteck mit 50% Tastverhältnis.

Wir nehmen nun an, dass Die Eingangsspannung 0.5 V sein soll. Wir erhalten das folgende Resultat.

- Der Ausgang des Integrators $A_2$ sei auf +1mV. Dann ist $S_1$ eingeschaltet und am Eingang B von $A_1$ liegt -1V. Die Ausgangsspannung von $A_1$ ist dann $0.5V - (1V) = -0.5V$, der Integrierer integriert mit einer Verstärkung von 10V/Vs gegen -1mV, der unteren Umschaltschwelle von $S_1$.

- Die Integrationsrate ist $|-0.5V|10V/Vs = 5V/s$. Daraus ergibt sich die Integrationszeit zu $t_{int} = 2mV/(5V/s) = 0.4ms$.

- Dann schaltet $S_1$ auf +1V. Der Ausgang von $A_1$ ist nun auf $0.5V - (-1V) = 1.5V$. Der Integrator $A_2$ integriert nun von -1mv auf 1mV, nun mit einer Rate von $15V/s$. Die Integrationszeit ist 0.133 ms. Das Ausgangssignal ist also ein 1.875kHz Rechteck mit 75% Tastverhältnis. Das heisst, das Ausgangssignal entspricht $1V * 0.75 + (-1V) * 0.25 = 0.5V$

Schliesslich soll die Eingangsspannung -0.5V sein.

- Der Ausgang des Integrators $A_2$ sei auf +1mV. Dann ist $S_1$ eingeschaltet und am Eingang B von $A_1$ liegt -1V. Die Ausgangsspannung von $A_1$ ist dann $-0.5V - (1V) = -1.5V$, der Integrierer integriert mit einer Verstärkung von 10V/Vs gegen -1mV, der unteren Umschaltschwelle von $S_1$.

- Die Integrationsrate ist $|-1.5V|10V/Vs = 15V/s$. Daraus ergibt sich die Integrationszeit zu $t_{int} = 2mV/(15V/s) = 0.133ms$.

- Dann schaltet $S_1$ auf +1V. Der Ausgang von $A_1$ ist nun auf $-0.5V - (-1V) = 0.5V$. Der Integrator $A_2$ integriert nun von -1mv auf 1mV, nun mit einer Rate von $5V/s$. Die Integrationszeit ist 0.4 ms. Das Ausgangssignal ist also ein 1.875kHz Rechteck mit 25% Tastverhältnis. Das heisst, das Ausgangssignal entspricht $1V * 0.25 + (-1V) * 0.75 = -0.5V$

Die Schaltung nach Abb. 4.85 zeigt diese Eigenschaften. Der untere Teil der Abbildung zeigt auf der oberen Oszilloskopspur das Ausgangssignal von $S_1$ und unten das Eingangssignal. Die Schaltung erzeugt also ein Rechtecksignal, bei dem das Tastverhältnis

$$\frac{t_{on}}{t_{on} + t_{off}} = \frac{U_{ein} - U_{unten}}{U_{oben} - U_{unten}} \tag{4.1.95}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Die Wandlung geschieht nun, indem man mit einer höheren Taktfrequenz zählt, wie oft Einsen und Nullen im Rechtecksignal auftreten. Hier könnte man zum Beispiel mit einem MHz zählen. Man würde folgendes erhalten

| Spannung | Anzahl 1 | Anzahl 0 |
|---|---|---|
| 0V | 500 | 500 |
| -0.5V | 250 | 750 |
| 0.5V | 750 | 250 |

Die Schaltung hat also, ohne grossen Aufwand 10 Bit Präzision. Sie hat folgende Eigenschaften:

- es gibt kein Aliasing bei der Umwandlung in den Bitstrom. Da bei der Digitalisierung nicht abgetastet wird, gibt es kein Aliasing.
- Prinzipbedingt gibt es keine fehlenden Codes.
- Das Wandlerverhalten ist absolut monoton und linear.
- Der Wandler ist unempfindlich gegen steile Flanken, gegen Rauschen und gegen hochfrequente Störungen.
- Aliasing kann auftreten, wenn die digitale Abtastfrequenz ungeschickt (zu tief) gewählt ist.

Die Funktion des Integrators ist, eine Tiefpassfilterung zur Verfügung zu stellen. Wir haben die Schaltung mit einem Baustein aufgebaut, sie wird als Sigma-Delta-Wandler ($\Sigma\Delta$-Wandler) erster Ordnung bezeichnet. Mit einer höheren Ordnung kann das Tastverhältnis[9] schneller an das Eingangssignal angepasst werden[**?**].

### 4.1.9. Lock-In Verstärker am Beispiel des AD630 Chips

Zur Messung von periodischen Signalen verwendet man häufig Lock-In Verstärker. Der Kern jedes Lock-In-Verstärkers ist ein synchroner Gleichrichter. der synchrone Gleichrichter in Abb. 4.86 ist ein Multiplizierer. Es können aber auch einfache Umschalter für die Polarität verwendet werden.
Die Ausgangsspannung wird durch das folgende Integral berechnet:

$$U_{Lock-In} = \frac{1}{T} \int_{t-T}^{t} U(\tau) \sin(\omega_0 \tau) d\tau \tag{4.1.96}$$

Hier ist $sin(\omega_0 \tau)$ die Referenzspannung und $U_e(t)$ die Eingangsspannung. Die Integration wird normalerweise, wie auch in der Abbildung 4.86, mit Tiefpassfiltern durchgeführt. Die untere Hälfte von Abbildung 4.86 zeigt das Ausgangssignal des Lock-In-Verstärkers. Dabei wird eine Signalspannung von 2V bei 1 kHz und eine Störspannung von 2 V bei 5 kHz als Eingangssignal verwendet. Die Integrationszeit ist $\tau = 3k\Omega \times 1\mu F = 3ms$. Die Nutz-Eingangsspannung wird alle 15 ms von ihrer Spannung von 2V auf 0V geschaltet. 15 ms später wird sie wieder eingeschaltet.
Meistens wird in Lock-In-Verstärkern das Eingangssignal nicht mit dem Referenzsignal multipliziert, sondern nur die Verstärkung zwischen den Werten +1 und -1 umgeschaltet.

[9] http://kabuki.eecs.berkeley.edu/ kelvink/thesis/thesis.pdf

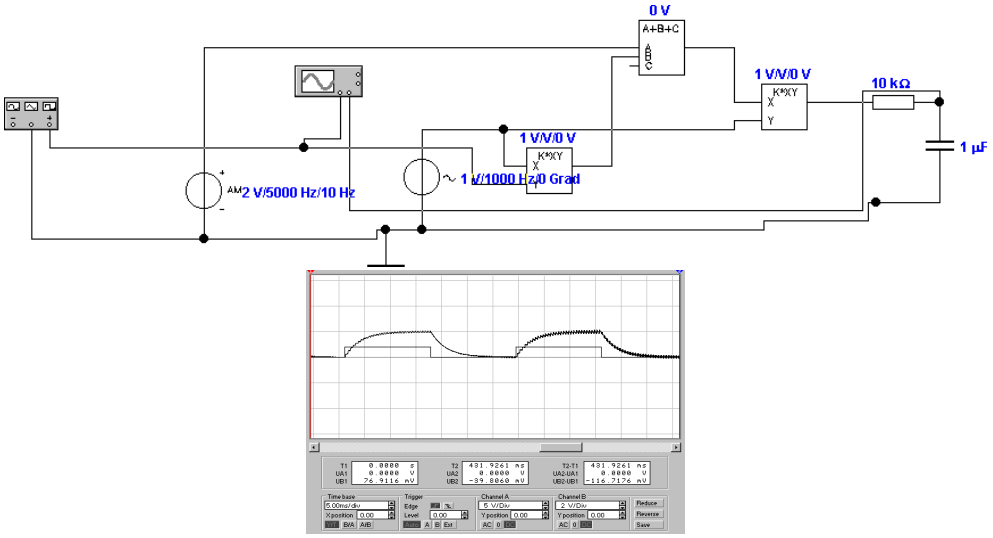



Abbildung 4.86.: Prinzipbild eines Lock-In-Verstärkers. Oben die Schaltung, unten die Signalformen für eine Signalspannung von 2V bei 1 kHz und eine Störspannung von 2 V bei 5 kHz. Die Integrationszeit ist $\tau = 3k\Omega \times 1\mu F = 3ms$

Die Funktion des Lock-In-Verstärkers mit periodischer Umschaltung kann man mathematisch wie folgt formulieren:

$$U_a = U(t)S(t)$$

$$S(t) = \begin{cases} 1 & \text{für} \quad U_{st} > 0 \\ -1 & \text{für} \quad U_{st} < 0 \end{cases}$$

Die Schaltspannung $S(t)$ wird in eine Fourierreihe entwickelt:

$$S(t) = \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{1}{2n+1} \sin(2n+1)\omega_{st}t \tag{4.1.97}$$

Wir nehmen im weiteren an, dass die Eingangsspannung $U_a(t)$ sinusförmig und ein ganzzahliges Vielfaches der Referenzfrequenz $\omega_e = m\omega_{st}$ ist.

$$U_a(t) = U_e \sin(m\omega_{st}t + \varphi_m) \frac{4}{\pi} \sum_{n=0}^{\infty} \frac{1}{2n+1} \sin(2n+1)\omega_{st}t \tag{4.1.98}$$

Für sinusförmige Schwingungen gelten die folgenden Beziehungen:

$$\frac{1}{T} \int_0^T \sin(m\omega_{st}\tau + \varphi_m) d\tau = 0 \tag{4.1.99}$$

$$\frac{1}{T} \int_0^T \sin(m\omega_{st}\tau + \varphi_m) \sin(\ell\omega_{st}\tau) d\tau = \begin{cases} 0 & \text{für} \quad m \neq \ell \\ \frac{1}{2}\cos\varphi_m & \text{für} \quad m = \ell \end{cases} \tag{4.1.100}$$

Damit wird die gemittelte Ausgangsspannung des Lock-In-Verstärkers

$$< U_a >= \begin{cases} \frac{2}{\pi m} U_e \cos\varphi_m & f\ddot{u}r \quad m = 2n+1 \\ 0 & f\ddot{u}r \quad m \neq 2n+1 \end{cases} \tag{4.1.101}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

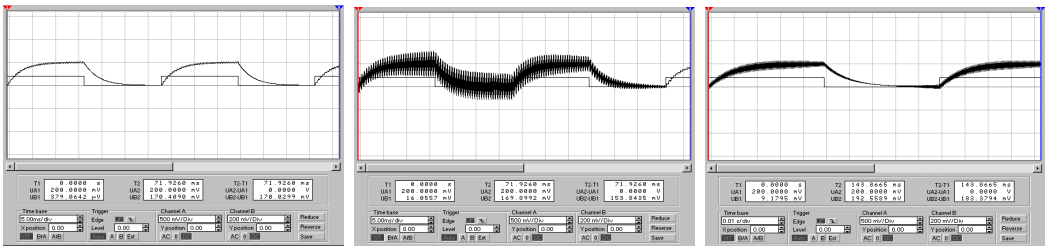


Abbildung 4.87.: Signalformen des Lock-In-Verstärkers nach Abb. 4.86. Links: Signalspannung von 0.2V bei 1 kHz und eine Störspannung von 2 mV bei 5 kHz. Die Integrationszeit ist $\tau = 3k\Omega\, 1\mu F = 3ms$. Mitte: Signalspannung von 0.2V bei 1 kHz und eine Störspannung von 2 V bei 5 kHz. Die Integrationszeit ist $\tau = 3k\Omega\, 1\mu F = 3ms$. Rechts: Signalspannung von 0.2V bei 1 kHz und eine Störspannung von 2 V bei 5 kHz. Die Integrationszeit ist $\tau = 10k\Omega\, 1\mu F = 10ms$.

Bei einem Analogmultiplizierer erhält man

$$< U_a >= \begin{cases} \frac{1}{2} U_e \cos\varphi & für \quad m = 1 \\ 0 & für \quad m \neq 1 \end{cases} \tag{4.1.102}$$

Anders als bei dem Detektor mit Umschalter ist die Ausgangsspannung nur dann ungleich Null, wenn die Frequenzen von Eingangsspannung und Referenzspannung gleich sind. Bei Umschaltern ist der Lock-In-Verstärker auch auf die Harmonischen des Eingangssignals empfindlich.
Die Abb. 4.87 zeigt den Einfluss von Störspannungen und der Filterzeitkonstanten. Die linke Seite zeigt das Ausgangssignal, wenn das Eingangssignal auf 0.2 V gesetzt wird und wenn das Störsignal quasi ausgeschaltet ist. Das 10 mal stärkere Störsignal ist in der Mitte wieder eingeschaltet. Durch Verlängerung der Integrationszeit in der rechten Seite kann das Signal-zu Störsignal-Verhältnis verbessert werden.
Die Bandbreite des Detektors hängt von der Integrationszeit $\tau$ ab. Es gilt

$$\Delta f = \frac{1}{\tau} \tag{4.1.103}$$

Man kann durch Integration über mehrere Sekunden in einem Lock-In-Verstärker leicht Frequenzen von einigen kHz mit Bandbreiten im mHz-Bereich messen.
Multipliziert man das Eingangssignal mit $\sin\omega_{st}t$ und auch $\cos\omega_{st}t$, so kann man sowohl die Amplitude wie auch die Phase über die Beziehungen im rechtwinkligen Dreieck bestimmen.
Abbildung 4.88 zeigt den Aufbau und die Verwendung des Verstärkers *AD630*. Die linke Seite zeigt den inneren Aufbau. Zwei Verstärker, Amp A und Amp B können über den vom Komparator Comp gesteuerten Schalter an den Ausgangsverstärker gelegt werden. Der Chip beinhaltet Widerstände, so dass der eine Eingangsverstärker als invertierender und der andere als nicht-invertierender Verstärker aufgebaut ist. Dann arbeitet der AD630 als Lock-In-Verstärker, mit der Referenzfrequenz am Eingang des Komparators.
Der mittlere Teil von Abbildung 4.88 zeigt die Schaltung des AD630 als Lock-In-Verstärker. Der linke AD542 arbeitet als Eingangs-Pufferverstärker, der rechte als

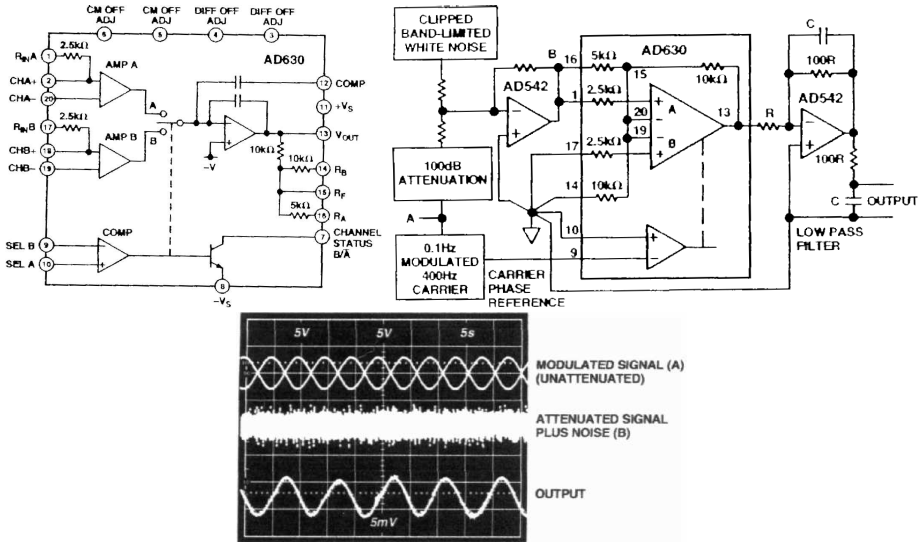



Abbildung 4.88.: AD630 Links: Schaltschema. Rechts: Einsatz der Schaltung als Lock-In-Verstärker. Unten: Signalformen

Ausgangsintegrator. Die Ganze Schaltung ist ein die Verstärkung umschaltender Synchrongleichrichter.
Die rechte Seite von Abbildung 4.88 zeigt, dass der Lock-In-Verstärker ein *Signal*, das etwa 100000 schwächer als das Störsignal (weisses Rauschen) ist, detektieren und anzeigen kann.
Viele Lock-In-Verstärker werden heute mit DSPs aufgebaut (siehe auch den Abschnitt 2.9).

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

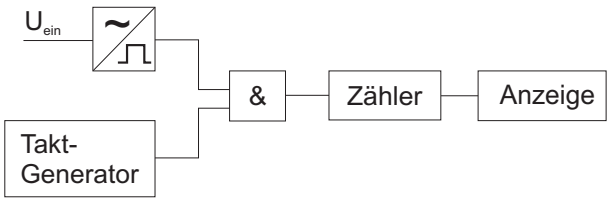



Abbildung 4.89.: Messung von Zeit oder Periodendauer.

| $\tau$ | $f_m$ | 10Hz | 100Hz | 1kHz | 10kHz | 100kHz | 1MHz | 10MHz |
|---|---|---|---|---|---|---|---|---|
| 10$\mu$s | 100kHz | - | - | - | - | 1 | 0.1 | 0.01 |
| 100$\mu$s | 10kHz | - | - | - | 1 | 0.1 | 0.01 | $10^{-3}$ |
| 1ms | 1kHz | - | - | 1 | 0.1 | 0.01 | $10^{-3}$ | $10^{-4}$ |
| 10ms | 100Hz | - | 1 | 0.1 | 0.01 | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ |
| 100ms | 10Hz | 1 | 0.1 | 0.01 | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ |
| 1s | 1Hz | 0.1 | 0.01 | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ | $10^{-7}$ |

Tabelle 4.4.: *Auflösung* $\varepsilon$ der Periodendauermessung. Horizontal ist die Taktfrequenz angegeben, vertikal die zu messende Periode $\tau$oder die dazugehörige Frequenz $f_m = \frac{1}{\tau}$.

## 4.2. Messung weiterer physikalischer Grössen

### 4.2.1. Frequenzmessung

Frequenz und Zeitmessungen werden auf die Bestimmung einer Periodendauer, beziehungsweise auf die Zählung von Pulsen eines stabilen Generators zurückgeführt. Zeit und damit auch Frequenz ist die am genauesten bestimmbare physikalische Grössen.
Abbildung 4.89 zeigt die Messung einer Zeit oder einer Periodendauer. Die Eingangsspannung wird in eine Rechteckspannung übergeführt (durch eine Triggerstufe). Diese Rechteckspannung steuert das Tor (AND-Gatter), das die Taktimpulse eines stabilen Oszillators auf einen Zähler und damit auf eine Anzeige schaltet.
Die Periodendauermessung ist umso genauer, je länger die Periodendauer ist. Achtung!
Wenn die Taktfrequenz $f_0$ ist und die Periodendauer $\tau$, dann gilt für die *Auflösung* $\varepsilon$

$$\varepsilon = \frac{1}{f_0 \tau} \tag{4.2.1}$$

Typischerweise ist $f_0$ 1 MHz oder 10 MHz.
Gleichung (4.2.1) kann für Frequenzen $f_m$ so umgeschrieben werden:

$$\varepsilon = \frac{f_m}{f_0} \tag{4.2.2}$$

Die *Auflösung* der Periodendauermessung ist in der Tabelle 4.4 zusammengefasst. Kehrt man das Messprinzip aus Abb. 4.89 um, so erhält man den Frequenzmesser nach Abb. 4.93. Hier wirkt der Taktgenarator als Schalter für den Impulsstrom, der aus der Eingangsspannung durch einen Rechteckformer abgeleitet wurde. Die restliche Schaltung mit der Anzeige bleibt gleich.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

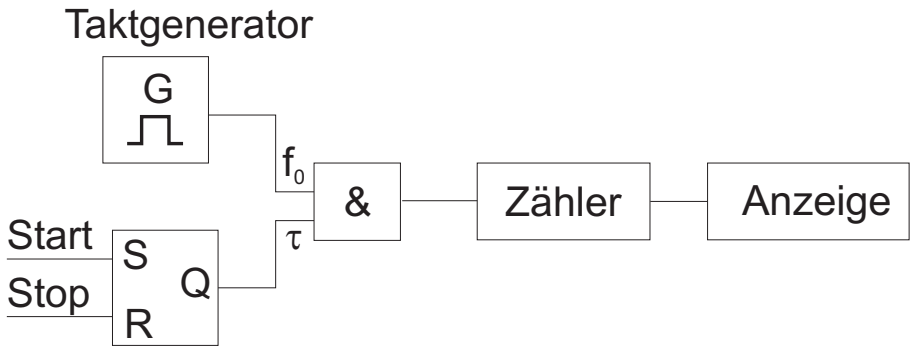



Abbildung 4.90.: Prinzip der Schaltung für eine Zeitmessung

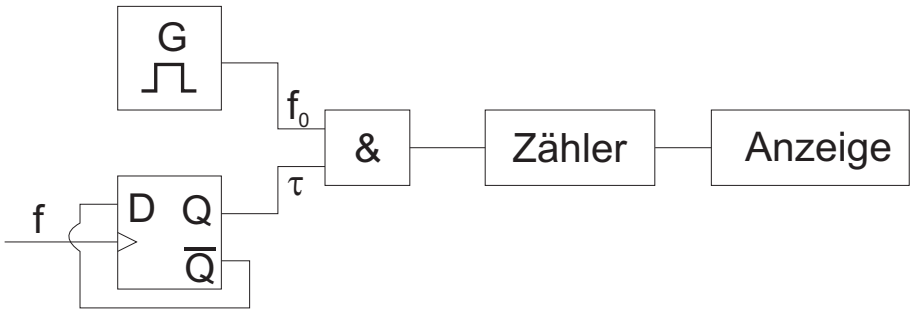



Abbildung 4.91.: Prinzip der Schaltung für eine Periodendauermessung

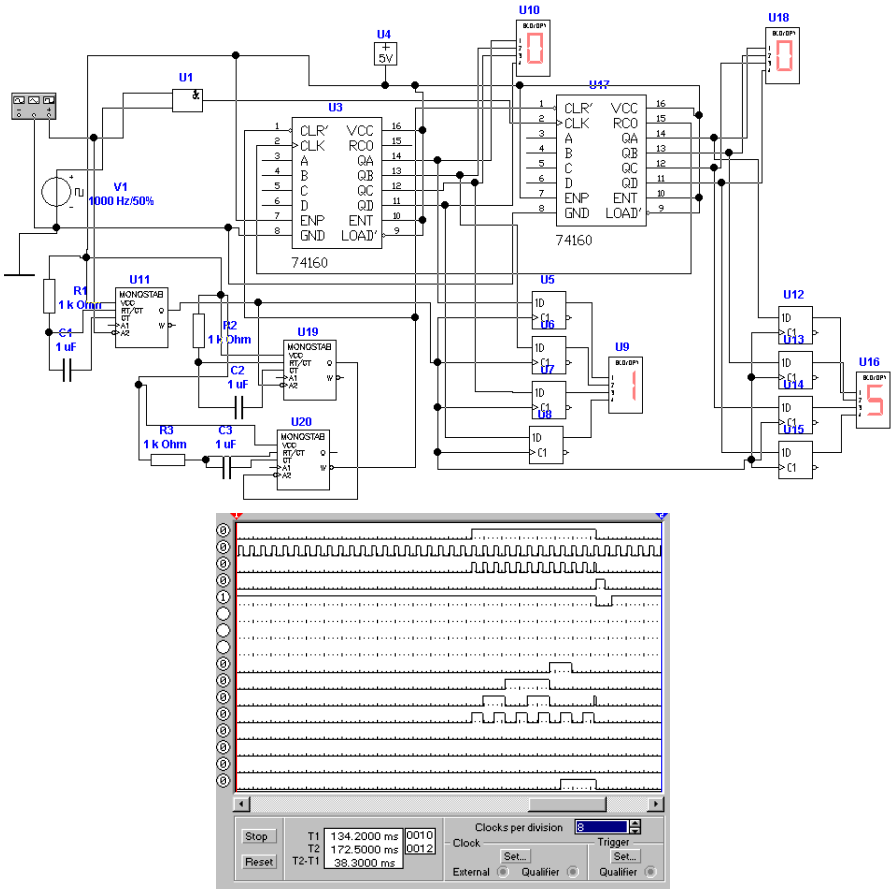



Abbildung 4.92.: Schaltung für eine Periodendauermessung

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

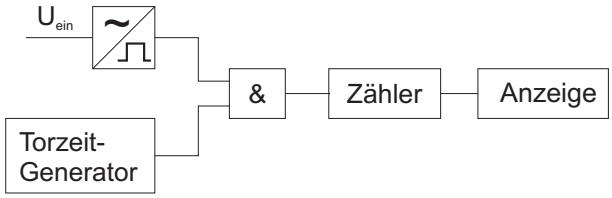



Abbildung 4.93.: Messung der Frequenz

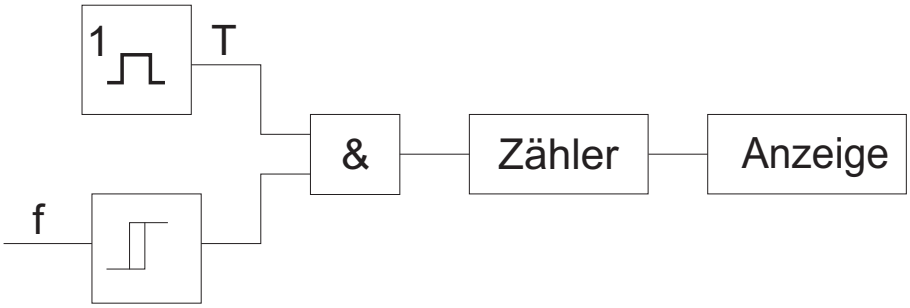



Abbildung 4.94.: Schaltung für eine Frequenzmessung

bei einer Torzeit $T$ ist die *Auflösung* $\varepsilon$ der Messung der Frequenz $f_m$ durch

$$\varepsilon = \frac{1}{f_m T} \tag{4.2.3}$$

gegeben. Für die *Auflösung* der Frequenzmessung gilt Tabelle 4.4, aber mit vertauschten Benennungen. Horizontal ist nun die zu messende Frequenz, Vertikal die Torzeit des *Messgerätes*. Eine Implementation dieser Schaltung findet man in Abb. 4.94 und Abb. 4.95.
Aus der Tabelle 4.4 kann man ableiten, dass bei vorgegebener Messdauer hohe und tiefe Frequenzen sehr genau gemessen werden können, mittlere Frequenzen jedoch nicht. Wenn $T$ die Messdauer und $f_0$ die Taktfrequenz des Messoszillators ist, gilt für die Genauigkeit

$$\varepsilon = \min\left(\frac{f_m}{f_0}, \frac{1}{f_m T}\right) \tag{4.2.4}$$

Aus Gleichung (4.2.4) wird Tabelle 4.5 berechnet.
Das Verhältniszählverfahren nach Abb. 4.96 umgeht dieses Problem. Das *Signal* des Taktgenerators $f_0$ wird nach einem Teiler durch $m$ auf den Daten- (D-) Eingang

| T | 1 Hz | 10 Hz | 100 Hz | 1 kHz | 10 kHz | 100 kHz | 1 MHz |
|---|---|---|---|---|---|---|---|
| 10 s | $10^{-6}$ | $10^{-5}$ | $10^{-4}$ | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ | $10^{-7}$ |
| 1 s | $10^{-6}$ | $10^{-5}$ | $10^{-4}$ | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ | $10^{-6}$ |
| 0.1 s | - | $10^{-5}$ | $10^{-4}$ | $10^{-3}$ | $10^{-3}$ | $10^{-4}$ | $10^{-5}$ |
| 0.01 s | - | - | $10^{-4}$ | $10^{-3}$ | $10^{-2}$ | $10^{-3}$ | $10^{-4}$ |
| 1 ms | - | - | - | $10^{-3}$ | $10^{-2}$ | $10^{-2}$ | $10^{-3}$ |

Tabelle 4.5.: Genauigkeit der Frequenzmessung für eine Messdauer $T$ bei einer Taktfrequenz von 1 MHz.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

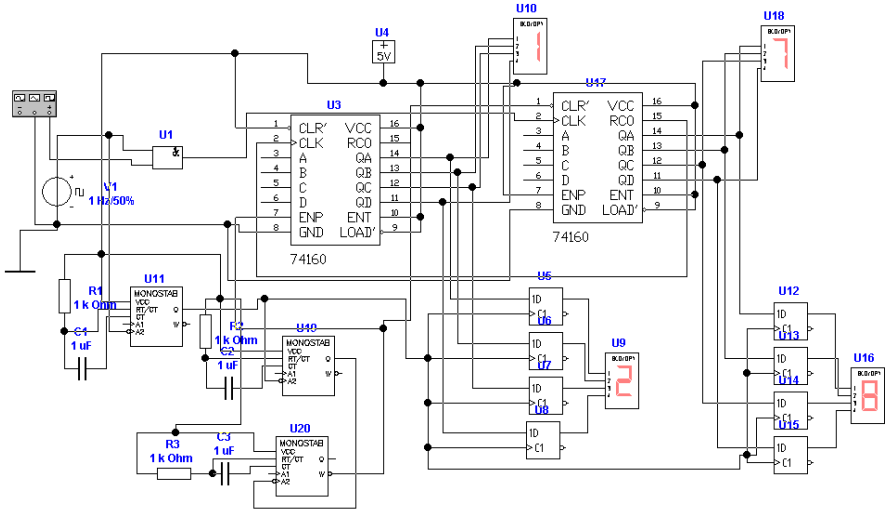



Abbildung 4.95.: Schaltung für eine Frequenzmessung. Der Unterschied zur Periodendauermessung liegt alleine in der Taktfrequenz des Referenzoszillators V1!

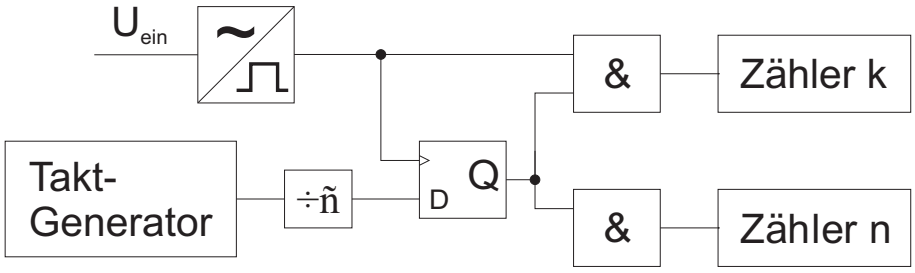



Abbildung 4.96.: Messung der Frequenz mit dem Verhältniszählverfahren.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

| $f$ | $k$ | $n$ | $f_{gemessen}$ |
|---|---|---|---|
| 5.6789123 Hz | 6 | 1056540 | $5.6789141 \pm 0.000006 Hz$ |
| 3456.1234 Hz | 3457 | 1000253 | $3456,1256 \pm 0.004 Hz$ |
| 876985.134 Hz | 876986 | 1000000 | $876986 \pm 0.9 Hz$ |

Tabelle 4.6.: Beispiele für die Verhältnismessung mit einer Torzeit von 1 s und einer Taktfrequenz von $f_0 = 1 MHz$

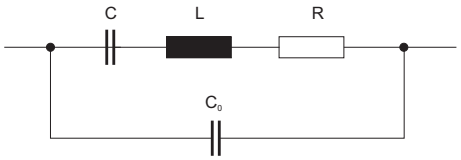



Abbildung 4.97.: Ersatzschaltbild eines Schwingquarzes

eines Flip-Flops gegeben. Die zu messende Frequenz $f$ wird über einen Rechteckformer an den Clockeingang des D-Flip-Flops gegeben. Wenn der Ausgang des Teilers von 0 auf 1 wechselt, wird dieser Zustand bei der nächsten steigenden Flanke des Eingangssignals auf den Ausgang übertragen. Damit werden über die zwei AND-Gatter die Zähler freigegeben. Der untere Zähler zählt nun die Pulse des Taktoszillators $n$. Der obere Zähler zählt die Pulse des Eingangssignals $k$. Wenn das Eingangssignal sehr schnell ist, zählt der obere Zähler im wesentlichen den Teilerfaktor $\tilde{n}$ des Teilers. Bei langsamen Eingangssignalen können noch zusätzliche Werte dazukommen. Die gesuchte Eingangsfrequenz $f$ ist

$$f = f_0 \frac{k}{n} \tag{4.2.5}$$

Tabelle 4.96 zeigt für drei exemplarische Freqenzen die Funktionsweise des Zählers. Der Zähler, $k$, ist immer die Eingangsfrequenz $f$ aufgerundet auf die nächste ganze Zahl $k = \sup f$. Der Nenner ist dann $n = \inf\left(f_0 \frac{k}{f}\right)$. Also ist letztlich

$$f_{gemessen} = f_0 \frac{k}{n} = f_0 \frac{\sup f}{\inf\left(f_0 \frac{\sup f}{f}\right)} \tag{4.2.6}$$

#### 4.2.1.1. Quarzbasierte Messung

Als frequenzbestimmendes Glied in Zählern verwendet man häufig Quarze. Das Ersatzschaltbild eines Quarzes (Abb. 4.97) besteht aus der kapazität $C$ und der Induktivität $L$. Die beiden Impedanzen bilden einen Serienschwingkreis, der die mechanische Resonanz nachbildet. $R$ ist der Dämpfungswiderstand und $C_0$ die Kapazität der Zuleitungen. Die obigen Werte können aus den Werten der mechanischen Resonanz abgeleitet werden[?]. Tietze-Schenk[?] geben als typische Werte für einen 4MHz-Quarz an, dass $L = 100 mH$, $R = 100\Omega$, $C = 0.015 pF$, $Q = 25000$ und $C_0 = 5 pF$. Bei Schwingkreisen ist die Dämpfung immer dann besonders klein, wenn L gross und C klein ist.
Jeder Quarz hat zwei Resonanzen, die Serienresonanz sowie die Parallelresonanz, in der auch die Anschlusskapazitäten eingehen. Die Impedanz des Schwingquarzes

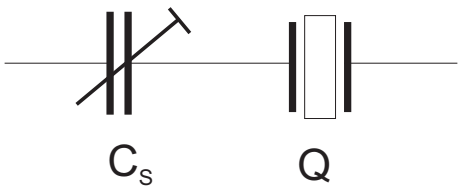



Abbildung 4.98.: Feineinstellung der Quarzfrequenz

ist

$$\underline{Z}_Q = \frac{j}{\omega} \frac{\omega^2 LC - 1}{C_0 + C - \omega^2 LCC_0} \tag{4.2.7}$$

Es gibt also zwei Extremalwerte: einmal wird $\underline{Z}_Q = 0$ und einmal $\underline{Z}_Q = \infty$. Die erste Resonanz heisst Serienresonanz. Für sie gilt, sofern man $R$ vernachlässigt,

$$\omega_s = \frac{1}{\sqrt{LC}} \tag{4.2.8}$$

Im Gegensatz dazu ist die Parallelresonanz (Nullstelle im Nenner)

$$\omega_p = \frac{1}{\sqrt{LC}} \sqrt{1 + \frac{C}{C_0}} \tag{4.2.9}$$

Aus Gleichung (4.2.9) ist ersichtlich, dass die Parallelresonanz höher liegt. Beim oben erwähnten 4 MHz-Quarz ist dies 0,15 %. Die Serienresonanz hängt, anders als die Parallelresonanz, nicht von den Anschlusskapazitäten ab. Sie ist also in jedem Falle stabiler.
Abb. 4.98 zeigt, dass man mit einem Kondensator in Serie Die Resonanzfrequenz einstellen kann. Die modifizierte Impedanz des Quarzes ist

$$\underline{Z'}_Q = \frac{1}{j\omega C_S} \frac{C + C_0 + C_S - \omega^2 LC\,(C_0 + C_S)}{C_0 + C - \omega^2 LCC_0} \tag{4.2.10}$$

Mit der Einstellbarkeit der Serienresonanzfrequenz hängt sie nun auch von $C_0$, der Streukapazität, ab. Allerdings ist schnell ersichtlich, dass für $C_S \to \infty$ Gleichung (4.2.10) in Gleichung (4.2.9) übergeht. Die Parallelresonanzfrequenz wird übrigens bei der Abstimmung nicht verändert. Für die einstellbare Serienresonanzfrequenz bekommt man

$$\omega'_s = \frac{1}{\sqrt{LC}} \sqrt{1 + \frac{C}{C_0 + C_S}} \tag{4.2.11}$$

Da bei allen Quarzen $C \ll C_0 + C_S$ ist bekommt man

$$\frac{\Delta\omega_s}{\omega_s} = \frac{C}{2\,(C_0 + C_S)} \tag{4.2.12}$$

Wenn die Kapazität $C_S \to 0$ geht, ist die serienresonanzfrequenz gleich der Parallelresonanzfrequenz, mit all den Stabilitätsproblemen. Es ist also sinnvoll, dass man $C_S$ möglichst gross lässt.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

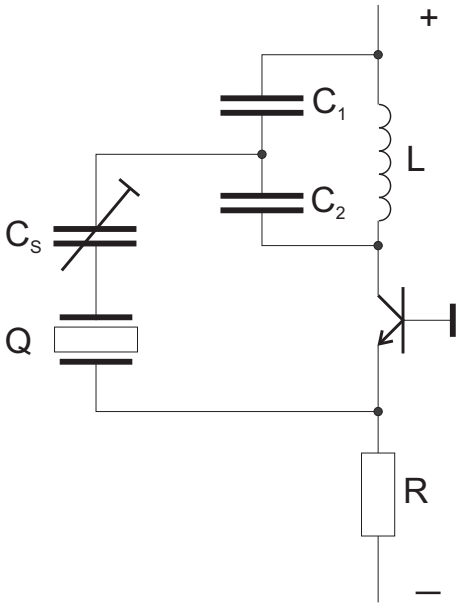



Abbildung 4.99.: Quarzoszillator nach Colpitts

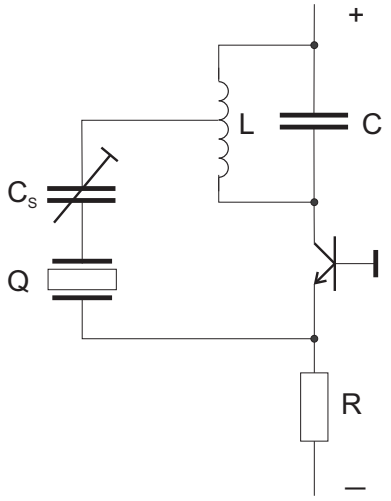



Abbildung 4.100.: Quarzoszillator nach Hartley

**4.2.1.1.1. Quarzoszillatoren** Als Taktgeneratoren für die Frequenz- bzw. Periodendauermessgeräte kommen Quarz-Oszillatoren in Frage. Abb. 4.99 zeigt einen solchen Oszillator nach Colpitts. Wichtig bei diesem Oszillator, wie auch beim Hartley-Oszillator in der Abb. 4.100 ist, dass die Impedanzen klein gegen den Dämpfungswiderstand $R$ des Quarzes sind.
Der Colpitts- und der Hartley-Oszillator unterscheiden sich durch die Art der Spannungsrückkopplung, Bei Colpitts wird die Kapazität geteilt, während bei Hartley Die Spule angezapft wird. Dieses ist sehr viel teurer als zwei Kondensatoren, ausser bei sehr hohen Frequenzen„ bei denen man die Leitungsinduktivitäten auf der Printplatte ausnutzt.
Abb. 4.101 zeigt die heute übliche Bauart von Quarzoszillatoren mit Invertern als Verstärkern. Das Ausgangssignal des Inverters wird über ein RC-Glied und den Quarz an den Eingang zurückgekoppelt. Der einstellbare Kondensator dient, wie schon bei den vorherigen Aufbauten dazu, die Frequenz im Promille-Bereich

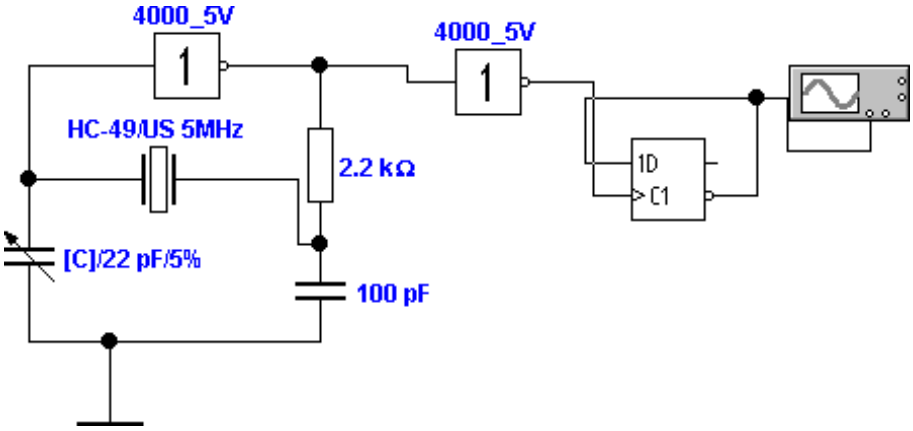



Abbildung 4.101.: Quarzoszillator mit Inverter-Gatter

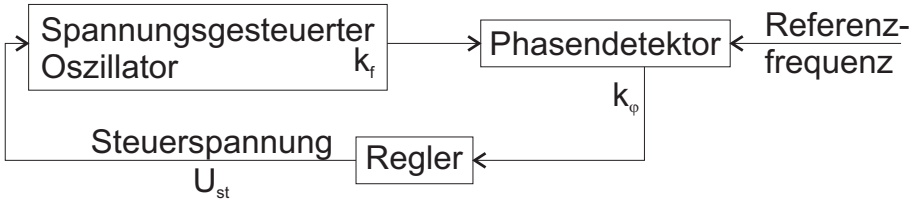



Abbildung 4.102.: Prinzipschaltbild eines Phasenregelkreises

abzustimmen.
Wenn Schaltungen aufgebaut werden müssen, kann man spezielle Treiberschaltungen für Quarze verwenden. Teilweise sind die Treiberschaltungen auch in den Bausteinen für Zähler usw. eingebaut.

#### 4.2.1.2. PLL

Wenn Frequenzen schnell gemessen werden sollen, oder wenn ein Oszillator mit einem bestimmten Teilerverhältnis an einen Referenzoszillator gekoppelt werden soll, verwendet man Phasenregelkreise, englisch Phase Locked Loop oder PLL. Abb. 4.102 zeigt das Prinzipbild, wenn die Ausgangsfrequenz des spannungsgesteuerten Oszillators gleich der Referenzfrequenz sein soll. Die Frequenz $f_s$ des spannungsgesteuerten Oszillators ist

$$f_s = f_0 + k_f U_s t \tag{4.2.13}$$

Am Phasendetektor entsteht eine Ausgangsspannung, die, zumindestens in der Nähe des Nullpunktes (bei dem beide Frequenzen gleich wären) linear ist.

$$U_\varphi = K_\varphi \varphi \tag{4.2.14}$$

Wenn die Referenzfrequenz $f_{ref}$ und die Frequenz $f_s$ unterschiedlich sind, nimmt die Phasenverschiebung $\varphi$ mit der Zeit zu: die Strecke hat ein integrierendes Verhalten. Im eingeschwungenen Zustand sind die Frequenzen exakt gleich, die verbleibende Phasenverschiebung ist:

$$\alpha - \varphi = \frac{f_{ref} - f_s}{A_R k_r k_\varphi} \tag{4.2.15}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

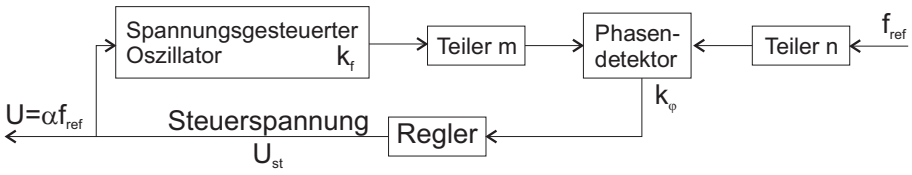



Abbildung 4.103.: Prinzipschaltbild eines Phasenregelkreises für beliebige Frequenzverhältnisse

wobei $A_R$ die Schleifenverstärkung ist. $\alpha$ ist eine gewollt eingeführte, konstante Phasenverschiebung. Der Frequenzgang dieser Regelschleife kann wie folgt berechnet werden:

$$\varphi = \int_0^t \omega_s d\tau - \int_0^t \omega_{ref} d\tau = \int_0^t \Delta\omega d\tau \tag{4.2.16}$$

Nun wird die Frequenz $f_s$ sinusförmig mit $\omega_m$ moduliert.
Mit $\Delta\omega(t) = \widehat{\Delta\omega} \cos \omega_m t$ bekommt man für die Phase

$$\varphi(t) = \frac{\widehat{\Delta\omega}}{\omega_m} \sin \omega_m t \tag{4.2.17}$$

In komplexer Schreibweise wird Gleichung (4.2.17)

$$\frac{\underline{\varphi}}{\underline{\Delta\omega}} = \frac{1}{j\omega_m} \tag{4.2.18}$$

Dies ist der Frequenzgang eines Integrators. Schliesslich ist die komplexe Schleifenverstärkung

$$\underline{A_s} = \frac{k_f k_\varphi}{j f_m} \tag{4.2.19}$$

Gleichung (4.2.19) suggeriert, dass das Integratorverhalten der Strecke die Regelung sehr einfach macht. Nun sind aber alle realen Phasendetektoren mit mehr oder weniger langen Verzögerungszeiten behaftet. Wie im Abschnitt 2.4.3 gezeigt, folgt daraus, dass im Regelkreis die Phasenverschiebung (nicht die zwischen den Frequenzen $f_{ref}$ und $f_s$!) proportional zur Frequenz ist. Damit wird es sehr schwierig, den Regelkreis zu stabilisieren. Die Dimensionierung von Phasenregelkreisen gehört somit nicht zu den einfachsten Aufgaben.
In Abb. 4.103 wird gezeigt, wie mit zwei zusätzlichen Teilern beliebige Frequenzverhältnisse gelockt werden können. Der Phasenregelkreis erzwingt, dass $\frac{f_s}{m} = \frac{f_{ref}}{n}$ ist. Damit erhält man für die Frequenz des spannungsgesteuerten Oszillators

$$f_s = \frac{m}{n} f_{ref} \tag{4.2.20}$$

Bei PLL-UKW-Empfängern könnte zum Beispiel $f_{ref} = 4MHz$ sein Das Frequenzraster ist 50 kHz, also muss $n = 4000/50 = 80$ sein. Wenn m einstellbar ist mit $m = 1940 \ldots 2160$ kann der gesamte UKW-Bereich eingestellt werden.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

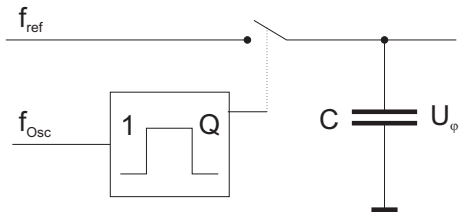



Abbildung 4.104.: Sample/Hold als Phasendetektor

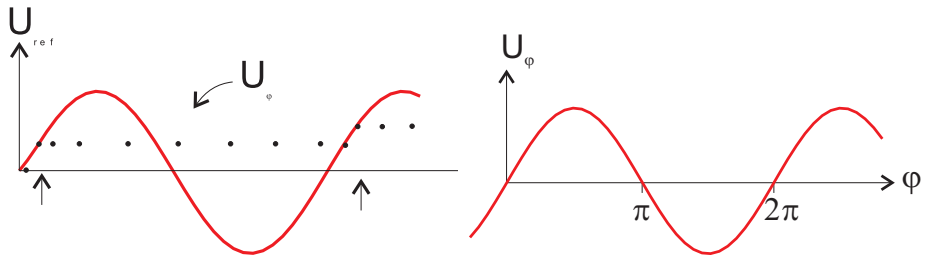



Abbildung 4.105.: Funktionsweise (links) und Kennlinie (rechts) des Sample/Hold als Phasendetektor

**4.2.1.2.1. Phasendetektoren** Abb. 4.104 zeigt, dass man ein *Abtast-Halteglied* oder Sample/Hold-Glied eine phasenempfindliche Detektion durchführen kann. Der spannungsgesteuerte Oszillator $f_{osc}$ triggert den Monoflop, der seinerseits einen Schalter betätigt, der die momentane Spannung der Referenz (Frequenz $f_{ref}$) auf einem Kondensator speichert. Abb. 4.105 zeigt auf der linken Seite die entsprechenden Spannungsverläufe.

Wenn die beiden Frequenzen gleich sind, und nur die Phase nicht übereinstimmt, dann misst so das Sample/Hold-Glied die zu dieser Phasenlage gehörige Spannung. Voraussetzung ist, dass das Eingangssignal sinusförmig oder dreiecksförmig ist. Die resultierende Kennlinie ist in Abb. 4.105 auf der rechten Seitze gezeigt. Dadurch, dass positive und negative Ausgangswerte für beide Polaritäten der Phase auftreten, hat dieser Detektor einen sehr eingeschränkten Fangbereich.

Besser in dieser Beziehung ist der Phasendetektor nach Abb. 4.106. Die beiden Eingangsfrequenzen $f_1$ und $f_2$ werden auf den Setz- (S-)Eingang jeweils eines RS-Flip-Flops gegeben. **Voraussetzung, dass diese Schaltung funktioniert ist, dass die Wellenform der beiden Eingangsfrequenzen jeweils kurze Pulse sind.** Wenn zum Beispiel $f_1$ in der Phase früher ist als $f_2$ dann wird das obere SR-Flip-Flop gesetzt. Sein Ausgangssignal geht nach 1. Da das untere RS-Flip-

Hinweis

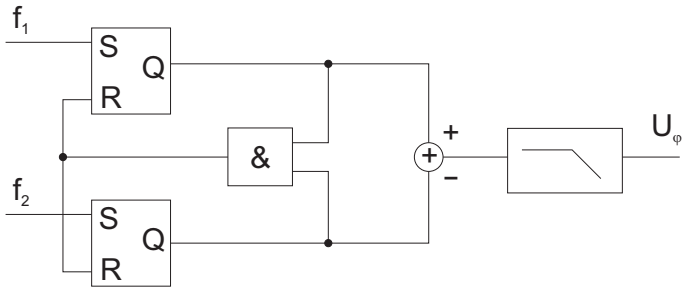



Abbildung 4.106.: Vorzeichenrichtiger Phasendetektor

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

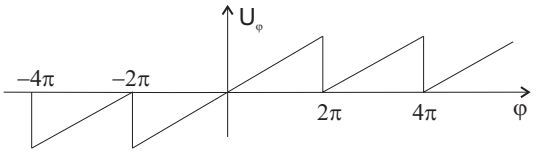



Abbildung 4.107.: Kennlinie des vorzeichenrichtigen Phasendetektor

Flop noch nicht gesetzt ist, ist das Ausgangssignal des AND-Gatters 0. Wenn nun der Puls der unteren Frequenz $f_2$ auch das untere Flip-Flop am Ausgang auf 1 setzt, dann setzt das AND-Gatter die beiden Flip-Flops zurück. Wenn die untere Frequenz in der Phase vorgeht, dann gibt entsprechend das untere Flipflop Pulse ab. Der Summierer gewichtet das obere Flip-Flop mit + und das untere mit -. Schliesslich werden die Pulse mit einem Tiefpassfilter geglättet.
Abb. 4.107 zeigt die Kennlinie dieses Detektors. Solange die Phase von $f_1$ voreilt, ist das Ausgangssignal positiv, solange sie nacheilt, negativ. Damit verhält sich diese Schaltung wesentlich gutmütiger als das Sample/Hold-Glied.
Abb. 4.108 zeigt schliesslich, wie so ein Phasendetektor aufgebaut wurde. Eine spannungsgesteuerte Rechteckquelle $V_1$ ist der Nachlaufoszillator, der Referenzoszillator wird durch den Funktionsgenerator realisiert. Die beiden Signale werden durch jeweils einen Monoflop, dessen internes Schaltbild man im Einschub oben rechts sehen kann, zu kurzen Pulsen geformt. Die RS-Flip-Flops $U_7$ und $U_8$ detektieren die Phase, zusammen mit dem AND-Gatter $U_9$. Anders als in der Literatur angegebn[?] muss man im Simulationsprogramm[?]eine Verzögerungsstreckt bestehend aus den beiden Invertern/Schmitt-Triggern $U_10$ und $U_12$ anwenden. Die Ausgangssignale von $U_7$ und $U_8$ werden in $A_1$ subtrahiert und in $A_2$ tiefpassgefiltert. Dieses *Signal* steuert den spannungsgesteuerten Rechteckoszillator.
Unten in Abb. 4.108 sind die beiden Kurvenformen der Oszillatoren zu sehen. es wird deutlich, dass die Phasenregelschlaufe das obere *Signal* an das untere Referenzsignal heranführt.

#### 4.2.1.3. Weitere Möglichkeiten der Zeitmessung

Wenn Zeiten im Bereich von $\mu$s bis 100 ps elektrisch gemessen werden sollen, dann wird häufig das Verfahren nach Abb. 4.109 verwendet. Dabei wird, nach einem Startpuls eine Rampe hochgefahren. Der Stoppuls triggert einen Sample/Hold, der zu diesem zeit die Amplitude misst. Aus der Anstigsrate $\alpha = \frac{dU}{dt}$ und der gemessenen Spannung $U_t$ errechnet man die Zeit

$$t = \frac{U_t}{\alpha} = \frac{U_t}{\frac{dU}{dt}} \tag{4.2.21}$$

Die Zeitmessung nach Abb. 4.109 funktioniert deshalb so gut, da es möglich ist, Sample/Hold-Verstärker zu bauen, deren Einschaltzeitpunkt auf wenige ps genau ist, auch wenn der ganze Schaltvorgang mehrere hundert ps dauern sollt. Diese lange Zeitdauer gibt einen systematischen, also korrigierbaren Fehler.

### 4.2.2. Magnetfelder

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

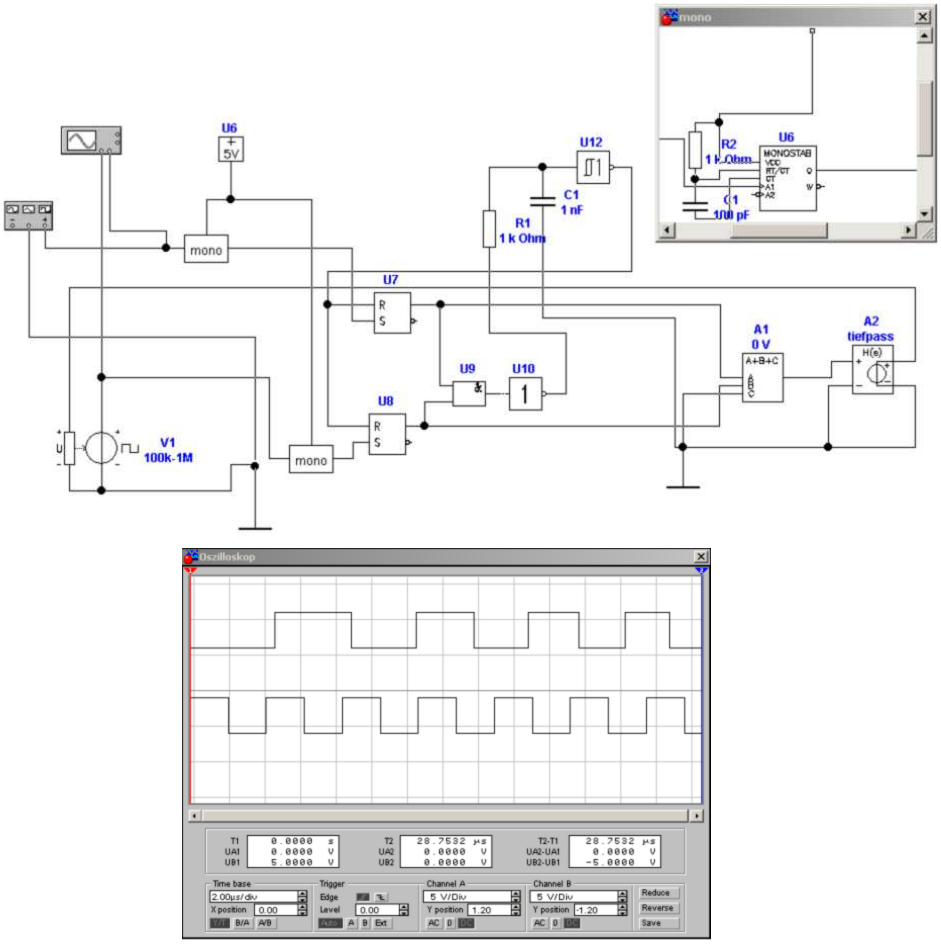



Abbildung 4.108.: Implementierung eines vorzeichenrichtigen Phasendetektor. Unten ist gezeigt, wie der gesteuerte Oszillator sich der Frequenz nähert.

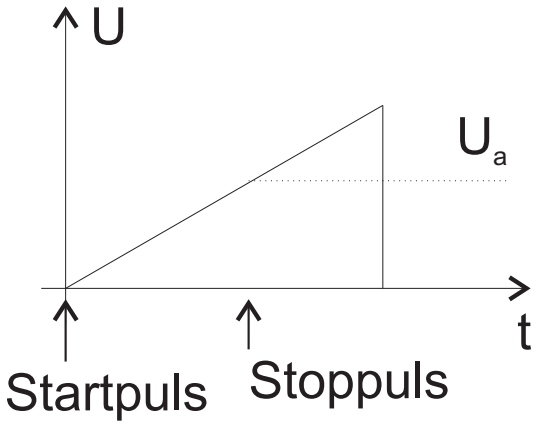



Abbildung 4.109.: Messung kurzer Zeiten mit definierten Rampen

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

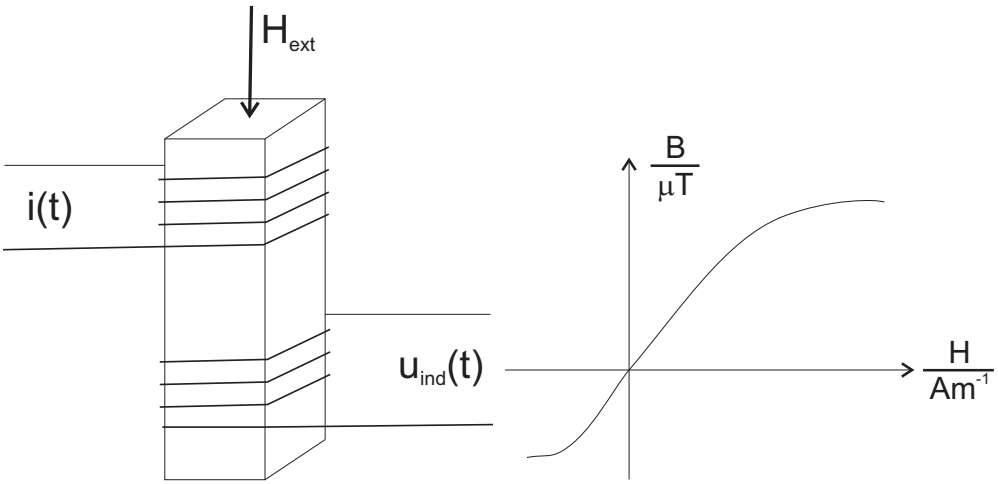



Abbildung 4.110.: Messung von Magnetfeldern mit Transformator. Rechts ideale Hysteresekurve

#### 4.2.2.1. Kernsondenmagnetometer

Die Messung von magnetischen Feldern kann über induzierte Spannungen

$$U_{ind} = -NA\frac{dB}{dt} \tag{4.2.22}$$

erfolgen. Hier ist $N$ die Windungszahl, $A$ die Querschnittsfläche und $B$ die magnetische Induktion. Eine einfache und für viele Zwecke ausreichende Möglichkeit sind rotierende Spulen. Die magnetische Induktion in Gleichung (4.2.22) wird dabei durch die Änderung der effektiven Fläche $A$ erreicht.
Bei nichtlinearen magnetischen Materialien, wie in Abb. 4.110, rechts, gezeigt kann mit einem Transformator das externe Feld gemessen werden. Abb. 4.110, links, zeigt einen nichtlinearen Transformator. Die Nichtlinearität sei durch die Gleichung $B(H) = \mu_0 \left[H + K \cdot H^3\right]$ modelliert. Moduliert man H mit der oberen Spule in Abb. 4.110 mit der Frequenz $\omega$ und existiert ein externes Feld $H_{ext}$ dann ist das effektive Feld $H(t) = H_{ext} + H_0 \sin(\omega t)$. Die magnetische Induktion wird dann

$$\begin{aligned} B(t) &\approx \mu_0 \left[H(t) + KH^3(t)\right] \\ &= \mu_0 \left[H_{ext} + H_0 \sin(\omega t) + K\left(H_{ext} + H_0 \sin(\omega t)\right)^3\right] \end{aligned} \tag{4.2.23}$$

Ausmultipliziert und nach Anwendung der Rechenregeln für trigonometrische Funktionen ergibt sich

$$\begin{aligned} U_{ind}(t) \approx \mu_0 NAH_0\omega \Bigg\{ &-\left[1 + 3K\left({H_{ext}}^2 + \frac{{H_0}^2}{2}\right)\right]\cos(\omega t) \\ &+3H_0K\left[-H_{ext}\sin(2\omega t) + \frac{H_0}{4}\cos(3\omega t)\right]\Bigg\} \end{aligned} \tag{4.2.24}$$

Gleichung (4.2.24) zeigt, dass die Frequenzkomponente bei $2\omega$ nur auftritt, wenn $H_{ext} \neq 0$ ist. Die Grösse des externen magnetischen Feldes $H_{ext}$ kann dann mit

$$H_{ext} \approx -\frac{1}{3\mu_0 NAH_0^2\omega K} U_{ind}\,|_{2\omega} \tag{4.2.25}$$

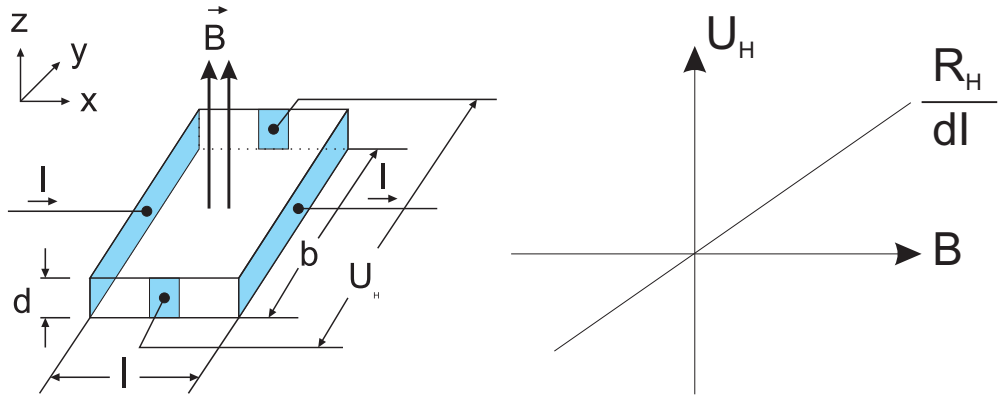



Abbildung 4.111.: Messung von Magnetfeldern mit dem Halleffekt. Rechts die Kennlinie des Halleffekts

angegeben werden. Da die Amplitude des modulierten Feldes quadratisch im Nenner erscheint, kann (in Grenzen) die Empfindlichkeit durch Erhöhung der Modulationsamplitude gesteigert werden.

#### 4.2.2.2. Hall-Effekt

Wenn in einem Magnetfeld senkrecht zur Magnetfeldrichtung ein Strom fliesst, dann bewirkt die Lorentzkraft, dass die zur Magnetfeldrichtung und zur Stromrichtung senkrechten Seiten des Leiters geladen werden. Dieser Effekt heisst Hall-Effekt. Zwischen den Elektroden in y-Richtung in Abb. 4.111 tritt dann die Hallspannung

$$U_H = \frac{R_H}{d} IB \tag{4.2.26}$$

auf. Dabei ist angenommen worden, dass die Länge sehr viel grösser als Breite sei. Damit kann man eine Betrachtung des stationären Zustandes durchführen.
Die Lorentzkraft auf ein bewegtes Ladungsteilchen ist $\boldsymbol{F}_L = q(\boldsymbol{v} \times \boldsymbol{B}$. Im Gleichgewicht wird sie durch die elektrostatische Kraft des Hall-Feldes

$$\boldsymbol{F}_H = q\boldsymbol{E}_H \tag{4.2.27}$$

kompensiert. Aus der Betrachtung des Kräftegleichgewichts folgt

$$\boldsymbol{E}_H = -(\boldsymbol{v} \times \boldsymbol{B}) \tag{4.2.28}$$

Gleichung (4.2.28 ist allgemeingültig. Für die folgende Betrachtung nehmen wir an, dass Magnetfeld und Stromflussrichtung orthogonal seien. Dann ist die induzierte Hallspannung $U_H = -bv_xB$. Aus der mittleren Geschwindigkeit $v_x$ der Ladungsträger kann, bei bekannter *Ladungsträgerdichte* $n$, die Stromdichte $j_{x,n} = nqv_x = \frac{I}{bd}$ berechnet werden. Diese Stromdichte ist die Folge des Stromes $I$, der über die Stirnfläche $b \cdot d$ eingekoppelt wird. Man hat also $j_x = \frac{I}{b \cdot d}$ und damit für negative Ladungsträger (Betrag der Ladung: $q$)

$$U_H = -\frac{1}{nq}\frac{1}{d} IB \tag{4.2.29}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

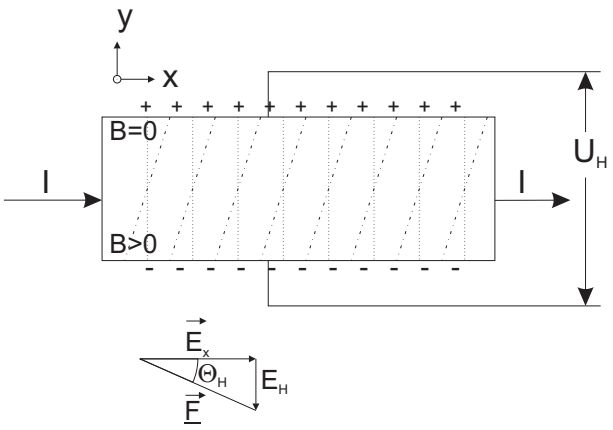



Abbildung 4.112.: Äquipotentialverlauf beim Halleffekt

Die Abhängigkeit von der *Ladungsträgerdichte* (n für negative Ladungen und p für positive Ladungen) und von der Ladung wir in einer Hall-Konstante $R_H$ zusammengefasst. Für die beiden Ladungsträgerpopulationen ergeben sich

$$R_{H,n} = -\frac{1}{nq} \tag{4.2.30}$$

$$R_{H,p} = \frac{1}{pq} \tag{4.2.31}$$

Die funktionale Abhängigkeit ist in Abb. 4.111 gezeigt.
In einem Widerstand mit konstantem Querschnitt und homogener Materialzusammensetzung sind die Äquipotentialflächen der Spannung senkrecht zur Stromrichtung. Liegt eine Hallspannung vor, addiert sich deren elektrisches Feld zum ursprünglichen elektrischen Feld. Die Äquipotentialflächen werden, wie in Abb. 4.112 gezeigt, um den Hall-Winkel $\Theta_h$ gekippt.

$$\tan(\Theta_H) = \frac{|\boldsymbol{E}_H|}{|\boldsymbol{E}_x|} \tag{4.2.32}$$

Unter Verwendung der Hallbeweglichkeit $\mu_H$ der Ladungsträger, die sich nicht sehr von der Driftbeweglichkeit $\mu_{drift}$ unterscheidet, ist der Hall-Winkel

$$\tan(\Theta_{H,n}) = -\mu_{H,n}B \tag{4.2.33}$$

$$\tan(\Theta_{H,p}) = \mu_{H,p}B \tag{4.2.34}$$

#### 4.2.2.3. Feldplatten, Gauss-Effekt

Feldplatten, wie sie in Abb. 4.113 gezeigt werden, verwenden den gleichen physikalischen Mechanismus wie die Hall-Sonden, aber in longitudinaler Weise. In einem Leiter, wie er in der Abbildung rechts gezeigt ist, ist der Widerstand ohne Magnetfeld

$$R_0 = \rho\frac{l_0}{b_0 d}$$

wenn $l_0$ wie üblich die Länge des Leiterstückes ist, $b_0$ die Breite und $d$ die Dicke.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

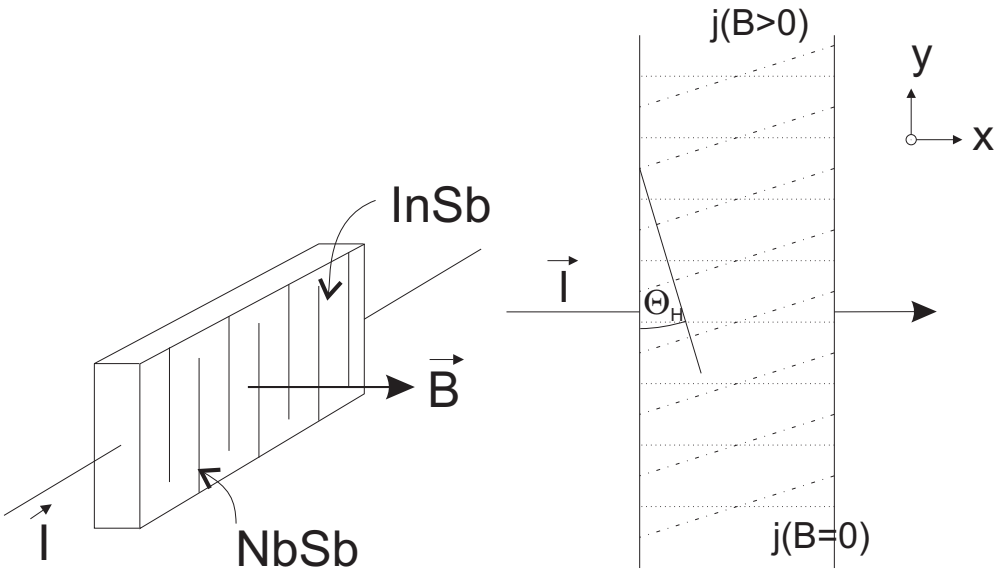



Abbildung 4.113.: Feldplatte. Rechts der Äquipotentialverlauf

Wenn andere Effekte des Magnetfeldes vernachlässigt werden, tritt immer noch die Verkippung der Äquipotentialflächen des elektrischen Feldes auf. Wenn die Probe breiter als lang ist, kann man davon ausgehen, dass das Magnetfeld die Wege um

$$l(\Theta_H) = \frac{l_0}{\cos(\Theta_H)} \tag{4.2.35}$$

$$b(\Theta_H) = b_0 \cos(\Theta_H) \tag{4.2.36}$$

verlängert. Wie beim Halleffekt ausgeführt, ist die Verlängerung eine Funktion des Hall-Winkels $\Theta_H$. Durch den längeren Weg erhöht sich der Widerstand um

$$R(\Theta_H) = R_0 \frac{1}{\cos^2(\Theta_H)} = R_0 \left(1 + \tan^2(\Theta_H)\right) \tag{4.2.37}$$

Wenn man die beim Halleffekt definierte Hall-Beweglichkeit verwendet, wird der Widerstand

$$R(B) = R_0 \left(1 + K \left(\mu_H B\right)^2\right) \tag{4.2.38}$$

Um die Forderungen nach einem kurzen Leiterstück und nach langen Wirkungswegen zu erfüllen, werden, wie in Abb. 4.113 gezeigt, mäanderierende Strukturen verwendet. Bei ihnen kann man für jedes Teilstück davon ausgehen, dass die in der gleichen Abbildung rechts gezeigte Situation vorliegt. Abb. 4.114 schliesslich zeigt die resultierende Kennlinie einer Feldplatte. Um eine hohe Empfindlichkeit zu erreichen, müssen Werkstoffe mit hoher Ladungsträgerbeweglichkeit verwendet werden. Deshalb werden, wie bei Hall-Generatoren InSb, InAs, *Si* und GaAs verwendet.

#### 4.2.2.4. Magnetoresistiver Effekt

Unter dem Einfluss eines externen magnetischen Feldes verändern gewisse ferromagnetische Werkstoffe ihre Leitfähigkeit. Eine unabdingbare Voraussetzung für diesen Effekt ist die Existenz einer Anisotropie der elektrischen Leitfähigkeit.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

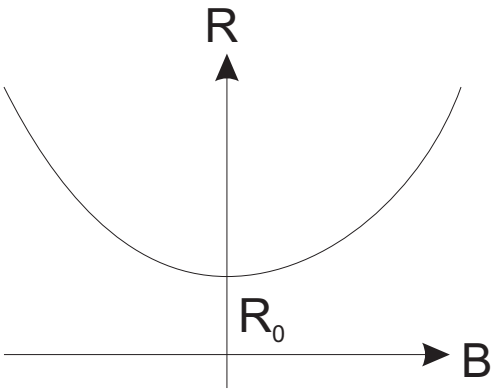



Abbildung 4.114.: Abhängigkeit des Widerstandes in longitudinaler Richtung vom Feld.

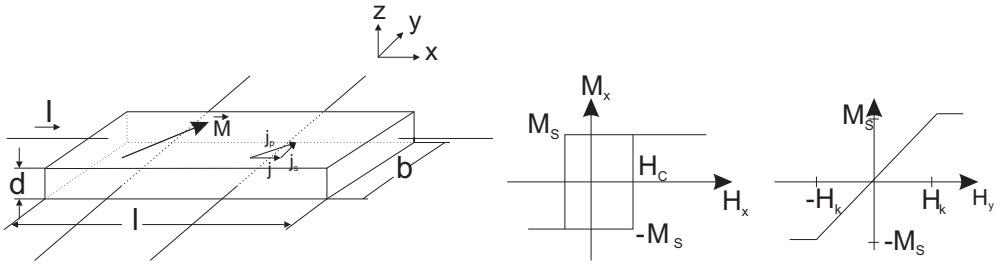



Abbildung 4.115.: Magnetoresistiver Effekt: Feldrichtungen. Rechts: Hysterese

Wenn man annimmt, dass ein Material die Leitfähigkeiten $\rho_p$ parallel zur Magnetisierungsrichtung und $\rho_s$ senkrecht dazu hat, kann eine einfache Ableitung nach[?] angegeben werden.
Wir nehmen an, dass die Magnetisierung zur Stromrichtung den Winkel $\Theta$ einschliesst. Dann ergeben sich für die Felder (siehe auch 4.116)

$$\begin{aligned} \boldsymbol{E} &= \boldsymbol{E}_p + \boldsymbol{E}_s \\ E_p &= E\cos\Theta \\ E_s &= E\sin\Theta \\ a &= E_p\cos\Theta \\ b &= E_s\cos\Theta \\ E &= a + b \\ E &= E_p\cos\Theta + E_s\sin\Theta \end{aligned} \tag{4.2.39}$$

und die Ströme

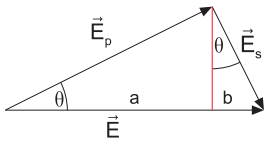



Abbildung 4.116.: Aufteilung von $\boldsymbol{E}$ beim magnetoresistiven Effekt

$$\begin{aligned} \boldsymbol{j} &= \boldsymbol{j}_p + \boldsymbol{j}_s \\ j_p &= j\cos\Theta \\ j_s &= j\sin\Theta \\ a &= j_p\cos\Theta \\ b &= j_s\cos\Theta \\ j &= a + b \\ j &= j_p\cos\Theta + j_s\sin\Theta \end{aligned} \tag{4.2.40}$$

Aus den Materialeigenschaften ergeben sich die Beziehungen

$$\begin{aligned} E_p &= \rho_p j_p \\ E_s &= \rho_s j_s \end{aligned} \tag{4.2.41}$$

Die Gleichung für das elektrische Feld Gleichung (4.2.39) wird mit Gleichung (4.2.41) kombiniert

$$E = \rho_p j_p\cos\Theta + \rho_s j_s\sin\Theta$$

Wir berechnen $j_s$ aus Gleichung (4.2.40)

$$j_s = \frac{j - j_p\cos\Theta}{\sin\Theta}$$

Durch Einsetzen erhalten wir

$$\begin{aligned} E &= \rho_p j_p\cos\Theta + \rho_s\frac{j - j_p\cos\Theta}{\sin\Theta}\sin\Theta \\ &= \rho_p j_p\cos\Theta + \rho_s\left(j - j_p\cos\Theta\right) \\ &= \rho_s j + \left(\rho_p - \rho_s\right) j_p\cos\Theta \\ &= \rho_s j + \left(\rho_p - \rho_s\right) j\cos^2\Theta \end{aligned}$$

Auf andere Weise geschrieben ist das Resultat

$$E(\Theta) = j\rho_s\left(1 + \frac{\rho_p - \rho_s}{\rho_s}\cos^2\Theta\right) \tag{4.2.42}$$

Diese Gleichung kann auf den Widerstand umgerechnet werden.

$$R(\Theta) = \frac{l}{bd}\rho_s + \frac{l}{bd}(\rho_p - \rho_s)\cos^2\Theta \tag{4.2.43}$$

Die Vorzugsrichtung des Detektors wird über die Achse der Magnetisierung eingestellt. Abb. 4.115 rechts zeigt Hysteresekurven entlang der magnetisch harten (rechts) und magnetisch leichten Achse (links).
Entlang der magnetisch harten Achse gilt für kleine Feldstärken $H_y < H_k$

$$M_y(H_y) = H_y\frac{M_s}{H_k} \tag{4.2.44}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

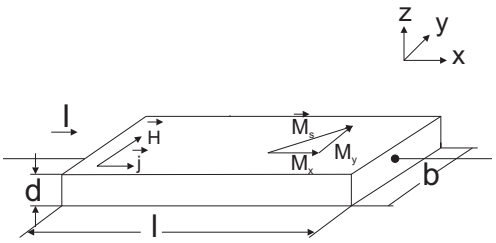



Abbildung 4.117.: Aufbau eines magnetoresistiven Sensors

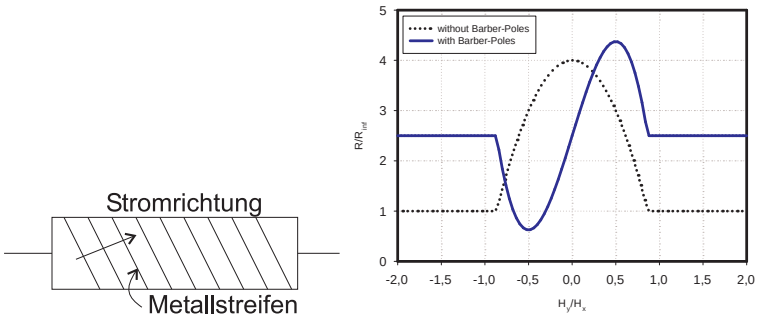



Abbildung 4.118.: Aufbau eines magnetoresistiven Sensors mit 'Barber Poles'. Rechts die Kennlinie mit (durchgezogen) und ohne 'Barber Poles' (gestrichelt)

Wenn die äussere magnetische Feldstärke nur eine Komponente in die y-Richtung aufweist, und die magnetisch harte Achse dieses Materials auch in diese Richtung zeigt, so bekommt man aus Abb. 4.117 unter Vernachlässigung von Randeffekten (Entmagnetisierung) für den Winkel $\Theta$ und die Magnetisierungen, bzw. die Feldstärken

$$\frac{M_y}{M_s} = \sin\Theta = \frac{H_y}{H_k} \text{ für } H_y < H_k \tag{4.2.45}$$

Umgeformt erhält man für kleine Feldstärken $H_y$

$$1 - \left(\frac{H_y}{H_k}\right)^2 = \cos^2\Theta \text{ für } H_y < H_k \tag{4.2.46}$$

Ausserhalb dieses Bereiches ergibt sich der gewöhnliche Widerstand. Zusammenfassend erhält man

$$\begin{aligned} R(H_y) = & \frac{l}{bd}\left(\rho_s + (\rho_p - \rho_s)\left[1 - \left(\frac{H_y}{H_k}\right)^2\right]\right) && \text{für } H_y < H_k \\ R = & \frac{l}{bd}\rho_s && \text{für } H_y > H_k \end{aligned} \tag{4.2.47}$$

Abb. 4.118 zeigt rechts die Kennlinie des Sensors. Die quadratische Abhängigkeit ist sehr schön zu sehen. Bei der Anwendung in Messgeräten stört diese quadratische Abhängigkeit jedoch. Deshalb versucht man, die Stromrichtung und die Magnetisierung M im $45^0$-Winkel festzulegen. Mit Metallstreifen in der gewünschten Richtung, wie in der Abb. 4.118, links, gezeigt, kann dies erreicht werden. Mit diesen sogenannten 'Barber Poles' ist die Linearisierung möglich. Die dazugehörige Kennlinie wird auf der rechten Seite gezeigt.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

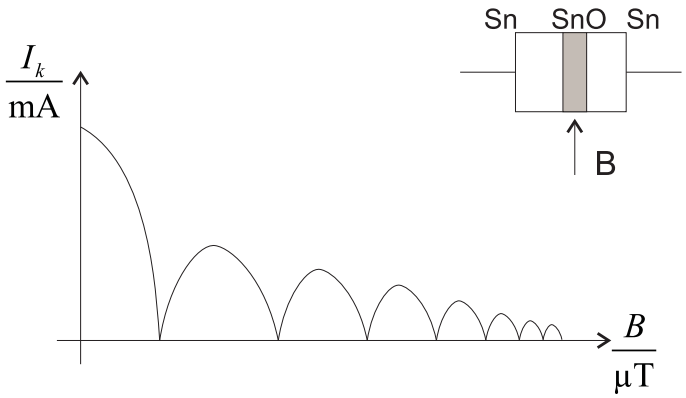



Abbildung 4.119.: Josephsoneffekt. Abhängigkeit des kritischen Stromes von der Flussdichte

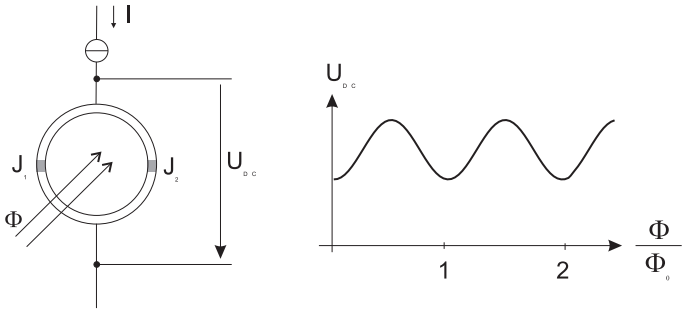



Abbildung 4.120.: Aufbau eines SQUID

#### 4.2.2.5. Josephson-Effekt

Mit dem Josephson-Effekt ist es möglich, sehr kleine Magnetfelder zu messen. Beim Gleichstrom-Josephson-Effekt fliessen Elektronen paarweise (Cooper-Paare) durch die Oxidschicht in Abb. 4.119, die als Tunnelübergang wirkt. Wenn eine kritische Stromstärke $I_K$ überschritten ist, tritt eine Potentialdifferenz auf. Sie rührt vom Tunneln einzelner Elektronen her. Die kritische Stromstärke $I_k$ ist abhängig von der magnetischen Flussdichte $\boldsymbol{B}$ in der Ebene des Tunnelüberganges.

$$I_k = I_{k,0} \frac{\sin \pi \frac{\Phi}{\Phi_0}}{\frac{\Phi}{\Phi_0}} \tag{4.2.48}$$

Dabei ist $\Phi$ magnetischer Fluss im Tunnelübergang. Das elementare Flussquant ist $\Phi_0 = \frac{h}{2e} = 2.07 \times 10^{-15}$Vs. $I_{k,0}$ ist der supraleitende Strom ohne $\boldsymbol{B}$.
Bei einem SQUID (Superconducting Quantum Interference Device) nach Abb. 4.120 beeinflusst der Fluss durch die Öffnung des Ringes die Einteilchen-Wellenzustände der supraleitenden Elektronen. Der Fluss $\Phi$ bewirkt bei einem geeigneten Gleichstrom $I$ eine periodische Abhängigkeit der Potentialdifferenz $U_{DC}$ an den Tunnelkontakten. Die Periode hängt vom magnetischen Flussquant $\Phi_0$ ab und ist, ausgedrückt im externen magnetischen Fluss $B$

$$B = \frac{\Phi_0}{A} = \frac{4\Phi_0}{\pi d^2} \tag{4.2.49}$$

wobei $d$ der Durchmesser des Ringes ist. Für einen Durchmesser von 1 mm er-

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

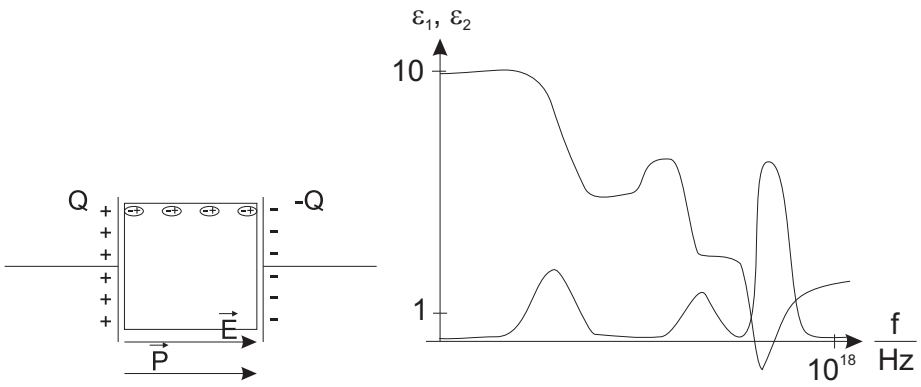



Abbildung 4.121.: Dielektrische Funktion. Links: Prinzipielle Messanordnung. Rechts typisches dielektrisches Spektrum

hält man für die Periode in $B$ den wert 2,6 nT, was etwa 19000 mal weniger als das Erdmagnetfeld ist. Vergrössert man den Durchmesser des Ringes, steigt die Empfindlichkeit. Die Empfindlichkeit des SQUID beruht darauf, dass die grössere Fläche mehr Fluss umfängt. Mit geeigneten techniken lassen sich mit einem SQUID auch Bruchteile eines Flussquantes detektieren.
Wenn das SQUID nicht direkt an der Messstelle sitzen kann, z.B. wegen der notwendigen Kühlung, kann das Magnetfeld mit einer Spule detektiert und über Drähte und eine zweite Spule zum SQUID gebracht werden.

## 4.2.3. Dielektrische Funktion

In Medien hängt die Polarisation $\boldsymbol{P}$ vom elektrischen Feld $\boldsymbol{E}$ ab. Im Allgemeinen gibt es eine nichtlineare Abhängigkeit, die tensoriellen Charakter hat. Für isotrope Materialien hat man $\boldsymbol{P} = \chi\varepsilon_0\boldsymbol{E}$. Mit $\varepsilon_r = 1 + \chi$ erhält man für die elektrische Flussdichte

$$\boldsymbol{D} = (1+\chi)\varepsilon_0\boldsymbol{E} = \varepsilon_r\varepsilon_0\boldsymbol{E} \tag{4.2.50}$$

$\varepsilon_r$ ist im isotropen Falle die dielektrische Funktion. Sie hängt vom Aufbau der Materie ab und reduziert das elektrische Feld im Innern sowie die Kräfte auf Ladungen um $\frac{1}{\varepsilon_r}$ und erhöht die Kapazität von Kondensatoren um $\varepsilon_r$.
In nichtpolaren Medien ist die Verschiebungspolarisation der einzige mögliche Mechanismus. Dabei werden die Schwerpunkte der positiven und negativen Ladungswolken in Atomen oder Molekülen gegeneinander verschoben. Bei polaren Molekülen kann das äussere Feld die bestehenden Dipole und Multipole ausrichten und so die Orientierungspolarisation hervorrufen. Da thermische Fluktuationen die Orientierung in einen zufälligen Zustand zu treiben versuchen, ist die Orientierungspolarisation stark temperaturabhängig. Bei zeitabhängigen äusseren Feldern wirkt sich weiter die Trägheit der zu bewegenden Ladungen aus. Resonanzen und ein frequenzabhängiger Response sind die Folge. Bei einem äusseren Feld $\boldsymbol{E}(\omega) = \boldsymbol{E}_0\cos\omega t$ muss Gleichung (4.2.50) umgeschrieben werden.

$$\boldsymbol{D}(\omega) = \boldsymbol{D}_0\cos(\omega t - \delta) = \varepsilon_0\varepsilon_1\boldsymbol{E}_0\cos\omega t + \varepsilon_0\varepsilon_2\boldsymbol{E}_0\cos\omega t \tag{4.2.51}$$

Amplitude und Phase sind $D_0 = \sqrt{\varepsilon_1^2+\varepsilon_2^2}\varepsilon_0 E_0$ und $\tan\delta = \frac{\varepsilon_2}{\varepsilon_1}$. Dies kann auch

ausgedrückt werden, indem in Gleichung (4.2.50) $\varepsilon_r = \varepsilon_1 - j\varepsilon_2$ gesetzt wird und Flussdichte und Feld frequenzabhängig angesehen werden.
Nach Maxwell (Gleichung (A.1.2) erzeugt eine zeitlich sich ändernde Flussdichte eine Stromdichte $\boldsymbol{j} = \frac{d\boldsymbol{D}}{dt} = j\omega\varepsilon_0\varepsilon_r\boldsymbol{E}$. $\varepsilon_2$ bestimmt also den Strom, der in Phase mit dem elektrischen Feld ist. Dieser Strom verursacht durch Stösse mit den Atomrümpfen, Defekten etc. Dissipation. $\varepsilon_1$ ist ein Mass für die gespeicherte Ladung.
Ist das zu untersuchende Material leitfähig $\boldsymbol{j} = \sigma\boldsymbol{E}$ , dann ist der Gesamtstrom

$$\boldsymbol{j}_{ges} = \sigma\boldsymbol{E} + \frac{\partial\boldsymbol{D}}{\partial t} = \sigma\boldsymbol{E} + j\omega\varepsilon_0\varepsilon_r\boldsymbol{E} = j\omega\varepsilon_0\left(\varepsilon_r - \frac{j\sigma}{\varepsilon_0\omega}\right)\boldsymbol{E} \tag{4.2.52}$$

Man bezeichnet $\varepsilon_r - \frac{j\sigma}{\varepsilon_0\omega} = \varepsilon_1 - j\left(\varepsilon_2 + \frac{j\sigma}{\varepsilon_0\omega}\right)$ als verallgemeinerte dielektrische Funktion. Die rechte Seite von Abb. 4.121 zeigt ein typisches dielektrisches Spektrum.

## 4.2.4. Temperaturmessungen

### 4.2.4.1. Thermowiderstand

Die Leitfähigkeit für elektrischen Strom hängt von der Materialzusammensetzung, der Kristallinität und der Temperatur ab. Die Streuung von Ladungsträgern an Störstellen, Korngrenzen und die Anzahl der beweglichen Ladungsträger bestimmen die Leitfähigkeit.
In isotropen Metallen wirkt nach Anlegen einer Spannung an jedem Punkt eine Feldstärke, die die Elektronen beschleunigt. Streuung und Stösse bremsen die Elektronen wieder ab, so dass der Strom nicht über alle Grenzen wächst. Das elektrische Feld und der Strom hängen über die Leitfähigkeit $\sigma$ (bei anisotropen Materialien ein Tensor) zusammen.

$$\boldsymbol{j} = \sigma\boldsymbol{E} = \frac{1}{\rho}\boldsymbol{E} \tag{4.2.53}$$

Der messbare Widerstand $R$ bei einem Leiter der Länge $l$ und mit dem Querschnitt $A$ ist

$$R = \frac{U}{I} = \frac{E \cdot l}{j \cdot A} = \rho\frac{l}{A} \tag{4.2.54}$$

Nach dem Drude-Modell ist die Leitfähigkeit bei einem idealen Gitter gegeben durch

$$\sigma = qn\mu = qn\frac{\tau_F q}{m^*} \tag{4.2.55}$$

Mit wachsender Temperatur schwingen die Atomrümpfe stärker um ihre Ruhelagen. Dadurch behindern sie den Strom. Die mittlere freie Flugzeit $\tau_F$ nimmt ab. Letztlich nimmt der Widerstand zu.
Nach der Regel von Matthiessen $\rho = \rho_G + \rho_P(T)$ erhält man den spezifischen Widerstand eines Metalls aus dessen Restwiderstand $\rho_G$, der die Wechselwirkung der Elektronen mit **statischen** Defekten beschreibt und aus einem temperaturabhängigen Teil.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Für viele nichtferromagnetische Metalle kann die Temperaturabhängigkeit des spezifischen Widerstandes aus der Debye-Temperatur $\Theta_D$ berechnet werden.

$$\rho_p(T) = \rho(T = \Theta_D)\left[1.17\frac{T}{\Theta_D} - 0.17\right] \quad \text{für} \quad T > 0.15\Theta_D \tag{4.2.56}$$

Bei sehr tiefen Temperaturen hängt der temperaturabhängige Teil des spezifischen Widerstandes wie $T^5$ von der Temperatur ab: der spezifische Widerstand wird konstant beim spezifischen Restwiderstand des Materials.
Für sehr hohe Temperaturen kann Gleichung (4.2.55) linearisiert werden.

$$\rho(T_2) = \rho(T_1)\left[1 + \alpha_{T_1}(T_2 - T_1)\right] \tag{4.2.57}$$

$$\alpha_{T_1} = \frac{1}{\rho(T_1)}\frac{\rho(T_2) - \rho(T_1)}{T_2 - T_1}$$

$$= \frac{1}{\rho(T_1)}\frac{d\rho}{dT} \tag{4.2.58}$$

$$\alpha_{T_1} = \frac{1}{\rho_G + \rho_P(T_1)}\frac{\rho_P(T_2) - \rho_p(T_1)}{T_2 - T_1} \tag{4.2.59}$$

Dabei sind die Widerstandswerte bei den einzelnen Temperaturen mit Gleichung (4.2.56 berechnet worden. Wenn man $T_2 = \Theta_D$ setzt, ergibt sich

$$\alpha_{T_1} = \frac{\rho_p(\Theta_D)\left(0.83 - 1.17\frac{T}{\Theta_D}\right)}{\rho_g + \rho_p(\Theta_D)\left[1.17\frac{T}{\Theta_D} - 0.17\right]} \cdot \frac{1}{\Theta_D - T_1} \quad \text{für} \quad T > 0.15\Theta_D \tag{4.2.60}$$

Bei Vernachlässigung des spezifischen Restwiderstandes $\rho_g$ erhält man für hohe Temperaturen für den Temperaturkoeffizienten des Widerstandes

$$\alpha_{T_1} = \frac{1}{T_1 - 0.145\Theta_D} \quad \text{für} \quad T \gg \Theta_D \tag{4.2.61}$$

Aus den Debye-Temperaturen zwischen 50 K und 400 K ergeben sich bei 293 K Temperaturkoeffizienten zwischen $3.5 \times 10^{-3} \ldots 4.26 \times 10^{-3}/K$. Tabelle I.4 gibt eine Zusammenfassung. Sie zeigt, dass auch bei ziemlich unterschiedlichen Debye-Temperaturen der Temperaturkoeffizient sich nicht sehr viel unterscheidet.
Bei Metallen, bei denen die obige Regel nicht so genau gilt, wird dies

- den Einbau von Fremdatomen, die die Frequenzen der Gitterschwingungen ändern
- Die Gitterschwingungen der Fremdatome
- die durch die Wärmeausdehnung reduzierte *Fermi-Energie*
- die nicht isotrope Streuung an Gitterdefekten

zurückgeführt. Für Metall-Widerstandsthermometer werden Metalle mit hohem Temperaturkoeffizienten und guter Stabilität verwendet, so vor allem Pt, Ni, Ir und Mo. Normiert sind die PT-100 Widerstände, die beider Referenztemperatur auf 100 Ω normiert sind. Abb. 4.122 zeigt die typische Kennlinie eines Metallwiderstandes.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

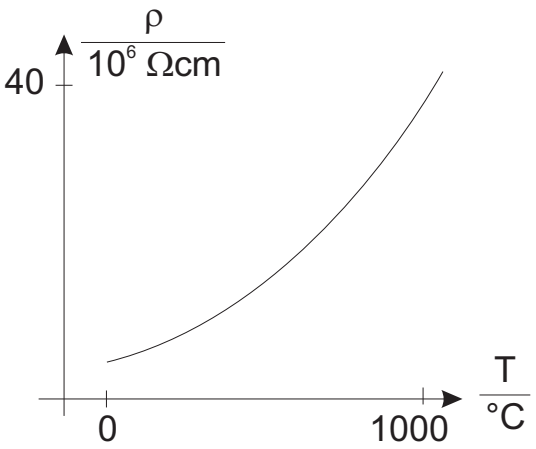



Abbildung 4.122.: Temperaturabhängigkeit des elektrischen Widerstandes

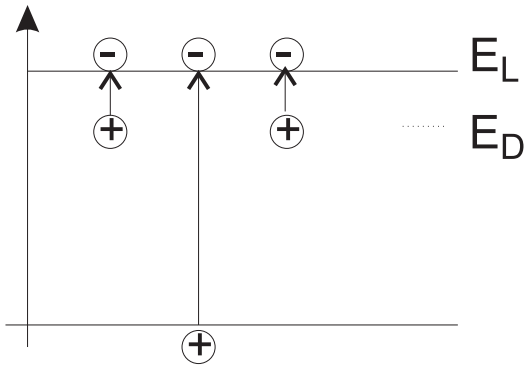



Abbildung 4.123.: Zusammenfassung Bändermodell

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

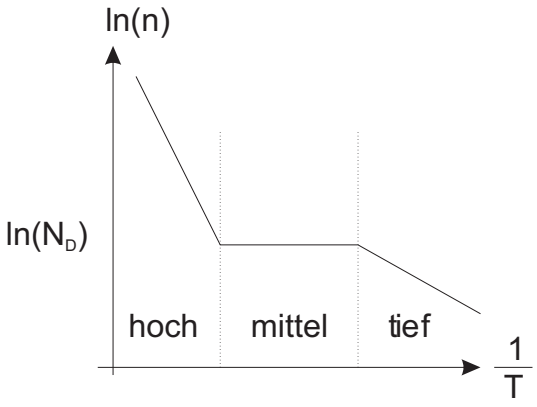



Abbildung 4.124.: Temperaturabhängigkeit der Ladungsträgerkonzentration

#### 4.2.4.2. Temperaturabhängigkeit von Halbleiterübergängen

Bei Halbleitern ist neben der Temperaturabhängigkeit der Streuung vor allem die Konzentration sowohl der positiven wie auch der negativen Ladungsträger als Funktion der Temperatur wichtig. Aus der Gleichung (4.2.55)für die Leitfähigkeit von Metallen kann für Halbleiter die folgende Beziehung postuliert werden, indem man annimmt, dass die Leitung der positiven und negativen Ladungsträger unabhängig ist

$$\sigma(T) = q\left[n(T)\mu_n(T) + p(T)\mu_p(T)\right] \tag{4.2.62}$$

Das Bändermodell in Abb. 4.123 stellt beispielhaft am Leitungsband und am Donatorenband dar, wie die Energielandschaft in der Bandlücke ist. Die Leitfähigkeit von Halbleitern setzt sich aus zwei Teilen, der Störstellenleitung

$$n(T, N_D) = \sqrt{\frac{N_L(T)\cdot N_D}{2}}\cdot e^{\left(-\frac{\Delta E_D}{2kT}\right)} \tag{4.2.63}$$

$$p(T, N_A) = \sqrt{\frac{N_V(T)\cdot N_A}{2}}\cdot e^{\left(-\frac{\Delta E_A}{2kT}\right)} \tag{4.2.64}$$

und der Eigenleitung

$$n(T) = p(T) = n_i(T) = \sqrt{N_L(T)N_V(T)}\cdot e^{\left(-\frac{E_G}{2kT}\right)} \tag{4.2.65}$$

mit $N_L(T) = 2\left(\frac{2\pi m_L^* kT}{h^2}\right)^{\frac{2}{3}}$ und $N_V(T) = 2\left(\frac{2\pi m_V^* kT}{h^2}\right)^{\frac{2}{3}}$ zusammen. In erster Näherung dominieren die exponentiellen Boltzmann-Terme gegen die $T^{3/2}$ Abhängigkeit der Vorfaktoren.
Es gibt im Widerstandsverhalten von Halbleitern drei Bereiche

1. Bei niedrigen Temperaturen ist nur ein Teil der Störstellen ionisiert. Die Zahl der ionisierten Störstellen steigt exponentiell mit $-\frac{E_{D,A}}{2kT}$ an.

2. Bei mittleren Temperaturen sind alle Störstellen ionisiert, die Ladungsträgerkonzentration ist konstant.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

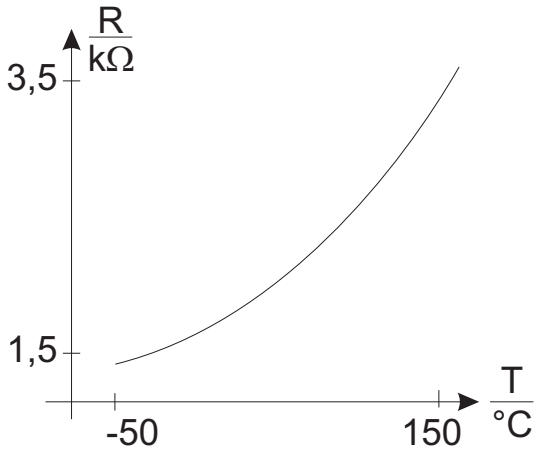



Abbildung 4.125.: Temperaturabhängigkeit der Spreading-Resistance

3. Bei höheren Temperaturen setzt die Eigenleitung ein. Die Ladungsträgerzahl steigt mit $-\frac{E_G}{2kT}$ an.

Abb. 4.124 zeigt eine Skizze dieses Verhaltens.
Im mittleren Bereich, bei einer konstanten Ladungsträgerkonzentration, sollte sich die Leitfähigkeit wie $T^{-3/2}$ verhalten. Tatsächlich beobachtet man aber Exponenten zwischen -1.5 und -2.5. Thermowiderstände aus Silizium werden typischerweise zwischen 220 K und 420 K eingesetzt.
Der Spreading-Widerstand, dessen Kennlinie in Abb. 4.125 gezeigt ist, wird üblicherweise zur Messung des temperaturabhängigen spezifischen Widerstandes eingesetzt. Dabei handelt es sich um einen kreisförmigen, ebenen Kontakt mit dem Durchmesser $d$. Wenn $d$ klein ist gegen die Dicke des Halbleitermaterials erhält man für den Spreading-Resistance

$$R(T) = \frac{\rho(T)}{2d} \quad \text{für} \quad d \gg h \tag{4.2.66}$$

**4.2.4.2.1. Heissleiter** Die Leitungsmechanismen in Halbleitern sind anders als in Metallen. Bei gewissen halbleitenden Keramikwerkstoffen ergibt sich die Leitung durch das Hüpfen von Ladungsträgern von einem Korn zum nächsten. Diese keramischen Werkstoffe werden durch *Sintern* hergestellt. Dabei werden pulverförmige Materialien unter Druck soweit erhitzt, dass die einzelnen Pulverkörner zusammenbacken. Der Widerstand zwischen den Körnern ist dabei sehr viel grösser als in den Körnern. Durch das Vorhandensein dieser Barriere ist die Elektronenleitung ein thermisch aktivierter Prozess. Deshalb ist die Leitfähigkeit bei hohen Temperaturen besser ist als bei tiefen.
Viele Heissleiter sind nach der Strukturformel $A^{2+}B_2^{3+}O_4^{2-}$ aufgebaut. Dabei sind A zweiwertige und B dreiwertige Metalle. Die Metallkationen auf der position A werden tetraedrisch von 4 Sauerstoffanionen umgeben, während B oktaedrisch von 8 Sauertoffanionen umgeben ist. Diese Spinell-Struktur wird durch die Einlagerung von Oxiden so verändert, dass auf B zwei- und drei-wertige Metalle sitzen: damit kann ein Hopping-Prozess durchgeführt werden. Die Bewegung wird durch die Rate des Ablöse- und Einfang-Prozesses bestimmt. Sie wird analog zur Diffusion

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

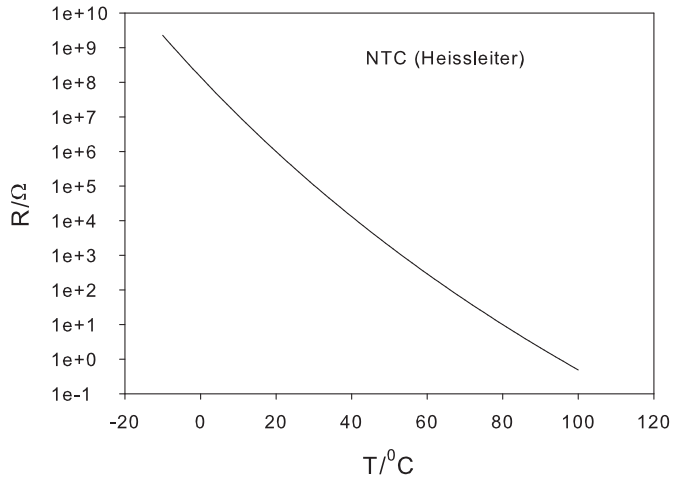



Abbildung 4.126.: Kennlinie eines Heissleiters (NTC)

von Atomen in Festkörpern beschrieben

$$D(T) = D_0(T) \cdot e^{\left(-\frac{E_A}{kT}\right)} \tag{4.2.67}$$

Mit der Einsteinbeziehung wird die Beweglichkeit

$$\mu_n(T) = q\frac{D_0(T)}{kT} \cdot e^{\left(-\frac{E_A}{kT}\right)} \tag{4.2.68}$$

Unter Vernachlässigung des Vorfaktors ergibt sich

$$\rho(T) = \rho_{T\to\infty} \cdot e^{\left(\frac{E_A}{kT}\right)} \tag{4.2.69}$$

oder für eine vorgegebene Widerstandsgeometrie

$$R(T) = A \cdot e^{\left(\frac{B}{T}\right)} \tag{4.2.70}$$

Hier wurden zwei Materialkonstanten A und B eingeführt. man kann sie bestimmen, indem man den Widerstandswert $R$ für zwei feste Temperaturen bestimmt. Die eine dieser festen Temperaturen ist meistens $T_{20} = 293,15K$, der dazugehörige Widerstandswert sei $R_{20} = A\exp(B/T_{20})$. damit wird auch $A = R_{20}\exp(-B/T_{20}$. Die zweite feste Temperatur sei $T_x$ mit dem Widerstandswert $R_x$. Daraus erhält man

$$R_x = R_{20}e^{\left[B\left(\frac{1}{T_x}-\frac{1}{T_{20}}\right)\right]} \tag{4.2.71}$$

Löst man diese Gleichung nach $B$ auf, so ergibt sich

$$B_{20,x} = \frac{\ln\left(\frac{R_x}{R_{20}}\right)}{\frac{1}{T_x}-\frac{1}{T_{20}}} \tag{4.2.72}$$

Damit ist es nun möglich, den Temperaturgang des Widerstandes anzugeben.

$$R(T) = R_{20}e^{\left[B_{20,x}\left(\frac{1}{T}-\frac{1}{T_{20}}\right)\right]} \tag{4.2.73}$$

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

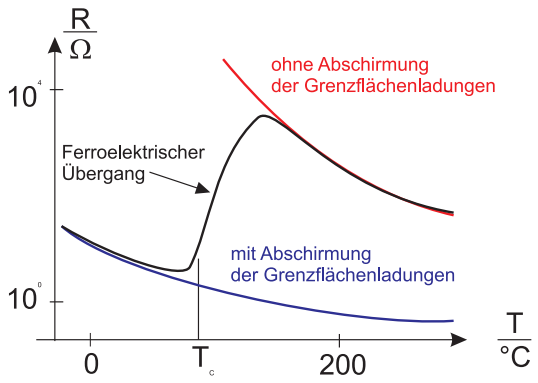



Abbildung 4.127.: Kennlinie eines Kaltleiters (PTC)

Aus Gleichung (4.2.58) folgt für den temperaturabhängigen Temperaturkoeffizienten des Widerstandes

$$\alpha = \frac{B}{T^2} \tag{4.2.74}$$

Heissleiter sind stark nichtlineare Widerstände. Ein typischer Widerstandsverlauf ist in Abb. 4.126 gezeigt. Damit kann man sie, zum Beispiel, als Übertemperatursicherungen verwenden, die direkt am Verbraucher die Versorgungsspannung teilweise kurzschliessen. Heissleiter sind über die Materialzusammensetzung sehr leicht auf einen bestimmten Grundwiderstand und einen gewünschten Temperaturverlauf einstellbar. Typische Heissleitermaterialien sind $Fe_3O_4$, $Zn_2TiO_4$ und viele andere mehr.

**4.2.4.2.2. Kaltleiter** Abb. 4.127 zeigt die Kennlinie eines Kaltleiters. Diese bestehen aus Mischkristallen und Metalloxiden, wie zum Beispiel BaO, CaO, SrO, $ZrO_2$. Die Kaltleiter sind ferroelektrisch. Viele dieser Keramiken haben eine Perowskitstruktur (wie die Hochtemperatursupraleiter) mit der Strukturformel $A^{2+}B^{4+}O_3^{6-}$. Dabei wird die A-Position durch zweiwertige Metalle mit Oxiden des Typs AO besetzt. Die B- Position muss mit vierwertigen Metallen und damit den Oxiden des Typs $BO_2$ besetzt werden. Der Stromfluss in diesen Materialien wird durch die Potentialbarrieren an den Korngrenzen bestimmt. Die dort als *Akzeptoren* chemisorbierten Sauerstoffatome[?] führen zu Verarmungszonen der Weite w. Diese ergibt sich aus aus der Anzahl besetzter Sauerstoffplätze $\tilde{N_S}$ und der Dotierung $n_D$ zu

$$w = \frac{\tilde{N_S}}{n_D} \tag{4.2.75}$$

Aus der Poissongleichung berechnet man die Potentialhöhe $\Psi_0$ der Barriere.

$$\Psi_0 = q^2 \frac{n_D}{2\varepsilon_0\varepsilon_r} d^2 \tag{4.2.76}$$

Der elektrische Widerstand zeigt deshalb eine exponentielle Abhängigkeit von der Temperatur

$$R(T) \propto R_0 e^{\left(\frac{\Psi_0}{kT}\right)} \tag{4.2.77}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Der elektrische Widerstand eines Kaltleiters steigt bei der Curie-Temperatur des Ferroelektrikums abrupt an. Unterhalb der Curie-Temperatur gibt es in den einzelnen Körnern spontane Polarisation. Die negative Korngrenzenladung wird dadurch abgeschirmt. Damit verringert sich bei tiefen Temperaturen (unter der Curie-Temperatur) die Potentialbarrieren zwischen den Körnern. Oberhalb der Curie-Temperatur ist die Dielektrizitätszahl sehr viel geringer als unterhalb. Oberhalb existiert keine ferroelektrische Ordnung und keine spontane Polarisation. Die Abschirmung der Raumladungszonen wird sehr viel ineffektiver, die Potentialbarrieren steigen. damit steigt auch der Widerstand beim Übergang von tiefen zu hohen Temperaturen um 3 bis 6 Grössenordnungen. Der weiteren Zuname wirken die an Korngrenzen bei hohen Temperaturen freigesetzten Ladungsträger entgegen.

**4.2.4.2.3. Integrierte Temperatursensoren** Die Temperaturabhängigkeit des Stromes durch pn-Übergänge oder die Temperaturabhängigkeit der Spannung an solchen Übergängen lässt sich zur Temperaturwmessung ausnutzen.
Nach Shockley ist die Strom-Spannungskennlinie einer $p^+n$-Diode

$$I = I_S \left[ e^{\left(q\frac{U}{KT}\right)} - 1 \right] \tag{4.2.78}$$

$$I_S = qA\frac{D_p p_{n0}}{L_p} = qA\sqrt{\frac{D_p}{\tau_p}}\frac{n_i^2}{N_D}$$

$$\propto T^{3+\gamma/2} e^{\left(-\frac{E_G}{kT}\right)} \tag{4.2.79}$$

dabei ist $I_S$ der Sättigungsstrom und $D_p$ die Diffusionskonstante der Ladungsträger. Näherungsweise ist die Leitfähigkeit eine exponentielle Funktion von $-\frac{E_G}{kT}$.
Mit Gleichung (4.2.65) erhält man für den Strom durch eine in Flussrichtung betriebene Diode

$$I_f(T) \propto e^{\left(q\frac{U_f}{kT} - \frac{E_G}{kT}\right)} \tag{4.2.80}$$

$$U_f(T) \propto \frac{E_G}{q} + \frac{kT}{q}\ln(I_f) \tag{4.2.81}$$

mit der Nebenbedingung $U_f \gg kT/q$. Da der Strom $I_f$ ebenfalls temperaturabhängig ist, erhält man die folgende nichtlineare Kennlinie

$$\alpha^{U_f}(T) = \left.\frac{\partial U_f}{\partial T U_f}\right|_{I_f=const} = -\frac{1}{T}\left(\frac{E_G}{qU_f} - 1\right) \tag{4.2.82}$$

In der Praxis verwendet man als Sensor anstelle einer Diode die Basis-Emitter-Strecke eines Transistors, bei dem Kollektor und Basis kurzgeschlossen sind. Diese Anordnung hat eine Charakteristik, die sehr viel besser mit der Shockley-Gleichung beschrieben werden kann als die einer Diode.
Abb. 4.128 zeigt die Temperaturabhängigkeit der Basis-Emitterspannungsdifferenz zweier gekoppelter Transistoren. Das Verhältnis der Stromabhängigkeit von zweier auf dem gleichen Chip hergestellter Transistoren ist einfach das Verhältnis ihrer

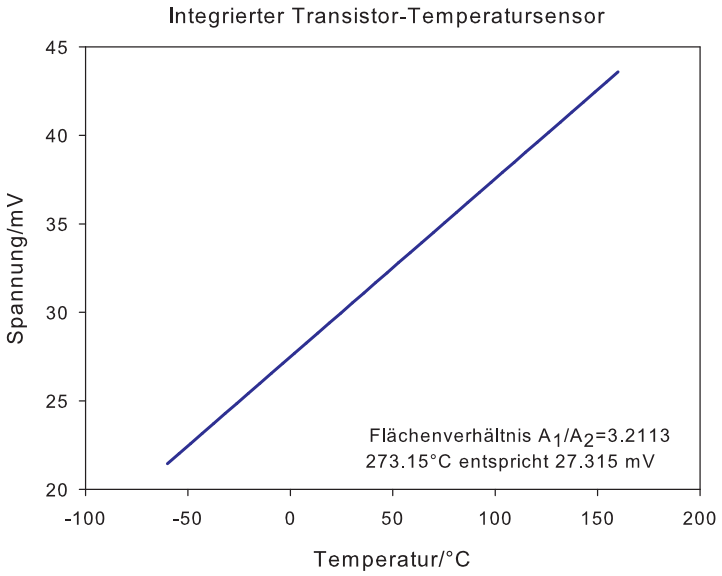



Abbildung 4.128.: Kennlinie eines integrierten Temperatursensors. Das Flächenverhältnis ist so eingestellt, dass die resultierende Spannung in $mV$ ein Zehntel der Temperatur in $K$ ist. Die Temperaturablesung mit einem digitalen Voltmeter wird dann sehr einfach.

Basisflächen $A_1$ und $A_2$. Man erhält in sehr guter Näherung, sofern die Flächen nicht gleich sind, die lineare Kennlinie

$$\Delta U_{BE} = U_{BE_2} - U_{BE_1} = \frac{kT}{q} \ln\left(\frac{I_{f_2}}{I_{f_1}}\right) = \frac{kT}{q} \ln\left(\frac{A_2}{A_1}\right) \tag{4.2.83}$$

Wählt man $A_2/A_1 = 3.2113$ ist $\Delta U_{BE} = 0.1(mV/^0C) \cdot T$.

### 4.2.4.3. Thermoelektrischer Effekt

Wenn über einem Leiterstück der Länge $l$ die Temperatur $\Delta T$ abfällt, dann entsteht die Thermospannung

$$U_{TH}(T) = \int_0^l E(x,T)dx \tag{4.2.84}$$

Bei homogenen Materialien definiert man den Seebeck-Koeffizienten $\alpha_S$. Die Thermospannung ist dann

$$U_{TH} = \alpha_S \int_{T_1}^{T_2} dT \tag{4.2.85}$$

Bei der hochohmigen Messung der Thermospannung nach Abb. 4.129 muss man zwei Materialien A und B mit unterschiedlichen Seebeck-Koeffizienten einsetzen. Man erhält

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

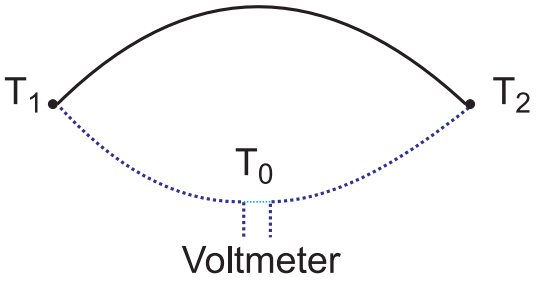



Abbildung 4.129.: Thermospannung

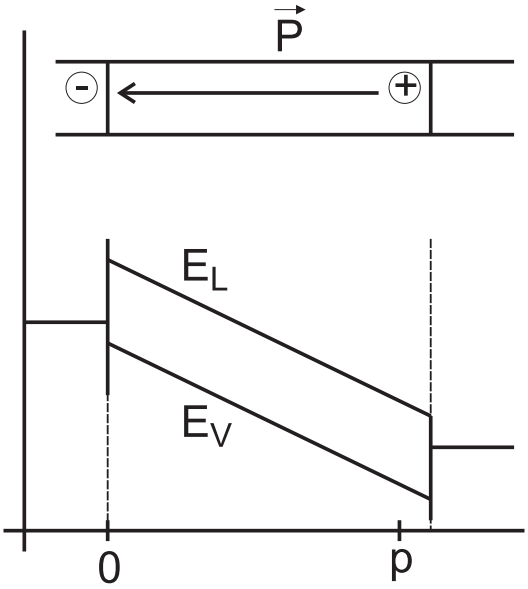



Abbildung 4.130.: Entstehung des pyroelektrischen Effektes

$$U_{TH}(T_1,T_2,T_0) = U_1(T_1,T_0)+U_2(T_2,T_1)+U_3(T_0,T_2) \quad (4.2.86)$$
$$U_{TH}(T_1,T_2,T_0) = \alpha_{S,B}(T_1-T_0)+\alpha_{S,A}(T_2-T_1)+\alpha_{S,B}(T_0-T_2) \quad (4.2.87)$$
$$U_{TH}(T_1,T_2) = (\alpha_{S,A}-\alpha_{S,B})\,(T_2-T_1) \quad (4.2.88)$$

Die Wahl der Materialien von Thermopaaren hängt vom Temperaturbereich, der verlangten Genauigkeit und nicht zuletz auch vom Preis ab. Tabelle I.3 im Anhang zeigt die thermoelektrische Spannungsreihe und die Seebeck-Koeffizienten.

#### 4.2.4.4. Pyroelektrischer Effekt

Abbildung 4.130 zeigt die Messanordnung der Temperatur mit einem Pyroelektrikum. In gewissen unsymmetrisch aufgebauten Kristallen (Triglyzinsulfat (TGS), Lithiumtantalat, Blei-Zirkonat-Titanat (PZT) und PVDF-Folien) mit polarer Achse können spontane elektrische Polarisationen auftreten. Die Änderung der elektrischen Polarisation aufgrund der Änderung der Temperatur nennt man den *pyroelektrischen Effekt*. Zusätzlich werden Oberflächenladungen erzeugt.
Mit den pyroelektrischen Koeffizienten $\boldsymbol{p}(T)$ und unter der Annahme kleiner Tem-

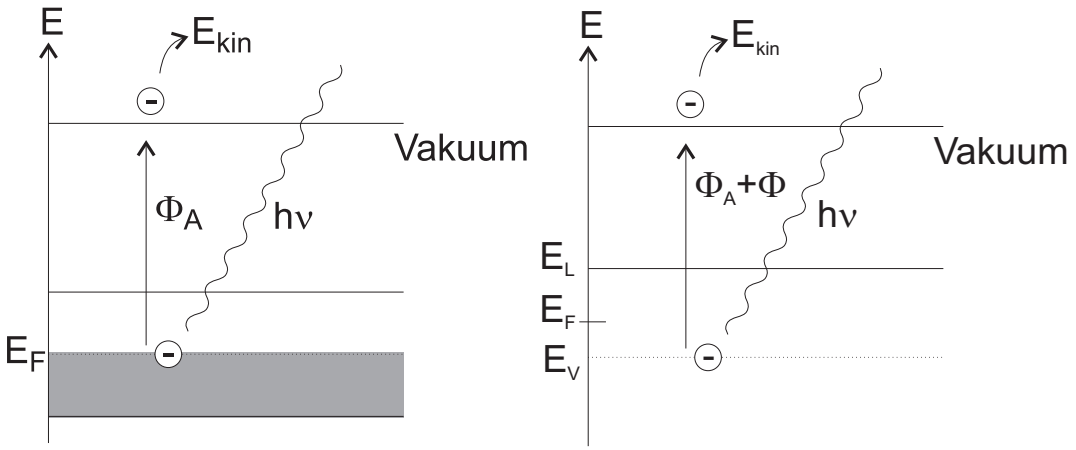



Abbildung 4.131.: Äusserer Photoeffekt: Bändermodelle für Metalle und Halbleiter.

peraturänderungen ($\boldsymbol{p}(T) \approx \boldsymbol{p}_T$) bekommt man für die Polarisation

$$\Delta \boldsymbol{P} = \boldsymbol{p}_T \Delta T \tag{4.2.89}$$

ist das Dielektrikum geeignet orientiert, das heisst, nur die x-Achse ist involviert, erhält man

$$|\Delta Q_{Dip}| = p_{T,x} A \Delta T \tag{4.2.90}$$

$$|\Delta U| = \frac{p_{T,x} A}{C} \Delta T \tag{4.2.91}$$

Um den Einfluss der Kriechströme auf die Ladungsmessung zu minimieren, muss die Wärmestrahlung zerhackt werden. Da alle Pyroelektrika auch Piezoelektrika sind, muss beim experimentieren geachtet werden, dass man nicht den Piezoeffekt fälschlicherweise für den pyroelektrischen Effekt hält.

## 4.2.5. Licht

Licht kann durch seine thermischen, energetische oder mechanischen Wirkungen gemessen werden. Thermische Wirkungen nutzt man aus, wenn die aus Licht absorbierte Wärmemenge (zum Beispiel bei einem Bolometer) oder die absorbierte Leistung (zum Beispiel ein Schwarzer Körper mit einem definierten Wärmeleck an die Umgebung) zu einer Temperaturerhöhung führt, die dann wie im Kapitel 4.2.4 gemessen werden kann.
Licht hat eine Energiedichte, das heisst, ein federnd gelagerter Spiegel wird durch den der Energiedichte äquivalenten Druck ausgelenkt. Dieser Effekt hat nur bei sehr präzisen Messungen oder sehr kleinen Spiegeln einen Einfluss.
Meistens wird Licht über den äusseren oder den inneren Photoeffekt detektiert. Mit dem äusseren Photoeffekt bezeichnet man die Anregung von Ladungsträgern aus dem Leitungs- oder Valenzband über die Vakuumenergie hinaus, wie es in Abb. 4.131 schematisch dargestellt ist. Dass Metalle stabil sind hängt damit zusammen, dass die *Fermi-Energie* um eine *Austrittsarbeit* $\Phi_A$ genannte Energie

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

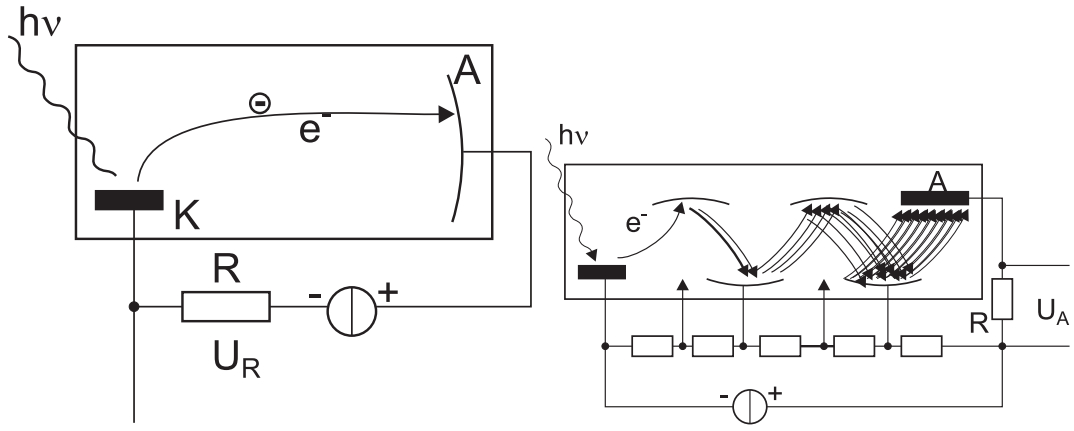



Abbildung 4.132.: Photozelle (links) und Photovervielfacher (rechts).

unter dem Vakuumenerginiveau liegt. Bei den Halbleitern kommt noch die Elektronenbindungsenergie $\phi$ hinzu, die den Abstand der Valenzbandoberkante von der *Fermi-Energie* $E_F$ beschreibt. Folgende Ungleichung

$$E = \hbar\omega \geq \Phi_A + \phi \tag{4.2.92}$$

muss erfüllt sein. Wir können die kinetische Energie der Elektronen $E_{kin}$ oder das dazu äquivalente Potential $U_e$ ausrechnen.

$$E_{kin} = \hbar\omega - \Phi_A - \phi \tag{4.2.93}$$

$$U_e = \frac{E_{kin}}{e} = \frac{\hbar\omega - \Phi_A - \phi}{e} \tag{4.2.94}$$

Schliesslich ergibt sich für die langwellige Grenze

$$\lambda_G = \frac{h \cdot c}{\Phi_A + \phi} = \frac{1.24\,\mu m}{(\Phi_A + \phi)\,[eV]} \tag{4.2.95}$$

Die Fotokathoden sind demnach nur bis zu einer bestimmten Wellenlänge empfindlich. Je länger die noch zu detektierende Wellenlänge sein soll, desto niedriger muss die *Austrittsarbeit* des Kathodenmateriales sein. Häufig werden für den sichtbaren Bereich CsSb und Na/K/Sb- Verbindungen verwendet.

#### 4.2.5.1. Fotozelle und Fotovervielfacher

Abb. 4.132 zeigt links den Aufbau einer Fotozelle. Die durch das Licht aus der Fotokathode K herausgelösten Elektronen werden durch die Spannung zur Anode hin beschleunigt. der entstehende Strom wird als Spannungsabfall über dem Widerstand R gemessen. Um Stösse der Elektronen zu vermeiden muss der Raum zwischen Fotokathode und Anode evakuiert sein. Die spektrale Empfindlichkeit hängt vom verwendeten Kathodenmaterial ab. Im Durchschnitt werden für jedes absorbierte Photon etwa 0.1 Elektronen emittiert. Um einen Strom von 1 pA zu bekommen, müssen also $10^{-12}/(1.6 \times 10^{-19}) = 6.25 \times 10^7$ Photonen absorbiert werden, was bei einer Photonenenergie von 2 eV einer Lichtleistung von 20 pW entspricht. Da es schwierig ist kleinere Ströme zu messen, ist dies die praktische Grenze der Empfindlichkeit.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

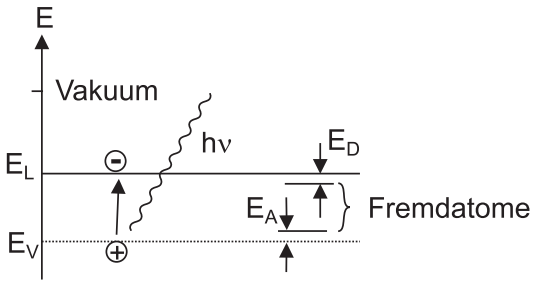



Abbildung 4.133.: Innerer Photoeffekt für Halbleiter.

Um auch einzelne Photonen detektieren zu können verwendet man Photovervielfacher (Photo Multiplier), wie in Abb. 4.132 gezeigt. Dabei werden die emittierten Photoelektronen über eine Spannung von etwa 100 V beschleunigt und auf eine Zwischenelektrode geschickt. Ihre kinetische Energie bewirkt, dass mehr al 1 Elektron, im allgemeinen $\upsilon$ Elektronen, für jedes eintreffende Elektron freigesetzt werden. Die von der Anode aufgefangene Elektronenzahl ist für n solcher Verstärkungsstufen $\upsilon^n$. Bei 10 Stufen und einem $\upsilon = 4.78$ würde für jedes Photon 6250000 Elektronen erzeugt.
da bei einer Photozelle oder einem Photovervielfacher nur sehr kleine Kapazitäten vorkommen, können Frequenzen bis zu 10 GHZ detektiert werden. dabei ist zu beachten, dass die Laufzeiten im Detektor einige Nanosekunden betragen können. Abb. 4.133 zeigt das Bändermodell für den innern Photoeffekt. Neben der Anregung vom Valenzband ins Leitungsband mit der Energie grösser als $E_g$ können auch Anregungen vom Valenzband in Fremdatomzustände (Energie $E_A$) oder von von Fremdatomzuständen ins Leitungsband (hier auch mit $E_A$) auftreten.

### 4.2.5.2. Photowiderstand

Bei einem Photowiderstand ändert sich sein Leitwert mit der Anzahl vorhandener Ladungsträger. Wenn das Licht vollständig absorbiert wird, der Quantenwirkungsgrad $\eta$, die Strahlungsleistung $P$, der Querschnitt des Widerstandes $A$ und seine Länge $l$ sind, ist die Rate, mit der Ladungsträger ins Leitungsband angeregt werden

$$g = \eta \frac{1}{A \cdot l} \frac{P}{\hbar\omega} \tag{4.2.96}$$

Die Rekombinationsrate $r$ hängt von der mittleren Lebensdauer $\tau$ und der Ladungsträgerkonzentration $n$ ab

$$r = \frac{n}{\tau} \tag{4.2.97}$$

Im Gleichgewicht ist die Rekombinationsrate $r$ gleich der Generationsrate $g$. Die Anzahl Ladungsträger im Leitungsband ist also

$$n = \tau\eta \frac{1}{A \cdot l} \frac{P}{\hbar\omega} \tag{4.2.98}$$

Andererseits kann man für den Strom im Halbleiter bei bekannter Beweglichkeit $\mu_n$ schreiben

$$I = \frac{E}{\rho} A = e\mu_n n E A \tag{4.2.99}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

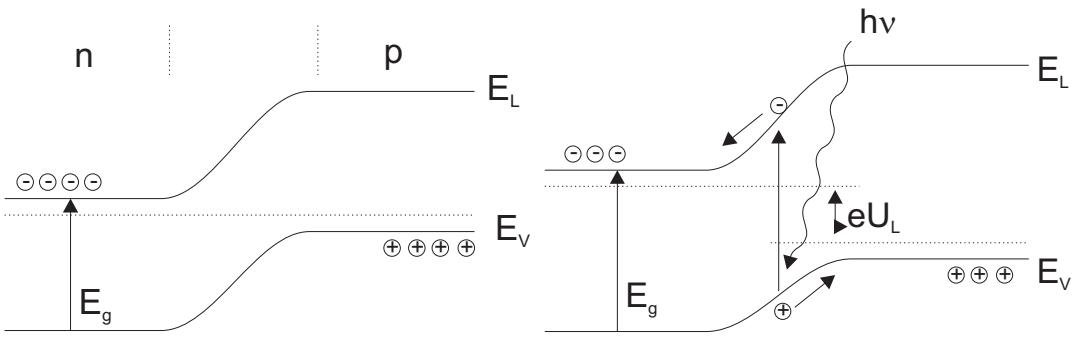



Abbildung 4.134.: Bändermodelle für den inneren Fotoeffekt: links unbeleuchtet und rechts beleuchtet.

Die Kombination von Gleichungen (4.2.98) und (4.2.99) gibt

$$I(P) = e\left(\eta\frac{P}{\hbar\omega}\right)\left(\frac{\mu_n\tau E}{l}\right) = I_{Ph}\left(\frac{\mu_n\tau E}{l}\right) \tag{4.2.100}$$

Die Transitzeit (Zeit zum Durchqueren des Widerstandes) sei $t_{tr} = \frac{l}{\mu_n E}$. Dann ist der Verstärkungsfaktor

$$M_0 = \frac{I}{I_{Ph}} = \frac{\mu_n\tau E}{l} = \frac{\tau}{t_{tr}} \tag{4.2.101}$$

und hängt nur vom Verhältnis der mittleren Ladungsträgerlebensddauer und zur Transitzeit zwischen den Elektronen ab.
Die spektrale Empfindlichkeit eines Photoleiters basierend auf dem inneren Photoeffekt ist

$$S_\lambda^I = \frac{I}{P} = \frac{M_0 I_{Ph}(P)}{P} = \frac{e\eta}{\hbar\omega}M_0 = \frac{e\eta}{h\cdot c}M_0\lambda \tag{4.2.102}$$

Die Empfindlichkeit steigt also linear mit der Wellenlänge an, solange die Grenzwellenlänge, die durch die minimalen Energiesprünge gegeben ist, nicht erreicht werden. Photoleiter haben demnach ein ideales Quantenverhalten. Sie werden für das sichtbare Licht aus CdSe und CdS und für den infraroten Bereich aus PbS, PbSe, PbTe und InSb hergestellt.

### 4.2.5.3. Photodiode

In der *Raumladungszone eines pn-Überganges* entstehen bei der Beleuchtung Elektron-Loch-Paare. Wie die linke Seite von Abb. 4.134 zeigt, werden die Elektronen zur n-Zone und die Löcher zur p-Zone beschleunigt. Dadurch nimmt die elektrische Feldstärke in der *Raumladungszone* und damit auch die Barrierenhöhe ab. Die Fermienergien verschieben sich, es entsteht eine elektromotorische Kraft. Diese wird als fotovoltaische Spannung $U_L$ aussen abgegriffen.
Der generierte Fotostrom ist

$$I_{Ph}(P) = e\cdot g(P)\cdot A\cdot l = e\cdot\eta\frac{P}{\hbar\omega} \tag{4.2.103}$$

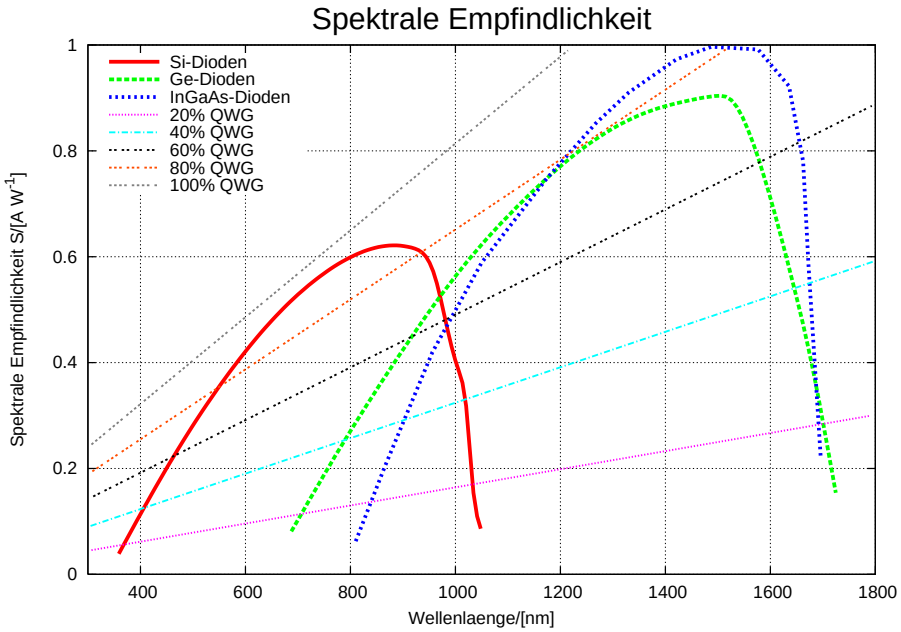



Abbildung 4.135.: Spektrale Empfindlichkeit von verschiedenen Materialien im Vergleich zu Quanteneffizienzen.

Analog ist die spektrale Empfindlichkeit (siehe auch Abb. 4.135)

$$S_\lambda^I = \frac{I}{P} = \frac{I_{Ph}(P)}{P} = \frac{e\eta}{\hbar\omega} = \frac{e\eta}{h \cdot c}\lambda \tag{4.2.104}$$

Anders als bei Photowiderständen gibt es hier keinen durch die aussen angelegte Spannung bedingten Verstärkungsfaktor $M_0$. Photodioden sind deshalb schneller und weniger empfindlich. Da die *Raumladungszone* eine sehr geringe Tiefe $l$ hat, muss der Quantenwirkungsgrad $\eta$ als wellenlängenabhängig angenommen werden. Wird die Photodiode in Sperrichtung vorgespannt, dann ist der Potentialverlauf in der *Raumladungszone* steiler, entsprechend werden die von Licht generierten Ladungsträger schneller zu den Anschlüssen befördert: die Photodiode wird schneller. Wird die Vorspannung so gross, dass die Energie der vom Licht generierten La-

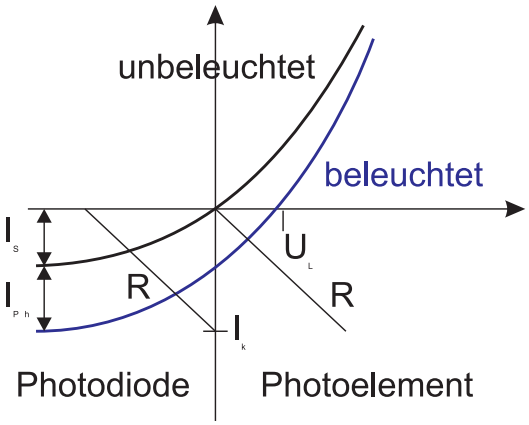



Abbildung 4.136.: Kennlinie einer Photodiode (schwarz, unbeleuchtet und blau, beleuchtet).

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

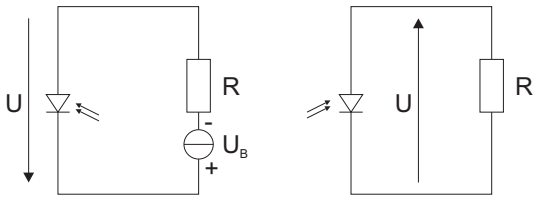



Abbildung 4.137.: Photodiode (links) und Photoelement (rechts)

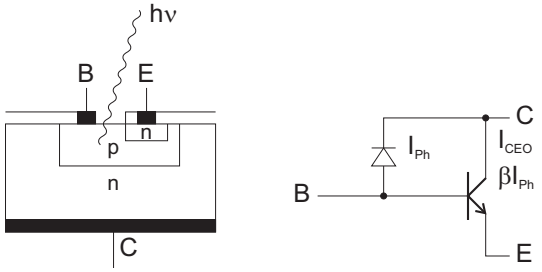



Abbildung 4.138.: Phototransistor: links Aufbau und rechts Ersatzschema.

dungsträger ausreicht, weiter Ladungsträger zu generieren (Avalanche-Effekt oder Lawineneffekt) dann hat die dann Avalanche-Photodiode genannte Diode eine innere Verstärkung, die in besonderen Fällen zum Zählen einzelner Photonen ausreicht. Die Strom-Spannungskennlinie (siehe Abb. 4.136) einer Photodiode lässt sich analog zu der einer gewöhnlichen Diode als

$$I = I_S \cdot \left(e^{e\frac{U}{kT}} - 1\right) - I_{Ph}(P) \tag{4.2.105}$$

Je nach äusserer Beschaltung betreibt man die Photodiode als Photodiode (Abb. 4.137, links) oder als Photoelement (Abb. 4.137, links). Die Photodiode arbeitet im 3. Quadranten des Kennlinienfeldes in Abb. 4.136, das Photoelement im vierten. Die Beschaltung mit einem Widerstand $R$ ist in Abb. 4.137 eingezeichnet. Der Schnittpunkt der jeweiligen Arbeitsgeraden mit der Widerstandskennlinie ergibt den Arbeitspunkt. Als Material für Photodioden wird bevorzugt *Si* verwendet, für den Infrarotbereich auch Ge und InSb.

#### 4.2.5.4. Phototransistor

Abb. 4.138 zeigt, wie man die Empfindlichkeit einer Photodiode steigern kann, indem man sie als Stromquelle an der Basis eines Transistors verwendet. Die linke Seite zeigt einen Querschnitt durch einen *Phototransistor*, der sich von einem gewöhnlichen *Transistor* durch seinen grösseren Basis-Emitterbereich unterscheidet. Wenn der *Transistor* die Stromverstärkung $\beta$ hat so ist der Kollektorstrom

$$I_{CEO} = I_{Ph}(P) + \beta I_{Ph}(P) = (1+\beta)\, I_{Ph}(P) \tag{4.2.106}$$

Entsprechend ist die spektrale Empfindlichkeit

$$S_\lambda^I = \frac{I}{P} = \frac{(1+\beta)\, I_{Ph}(P)}{P} = \frac{e\eta\,(1+\beta)}{\hbar\omega} = \frac{e\eta\,(1+\beta)}{h\cdot c}\lambda \tag{4.2.107}$$

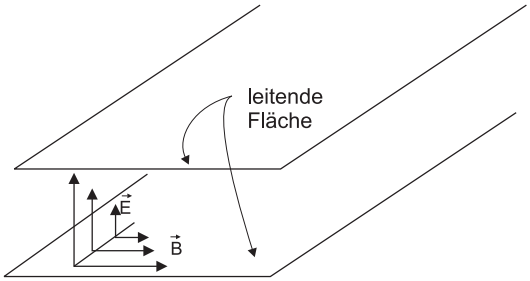



Abbildung 4.139.: Schematische Darstellung einer plattengeführten elektromagnetischen Welle.

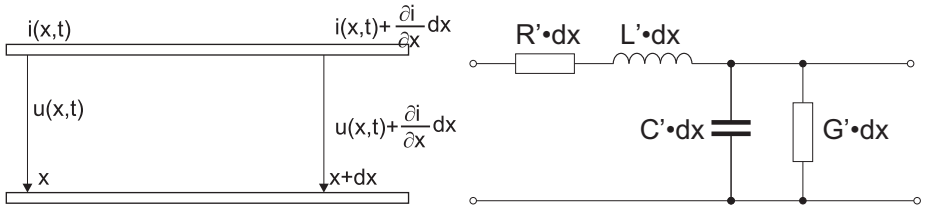



Abbildung 4.140.: Links die Darstellung eines Wellenleiters. Spannung und Strom ändern sich bei einer Verschiebung um $dx$ um die entsprechenden infinitesimalen Grösse. Rechts das Ersatzschaltbild.

## 4.3. Leitungen

Ebene Wellen (Siehe im Anhang Abschnitt A.4) können zwischen zwei Metallplatten problemlos geführt werden. Abb. 4.139 zeigt, dass, wenn das elektrische Feld senkrecht zu den Platten steht, dies mit den Randbedingungen vereinbar ist. Der Abstand der Platten kann so klein man will gewählt werden, Leitung hat man immer noch. Die in der Abbildung gezeigte Welle heisst TEM-Plattenwelle.
Tabelle 4.7 zeigt eine Zusammenstellung verschiedener Wellenleiter, bei denen die Dimensionen klein gegen die Wellenlänge sind.

### 4.3.1. Leitungsgleichungen

Nach Abb. 4.140 kann man Zweidrahtleitungen, Koaxialleitungen und Streifenleitungen mit folgendem phänomenologischen Ansatz[**?**] behandeln: Die Änderung der Spannung $u(x,t)$ und des Stromes $i(x,t)$ längs der Leiterstrecke $dx$ geschieht wegen

- des Längswiderstandes $R = R' \cdot dx$, wobei $R'$ der Widerstandsbelag ist
- der Längsinduktivität $L = L' \cdot dx$, wobei $L'$ der Induktivitätsbelag ist
- des Querleitwertes $G = G' \cdot dx$, wobei $G'$ der Leitwertbelag ist
- des Querkapazität $C = C' \cdot dx$, wobei $C'$ der Kapazitätsbelag ist

Dies ist das in Abb. 4.140, rechts, angegebene Ersatzschaltbild. Für die Strom-und Spannungsänderung erhalten wir

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

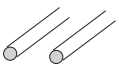
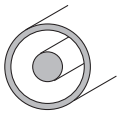
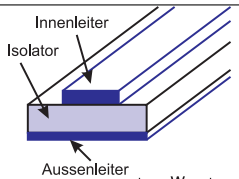
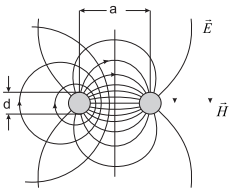
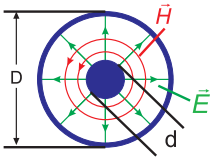
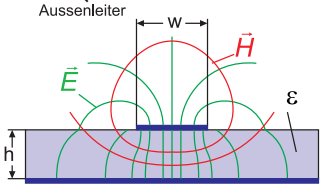
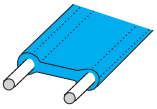
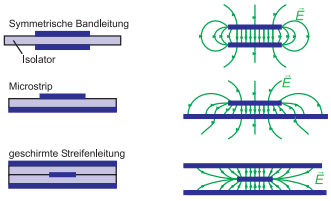
| TEM Wellentyp | (symmetrische) Paralleldrahtleitung (Lecherleitung) | Koaxialleitung | Streifenleitung |
|---|---|---|---|
| Leiter-Grundform |  |  |  |
| elektrische und magnetische Feldlinien |  |  |  |
| Leitungs-Wellenwiderstand | $Z = \frac{120}{\sqrt{\varepsilon}} \cdot \ln \frac{2a}{d} [\Omega]$ für $a > 2.5 \cdot d$ | $Z = \frac{60}{\sqrt{\varepsilon}} \cdot \ln \frac{D}{d} [\Omega]$ | $Z = \frac{120\pi}{\sqrt{\frac{\varepsilon+1}{2} + \frac{\varepsilon-1}{2\sqrt{1+12 \cdot h/w}}}} \cdot \frac{1}{\frac{w}{h}+1.393+0.667 \cdot \ln\left(\frac{w}{h}+1.444\right)}$ $[\Omega]$ für $w \geq h$ |
| Variationen in den Ausführungsformen |  Anwendung als Antennenleitung bei hohen Frequenzen ($Z = 240\Omega, Z = 300\Omega$) | Der Aussenleiter wird meist als Drahtgeflecht ausgeführt. ($Z = 50\Omega, Z = 60\Omega, Z = 75\Omega, Z = 200\Omega$) |  |

Tabelle 4.7.: Homogene Leitungen, deren Abmessungen klein gegen die Wellenlänge sind

$$u(x,t) - u(x+dx,t) = -\frac{\partial u}{\partial x}\cdot dx \qquad = R'\cdot dx\cdot i(x,t) + L'\cdot dx\cdot\frac{\partial i}{\partial t}$$
$$i(x,t) - i(x+dx,t) = -\frac{\partial i}{\partial x}\cdot dx \qquad = G'\cdot dx\cdot u(x,t) + C'\cdot dx\cdot\frac{\partial u}{\partial t}$$

Die Leitungsgleichungen in differentieller Form lauten also

$$-\frac{\partial u}{\partial x} = R'\cdot i(x,t) + L'\cdot\frac{\partial i}{\partial t} \tag{4.3.1}$$
$$-\frac{\partial i}{\partial x} = G'\cdot u(x,t) + C'\cdot\frac{\partial u}{\partial t} \tag{4.3.2}$$

Wenn Gleichung (4.3.1) nach x und Gleichung (4.3.2) nach t abgeleitet werden, können die Gleichungen kombiniert werden zur Leitungs-Wellengleichung[10] oder Telegraphengleichung.

$$\frac{\partial^2 u}{\partial x^2} = R'\cdot G'\cdot u(x,t) + (R'\cdot C' + G'\cdot L')\cdot\frac{\partial u(x,t)}{\partial t} + L'\cdot C'\cdot\frac{\partial^2 u(x,t)}{\partial t^2} \tag{4.3.3}$$

oder für den Strom

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

#### 4.3.1.1. Fortschreitende, gedämpfte Wellen

Lösungen der Leitungs-Wellengleichung bekommt man mit dem Ansatz $u(x,t) = \underline{u}(\omega)\cdot e^{-\gamma x}\cdot e^{i\omega t} + c.c.$ sowie $i(x,t) = \underline{i}(\omega)\cdot e^{-\gamma x}\cdot e^{j\omega t} + c.c.$
Damit wird die Fortpflanzungskonstante

$$\gamma = \alpha + j\beta = \sqrt{(R' + j\omega L')\cdot(G' + j\omega C')} \tag{4.3.4}$$

Setzt man den Ansatz in Gleichungen (4.3.1) und (4.3.2) ein, so sind die komplexen Amplituden

$$\gamma\cdot\underline{u}(\omega) = (R' + j\omega L')\cdot\underline{i}(\omega)$$
$$\gamma\cdot\underline{i}(\omega) = (G' + j\omega C')\cdot\underline{u}(\omega)$$

Der Quotient aus komplexer Spannungsamplitude und Stromamplitude ist der Leitungs-Wellenwiderstand

$$\underline{z}(\omega) = \frac{\underline{u}(\omega)}{\underline{i}(\omega)} = \sqrt{\frac{R' + j\omega L'}{G' + j\omega C'}} \tag{4.3.5}$$

Bei einer verlustlosen Leitung ($R' = 0$, $G' = 0$) ist $\gamma = \alpha + j\beta = j\omega\sqrt{L'C'}$, also rein imaginär. Die Dämpfung ist null ($\alpha = 0$) und $\beta \equiv \omega/c_{ph} = \omega\sqrt{L'C'}$. Damit ist der Wellenwiderstand der verlustlosen Leitung

$$\underline{z}(\omega) \equiv Z = \sqrt{\frac{L'}{C'}} \tag{4.3.6}$$

[10] O. Heaviside 1892

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

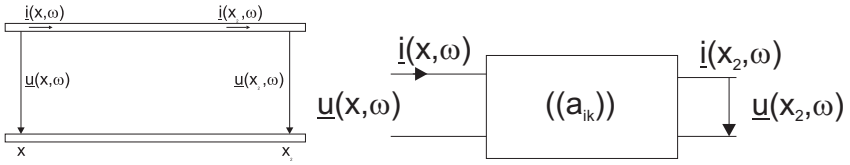



Abbildung 4.141.: Koordinaten am Leitungsende (links) und Vierpoldarstellung eines Leiterstückes.

und die Phasengeschwindigkeit der Leitungswelle

$$c_{ph} = \frac{1}{\sqrt{L' \cdot C'}} = \frac{1}{C' \cdot Z} = \frac{Z}{L'} \tag{4.3.7}$$

#### 4.3.1.2. Wellenfeld auf einem endlichen Leitungsstück mit reflektierter Welle

Als allgemeine Lösung setzen wir eine hin- und eine herlaufende Welle an.

$$\begin{aligned} u(x,t) &= \left[\underline{u}_r(\omega) \cdot e^{-\gamma x} + \underline{u}_l(\omega) \cdot e^{\gamma x}\right] \cdot e^{i\omega t} + c.c. &&= \underline{u}(x,\omega) \cdot e^{j\omega t} \\ i(x,t) &= \frac{1}{\underline{z}(\omega)} \left[\underline{u}_r(\omega) \cdot e^{-\gamma x} - \underline{u}_l(\omega) \cdot e^{\gamma x}\right] \cdot e^{j\omega t} + c.c. &&= \underline{i}(x,\omega) \cdot e^{j\omega t} \end{aligned}$$

Man erhält diese Gleichungen, indem Gleichung (4.3.5) benutzt wird. Eine nach links laufende Welle kehrt dabei die Stromrichtung um.
Unter Verwendung der Koordinaten in Abb. 4.141 bekommt man für $x_2$

$$\begin{aligned} \underline{u}(x_2,\omega) &= \underline{u}_r \cdot e^{-\gamma x_2} + \underline{u}_l \cdot e^{\gamma x_2} \\ \underline{z}(\omega) \cdot \underline{i}(x_2,\omega) &= \underline{u}_r \cdot e^{-\gamma x_2} - \underline{u}_l \cdot e^{\gamma x_2} \end{aligned}$$

Wir können diese Gleichungen nach $\underline{u}_r$ und $\underline{u}_l$ auflösen und erhalten

$$\underline{u}_r(\omega) = \left[\underline{u}(x_2,\omega) + \underline{z}(\omega) \cdot \underline{i}(x_2,\omega)\right] \cdot e^{\gamma \frac{x_2}{2}} \tag{4.3.8}$$

$$\underline{u}_l(\omega) = \left[\underline{u}(x_2,\omega) - \underline{z}(\omega) \cdot \underline{i}(x_2,\omega)\right] \cdot e^{-\gamma \frac{x_2}{2}} \tag{4.3.9}$$

Eingesetzt in unseren Ansatz ergibt sich

$$\underline{u}(x,\omega) = \underline{u}(x_2,\omega) \cosh \gamma(x_2 - x) + \underline{z}(\omega) \cdot \underline{i}(x_2,\omega) \sinh \gamma(x_2 - x) \tag{4.3.10}$$

$$\underline{i}(x,\omega) = \underline{i}(x_2,\omega) \cosh \gamma(x_2 - x) + \underline{u}(x_2,\omega) \sinh \gamma(x_2 - x) \frac{1}{\underline{z}(\omega)} \tag{4.3.11}$$

Diese Gleichung kann auch in Matrixschreibweise angegeben werden

$$\begin{pmatrix} \underline{u}(x,\omega) \\ \underline{i}(x,\omega) \end{pmatrix} = \begin{pmatrix} \cosh \gamma(x_2 - x) & \underline{z}(\omega) \sinh \gamma(x_2 - x) \\ \frac{\sinh \gamma(x_2 - x)}{\underline{z}(\omega)} & \cosh \gamma(x_2 - x) \end{pmatrix} \cdot \begin{pmatrix} \underline{u}(x_2,\omega) \\ \underline{i}(x_2,\omega) \end{pmatrix} \tag{4.3.12}$$

Damit stehen die Leitungsgleichungen in Vierpol-Kettenform da (siehe auch Abb. 4.141, rechts).

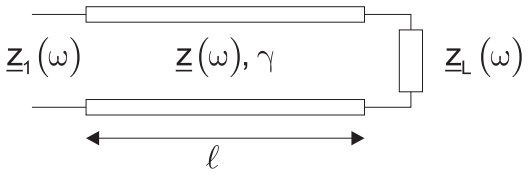



Abbildung 4.142.: Einfluss des Abschlusswiderstandes.

#### 4.3.1.3. Anwendung der Leitungsgleichungen

Mit den Leitungsgleichungen können

- durch ein Leitungsstück gegebener Länge $\ell$ Strom in Spannung oder Spannung in Strom transformiert werden.
- eine Abschlussimpedanz $\underline{z}_L(\omega)$ in eine Eingangsimpedanz $\underline{z}_1(\omega)$ transformiert werden.

Im zweiten Fall wendet man die Gleichung (4.3.12) an und erhält für die Anordnung nach Abb. 4.142

$$\underline{z}_1(\omega) = \left.\frac{\underline{u}(x)}{\underline{i}(x)}\right|_{x=x_2-\ell} = \frac{\underline{z}_L(\omega)\cdot\cosh\gamma\ell + \underline{z}(\omega)\cdot\sinh\gamma\ell}{\underline{z}_L(\omega)\cdot\sinh\gamma\ell + \underline{z}(\omega)\cdot\cosh\gamma\ell}\cdot\underline{z}(\omega) \tag{4.3.13}$$

Für ein verlustloses Leitungsstück gilt die Gleichung (4.3.6), $\underline{z}(\omega) \equiv Z$ und die dazu führenden Überlegungen so dass die Gleichungen (4.3.10) und (4.3.11)

$$\underline{u}(x,\omega) = \underline{u}(x_2,\omega)\cos\beta(x_2-x) + Z\cdot\underline{i}(x_2,\omega)\sin\beta(x_2-x) \tag{4.3.14}$$

$$\underline{i}(x,\omega) = \underline{i}(x_2,\omega)\cos\beta(x_2-x) + \underline{u}(x_2,\omega)\sin\beta(x_2-x)\frac{1}{Z} \tag{4.3.15}$$

wird. Analog ergibt sich für die Widerstandstransformation nach Gleichung (4.3.13)

$$\underline{z}_1(\omega) == \frac{\underline{z}_L(\omega)\cdot\cos\beta\ell + j\cdot Z\cdot\sin\beta\ell}{j\cdot\underline{z}_L(\omega)\cdot\sin\beta\ell + Z\cdot\cos\beta\ell}\cdot Z \tag{4.3.16}$$

#### 4.3.1.4. Kurzgeschlossene Leitung

Bei einer kurzgeschlossenen, verlustlosen Leitung (dann ist $\underline{z}_L(\omega) = 0$) ist die Eingangsimpedanz mit Gleichung (4.3.16) durch

$$\underline{z}_1^K = j\cdot Z\cdot\tan\beta\ell = j\cdot Z\tan\frac{2\pi\ell}{\lambda} \tag{4.3.17}$$

gegeben. Die Schaltung und der *Eingangswiderstand* sind in Abb. 4.143 angegeben. Die Eingangsimpedanz wird zweckmässigerweise als Funktion von $\ell/\lambda$ angegeben. Man findet folgendes Verhalten:

| | |
|---|---|
| Leitungen mit $\ell < \lambda/4$: | induktives Verhalten |
| Leitungen mit $\ell = \lambda/4$: | Verhalten wie beim *Parallelschwingkreis* |
| Leitungen mit $\lambda/4 < \ell < \lambda/2$: | kapazitives Verhalten |
| Leitungen mit $\ell = \lambda/2$: | Verhalten wie beim Serienschwingkreis |

Dann wiederholt sich dieses Verhalten.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

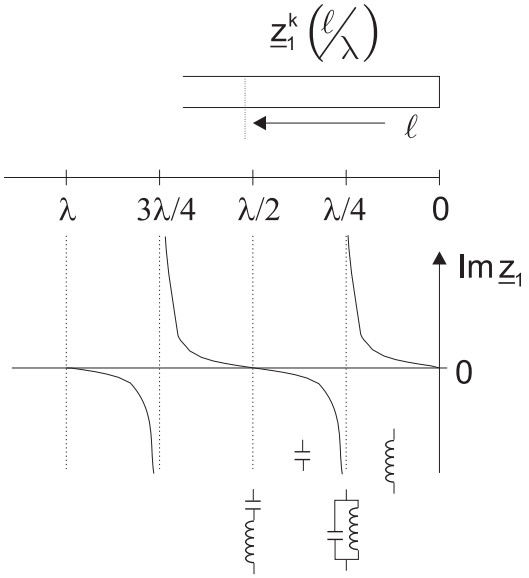



Abbildung 4.143.: Kurzgeschlossene Leitung. Unten ist der Imaginärteil der Impedanz gezeichnet.

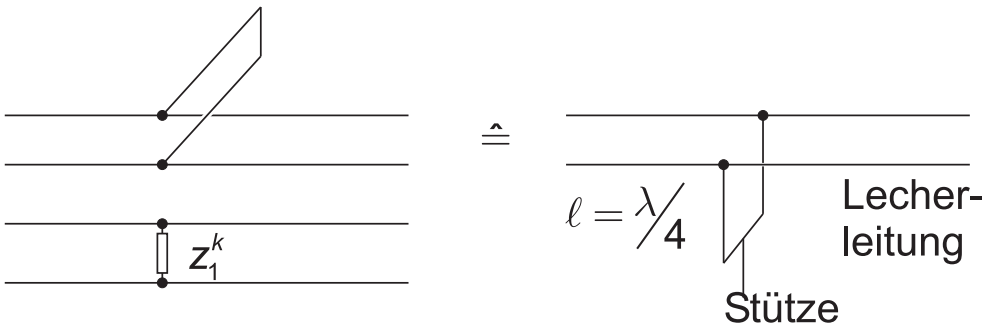



Abbildung 4.144.: Stichleitung als Impedanz und, rechts, Stütze für Lecherleitungen.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

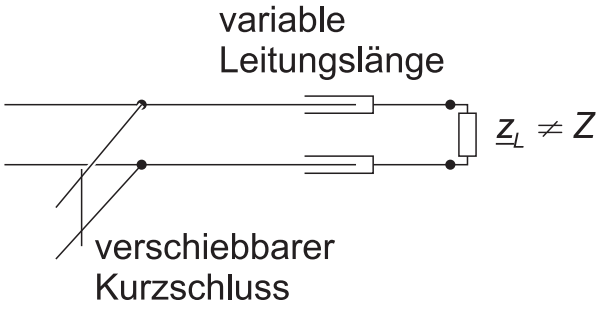



Abbildung 4.145.: Impedanzanpassung mit einer Stichleitung mit verschiebbarem Kurzschluss.

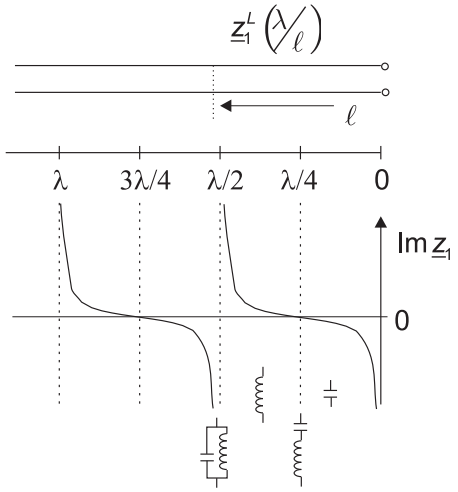



Abbildung 4.146.: Offene Leitung. Unten ist der Imaginärteil der Impedanz gezeichnet.

In Abb. 4.144 links ist gezeigt, dass man mit einer einzelnen Stichleitung eine konzentrierte Impedanz $\underline{z}_1^K$ erzeugen kann. Auf der rechten Seite in dieser Abbildung sieht man, dass eine $\ell = \lambda/4$-Leitung als verlustfreie Stütze für eine Lecherleitung dienen kann.
Wenn eine Leitung der Impedanz $Z$ an einen Verbraucher der Impedanz $\underline{z}_L$ angeschlossen werden muss, dann kann man diese Anpassung erreichen, indem man wie in Abb. 4.145 eine Leitung variabler Länge und eine Stichleitung mit verschiebbarem Kurzschluss verwendet. Eine vollständige Anpassung an den komplexen Verbraucher $\underline{z}_L$ ist mit drei Stichleitungen im Abstand $\lambda/4$ möglich. Dabei muss, anders als in Abb. 4.145 der Abstand zum Verbraucher nicht geändert werden. Die Berechnung erfolgt, indem man die obigen Formeln abschnittsweise anwendet. Alternativ kann man mit einem Smith-Diagram die Aufgabe graphisch lösen.

#### 4.3.1.5. Offene Leitung

Bei einer offenen, verlustlosen Leitung (dann ist $\underline{z}_L(\omega) = \infty$) ist die Eingangsimpedanz mit Gleichung (4.3.16) durch

$$\underline{z}_1^L = -j \cdot Z \cdot \cot \beta\ell = -j \cdot Z \cot \frac{2\pi\ell}{\lambda} \tag{4.3.18}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

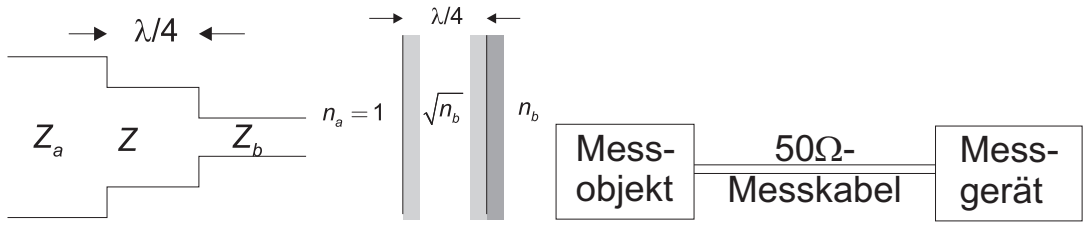



Abbildung 4.147.: Links ist ein Impedanz-Transformator gezeigt. In der Mitte folgt eine Skizze einer vergüteten Linse. Rechts ist die Verwendung eines 50 Ω Messkabels eingezeichnet.

gegeben. Die Schaltung und der *Eingangswiderstand* sind in Abb. 4.146 angegeben. Die Eingangsimpedanz wird zweckmässigerweise als Funktion von $\ell/\lambda$ angegeben. man findet folgendes Verhalten:

| | |
|---|---|
| Leitungen mit $\ell < \lambda/4$: | kapazitives Verhalten |
| Leitungen mit $\ell = \lambda/4$: | Verhalten wie beim Serienschwingkreis |
| Leitungen mit $\lambda/4 < \ell < \lambda/2$: | induktives Verhalten |
| Leitungen mit $\ell = \lambda/2$: | Verhalten wie beim *Parallelschwingkreis* |

Dann wiederholt sich dieses Verhalten.
Wie Abb. 4.147 zeigt, kann man eine vollständige Anpassung einer Leitung mit der Impedanz $Z_a$ an eine Leitung der Impedanz $Z_b$ erreichen, wenn man, nach Gleichung (4.3.16) ein Zwischenstück der Länge $\lambda/4$ mit dem Wellenwiderstand

$$Z = \sqrt{Z_a \cdot Z_b} \tag{4.3.19}$$

einfügt. Ein analoges Beispiel ist die Vergütung von Linsen. Zur Entspiegelung bringt man, wie in Abb. 4.147, Mitte, gezeigt eine Schicht der Dicke $\lambda/4$ mit dem *Brechungsindex* $n = \sqrt{1 \cdot n_b} = \sqrt{n_b}$. Dies ist äquivalent zur Gleichung (4.3.18), da für ebene elektromagnetische Wellen gilt:

$$Z_n = \sqrt{\frac{\mu}{\varepsilon}} \cdot Z_0 \underbrace{=}_{\mu = 1} \frac{Z_0}{n} \tag{4.3.20}$$

Achtung

Weiter kann man aus den obigen Gleichungen ableiten, dass bei hohen Frequenzen 50 Ω-Messkabel stets mit der Nennimpedanz abgeschlossen werden müssen. Wenn zum Beispiel ein 1-Meter-Kabel mit der Dielektrizitätszahl $\varepsilon = 1$ mit einem hochomigen Anschluss (z.B. ein Oszilloskop) verbunden wird, dann liegt am Eingang der Leitung bei etwa 53 MHz ein Kurzschluss vor.
Aus den Maxwellgleichungen (siehe Anhang A.1) sowie den Materialgleichungen für isotrope Materialien

$$\begin{aligned} \rho &= 0 \\ \boldsymbol{j} &= 0 \\ \boldsymbol{D} &= \varepsilon\varepsilon_0 \boldsymbol{E} \\ \boldsymbol{B} &= \mu\mu_0 \boldsymbol{H} \end{aligned}$$

erhält man die skalaren Gleichungen

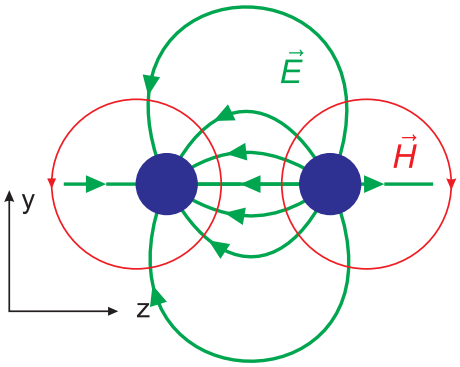



Abbildung 4.148.: Koordinatensystem zur Berechnung der Leitereigenschaften in Zweidrahtleitungen.

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = -\mu\mu_0 \frac{\partial H_x}{\partial t} \tag{4.3.21}$$

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -\mu\mu_0 \frac{\partial H_y}{\partial t} \tag{4.3.22}$$

$$\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y} = -\mu\mu_0 \frac{\partial H_z}{\partial t} \tag{4.3.23}$$

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = \varepsilon\varepsilon_0 \frac{\partial E_x}{\partial t} \tag{4.3.24}$$

$$\frac{\partial H_x}{\partial z} - \frac{\partial H_z}{\partial x} = \varepsilon\varepsilon_0 \frac{\partial E_y}{\partial t} \tag{4.3.25}$$

$$\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} = \varepsilon\varepsilon_0 \frac{\partial E_z}{\partial t} \tag{4.3.26}$$

$$\frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0 \tag{4.3.27}$$

$$\frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z} = 0 \tag{4.3.28}$$

Unter der Annahme, dass sich die Wellen in die x-Richtung ausbreiten (Siehe Abb. 4.148), dass eine TEM-Welle und dass eine homogene Leitung vorliegt (d.h. $\boldsymbol{E}(x,y,z,t) = \underline{\boldsymbol{E}}(y,z)\, e^{-\gamma x + j\omega t} + c.c.$ sowie $\boldsymbol{H}(x,y,z,t) = \underline{\boldsymbol{H}}(y,z)\, e^{-\gamma x + j\omega t} + c.c.$ mit $\gamma = \alpha + j\beta$ der Fortpflanzungskonstanten, wobei $\alpha$ die Dämpfungskonstante und $\beta$ der Wellenvektor ist) reduzieren sich die Gleichungen auf

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

$$\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z} = 0 \tag{4.3.29}$$

$$\gamma \underline{E}_z = -\mu\mu_0 j\omega \underline{H}_y \tag{4.3.30}$$

$$\gamma \underline{E}_y = -\mu\mu_0 j\omega \underline{H}_z \tag{4.3.31}$$

$$\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} = 0 \tag{4.3.32}$$

$$\gamma \underline{H}_z = \varepsilon\varepsilon_0 j\omega \underline{E}_y \tag{4.3.33}$$

$$-\gamma \underline{H}_y = \varepsilon\varepsilon_0 j\omega \underline{E}_z \tag{4.3.34}$$

$$\frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0 \tag{4.3.35}$$

$$\frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z} = 0 \tag{4.3.36}$$

Die Fortpflanzungskonstante $\gamma$ erhalten wir durch die Kombination der Gleichungen (4.3.30), (4.3.31), (4.3.33) und (4.3.34).

$$\gamma^2 = -\mu\mu_0\varepsilon\varepsilon_0\omega^2 \tag{4.3.37}$$

oder

$$\gamma = j\omega\frac{\sqrt{\varepsilon\mu}}{c_0} \tag{4.3.38}$$

wobei $c_0$ die Vakuumlichtgeschwindigkeit ist.
Weiter wird

$$\underline{E}_y(y,z) = \sqrt{\frac{\mu}{\varepsilon}}\sqrt{\frac{\mu_0}{\varepsilon_0}}\underline{H}_z(y,z) \tag{4.3.39}$$

$$\underline{E}_z(y,z) = -\sqrt{\frac{\mu}{\varepsilon}}\sqrt{\frac{\mu_0}{\varepsilon_0}}\underline{H}_y(y,z) \tag{4.3.40}$$

Also ist das Amplitudenverhältnis an einem beliebigen Ort.

$$\frac{|\underline{\boldsymbol{E}}(y,z)|}{|\underline{\boldsymbol{H}}(y,z)|} = \frac{\sqrt{|\underline{E}_y|^2 + |\underline{E}_z|^2}}{\sqrt{|H_y|^2 + |H_z|^2}} = \sqrt{\frac{\mu}{\varepsilon}}\sqrt{\frac{\mu_0}{\varepsilon_0}} = \sqrt{\frac{\mu}{\varepsilon}}Z_0 \tag{4.3.41}$$

mit $Z_0 = \sqrt{\frac{\mu_0}{\varepsilon_0}} = 120\pi\,[\Omega] \approx 377\Omega$ der Wellenwiderstand des Vakuums. Also ist in einer TEM-Welle das Amplitudenverhältnis zwischen elektrischem und magnetischem Feld überall gleich dem einer ebenen Welle!
Das $\boldsymbol{E}$-Feld ist rotationsfrei in der $yz$-Ebene, da die $x$-Komponente

$$(\mathbf{rot}\ E)_x = -\dot{B}_x = 0$$

und da auch $H_x = 0$ ist.
Damit kann man das $\boldsymbol{E}$ -Feld mit einem elektrostatischen Potential $\varphi$ darstellen, also

$$\boldsymbol{E} = -\,\mathbf{grad}\ \varphi = -\boldsymbol{\nabla}\varphi \tag{4.3.42}$$

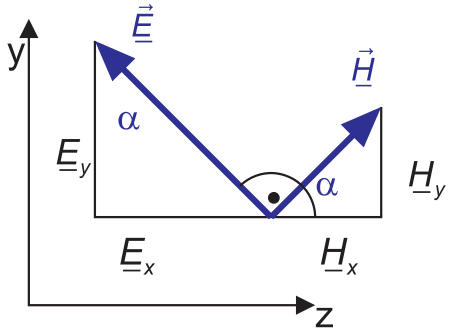



Abbildung 4.149.: Relative Lage von $\boldsymbol{E}$-Feld und $\boldsymbol{H}$-Feld.

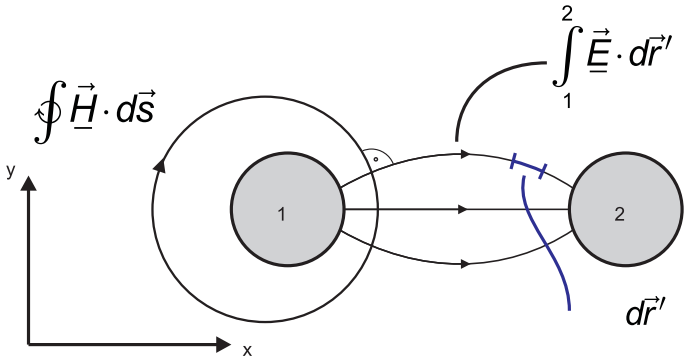



Abbildung 4.150.: Integrationswege um Strom und Spannung zu erhalten.

Zusammen mit $div\, \boldsymbol{E} = \boldsymbol{\nabla} \cdot \boldsymbol{E} = 0$ bekommt man die Potentialgleichung

$$\boldsymbol{\nabla} \cdot \boldsymbol{\nabla} \varphi = 0 = \Delta \varphi \tag{4.3.43}$$

Also ist das $\boldsymbol{E}$ -Feld bei TEM-Wellen einem statischen $\boldsymbol{E}$ -Feld, beschrieben durch die Elektrostatik und die *Potentialtheorie*, äquivalent (Siehe auch Abb. 4.149) .
Das $\boldsymbol{H}$ -Feld steht überall senkrecht zum $\boldsymbol{E}$ -Feld, da

$$\frac{\underline{E}_z\,(y,z)}{\underline{E}_y\,(y,z)} = -\frac{\underline{H}_z\,(y,z)}{\underline{H}_y\,(y,z)} \tag{4.3.44}$$

Jeder der beiden Koeffizienten kann mit $\tan\alpha$ dargestellt werden. Da die $\boldsymbol{H}$-Feldlinien stets senkrecht zu den $\boldsymbol{E}$-Feldlinien stehen, verlaufen sie entlang der *Potentiallinien* des elektrostatischen Potentials.
Wir gehen nun zu integralen Grössen über (Die Integrationswege sind in 4.150 gezeigt). Es liegt, da aus $E_x = 0$ auch $\dot{D}_x = 0$ folgt, ein reiner Leitungsstrom vor. An der Stelle $x$ ist er

$$\underline{i}\,(x,\omega) = \int\limits_{\substack{Leitungs-\\querschnitt}} \underline{i}_x\,(y,z,\omega)\boldsymbol{n}da \;=\; \oint \underline{\boldsymbol{H}}\,(y,z)\cdot d\boldsymbol{s} \tag{4.3.45}$$

Für die Spannung folgt

$$\underline{u}\,(x,\omega) = \int_1^2 \underline{\boldsymbol{E}}\,(y,z)d\boldsymbol{r}\,' \tag{4.3.46}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

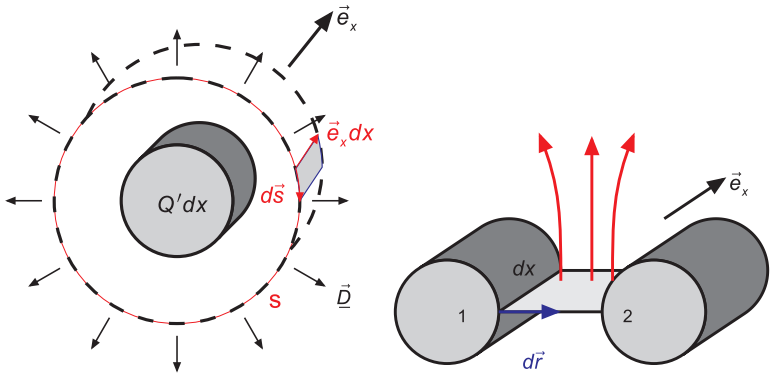



Abbildung 4.151.: Berechnung des Kapazitätsbelags (links) und des Induktivitätsbelags (rechts).

Das Vierpol-Ersatzschaltbild eines verlustfreien Leiterstückes wird durch einen Kapazitätsbelag $C'$ und einen Induktivitätsbelag $L'$ charakterisiert. Der Wellenwiderstand ist dann

$$Z = \sqrt{\frac{L'}{C'}} \tag{4.3.47}$$

und die Phasengeschwindigkeit

$$c_{ph} = \frac{1}{\sqrt{L' \cdot C'}} \tag{4.3.48}$$

Zur Berechnung des Kapazitätsbelages (siehe Abb. 4.151, links) ermitteln wir auf einem Leiterstück der Länge $dx$ die Ladung $Q = Q'dx$. Aus den Maxwellschen Gleichungen in Integralform

$$\iiint\limits_{V(A)} \rho dV = \iiint\limits_{V(A)} \mathbf{div}\ \boldsymbol{D} dV = \iint\limits_{A} \boldsymbol{D} \cdot d\boldsymbol{a}$$

Das erste Integral ergibt (siehe Abb. 4.151, links)

$$Q'dx = \iiint\limits_{V(A)} \rho dV$$

Das Oberflächenintegral über $A$ kann in drei Teile aufgeteilt werden. Die zwei Zylinderendflächen tragen nichts zum Integral bei, da dort $\boldsymbol{D}$ tangential zur Oberfläche ist. Allein die Zylindermantelfläche trägt zum Integral bei. Das in Abb. 4.151, links, eingezeichnete Flächenelement hat die Seiten $\boldsymbol{e}_x dx$ und $d\boldsymbol{s}$, wobei $d\boldsymbol{s}$ das Linienelement entlang des Zylinderumfangs ist. Die Richtung von $d\boldsymbol{s}$ folgt mit der rechten-Hand-Regel aus der Richtung von $\boldsymbol{e}_x$. Der Normalenvektor auf das Flächenelement steht senkrecht auf $d\boldsymbol{s}$ und $\boldsymbol{e}_x$. Also ist der Normalenvektor

$$\boldsymbol{n} = \frac{d\boldsymbol{s}}{ds} \times \boldsymbol{e}_x$$

Die Fläche ist $da = dsdx$. Also erhalten wir

$$\begin{aligned} Q'dx &= \iint_A \underline{\boldsymbol{D}} \cdot \boldsymbol{n} da & &= \iint_A \underline{\boldsymbol{D}} \cdot \left( \frac{d\boldsymbol{s}}{ds} \times \boldsymbol{e}_x \right) ds dx \\ &= \iint_A \underline{\boldsymbol{D}} \cdot (d\boldsymbol{s} \times \boldsymbol{e}_x)\, dx & &= \iint_A \boldsymbol{e}_x \cdot (\underline{\boldsymbol{D}} \times d\boldsymbol{s})\, dx \\ &= \iint_A (\underline{\boldsymbol{D}} \times d\boldsymbol{s}) \cdot \boldsymbol{e}_x dx & &= \left( \oint_s \underline{\boldsymbol{D}} \times d\boldsymbol{s} \right) \cdot \boldsymbol{e}_x dx \end{aligned} \tag{4.3.49}$$

Mit $C' \cdot dx = Q' \cdot dx / \underline{u}$ wird die Kapazität

$$C' = \varepsilon\varepsilon_0 \frac{\oint (\underline{\boldsymbol{E}} \times d\boldsymbol{s}) \cdot \boldsymbol{e}_x}{\int_1^2 \underline{\boldsymbol{E}} \cdot d\boldsymbol{r}\,'} = \frac{\sqrt{\varepsilon\mu}}{c_0} \frac{\oint \underline{\boldsymbol{H}} \cdot d\boldsymbol{s}}{\int_1^2 \underline{\boldsymbol{E}} \cdot d\boldsymbol{r}\,'} \tag{4.3.50}$$

da $\underline{\boldsymbol{E}} = \sqrt{\frac{\mu\mu_0}{\varepsilon\varepsilon_0}} \{\underline{\boldsymbol{H}} \times \boldsymbol{e}_x\}$ ist. Weiter haben wir die Vektoridentität

$$(\boldsymbol{a} \times \boldsymbol{b}) \times \boldsymbol{c} = \boldsymbol{b}\,(\boldsymbol{a} \cdot \boldsymbol{c}) - \boldsymbol{a}\,(\boldsymbol{b} \cdot \boldsymbol{c})$$

verwendet. Wir setzen $\boldsymbol{a} \equiv \underline{\boldsymbol{H}}$, $\boldsymbol{b} \equiv \boldsymbol{e}_x$ und $\boldsymbol{c} \equiv d\boldsymbol{s}$ und erhalten

$$(\underline{\boldsymbol{H}} \times \boldsymbol{e}_x) \times d\boldsymbol{s} = \boldsymbol{e}_x\,(\underline{\boldsymbol{H}} \cdot d\boldsymbol{s})$$

da $\boldsymbol{e}_x \perp d\boldsymbol{s}$ ist.
Um den Induktivitätsbelag zu ermitteln (siehe Abb. 4.151, rechts), berechnet man den Fluss $\phi$ zwischen der Stelle $x$ und $x + dx$

$$\phi = \iint_A \boldsymbol{B} \cdot \boldsymbol{n} da \tag{4.3.51}$$

wobei über eine fiktive Fläche $A$ zwischen den beiden Leitern integriert wurde. Das Flächenelement hat die Grösse $da = dxdr$. Der Normalenvektor wird durch

$$\boldsymbol{n} = \frac{d\boldsymbol{r}}{dr} \times \boldsymbol{e}_x$$

gegeben. Also ist

$$\begin{aligned} \phi &= \iint_A \boldsymbol{B} \cdot \boldsymbol{n} da & &= \iint_A \boldsymbol{B} \cdot \left( \frac{d\boldsymbol{r}}{dr} \times \boldsymbol{e}_x \right) dr\, dx \\ &= \iint_A \boldsymbol{B} \cdot (d\boldsymbol{r} \times \boldsymbol{e}_x)\, dx & &= \left( \int_1^2 \boldsymbol{B} \cdot (d\boldsymbol{r} \times \boldsymbol{e}_x) \right) dx \\ &= \left( \int_1^2 \boldsymbol{e}_x\,(\boldsymbol{B} \times d\boldsymbol{r}) \right) dx & &= \left( \int_1^2 \boldsymbol{B} \times d\boldsymbol{r} \right) \boldsymbol{e}_x dx \end{aligned} \tag{4.3.52}$$

Da die Induktivität L längs der Strecke $dx$ gegeben ist durch $L = L' \cdot dx = \underline{\phi}(x, w) \cdot dx / \underline{i}(x, \omega)$ folgt für den Induktivitätsbelag

$$L' = \mu\mu_0 \frac{\int_1^2 (\underline{\boldsymbol{H}} \times d\boldsymbol{r}) \cdot \boldsymbol{e}_x}{\oint \underline{\boldsymbol{H}} \cdot d\boldsymbol{s}} = \frac{\sqrt{\mu\varepsilon}}{c_0} \cdot \frac{\int_1^2 \underline{\boldsymbol{E}} \cdot d\boldsymbol{r}}{\oint \underline{\boldsymbol{H}} \cdot d\boldsymbol{s}} \tag{4.3.53}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

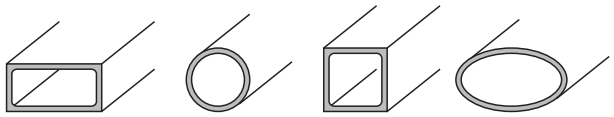


Abbildung 4.152.: Bauformen von Wellenleitern.

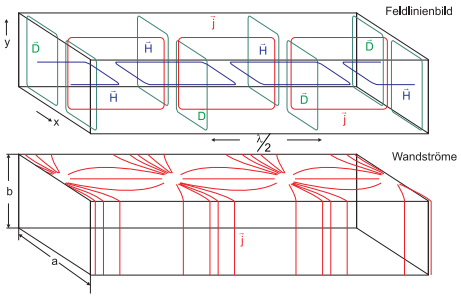



Abbildung 4.153.: Feldlinienbild in rechteckförmigen Wellenleitern.

weil $\underline{\boldsymbol{H}} = \sqrt{\frac{\varepsilon\varepsilon_0}{\mu\mu_0}}\,\{\boldsymbol{e}_x \times \underline{\boldsymbol{E}}\}$ ist.
Für den Leitungswellenwiderstand erhalten wir

$$Z = \sqrt{\frac{L'}{C'}} = \frac{\sqrt{L'C'}}{C'} = \frac{\sqrt{\varepsilon\mu\varepsilon_0\mu_0}}{\varepsilon_0\mu_0} \frac{\int_1^2 \boldsymbol{E}\cdot d\boldsymbol{r}}{\oint (\underline{\boldsymbol{E}} \times d\boldsymbol{s})\cdot \boldsymbol{e}_x}$$
$$= Z_0\sqrt{\frac{\mu}{\varepsilon}} \cdot \frac{\int_1^2 \boldsymbol{E}\cdot d\boldsymbol{r}}{\oint (\boldsymbol{E} \times d\boldsymbol{s})\cdot \boldsymbol{e}_x} \tag{4.3.54}$$

Weiter ist die Phasengeschwindigkeit

$$c_{ph} = \frac{1}{\sqrt{L'C'}} = \frac{C_0}{\sqrt{\varepsilon\mu}} \tag{4.3.55}$$

## 4.3.2. Elektrische Leitungen bei hohen Frequenzen

Die Leitung von elektromagnetischen Wellen bei hohen Frequenzen ist mit gewöhnlichen Kabeln nicht mehr möglich. Üblich für die niedrigeren Frequenzen sind Hohlleiter und für die ganz hohen Frequenzen Streifenleiter.
Einige Bauformen sind in der Abb. 4.152 dargestellt. Prinzipiell sind alle Bauformen möglich. In der Praxis werden jedoch nur

- Rechteck-Hohlleiter mit einem Seitenverhältnis von 2:1 und
- Rundhohlleiter verwendet.

Die letzteren werden aber nur für sehr spezielle Leitungsprobleme verwendet, unter anderem da durch ihre hohe Symmetrie keine Polarisationserhaltung garantiert ist. Aus den Maxwell-Gleichungen und den üblichen Randbedingungen lassen sich leicht die möglichen Wellenleitermoden bestimmen. Da die $\boldsymbol{E}$-Felder an den Wänden eine Knotenlinie haben müssen, muss transversal mindestens eine halbe Wellenlänge in den Hohlleiter passen. Deshalb gibt es eine **untere Grenzfrequenz**,

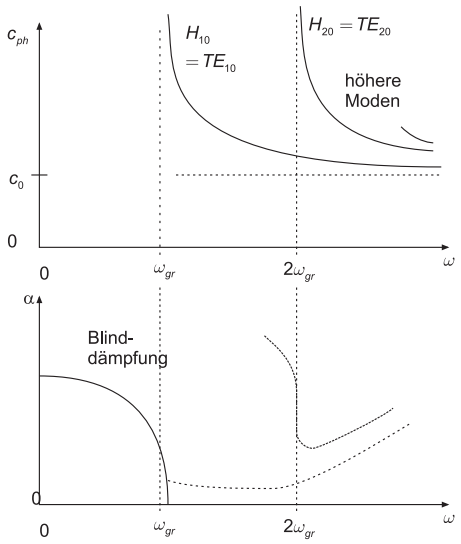



Abbildung 4.154.: Phasengeschwindigkeit und Dämpfung im rechteckförmigen Wellenleitern.

unter der eine Wellenführung nicht möglich ist[11]. Die unterste Mode (die mit der längsten Wellenlänge) in einem rechteckförmigen Wellenleiter wird die **$TE_{10}$-Mode** oder auch die **$H_{10}$-Mode** genannt. Die erste Bezeichnung stammt daher, dass das $\boldsymbol{E}$-Feld senkrecht zur Ausbreitungsrichtung steht, also transversal ist. Die zweite Bezeichnung besagt, dass das $\boldsymbol{H}$-Feld eine **longitudinale** Komponente hat, eine Komponente die nur in einer geführten Welle existieren kann. Bei beiden gibt der erste Index die Zahl der halben Sinusbögen über der längeren Seite an, der zweite die Zahl der halben Sinusbögen über der kürzeren Seite. Abb. 4.153 zeigt die dazugehörigen Feldlinienbilder.

Durch die Wechselwirkung mit den Wänden ist die Wellenausbreitung in einem Wellenleiter dispersiv. Abb. 4.154 zeigt die Phasengeschwindigkeit als Funktion der Wellenlänge. Unterhalb der unteren Grenzfrequenz $\omega_{gr}$, bei der $\lambda_0/2 = a$ ist, gibt es keine Wellenausbreitung. Die Amplitude wird exponentiell gedämpft. Oberhalb der der Grenzfrequenz wird die Wellenausbreitung durch die Wandströme und deren resistive Verluste gedämpft. Die Verluste sind für die Grundmode $TE_{10}$ minimal. Die beste Transmission erreicht man mit supraleitenden Wellenleitern. Ab $2\omega_{gr}$ wird die $TE_{20}$-Mode auch geführt. Typischerweise verwendet man bei Wellenleitern nicht den ganzen möglichen Bereich, in dem nur die Grundmode geführt wird, sondern nur $1.25 \cdot \omega_{gr} \ldots 1.9 \cdot \omega_{gr}$. Die untere Grenzfrequenz rührt von den divergierenden Verlusten für $\omega \to \omega_{gr}^{+}$ her, die obere von der Tatsache, dass auch unterhalb von $2\omega_{gr}$ die $TE_{20}$-Mode eine merkbare Amplitude bekommt. Weiter ist in diesem Bereich die Änderung der Phasengeschwindigkeit minimal, die Dispersion also gering[12].

[11] Beim Koaxialkabel löst der zentrale Wellenleiter das bei den Hohlleitern vorhandene Problem, dass $\boldsymbol{\nabla E} = 0$ sein muss

[12] Dies ist eine Voraussetzung für die Übertragung hoher Frequenzen über lange Distanzen

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

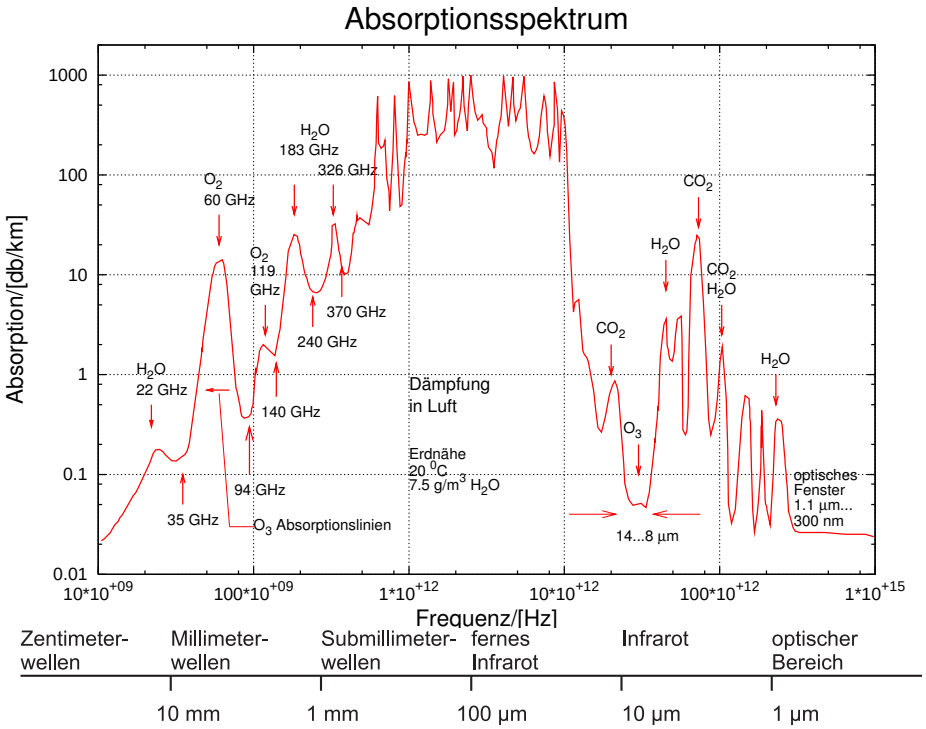



Abbildung 4.155.: Dämpfung elektromagnetischer Wellen in Luft.

Für die Wellenlänge $\lambda_L$ im Hohlleiter gilt die Formel

$$\lambda_L = \frac{\lambda}{\sqrt{1 - \left(\frac{\lambda}{\lambda_{gr}}\right)^2}} \tag{4.3.56}$$

wobei $\lambda_{gr}$ die zur Grenzfrequenz $\omega_{gr}$ gehörige Wellenlänge ist.
Elektromagnetische Wellen werden einerseits durch die in den Hohlleitern vorhandene Luft, andererseits aber auch durch die Absorption in den Metallwänden der Hohlleiter oder in den Metallstreifen der Streifenleiter gegeben. Abb. 4.155 zeigt das Absorptionsspektrum in Luft. Bei tiefen Frequenzen ist es vor allem die Absorption durch Wasser und durch Sauerstoff (mit magnetischem Dipolmoment!), die dominiert. Im Infrarotbereich kommt die Absorption durch $CO_2$ hinzu. Zwischen 14 und 8 $\mu$m ist ein Absorptionsfenster, wie auch zwischen 1100 und 300 nm.
Die Absorption steigt in Metall-Hohlleitern stark mit der Frequenz an. Zwei Gründe gibt es:

- Die effektive Leiterdicke nimmt wegen des Skineffektes mit $1/\sqrt{\omega}$ ab. Dadurch wird die Wellenleitereigenschaft des Metalls schlechter (Siehe Abb. 4.156).

- Um auch bei höheren Frequenzen als einzige geführte Mode nur die Grundmode zu haben, muss die lineare Dimension des Wellenleiters proportional zur Wellenlänge sein. Da dadurch die zur Führung beitragende Oberfläche des Wellenleiters proportional zu $1/\omega$ ist, nimmt die Dämpfung entsprechend zu.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

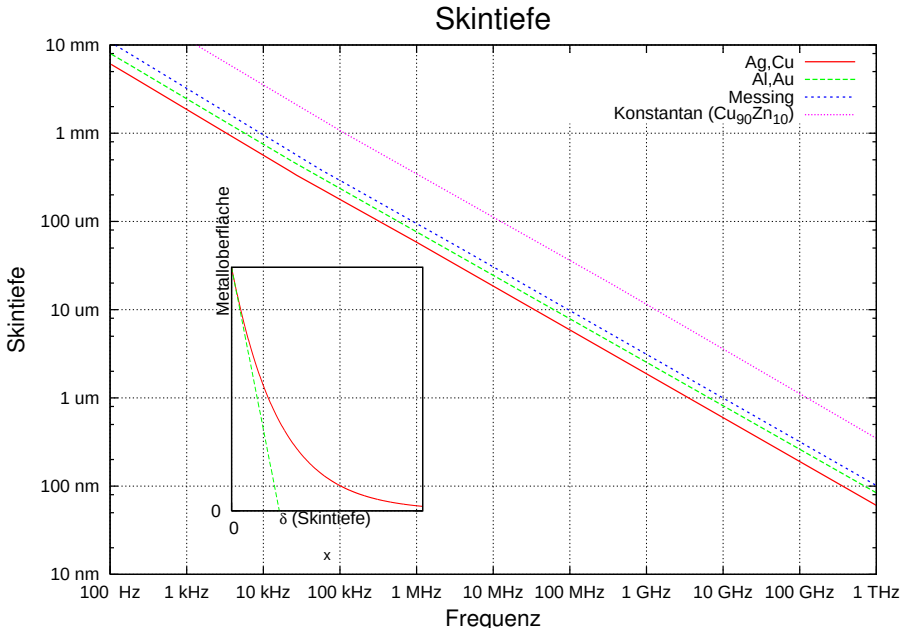



Abbildung 4.156.: Skintiefe elektromagnetischer Wellen.

Insgesamt ergibt sich eine zu $\omega^{3/2}$ proportionale Dämpfung[13].

Abb. 4.157 zeigt die Dämpfungseigenschaften für Wellenleiter. Das für viele physikalische Experimente wichtigste Frequenzband ist das X-Band zwischen 8.2 und $12.4GHz$.

Für höhere Frequenzen verwendet man oft Streifenleiter. Sie können mit photolithographischen Verfahren hergestellt werden, sind also in kleinen Abmessungen wesentlich präziser herzustellen als die Hohlwellenleiter. Die Dämpfung ist im Allgemeinen höher bei Streifenleitern als bei Hohlleitern. Andererseits hat dies, zum Beispiel bei der Mobilkommunikation, wegen den geringen Grössen der Geräte, kaum einen Einfluss. Zusammen mit SMD-Bauteilen (Surface Mounted Device) lassen sich sehr effizient Mikrowellenschaltungen industriell herstellen. Der Streifenleiter in der Abb. 4.158 hat die folgenden Eigenschaften:

- Es gibt zwei benachbarte Leiteroberflächen, die unterschiedliche Ladungen tragen, also den Anfang und das Ende von $\boldsymbol{D}$-Feldlinien darstellen.

- Es gibt Grenzflächen zwischen verschiedenen Dielektrika, die von Feldlinien durchsetzt sind.

Im Gegensatz zu Hohlleitern gibt es bei Streifenleitern sowohl beim $\boldsymbol{H}$-Feld als auch beim $\boldsymbol{E}$-Feld longitudinale Komponenten. Dies rührt daher, dass an Grenzflächen neben der Normalkomponente von $\boldsymbol{D}$ auch die Tangentialkomponente von $\boldsymbol{E}$ stetig sein muss. Man spricht deshalb von **Quasi-TEM**-Moden oder von Hybrid-Moden.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

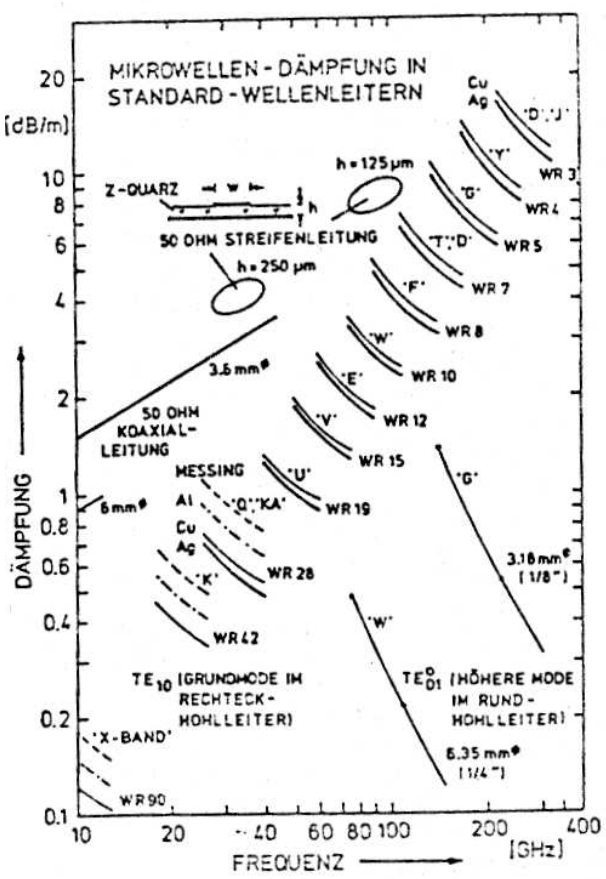



Abbildung 4.157.: Dämpfungseigenschaften von Wellenleitern.

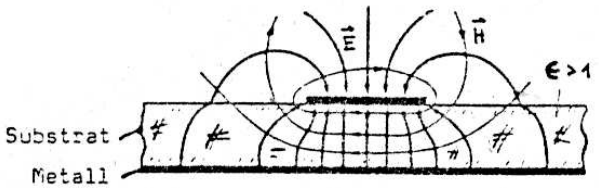



Abbildung 4.158.: Streifenleiter.

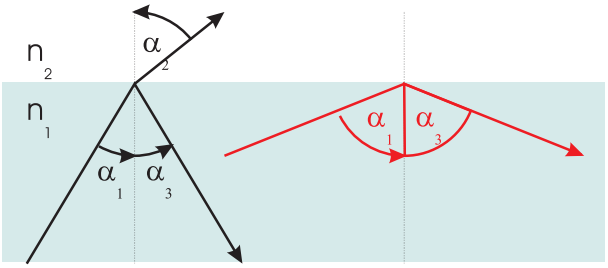



Abbildung 4.159.: Brechungsgesetz von Snellius.

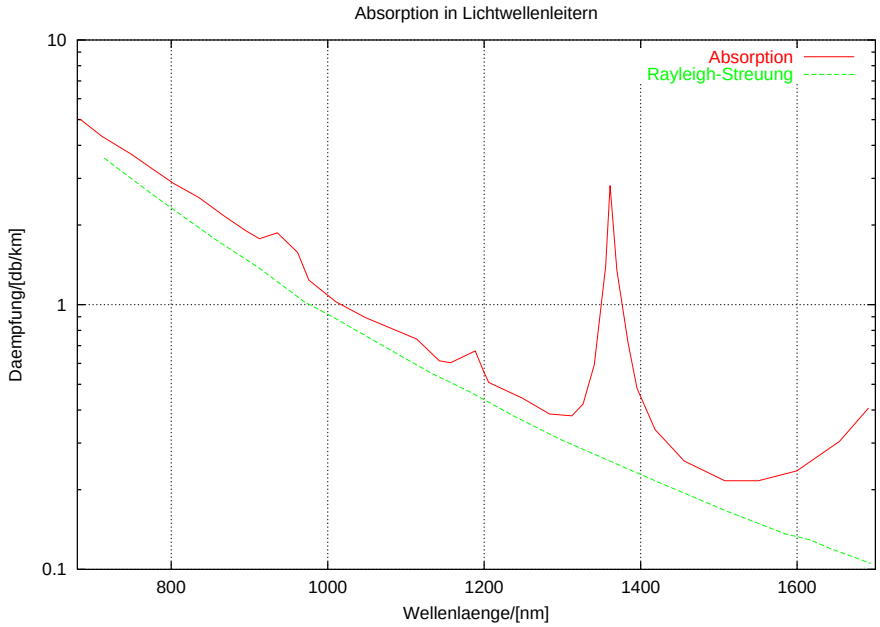



Abbildung 4.160.: Verlauf der Dämpfung in einem Lichtwellenleiter als Funktion der Wellenlänge $\lambda$.

## 4.3.3. Optische Leitungen

Bei dielektrischen Wellenleitern, zu denen auch optische Fasern gehören, werden Wellen im Medium mit dem grössten *Brechungsindex* geführt. Die Ausbreitungsgeschwindigkeit hängt von der Brechzahl ab, also $c_n = c_0/n$. Beim Übergang vom optisch dichteren Medium mit $n_1$ nach dem optisch dünneren Medium mit $n_2 < n_1$ kann bei flachem Einfall Totalreflexion auftreten. Das Snellius'sche Brechungsgesetz (siehe auch Abb. 4.159) lautet

$$n_1 \cdot \sin\alpha_1 = n_2 \cdot \sin\alpha_2 \qquad \text{für den gebrochenen Strahl bei } \alpha_1 < \alpha_0 \qquad (4.3.57)$$

$$\alpha_1 = \alpha_3 \qquad \text{für jeden gespiegelten Strahl} \qquad (4.3.58)$$

$$\alpha_0 = \arcsin(\frac{n_2}{n_1}) \qquad \text{Grenzwinkel der Totalreflexion} \qquad (4.3.59)$$

für den gebrochenen sowie den reflektierten Strahl. Nach diesem Prinzip wird Licht in Multimoden-Wellenleitern geführt. Die unterschiedlichen Brechungsindizes werden mit Dotierstoffen erreicht. So vergrössert, zum Beispiel, eine Dotierung mit Ge den *Brechungsindex*. Eine Dotierung mit F verringert ihn.

### 4.3.3.1. Verlustmechanismen in Wellenleitern

Optische Wellenleiter werden in der Kommunikationstechnik vor allem wegen ihren geringen Verlusten[14] (für eine Übersicht über den Dämpfungsverlauf siehe Abb. 4.160) und ihrer geringen Anfälligkeit auf externe Störungen. An der Universität

[13] Die $TE_{01}^0$-Mode, eine höhere Mode des Rundwellenleiters hat mit ihren konzentrischen $\boldsymbol{E}$-Feldlinien die kleinste bekannte Dämpfung, da bei ihr die Wandströme minimal sind.

[14] Bei Kupferkabeln braucht man alle 2 km einen Verstärker, bei Lichtwellenleitern kann der Abstand zwischen den Verstärkern über 100 km betragen

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

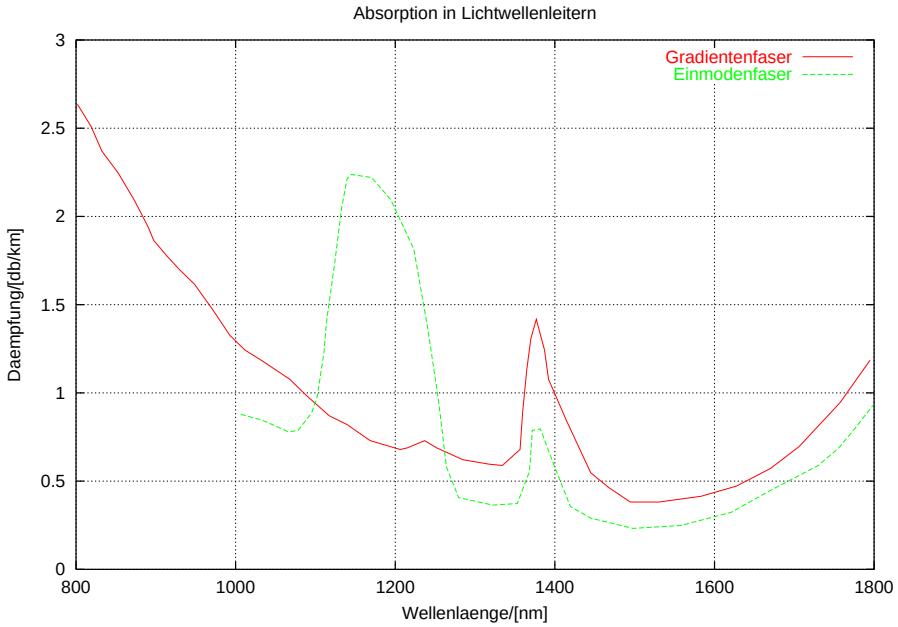



Abbildung 4.161.: Spektrale Dämpfungswerte von verschiedenen Glasfasertypen.

Ulm hat ein Wellenleiter, der mitten durch einen Brand führte, während dem Brand anstandslos Daten übertragen!
Die Dämpfung bewirkt einen exponetiellen Abfall der übertragenen Leistung. Üblicherweise wird die Dämpfung $\alpha$ in dB (deziBel) angegeben. Der Leistungsabfall ist also

$$P(z) = P_0 \cdot 10^{\alpha_z [dB]/10} \tag{4.3.60}$$

wobei $P_0$ die Leistung am Eingang der Faser ist und $\alpha$ der Dämpfungsfaktor in dB. Die Dämpfung setzt sich aus drei Komponenten zusammen:

- der Streuung
- der Absorption
- der Biegedämpfung

zusammen. Die Biegedämpfung rührt von der Krümmung der Glasfaser her und ist, zumindestens für Licht das über Totalreflexion geleitet wird, einfach zu verstehen. Wird das Kabel in einer zu engen Schleife gelegt, so ist die Bedingung der Totalreflexion nicht mehr erfüllt und die Verluste steigen. Zusätzlich zu dieser Makrodämpfung kommt die Mikrodämpfung, deren Ursache Spannungen in der Faser und Schwankungen in der Materialzusammensetzung, zum Beispiel durch eine nicht konstante Dichte der Dotierstoffe, sind.
Die Absorption hängt von der Reinheit des Materials ab. Insbesondere störend ist die $OH^-$-Bande bei 1380 nm und bei 1240 nm sowie die Infrarotabsorption über 1600 nm.
Der hauptsächliche Dämpfungsmechanismus ist jedoch die Rayleigh-Streuung, die bis zu 95% der gesamten Dämpfung ausmacht. Sie rührt daher, dass im Glas

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

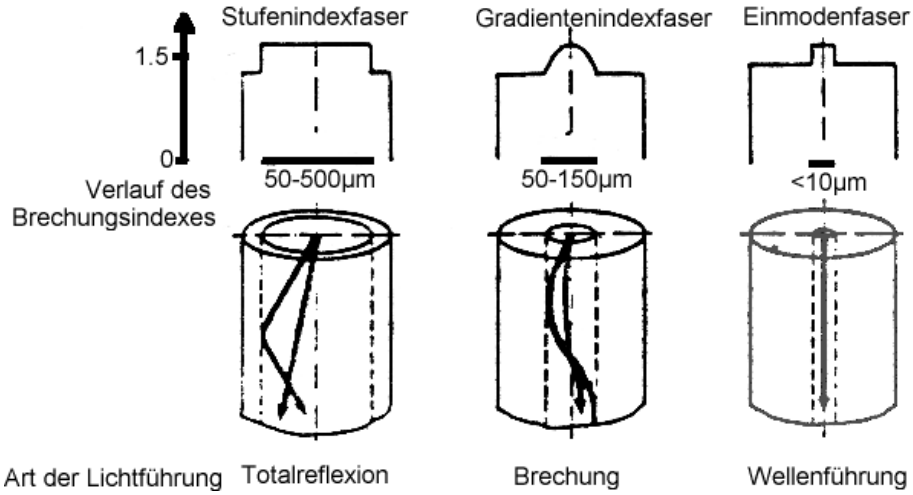



Abbildung 4.162.: Optische Wellenleiter (Glasfasern) werden in drei Kategorien eingeteilt: Stufenindexfasern, Gradientenindexfasern und Einmodenfasern.

mikroskopische Dichteschwankungen existieren, die sich aus Gründen der Thermodynamik auch nicht komplett eliminieren lassen, sowie wegen der notwendigen Dotierstoffe.
Abb. 4.161 zeigt Dämpfungsspektren von Gradientenfasern und von Einmodenfasern. Die Einmodenfaser (auch Monomode-Faser genannt) hat eine geringere Dämpfung, da ihr Kern weniger dotiert werden muss. Der starke anstieg der Dämpfung unter 1250 nm Wellenlänge rührt daher, dass die Faser für kurze Wellenlängen nicht mehr einmodig ist, dass also die Führungseigenschaften nicht mehr so perfekt sind. Infrarotabsorption, Rayleigh-Streuung und Wasserabsorption ($OH^-$-Absorption) sind bei beiden Typen zu erkennen.

#### 4.3.3.2. Typen von optischen Wellenleitern

Es sind drei Typen von optischen Wellenleitern üblich (siehe auch Abb. 4.162)

- Stufenindexfasern
- Gradientenindexfasern
- Einmodenfasern

Bei allen dreien kann das Indexprofil mit

$$n^2(r) = n_1^2 \left(1 - 2\Delta \cdot f\left(\frac{r}{a}\right)\right) \tag{4.3.61}$$

angegeben werden. Dabei ist $a$ der Kernradius, $r$ der Abstand vom Fasermittelpunkt und $n_1$ der *Brechungsindex* im Kern. $\Delta$ ist die relative Brechzahldifferenz zwischen Kern und Mantel. Ausserhalb des Kerns, also für $r > a$ hat man

$$n^2(r) = n_1^2 \left(1 - 2\Delta\right) \tag{4.3.62}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

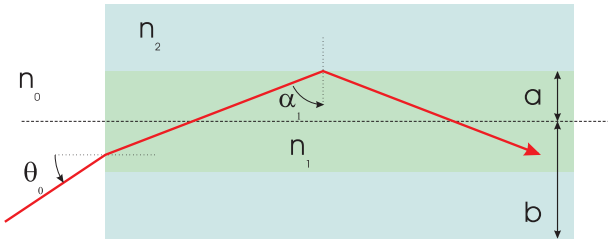



Abbildung 4.163.: Geometrie zur Berechnung von Stufenindexfasern.

Diese allgemeine Funktion, die auch über den ganzen Faserquerschnitt gilt, ist für eine Stufenindexfaser

$$f\left(\frac{r}{a}\right) = \begin{cases} 0 & \text{für} \quad r < a \\ 1 & \text{für} \quad r \geq a \end{cases} \tag{4.3.63}$$

**4.3.3.2.1. Stufenindexfaser** In der Stufenindexfaser werden alle Lichtstrahlen, für die $\sin\alpha_1 \geq N_2/n_1$ gilt total reflektiert, das heisst geführt[?]. Diese aus Abb. 4.163 ablesbare Bedingung kann auch als

$$\cos\alpha_1 \leq \frac{\sqrt{n_1^2 - n_2^2}}{n_1} \tag{4.3.64}$$

geschrieben werden. Strahlt man aus der Umgebung mit dem *Brechungsindex* $n_0$ Licht unter dem Winkel $\Theta_0$ ein, so gilt an der Eintrittsfacette $n_0 \sin\Theta_0 = n_1 \cos\alpha_1$. Daraus folgt $n_0 \sin\Theta_0 \leq \sqrt{n_1^2 - n_2^2}$ und somit

$$\sin\Theta_0 \leq \frac{\sqrt{n_1^2 - n_2^2}}{n_0} \tag{4.3.65}$$

Damit werden alle Lichtstrahlen, für die die obige Bedingung gilt, geführt. Diese Bedingung ist aber auch äquivalent zur Definition der Numerischen Apertur eines Objektivs. Also sagt man, dass

$$N.A. \equiv \sqrt{n_1^2 - n_2^2} \tag{4.3.66}$$

sei die numerische Apertur der Faser. Der maximale Wert von $\Theta_0$ heisst der Akzeptanzwinkel und ist

$$\Theta_0 = \arcsin\frac{\sqrt{n_1^2 - n_2^2}}{n_0} \tag{4.3.67}$$

Für einen *Brechungsindex* des Kerns $n_1 = 1.57$ und einen *Brechungsindex* des Mantels $n_2 = 1.51$ in Luft ($n_0 = 1$) ergibt sich $N.A. = 0.43$ und damit der Akzeptanzwinkel $\Theta_0 = 25.5^0$.
Wenn es darum geht, Licht aus einem räumlich eng begrenzten Gebiet mit einer relativ hohen numerischen Apertur zu sammeln, ohne dass eine Abbildung gewünscht wird, kann man vielfach anstelle von Linsen Fasern mit ähnlichen oder sogar grösseren numerischen Aperturen verwenden. Mit einigen Fasern lässt sich so sehr viel effizienter emittiertes Licht sammeln als mit einer einzelnen Linse.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

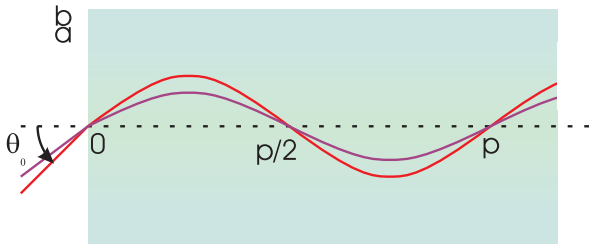



Abbildung 4.164.: Geometrie zur Berechnung von Gradientenindexfasern.

**4.3.3.2.2. Gradientenindexfaser** In Gradientenindexfasern gilt für die Indexfunktion nach Gleichung (4.3.61) analog zu Gleichung (4.3.63)

$$f\left(\frac{r}{a}\right) = \begin{cases} \left(\frac{r}{a}\right)^{\alpha} & \text{für} \quad r < a \\ 1 & \text{für} \quad r \geq a \end{cases} \tag{4.3.68}$$

Dabei ist $r$ der Abstand vom Kernzentrum. Der Unterschied im *Brechungsindex* in Gleichung (4.3.61) ist meistens klein, d.h. $\Delta \ll 1$. Der Exponent ist andererseits häufig $\alpha = 2$. Also kann für den *Brechungsindex* als Funktion der Position näherungsweise angenommen werden

$$n(r) \approx n_1 \left(1 - \Delta \frac{r^2}{a^2}\right) = n_1 \left(1 - \frac{r^2}{2\rho^2}\right) \tag{4.3.69}$$

wobei $\rho \equiv a/\sqrt{2\Delta}$ ist. Zur Berechnung der Bahnkurve nehmen wir an, dass $n_1$ entlang der Faser nicht variiert.
Um den Lichtweg, wie er in Abb. 4.164 angegeben ist, zu berechnen, gehen wir nach Pérez[**?**] von der vektoriellen Gleichung

$$\frac{d}{ds}(n\boldsymbol{u}) = \boldsymbol{\nabla} n \tag{4.3.70}$$

Mit der Gaussschen Näherung $ds \sim dz$ erhält man für eine radiale Achse $x$

$$\frac{d}{dz}\left(n\frac{dx}{dz}\right) = \frac{\partial n}{\partial x} \tag{4.3.71}$$

Mit

$$\frac{\partial n}{\partial x} = \frac{dn}{dr}\frac{\partial r}{\partial x} = \left(-\frac{n_1 4}{\rho^2}\right)\frac{x}{r} = -\frac{N_1 x}{\rho^2}$$

und da $n_1$ nicht entlang der Faser ($z$!) nicht ändert, bekommt man

$$n\frac{d^2x}{dz^2} + n_1\frac{x}{\rho^2} = 0 \tag{4.3.72}$$

Da bei den meisten Glasfasern $\Delta \ll 1$ ist ist auch $n \approx n_1$. Deshalb erhält man schliesslich für die Differntialgleichung des Lichtweges

$$\frac{d^2x}{dz^2} + \frac{x}{\rho^2} = 0 \tag{4.3.73}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Der Lichtweg durch eine Gradientenindexfaser mit paraboloidem Indexprofil wird durch eine der Schwingungsgleichung ähnliche Gleichung beschrieben. Die allgemeine Lösung ist also

$$x = A\cos\left(\frac{z}{\rho}\right) + B\sin\left(\frac{z}{\rho}\right) \tag{4.3.74}$$

Zur Berechnung der Schwingungsform nehmen wir an, dass ein Lichtstrahl im Abstand $x_e$ von der Faserachse mit der Steigung $x'_e = dx_e/dz$ in die Faser eintritt. Dann haben wir im Innern der Faser

$$x_e = x(0) = A \tag{4.3.75}$$

$$(n_1 x_e^{'}) = \frac{n_1}{\rho}\left[-A\sin\left(\frac{z}{\rho}\right) + B\cos\left(\frac{z}{\rho}\right)\right]_{z=0} = \frac{n_1 B}{\rho} \tag{4.3.76}$$

Nach Pérez[**?**] kann damit die optische Transfermatrix bestimmt werden. Die Lösung für den speziellen Fall $x_e = 0$ und $n_1 x'_e \neq 0$ ist

$$x(z) = \rho x'_e \sin\left(\frac{z}{\rho}\right) = \frac{p x'_e}{2\pi}\sin\left(2\pi\frac{z}{p}\right) \tag{4.3.77}$$

wobei $p = 2\pi\rho$ gesetzt wurde. Man ersieht aus Gleichung (4.3.77), dass alle Lichtstrahlen in Achsennähe sich periodisch im Abstand $p$ schneiden[15].
Die numerische Apertur einer Gradientenfaser ist eine Funktion von $a$, $n_1$ und $\rho = a/\sqrt{2\Delta}$, also von der maximalen Differenz des Brechungsindexes $\Delta$. An der Eintrittsfläche haben wir

$$n_0 \left|\sin\Theta_0\right| = \left|n_1 x'_e\right|_{x_e=0}$$

Die Bedingung, dass ein Lichtstrahl nicht aus dem Kernbereich herausläuft, also dass $|x(z)| < a \quad \forall z$ ist wegen $|x| = |\rho x'_e \sin(z/\rho)| \leq |\rho x'_e|$ immer dann erfüllt, wenn $|\rho x'_e| \leq a$ ist. Dann gilt

$$n_0 \left|\sin\Theta_0\right| \leq \frac{n_1 a}{\rho} = n_1\sqrt{2\Delta} = N.A. \tag{4.3.78}$$

Damit ist die Numerische Apertur berechnet. Für eine Beispielfaser mit $n_1 = 1.57$, einem Kerndurchmesser von $a = 40\mu m$ und einem Indexsprung von $\Delta = 0.06$ erhält man $\rho = 115\mu m$ und damit die numerische Apertur $N.A. = 0.54$. Zur Illustration zeigt Abb. 4.164 ein gemessenes Brechzahlprofil. Der Knick unten links und rechts zeigt den Durchmesser des Kerns an. Der Dip in der Mitte ist produktionsbedingt.

**4.3.3.2.3. Einmodenfasern** Da die Kerne bei den Gradientenfasern und den Stufenindexfasern meistens so weit sind, dass mehrere Moden übertragen werden, können sie kurze Impulse im ns-Bereich oder kürzer nicht über lange Strecken übertragen. Lichtstrahlen, die unter verschiedenen Winkeln eintreten, legen unterschiedlich lange Wege zurück. damit verbreitern sich Impulse proportional zu

[15] Dieses Verhalten ist die Grundlage für die Konstruktion von GRIN-Linsen

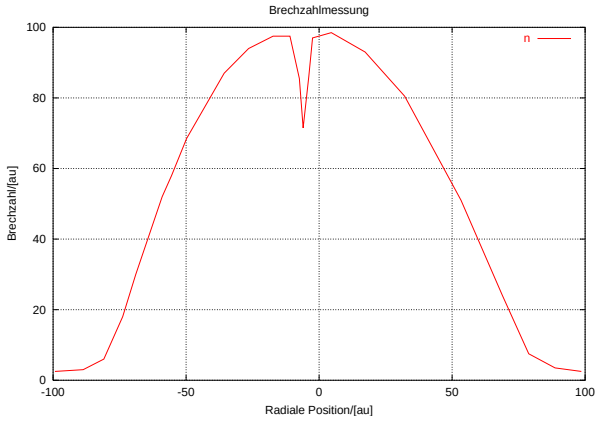



Abbildung 4.165.: Gemessenes Brechzahlprofil einer Gradientenindexfaser.

der Länge der Faser. Wenn nun der Kerndurchmesser auf wenige Mikrometer verkleinert werden, kann die Faser nur noch eine Mode übertragen. Die Lösung des Laufzeitproblems erkauft man sich mit grossen Schwierigkeiten bei der Justage von Faserspleissen.

**4.3.3.2.4. Einkopplung in optische Wellenleiter** Zur Charakterisierung der Einkopplung verwendet man einerseits den Kopplungswirkungsgrad $\eta_K = P_2/P_1$, der das Verhältnis von eingekoppelter Leistung zu angebotener Leistung anzeigt, oder, andererseits, die Kopplungsdämpfung $\alpha_K = 10\,\mathrm{lg}(P_2/P_1)$, die in dB gemessen wird. Wenn eine Laserdiode oder eine LED in eine Faser gekoppelt wird, setzt man für $P_1$ die Leistung des Senders ein. $P_2$ ist dann die in der Faser transportierte Leistung. Der Kopplungswirkungsgrad zwischen optischem Sender und dem optischen Lichtwellenleiter hängt von folgenden Grössen ab:

- Strahlungscharakteristik des Senders
- lokaler Akzeptanzwinkel des optischen Wellenleiters
- Abstand zwischen dem Sender und dem optischen Wellenleiter
- Versatz der optischen Achsen von Sender und optischem Wellenleiter
- Neigung der optischen Achsen von sender und Wellenleiter.

Als Beispiel betrachten wir die Kopplung einer flächigen LED an einen Wellenleiter[**?**]. Die LED wird als Lambert-Strahler modelliert.

$$P_s = \pi^2 \cdot r_{LED}^2 \cdot L_{LED} \cdot \Omega_0 \tag{4.3.79}$$

wobei $r_{LED}$ der Radius der emittierenden Fläche der LED ist, $L_{LED}$ die Strahldichte der LED und $\Omega_0$ der Raumwinkel, in den sie abstrahlt. Die in eine Gradientenfaser mit dem Profilparameter $\alpha$ eingestrahlte Leistung ist

$$P_{LWL} = (\pi \cdot r_{max} \cdot N.A.)^2 \cdot L_{LED} \cdot \Omega_0 \cdot \left[1 - \frac{2}{\alpha + 2}\left(\frac{r_{max}}{a}\right)^{\alpha}\right] \tag{4.3.80}$$

dabei ist

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

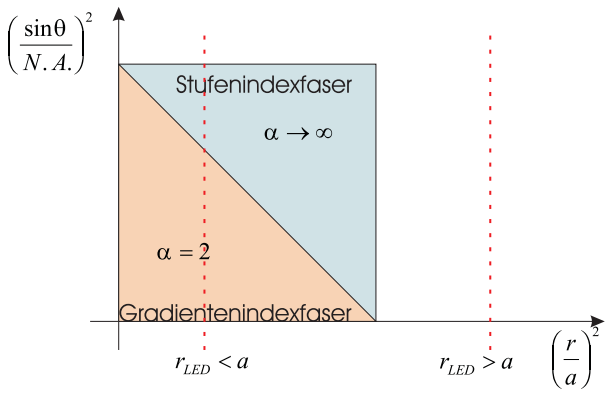



Abbildung 4.166.: Phasenraumdiagramm zur Abschätzung des Kopplungswirkungsgrades von optischen Wellenleitern.

$$r_{max} = \begin{cases} r_{LED} & \text{für} \quad r_{LED} \leq a \\ a & \text{für} \quad r_{LED} \geq a \end{cases}$$

Wie weiter oben eingeführt ist $a$ der Radius des Wellenleiterkerns, $N.A.$ die numerische Apertur und $\alpha$ der Profilparameter. Aus den obigen Gleichungen errechnet man, dass der Koppelwirkungsgrad

$$\eta_K = \left(\frac{r_{max}}{r_{LED}}\right) \cdot N.A.^2 \cdot \left[1 - \frac{2}{2+\alpha}\left(\frac{r_{max}}{a}\right)^{\alpha}\right] \tag{4.3.81}$$

Wenn der Durchmesser des LED-Chips an den Durchmesser des Wellenleiters angepasst ist, erhält man

$$\eta_K = N.A.^2 \left(\frac{2}{2+\alpha}\right) \tag{4.3.82}$$

Damit bekommt man

| Wellenleitertyp | Kopplungswirkungsgrad |
|---|---|
| Stufenindexfaser ($\alpha \to \infty$) | $\eta_K = N.A.^2$ |
| Gradientenindexfaser ($\alpha = 2$) | $\eta_K = \frac{N.A.^2}{2}$ |

Man ersieht daraus, dass der Kopplungswirkungsgrad bei angepassten Durchmessern für Stufenprofilfasern mit $N.A. = 0.5$ $\eta_K = 0.25$ und für $N.A. = 0.24$ $\eta_K = 0.0576$ ist. Für Gradientenindexfasern mit dem gleichen Kerndurchmesser ist die Einkoppeleffizient jeweils halb so gross.
Der Kopplungswirkungsgrad kann über ein sogenanntes Phasenraumdiagramm wie in Abb. 4.166 gezeigt, abgeschätzt werden. Bei diesem wird der Sinus des Enfallswinkels relativ zur numerischen Apertur quadriert ($\left(\frac{\sin\Theta}{N.A.}\right)^2$ gegen die Fläche der LED relativ zur Fläche des Wellenleiterkerns aufgetragen. Man ersieht aus dem Diagramm, dass für den Fall dass der Kerndurchmesser sehr gross wird, die Gradientenindexfaser fast den gleichen
Kopplungswirkungsgrad hat wie die Stufenindexfaser.
Folgende Regeln können abgeleitet werden:

- Die numerische Apertur der Faser sollte so gross wie möglich sein

| Art der Koppeloptik zwischen Sender und Faser | Koppeldämpfung |
|---|---|
| Stirnflächenkopplung | 5 dB ... 8 dB |
| Kugellinse oder Zylinderlinse | 1.5 dB ... 5 dB |
| Faserende dachförmig angeschliffen | 1.5 ... 2 dB |
| Faserende sphärisch angeschmolzen | 0.2 ... 1 dB |
| Faserende als taper ausgezogen | 0.2 ... 1 dB |

Tabelle 4.8.: Koppeldämpfung bei Faser-Faser-Kopplung

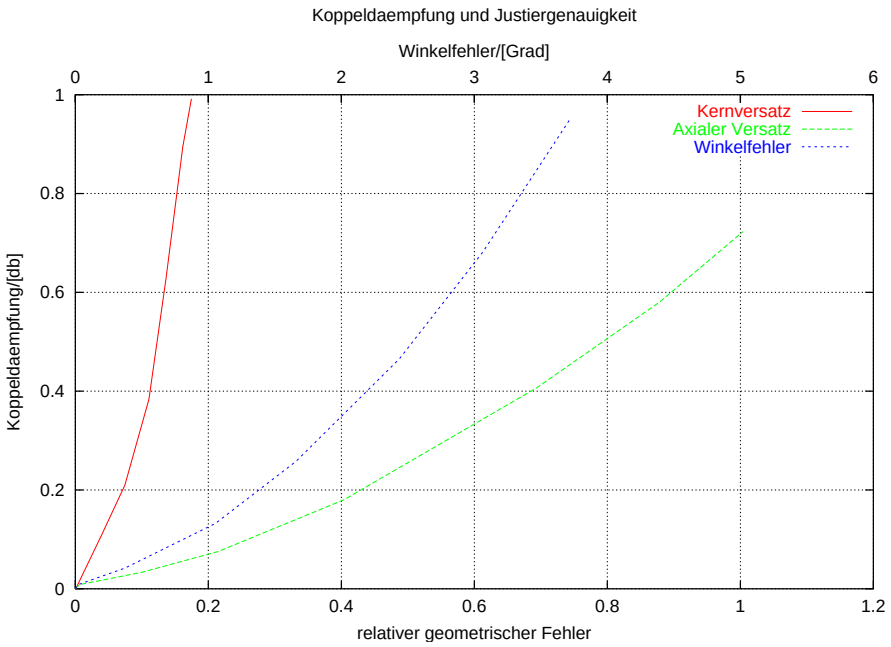



Abbildung 4.167.: Koppeldämpfung $\alpha_K$ bei der Verbindung von Glasfasern.

- Der Kerndurchmesser des optischen Wellenleiters sollte so gross wie möglich sein, mit der Nebenbedingung, dass die Modendispersion ein vorher festgelegtes Mass nicht überschreiten darf.
- Der Profilparameter $\alpha$ sollte so gross wie möglich sein, also eine Stufenindexfaser[16].

Neben der direkten End-zu End-Kopplung werden auch Koppeloptiken verwendet. Die Koppeldämpfungen der gebräuchlichsten Bauarten sind in der Tabelle 4.8 zusammengefasst.
Die Abbildung 4.167 zeigt den Einfluss von Fehlern auf die Koppeldämpfung dargstellt ist

**Kernversatz** Unter dieser Grösse ist der Abstand der Symmetrieachsen der beiden optischen Wellenleiter zu verstehen. Die Darstellung zeigt, dass um die Dämpfung klein zu halten dieser Fehler kleiner als ein zehntel des Kerndurchmessers sein muss. Dies heisst für Multimodefasern eine radiale Positioniergenauigkeit von etwa 2 $\mu$m und für Einmodenfasern von etwa 200 nm!

[16] Stufenindexfasern sind wegen ihrer grossen Modendispersion für Langstreckenübertragungen nicht einsetzbar!

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

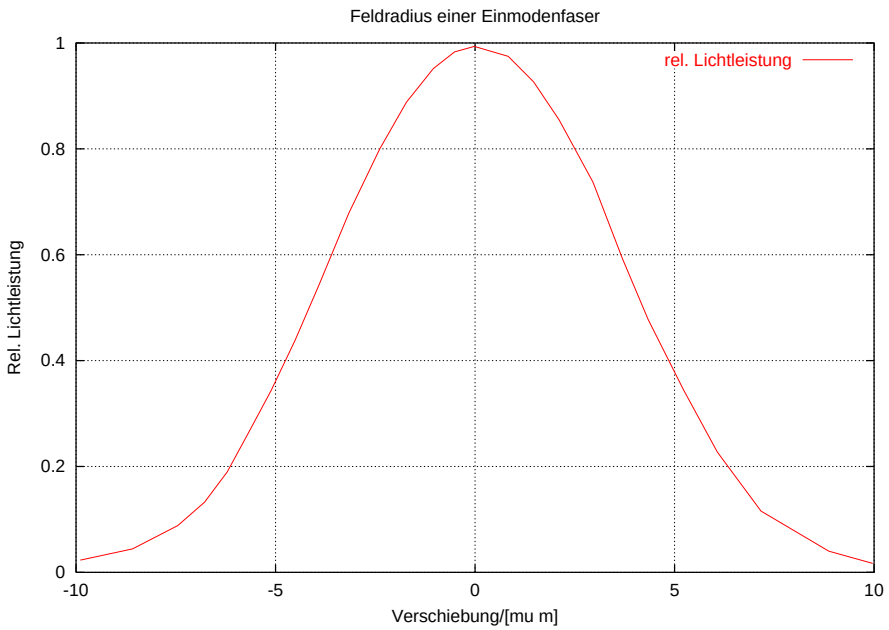



Abbildung 4.168.: Feldradiusbestimmung bei einer Einmoden-Glasfaser.

**Axialer Versatz** Unter dieser Grösse versteht man den Abstand der beiden Faserendflächen. Dieser parameter ist weniger kritisch. Für den gleichen Fehler wie beim Kernversatz darf der Abstand bei Multimodefasern etwa 15 $\mu$m und bei Einmodenfasern 1.5 $\mu$m betragen.

**Winkelfehler** Mit dieser Grösse ist die Verkippung der Faserachsen gegeneinander gemeint. Um den gleichen Fehler wie beim Kernversatz oder beim axialen Abstand zu haben, muss der Winkelfehler kleiner $1^0$ sein.

**4.3.3.2.5. Modenverteilung bei Glasfasern** Abb. 4.168 zeigt das Modenprofil eines Einmoden-Wellenleiters. Die Breite bei $1/e$ ist hier etwa 10 $\mu$m. Abb. 4.169 zeigt den Felddurchmesser als Funktion desr Wellenlänge. Sehr schön sieht man den Einmodenbereich rechts mit einem Minimum kurz bevor die Faser zweimodig wird.

### 4.3.3.3. Bragg-Gitter und Bragg-Sensoren

Wenn durch germaniumdotierte optische Wellenleiter hohe Leistungen gesandt werden kann das Licht Modulationen des Brechungsindexes im Faserkern erzeugen. Diese periodischen Störungen des Brechungsindexes wirken wie ein Bragg-Gitter, analog zur Streuung von Röntgenstrahlen in Kristallen. Heutzutage werden Faser-Bragg-Gitter als Sensoren und Spiegel verwendet[**?**].

**4.3.3.3.1. Herstellung** Die Herstellung von Bragg-Gittern in Quarz-Fasern beruht auf der Lichtempfindlichkeit von Germanium-dotiertem Quarz. Ein einzelnes Photon ($\lambda = 146nm$) kann eine Indexändrung auslösen. Man glaubt, dass oxidierte Germanium-Dimere ($O_3Ge - GeO_3$) durch das Licht aufgespalten werden und dass sich so ein Farbzentrum bildet. Wichtig ist dabei, dass ein Sauerstoffdefizit um dieses Farbzentrum herrscht.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

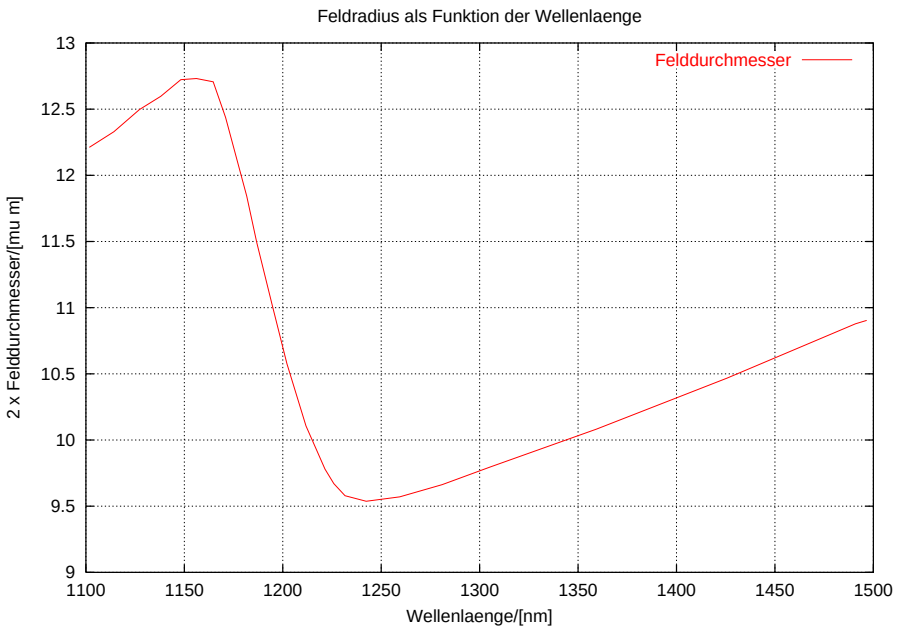



Abbildung 4.169.: Felddurchmesser bei einer Einmoden-Glasfaser als Funktion der Wellenlänge.

Die Photoempfindlichkeit kann gesteigert werden, indem die Faser mit Wasserstoff beladen wird, indem sie mit einer Wasserstoffflamme erhitzt werden und indem Bor zusätzlich zum Germanium dotiert wird.
In einem von verschiedenen diskutierten Modellen wird die Indexvariation im nahen Infrarot auf Absorptionsänderungen im ultravioletten zurückgeführt. Die dielektrische Funktion eines Materials besteht aus einem Realteil und einem Imaginärteil

$$\varepsilon = \varepsilon_r + j\varepsilon_i = (n + j\kappa)^2 \tag{4.3.83}$$

Dabei ist n der *Brechungsindex* und $\kappa$ die Absorptionskonstante. Aus der Kausalität der Physik hatten Kramers und Kronig ihre Beziehung

$$\varepsilon_r(\lambda) = 1 + \int \frac{\varepsilon_i(\lambda')}{\lambda' - \lambda} d\lambda' \tag{4.3.84}$$

zwischen dem Real- und dem Imaginärteil abgeleitet. Wenn nun in einem Frequenzbereich der Imaginärteil (oder auch der Realteil) sich ändert, hat dies einen Einfluss auf den Realteil (oder Imaginärteil) bei **allen** anderen Frequenzbereichen. daraus kann man schliessen, dass ein Farbzentrum im UV-Bereich (ändert $\kappa$) den *Brechungsindex* im Infraroten beeinflusst. Da der Effekt im Imaginärteil über einen weiten Frequenzbereich im Realteil ausgeschmiert wird ist die Änderung des Brechungsindexes gering.
In einem anderen Modell wird angenommen, dass die durch die Photoionisation der Ge-Ge-Bindung freiwerdenenden Elektronen in der Nähe getrappt werden und so Dipolfelder erzeugen. Durch das statische elektrische Feld würden die Suszeptibilität dritter Ordnung moduliert werden und so die Variation des Brechungsindexes hervorrufen.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

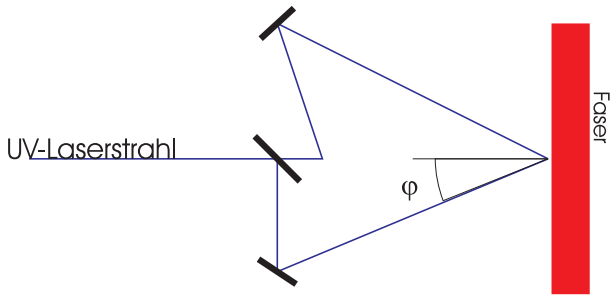



Abbildung 4.170.: Herstellung eines Bragg-Gitters mit Interferometrie.

Ein drittes Modell nimmt an, dass durch die Wechselwirkung mit dem laserlicht die Dichte des Materials des Faserkerns verändert wird. Dadurch würden plastische Verformungen entstehen, die nicht mehr relaxieren könnten.
Ein viertes Modell schliesslich führt die Indexmodulation auf Spannungen zurück, die durch die UV-Beleuchtung entstanden seien. Dabei würde Zugspannung den *Brechungsindex* erniedrigen und Druckspannung ihn erhöhen.

**4.3.3.3.2. Bauformen** Bragg-Gitter können entweder extern oder intern geschrieben werden. Eine häufige Methode bei der externen Generierung ist die Interferometrie. Abbildung 4.170 zeigt, wie man mit einem aufgespaltenen Strahl das Gitter herstellen kann. Die Bragg-Giterkonstante hängt vom halben Öffnungswinkel der beiden Strahlen $\varphi$ sowie von der Wellenlänge $\lambda_w$ des Schreibstrahls ab und ist

$$\Lambda = \frac{\lambda_w}{2 \sin \varphi} \tag{4.3.85}$$

Die Bragg-Wellenlänge in der Faser ist $\lambda_B = 2n\Lambda$. Weiter der Abstand der Indexmaxima in der Faser gleich wie ausserhalb, da die Flächen gleicher Intensität senkrecht zur Faser stehen. Zusammenfassend ergibt sich für die Bragg-Wellenlänge also

$$\lambda_B = \frac{n\lambda_W}{\sin \varphi} \tag{4.3.86}$$

Bei einer Schreibwellenlänge von $\lambda_W = 157nm$, einem Winkel $\varphi = 45°^0$ und einem *Brechungsindex* $n = 1.5$ wäre die Bragg-Wellenlänge $\lambda_B = 333\,nm$. Die interferometrische Methode nach Abb. 4.170 hat zum Vorteil, dass die Wellenlänge sehr leicht geändert werden kann. Nachteilig ist, dass der gesamte Aufbau interferometrische Stabilität benötigt.
Die in der Abb. 4.171 gezeigten Interferometer haben die notwendige Stabilität. Beide Interferometer sind sehr stabil, einfach herzustellen und haben einen einstellbaren Einfallswinkel. Anders als beim Prismenspektrometer geht das Licht beim Lloyd-Spektrometer nicht durch ein Dielektrikum. Dieses Spektrometer ist also weitgehend frei von Dispersionseffekten. Beide Spektrometer können nur Gitter von sehr beschränkter Länge in die Fasern einschreiben. Dies ist ihr hauptsächlicher Nachteil.
Wenn man Gitter mit variabler Tiefe oder Periode der Indexmodulation schreiben will, bedient man sich Häufig der Phasenmasken (Siehe auch abb. 4.172). Diese diffraktiven Masken können entweder holographisch oder lithographisch hergestellt

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

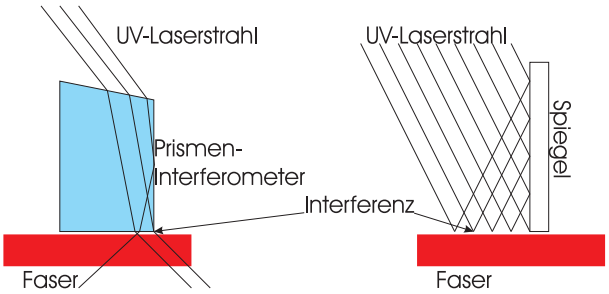



Abbildung 4.171.: Herstellung eines Bragg-Gitters mit einem Prismen-Interferometrie (links) und einem Lloyd-Interferometer (rechts).

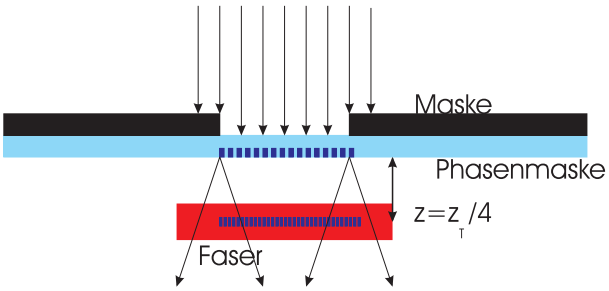



Abbildung 4.172.: Herstellung eines Bragg-Gitters mit einer Phasenmaske.

werden. Die Phasenmasken werden so konstruiert, dass der Interferenzstrahl nullter Ordnung unterdrückt wird (Seine Intensität ist weniger als 5%) . Man versucht etwa 35 % der Intensität in die beiden ersten Ordnungen zu transferieren. Das Nahfeld-Interferenzmuster hat so eine Periode von der Hälfte der Periode der Phasenmaske (Talbot-Effekt). Eine Einführung in den Talboteffekt findet man in der Doktorarbeit von Eero Noponen[17][?]. Durch die Fresnel-Beugung werden periodische Strukturen in ganzzahligen Vielfachen der Talbotdistanz

$$z_T = \frac{2d^2}{\lambda} \tag{4.3.87}$$

exakt abgebildet. Neben dem ganzzahligen Talboteffekt existiert auch der gebrochenzahlige. Mehrfache Bilder des ursprünglichen Gitters werden bei den Distanzen

$$z = \left(q + \frac{p}{n}\right) z_T \tag{4.3.88}$$

gebildet. Dabei sind $n$, $p$ und $q$ ganzzahlig. Zum Beispiel erhält man in der Distanz $z = z_T/(2 * 2) = z_T/4$ zwei um eine halbe Gitterperiode gegeneinander verschobene Phasengitter, wenn das ursprüngliche Gitter ein Amplitudengitter war. Analog erhält man in diesem Abstand zwei um eine halbe wellenlänge gegeneinander verschobene Amplitudengitter, wenn das ursprüngliche Gitter ein Phasengitter war. Zum Beispiel würde ein Gitter der Periode $d = 1\mu m$, beleuchtet mit $\lambda = 500nm$ eine Talbotdistanz von $z_T = 4\mu m$ haben. Das heisst, Im Abstand $z = z_T/4 = 1\mu m$ befindet sich nun ein Gitter mit der Periode $500nm$.

[17] http://focus.hut.fi/ en/dr/node19.html

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

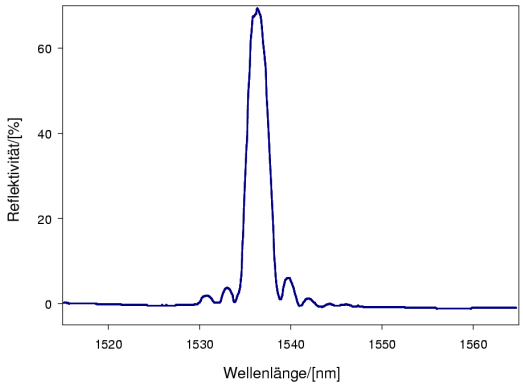



Abbildung 4.173.: Reflektionsspektrum eines Bragg-Gitters dritter Ordnung (Abbildung nach Malo et al.[**?**].

Durch eine Verkippung der Maske kann man, in Grenzen, die Periodendauer einstellen.
Die umfassendste Kontrolle über die Form des Gitters hat man, wenn man dieses mit einem konfokalen Laser-Scanning-Mikroskop schreibt. Dort kann man die Lage und die Modulationstiefe von jedem einzelnen Strahl einstellen. Abb. 4.173 zeigt ein Reflexionsspektrum eines so hergestellten Gitters.

**4.3.3.3.3. Berechnung** Die Streuung an Faser-Bragg-Gittern wird analog zur Braggstreuung in Kristallen behandelt[**?**]. Die Energieerhaltung sagt, dass die Frequenz des einfallenden Lichtes $\omega_i$ und jene des reflektierten Lichtes $\omega_r$ gleich sein müssen. Die Impulserhaltung andererseits liefert die Bedingung, dass

$$\boldsymbol{k}_i + \boldsymbol{K} = \boldsymbol{k}_f \tag{4.3.89}$$

sein muss. Dabei ist $\boldsymbol{K}$ der Gittervektor mit $|\boldsymbol{K}| = 2\pi/\Lambda$, wobei $\Lambda$ die Periodenlänge des Gitters ist.
Da in einer optischen Faser die Ausbreitungsrichtungen vorgegeben sind (entlang einer Achse) erhält man aus Gleichung (4.3.89)

$$2\left(\frac{2\pi n}{\lambda_B}\right) = \frac{2\pi}{\Lambda} \tag{4.3.90}$$

oder, vereinfacht

$$\lambda_B = 2n\Lambda \tag{4.3.91}$$

wobei $\lambda_B$ die Wellenlänge des Lichtes im Vakuum und $n$ der *Brechungsindex* der Faser im Kern ist. Wir nehmen nun an, dass das Bragg-Gitter über die Länge $l$ die Brechzahlmodulation

$$n(x) = n_0 + \Delta n \cos\left(\frac{2\pi x}{\Lambda}\right) \tag{4.3.92}$$

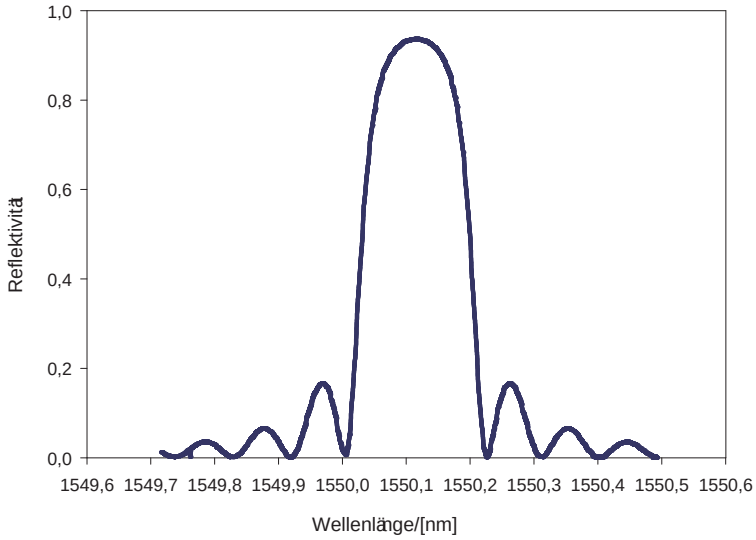



Abbildung 4.174.: Berechnetes Reflexionsspektrum eines Bragg-Gitters (Abbildung nach Othonos.[**?**].

sei, wobei die Modulation $\Delta n$ typischerweise $10^{-5} \ldots 10^{-7}$ ist. Die Reflektivität des Bragg-Gitters ist nun

$$R(\ell, \lambda) = \frac{\Omega^2 \sinh^2(s\ell)}{\Delta k^2 \sinh^2(s\ell) + s^2 \cosh^2(s\ell)} \tag{4.3.93}$$

Die Reflektivität $R(\ell, \lambda)$ ist eine Funktion der Gitterlänge $\ell$ und der Wellenlänge $\lambda$. $\Omega$ ist die Kopplungskonstante, $\Delta k = k - \pi/\lambda$ ist der Wellenvektor der Verstimmung, $k = 2\pi n_0/\lambda$ ist der Wellenvektor des Lichtes und $s = \sqrt{\Omega^2 - \Delta k^2}$. Die Kopplungskonstante ist

$$\Omega = \frac{\pi \Delta n \eta(V)}{\lambda} \tag{4.3.94}$$

Hier ist $\eta(V) \approx 1 - 1/V^2$, $V \geq 2.4$, ist eine Funktion des Faserfüllfaktors $V$, der angibt, wieviel der Faserintensität der Grundmode im Kern (mit dem Bragg-Gitter) lokalisiert ist. Abb. 4.174 zeigt ein berechnetes Reflexionsspektrum. Auf der Mittenfrequenz des Bragg-Gitters ist $\Delta k = 0$. Also ist die Reflektivität

$$R(\ell, \lambda) = \tanh^2(\Omega \ell) \tag{4.3.95}$$

Die Halbwertsbreite des Reflexionsmaximums ist gegeben durch[**?**]

$$\Delta\lambda = \lambda_B \alpha \sqrt{\left(\frac{\Delta n}{2n_0}\right)^2 + \left(\frac{1}{N}\right)^2} \tag{4.3.96}$$

Abb. 4.175 zeigt, dass Bragg-Gitter die in sehr empfindliche Fasern geschrieben werden, die also eine starke Modulation des Brechungsindexes haben, auf der höherfrequenten seite des Bragg-Peaks ein ausgeprägtes Spektrum haben, das von Mantelmoden herrührt. Die spektralen Eigenschaften werden von Licht, das die Faser seitwärts verlässt, bestimmt.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

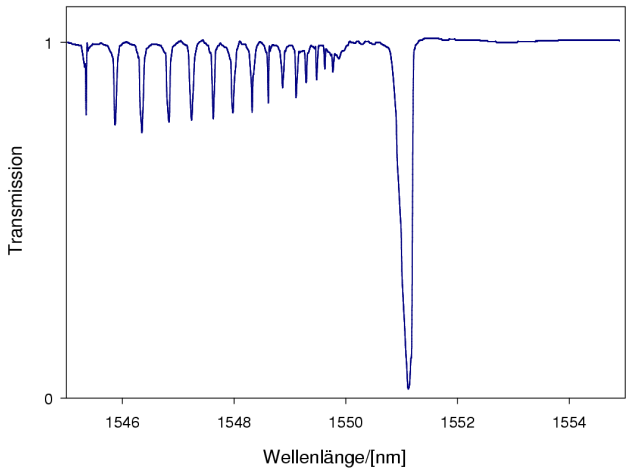



Abbildung 4.175.: Transmission durch ein starkes Bragg-Gitter. Dabei ist klar ersichtlich dass Strahlung in die Mantelmoden gekoppelt wird. (Abbildung nach Othonos[**?**]).

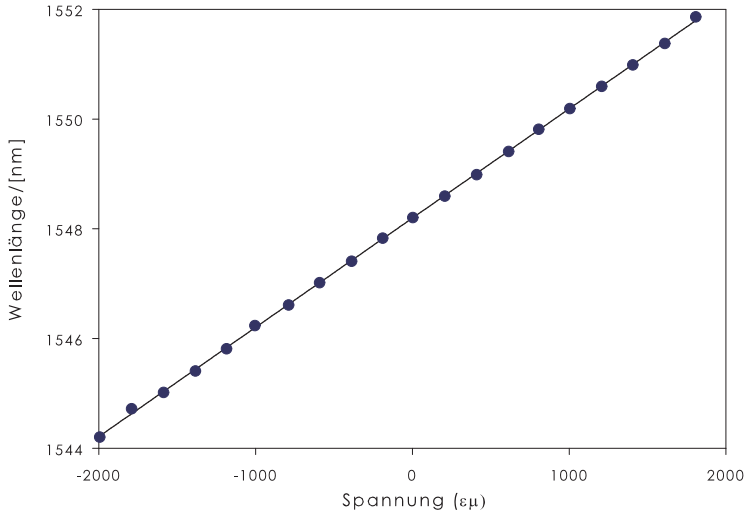



Abbildung 4.176.: Bragg-Wellenlänge eines Bragg-Gitters als Funktion der angelegten mechanischen Spannung (Abbildung nach Othonos[**?**]). Das Bragg-Gitter war in eine Erbium-Dotierte Faser eingeschrieben und arbeitete als Auskoppelgitter.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

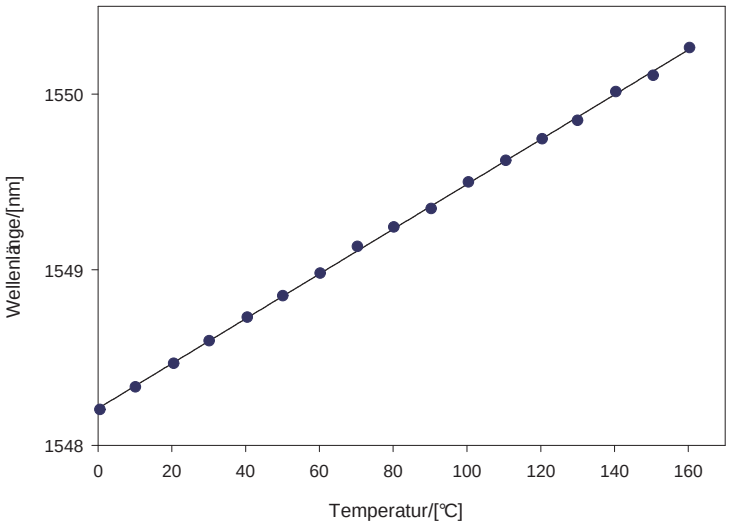



Abbildung 4.177.: Bragg-Wellenlänge eines Bragg-Gitters als Funktion der Temperatur (Abbildung nach Othonos[**?**]). Das Bragg-Gitter war in eine Erbium-Dotierte Faser eingeschrieben und arbeitete als Auskoppelgitter.

**4.3.3.3.4. Anwendungen als Sensor** Die Mittenfreqquenz des Faser-Bragg-Gitters hängt vom *Brechungsindex* und der Periodenlänge ab. Beide Grössen werden jedoch durch externe Parameter verändert. Sowohl die Temperatur wie auch Zug auf die Faser können die Mittenfrequenz verschieben. Aus Gleichung (4.3.91) ermöglicht eine Berechnung der Verschieb3ung der Mittenwellenlänge des Gitters.

$$\Delta\lambda_B = 2\left(\Lambda\frac{\partial n}{\partial \ell} + n\frac{\partial \Lambda}{\partial \ell}\right)\Delta\ell + 2\left(\Lambda\frac{partialn}{\partial T} + n\frac{\partial \Lambda}{\partial T}\right)\Delta T \tag{4.3.97}$$

Der erste Summand in Gleichung (4.3.97) stellt den Einfluss von Zugspannungen dar (Eine Messung ist in Abb. 4.176 zu sehen). Man kann diesen Effekt auch als

$$\delta\lambda_B = \lambda_B(1 - p_e)\epsilon_z \tag{4.3.98}$$

darstellen. In dieser Gleichung ist $p_e$ die effektive spannungsoptische Konstante. Sie ist wie folgt definiert:

$$p_e = \frac{n^2}{2}\left[p_{12} - \nu(p_{11} + p_{12})\right] \tag{4.3.99}$$

Dabei sind $p_{11}$ und $p_{12}$ Komponenten des spannungsoptischen Tensors. $n$ ist der *Brechungsindex* im Kern der Faser und $\nu$ ist die Poisson-Zahl. Bei einer typischen optischen Faser ist nach Othonos[**?**] $p_{11} = 0.113$, $p_{12} = 0.252$, $\nu = 0.16$ und $n = 1.482$. Man erwartet dann eine Empfindlichkeit von $0,001pm$ für eine Spannung von $10^{-6}$.
Der zweite Teil von Gleichung (4.3.97) beschreibt den Einfluss der Temperatur. Einerseits ändert die Temperatur den Abstand der Indexschwankungen, also die Periodenlänge, und andererseits ändert sich der *Brechungsindex*. Wir können für

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

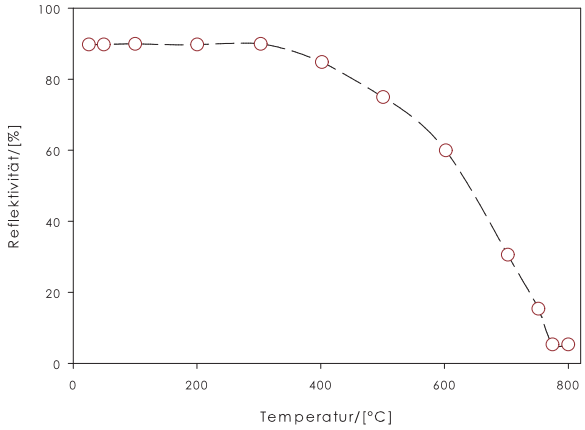



Abbildung 4.178.: Temperaturabhängigkeit der Reflektivität (Abbildung nach Meltz[**?**]).

die Änderung der Bragg-Wellenlänge schreiben:

$$\Delta\lambda_B = \lambda_B \left(\alpha + \zeta\right) \Delta T \tag{4.3.100}$$

Dabei ist $\alpha = (1/\Lambda)(\partial\Lambda/\partial T)$ der thermische Ausdehnungskoeffizient[18] und $\zeta = (1/n)(\partial n 7 \partial T)$ der thermo-optische Koeffizient[19]. Der Temperatureffekt ist also durch die Änderung des Brechungsindexes dominiert. Der Zahlenwert für Quarzglas ist 14 pm/$^{\circ 0}$C. Abb. 4.177 zeigt den Einfluss der Temperatur auf die Bragg-Wellenlänge eines Bragg-Gitters.
Die Änderung der Temperatur bewirkt nicht nur eine Verschiebung der Bragg-Wellenlänge, sondern auch eine Erniedrigung der Reflektivität, wie es schön aus Abbildung 4.178 ersichtlich ist.
Da jedes Bragg-Gitter in einer Faser mit mehreren Sensorstellen eine eigene, klar von den anderen trennbare Resonanzfrequenz haben kann, können einzelne Temperatur- oder Spannungssensoren über eine Auswahl der Wellenlänge adressiert werden.
Abb. 4.179 zeigt einen Fabry-Perot-Resonator in einer Faser. Der rechte Teil der Abbildung zeigt, dass dieses Fabry-Perot in einer Faser eine exzellente Linienbreite hat.
Abb. 4.180 zeigt, dass man bei Faser-Bragg-Gittern mehrere Gitter übereinander einbringen kann. Dies ist einsichtig, wenn man bedenkt, dass ein Faser-Bragg-Gitter eigentlich mit Hologrammen verwandt ist. Auch bei Hologrammen können mehrere von ihnen in der gleichen Fotoschicht gespeichert werden. Mit Faser-Bragg-Gittern lassen sich so ganz neuartige Interferometer aufbauen.

[18] Für Quarz etwa $\alpha = 5.5 \times 10^{-7}$
[19] Für Germanium-dotierten Quarz etwa $\zeta = 8.6 \times 10^{-6}$.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

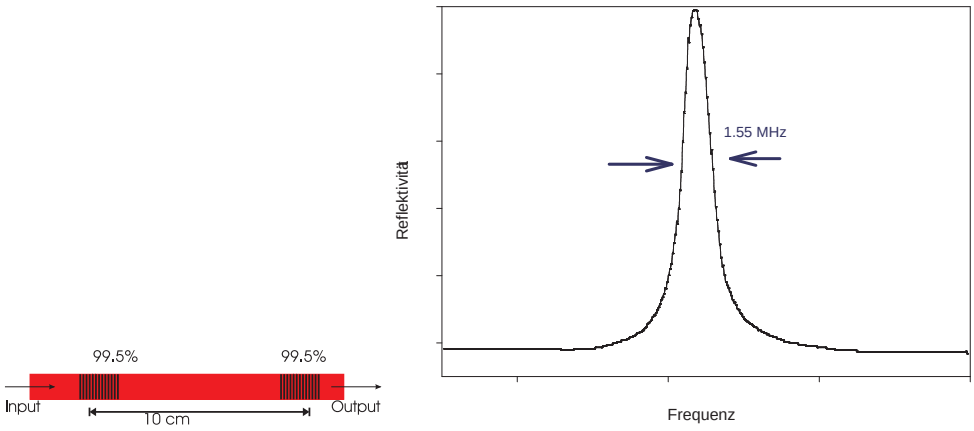



Abbildung 4.179.: Zwei Bragg-Gitter als Fabry-Perot-Resonatoren. Rechts ist das Transmissionsspektrum gezeigt (Abbildung nach Othonos[**?**]).

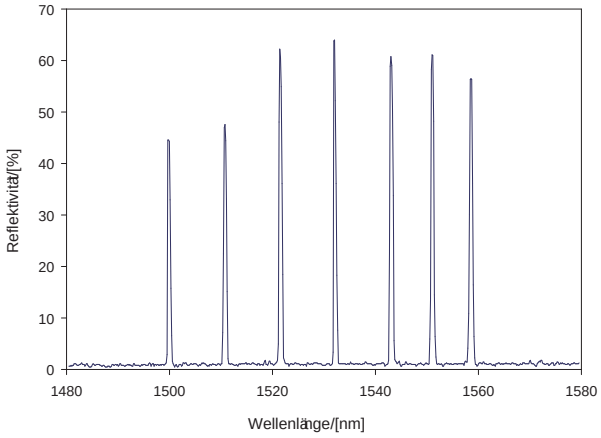



Abbildung 4.180.: Reflexionsspektrum für sieben am gleichen Ort eingebrannte Bragg-Gitter (Abbildung nach Othonos[**?**]).

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

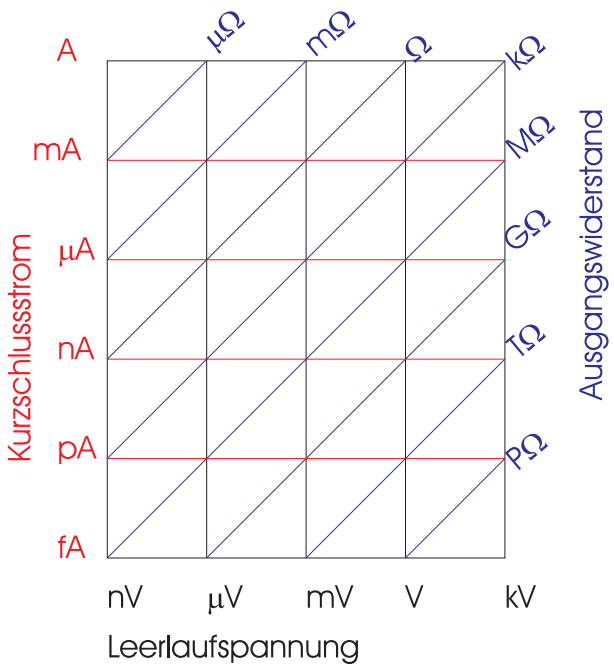



Abbildung 4.181.: Testfeld für eine *Spannungsquelle*[**?**].

## 4.4. Messungen kleiner Pegel

In diesem Abschnitt werden Fehlerquellen bei der Messung kleiner Signale diskutiert. Viele der hier behandelten Effekte sind sehr schön in einer Broschüre von Keithley[**?**] dargestellt.

### 4.4.1. Testfelder

Wenn eine elektrische Messung mit kleinen Pegeln, hohen Impedanzen oder kleinen Strömen durchgeführt wird und gleichzeitig eine bestimmte G, so müssen sogenannte Testfelder bestimmt werden.
Im allgemeinen hängen der Kurzschlussstrom $I_K$ einer *Spannungsquelle* und ihre Leerlaufausgangsspannung $U_L$ über einen äquivalenten Quellenwiderstand $R_S$

$$R_S = \frac{U_L}{I_K} \tag{4.4.1}$$

In Abb. 4.181 ist nun ein Testfeld gezeigt. Auf der horizontalen Achse ist dabei die Leerlaufausgangsspannung $U_L$ angegeben. Die vertikale Achse ist der Kurzschlussstrom $I_K$. Die schrägen Linien zeigen den dazugehörigen Ausgangswiderstand. Um herauszufinden, wie eine Messung durchgeführt werden muss, werden die folgenden Schritte abgearbeitet:

- Man bestimmt den Kurzschlussstrom $I_K$ und die Leerlaufspannung $U_L$. Dies ergibt Punkt **A** in Abb. 4.182.
- Es wird die gewünschte Genauigkeit festgelegt.
- Vom Punkt **A** aus zeichnet man eine Linie, deren Länge der gewünschten Genauigkeit entspricht, nach unten und kommt so zum Punkt **B**. Dabei

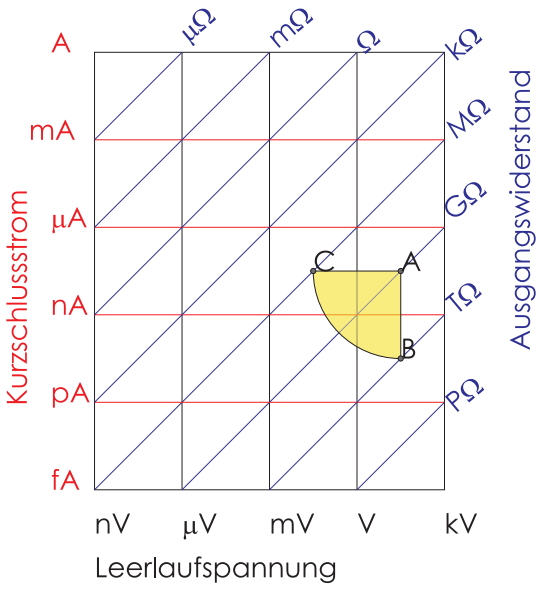



Abbildung 4.182.: Beispiel für eine Testhülle[**?**].

entspricht eine Genauigkeit von 1 % zwei Dekaden. Ein kleinerer Fehler entspricht mehr Dekaden.

- Ebenso wird eine horizontale Linie mit der gleichen Länge vom Punkt **A** zum Punkt **C** gezeichnet.
- Der Viertelkreis zwischen den Punkten **B** und **C** umschliesst zusammen mit den beiden Geraden das Testfeld.

Dieses Testfeld bedeutet nun, dass Parallelwiderstände zur Quelle, die grösser als der Widerstand am Punkt **B** sind, die gewünschte Genauigkeit nicht beeinträchtigen. Dieser minimale Parallelwiderstand zeigt, wie gross der Innenwiderstand eines Spannungsmessers sein muss, damit die Spannungsmessung die geforderte Genauigkeit ermöglicht. Ebenso zeigt der Punkt **C**, wie klein ein Serienwiderstand sein muss (links von diesem Punkt), damit eine Strommessung nicht durch das *Messgerät* verfälscht wird.
Tabelle 4.9 zeigt eine Aufstellung verschiedener Messarten, häufige Fehlerquellen und Möglichkeiten ihrer Beseitigung.
Abbildung 4.183 stellt die durch verschiedene Störmechanismen unzugänglichen Messbereiche bei einer Spannungsmessung dar. Blau ist der Bereich, in dem Thermospannungen das Messresultat verfälschen. Die in der Abbildung gezeigten Bereiche hängen von der Temperatur und den Materialkombinationen ab. Sie sind im Einzelfall neu zu berechnen. Die rotbraune Farbe zeigt den Bereich von Ausgangswiderständen (oder, äquivalent, von Kombinationen von Spannungen und Strömen) an, bei denen ein *Messgerät* mit hier $10M\Omega$ *Eingangswiderstand* 10% Fehler erzeugt. Grün ist der Bereich, der wegen Eingangsströmen im *Messgerät* nicht zugänglich ist. Schwarz schliesslich ist der Bereich des weissen Rauschens oder des thermischen Rauschens.
Abbildung 4.184 stellt die durch verschiedene Störmechanismen unzugänglichen Messbereiche bei einer Strommessung dar. Blau ist der Bereich, der durch den Spannungsabfall am *Messwiderstand* einen Fehler erzeugt. Die rotbraune Farbe

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

| **Art der Messung** | **Messbereich** | **Anzeichen für Fehler** | **Wahrscheinliche Ursachen** | **Massnahmen zur Vermeidung** |
|---|---|---|---|---|
| Kleine Spannungen | $< 1\mu V$ | Offsetspannung | Thermospannungen | Alle Anschlüsse auf der gleichen Temperatur halten<br>Gekrimpte Cu-Cu-Verbindungen |
| | | Störspannungen | Magnetische Interferenzen | Verdrillte Leitungen<br>Leitungen von Magnetfeldern entfernen oder abschirmen |
| Kleine Ströme | $< 1\mu A$ | Offsetstrom | Leckströme in Isolatoren | Gute Isolatoren verwenden, gut reinigen |
| | | | Messstrom im Messwerk (Bias) | Picoampèremeter oder Elektrometer verwenden |
| | | | Dunkelstrom im Detektor | Dunkelstrom unterdrücken oder kompensieren |
| | | Störströme | Elektrostatische Kopplung | Hohe Spannungen und Relativbewegungen der Kabel dazu vermeiden<br>Abschirmung |
| | | | Vibration / Deformation | Vibrationen fernhalten<br>Rauscharme Kabel verwenden |
| Niedrige Widerstände | $< 100m\Omega$ | Widerstandsoffsets | Kabelwiderstand | 4-Draht Methode (Kelvin-Methode) verwenden |
| | | Drift | Thermospannungen | Pulsförmige Testsignale mit Offsetkompensation |
| | | Schwankende Messwerte | Magnetische Interferenz | Von Magnetfeldern fernhalten oder abschirmen<br>Verdrillte Leitungen verwenden |
| Hohe Widerstände | $> 1G\Omega$ | Ablesung zu klein | Belastungswiderstand (Shunt) | Anschlüsse und Kabel mit höherem Isolationswiderstand verwenden<br>Guard-Techniken verwenden |
| | | | Niedriger $R_{ein}$ des Voltmeters | *Spannungsquelle* und Strommessung verwenden |
| | | | Offsetströme | Offsetströme bei abgeschalteter Testspannung kompensieren |
| | | Schwankende Werte | Elektrostatische Kopplung | Hohe Spannungen in der Nähe sowie Bewegung des Kabels vermeiden |
| Spannung aus einer Quelle mit hoher Impedanz | $> 1M\Omega$ | Ablesung zu klein | Belastungswiderstand (Shunt) | Anschlüsse und Kabel mit höherem Isolationswiderstand verwenden<br>Guard-Techniken verwenden |
| | | | Offsetströme | Offsetströme bei abgeschalteter Testspannung kompensieren |
| | | Schwankende Werte | Elektrostatische Kopplung | Hohe Spannungen in der Nähe sowie Bewegung des Kabels vermeiden |

Tabelle 4.9.: Übliche Fehlerquellen und Massnahmen, um ihren Einfluss zu vermindern (nach Keithley[**?**]

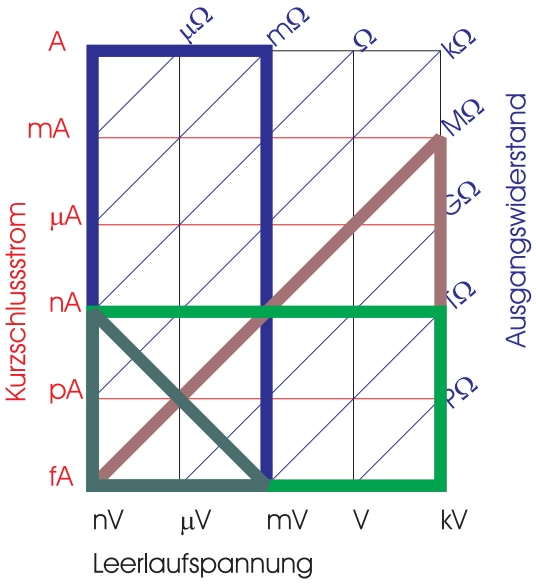



Abbildung 4.183.: Fehlerquellen bei der Spannungsmessung[**?**]. Blau: Thermospannungen, rotbraun: Fehler durch den zu grossen Ausgangswiderstandes der Quelle, grün: Eingangsströme des des Voltmeters und schwarz: thermisches Rauschen.

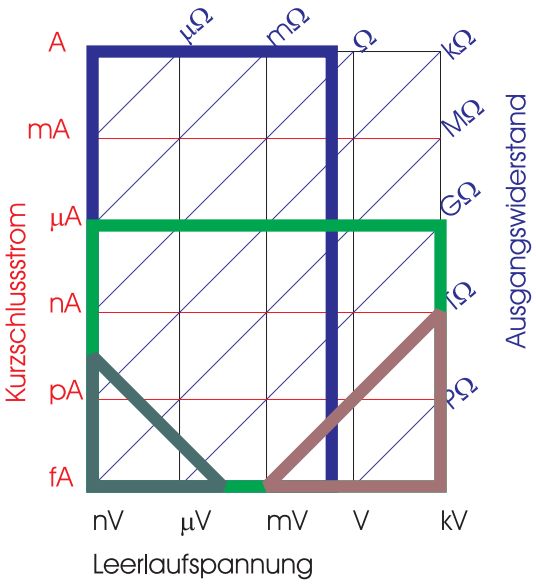



Abbildung 4.184.: Fehlerquellen bei der Strommessung[**?**]. Blau: Spannungsabfall am *Messgerät*, rotbraun: Fehler durch Parallelwiderstände zur Quelle, grün: Induzierte oder generierte Ströme und schwarz: thermisches Rauschen.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

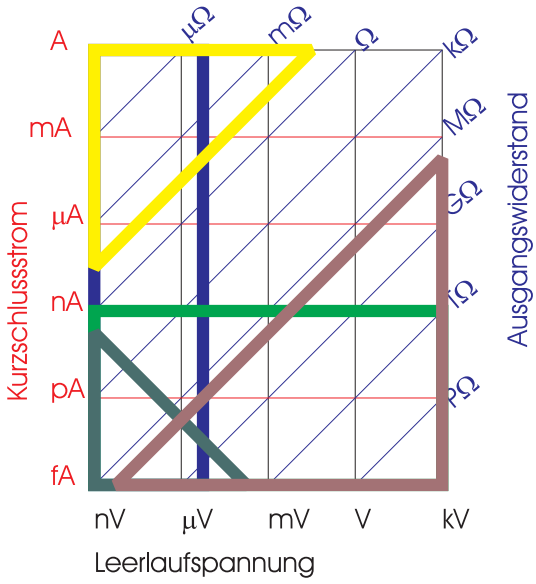



Abbildung 4.185.: Fehlerquellen bei der Widerstandsmessung[**?**]. Gelb: Leitungswiderstand in Serie zum Testobjekt, blau: Thermospannungen, rotbraun: Isolationswiderstände parallel zum Testobjekt, grün: generierte oder induzierte Ströme und schwarz: thermisches Rauschen.

zeigt den Einfluss von Parallelwiderständen zur zu messenden Quelle dar. Grün ist der Bereich, der wegen induzierten oder generierten Strömen nicht zugänglich ist. Schwarz schliesslich ist der Bereich des weissen Rauschens oder des thermischen Rauschens.
Abbildung 4.185 stellt schliesslich die durch verschiedene Störmechanismen unzugänglichen Messbereiche bei einer Widerstandsmessung dar. Gelb ist der Bereich, der wegen den Widerständen des Messkabels nicht zugänglich ist. Blau ist der Bereich der Thermospannungen. Die rotbraune Farbe zeigt den Einfluss von Isolationswiderständen parallel zur zu messenden Quelle dar. Grün ist der Bereich, der wegen induzierten oder generierten Strömen nicht zugänglich ist. Schwarz schliesslich ist der Bereich des weissen Rauschens oder des thermischen Rauschens.
Allgemeine Fehlerquellen bei elektrischen Messungen sind

**Rauschen** Das thermische Rauschen, wie es im Abschnitt 2.8 dargestellt wurde, hat eine Leistung von $P = 4kT\Delta f$ in der Bandbreite $\Delta f$.

**Drift** *Messgeräte* sind im allgemeinen nicht stabil. Ihre Anzeigen ändern sich langsam mit der Zeit. Diesen Fehler nennt man Drift.

**Geschwindigkeit** Jede Messung braucht eine bestimmte Zeit. Wird von einem Messsystem wie zum Beispiel einem Lock-In-Verstärker (sie auch den Unterabschnitt 4.1.9 eine im Vergleich zu seiner Bandbreite zu hohe Datenrate verlangt, so sind die Messungen wegen der Bandbreitenbegrenzung fehlerbehaftet.

**Umgebungsbedingungen** Temperatur und Luftfeuchte können Messfehler verursachen. Einerseits verursachen Temperaturänderungen Änderungen der Leit-

fähigkeit der Widerstände der eingesetzten *Messgeräte*. Andererseits bedeutet eine hohe Luftfeuchtigkeit dass die Isolationswiderstände in den Messgeräten niedriger werden und so Messfehler verursachen.

**Ionisierende Strahlung** Ionisierende Strahlen können durch ihre Energie Elektronen aus einem Leiter oder Isolator herauslösen. Dadurch fliessen zusätzliche Ströme. Während ionisierende Strahlung als Folge der menschlichen Aktivitäten Teilchen mit relativ niedrigen Energien emittiert, bestehen die kosmischen Strahlen im allgemeinen aus Teilchen mit sehr hohen Energien. Deshalb können letztere nicht und erstere nur unter grösserem Aufwand abgeschirmt werden.

**Netzstörungen** Netzstörungen, das heisst das Übersprechen der Spannungsversorgung des *Messgerätes*, des Prüflings oder von anderen Geräten sind im allgemeinen die häufigsten Störungen und mit von den am schlechtesten zu eliminierenden. Die einzige Massnahme ist meistens, die Erdverbindungen neu zu konfigurieren.

**Mechanische Störungen** Unter den mechanischen Störungen versteht man die Einflüsse von Vibrationen und auch die Wirkungen von Verbiegungen von Kabeln.

**Schutzerde und Schutzleiter** Diese für einen sicheren Betrieb von Geräten notwendigen Einrichtungen erzeugen vor allem in Verbindung mit Oszilloskopen Brummschleifen. Diese können nur aufgebrochen werden, indem man die Oszilloskope für differentielle Messungen einrichtet, also nur die Hälfte der Kanäle verwendet. Alternativ kann man eigene Differenzverstärker den Oszilloskopen vorschalten. Die letzte Möglichkeit ist, **zugelassene Trenntransformatoren** zu verwenden.

Achtung! Nie Schutzleiter auftrennen! Immer Trenntransformatoren oder Differenzverstärker verwenden!

## 4.4.2. Spannungen

Bei Spannungsmessungen treten insbesondere die folgenden Fehlerquellen auf

**Anschlüsse mit hohen Impedanzen** Leck- Streuströme verfälschen eine Spannungsmessung. Wenn gilt, dass $V_s < 1\mu A \cdot R_s$ ist, so muss bei der Spannungsmessung besonders vorsichtig vorgegangen werden. Dabei ist $V_s$ die geforderte Spannungsempfindlichkeit, 1 $\mu$A der Strom, der durch den Innenwiderstand des Voltmeters fliesst. Wenn die geforderte Empfindlichkeit 1 mV ist, dann hat müssen Quellen mit $R_s > 1k\Omega$ mit besonderen Vorsichtsmassnahmen gemessen werden.

**Isolationswiderstand** Durch schlechte Isolationsmaterialien werden Spannungsmessungen verfälscht. Die Wahl des Isolationsmaterials entscheidet über die Qualität von Spannungsmessungen. Im Abschnitt I.1 ist eine Tabelle von Isoltionsmaterialien angegeben.

**Eingangswiderstand** Jedes reale Voltmeter kann als ein ideales Voltmeter mit dem *Eingangswiderstand* $R_{ein}$ des realen Voltmeters in parallel ersetzt werden. Die gemessene Spannung ist $V_{mess} = V_s \frac{R_{ein}}{R_{ein}+R_s}$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

**Offsetspannung** Jede Spannung in Serie mit der zu messenden Spannung und der Spannung am Voltmeter verfälscht die Ablesung. Zu den Offsetspannungen gehören die Thermospannungen und die durch wechselnde Magnetfelder induzierten Spannungen.

**Offsetstrom** Der Offsetstrom bei einem idealen Voltmeter und einer *Spannungsquelle* mit dem Innenwiderstand $R_s$ verfälscht die Quellenspannung $V_s$ zu $V_m = V_s \pm I_{offset} \cdot R_s$. Offsetströme entstehen durch die Eingangstransistoren der *Messgeräte*. Jeder *bipolare Transistor* benötigt einen minimalen Eingangsstrom um zu funktionieren. Bei digitalen Voltmetern und bei Nanovoltmetern beträgt der Offsetstrom etwa $10pA$ bis $10nA$. Der Offsetstrom von Elektrometerverstärkern kann von $10fA$ bis hinunter zu $50aA$[20] betragen.

**Belastungswiderstand** Widerstände parallel zum Voltmeter verfälschen die Messung. So wird bei einem Quellwiderstand $R_s$ und bei einem Querwiderstand $R_Q$ die gemessene Spannung $U_m = U_s \left(\frac{R_Q}{R_Q+R_s}\right)$. Vielfach ist der Isolationswiderstand des Kabels der unerwünschte Querwiderstand. Der Einfluss des Kabelisolationswiderstandes kann durch Guard-Techniken vermindert werden. Wenn die Verstärkung des Guard-Verstärkers $A_{guard}$ ist, dann ist die gemessene Spannung $U_m = U_s \left(\frac{A_{guard} \cdot R_Q}{A_{guard} \cdot R_Q+R_s}\right)$

**Kapazität zur Schirmleitung** Die Kapazität zur Abschirmung des Messkabels $C_k$ bewirkt zusammen mit dem Ausgangswiderstand $R_s$ eine Zeitkonstante $\tau = R_s C_k$. Damit ist der Zeitverlauf der Messspannung $U_m = U_s \left[1 - \exp\left(-\frac{t}{\tau}\right)\right] = U_s \left[1 - \exp\left(-\frac{t}{R_s C_k}\right)\right]$. Dabei wird die Ladung $Q = U_s C_k$ auf die Kabelkapazität $C_k$ übertragen. Wenn die Abschirmung mit einem Verstärker (Verstärkung $A_{guard}$ auf dem Potential der Eingangsspannung gehalten wird, ist die Zeitkonstante $\tau_{guard} = \frac{\tau}{A_{guard}}$. Die auf der Kabelkapazität gespeicherte Ladung ist dann $Q = U_s \frac{C_k}{A_{guard}}$.

**Thermospannungen** Thermospannungen werden zur zu messenden Spannung hinzu- oder abgezählt. Die Grösse der Thermospannungen hängt von der Materialkombination und von den Temperaturen entlang des Messkreises ab. Tabelle I.2 gibt die wichtigsten thermoelektrischen Koeffizienten an.

**Thermospannungen in Steckern** Die Thermospannungen in Steckern werden meistens vergessen. Anders als im Rest des Messkreises sind die Materialien und die Temperaturen sehr viel schlechter kontrollierbar. Durch Übergangswiderstände zwischen Stecker und Kupplung kann sich die Kontaktstelle unbemerkt und unkontrolliert erwärmen. Zu den Steckern gehören auch geräteinterne Stecker. Bei externen Verbindungen kann der Einfluss der Thermospannungen untersucht werden, indem man die Steckerverbindungen (sofern möglich!) umkehrt.

**Gleichtaktstrom und daraus resultierende Fehler** Gleichtaktströme können insbesondere Messungen von sehr kleinen Spannungen beeinflussen. Zu den Gleichtaktstromquellen gehört unter anderem die Ströme die zwischen der

[20] $50aA$ entsprechen 300 Elektronen pro Sekunde.

Netzerde und dem Erd-Eingangspol (Buchse "0") des *Messgerätes* fliesst. Vielfach rührt dieser Strom von der kapazitiven Kopplung zwischen der Primär- und Sekundärspule des Netztransformators her[**?**]. Noch schlimmer sind die Fehler, wenn die Buchse "0" mit dem empfindlichen Teil des Messobjektes verbunden ist. Dies ist unter allen Umständen zu vermeiden.

**Magnetfelder** Magnetfelder induzieren Spannungen in alle von den Messleitungen eingeschlossenen Flächen. Die Maxwell'schen Gleichungen ergeben, dass die induzierte Spannung $U_B = \frac{d\Phi}{dt} = \frac{d\boldsymbol{BA}}{dt} = \boldsymbol{B}\frac{d\boldsymbol{A}}{dt} + \boldsymbol{A}\frac{d\boldsymbol{B}}{dt}$. Verdrillte Kabel und eine gut überlegte Führung der Kabel minimieren die induzierten Spannungen von variierenden Magnetfeldern. Bewegte Kabel (zum Beispiel zu einer beweglichen Messstelle) können auch bei statischen Magnetfeldern induzierte Spannungen bewirken.

**Erdschleifen** Einer der häufigsten Fehler sind Erdschleifen. Sie rühren daher, dass netzbetriebene *Messgeräte* einerseits über ihre "0"-Buchsen und andererseits über den Schutzleiter verbunden sind. Dieser fehler kann vermindert werden, indem differentielle Eingänge verwendet werden. Bei Oszilloskopen müssen gesonderte Differenzverstärker vorgeschaltet werden. Bei Datenerfassungskarten für Computer sollte immer der Variante mit differentiellen Eingängen der Vorzug vor der Variante mit den einfachen Eingängen gegeben werden.

Achtung! Nie Schutzleiter auftrennen! Immer Trenntransformatoren oder Differenzverstärker verwenden!

**Abschirmung** Wenn man bei einem Elektrometerverstärker an den empfindlichen Eingang 2 cm Draht anschliesst, den 2V-Bereich einstellt, ein Stück Kunststoff an Wolle reibt und dieses Stück etwa einen Meter vom Eingang entfernt hin und her bewegt, dann schlägt das Elektrometer merklich aus. Ähnliche Experimente kann man auch mit Wechselfeldern durchführen. In beiden Fällen hilft nur eine Abschirmung. Diese Abschirmung sollte

- die zu testende Schaltung, das Elektrometer und die die Messung durchführende Person einschliessen, oder
- die zu testende Schaltung und das Elektrometer, oder
- nur die zu messende Schaltung sowie abgeschirmte Kabel

umfassen. Es ist vorteilhaft für solche Messungen triaxiale Kabel zu verwenden, wobei die innere Schirmung mit einer Guardschaltung auf dem Messpotential gehalten werden sollte.

### 4.4.3. Ströme

Präzise Strommessungen werden ebenso wie Spannungsmessungen durch Fehlerquellen verfälscht. Diese Fehler sind

**Spannungsbelastung** Ein realer Strommesser kann als ein idealer Strommesser in Serie mit einem *Messwiderstand* $R_m$ angesehen werden. Der Kurzschlussstrom einer *Spannungsquelle* $U_s$ mit dem Innenwiderstand $R_s$ wäre $I_s = \frac{U_s}{R_s}$. Durch den endlichen Innenwiderstand des Strommessers ist der gemessene Strom aber $I_m = \frac{U_s}{R_s+R_m} = I_s\frac{R_s}{R_s+R_m}$. Bei digitalen Messgeräten entspricht der Messstrom einer Spannung $U_b$. Dann ist der gemessene Strom $I_m = \frac{U_s-U_b}{R_s} = I_s - \frac{U_b}{R_s}$.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

**Leitungswiderstände** Trotzdem die Leitungswiderstände $R_L$ meistens kleiner als der Ausgangswiderstand $R_s$ oder der *Messwiderstand* $R_m$ ist, können Sie die Strommessung verfälschen. Der gemessene Strom ist $I_m = \frac{U_s}{R_s}\frac{R_s}{R_s+R_L}$.

### Rauschen und Interferenzen

**Quellenwiderstand** Durch die Abschirmung kann der Einfluss einer externen Störspannung auf die Strommessung minimiert werden. Jedoch bewirkt die Schirmung zusammen mit der bei einer Stromquelle notwendigerweise hohen Ausgangsimpedanz eine Zeitkonstante. Bei empfindlichen Strommessungen müssen deshalb neben der Abschirmung auch Guard-Techniken verwendet werden.

**Eingangskapazität** Bei Strom/Spannungswandlern ist die Verstärkung für Rauschen grösser als für das Nutzsignal. Wenn wir eine Quelle mit der Spannung $U_s$ und dem Ausgangswiderstand $R_s$ betrachten, die an einen Strom/Spannungswandler mit dem Rückkopplungswiderstand $R_f$ angeschlossen ist, dann ist die Ausgangsspannung $U_a = -U_s\frac{R_f}{R_s}$. Dabei gilt bei Stromquellen meistens, dass $R_s > R_f$ ist, ist $U_a$ kleiner als $U_s$. Formal koppelt eine Rauschspannung an den nichtinvertierenden Eingang des Operationsverstärkers des Strom/Spannungswandlers. Damit ist die Verstärkung der Rauschspannung $U_r$ durch $U_{a,r} = U_r\left(1+\frac{R_f}{R_s}\right)$. Deshalb wird bei Strom/Spannungswandlern bei kleinen Quellwiderständen $R_f$ das Rauschen überproportional verstärkt. Deshalb empfiehlt Keithley [**?**] als minimale Quellwiderstände

| Strombereich | Minimale Quellwiderstände |
|---|---|
| pA | $1G\Omega \ldots 100G\Omega$ |
| nA | $1M\Omega \ldots 100M\Omega$ |
| $\mu$A | $1k\Omega \ldots 100k\Omega$ |
| mA | $1\Omega \ldots 100\Omega$ |

Besonders bei der Rastertunnelmikroskopie bei kleinen Strömen und kleinen Spannungen[21] kann diese Empfehlung nicht eingehalten werden. Weiter sind zu den vorhandenen Widerstände immer Parallelkapazitäten vorhanden. In diesem Falle muss mit den Beträgen der Impedanzen gerechnet werden.

**Offsetströme** Offsetströme entstehen wie bei den Spannungsmessungen durch die notwendigen Eingangsströme der Eingangsverstärker. Das für Spannungsmessungen gesagte gilt analog auch für die Strommessungen. Wenn die Offsetströme zeitlich konstant sind, können sie auch kompensiert werden. Besonders einfach ist dies mit einer externen Stromquelle.

**Triboelektrische Effekte** Triboelektrische Ströme entstehen, wenn unterschiedliche Materialien sich gegeneinander bewegen und Reibungskräfte vorhanden sind. Triboelektrizität entsteht vor allem in Kabeln, deren Biegung wechselt oder die unterschiedlichen mechanischen Spannungen ausgesetzt sind. Die Auswahl der Materialien, aus denen die Kabel zusammengesetzt sind, beeinflusst die Höhe der generierten triboelektrischen Ströme.

[21] Wenn mit einem STM Supraleiter untersucht werden sollen, dann ist die angelegte Spannung meistens bei 1 mV und der Tunnelstrom bei $1 \ldots 5$ nA. Dann ist die Bedingung, die Keithley[**?**] angibt, nicht erfüllt

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

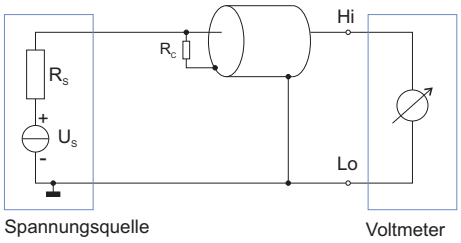



Abbildung 4.186.: Einfluss des Kabelleckstromes durch $R_C$ auf eine Spannungsmessung

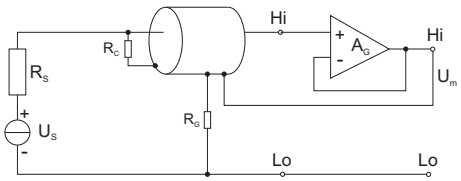



Abbildung 4.187.: Messung kleiner Spannungen mit einer Guard-Konfiguration

**Piezoelektrische Effekte, gespeicherte Ladung** Wenn an gewisse Materialien mit nicht zentrosymmetrischem Kristallbau mechanische Spannungen angelegt werden, entstehen durch den Piezoeffekt Ladungen. Bei periodischer mechanischer Beanspruchung der Isolationsmaterialien bewirkt dies auch einen Wechselstrom im Takt der mechanischen Anregung. Um piezoelektrische Effekte zu vermeiden sollten Isolatoren nicht mechanisch belastet werden.

**Elektrochemische Effekte** Wenn die Oberflächen von Isolatoren verschmutzt sind kann Strom über elektrochemische Prozesse geleitet werden. Flussmittelrückstände oder Rückstände von Lösungsmitteln wie auch Schmutz und Fett von Fingern sind vielfach auf Isolatoren zu finden. Bei ungeeigneter Kombination dieser Rückstände mit dem Basismaterial können so auch Lokalelemente, also Batterien entstehen. Diese Effekte können mit Guard-Techniken minimiert werden.

## 4.4.4. Techniken zur Verhinderung von Fehlmessungen

Die im folgenden beschriebenen Techniken zur Kompensation oder Verhinderung von Fehlern stammen einerseits aus dem Handbuch von Keithley[**?**] und andererseits aus der eigenen Erfahrung.

### 4.4.4.1. Einfluss von Schirmungen

Abbildung 4.186 zeigt die Messung von Spannungen mit einem Quellwiderstand, bei dem der Isolationswiderstand des Kabels nicht mehr vernachlässigt werden kann. Die gemessene Spannung ist

$$U_M = U_S \left( \frac{R_C}{R_S + R_C} \right) \tag{4.4.2}$$

Wenn man das die Schirmung des Kabels mit einem Operationsverstärker auf dem

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

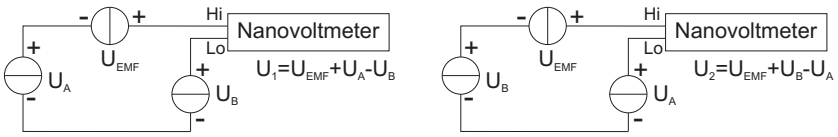



Abbildung 4.188.: Kompensation von ungewollten Thermospannungen durch Vertauschen der Anschlüsse

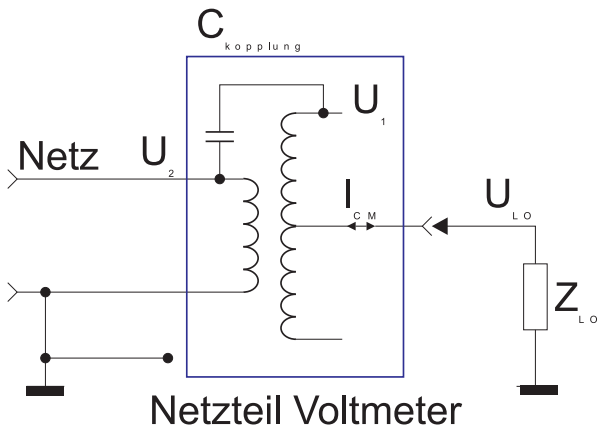



Abbildung 4.189.: Gleichtaktstrom hervorgerufen durch die Ankopplung an die Netzspannung

Potential der zu messenden Spannung hält, dann wird

$$U_M = U_S \left( \frac{A_G R_C}{R_S + A_G R_C} \right) \tag{4.4.3}$$

wobei $A_G$ die Verstärkung des Operationsverstärkers bei der betrachteten Frequenz ist.

#### 4.4.4.2. Thermospannungen

Wenn eine ungewollte Thermospannung $U_{EMF}$ oder eine andere das Vorzeichen behaltenden Störspannung bei der Messung der Differenzspannung von zwei Thermoelementen $U_A$ und $U_B$ stört, kann man mit zwei Messungen mit jeweils vertauschten Kabeln diesen Einfluss kompensieren.

$$\begin{aligned} U_1 &= U_{EMF} + U_A - U_B \\ U_2 &= U_{EMF} + U_B - U_A \\ \frac{U_1 - U_2}{2} &= \frac{U_{EMF} + U_A - U_B - (U_{EMF} + U_B - U_A)}{2} = U_A - U_B \end{aligned} \tag{4.4.4}$$

Die letzte Zeile von Gleichung (4.4.4) gibt das Schlussresultat.

#### 4.4.4.3. Störungen in Netzteilen

Abbildung 4.189 zeigt wie durch die Ankopplung an die Netzspannung ein Gleichtaktstrom hervorgerufen wird. Die Kapazität $C_{Kopplung}$ koppelt die Netzspannung

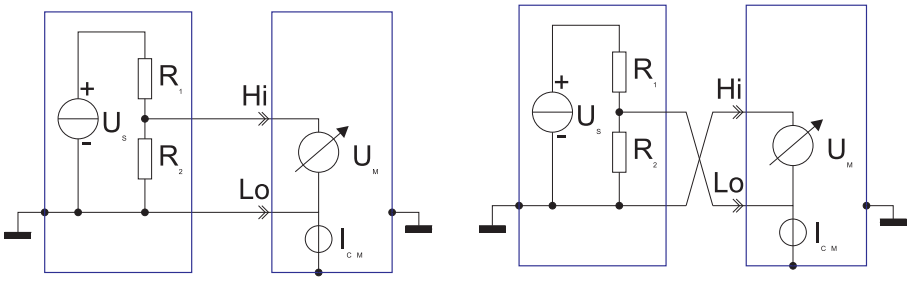



Abbildung 4.190.: Gleichtaktstrom hervorgerufen durch den falschen Anschluss der Messkabel

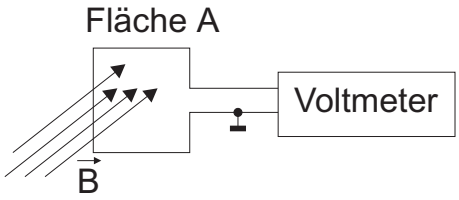



Abbildung 4.191.: Durch Magnetfelder induzierte Spannungen

von der Primärseite auf die Sekundärseite. Der durch diese Kapazität in der Sekundärseite induzierte Strom ist

$$I_{CM} = 2\pi f C_{Kopplung}(U_2 \pm U_1) \tag{4.4.5}$$

Die magnetiche Kopplung zwischen der Primär- und der Sekundärseite wird optimal, wenn die beiden Wicklungen abwechselnd übereinandergelegt werden. Dies führt aber zu einer grossen Koppelkapazität, die das Störspannungsniveau auf der Sekundärseite erhöht. Zusätzlich sind die notwendigen Kriechwege für Ströme bei einer solchen, magnetisch effiziente Schaltung ungenügend lang.

#### 4.4.4.4. Fehler durch falschen Anschluss der Messkabel

Wenn, wie in Abb. 4.190 rechts das Messkabel an die Erde des Messobjektes und das Nullkabel an die empfindliche Stelle angeschlossen wird, fliesst vom Punkt 'Lo' ein Strom zur Erde. Der korrekte Anschluss in der Abbildung links vermeidet diesen Fehlerstrom.

#### 4.4.4.5. Verringerung des Einflusses von zeitabhängigen Magnetfeldern

Der durch eine (veränderliche) Fläche $A$ fliessende (veränderliche) Fluss induziert die Spannung

$$U_B = \frac{d\phi}{dt} = \frac{d\boldsymbol{B}\boldsymbol{A}}{dt} = \boldsymbol{B}\frac{d\boldsymbol{A}}{dt} + \frac{d\boldsymbol{B}}{dt}\boldsymbol{A} \tag{4.4.6}$$

Der Einfluss von Magnetfeldern $\boldsymbol{B}$ kann minimiert werden, wenn

- Man minimiert die Fläche $\boldsymbol{A}$
- Bei der Kabelführung sollen Bereiche mit hohen Magnetfeldern $\boldsymbol{B}$ gemieden werden.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

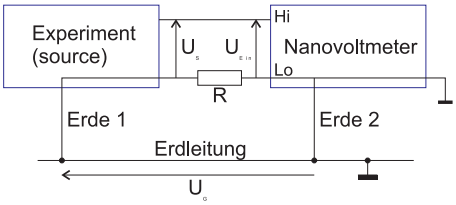



Abbildung 4.192.: Erdschleifen hervorgerufen durch mehrfache Erdung

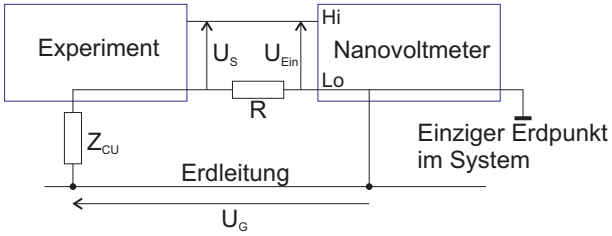



Abbildung 4.193.: Verhinderung des Einflusses von Erdströmen durch Erdung an einem einzelnen Punkt

- Der Betrag und die Orientierung der Fläche $\boldsymbol{A}$ müssen konstant gehalten werden. Insbesondere müssen mechanische Vibrationen und die Bewegung der Kabel vermieden werden.
- Der Betrag und die Richtung des Magnetfeldes $\boldsymbol{B}$ muss konstant gehalten werden.

#### 4.4.4.6. Erdschleifen und ihre Verhinderung

Wenn wie in Abb. 4.192 in der Erdleitung eine Spannung von $U_G$ induziert wird (zum Beispiel durch hohe Ströme oder durch magnetische Induktion) dann liegt dadurch am Leitungswiderstand $R$ des unteren Messkabels die Spannung $U_G$. Die gemessene Spannung ist dann

$$U_{ein} = U_S + U_G = U_S + IR \tag{4.4.7}$$

Dabei haben die Grössen typischerweise die Werte $R \approx 100m\Omega$, $I \approx 1A$. Dabei kann $U_G = IR$ sehr viel grösser als $U_S$ sein.
Abb. 4.193 zeigt eine Erdung an einem einzelnen Punkt. Wieder gilt

$$U_{ein} = U_S + IR \tag{4.4.8}$$

wobei aber jetzt der Strom $I$ durch den Isolationsimpedanz $Z_{CM}$ fliesst. Deshalb ist er nicht in der Grössenordnung von Ampères, sondern um Nanoampères. Somit werden Erdfehler vermieden.

#### 4.4.4.7. Kapazitive Fehler in geschirmten Kabeln

Die durch eine elektrostatische Störquelle $U$ (Abb. 4.194 induzierte Spannung ist

$$I = C\frac{dU}{dt} + U\frac{dC}{dt} \tag{4.4.9}$$

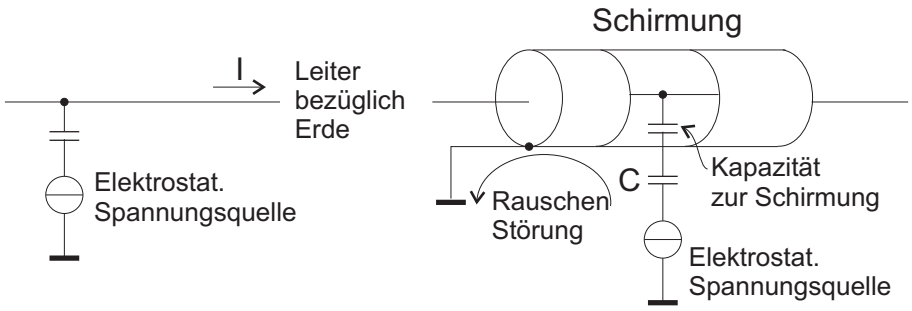



Abbildung 4.194.: Elektrostatische Kopplung (links) und elektrostatische Abschirmung

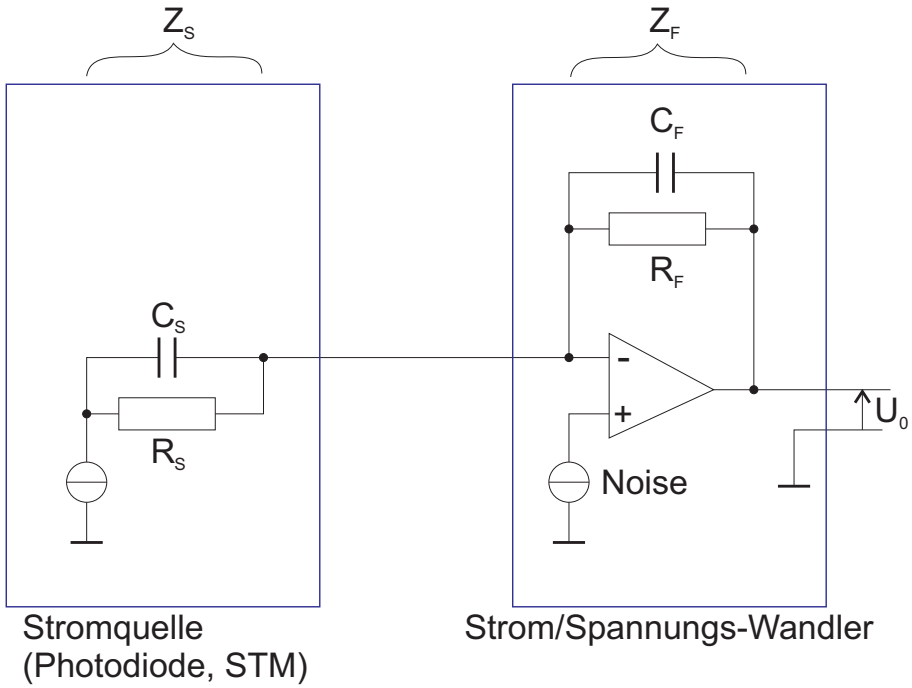



Abbildung 4.195.: Einfluss des Ausgangswiderstandes und der kapazitiven Belastung einer Stromquelle

Durch die elektrostatische Schirmung wie in Abb. 4.194, rechts, wird die Kopplung verhindert.

#### 4.4.4.8. Kapazitiv belastete Stromquelle

Bei Stromquellen wie zum Beispiel bei Rastertunnelmikroskopen oder bei Photodioden beeinflusst der Ausgangswiderstand $R_S$ und die Ausgangskapazität $C_S$ der Quelle sowie der Rückkopplungswiderstand $R_F$ und die Störkapazitäten $C_F$ des Strom/Spannungswandlers die Messresultate (Sie Abb. 4.195). Die Rauschspannung des Operationsverstärkers vergrössert sich so um

$$U_{Noise,Output} = U_{Noise,Input}\,(1 + R_F/R_S) \tag{4.4.10}$$

Das heisst, das Ausgangsrauschen vergrössert sich bei einer Verkleinerung des Quellwiderstandes $R_S$. Zum Beispiel heisst das, dass in einem Rastertunnelmikroskop bei kleineren Distanzen das Rauschen zunimmt. Auch durch die Kapazitäten

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

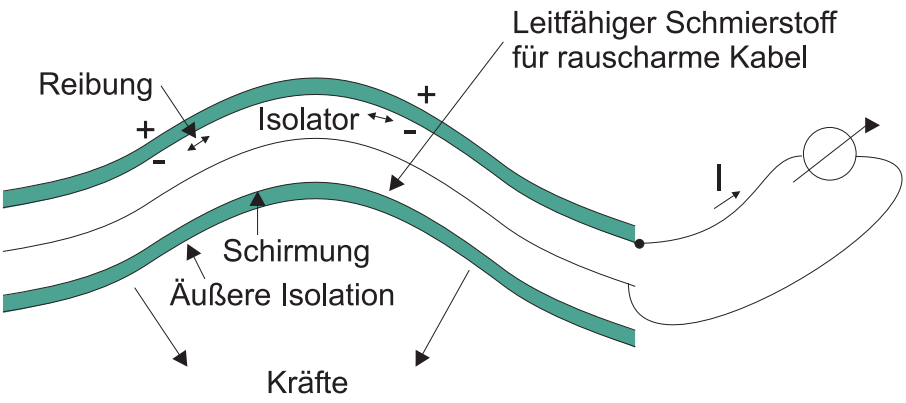



Abbildung 4.196.: Triboelektrische Effekte durch die Verbiegung von Kabeln

nimmt die Rauschspannung zu.

$$U_{Noise,Output} = U_{Noise,Input}\,(1 + Z_F/Z_S) \tag{4.4.11}$$

Hier ist $Z_F$ die Impedanz der Parallelschaltung aus $R_F$ und $C_F$.

$$Z_F = \frac{R_F}{\sqrt{(2\pi f C_F R_F)^2 + 1}} \tag{4.4.12}$$

Analog gilt für $Z_S$

$$Z_S = \frac{R_S}{\sqrt{(2\pi f C_S R_S)^2 + 1}} \tag{4.4.13}$$

oder zusammen:

$$U_{Noise,Output} = U_{Noise,Input}\left(1 + \frac{R_F}{R_S}\sqrt{\frac{(2\pi f C_F R_F)^2 + 1}{(2\pi f C_S R_S)^2 + 1}}\right) \tag{4.4.14}$$

#### 4.4.4.9. Triboelektrische Effekte in abgeschirmten Kabeln

Handelsübliche Messkabel sind aus mehreren Schichten aufgebaut. Wenn Kabel verbogen werden (Abb. 4.196) gleiten diese Schichten aneinander vorbei. Wie wenn man Kunststoffe mit Fellen reibt, entstehen auch bei Kabeln Ladungen. Da die Ladungen durch die Oberflächenwiderstände abgebaut werden und durch Bewegung wieder erzeugt werden. Periodische Bewegungen erzeugen also auch periodische Spannungen oder Ströme, die wiederum empfindliche Messungen stören oder gar verunmöglichen können.
Um triboelektrische Effekte zu vermeiden sollten für Experimente

- nur rauscharme Kabel verwendet werden
- mechanische Vibrationen an der Quelle bedämpft werden
- die Messkabel fixiert sein

Mit diesen Vorsichtsmassnahmen hat man optimale Messbedingungen.

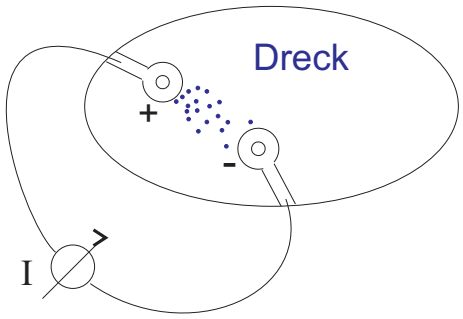



Abbildung 4.197.: Elektrochemische Effekte und Leckströme an Oberflächen

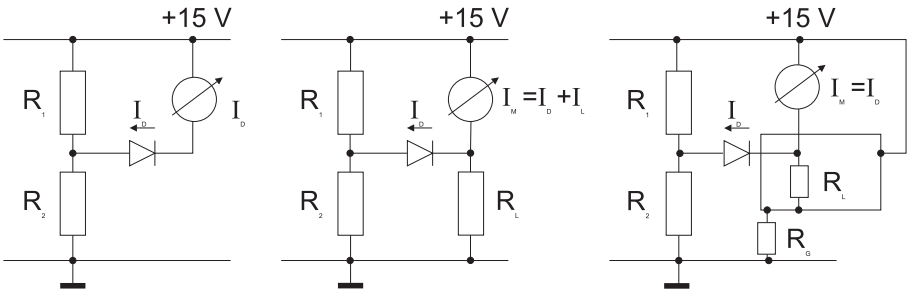



Abbildung 4.198.: Messung des Sperrstromes einer Diode. Links der Originalaufbau, in der Mitte der Messaufbau mit Leckströmen und rechts der Aufbau mit Guard-Ringen

#### 4.4.4.10. Leckströme an Oberflächen, Verwendung von 'Guard'-Ringen

In Abb. 4.198 zeigt die Messung des Leckstromes einer Diode. Wenn das Ampèremeter in der obigen Abbildung (Mitte) einen Isolationswiderstand von $1G\Omega$ hat, dann fliesst ein Leckstrom von $15nA$. Der Guard-Ring in der Abbidung rechts verringert die am Isolationswiderstand von $1G\Omega$ liegende Spannung auf $200\mu V$. Dadurch wird der Leckstrom etwa $0.2pA$. Der ursprüngliche Leckstrom von $15nA$ wird durch die $15V$-*Spannungsquelle* geliefert.

#### 4.4.4.11. Spektrum der Störsignale

Abbildung 4.199 zeigt ein typisches Spektrum von Störstrahlungen. Das Störstrahlungsspektrum enthält die folgenden Bestandteile

**1/f-Rauschen** Bei sehr tiefen Frequenzen dominiert das *1/f-Rauschen*. Diese Art Rauschen, im Abschnitt 2.8.2.3 besprochen, rührt von statistischen Schwankungen beim Stromtransport her. Typischerweise hat das 1/f-Rauschen über 1 kHz keine Bedeutung mehr.

**Weisses Rauschen** Oberhalb von 1 kHz ist *weisses Rauschen* dominierend. Dieses rührt von den statistischen Schwankungen der Ladungsträgerkonzentration her (siehe Abschnitt 2.8.1).

**Netzinterferenzen** Zwischen 50 Hz (USA und andere: 60 Hz) und etwa einem kHz macht sich die *Netzfrequenz* und ihre *Oberfrequenz*en störend bemerkbar. Bei Netzteilen ist weniger die Komponente bei 50 Hz als vielmehr wegen der Gleichrichtung die Komponente bei 100 Hz wichtig.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

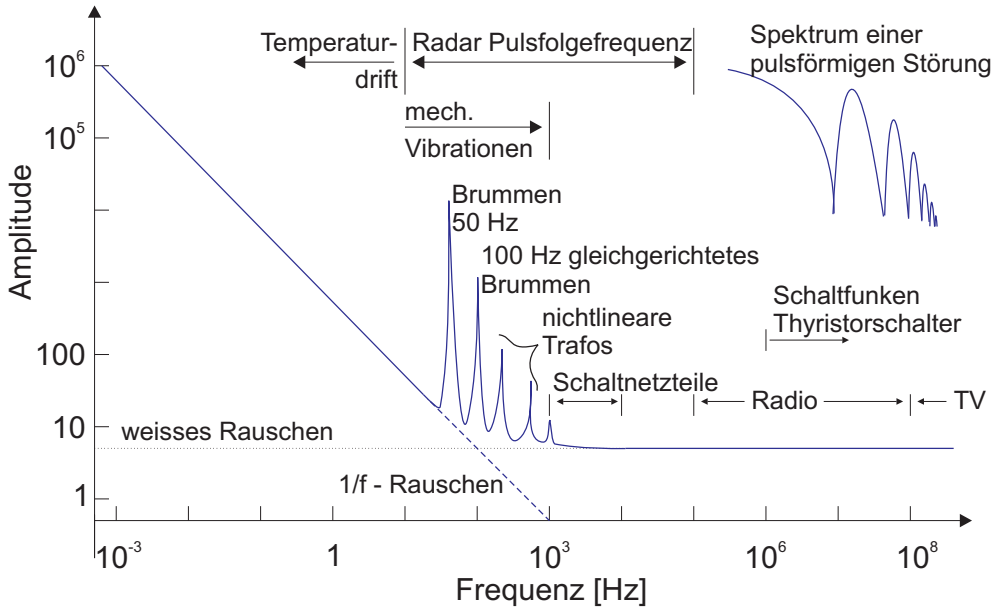



Abbildung 4.199.: Störspektrum in Schaltungen

**Mechanische Vibrationen** Die mikroskopisch kleinen Bewegungen der Erde, von Gebäuden und von Geräten können die Funktion von empfindlichen Messgeräten stören. Immer da wo eine kapazitive oder induktive Kopplung zwischen mehreren Schaltkreisen vorliegt, induzieren *mechanische Vibrationen* Störungen. Bei *optischer Datenübertragung* bewirkt eine mechanische Bewegung, dass die auf den Detektor fallende Lichtintensität schwankt.

**Temperaturdrift** Die *Temperaturdrift* beeinflusst Messungen unter 10 Hz. Durch den Tagesgang der Temperatur, durch die Änderung des Abstandes der ExperimentatorIn können *Temperaturschwankung*en von einigen Bruchteilen von Kelvin induziert werden. Dies reicht, um bei sehr empfindlichen Messungen Störungen hervorzurufen.

**Radarpulsfolgen** Zwischen 10 Hz und etwa 100 kH können Störungen durch die Pulsfolgefrequenz von Radargeräten auftreten. Diese dürften in der Nähe von Flugplätzen und von Flugkontrolleinrichtungen häufiger auftreten. Radargeräte sind als Störquellen nicht zu vernachlässigen, da es Geräte gibt, die mit einigen 10 kW mittlerer Leistung senden. Die Spitzenleistungen sind dann im Megawatt-Bereich.

**Schaltnetzteile** Schaltnetzteile sind bei modernen Geräten eine oft übersehene Qualle von Störungen. Da die Schaltfrequenzen zwischen 1 kHz und einigen 10 kHz liegen, sind sie oftmals schwer von Nutzsignalen zu trennen.

**Rundsteuerungen der Elektrizitätswerke** Elektrizitätswerke schalten mit Rundsteueranlagen grosse Verbraucher in Nebenzeiten ein und zu den Hauptlastzeiten wieder aus. Die Steuerung erfolgt mit einem *Signal* von etwa 1000 Hz und einer Amplitude von 20 V bis 30 V. Diese Störspannungen koppeln hervorragend durch die Kapazität zwischen der Primärwindung und der Sekundärwindung auf die restliche Schaltung. Sie sind nur schwer herauszufil-

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

tern. Die kommende Datenübertragung über Spannungsnetze könnte einen ähnlichen Einfluss auf empfindliche Geräte haben.

**Thyristorschalter** Thyristorschalter haben eine sehr grosse Flankensteilheit beim Schalten. Entsprechend wird durch *Thyristoren* vor allem das Spektrum über 1 MHz gestört.

**Radiowellen** Radiosender strahlen zwischen 100 kHz bis zu 100 MHz elektromagnetische wellen ab. Die Sendeleistung reicht von wenigen Watt bis zu Megawatt. Die leistungsfähigen Sender können in einem weiteren Umfeld massive Gerätestörungen bewirken. So ist es zum Beispiel im abstand von einigen 100 m von einem Mittelwellensender (500 kHz, 500 kW) möglich, Fluoreszenzlampen mit dem abgestrahlten elektrischen Feld zum leuchten zu bringen[22].

**TV** Fernsehsender strahlen Störsignale im Bereich über 100 MHz ab.

**Mobiltelefone** Mobiltelefone arbeiten mit Frequenzen von einigen GHz. Da die Daten in Paketen abgesendet werden, können die Störungen durch nichtlineare Effekte auch bei tieferen Frequenzen auftreten.

Beim Bau und Betrieb von Messgeräten muss sichergestellt werden, dass diese weder elektromagnetische Strahlung über den zugelassenen Werten abstrahlen noch durch elektromagnetische Strahlung gestört werden.

## 4.5. Lichtquellen für optische Messverfahren

*Optische Messverfahren* werden in der modernen Laborpraxis immer wichtiger. Sie ermögliche eine extrem hohe zeitliche und spektrale (energetische) *Auflösung*. Licht wird auf seinem Weg durch andere elektromagnetische Strahlung nicht beeinflusst, sofern die Intensitäten klein genug sind, dass keine nichtlinear-optischen Effekte auftreten.
Im folgenden Abschnitt wird auf die Grundlagen der Lasertechnik eingegangen. Es folgt eine Darstellung von Systemen zur Erzeugung ultrakurzer Pulse. Diese werden einerseits mit Pump-Probe-Techniken und andererseits mit elektrooptischen Verfahren detektiert. Schliesslich werden Geräte zur Messung von Spektren besprochen.

### 4.5.1. Grundlagen der Lasertechnik

Wenn sich Materie in optisch angeregten Zuständen befindet, wird diese Anregung durch Emission abgebaut. Wenn es gelänge, alle Atome oder Moleküle in einem bestimmten Volumen kohärent strahlen zu lassen, dann würde man eine Lichtquelle mit einzigartigen Eigenschaften gewinnen.
Der Laser, am Anfang der 60-er Jahre erfunden wurde, erfüllt genau diese Bedingungen. Die Abbildung 4.200 zeigt den schematischen Aufbau. Ein aktives Medium befindet sich in einem *Fabry-Perot-Resonator*[**?**][**?**]. Das Licht im *Resonator* wird

[22] Da diese Sender mit einer sehr hohen Güte gefahren werden, stört die Anwesenheit einer leuchtenden Fluoreszenzröhre den Sender so stark, dass eine Verstimmung der Schwingkreise resultiert.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

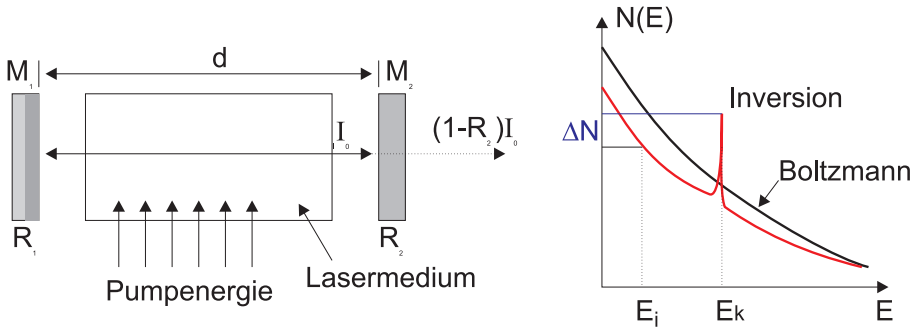



Abbildung 4.200.: Aufbau eines Lasers (links) sowie schematische Darstellung der Inversion im Vergleich zur thermischen Verteilung.

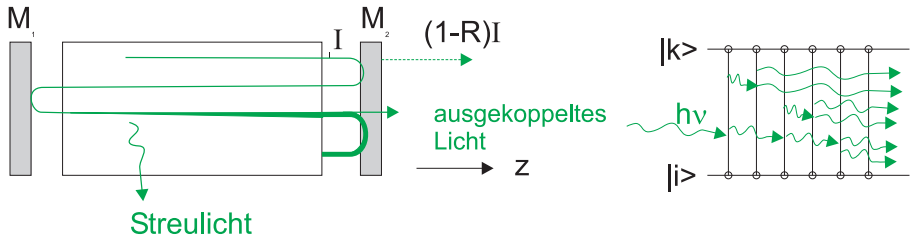



Abbildung 4.201.: Schematische Darstellung der Verstärkung und der Verluste in einem *Resonator*

durch das aktive Medium bei jedem Durchgang verstärkt. Die Verstärkung erfolgt durch stimulierte Emission. Ein kleiner Teil des Lichtes wird durch die Spiegel des Fabry-Perot-Resonators ausgekoppelt und steht für Experimente zur Verfügung.
Die rechte Seite der Abb. 4.200 zeigt die Besetzungsverteilung. Im Vergleich zu einer thermischen Verteilung, gegeben durch die Boltzmannverteilung $N(E) = \exp(-E/kT)$, sind die Zustände bei hohen Energien deutlich stärker besetzt als im thermischen Fall. Diese sogenannte Besetzungsinversion ist für die Funktionsweise des Lasers notwendig.
Die Diskussion der Wirkungsweise von Lasern beruht auf dem exzellenten Lehrbuch von Demtröder[?].

#### 4.5.1.1. Schwellwertbedingung

Um die Intensität der in $z$-Richtung laufenden Welle in Abb. 4.201 zu berechnen setzen wir für die Intensität an

$$I(\nu, z) = I(\nu, z = 0)e^{-\alpha(\nu)z} \tag{4.5.1}$$

$I(\nu, z = 0)$ ist dabei die Refernzintensität bei $z = 0$.
Hier ist der frequenzabhängige Absorptionskoeffizient durch

$$\alpha(\nu) = [N_i - (g_i/g_k)N_k]\,\sigma(\nu) \tag{4.5.2}$$

gegeben. $\alpha\nu$ hängt von den Besetzungsdichten $N_i$ des unteren Laserniveaus und $N_k$ des oberen Laserniveaus, von den statistischen Gewichten $g_i$ und $g_j$[23] und vom

[23] Für einen elektronischen Zustand $E_i$ eines freien Atoms mit der Drehimpulsquantenzahl $J$ ist $g_i = 2J + 1$

optischen Wirkungsquerschnitt $\sigma\nu$ ab.
Wenn $(g_i/g_k)N_k > N_k$ ist, wird der Absorptionskoeffizient in Gleichung (4.5.2) negativ. Aus der Dämpfung ist also, analog wie bei der Phasendrehung von Operationsverstärkern, eine Verstärkung geworden. Der Verstärkungsfaktor ist

$$G_0(\nu, z) = \frac{I(\nu, z)}{I(\nu, z = 0)} = e^{-\alpha(\nu)z} \tag{4.5.3}$$

Die gesamte Abschwächung oder, bei negativen Werten von $\gamma$ kann in eine Gleichung mit einem Exponentialfaktor zusammengefasst werden.

$$I/I_0 = e^{-\gamma} \tag{4.5.4}$$

In der Regel wird das zur Verstärkung verwendete optische Medium in einen *Resonator* gebracht (analog zur Abb. 4.201, links). An den beiden Endspiegeln treten Verluste auf. Einerseits ist es nicht möglich, einen Spiegel mit einer Reflektivität von 100% zu bauen, der zudem noch eine unendliche Ausdehnung hat um Beugungsverluste zu minimieren. Andererseits muss an einem Spiegel die Reflektivität kleiner als 1 sein, damit Laserlicht ausgekoppelt werden kann. Die Verstärkung, Beugungs-, Auskopplings- Reflexionsverluste beim beim Durchgang durch einen *Resonator* können als Intensitätsänderung pro Umlauf geschrieben werden

$$G = I/I_0 = exp[-2\alpha(\nu)L - \gamma] \tag{4.5.5}$$

Bei der Berechnung der Verstärkung nach einem Umlauf ist angenommen worden, dass das Medium die Länge $L$ hat. Wenn $G$ grösser als 1 ist, beginnt die stimulierte Emission im Lasermedium die spontane Emission zu dominieren. Damit dies möglich ist, muss $-2\alpha(\nu)L > \gamma$ sein. Zusammen mit Gleichung (4.5.2) bekommt man die Schwellwertbedingung

$$\Delta N = N_k(g_i/g_k) - N_i > \Delta N_S = \frac{\gamma}{2\sigma(\nu)L} \tag{4.5.6}$$

für die minimale Besetzungsinversion $\Delta N_S$.
Die Laseremission beginnt immer mit einer spontanen Emission aus dem oberen Laserniveau in eine Resonatormode. Dabei werden die Photonen, deren Frequenz nahe der Resonator-Mittenfrequenz liegt, bevorzugt verstärkt. Durch die beginnende stimulierte Emission wird wird die Besetzungsinversion abgebaut bis ein Gleichgewicht erreicht wird. Unabhängig von der Pumpleistung ist die Inversion in einem Laser beim stationären Betrieb immer gleich der Schwellwertinversion $\Delta N_S$.

#### 4.5.1.2. Die Bilanzgleichungen

Der stationäre Laserbetrieb kann durch Bilanzengleichungen beschrieben werden. Anhand des Termschemas in Abb 4.202 ist ersichtlich, dass aus einem Pumpprozess $P$ das obere Laserniveau $|2>$ gespiesen wird. Zusätzlich wird die Besetzungszahl dieses Niveaus durch die Absorption aus dem unteren Laserniveau $|1>$ mit der Rate $N_1 B_{12} \cdot n \cdot h \cdot \nu$ erhöht. Es gibt drei Verlustkanäle, die spontane Emission mit der Rate $N_2 A_{21}$, die induzierte Emission mit der Rate $N_2 B_{21} \cdot n \cdot h \cdot \nu$ und die verlustrate $N_2 R_2$, zum Beispiel in Triplettzustände. Das untere Laserniveau $|1>$ wird durch den Relaxationsprozess mit der Rate $N_1 R_1$ entvölkert.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

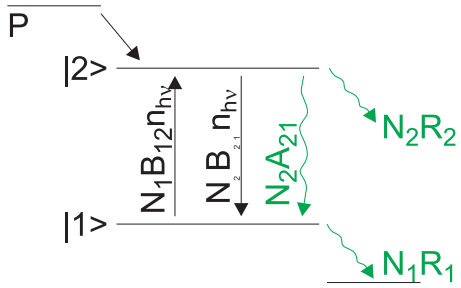



Abbildung 4.202.: Funktion eines Lasers: Pumpprozess P, Relaxationsraten, induzierte und spontane Emission.

Wenn man annimmt, dass die statistischen Gewichte gleich sind ($g_1 = g_2$), bekommt man die Ratengleichungen

$$\frac{dN_1}{dt} = (N_2 - N_1)B_{21}nh\nu + N_2A_{21} - N_1R_1 \tag{4.5.7}$$

$$\frac{dN_2}{dt} = P - (N_2 - N_1)B_{21}nh\nu + N_2A_{21} - N_2R_2 \tag{4.5.8}$$

$$\frac{dn}{dt} = -\beta n + (N_2 - N_1)B_{21}nh\nu \tag{4.5.9}$$

Der Laserresonator hat seine eigene Verlustrate. Wenn man $N_1 = N_2$ setzt erhält man aus (4.5.9) den Verlustfaktor $\beta$

$$n = n_0e^{-\beta t} \tag{4.5.10}$$

Durch Vergleich erhält man für den Verlustfaktor $\gamma$

$$\gamma = \beta T = \beta(2d/c) \tag{4.5.11}$$

wobei $d$ die Resonatorlänge ist.
Im stationären Betrieb müssen die in den obigen Gleichungen vorkommenden Ableitungen verschwinden. Aus den Gleichungen (4.5.7) und (4.5.8) bekommt man in diesem Falle

$$P = N_1R_1 + N_2R_2 \tag{4.5.12}$$

Die Pumprate muss also im stationären Betrieb die beiden Verlustraten $N_1R_12$ und $N_2R_2$ aus dem unteren, beziehungsweise aus dem oberen Laserniveau ausgleichen. Andererseits bekommt man durch Addition aus (4.5.8) und (4.5.9) die Gleichung

$$P = \beta n + N_2(A_{21} + R_2) \tag{4.5.13}$$

Die Pumprate $P$ ersetzt also die Resonatorverluste (4.5.11) sowie die durch spontane Emission und Relaxation aus dem oberen Laserniveau verschwindenden Photonen. Die Relaxationsrate des unteren Niveaus ist im stationären Betrieb

$$N_1R_1 = N_2A_{21} + \beta n \tag{4.5.14}$$

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

Sie kompensiert gerade die spontane Emission und die Verlustrate der induzierten Photonen. Deshalb ist sie immer grösser als die Auffüllrate aus dem Niveau $|2>$ durch spontane Emission.
Wir multiplizieren Gleichung (4.5.7) mit $R_2$ und Gleichung (4.5.8) mit $R_1$ und können für den stationären Zustand ($d/dt = 0$) mit der Definition $\Delta N_{stat} = N_2 - N_1$ die folgende Umformung

$$\begin{aligned} 0 &= -R_1 P + (R_2 + R_1)(N_2 - N_1) B_{21} n h\nu + \\ &\quad (R_2 + R_1) N_2 A_{21} + (N_2 - N_1) R_1 R_2 \\ &= -R_1 P + (R_2 + R_1) \Delta N_{stat} B_{21} n h\nu + \\ &\quad (R_2 + R_1) N_2 A_{21} + \Delta N_{stat} R_1 R_2 \end{aligned} \tag{4.5.15}$$

durchführen.
Mit der Gleichung (4.5.12) erhält man die stationäre Besetzungsinversion

$$\Delta N_{stat} = \frac{(R_1 - A_{21}) P}{B_{21} n h\nu (R_1 + R_2) + A_{21} R_1 + R_1 R_2} \tag{4.5.16}$$

Aus (4.5.16) folgt, dass eine stationäre Besetzungsinversion $\Delta N_{stat} > 0$ nur für Medien mit $R_1 > A_{21}$ möglich ist. Dies bedeutet, dass das untere Laserniveau sich schneller entleeren muss als das obere sich durch spontane Emission entvölkert.
Im realen Laserbetrieb wird das untere Laserniveau zusätzlich durch die induzierte Emission bevölkert. Die Relaxationsrate des unteren Laserniveaus muss deshalb der Bedingung

$$R_1 > A_{21} + B_{21} \rho \tag{4.5.17}$$

genügen.

#### 4.5.1.3. Optische Resonatoren

Wenn der Energieverlust der k-ten Mode mit der Zeit wie

$$dE_k = -\beta_k E_k dt \tag{4.5.18}$$

ist dann ist

$$E_k(t) = E_k(0) e^{-\beta_k t} \tag{4.5.19}$$

Die Resonatorgüte ist als

$$Q_k \equiv -2\pi\nu \frac{E_k}{dE_k/dt} = 2\pi\nu/\beta_k \tag{4.5.20}$$

definiert. Für einen *Resonator* der Länge $d$ ist der Verlustfaktor durch

$$\gamma = (2d/c)\beta \tag{4.5.21}$$

gegeben. Der Verlustfaktor setzt sich aus Beugungsverlusten, Absorptionsverlusten, Reflexionsverluste und die Verluste durch Lichtstreuung zusammen.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

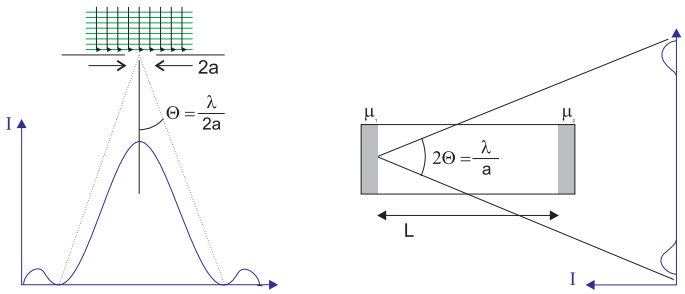



Abbildung 4.203.: Beugung einer ebenen Welle an einer Blende

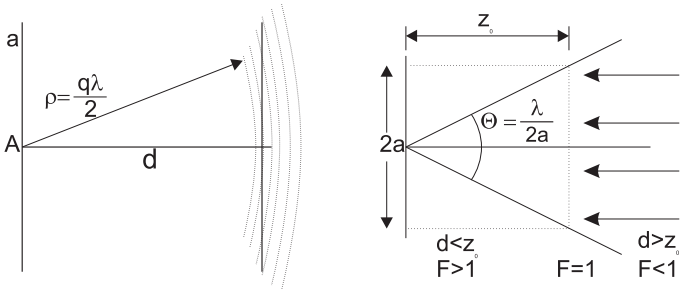



Abbildung 4.204.: Erklärung der Fresnelzahl

Intensität und Reflexionsverluste

$$I = I_0 R_1 R_2 = I_0 e^{-\gamma_R} \text{ mit } \gamma_R = -\ln(R_1 R_2) \tag{4.5.22}$$

Mit der Umlaufzeit $T = 2d/c$ wird die Abklingkonstante $\beta_R = \gamma_R/T = \gamma_R \cdot c/2d$. Die mittlere Verweilzeit der Photonen im *Resonator* ist

$$\tau = -\frac{2d}{c \ln(R_1 R_2)} \tag{4.5.23}$$

Die Beugung wird durch die Fresnel-Zahl charakterisiert.

$$F = a^2/(d\lambda) \tag{4.5.24}$$

Sie gibt an, wieviele Fresnelzonen auf dem gegenüberliegenden Spiegel entstehen, wenn man im Abstand $\rho_q = q\lambda/d$ (q ganzzahlig). Wenn $d < z_0$ ist, ist $F > 1$ und die Beugungsverluste minimal. Damit bei planparallelen Spiegeln ein Photon m-Umläufe machen kann, muss der Beugungswinkel $\Theta < a/(md)$ sein. Also muss

$$F > m \tag{4.5.25}$$

sein. Resonatoren mit der gleichen *Fresnelzahl* haben die gleichen Verluste.
Um die *Beugungsverluste* eines Resonators zu berechnen, kann man den *Resonator* durch eine Folge von Linsen und Blenden ersetzen (siehe Abb. 4.205). Dabei entsprechen ebene Spiegel einer Apertur. Gekrümmte Spiegel müssen entsprechend durch Sammel- oder Zerstreuungslinsen ersetzt werden. Aus der Abbildung 4.205 ist sofort ersichtlich, dass *ebene Wellen* keine Lösung des resonatorproblems sein können.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

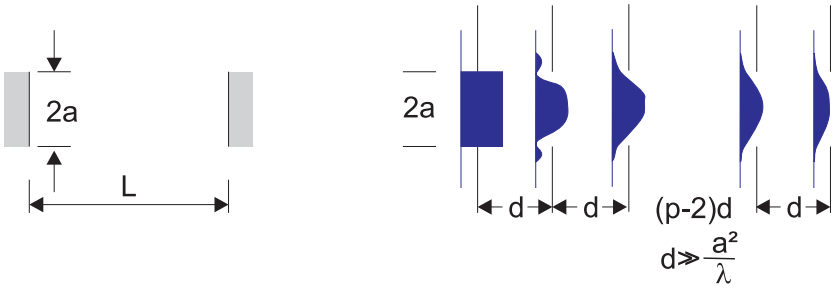



Abbildung 4.205.: Anschauliche Erklärung, dass ein ebener Spiegelresonator mit einer Folge von Blenden äquivalent ist.

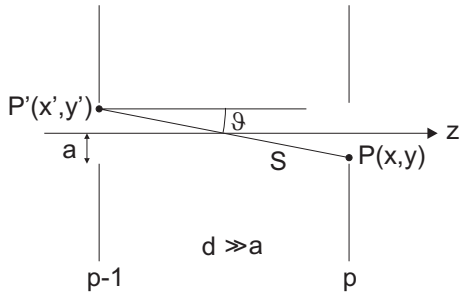



Abbildung 4.206.: Die Feldamplitude $P(x, y)$ kann aus den Amplituden in der Ebene $P'(x', y')$ bestimmt werden.

#### 4.5.1.4. Fourieroptik

Um die Beugungserscheinungen an einer Folge von Aperturen handhaben zu können, wird die *Kirchhoff-Fresnel*'sche Beugungstheorie auf die Geometrie in Abb. 4.205. Die Feldverteilung bei der $A_p$-ten Apertur wird aus der Feldverteilung in der $A_{p-1}$-ten Apertur mit Hilfe der Gleichungen der *Fourieroptik* berechnet.
Die Amplitude am Punkt $P(x, y)$ in der Apertur $A_p$ ist durch

$$A_p(x,y) = -\frac{j}{2\lambda}\int_{x'}\int_{y'} A_{p-1}(x',y')\frac{1}{\rho}e^{-jk\rho}(1+\cos\vartheta)dx'dy' \tag{4.5.26}$$

gegeben (Siehe Abb. 4.206). Die stationäre Feldverteilung muss die beiden folgenden Eigenschaften haben:

- Da der *Resonator* als lineares System betrachtet wird, wirken sich die Beugungsverluste als Multiplikation mit einem reellen Faktor $0 < \sqrt{1-\gamma_B} < 1$ aus.
- Der Lichtweg zwischen zwei Aperturen (Spiegeln) wird durch einen Phasenfaktor $e^{j\phi}$ beschrieben.

Für die Amplitude gilt also

$$A_p(x,y) = CA_{p-1}(x,y) \text{mit } C = e^{j\phi}\sqrt{1-\gamma_B} \tag{4.5.27}$$

wobei wie oben diskutiert, der Faktor $|C|^2 = 1-\gamma_B$ den ortsunabhängigen Intensitätsverlust durch Beugung beschreibt. Die Modenverteilung ist die Lösung der

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

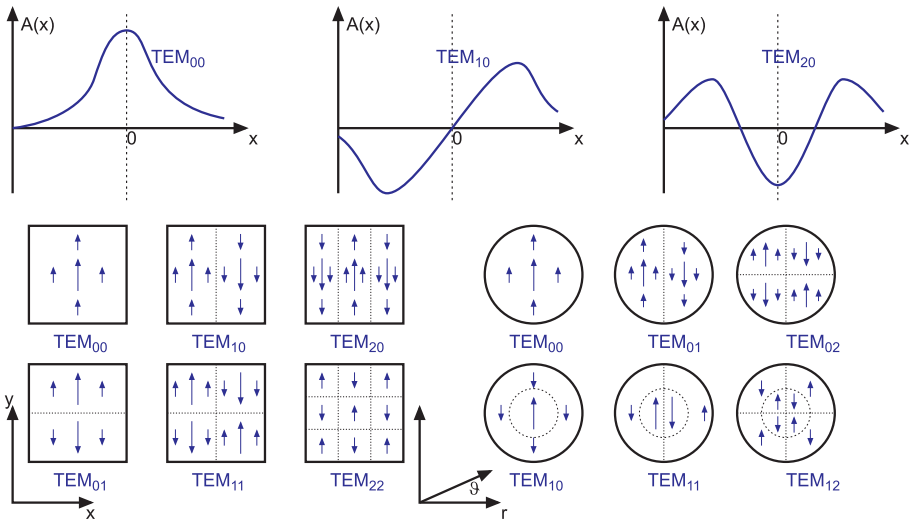



Abbildung 4.207.: Oben die eindimensionale Modenverteilung unten links die *Modenverteilung* in kartesischen Koordinaten und unten rechts in Zylinderkoordinaten.

Gleichung, die entsteht, wenn man (4.5.27) in (4.5.26) einsetzt. Diese Gleichungen sind im allgemeinen nicht analytisch lösbar.
Nur für den symmetrischen konfokalen Resonator kann eine Näherungslösung[?] angegeben werden. Dazu muss der Ursprung des Koordinatensystems in das Zentrum des Resonators gelegt werden. Dann ist für eine beliebige Ebene die Intensitätsverteilung

$$A_{m,n}(x,y,z) = C \cdot H_m(x^*) \cdot H_n(y^*) \cdot e^{-\left(x^{*2}+y^{*2}\right)/4} \cdot e^{-j\phi(x,y,z)} \tag{4.5.28}$$

$H_m$ und $H_n$ sind die Hermitschen Polynome $m$-ter und $n$-ter Ordnung. $C$ ist ein Normierungsfaktor und $x^* = \sqrt{2}\frac{x}{w}$ und $y^* = \sqrt{2}\frac{y}{w}$ sind normierte Koordinaten. Die Normierungsgrösse $w$ ist ein Mass der radialen Amplitudenverteilung und durch

$$w^2(z) = \frac{\lambda d}{2\pi}\left[1 + \left(\frac{2z}{d}\right)^2\right] \tag{4.5.29}$$

gegeben. $d$ ist hier die Länge des Resonators. Unter Verwendung der Abkürzung $\xi = 2z/d$ bekommt man für die Phase

$$\phi(x,y,z) = \frac{2\pi}{\lambda}\left[\frac{b}{2}\left(1+\xi^2\right) + \frac{\left(x^2+y^2\right)\xi}{d\left(1+\xi^2\right)}\right] - (1+m+n)\left(\frac{\pi}{2} - \arctan\frac{1-\xi}{1+\xi}\right) \tag{4.5.30}$$

Abbildung 4.207 zeigt einige Modenverteilungen. Sie werden *TEM-Moden* genannt, da sie in guter Näherung transversal-elektromagnetische Wellen darstellen. Die Zahlen $m$ und $n$ geben die Anzahl *Knoten* der Feldverteilung an.
Ist $n = m = 0$ so hat man die Grundmode. Ihre Intensitätsverteilung ist

$$I_{00}(x,y) = I_0 e^{-(x^2+y^2)/w^2} \tag{4.5.31}$$

Sie haben deshalb eine Gausssche Intensitätsverteilung. Die Grösse $w$ gibt an, bei welchem Radius die Intensität auf den Faktor $1/e^2$ bezogen auf das Strahlzentrum

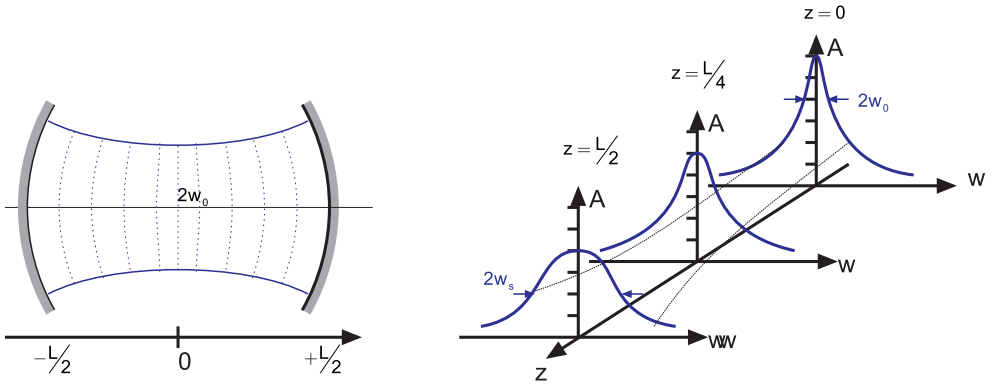



Abbildung 4.208.: Radiale Amplitudenverteilung in konfokalen Resonatoren

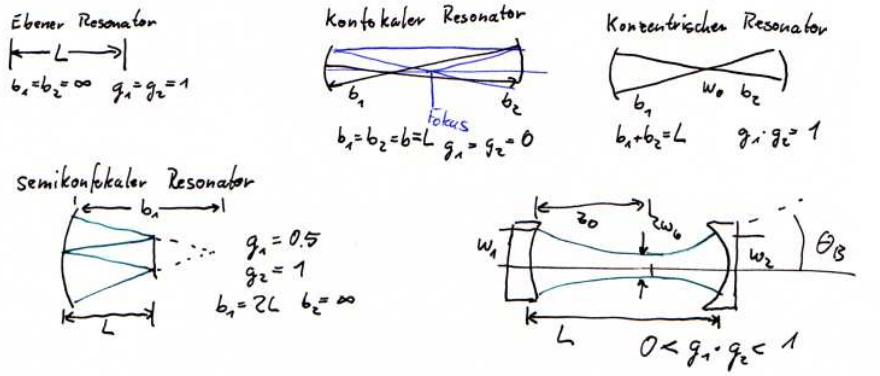



Abbildung 4.209.: Beispiele für Laserresonatoren

abgefallen ist. Der minimale Strahldurchmesser

$$w_0 = \sqrt{\lambda d/2\pi} \tag{4.5.32}$$

heisst auch Strahltaille. Eine exemplarische Amplitudenverteilung ist in der Abbildung 4.208 gezeigt. Rsonatoren, deren Spiegel sich in die Wellenfronten eines symmetrischen konfokalen Resonators einpassen lassen, können ebenfalls mid der hier gezeigten Theorie beschrieben werden.
Die Abbildung 4.209 zeigt Beispiele von *Laserresonator*en.
Die Beugungsverluste von offenen Resonatoren hängen von der betrachteten Lasermode ab. Abbildung 4.210 zeigt einen Graphen der Beugungsverluste. Als Ordinate ist die *Fresnel-Zahl* angegeben. Durch eine Verringerung der *Fresnel-Zahl* können die Verluste der höheren Modenordnungen so vergrössert werden, dass sie nicht mehr anschwingen können.
Die Stabilität eines Resonators folgt aus der Forderung, dass die Strahlparameter eines zu den Spiegeln passenden Gaussstrahls nach einem Umlauf auf sich selber abgebildet werden soll. Aus der Mathematik der Gaussstrahlen erhält man mit

$$g_i = 1 - \frac{d}{b_i} \tag{4.5.33}$$

den Durchmesser des Strahls auf den Spiegeln $M_1$ und $M_2$. Der Strahldurchmesser ist jeweils

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

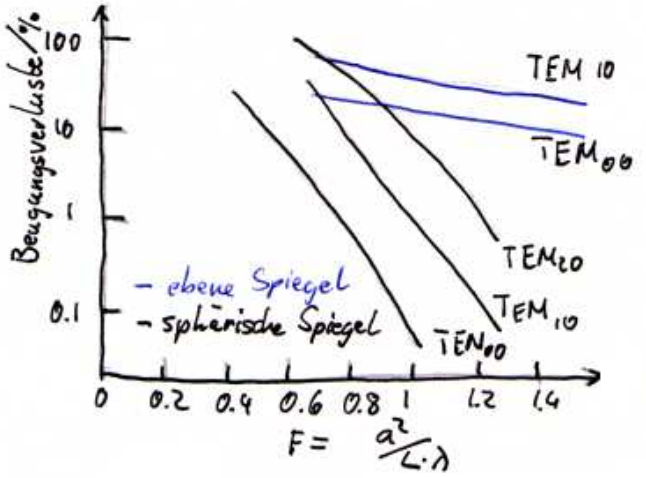



Abbildung 4.210.: Beugungsverluste von $TEM_{n,m}$-Moden

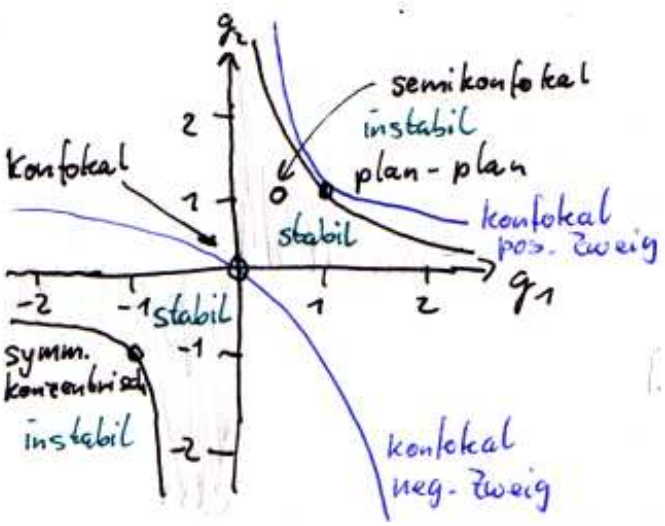



Abbildung 4.211.: Stabilitätsdiagramm für optische Resonatoren

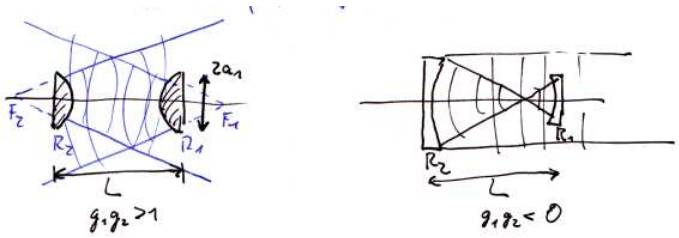


Abbildung 4.212.: Beispiele von instabilen Resonatoren

| Typ | Spiegelradien | Stabilitätsparameter |
|---|---|---|
| konfokal | $b_1 + b_2 = 2d$ | $g_1 + g_2 = 2g_1 \cdot g_2$ |
| konzentrisch | $b_1 + b_2 = d$ | $g_1 \cdot g_2 = 1$ |
| symmetrisch | $b_1 = b_2$ | $g_1 = g_2 = g$ |
| symmetrisch konfokal | $b_1 = b_2 = d$ | $g_1 = g_2 = 0$ |
| symmetrisch konzentrisch | $b_1 = b_2 = \frac{1}{2}d$ | $g_1 = g_2 = -1$ |
| semikonfokal | $b_1 = \infty$, $b_2 = 2d$ | $g_1 = 1$, $g_2 = \frac{1}{2}$ |
| eben | $b_1 = b_2 = \infty$ | $g_1 = g_2 = +1$ |

Tabelle 4.10.: Klassifizierung von Resonatoren nach Demtröder[**?**]. Die $b_i$ sind die Krümmungsradien der Spiegel, deren Abstand $b$ ist.

$$\pi w_1^2 = \lambda d \left( \frac{g_2}{g_1(1 - g_1 g_2)} \right)^{1/2} \tag{4.5.34}$$

$$\pi w_2^2 = \lambda d \left( \frac{g_1}{g_2(1 - g_1 g_2)} \right)^{1/2} \tag{4.5.35}$$

Also divergieren die Strahldurchmesser für $g_1 g_2 = 1$ sowie für $g_1 = 0$ und $g_2 = 0$. Die Stabilitätsbedingung folgt aus (4.5.34) und (4.5.35) und ist

$$0 < g_1 g_2 < 1 \tag{4.5.36}$$

Das resultierende Stabilitätsdiagramm ist in der Abbildung 4.211 gezeigt. Eine Liste der Bezeichnungen zeigt Tabelle 4.10.
Instabile Resonatoren, wie sie in der Abbildung 4.212 gezeigt sind, werden bevorzugt bei Verstärkermedien mit sehr hoher Verstärkung verwendet. Ebenso werden sie oft bei Kurzpuls-Lasern eingesetzt. Dadurch dass der Strahl divergiert, ist die Intensitätsverteilung des Laserlichts gleichmässiger über alle Moden verteilt.
Die Frequenzen der in einem Resonator möglichen Moden hängen, wie in Abbildung 4.213 gezeigt, vom Resonatortyp an. Beim konfokalen Resonator sind die Eigenfrequenzen durch

$$\nu = \frac{c}{\lambda} = \frac{c}{2d} \left[ q + \frac{1}{2}(m + n + 1) \right] \tag{4.5.37}$$

gegeben. $q$ ist der Index der longitudinalen Modenverteilung, $m$ und $n$ die Indices der transversalen Modenverteilung. Der Spiegelabstand

$$d = p \cdot \frac{\lambda}{2} \quad \text{wobei} \quad p = q + \frac{1}{2}(m + n + 1) \tag{4.5.38}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

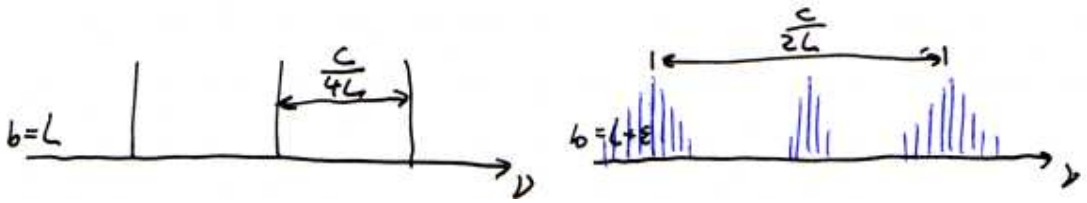



Abbildung 4.213.: Frequenzspektrum eines konfokalen Resonators (links) und eines nicht-konfokalen Resonators (rechts). Für den rechten Fall ist der Resonator nur wenig ($b = (1+\varepsilon)\cdot d$ mit $|\varepsilon| \ll 1$) vom konfokalen Resonator ($b = d$) unterschieden.

Das heisst, dass höhere transversale Moden mit $q_1 = q$ und $q_2 = m + n$ die gleich Frequenz haben wie eine transversale Grundmode ($m + n = 0$) mit dem longitudinalen Modenindex $q = q_1 + q_2$. Das Frequenzspektrum eines konfokalen Resonators ist also entartet. Der Modenabstand für die longitudinalen Moden ist

$$\delta\nu = \frac{c}{2d} \tag{4.5.39}$$

während transversale Moden mit $q_1 = m + n$ und $q_2 = q_1 + 1$ um

$$\delta\nu_{konfokal} = \frac{c}{4d} \tag{4.5.40}$$

voneinander entfernt sind.
Bei nichtkonfokalen Resonatoren, bei denen der Krümmungsradius der Spiegel $b$ nicht gleich dem Spiegelabstand $d$ ist, ist das Frequenzspektrum nicht mehr entartet

$$\nu = \frac{c}{2d}\left[q + \frac{1}{2}(1+m+n)\left(1 + \frac{4}{\pi}\arctan\frac{d-b}{d+b}\right)\right] \tag{4.5.41}$$

Die transversalen Moden liegen in einem Bereich um die transversale Grundmode mit dem gleichen longitudinalen Modenindex. Dies ist in der rechten Seite von Abbildung 4.213 gezeigt.
Bei einer endlichen Güte des Laserresonators verringert sich die Intensität des Lichtes mit jedem Umlauf um einen kleinen Wert. Nach der Zeit $\tau = \frac{Q}{2\pi\nu}$ ist sie auf den Wert $1/e$ gesunken. Die daraus resultierende Frequenzunschärfe ist

$$\Delta\nu = \frac{1}{2\pi\nu} = \frac{\nu}{Q} \tag{4.5.42}$$

oder, umgeschrieben,

$$\frac{\Delta\nu}{\nu} = \frac{1}{Q} \tag{4.5.43}$$

Wenn die Verluste im Laserresonator vorwiegend durch die Auskopplung von Licht an den Spiegeln stammen, können die Gleichungen für Fabry-Perot-Interferometer verwendet werden. Dort ist die transmittierte Intensität durch

$$I_T = I_0 \frac{T^2}{(1-R)^2 \cdot (1 + F\sin^2\frac{\delta}{2})} \tag{4.5.44}$$

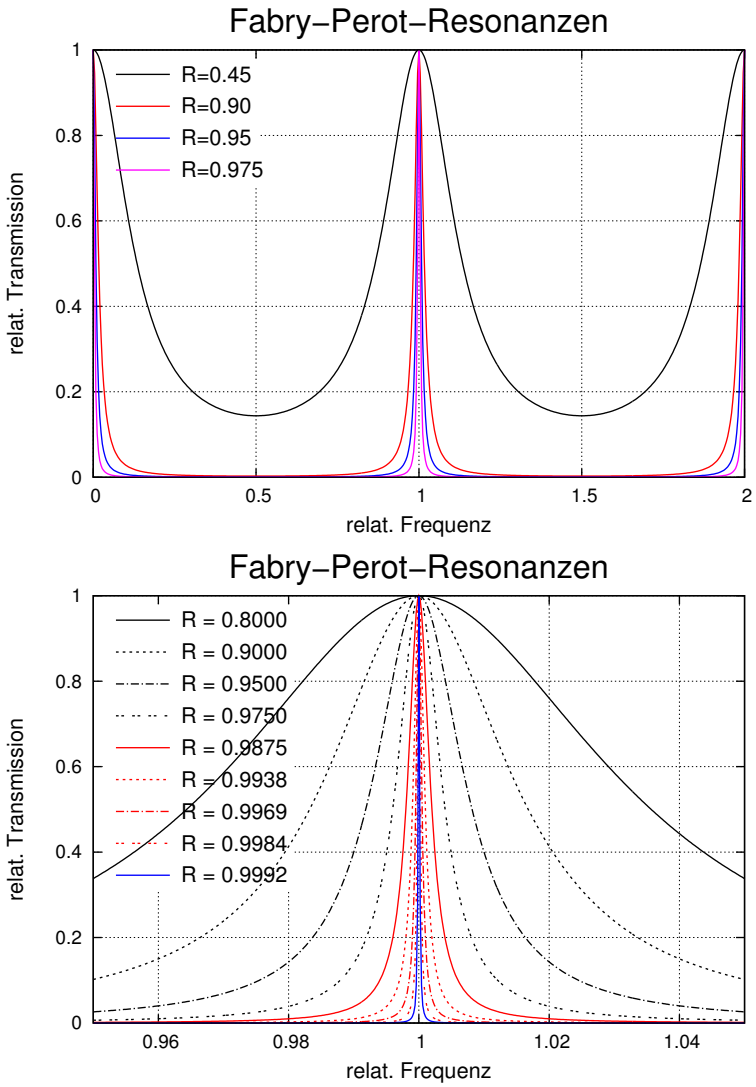



Abbildung 4.214.: Fabry-Perot-Resonanzen: oben ist ein Überblick gezeigt, unten die Vergrösserung um 1. Die Kurven sind auf einen Frequenzabstand von 1 normiert.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

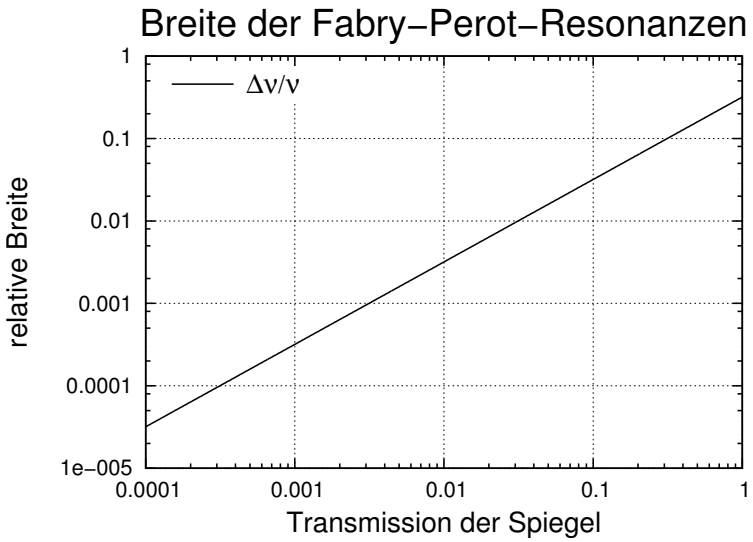



Abbildung 4.215.: Normierte Linienbreite als Funktion von $T = 1 - R$. Der Modenabstand im Fabry-Perot-Resonator ist 1.

gegeben (siehe auch Abb. 4.214), wobei die *Finesse* $F = \frac{4R}{(1-R)^2}$ ist. Die Reflektivität $R$ der Spiegel, die Absorption $A$ in den Spiegeln und ihre Transmission hängen über $T = 1 - A - R$ zusammen. Die Intensität im Resonator ist $I_{int} = \frac{I_T}{1-R}$. Resonanzen treten bei $\delta = 2m\pi$ auf. Die Halbwertsbreite ist dann

$$\Delta\nu = \frac{c}{2d}\frac{1-R}{\pi\sqrt{R}} = \frac{\delta\nu}{F^*} \tag{4.5.45}$$

Hier ist $F^* = \frac{\pi\sqrt{R}}{1-R}$ die Reflexionsfinesse. Haben die beiden Spiegel unterschiedliche Reflektivitäten $R_1$ und $R_2$, so wird für $R = \sqrt{R_1 \cdot R_2}$ gesetzt. Die in diesem Abschnitt berechneten Linienbreiten sind die Linienbreiten eines passiven Resonators. Durch das aktive Medium werden die Resonatoren entdämpft: die Linienbreiten werden geringer.
Mit einem aktiven Medium im Resonator werden diejenigen Moden verstärkt, für die die Nettoverstärkung pro pro Umlauf $G(\nu) = I/I_0 = exp[-2\alpha(\nu)L - \gamma]$ nach Gleichung (4.5.5) maximal ist. Nach Demtröder[?] ist die transmittierte Intensität

$$I_T = I_0 \frac{(1-R)^2 G(\nu)}{[1-G(\nu)]^2 + 4G(\nu)\sin^2\frac{\delta}{2}} \tag{4.5.46}$$

In Abbildung 4.216 ist das damit berechnete Verstärkungsprofil eingezeichnet. Wenn die Verstärkung gegen 1 geht (hier mit einer Gauss-Funktion[24], die ihr Maximum bei 53 und eine Breite von 14.34 hat) geht die Gesamtverstärkung $I_T/I_0 \to \infty$. Dieses Maximum wird bei $\delta = q \cdot 2\pi$ erreicht. Dabei muss anstelle der Resonatorlänge $d$ die effektive Resonatorlänge

$$d^* = (d - L) + n(\nu)L = d + (n-1) \cdot L \tag{4.5.47}$$

[24]Nach Demtröder[?] ist das Linienprofil gaussförmig, wenn die Dopplerverbreiterung, wie bei Gaslasern im sichtbaren Wellenlängenbereich, dominierend ist.

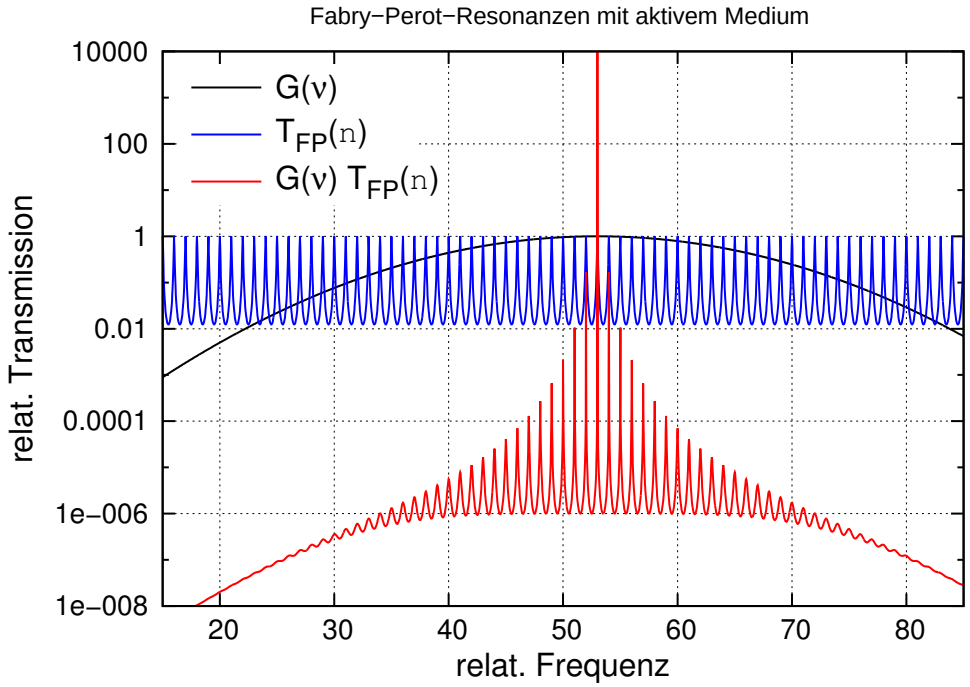



Abbildung 4.216.: Verstärkungsprofil (rot) eines Laserüberganges und die Resonatormoden (blau). Das kombinierte Verstärkungsprofil nach Gleichung (4.5.46) ist grün eingezeichnet.

verwendet werden. $L$ ist die Länge des Lasermediums und $n(\nu)$ der (frequenzabhängige) *Brechungsindex*. Die Frequenzbreite des aktiven Resonators wird

$$\Delta\nu = \delta\nu \frac{1 - G(\nu)}{2\pi\sqrt{G(\nu)}} = \frac{\delta\nu}{F_\alpha^*} \tag{4.5.48}$$

Die Finesse $F_\alpha^*$ des aktiven Resonators wird unendlich, wenn die Verstärkung $G(\nu) \to 1$ wird.
Die Abbildung 4.217 zeigt, wie das Modenprofil sich in Funktion der Verstärkung ändert. Während bei niedrigen Verstärkungen die Transmission für viele Moden etwa gleich ist, beginnt eine einzelne Mode zu dominieren, wenn die Verstärkung $G(\nu)$ gegen 1 geht.
Im Gegensatz zu den der Abbildung 4.216 zugrunde liegenden Annahmen ist das Verstärkungsprofil des Lasermediums meistens sehr viel breiter als der longitudinale Modenabstand. Deshalb ist die Anzahl schwingungsfähiger Moden meistens wie in der Abbildung 4.218 gezeigt, grösser als 1. Ausnahmen sind Laserdioden wegen ihrem sehr kurzen Resonator und gewisse sehr hochgezüchtete Laseranordnungen.

## 4.5.2. Kurzzeitlaser

Kurze Lichtpulse könnten erzeugt werden, indem die Betriebsspannung der Lichtquelle kurzzeitig eingeschaltet wird. Die kürzesten erreichbaren Zeiten hängen von den Schaltkapazitäten und den möglichen Schaltströmen ab. Es ist schwierig, Spannungen oder Ströme kürzer als in etwa 100 ps einzuschalten.
Deshalb werden kurze Lichtpulse ausschliesslich auf optischem Wege erzeugt. Man

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

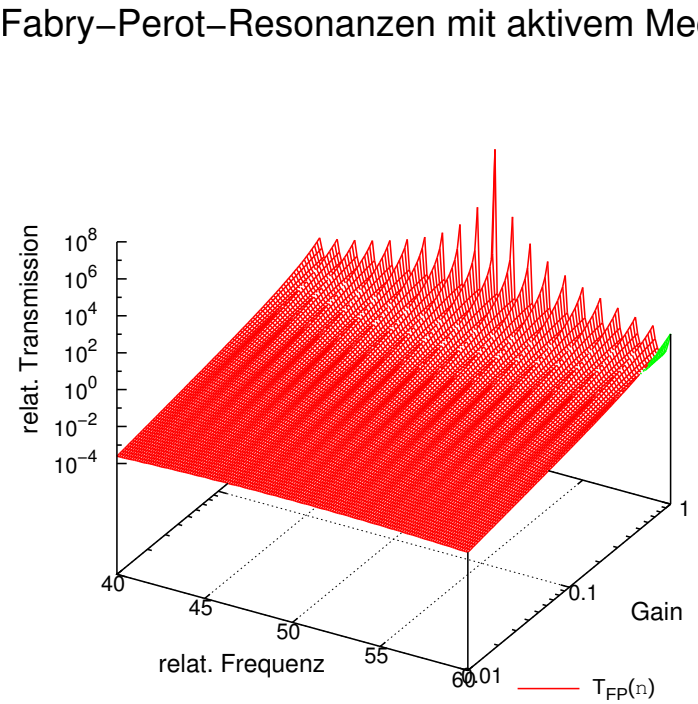



Abbildung 4.217.: Modenprofil des aktiven Resonators in Abhängigkeit der Verstärkung.

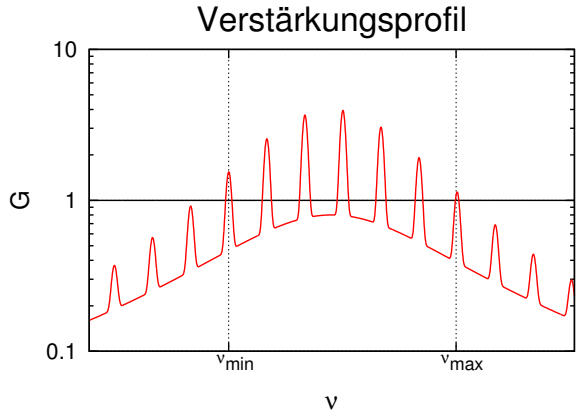



Abbildung 4.218.: Verstärkungsprofil des aktiven Mediums

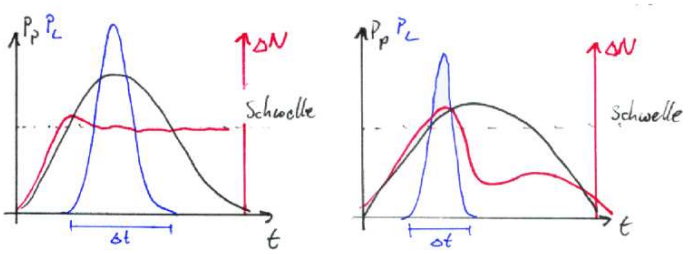



Abbildung 4.219.: Zeitliche Beziehung zwischen Pumppuls, Laserpuls und Besetzungsinversion. Links die Kurvenformen, wenn die Lebensdauer des unteren Laserniveaus genügend klein sind, andernfalls (rechts) wird die Pulsdauer und -energie limitiert.

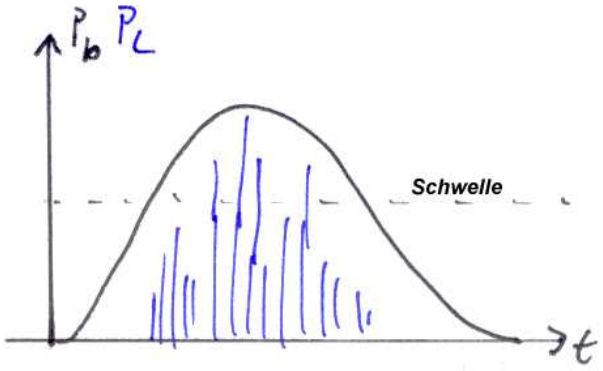



Abbildung 4.220.: Auch bei Kurzpulslasern treten Relaxationsschwingungen (Spikes) auf.

nutzt aus, dass das Einschalten eines Lasers mit grossen Relaxationsschwingungen verbunden ist. Diese Schwingungen entstehen, weil die für eine Lasertätigkeit notwendige Inversion im Dauerbetrieb wesentlich geringer ist als im Einschaltmoment. Die die Relaxationsschwingungen beschreibenden Differentialgleichungen sind nichtlinear: der Laser ist in vielen Betriebszuständen ein chaotisches System. Die Abbildung 4.219 zeigt den Zusammenhang der Laserleistung, der Inversion und der Pumpleistung. Wenn die Pumpe eingeschaltet wird, baut sich die Inversion parallel zum Anstieg der Pumpleistung auf. Wenn die Schwelle überschritten wird, wird die Besetzungszahl auf einem Wert, der nur unwesentlich über der Schwellinversion liegt, begrenzt. Die Laserleistung steigt rapide an und die Besetzungsinversion wird, wenn die Pumpleistung wieder abnimmt, wieder abgebaut. Der resultierende Laserpuls ist kürzer als der Pumppuls. Auf der rechten Seite der Abbildung 4.219 wird gezeigt, was passiert, wenn das untere Laserniveau nicht schnell genug entleert wird. Dann nimmt die Möglichkeit zu spontaner und induzierter Emission sehr viel schneller beschränkt. Die Besetzungszahlinversion baut sich ab, auch wenn die Pumpleistung hoch bleibt. Im Verhältnis zum Pumppuls ist der Laserpuls kürzer. Ein nächster Pumppuls kann jedoch erst dann folgen, wenn die Besetzung des unteren Laserniveaus wieder in die Nähe des Ursprungswertes abgebaut ist.

Wenn die induzierte Emission sehr stark verstärkt wird, wie zum Beispiel in Blitzlampen gepumpten Rubinlasern aber auch in Laserdioden, dann treten Relaxationsschwingungen auf. Während der Dauer des Pumppulses treten einige bis viele

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

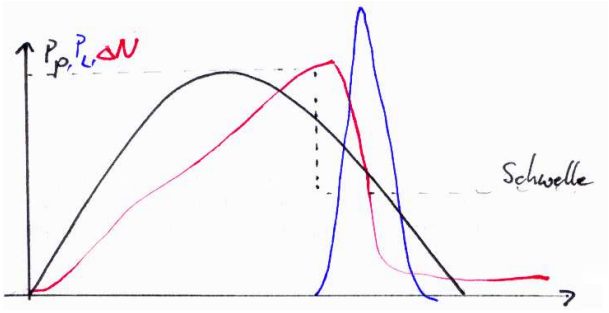



Abbildung 4.221.: *Güteschaltung* bei einem *Kurzpulslaser*. Die Dauer des Laserpulses und des Pumppulses sind so entkoppelt.

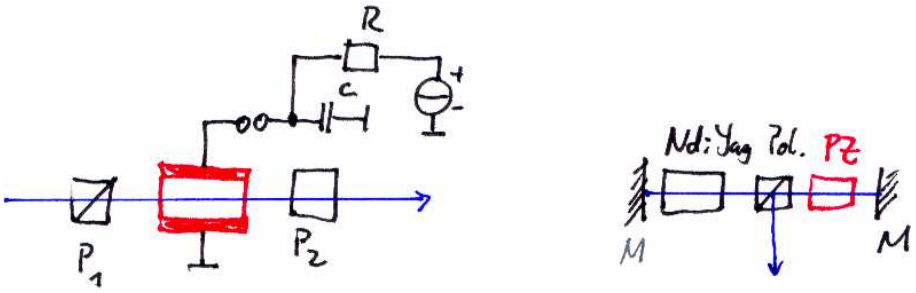



Abbildung 4.222.: Links die prinzipielle Schaltung einer Pockelszelle, rechts eine Implementation in einem gepulsten Nd-Yag-Laser.

sogenannte Spikes, also Relaxationsschwingungen auf. Die Einhüllende der Amplitude dieser Spikes folgt der Amplitude des Pumppulses.
Ein Nachteil dieser *Relaxationsschwingung*en ist, dass der Zeitpunkt der einzelnen Pulse nicht gut bestimmt ist. Indem man die Verluste im Resonator gross macht, verhindert man das Anschwingen der Laserschwingung. In der Abbildung 4.221 ist gezeigt, dass, wenn man die Verluste in kurzer Zeit $< 1ns$ erniedrigt, zu einem genau definierten Zeitpunkt ein einzelner Laserpuls entsteht.
Das Schalten der Verluste kann entweder über akusto-optische Schalter, elektro-optische Schalter oder durch sättigbare Absorber geschehen. Eine Implementation eines elektrooptischen Schalters ist die Pockelszelle. Die Transmission der Pockelszelle in Abb. 4.222 ist durch die Funktion

$$T = T_0 \left(1 - \cos^2 \Theta\right) \tag{4.5.49}$$

gegeben. Dabei ist $\Theta$ der Winkel der Drehung der Polarisationsebene. Dieser ist proportional zur an der Pockelszelle angelegten Spannung. Abb. 4.223 zeigt den Kurvenverlauf der Resonatorverluste, der Transmission durch die Pockelszelle in Relation zur angelegten Spannung.

#### 4.5.2.0.1. Akusto-optischer Modulator und Pulslaser mit Cavity Dumping

Im akusto-optischen Modulator wird eine Schallwelle unter schiefem Winkel zur Ausbreitungsrichtung des Lichtstrahles in einen Kristall eingestrahlt (sieh Abb. 4.225). Durch die laufende Schallwelle wird ein sich mit Schallgeschwindigkeit bewegendes moduliertes Dichteprofil erzeugt. Dieses bewirkt eine Modulation des Brechungsindexes und somit eine Bragg-Streuung am optischen Gitter.

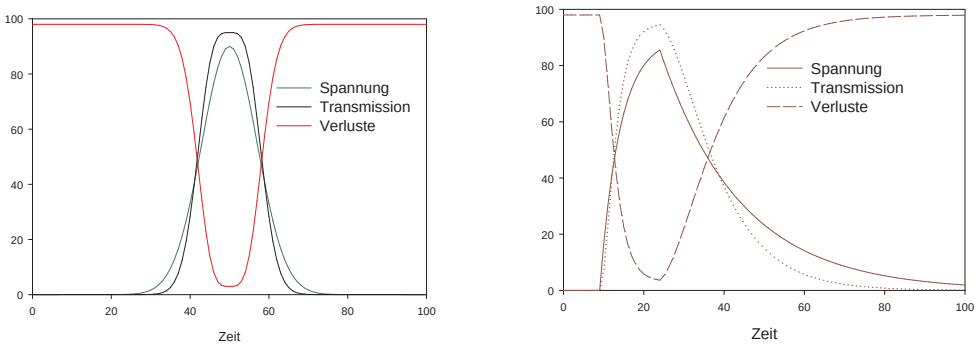



Abbildung 4.223.: Links sind für einen gaussförmigen Spannungspuls der Spannungsverlauf, die Transmission und die Verluste angegeben. Rechts das gleiche für einen exponentiell ansteigenden und abfallenden Puls.

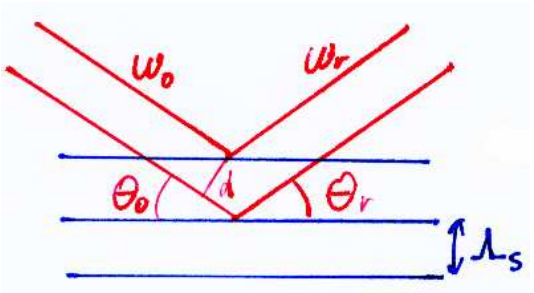



Abbildung 4.224.: Schematische Darstellung der Bragg-Reflexion von Licht an Schallwellen.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

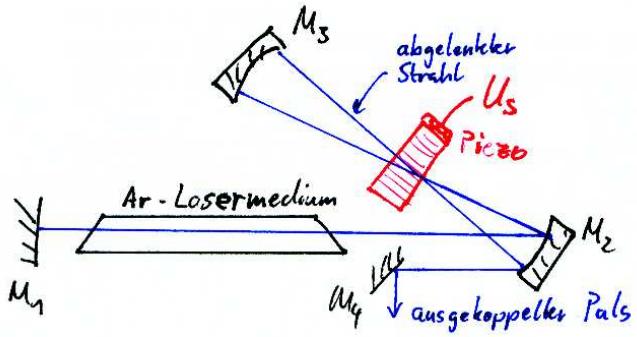



Abbildung 4.225.: Schematischer Aufbau der Auskopplung aus einem gütegeschalteten Laser (`cavity dumping`).

Wir nehmen nun an, dass in diesem Kristall mit dem *Brechungsindex* $n$ eine Schallwelle mit der Frequenz $\Omega$, der Schallgeschwindigkeit $c_S$ und der Wellenlänge $\Lambda_S = c_S/\Omega$ vorhanden ist. Wenn die Bragg-Bedingung

$$2\Lambda_S \sin\Theta = \frac{\lambda}{n} \tag{4.5.50}$$

erfüllt ist, dann wird der Bruchteil $\eta$ der eingestrahlten Intensität in die erste Beugungsordnung abgelenkt. Hier ist $\lambda$ die Wellenlänge des Lichtes. Die Beugungseffizienz $\eta$ hängt von der Tiefe der Brechzahlmodulation $\Delta n$ und somit von der Amplitude der Schallwelle ab. Dadurch dass das Licht durch eine **laufende** Schallwelle abgelenkt wird, wird seine Wellenlänge und Frequenz moduliert. Der ungebeugte Lichtstrahl hat die Frequenz $\omega = \lambda/c$, während der gebeugte Lichtstrahl um

$$\Delta\omega = 2\frac{nc_S}{c}\omega\sin\Theta = 2n\Lambda_S\frac{\Omega}{\omega\lambda}\omega\sin\Theta = \Omega \tag{4.5.51}$$

in der Frequenz Doppler-verschoben wird. Wenn die Amplitude des eingestrahlten Lichtes $E_0$ ist, sind die Amplituden des transmittierten und gebeugten Anteils

| | |
|---|---|
| transmittiert | $\sqrt{1-\eta}E_0\cos\omega t$ |
| gebeugt | $\sqrt{\eta}E_0\cos(\omega+\Omega)\,t$ |

Abb. 4.225 zeigt den Aufbau eines gepulsten Lasers, bei dem der akusto-optische Modulator die Auskopplung aus der Laser-Cavity steuert. Das vom Spiegel $M_2$ herkommende Licht passiert den akusto-optischen Modulator und wird mit der Effizienz $\eta$ abgelenkt. Auf dem Rückweg muss das ausgekoppelte Licht unabgelenkt durch den Modulator gehen (Effizienz $1-\eta$). Der Strahl, der unabgelenkt vom Spiegel $M_2$ her kommend durch den akusto-optischen Modulator gegangen wird, wird auf dem Rückweg mit der Effizienz $\eta$ abgelenkt. Im ersten Fall wird die Schallfrequenz von der Lichtfrequenz abgezählt, im zweiten Fall dazugezählt.
In der Auskoppelrichtung überlagern sich die Amplituden

$$\begin{aligned} E_{tot} &= \sqrt{\eta}\sqrt{1-\eta}E_0\cos(\omega-\Omega)t + \sqrt{\eta}\sqrt{1-\eta}E_0\cos(\omega+\Omega)t \\ &= \sqrt{\eta}\sqrt{1-\eta}E_0\left[\cos(\omega-\Omega)t + \cos(\omega+\Omega)t\right] \end{aligned} \tag{4.5.52}$$

Der ausgekoppelte Puls hat dann die Leistung

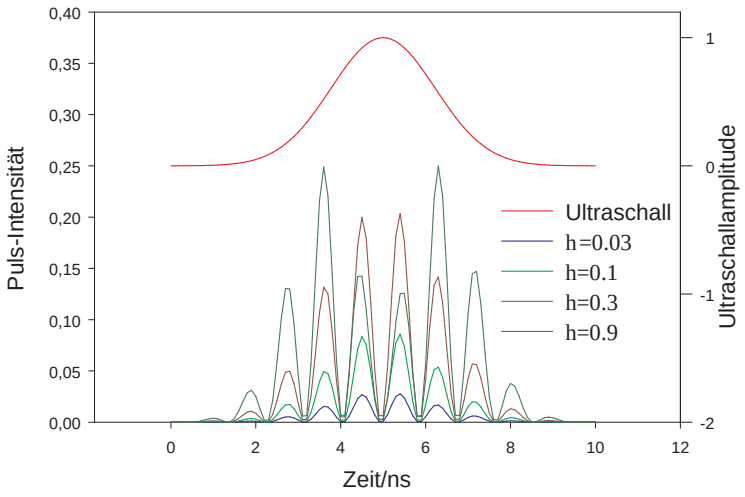



Abbildung 4.226.: Dargestellt ist der Verlauf des Ultraschallpulses und des Laserpulses für vier Modulationstiefen $\eta$ im akusto-optischen Modulator.

$$\begin{aligned} P_a(t) &= |\langle \boldsymbol{S}_t \rangle| \\ &= |\langle \boldsymbol{E}_{tot} \times \boldsymbol{H}_{tot} \rangle| \\ &= \frac{1}{2Z_0} E_{tot}^2 = 2c\varepsilon\eta t \left(1-\eta(t)\right) E_{tot}^2 \cos^2 \Omega t \end{aligned} \tag{4.5.53}$$

Hier ist $\boldsymbol{S}_t$ der Poynting-Vektor und $Z_0 = \sqrt{\mu_0/\varepsilon_0}$ der Wellenwiderstand des Vakuums. Während der zeit des Ultraschallimpulsdes wird $\eta(t)\left(1-\eta(t)\right)$ der in der Laserkavität eingeschlossenen optischen Leistung ausgekoppelt. Abb. 4.226 zeigt die Ultraschallamplitude und für vier verschiedene Beugungseffizienzen $\eta$ den zeitlichen Verlauf des ausgekoppelten Pulses. Interessant ist, dass für $\eta = 0.5$ ein Maximum erreicht wird. Bei der in Abb. 4.226 gezeigten Kurve für $\eta = 0.9$ resultieren deshalb zwei Intensitätsmaxima.
Mit dem Verfahren des Cavity-Dumping erreicht man bei Ionenlasern oder bei Farbstofflasern Pulslängen von $10 - 100ns$ mit Pulsfolgefrequenzen zwischen null und 4 MHz.

**4.5.2.0.2. Modenkopplung** Wenn, wie in Abbildung 4.227 gezeigt, ein akusto-optischer Modulator in den Laserresonator eingefügt wird, dann entstehen im Frequenzspektrum Nebenfrequenzen. Ist die Modulationsfrequenz $f$, dann existieren neben der Grundfrequenz des Lasers $\nu$ auch die Frequenzen $\nu \pm f$. Wenn die Modulationsfrequenz gleich dem Modenabstand im *Resonator* ist, das heisst wenn $f = c/2d$ ist, dann können die Seitenbänder auch an der Laseroszillation teilnehmen. Diese Seitenbänder werden auch moduliert, so dass alle vom Verstärkungsprofil des Lasermediums her möglichen Moden anschwingen.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

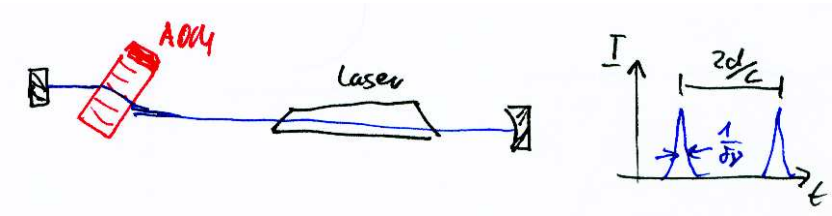



Abbildung 4.227.: Mit einem akusto-optischen Modulator im Ultraschallbereich kann eine aktive Modenkopplung erreicht werden. Die Lasermoden in einem

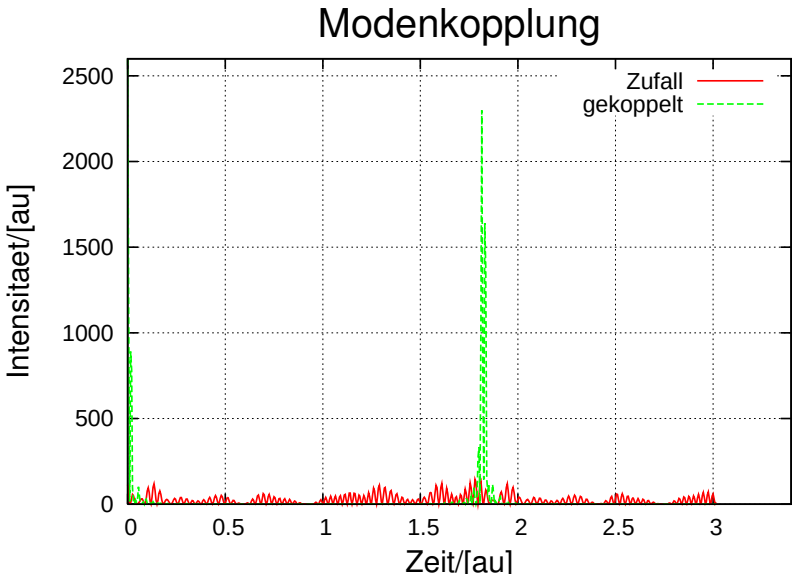



Abbildung 4.228.: Dargestellt einerseits die Überlagerung von 51 Moden mit zufälliger Phase und gleicher Amplitude sowie die Überlagerung von 51 modengelockter Moden. Die resultierende Pulsüberhöhung ist augenfällig.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

Durch die Modulation schwingen die Lasermoden nicht unabhängig, da ihre Phasen durch den Modulator gekoppelt sind. Abb. 4.228 zeigt, die resultierende Ausgangsamplitude für viele Lasermoden mit zufälligen Phasen sowie für gekoppelte Phasen. Die Intensität bei gekoppelten Phasen wird periodisch sehr gross. Andererseits zeigt das Ausgangssignal bei zufälligen Phasen das auch von Laserdioden her bekannten vergrösserte Rauschen.
Der akusto-optische Modulator moduliert die Transmission des Laserresonators mit

$$T = T_0 \left[1 - \delta \left(1 - \cos \Omega t\right)\right] = T_0 \left[1 - 2\delta \sin^2 \left(\frac{\Omega t}{2}\right)\right] \tag{4.5.54}$$

Unter der Annahme, dass alle Lasermoden die gleiche Amplitude $A_{k,0} = A_0$ haben wird bei einem kleinen Modulationsgrad $\delta \leq 1/2$ die instantane Amplitude der k-ten Mode zu

$$A_k(t) = T A_0 \cos \omega_k t = T_0 A_0 \left[1 - \delta \left(1 - \cos \Omega t\right)\right] \cos \omega_k t \tag{4.5.55}$$

Wenn nun die Modulationsfrequenz gleich der Umlaufzeit des Lichtes im resonator ist, also wenn $\Omega = 2\pi c/(2d)$ so wird die $k+1$-te Mode

$$A_{k+1} = \frac{A_0 T_0 \delta}{2} \cos \left(\omega_{k+1} t\right) \tag{4.5.56}$$

Diese Modulation wird, sofern sie innerhalb der Verstärkungsbandbreite des Lasermediums liegt, verstärkt. Die $k+1$-te Mode wird nun wieder moduliert, genauso wie alle nachfolgenden Moden. Das gleiche gilt auch für Moden mit abnehmenden Indizes. Durch die Modulation sind alle Phasen der verschiedenen Moden periodisch gleich. Dies tritt in der Gleichung (4.5.55) immer zu den Zeiten

$$t_j = j\frac{2d}{c} \quad \text{für } j = 0, 1, 2, \ldots \tag{4.5.57}$$

Ist die Bandbreite der verstärkbaren Moden (oberhalb der Laserschwelle) $\delta\nu$ und $\Delta\nu$ der Abstand der einzelnen Moden, dann ist die Anzahl der verstärkten Moden

$$N = \frac{\delta\nu}{\Delta\nu} = \frac{2\delta\nu d}{c} \tag{4.5.58}$$

Die Überlagerung von $2m + 1 = N$ Lasermoden mit gleicher Amplitude führt zur Gesamtamplitude

$$A(t) = A_0 \sum_{j=-m}^{j=m} \cos \left(\omega_0 + j\Omega\right) t \tag{4.5.59}$$

Die Laserintensität $I(t) = A^2(t)$ wird dann

$$I(t) \approx A_0^2 \frac{\sin^2 \left(N\Omega t/2\right)}{\sin^2 \left(\Omega t/2\right)} \cos^2 \omega_0 t \tag{4.5.60}$$

Wie auch aus Abbildung 4.228 ersichtlich ist, bekommt man eine Pulsfolgezeit $T$ und eine Pulsbreite $\Delta t$.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

| Lasermedium | Wellenlänge | Frequenzbreite $\delta\nu$ | Pulsbreite $\Delta T$ |
|---|---|---|---|
| HeNe | 633 nm | 1.5 GHz | 500 ps |
| Argon-Ionenlaser | 488 nm, 514 nm | 5-7 GHz | 150 ps |
| Nd-Glas-Laser | 1064 nm | 200 GHz | 5 ps |
| Farbstoff- oder Farbzentrenlaser | 600 nm | 30 THz | 30 fs |

Tabelle 4.11.: Demtröder [?] gibt die oben zusammengefassten Möglichkeiten zur Erzeugung kurzer Pulse an.

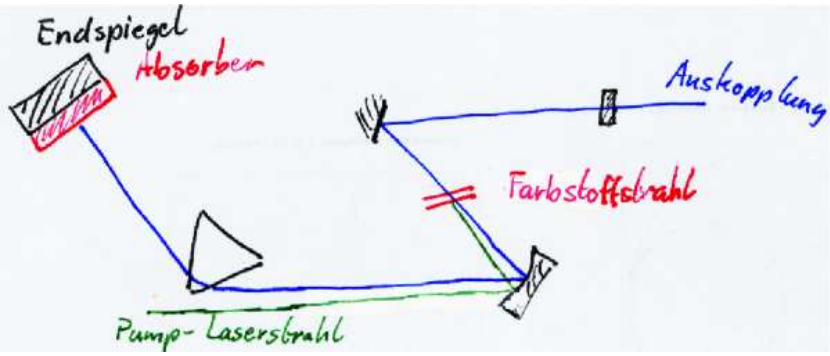



Abbildung 4.229.: Die Modenkopplung wird bei diesem Aufbau durch einen sättigbaren Absorber erreicht.

$$\text{Abstand der Pulse} \quad T = \frac{2d}{c} = \frac{1}{\Delta\nu} \tag{4.5.61}$$

$$\text{Pulsbreite} \quad \Delta T = \frac{1}{(2m+1)\Omega} = \frac{1}{N\Omega} = \frac{1}{\delta\nu} \tag{4.5.62}$$

Damit wird klar, dass die kürzest mögliche Pulsdauer von der Breite des Verstärkungsprofils abhängt. Lasermedien mit schmalen Linien wie zum Beispiel Gaslaser sind für Modenkopplung ungeeignet. Die Spitzenleistung eines modengekoppelten Lasers geht wie $N^2$, das heisst auch wieder mit der spektralen Bandbreite des Lasers. Die Eignung von Lasermedien zur Erzeugung kurzer Pulse wird in Tabelle 4.11 zusammengefasst.

**4.5.2.0.3. Passive Modenkopplung** Schneller als ein optischer Modulator schalten sättigbare Absorber. Wichtig ist, dass die Absorptionsniveaus des Absorbers eine möglichst kurze Abklingzeit haben. Abb. 4.229 zeigt den Aufbau eines Lasers mit einem sättigbaren Absorber. Dieser wird vor einem der Resonatorspiegel montiert, so dass nur an einem wohldefinierten Ort die Absorption sich ändern kann. Durch die Absorption im Medium werden die Verluste vergrössert. Die Verstärkung im Lasermedium muss so gross sein, dass das gesamte System die Schwellenverstärkung erreicht. Das Lasermedium emittiert vor dem Erreichen der Schwelle spontan und dann induziert verstärkt und in statistischen Abständen. Die Amplitude schwankt stark. Wenn einer dieser Pulse die Schwellenenergie erreicht, dann wird durch die Verstärkung die Absorption im sättigbaren Absorber leicht verringert. Dieser erste Puls löst also eine Photonenlawine aus, die einerseits die Verstärkung des Pulses erhöht und andererseits verhindert, dass die anderen Schwankungen

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

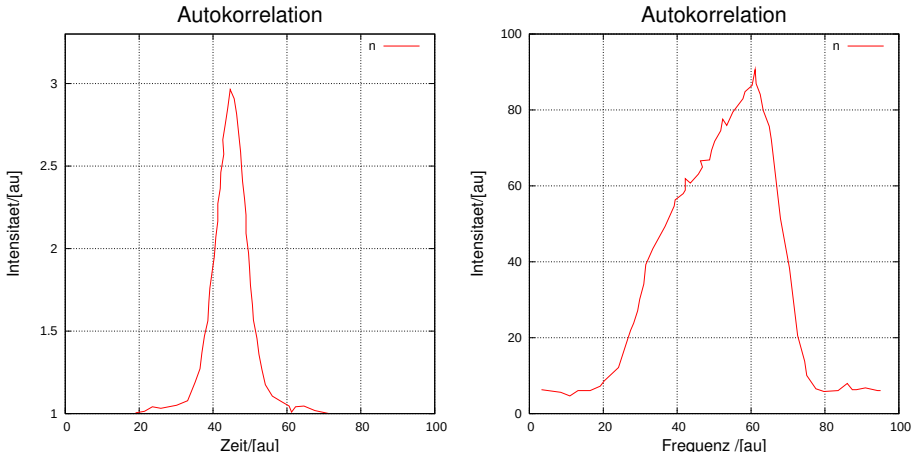



Abbildung 4.230.: Links wird die Autokorrelation, rechts das Spektrum eines modengekoppelten Pulses gezeigt (nach Demtröder [?]). Die Pulslänge ist 0.5 ps, die spektrale Breite 1nm.

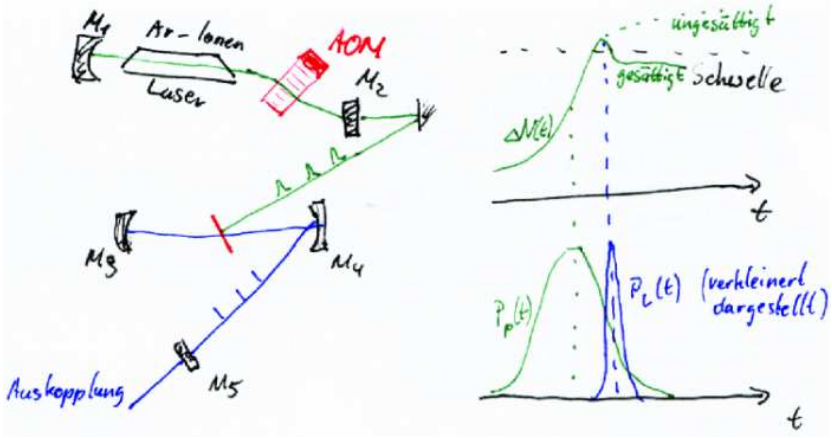



Abbildung 4.231.: Bei diesem Laser wird das Anregungslicht synchron zur Umlaufzeit im *Resonator* gepulst.

weiter verstärkt werden. Da das Absorptionsmedium eine sehr kurze Lebensdauer hat, ist es schon kurz nach dem Puls wieder in seinem hoch absorbierenden Zustand. Dieser umlaufende Puls ist der einzige, der verstärkt wird.
Die Pulsform und damit, über die *Fouriertransformation* auch das Spektrum, hängen von den Verstärkungseigenschaften des Mediums und von den spektralen Absorptionseigenschaften des Absorbers. Abbildung 4.230 zeigt links ein Beispiel für die Pulsform und rechts das Spektrum dieses Pulses. Die in Abb. 4.230 gezeigte Pulsbreite von $0.5ps$ ist die kürzeste, mit passiver Modenkopplung erreichbare Pulslänge.

**4.5.2.0.4. Synchron gepumpte Laser** Bei synchron gepumpten Lasern wird die Pumpleistung in einem Takt mit ganzahligem Verhältnis zur Umlaufszeit der Pulse im Resonator gepumpt. Die Abbildung 4.231 zeigt auf der linken Seite einen möglichen Aufbau eines synchron gepumpten Lasersystems[?]. Der Argon-Ionenlaser wird im Laserresonator mit einem akusto-optischen Modulator moduliert. Die Pumpleistung trifft mit der Umlauffrequenz der Pulse im Farbstofflaser auf das

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

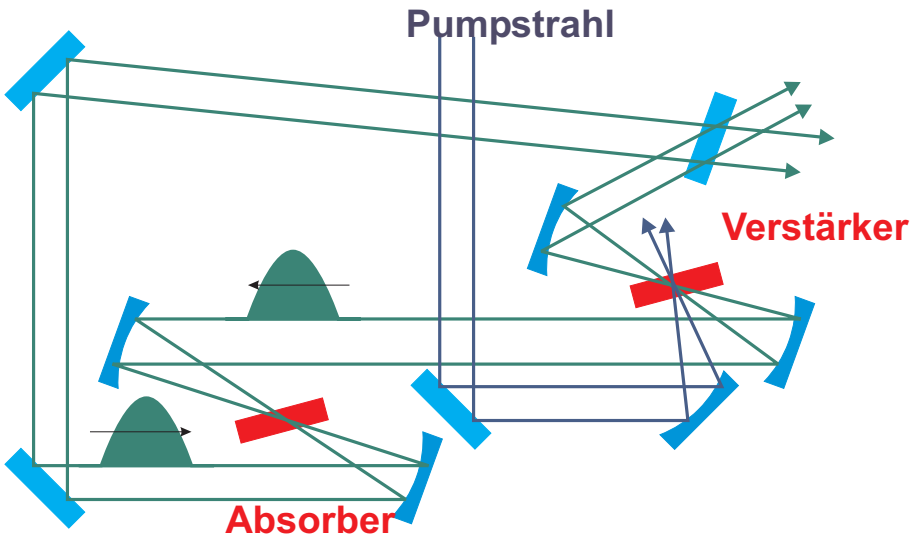



Abbildung 4.232.: Schematischer Aufbau eines CPM-Lasersystems.

Lasermedium, einen Farbstoffstrahl. Von allen möglichen, durch spontane Emission entstandenen Photonen werden nur diejenigen verstärkt, die synchron mit der Pumpleistung im Resonator umlaufen.
Die rechte Seite von Abbildung 4.231 den Verlauf der Verstärkung (oben) und die Intensitäten von Pumppuls und Laserpuls. Die Verstärkung würde bei sehr grossen Verlusten der gestrichelten Kurve folgen. Durch die Emission des Laserpulses, und da das synchrone Pumpen ähnlich wie ein Absorber im Resonator des Farbstofflasers wirkt, wird die Besetzungszahlinversion stark abgebaut. nur ein einzelner, aber sehr kurzer Laserpuls entsteht.
Die Umlaufszeit der Pulse im Laserresonator ist $T = 2d/c$ bei einem Resonator mit der Länge $d$. Typischerweise kann man mit einem synchron gepumpten Lasersystem Pulslängen von $0.5ps$ erreichen. Wenn der Resonator eine Länge von $1m$ hat, ist die Pulsfolgefrequenz $150MHz$. Ein Fehler von $1\mu m$ der Länge des Resonators führt zu einer Verbreiterung der Pulse auf $1ps$.
Durch einen akusto-optischen Modulator im Resonator des Pulslasers können die Verluste für alle ausser jeden k-ten Puls so erhöht werden, dass sie nicht anschwingen. Durch dieses Verfahren, das auch *Cavity Dumping* genannt wird, kann die Pulsfolgefrequenz erniedrigt werden. damit ist es möglich, auch längere Relaxationen auszumessen.

#### 4.5.2.1. fs-Laser

Sehr kurze Laserpulse erhält man mit sogenannten *CPM-Lasersystem*en. Eine mögliche Anordnung eines solchen Lasersystems ist in der Abbildung 4.232 gezeigt. Die Idee hinter dieser Anordnung ist die folgende:

- Zwei gegenläufige Pulse sollen den Verstärker im grösstmöglichen Abstand der halben Umlaufszeit $T/2$ passieren. Damit wird sichergestellt, dass die Verstärkung für beide Pulse gleich (aus Symmetriegründen) und maximal ist.
- Die Pulse sollen sich im sättigbaren Absorber überlagern. Jeder Puls schaltet für den anderen die Verluste auf einen niedrigeren Wert.

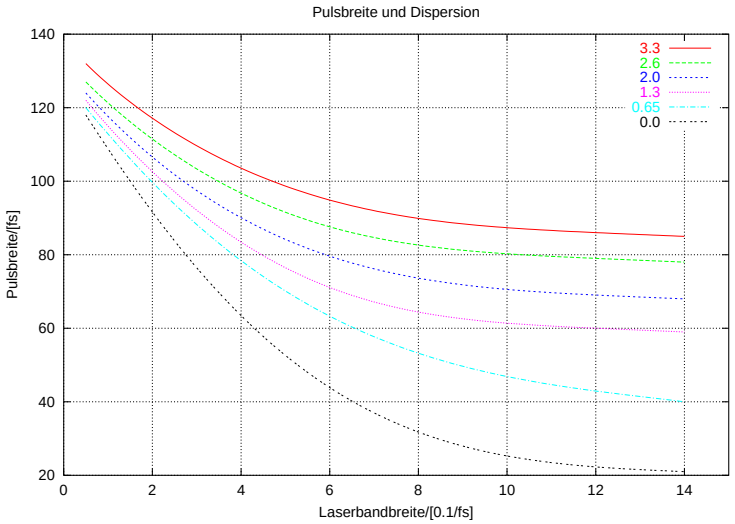



Abbildung 4.233.: Abhängigkeit der Pulsbreite von der Bandbreite eines Lasermediums unter Berücksichtigung der Dispersion.

Indem man die Dicke des Absorberstrahls sehr dünn ($< 100\mu m$) wählt, ist die Laufzeit durch das Medium kleiner als etwa $400fs$. Da nur die Überlagerung beider Pulse den Absorber auf niedrige Absorption schalten kann, ist dies nur bei einer perfekten Überlagerung der beiden Pulse, also wenn die Zeitunsicherheit sehr viel kleiner als $400fs$ ist, möglich.
Um die kürzesten möglichen Pulse zu erhalten, ist es notwendig, die *Dispersion* der Spiegel und der sonstigen optischen Elemente zu kompensieren[**?**]. Durch die CPM-Technik konnten Pulse mit einer Länge von unter $100fs$ erzeugt werden. Durch sättigbare Braggspiegel und eine *Dispersionskompensation* mindestens bis zur 3. Ordnung sind Pulse die kürzer als $10fs$ sind, möglich.

**4.5.2.1.1. Pulskompression** Wir nehmen an, dass ein optischer Puls mit der spektralen Energieverteilung $E(\omega)$ und der spektralen Breite $\delta\omega$ den zeitlichen Intensitätsverlauf

$$I(t) = \varepsilon_0 c \int |E(\omega, t)|^2 \, e^{j(\omega t - kz)} d\omega \tag{4.5.63}$$

hat. Dieser Puls läuft durch ein Medium mit dem *Brechungsindex* $n(\omega)$. Seine Form ändert sich, da die *Gruppenlaufzeit* für die verschiedenen spektralen Anteile verschieden lang ist. Wenn $v_{ph}$ die Phasengeschwindigkeit ist, bekommt man

$$v_g = \frac{d\omega}{dk} = \frac{d}{dk}(v_{Ph}k) = v_{Ph} + k\frac{dv_{Ph}}{dk} = \frac{c}{n}\left(1 + \frac{\lambda}{n}\frac{dn}{d\lambda}\right) \tag{4.5.64}$$

Diese Gruppengeschwindigkeit $v_g$ hat die Dispersion

$$\frac{dv_g}{d\omega} = \frac{\frac{dv_g}{dk}}{\frac{d\omega}{dk}} = \frac{1}{v_g}\frac{d^2\omega}{dk^2} \tag{4.5.65}$$

Bei Pulsen mit sehr hoher Intensität hängt der *Brechungsindex* von der Pulsleistung ab, ist also $n(\omega, I) = n_0(\omega) + n_1 I(t)$. Damit hängt die Phase auch von der

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

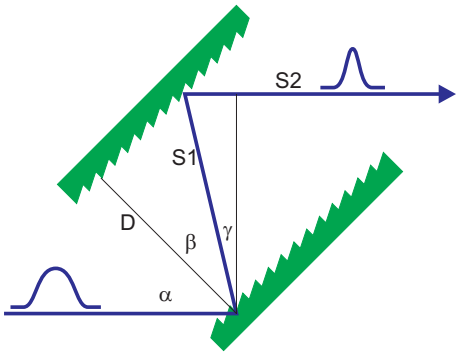



Abbildung 4.234.: *Dispersionskompensation* mit zwei Gittern. Der Wegunterschied $\Delta S = S_1 + S_2$ mit $S_1 = D/\cos\beta$ und $S_2 = S_1 \sin\gamma$

Intensität ab.

$$\phi = \omega t - kz = \omega t - \frac{\omega n z}{c} = \omega\left(t - \frac{n_0 z}{c}\right) - \frac{n_1 \omega z}{c} I(t) \tag{4.5.66}$$

Damit hängt aber auch die Frequenz eines Pulses von seiner instantanen Intensität ab. Mit $A = n_1 \omega z/c$ bekommt man

$$\omega = \frac{d\phi}{dt} = \omega_0 - \frac{A dI(t)}{dt} \tag{4.5.67}$$

Aus Gleichung (4.5.31) ersieht man, dass während des Intensitätsanstiegs eines Pulses seine Frequenz $\omega$ abnimmt. Zum Pulsende hin nimmt die Frequenz wieder zu. Durch diese Selbst-Phasenmodulation wird die spektrale Breite eines Pulses nach dem Durchgang durch ein dispersives Medium grösser.
Da der *Brechungsindex* $n$ bei normaler Dispersion $dn_0/d\lambda < 0$ die roten Anteile schneller propagieren lässt als die blauen Anteile, läuft der Puls auseinander. Das heisst wegen $n_0$ wird der Puls zeitlich breiter, wegen $n_1$ wird der Puls auch spektral breiter.
Unter der Annahme dass sich die Amplitude entlang der Ausbreitungsrichtung nur langsam ändert ($\lambda \partial^2 E/\partial z^2 \ll \partial E/\partial z$) wird die Wellengleichung

$$\frac{\partial E}{\partial z} + \frac{1}{v_g}\frac{\partial E}{\partial t} = \frac{j}{2v_g^2}\frac{\partial^2 E}{\partial t^2} - \frac{j\pi}{\lambda n} n_1 |E|^2 E \tag{4.5.68}$$

Ein Puls der Länge $\tau$ der mit der Geschwindigkeit $v_g$ durch ein Medium der Länge $L$ läuft, wird auf

$$\tau' = \tau\sqrt{1 + \left(\frac{\tau_c}{\tau}\right)^4} \tag{4.5.69}$$

verbreitert. dabei ist $\tau_c$ die kritische Pulsbreite

$$\tau_c = 2^{(5/4)}\sqrt{\frac{L}{\frac{\partial v_g}{\partial \omega}}} \tag{4.5.70}$$

Je kürzer der Puls ist, desto schneller läuft er auseinander. Zwei Beugungsgitter im Abstand $D$ können die unterschiedlichen Laufzeiten der roten und blauen Anteile

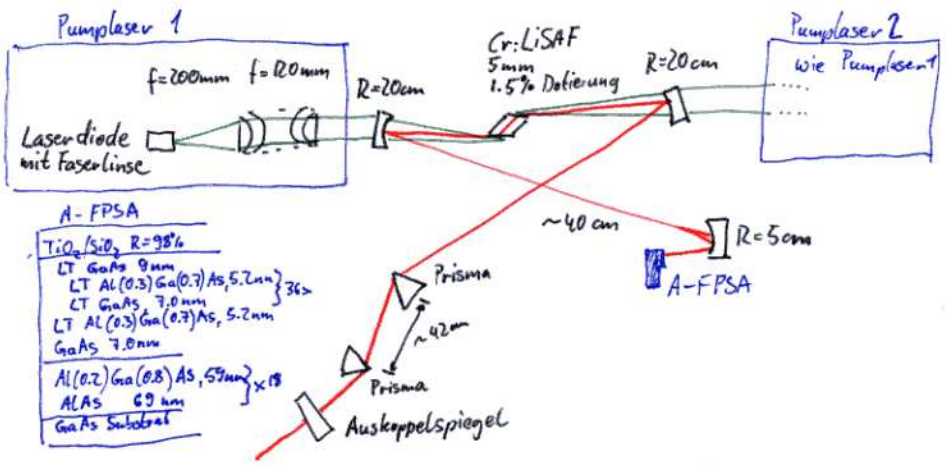



Abbildung 4.235.: Aufbau eines Cr:LiSAF-Lasers mit sättigbarem Bragg-Spiegel[**?**]

wieder kompensieren und so den Puls wieder komprimieren. Der optische Weg (siehe Abb. 4.234) ist dann

$$S(\lambda) = S_1 + S_2 = \frac{D}{\cos\beta}\,(1 + \sin\gamma) \tag{4.5.71}$$

dabei ist $\gamma = \pi - (\alpha + \beta)$. Nun verwenden wir das Additionstheorem für den Kosinus $\cos(\alpha + \beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$ wird Gleichung (4.5.71)

$$S(\lambda) = D\left(\cos\alpha + \frac{1}{\cos\beta} - \sin\alpha\tan\beta\right) \tag{4.5.72}$$

Die Dispersion eines Gitters ist $d\beta/d\lambda = 1/(d\cos\beta)$ wobei $d$ die die Gitterkonstante ist. Damit wird die Weglängendispersion

$$\frac{dS}{d\lambda} = \frac{dS}{d\beta}\frac{d\beta}{d\lambda} = \frac{D\lambda}{d^2\left[1 - \left(\sin\alpha - \frac{\lambda}{d}\right)^2\right)\Big]^{3/2}} \tag{4.5.73}$$

Nach Gleichung (4.5.73) nimmt der optische Weg mit zunehmender Wellenlänge zu. Damit lässt sich die normale Dispersion in Medien kompensieren. Ohne diese *Dispersionskompensation*, die unter Einbeziehung von Fasern und Prismen auch Effekte zweiter und dritter Ordnung kompensieren kann, wären fs-Laser nicht denkbar.

#### 4.5.2.2. Sättigbare Bragg-Spiegel als Anwendung von MQW-Schichten

Ein besonders eleganter Aufbau eines Kurzpuls-Lasersystems verwendet *sättigbare Bragg-Spiegel*[**?**] als sättigbares Medium. Konventionelle sättigbare Absorber haben eine Bandbreite und eine Mittenfrequenz, die vom Material abhängt. Andererseits ist bekannt, dass die Breite der Bandlücke bei Halbleitermaterialien durch die Einstellung des Mischungsverhältnisses bei ternären und quaternären Materialien in weiten Grenzen einstellbar ist. Durch die Verwendung von Schichtstrukturen können so hochwertige optische Schichten mit einstellbarer Bandbreite und einstellbarer Frequenz erzeugt werden.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

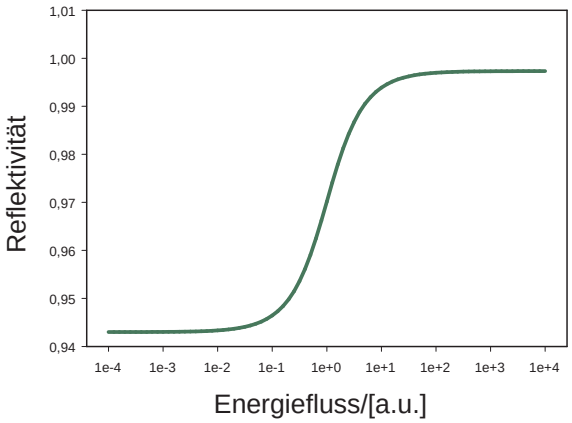



Abbildung 4.236.: Schematischer Verlauf der Reflektivität in einem *sättigbaren Bragg-Spiegel*

Wenn die optische Intensität bei der Beleuchtung eines Halbleitermaterials eine materialabhängige Schwelle überschreitet, befindet sich ein Grossteil der Elektronen des Valenzbandes in einem angeregten Zustand im Leitungsband. Das Material wird also transparent und ändert damit auch seinen *Brechungsindex*. Wenn nun ein Multischichtsystem so erzeugt wird, dass es bei hohen Intensitäten eine Reflektivität in der Nähe von 1 hat, dann kann dies wie ein sättigbarer absorber wirken.

Das in der Abbildung 4.235 gezeigte Lasersystem[**?**] verwendet einen *sättigbaren Bragg-Spiegel*, markiert mit *AFPSA* (antiresonant Fabri-Perot saturable absorber). Der Kurzpulslaser wird durch zwei Laserdionen über jeweils eine Strahlformungsoptik gepumpt. Als aktives Medium wird ein Cr:LiSAF-Kristall verwendet. Die Auskopplungseite des Laserresonators beinhaltet zwei Prismen zur *Dispersionskompensation*. Das andere Ende des Resonators wird durch einen *sättigbaren Bragg-Spiegel* gebildet. Die Schichtfolge in diesem Spiegel ist im Einsatz links angegeben.

Die schematische Kennlinie eines *sättigbaren Bragg-Spiegel* in der Abbildung 4.236 zeigt, dass die Reflektivität mit steigender Intensität zunimmt. Damit hat, wie bei den sättigbaren Absorbern der intensivste aller beim Einschalten anschwingenden Pulse die grösste Verstärkung. Nur dieser Puls wird im weiteren Verlauf durch den Laser verstärkt.

Ein *sättigbarer Bragg-Spiegel* aus $Al_xGa_{1-x}As/AlAs$ limitiert die Pulsweite auf 34 fs[**?**]. Der in der Abbildung 4.235 gezeigte *AFPSA sättigbare Bragg-Spiegel* ermöglicht durch eine geschicktere Ausnutzung der Materialien eine Erhöhung der Bandbreite und damit eine Pulslänge von 19 fs. Durch eine Kombination der Materialien $Al_{0.8}Ga_{0.2}As$ und $CaF_2$ sind Bandbreiten von $500nm$ um eine Mittenfrequenz von $800nm$ möglich[**?**]. Damit können mit einem Laser analog zur Abbildung 4.235 Pulse mit einer Länge von weniger als $10fs$ erzeugt werden.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

# 4.6. Optische Messverfahren

In diesem Abschnitt sollen einige grundlegende optische Messverfahren diskutiert werden. Der Abschnitt erhebt keinen Anspruch auf Vollständigkeit. Der interessierte Leser wird auf Werke wie das von Perez[**?**] oder Demtröder[**?**] verwiesen.

## 4.6.1. Absorptionsmessung

Absorption und Dispersion hängen in einem klassischen Modell[**?**] eng zusammen. Man beschreibt das optische Medium als eine Sammlung von getriebenen harmonischer Oszillatoren der Form

$$m\ddot{x} + 2\delta\dot{x} + kx = qE_0e^{j\omega t} \tag{4.6.1}$$

Wie üblich ist $m$ die Masse eines oszillierenden Teilchens, $k$ die Federkonstante, $q$ die Ladung auf der Masse, $\delta$ der Dämpfungsterm. Setzt man nun $\gamma = 2\delta/m$ und $\omega_0^2 = k/m$ und setzt als Lösung $x = x_0 \exp j\omega t$ an, so bekommt man

$$x_0 = \frac{qE_0}{m\left(\omega_0^2 - \omega^2 + j\gamma\omega\right)} \tag{4.6.2}$$

Durch die erzwungene Schwingung der Ladung $q$ entsteht ein induziertes elektrisches Dipolmoment

$$P = qx = \frac{q^2E_0e^{j\omega t}}{m\left(\omega_0^2 - \omega^2 + j\gamma\omega\right)} \tag{4.6.3}$$

Bei $N$ Oszillatoren pro Volumen ist die induzierte elektrische Polarisation $P$ pro Volumeneinheit durch

$$P = Nqx \tag{4.6.4}$$

gegeben. Die Polarisation ist jedoch in der Elektrodynamik auch mit der induzierenden elektrischen Feldstärke $E$ durch

$$P = (\varepsilon - 1)\,\varepsilon_0 E \tag{4.6.5}$$

verknüpft. Die relative Dielektrizitätszahl $\varepsilon$ hängt mit der Brechzahl $n$ über

$$n = \sqrt{\varepsilon} \tag{4.6.6}$$

zusammen. Durch Kombination von (4.6.3) bis (4.6.6) erhält man

$$n^2 = 1 + \frac{Nq^2}{\varepsilon_0 m\left(\omega_0^2 - \omega^2 + j\gamma\omega\right)} \tag{4.6.7}$$

Die Brechzahl $n(\omega)$ ist komplex und kann deshalb als

$$n(\omega) = n' - j\kappa \tag{4.6.8}$$

geschrieben werden. Beachtet man, dass die Lichtgeschwindigkeit $c$ im Medium von der Brechzahl $n$ abhängt ($c = c_0/n$) und setzt dies in die Gleichung einer

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

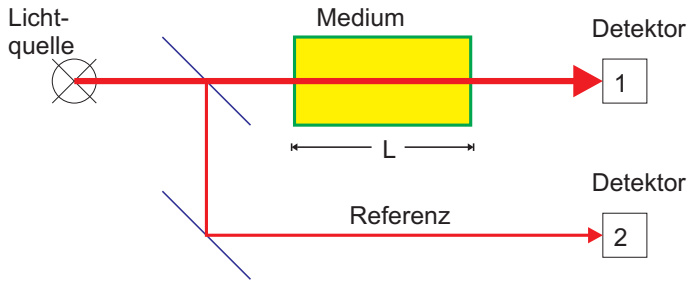



Abbildung 4.237.: Aufbau zur Messung einer Absorption

ebenen Welle $E = E_0 \exp[j(\omega t - Kz)]$ ein und berücksichtigt, dass $K_m = nK_0$ ist, so bekommt man

$$E = E_0 e^{-K_0 \kappa z} e^{j(\omega t - n' K_0 z)} = E_0 e^{-2\pi\kappa z/\lambda} e^{jK_0(c_0 t - n' z)} \tag{4.6.9}$$

Eine elektromagnetische Welle wird in einem Medium entsprechend dem Gesetz

$$E(z) = E_0 e^{-(2\pi\kappa/\lambda)z} \tag{4.6.10}$$

abgeschwächt. Der Imaginärteil des Brechungsindexes ist für die Absorption zuständig. Bei Gasen ist $n$ nur unwesentlich grösser als 1. Es gilt dann in guter Näherung, dass $n^2 - 1 \simeq 2(n-1)$ ist. betrachten wir den Real- und den Imaginärteil in der Nähe einer Resonanzfrequenz, ist also $|\omega - \omega_0| \ll \omega_0$ so erhält man[?]

$$\kappa = \frac{Nq^2}{8\varepsilon_0 m\omega_0} \cdot \frac{\gamma}{(\omega-\omega_0)^2 + \frac{\gamma^2}{4}} \tag{4.6.11}$$

$$n' = 1 + \frac{Nq^2}{4\varepsilon_0 m\omega_0} \cdot \frac{\omega-\omega_0}{(\omega-\omega_0)^2 + \frac{\gamma^2}{4}} \tag{4.6.12}$$

In der Literatur wird das Absorptionsgesetz üblicherweise mit Intensitäten formuliert. Die Lichtintensität wird beim Durchgang durch ein Medium wie

$$dI = -\alpha I dz \tag{4.6.13}$$

abgeschwächt. Aus dieser Gleichung folgt das Beer-Lambert'sche Absorptionsgesetz

$$I(z) = I_0 e^{-\alpha z} \tag{4.6.14}$$

Die aus einem klassischen Modell für eine Atom abgeleitete Absorption kann in den Beer-Lambert'schen Absorptionskoeffizienten durch quadrieren umgerechnet werden (die Intensität ist proportional zum Quadrat der Amplitude).

$$\alpha = \frac{4\pi\kappa}{\lambda} = 2 \cdot \frac{2\pi}{\lambda} \cdot \kappa = 2 \cdot K \cdot \kappa \tag{4.6.15}$$

Die Abbildung 4.237 zeigt einen Aufbau zur Messung der Absorption. Das Licht aus einer Quelle, entweder polychrom oder monochrom und eventuell durchstimmbar, wird durch das zu untersuchende Medium der Länge $L$ geschickt. Die Länge des

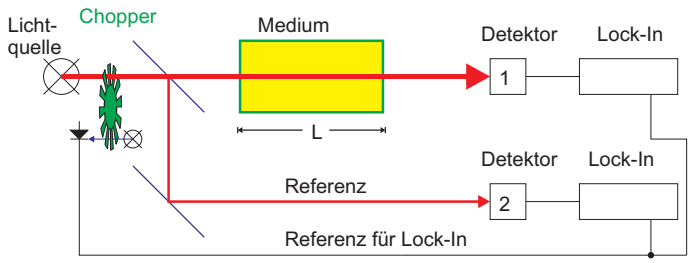



Abbildung 4.238.: Verbesserter Aufbau zur Messung einer Absorption mit Choppern

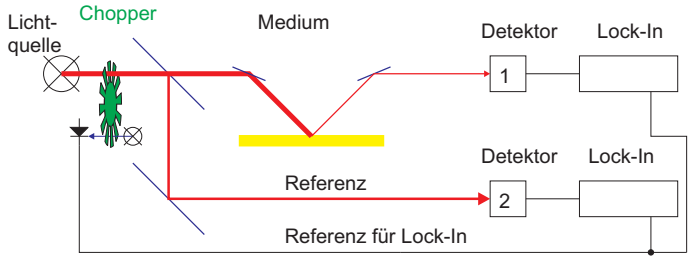



Abbildung 4.239.: Aufbau zur Messung der Reflektivität

Mediums muss so bemessen werden, dass der Detektor 1 in einem Bereich betrieben wird, in dem sein Signal-zu-Rausch-Verhältnis noch genügend ist. Bei Proben mit einem grossen $\alpha$ muss die Länge klein sein, bei Proben mit einer kleinen Absorption wie bei Gasen ist eine grosse Wirkungslänge notwendig.
Da die Lichtquellen nicht immer stabil arbeiten, oder da sie, wenn sie in der Frequenz durchgestimmt werden, ihre Intensität ändern, ist es oftmals notwendig, den in Abbildung 4.237 angegebenen Referenzzweig zu verwenden. Das Signal des Detektors 1 wird durch das Signal des Detektors 2 geteilt. Da die Kennlinie der verwendeten Detektoren nichtlinear sein kann, sollten beide mit etwa der gleichen Lichtintensität betrieben werden.
Bei einem empfindlichen Messaufbau können langsame Schwankungen des Umgebungslichtes wie auch nicht zu kontrollierende Streulichtquellen stören. Wie im Abschnitt 2.8 über Rauschen gezeigt, sind Messungen über lange Zeiten besonders vom $1/f$-Rauschen betroffen. Deshalb wird bei den meisten optischen Messungen das Licht mit einer Frequenz von etwa $1kHz$ moduliert. Oft wird dies wie in der Abbildung 4.238 gezeigt, ein Chopperrad verwendet. Mit einer Lichtschranke wird die Taktfreqenz gemessen und als Referenz in die den Detektoren nachgeschalteten Lock-In-Verstärker eingespiesen. Damit lassen sich die meisten Störquellen genügend stark unterdrücken.

## 4.6.2. Reflexionsmessung

Eine Reflektivitätsmessung (Abbildung 4.239) ist ähnlich aufgebaut wie eine Absorptionsmessung. Das Licht aus einer wird in einen Referenzstrahl und einen Teststrahl aufgeteilt. Beide werden mit einem *Chopper-Rad* getaktet. Das reflektierte Licht wird gesammelt und auf einen Detektor gebracht. Die Reflexion kann spekular oder diffus sein.
Eine mögliche Anwendung von Reflexionsmessungen ist die Bestimmung von Schad-

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

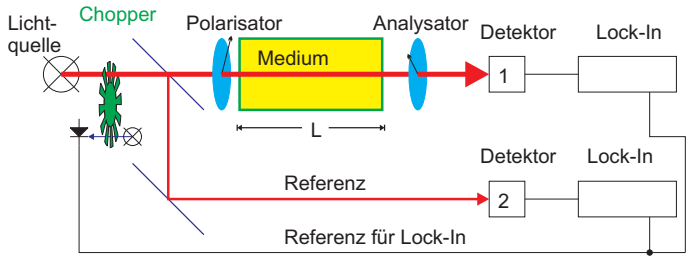



Abbildung 4.240.: Aufbau zur Messung der Polarisation

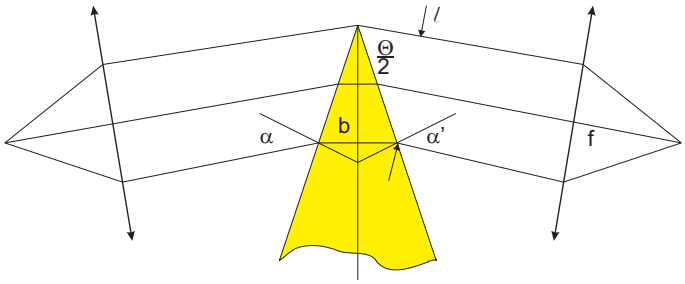



Abbildung 4.241.: Prismenspektrometer

stoffen in der Luft. Eine andere mögliche Anwendung ist die Messung der induzierten Transparenz oder der induzierten Brechzahländerung in Halbleiterproben. Dabei müssen jedoch sehr kleine Unterschiede der Reflektivität in der Gegenwart eines grossen Untergrundes bestimmt werden.

## 4.6.3. Polarisationsmessung

Auch der Aufbau für Polarisationsmessungen (Abbildung 4.240 ist sehr ähnlich dem Aufbau zur Messung der Absorption. Das Licht aus einer wird in einen Referenzstrahl und einen Teststrahl aufgeteilt. Beide werden mit einem *Chopper-Rad* getaktet. Der Polarisationszustand des Teststrahls wird in einem Polarisator vor der Probe festgelegt. nach der Transmission der Probe wird der Polarisationszustand mit einem Analysator ausgemessen. Der restliche Aufbau ist analog zu weiter oben diskutierten Versuchsaufbauten.
Die Messanordnung nach Abbildung 4.240 kann zum beispiel benutzt werden, um den Kerr-Effekt oder die Chiralität von Molekülen auszumessen.

## 4.6.4. Spektrometer und Polychromatoren

Spektrometer dienen zur Messung der Wellenlängenabhängigkeit der Intensität von Licht. Dieses Licht kann entweder direkt von einer Quelle abgeleitet sein, oder aber das Resultat eines optischen Experimentes sein.

### 4.6.4.1. Prismenspektrometer

Ein gebräuchliches Spektrometer wenn die *spektrale Auflösung* nicht allzu hoch sein soll, ist das Prismenspektrometer analog zur Abbildung 4.241. Nach Perez[**?**]

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

kann die Dispersion von Glas durch

$$n^2 = A_{-1}\lambda^2 + A_0 + A_1\lambda^{-2} + A_2\lambda^{-4} + A_3\lambda^{-6} + A_4\lambda^{-8} = \sum_{i=-1}^{\infty} A_i\lambda^{-2i} \tag{4.6.16}$$

beschrieben werden. Für eine bestimmte Glassorte sind die Parameter

$$\begin{aligned} A_{-1} &= -1.0108077 \cdot 10^{-2} \\ A_0 &= 2.2718929 \\ A_1 &= 1.0592509 \cdot 10^{-2} \\ A_2 &= 2.0816965 \cdot 10^{-4} \\ A_3 &= -7.6472538 \cdot 10^{-6} \\ A_4 &= 4.9240991 \cdot 10^{-7} \end{aligned} \tag{4.6.17}$$

Die Dispersion eines Materials $dn/d\lambda$ kann aus der Gleichung (4.6.16)berechnet werden. Wir setzen $f = n^2$ und erhalten

$$\frac{dn}{d\lambda} = \frac{d(\sqrt{f})}{d\lambda} = \frac{1}{2\sqrt{f}}\frac{df}{d\lambda} = \frac{\sum_{i=-1}^{\infty}(i-1)A_i\lambda^{-2i-1}}{\sqrt{\sum_{i=-1}^{\infty} A_i\lambda^{-2i}}} \tag{4.6.18}$$

Bezeichnet man mit $\delta$ den Winkel zwischen dem einfallenden Strahl vor dem Prisma und dem das Prisma verlassenden Strahl, so kann man die Winkeldispersion

$$D_w \equiv \frac{d\delta}{d\lambda} = \frac{d\delta}{dn}\frac{dn}{d\lambda} \tag{4.6.19}$$

Der zweite Term hängt nur vom Material ab, der erste Term beschreibt die Geometrie. Nach Perez[**?**] gilt für Prismen

$$\begin{aligned} \sin\alpha &= n\sin\beta \\ \sin\alpha' &= n\sin\beta' \\ \Theta &= \beta + \beta' \\ \delta &= \alpha + \alpha' - \Theta \end{aligned} \tag{4.6.20}$$

Da $\alpha$ und $\Theta$ konstant sind, gilt auch

$$\begin{aligned} \frac{d\delta}{dn} &= \frac{d\alpha'}{dn} \\ 0 &= dn\sin\beta + n\cos\beta d\beta \\ d\beta + d\beta' &= 0 \\ \cos\alpha' d\alpha' &= dn\sin\beta' + n\cos\beta' d\beta' \end{aligned} \tag{4.6.21}$$

Daraus folgt

$$\frac{d\delta}{dn} = \frac{\sin\beta'}{\cos\alpha'} - \frac{n\cos\beta'}{\cos\alpha'}\cdot\frac{d\beta}{dn} = \frac{\sin\beta'}{\cos\alpha'} - \frac{n\cos\beta'}{\cos\alpha'}\cdot\frac{\sin\beta}{\cos\beta} = \frac{\sin\Theta}{\cos\beta\cos\alpha'} \tag{4.6.22}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Im Minimum der Ablenkung hat eine Änderung des Einfallswinkels nur einen kleinen Einfluss auf die Ablenkung. Der Ausgangswinkel hängt wie

$$\frac{d\delta}{dn} = \frac{2\sin\left(\frac{\Theta}{2}\right)\cdot\cos\left(\frac{\Theta}{2}\right)}{\cos\left(\frac{\Theta}{2}\right)\cdot\cos\alpha'} = \frac{2\overline{OS}\sin\left(\frac{\Theta}{2}\right)}{\overline{OS}\cos\alpha'} = \frac{b}{\ell} \tag{4.6.23}$$

von den Eingangsgrössen ab. Die Winkeldispersion ist also

$$D_w = \frac{b}{\ell}\cdot\frac{dn}{d\lambda} \tag{4.6.24}$$

Wenn man das Spektrum in der Brennebene einer Linse mit einem Schirm oder einer CCD-Kamera beobachtet, benötigt man die Lineardispersion

$$D_\ell = fD_w = f\frac{b}{\ell}\cdot\frac{dn}{d\lambda} \tag{4.6.25}$$

Das Auflösungsvermögen eines Prismenspektrographen ist über

$$A \equiv \frac{\lambda}{\Delta\lambda} \tag{4.6.26}$$

definiert. Wenn die optischen Elemente von genügender Qualität sind, dann hängt das Auflösungsvermögen des Spektrometers von der Beugung des Lichtes am Eintrittsspalt ab. Bei einem (auch fiktiven) Spalt der Breite $B$ ist die Halbwertsbreite des Beugungsmusters nach Rayleigh

$$\Delta X_{\frac{1}{2}} = \frac{\lambda f}{B} \tag{4.6.27}$$

Damit wird

$$\Delta\lambda = \frac{X_{\frac{1}{2}}}{D_\ell} = \frac{\lambda f}{l\cdot f\cdot(b/l)\cdot(dn/d\lambda)} = \frac{\lambda}{b(dn/d\lambda)} \tag{4.6.28}$$

Somit ist das Auflösungsvermögen eines Prismenspektrometers durch

$$A_0 = \frac{\lambda}{\Delta\lambda} = \frac{\ell}{f}D_\ell = b\frac{dn}{d\lambda} \tag{4.6.29}$$

Wenn man das Bild des Spaltes vor der Lichtebene in der Detektionsebene vergrössert, dann wird

$$\Delta\lambda = \frac{s}{D_\ell} \tag{4.6.30}$$

über die Grösse des Bildes $s$ festgelegt. Damit ist das Auflösungsvermögen

$$A = \frac{\lambda}{\Delta\lambda} = \frac{\lambda}{s}D_\ell \tag{4.6.31}$$

Das Prismenspektrometer ist optimal eingestellt[**?**], wenn

$$\begin{aligned} s &= \frac{\lambda f}{\ell} \\ A &= \frac{\lambda D-\ell}{s} = b\frac{dn}{d\lambda} \end{aligned} \tag{4.6.32}$$

gilt. Typischerweise kann man ein Auflösungsvermögen (bei voll ausgeleuchtetem Prisma!) von $A = 2000$ erreichen.

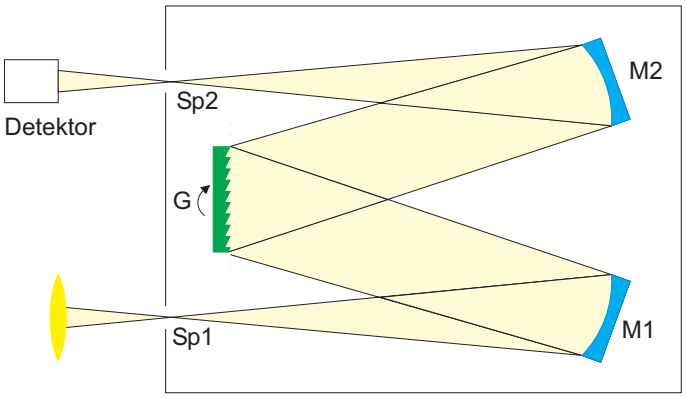



Abbildung 4.242.: Gitterspektrometer

#### 4.6.4.2. Gitterspektrometer

In einem *Gitterspektrometer* wird ein Beugungsgitter als dispersives Element verwendet. Abbildung 4.242 zeigt einen typischen Aufbau eines solchen Spektrometers. Licht tritt durch einen Eintrittsspalt ein und wird durch den Hohlspiegel $M1$ in paralleles Licht umgewandelt. Wie bei einer Linse wird der Öffnungswinkel durch eine *Numerische Apertur* charakterisiert. Das parallele Licht wird durch das Gitter $G$ in seine spektralen Anteile zerlegt. Eine bestimmte Richtung wird durch den Hohlspiegel $M2$ auf den Austrittsspalt fokussiert. Hinter diesem befindet sich der Detektor. Alternativ kann anstelle der Kombination aus Spalt und Detektor ein ein- oder zweidimensionaler Detektor verwendet werden. Dies kann eine Diodenzeile, eine CCD-Zeile oder ein flächenhafter CCD-Detektor sein.
Um eine optimale Empfindlichkeit zu erreichen müssen die folgenden Voraussetzungen erfüllt sein:

- Das zu untersuchende Licht muss einen zur Numerischen Apertur des Gitterspektrometers passenden Öffnungswinkel haben. Ist der Öffnungswinkel zu gering, wird das Gitter nicht voll ausgeleuchtet: die mögliche *Auflösung* wird nicht erreicht.
- Die Spaltöffnung muss so bemessen sein, dass genügend Licht passieren kann und dass die Beugung noch keinen bestimmenden Einfluss auf die Messung hat.
- Die Anzahl Linien pro Länge des Gitters muss zum gewünschten Wellenlängenumfang und zur gewünschten *Auflösung* passen.
- Der Austrittsspalt (oder die Grösse der Pixel der CCD) muss an den Eintrittsspalt angepasst sein.

Die Beugungserscheinungen an einem Gitter können mit einer Vektorgleichung beschrieben werden. Ist $\boldsymbol{a}$ der Gittervektor des Gitters und ist $\boldsymbol{k}_0$ der Wellenvektor des einfallenden Lichtes und $\boldsymbol{k}$ der Wellenvektor des gestreuten Lichtes, so gilt mit der Vereinbarung $\boldsymbol{K} = \boldsymbol{k} - \boldsymbol{k}_0$

$$\boldsymbol{a} \cdot \boldsymbol{K} = m \cdot 2\pi \quad m \in \mathbb{Z} \tag{4.6.33}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

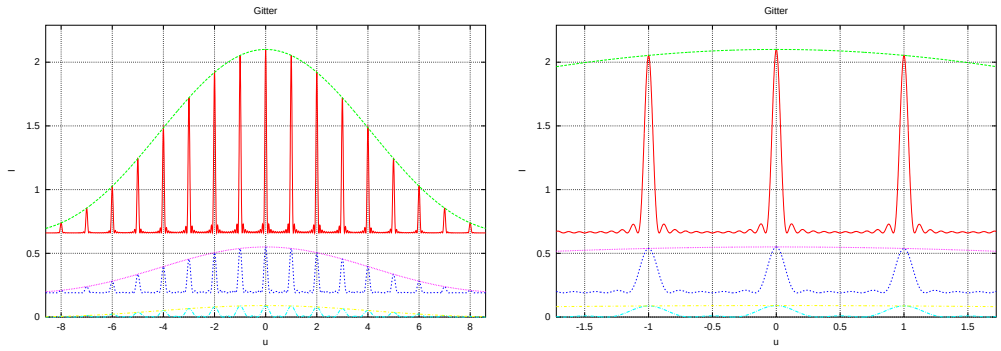



Abbildung 4.243.: Intensität des gebeugten Lichtes als Funktion der Anzahl beleuchteter Spalte. Links ist eine Übersicht, rechts die Detailansicht. Von unten nach oben sind $N = 3{,}6$ und 12 beleuchtete Spalte dargestellt. Die sonstigen Gitterparameter sind $a = 1$ und $\epsilon = 0.1$

Diese Gleichung kann auch mit Winkeln formuliert werden. Sei $\Theta_0$ der Winkel des einfallenden Lichtes zur Senkrechten auf das Gitter und sei $\Theta$ der entsprechende Winkel des gebeugten Lichtes und sei $\alpha \equiv \Theta_0 - \pi$, dann ist

$$a\,(\sin\Theta + \sin\alpha) = m\lambda \quad m \in \mathbb{Z} \tag{4.6.34}$$

Das Gitter soll nun $N$ Spalte haben. Wir setzen $u = \frac{\alpha-\alpha_0}{\lambda}$, wobei $\alpha = \sin\Theta$ und $\alpha_0 = \sin\Theta_0$ die Richtungskosinusse der gebeugten und der einfallenden Welle sind[?]. Die Spaltbreite eines einzelnen Spaltes sei $\epsilon$. Dann ist die gestreute Intensität hinter dem Gitter

$$I(u) = N^2\epsilon^2 \left[\frac{\sin(\pi u\epsilon)}{\pi u\epsilon}\right]^2 \cdot \left[\frac{\sin(N\pi ua)}{N\sin(\pi ua)}\right]^2 \tag{4.6.35}$$

Wenn das Gitter nicht voll ausgeleuchtet wird, muss für $N$ die Anzahl beleuchteter Gitterstriche eingesetzt werden.
Abbildung 4.243 zeigt den Intensitätsverlauf in Abhängigkeit der Anzahl beleuchteter Gitterstriche $N$. Aus (4.6.35) ist ersichtlich, dass die Intensität wie $N^2$ sich verhält. Wenn also ein Gitter in einem Gitterspektrometer nicht richtig ausgeleuchtet ist, verliert man sehr schnell sehr viel an Intensität auf dem Detektor.
Die Höhe der Beugungsmaxima ist durch

$$I_m(u) = N^2\epsilon^2 \left[\frac{\sin(\pi u\epsilon)}{\pi u\epsilon}\right]^2 \tag{4.6.36}$$

gegeben. Weiter kann man berechnen, dass die Breite eines Maximums proportional zur beleuchteten Breite $L = N \cdot a$ ist.
Die Winkeldispersion eines Gitters wird aus (4.6.34) berechnet:

$$a\cos\Theta d\Theta = md\lambda \quad m \in \mathbb{Z} \tag{4.6.37}$$

Im Gegensatz zu einem Prisma werden längere Wellenlängen stärker gebeugt. Die

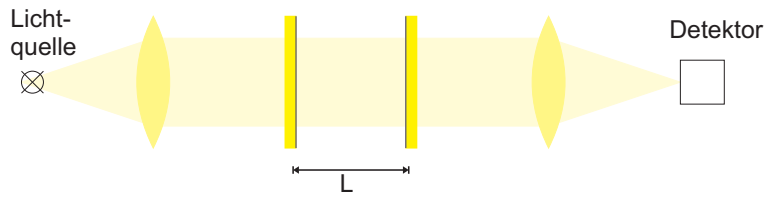



Abbildung 4.244.: Aufbau eines Fabri-Perot-Spektrometer

Winkeldispersion ist

$$D_w = \frac{d\Theta}{d\lambda} = \frac{m}{a\cos\Theta_m} = \frac{1}{\lambda}\cdot\frac{\sin\Theta_m - \sin\Theta_0}{\cos\Theta_m} \qquad m \in \mathbb{Z} \tag{4.6.38}$$

Für höhere Ordnungen ist die Winkeldispersion also grösser. Analog zum Prisma kann auch bei einem Gitter eine lineare Dispersion definiert werden.

$$D_\ell = fD_w = \frac{f}{\lambda}\cdot\frac{\sin\Theta_m - \sin\Theta_0}{\cos\Theta_m} \qquad m \in \mathbb{Z} \tag{4.6.39}$$

Auch beim Gitter gibt es einen Einfallswinkel, bei dem die Ablenkung minimal ist. Der Ablenkwinkel ist $\delta \equiv \Theta - \Theta_0$. Daraus folgt, dass

$$\frac{d\delta}{d\Theta_0} = \frac{d\Theta}{d\Theta_0} - 1 = \frac{\cos\Theta_0}{\cos\Theta} \tag{4.6.40}$$

sein muss. Man sieht, dass entweder $\Theta = \Theta_0$ oder $\Theta = -\Theta_0$ sein muss, um einen Extremwert von $\delta$ zu erhalten. Der erste Fall entspricht dem nicht gebeugten Strahl und ist nicht von Interesse. Eine Analyse der zweiten Ableitung zeigt, dass dort die Ablenkung des Strahls ein Minimum hat.
Das Auflösungsvermögen eines Gitters[**?**] ist

$$A = \frac{\lambda}{\Delta\lambda} = m\cdot N = \frac{L}{\lambda}\left(\sin\Theta_m - \sin\Theta_0\right) \qquad m \in \mathbb{Z} \tag{4.6.41}$$

Zum Beispiel hat ein Gitter mit einer Breite von $2cm$ und 500Linien/mm bekommt man $A = 20000$. Das Auflösungsvermögen hat ein Maximum. In (4.6.41) können die beiden Sinus dem Betrage nach maximal 1 werden. Also gilt

$$A = \frac{L}{\lambda}\left(\sin\Theta_m - \sin\Theta_0\right) \le \frac{L}{\lambda}\left(1+1\right) = \frac{2L}{\lambda} \tag{4.6.42}$$

Für das vorhin diskutierte Gitter ist also $A \le 80000$ bei einer Wellenlänge von $\lambda = 0.5\mu m$. Da bei höheren Ordnungen $m$ die maximale Intensität schnell abnimmt, wird man ein Gitter im allgemeinen mit einer Ordnung in der Nähe von eins betreiben.
Ein Gitterspektrometer hat bei gleicher Grösse von Gitter oder Prisma etwa eine zehn mal bessere *Auflösung*.

#### 4.6.4.3. Fabri-Perot-Spektrometer

Wenn eine sehr hohe *spektrale Auflösung* gefordert ist, verwendet man häufig *Fabri-Perot-Spektrometer*. Bei diesen Spektrometern wird, wie in der Abbildung 4.244 gezeigt, Licht in einen optischen Resonator eingekoppelt. Da das licht teilweise mit sich selber interferiert, können nur Eigenmoden des Resonators oder Frequenzen

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

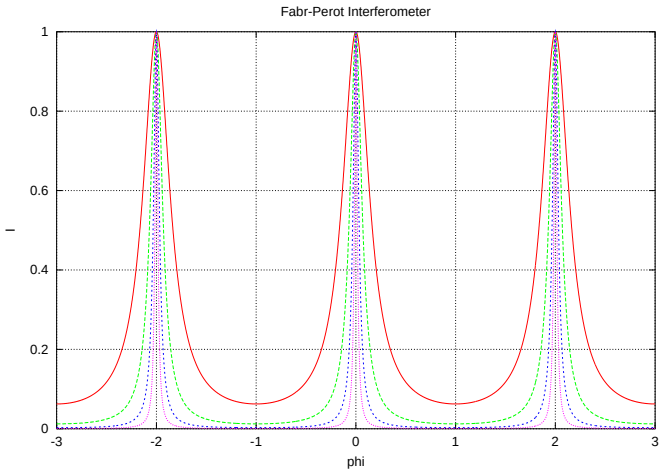



Abbildung 4.245.: Kennlinie eines Fabri-Perot-Inerferometers für Reflektivitäten von 0.6, 0.8, 0.9 und 0.95 beziehungsweise Werte von $M$ von 15, 80, 360 und 1520 (von oben)

in deren Nähe durch den Resonator propagieren. Wenn $R$ die für beide Spiegel gleiche Reflektivität ist, und $M = 4R/(1-R^2)$ist, dann wird die transmittierte Intensität

$$I_t(\phi) = \frac{I_{max}}{1 + M \cdot \sin^2(\phi)} \tag{4.6.43}$$

$\phi$ ist dabei die Phasenveschiebung beim Umlauf um den Resonator.
Abbildung 4.245 zeigt den Intensitätsverlauf als Funktion der Phase $\phi$ für $M = [15, 80, 360, 1520]$.
Wir betrachten nun einen Lichtstrahl, der unter dem Winkel $\alpha$ zur Achse des *Fabri-Perot-Interferometers* auf das Interferometer trifft. Wenn weiter der Abstand der beiden planparallelen Spiegel $L$ sei, ist die Phasendifferenz im Brennpunkt der Linse durch

$$\phi = \frac{2\pi}{\lambda} 2L \cdot \cos\alpha \tag{4.6.44}$$

gegeben. Die Phasendispersion (das Äquivalente zur Winkeldispersion) wird dann

$$D_\phi = -2\pi \frac{2L \cdot \cos\alpha}{\lambda^2} \tag{4.6.45}$$

Wenn man annimmt, dass $\Delta\phi_{\frac{1}{2}}$ die minimal messbare Phasendifferenz bei der Halbwertsbreite der Transmissionsfunktion ist, dann ist die minimal auflösbare Wellenlängendifferenz $\Delta\lambda = \Delta\phi_{\frac{1}{2}}/|D_\phi|$. Damit ist das Auflösungsvermögen durch

$$A = \frac{\Delta\lambda}{\lambda} = \frac{\pi M^{\frac{1}{2}} \cdot L \cdot \cos\alpha}{\lambda} \tag{4.6.46}$$

gegeben. Mit den Definitionen

| R | 0.60 | 0.80 | 0.90 | 0.95 | 0.99 |
|---|---|---|---|---|---|
| M | 15 | 80 | 360 | 1520 | 39600 |
| $\mathcal{F}$ | 6 | 14 | 30 | 61 | 313 |
| A | $0.24 \times 10^6$ | $0.56 \times 10^6$ | $1.2 \times 10^6$ | $2.45 \times 10^6$ | $125 \times 10^6$ |
| $\Delta\lambda(pm)$ | 2 | 0.89 | 0.42 | 0.21 | 0.004 |

Tabelle 4.12.: Auflösungsvermögen, kleinste auflösbare Wellenlängenänderung und Finesse eines Fabri-Perot-Interferometers[?]. Die konstanten Werte sind $\lambda = 0.5\mu m$, $L = 1cm$ und $\alpha \approx 0$. daraus folgt $p = 40000$.

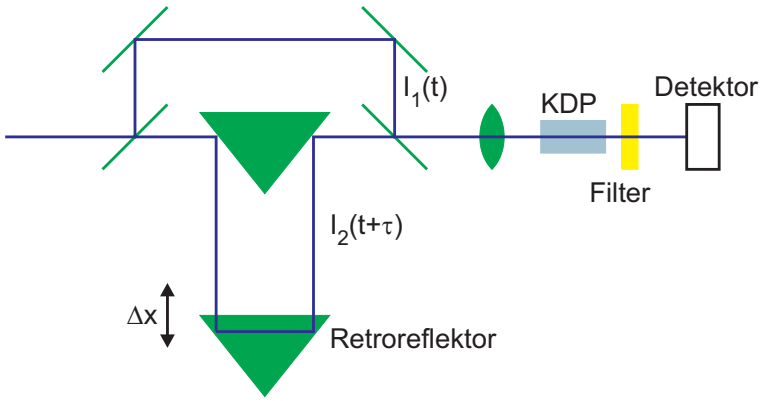



Abbildung 4.246.: Optischer Korrelator

$$\begin{aligned} p &= \frac{2L \cdot \cos\alpha}{\lambda} \\ \mathcal{F} &= \frac{\pi M^{\frac{1}{2}}}{2} = \frac{\pi R^{\frac{1}{2}}}{1-R} \end{aligned} \tag{4.6.47}$$

wird das Auflösungsvermögen

$$A = p\mathcal{F} \tag{4.6.48}$$

Die Grösse $\mathcal{F}$ heisst *Finesse*. $p$ ist die Interferenzordnung. Die Tabelle 4.12 zeigt eine Zusammenstellung der erreichbaren *Auflösung*. Diese ist für nicht allzu gute Spiegel etwa 100 bis 1000 mal besser als Gitterspektrometer und etwa 1000 bis 10000 mal besser als Prismenspektrometer.

## 4.6.5. Messverfahren für kurze Zeiten

Die Messung ultrakurzer Pulse sowie die Messung von Vorgängen, die kürzer als etwa eine ps sind, sind mit rein elektrischen Verfahren nicht möglich. Die beste Zeitauflösung erreichen optische Verfahren und elektrooptische Verfahren.
Ein optischer Korrelator, wie er in Abb. 4.246 gezeigt besteht aus einem Strahlteiler, in dem der ankommende Puls geteilt wird, einer Verzögerungsleitung, die einen Puls zeitlich verzögert wird, sowie einem nichtlinearen Medium, in dem die beiden Pulse überlagert werden. Der einkommende Puls soll die Intensität $I(t) = c\varepsilon_0 E^2(t)$ und die Halbwertsbreite $\Delta T$. Dieser Puls wird in zwei Teilpulse $I_1(t)$ und $I_2(t)$ aufgeteilt. Dabei durchlaufen die beiden Pulse unterschiedliche Wege $S_1$ und $S_2$. Der

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Wegunterschied $\Delta S = S_2 - S_1$ ist äquivalent zu einem Zeitunterschied $\tau = \Delta S/c$. Nach der Überlagerung der beiden Pulse ist die Intensität

$$\begin{aligned} I(t,\tau) &= c\varepsilon_0 \left[E_1(t) + E_2(t+\tau)\right]^2 \\ &= c\varepsilon_0 \left[E_0(t)\cos\omega t + E_0(t+\tau)\cos\omega(t+\tau)\right]^2 \end{aligned} \tag{4.6.49}$$

Ein linearer Detektor mit einer Integrationszeit $T$ würde das *Signal*

$$\begin{aligned} S^{(1)} &= \langle I(T,\tau)\rangle = \frac{1}{2T}\int_{-T}^{+T} I(t,\tau)dt \\ &= c\varepsilon_0 \left[\left\langle E_0^2\right\rangle + \frac{1}{T}\int_{-T}^{+T} E_0(t)E_0(t+\tau)\cos\omega t\cdot\cos\omega(t+\tau)dt\right] \end{aligned} \tag{4.6.50}$$

Für Zeiten $T$, die gross gegen die Pulsdauer $\Delta T$ sind, kann das Integral über alle Zeiten erweitert werden. Als Resultat wird die Korrelationsfunktion erster Ordnung $G^{(1)}$ erhalten.

$$G^{(1)}(\tau) = \int_{-\infty}^{+\infty} \frac{E(t)E(t+\tau)}{E^2(t)}dt = \frac{\langle E(t)\rangle\,\langle E(t+\tau)\rangle}{\langle E^2(t)\rangle} \tag{4.6.51}$$

messen. An den Grenzen hat die Korrelationsfunktion $G^{(1)}$ die Werte $G^{(1)}(0) = 1$ und $G^{(1)}(\infty) = 0$. Bei monochromatischem Licht würde die Funktion $G^{(1)}(\tau)$ mit der Periode $T = \lambda/(2c)$ oszillieren. Da jedoch kurze Pulse nicht monochromatisch sind, verschmiert sich die Korrelation. Ein langsamer Detektor, dessen Zeitkonstante viel grösser als die Pulsdauer ist, misst ein konstantes Signal. Dies muss so sein, da das Ausgangssignal des Detektors proportional zur eingestrahlten Energie und nicht proportional zur eingestrahlten Leistung ist.
Um eine *zeitliche Auflösung* zu bekommen, ist es notwendig, einen nichtlinearen Detektor zu verwenden. Eine Komponente der Übertragungsfunktion[**?**] könnte

$$I\left(2\omega,t,\tau\right) = A\left[I_1(t) + I_2(t+\tau)\right]^2 \tag{4.6.52}$$

sein. Die Konstante $A$ gibt die Stärke der Nichtlinearität an. Das Signal bei der Frequenz $2\omega$ ist

$$S(2\omega,t) = \frac{A}{T}\cdot\int I(2\omega,t,\tau)dt = A\left[\left\langle I_1^2\right\rangle + \left\langle I_2^2\right\rangle + 4\left\langle I_1(t)\cdot I_2(t+\tau)\right\rangle\right] \tag{4.6.53}$$

Die ersten beiden Terme sind von $\tau$ unabhängig. Sie ergeben ein konstantes Untergrundssignal. Der letzte Term enthält die Information über die Pulsform. Nach Demtröder rührt der Faktor 4 von der Addition der **Wahrscheinlichkeiten** der Photonen 1 und 2 und umgekehrt her. Die Korrelationsfunktion 2. Ordnung ist

$$G^{(2)}(\tau) = \frac{\int I(t)I(t+\tau)dt}{\int I^2(t)dt} = \frac{\langle I(t)\cdot I(t+\tau)\rangle}{\langle I^2(t)\rangle} \tag{4.6.54}$$

Wenn $I_1 = I_2 = \frac{1}{2}I$ ist, wird das detektierte Signal

$$S(2\omega,\tau) = A\left[G^{(2)}(0) + 2G^{(2)}(\tau)\right] = A\left[1 + 2G^{(2)}(\tau)\right] \tag{4.6.55}$$

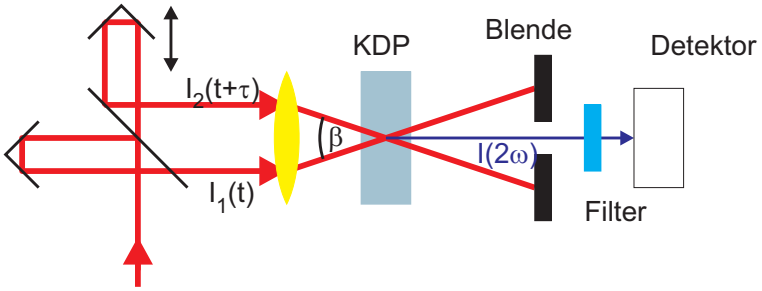



Abbildung 4.247.: Untergrundfreier Korrelator. Die Nichtlinearität in einem Material wie KDP (Kaliumdihydrogenphosphat) dient zur Korrelation.

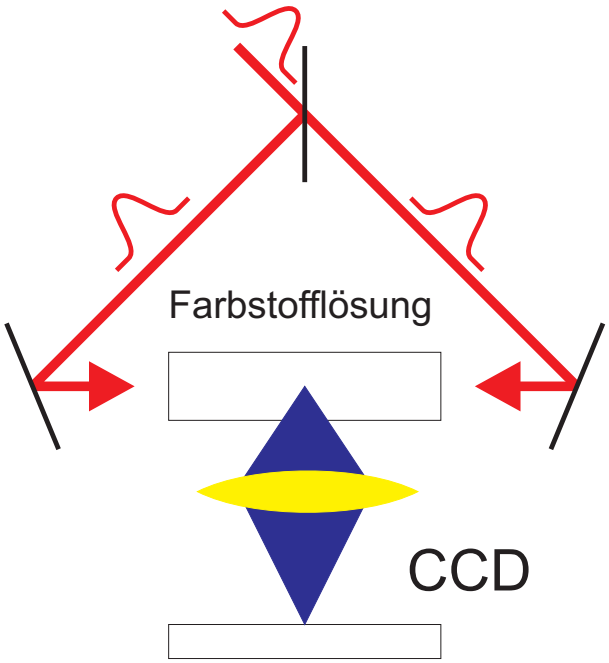



Abbildung 4.248.: Zweiphotonenfluoreszenz zur Messung von kurzen Pulsen

Für eine Zeitverzögerung von $\tau = 0$ wird das Signal $S(2\omega, \tau) = 3A$ maximal. Wenn die Zeitverzögerung sehr gross wird, geht der term $G^{(2)} \to 0$, und das Ausgangssignal wird konstant $S(2\omega, \tau) = A$, aber nicht null.
Die Messanordnung in der Abbildung 4.247 ermöglicht eine untergrundsfreie Messung. Um in einem Verdopplerkristall ein Signal bei der Frequenz $2\omega$ zu erzeugen, muss sowohl der Ort der beiden Impulse übereinstimmen wie auch der resultierende $\boldsymbol{k}$-Vektor. In der Abbildung 4.247 ist die Orientierung des Kristalls so gewählt, dass zwei Photonen aus dem gleichen Teilstrahl kein frequenzverdoppeltes Photon erzeugen können, da ihre $\boldsymbol{k}$-Vektoren nicht mit der Gitterorientierung kompatibel sind[**?**]. Nur wenn zwei Photonen aus je aus einem der beiden Teilstrahlen sich überlagern, stimmt die Impulsbedingung und ein frequenzverdoppeltes Photon kann entstehen.
Die Frequenzverdoppelung ist nicht der einzige nichtlinear-optische Effekt der zur Erzeugung einer Korrelation zweiter Ordnung benützt werden kann. Die Abbildung 4.248 zeigt schematisch einen Aufbau zur Zwei-Photonen-Absorption in einer Flüssigkeit. Wieder ist die Wahrscheinlichkeit, eine Absorption bei der halben

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

| Pulsform | Gleichung | $\frac{\Delta\tau}{\Delta T}$ | $\Delta T \cdot \Delta\nu$ |
|---|---|---|---|
| Rechteck | $I(t) = \begin{cases} 1 & \text{für} \quad 0 \le t \le \Delta T \\ 0 & \text{sonst} \end{cases}$ | 1 | 0.886 |
| Gauss-Profil | $\exp\left(-\frac{t^2}{0.36\Delta T^2}\right)$ | $\sqrt{2}$ | 0.441 |
| Hyperbolisches Sekansprofil | $\mathrm{sech}^2\left(\frac{t}{0.57\Delta T}\right)$ | 1.55 | 0.315 |

Tabelle 4.13.: Breiten von Autokorrelationsfunktion und Laserpuls nach Demtröder[?]

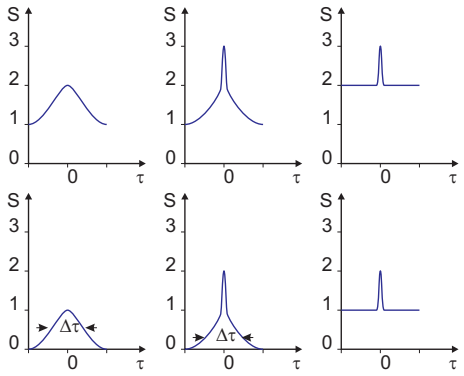



Abbildung 4.249.: Autokorrelationsprofile. Gezeigt werden a) ein Fourier-limitierter Puls, b) ein Rauschpuls, c) kontinuierliches Rauschen. Die obere Reihe zeigt das Messresultat bei einer Messung mit Untergrund, die untere bei einer untergrundfreien Messung

Wellenlänge des eingestrahlten Lichtes zu erhalten proportional zum Quadrat der Intensität.

Da in dem gezeigten Versuchsaufbau zwei gegenläufige Pulse sich überlagern, entsteht ein räumliches Interferenzmuster. Absorption bei der halben Wellenlänge und die daraus folgende Emission von Fluireszenzlicht ist auf den räumlichen Bereich der Überlappung beider Pulse beschränkt. Da ddiese pulse senkrecht zu ihrer Ausbreitungsrichtung ausgedehnt sind, ist es nicht möglich, eine Wechselwirkungszone, die kleiner als 0.3 mm ist, abzubilden. daher ist die *zeitliche Auflösung* der Messmethode für kurze Pulse nach Abbildung 4.247 auf Pulse mit einer Länge grösser als 1 ps beschränkt.

Die Gleichung (4.6.54) zeigt, dass $G^{(2)}$ symmetrisch in $\tau$ ist. deshalb lässt sich aus der Autokorrelationsfunktion keine aussage über die Pulsform machen. Auch zur Abschätzung der Pulsdauer muss eine Modellannahme über das Pulsprofil gemacht werden. Die Tabelle 4.13 zeigt, wie die 'gemessene' Pulsdauer von der profilform abhängt. Um die Pulsform zu bestimmen müssten Korrelationen höherer Ordnung (wie die Frequenzverdreifachung) gemessen werden.

Abbildung 4.249 zeigt schliesslich, dass auch weisses Rauschen (Siehe Abschnitt 2.8) ein Autokorrelationssignal erzeugt. Um Pulse eindeutig nachzuweisen muss die Funktion $1 + G^{(2)}(\tau)$ möglichst genau gemessen werden.

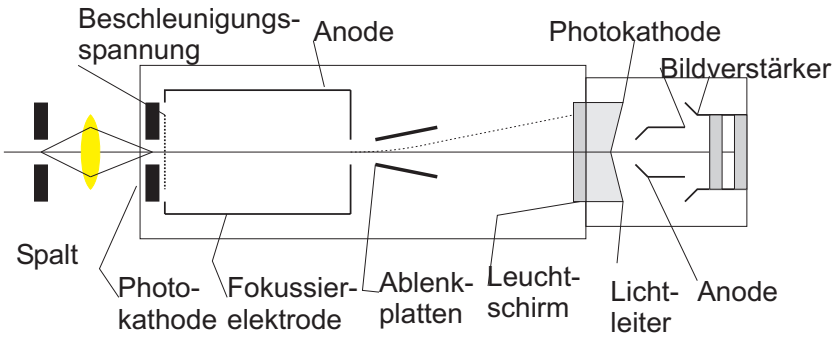



Abbildung 4.250.: Aufbau einer Streak-Kamera

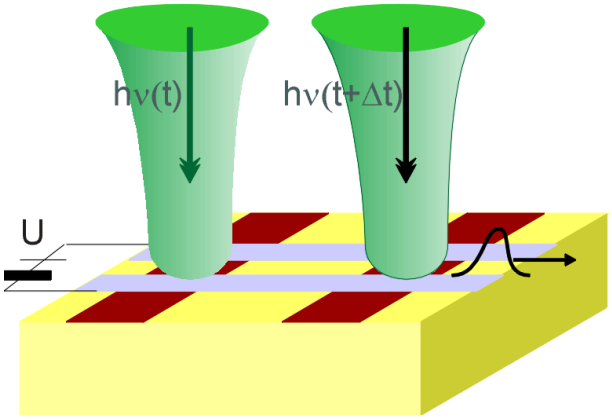



Abbildung 4.251.: Schalten eines Wellenleiters mit optischen Pulsen

## 4.7. Elektrooptische Messverfahren für kurze Zeiten

Eine kombinierte elektrooptische Methode zur Messung kurzer Pulse stellt die *Streakkamera* nach Abbildung 4.250 dar. Bei einer Streakkamera wird das zu messende Licht (entweder monochromatisch oder als Spektrum) linienförmig auf eine Photokathode gebracht. Die Photoelektronen werden durch eine Fokussierelektrode fokussiert und dann mit Ablenkplatten abgelenkt. Wenn an die Ablenkplatten eine schnell ansteigende Rampe angelegt wird, dann wird die zeitliche Abfolge des Signals auf dem leuchtschirm rechts in der Abbildung durch die Position kodiert. Das Bild auf dem leuchtschirm wird letztlich mit einem Bildverstärker soweit intensiviert, dass es auf einem Leuchtschirm mit genügender Intensität betrachtet oder dass es mit einer CCD-Kamera aufgezeichnet werden kann. Streakkameras haben eine *zeitliche Auflösung* von $400fs$ bis etwa $8ps$ abhängig von der Anstiegsgeschwindigkeit der Ablenkspannung.
Eine weitere Möglichkeit schnelle Signale zu generieren oder zu messen stellen Photoschalter dar. Abbildung 4.251 zeigt einen schematischen Aufbau eines Experimentes zur Messung der Ausbreitungseigenschaften von Pulsen auf einem Streifenleiter. Eine Spannung $U$ wird an den Streifenleiter angelegt. Der Streifenleiter soll auf einem hochohmigen Halbleitersubstrat aufgebracht werden. Die weinroten Streifen im Material stellen Stellen dar, in denen durch Licht Elektronen ins Leitungsband angeregt werden können. Wenn nun ein Lichtpuls auf der linken Seite des Substrates die beiden Leiter des Streifenleiters (siehe auch Abschnitt 4.3.2) kurzgeschlossen. Dadurch entsteht ein elektrischer Puls, der auf dem Strei-

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

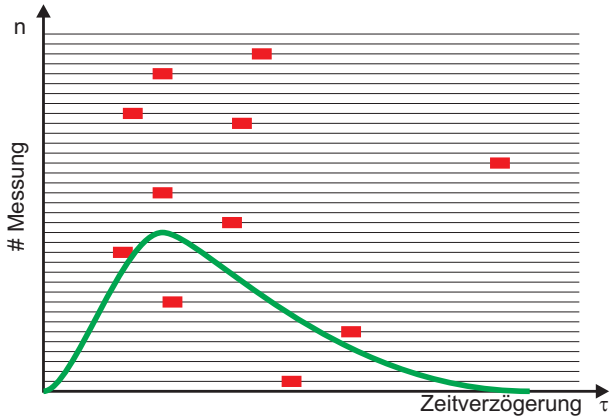



Abbildung 4.252.: Zeitkorreliertes Einzelphotonenzählen. Die vertikale Achse zeigt die einzelnen Messungen. Horizontal ist jeweils ein roter Balken eingezeichnet, wenn bei einer bestimmten Verzögerungszeit ein Photon gemessen wurde. Die grüne Kurve ist die Summenhäufigkeit.

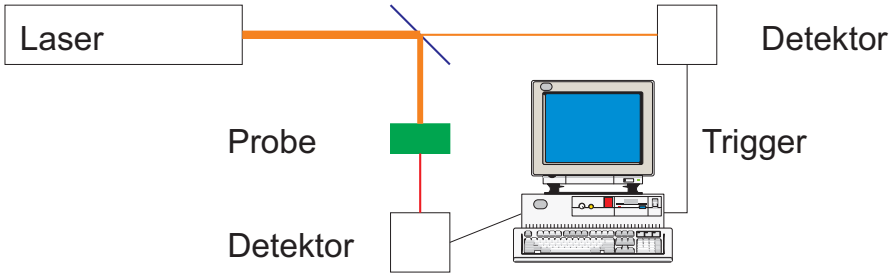



Abbildung 4.253.: Prinzip der Lumineszenzmessung mit einer zeitkorrellierten Messmethode[?].

fenlLeiter sowohl nach links wie auch nach rechts propagiert. Wenn nun mit einem zweiten Laserpuls, verzögert um $\Delta t$ der zweite Schalter kurzgeschlossen wird, ändert sich im Mittel die Ausgangsspannung am rechten Ende, je nachdem ob der durchlaufende Puls, erzeugt vom ersten Laserpuls gerade während des Kurzschlusses beim zweiten Schalter vorbeipropagiert oder nicht. Die Schaltung ist analog zu rein optischen Pump-Probe Experimenten.

## 4.8. Elektrische Messverfahren für kurze Zeiten

Abbildung 4.252 zeigt das Prinzip einer zeitkorrellierten Einzelphotonenmessung. Das *Messgerät*, in diesem Falle eine Messkarte[?] in einem PC startet bei jedem Triggerpuls aus einem Kurzpulslaser eine elektrische Rampenfunktion. Der Triggerpuls ist aus einem Puls abgeleitet, mit dem die Probe beleuchtet wird um die Lumineszenz anzuregen. Wird innerhalb des Messfensters, das heisst, innerhalb der Laufzeit der elektrischen Rampenfunktion, ein Lumineszenzphoton detektiert, dann wird die Rampenspannung zum Zeitpunkt des Eintreffens mit einem Sample/Hold-Glied gespeichert und später digitalisiert. In einem Speicher wird an der zur digitalisierten Spannung, also zur Zeit, gehörigen Stelle der Zählwert um eins erhöht.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

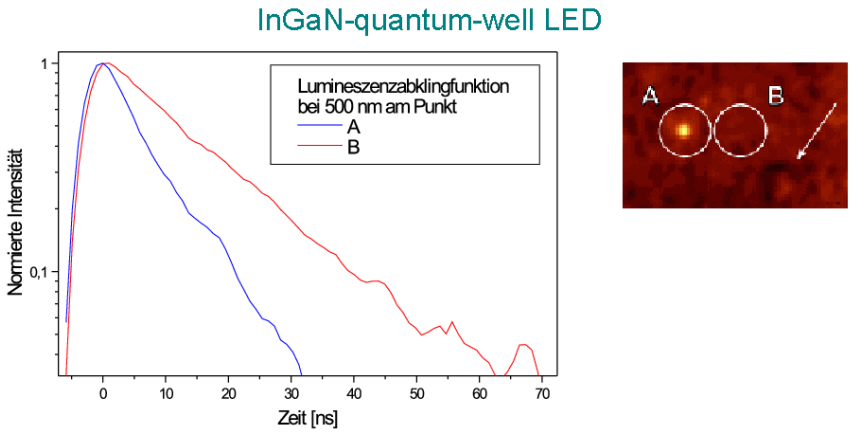



Abbildung 4.254.: Messung einer Lumineszenzabklingkurve mit der zeitkorrellierten Einzelphotonenzählung[?]

Nach einer genügend langen Integrationszeit steht in den Speichern die Zeitfunktion bereit. Abbildung 4.253 zeigt einen entsprechenden Messaufbau. Die Empfindlichkeit dieses Messaufbaus ist[?]

$$S = \frac{\sqrt{R_d \cdot N/T}}{Q} \tag{4.8.1}$$

wobei $R_d$ die Dunkelzählrate ist. $Q$ ist die Quanteneffizienz des Detektors, $N$ die Anzahl Kanäle und $T$ die Messzeit. Für eine Messzeit von $T = 1s$ und $N = 256$ Kanäle erhältm man folgende Empfindlichkeiten:

| Grösse | Photomultiplier | Avalanchediode 1 | Avalanchediode 2 |
|---|---|---|---|
| $Q$ | 0.1 | 0.5 | 0.8 |
| $R_d$ | 1 | 100 | 20 |
| $S$ | 160 | 320 | 90 |

Je nach Typ des Detektors ist die Detektionsgrenze zwischen $< 100$ und etwa 300 Photonen pro Sekunde. Die erreichbare Zeitauflösung hängt von der Geschwindigkeit der Detektoren und ihrem Jitter ab. Photomultiplier haben typischerweise eine Zeitauflösung von $200ps \ldots 1ns$. Mikrokanal-Photomultipliers sind schneller als $50ps$. Avalanche-Dioden haben eine Zeitauflösung zwischen $50ps \ldots 200ps$.
Abbildung 4.254 zeigt eine Messung der Lumineszenzabklingzeit von GaN mit der zeitkorrellierten Einzelphotonenmethode[?]. Als Anregungslicht wurde eine Pulsfolge aus einem TiSaphir-Laser mit etwa $200fs$ Pulsbreite und einer Wiederholfrequenz von $80Mhz$ verwendet. Gezeigt ist die Abklingzeit an zwei verschiedenen Orten.
Bei periodischen Signalen kann auch eine *Überabtastung*, englisch *Oversampling* erfolgen. Eine Überabtastung ist eine bewusste Verletzung des *Abtasttheorems von Nyquist*, wie es im Abschnitt 2.6.2.1 besprochen wurde. Wenn das zu untersuchende Signal eine relativ schmale Bandbreite besitzt, dann kann man durch eine tiefe Wahl der Abtastfrequenz das Eingangssignal zu tieferen Frequenzen hin verschieben und trotzdem mit formgetreu abtasten. Wichtig ist, dass der Sample/Hold-

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

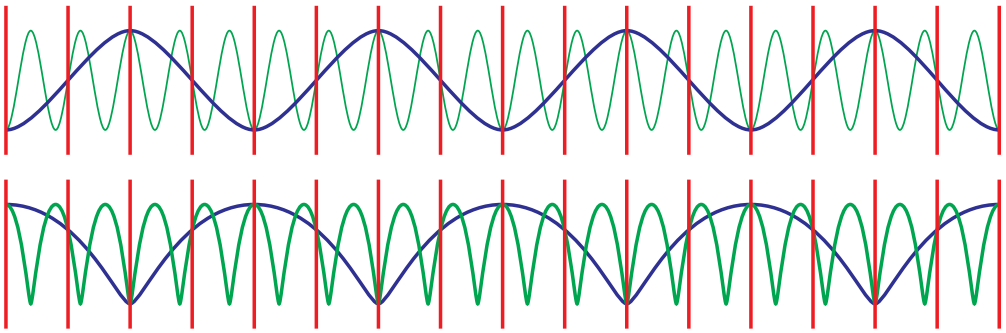


Abbildung 4.255.: Auswirkung des *Oversamplings* auf eine Messung.

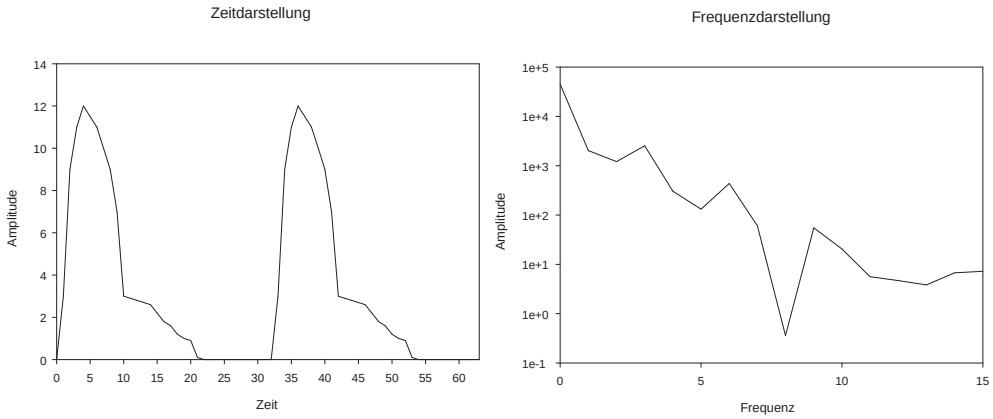



Abbildung 4.256.: Vergleich einer Frequenz- und einer Zeitbereichsmessung.

Verstärker am Eingang einen sehr kleinen *Jitter* haben muss.
Abbildung 4.255 zeigt zwei Beispiele einer Messung mit Überabtastung. Die senkrechten roten Striche zeigen den Zeitpunkt der jeweiligen Abtastung. Das grüne Signal ist das Eingangssignal. Aus der Abtastung resultiert das blaue Signal, das auch bei nicht sinusförmiger Amplitudenform die Signalform richtig wiedergibt.
Frequenz- und Zeitdarstellung sind im Normalfalle durch eine *Fouriertransformation* ineinander überführbar. Abbildung 4.256 links eine hypothetische Pulsform und rechts davon die *Fouriertransformation*. Es gilt, dass wenn Pulse kürzer werden dass dann ihre Bandbreite zunimmt. Das Produkt aus Bandbreite und Pulsbreite ist im Normalfall eine Konstante.

## 4.9. Elektrische Spektralanalyse und Netzwerkanalyse

Bei sehr hohen Frequenzen verwendet man häufig anstelle von Messungen im Zeitbereich Messungen im Frequenzbereich durchgeführt. Bis etwa zu 10 MHz werden *Fourieranalysatoren*, beruhend auf der schnellen *Fouriertransformation* (Siehe auch 2.4.2). Für höhere Frequenzen verwendet man Spektralanalysatoren, bei denen die Oszillatoren durchgestimmt werden. Abbildung 4.257 zeigt den Aufbau eines solchen *Spektralanalysator*s. Ein Sägezahnoszillator steuert einen in der Frequenz abstimmbaren Oszillator. Diese Abstimmung könnte zum Beispiel durch einen Schwingkreis mit Kapazitätsdioden realisiert sein. Dieses *Signal* wird in ei-

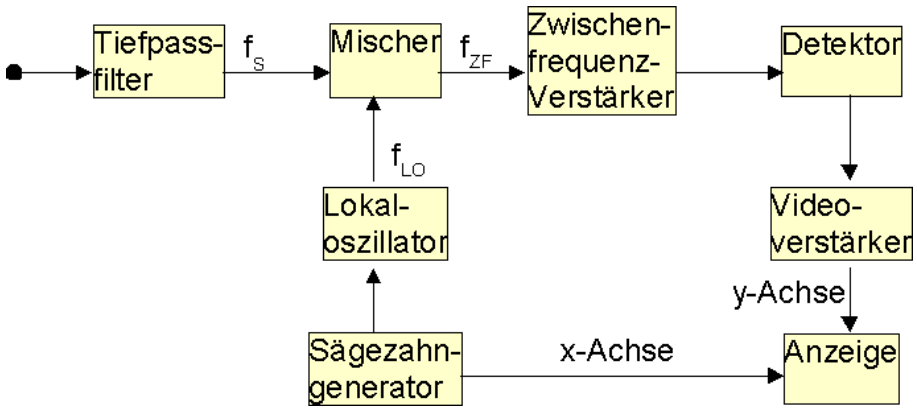



Abbildung 4.257.: Blockschaltbild eines Spektralanalysators.

nem Mischer mit dem tiefpassgefilterten Eingangssignal multipliziert. Aus allen Frequenzen $f_S$ des Eingangssignales wird nur derjenige Bereich, der in das Durchlassband des Zwischenfrequenzverstärkers fällt, weiter verstärkt. Dabei ist die Zwischenfrequenz

$$f_{ZF} = f_S + f_{LO} \tag{4.9.1}$$

Da die Zwischenfrequenz fest ist, wird bei einer Erhöhung der Oszillatorfrequenz die detektierte Signalfrequenz erniedrigt. Wie bei den Lock-In-Verstärkern (siehe auch 4.1.9) legt die Bandbreite des Zwischenfrequenzverstärkers die Bandbreite des Detektionssystems fest. Aus der Filterbandbreite ergibt sich die Messzeit. Die Zeit zur Überstreichung des Frequenzbereiches ist dabei umgekehrt proportional zum Quadrat der Filterbandbreite[**?**]. Die Quadratische Abhängigkeit kommt von Zwei Faktoren: Einerseits müssen bei einer halben Filterbandbreite doppelt so viele Messpunkte im zu überstreichenden Frequenzbereich. Gleichzeitig ist die Einstellzeit auf eine vorbestimmte Präzision doppelt so lange. Beide Effekte zusammen ergeben eine quadratische Abhängigkeit.

Der Aussteuerungsbereich des Spektrumanalysators wird bei niedrigen Pegeln durch das Geräterauschen und bei hohen Pegeln durch die nichtlinearen Verzerrungen im Mischer begrenzt. Wenn sehr hohe Frequenzen zu messen sind, dann verwendet man Oberwellen des lokalen Oszillators. Bei einer Messung mit der $n$-ten Oberwelle hängen die gemessene Frequenzkomponente $f_S$ und die Zwischenfrequenz $f_{ZF}$ wie

$$f_S = n \cdot f_{LO} + f_{ZF} \tag{4.9.2}$$

zusammen. Es ist damit möglich, bis zu 18 GHz Spektren zu messen. Da die Spiegelfrequenzen schlechter unterdrückt werden, ist die Qualität der Messung nicht so gut wie bei Spektralanalysatoren mit Grundfrequenzoszillatoren.

Neben der Darstellung von modulierten Signalen werden Spektralanalysatoren insbesonders zur Messung von Rauschsignalen (Abschnitt 2.8) verwendet. Bei Zufallsrauschen bedeutet eine Vervierfachung der Bandbreite eine Verdoppelung des Rauschpegels. Man bezieht die Rauschbandbreite eines Spektralanalysators auf die

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

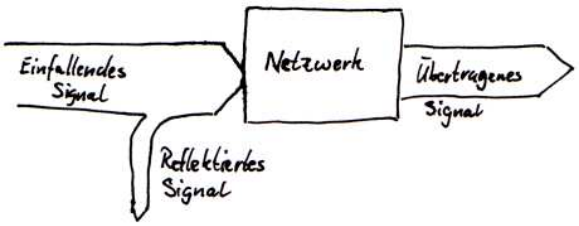



Abbildung 4.258.: Blockschaltbild eines Netzwerkanalysators.

Bandbreite eines Gauss-Filters. Diese ist etwa das 1.2-fache der 3dB-bandbreite[**?**].

$$\delta B = 10\,\lg\left(1.2\frac{B_{ZF}}{B_{ref}}\right) \tag{4.9.3}$$

Hier ist $B_{ZF}$ die Bandbreite des Zwischenfrequenz-Verstärkers. Da in Spektralanalysatoren meistens Spitzenwerte gemessen werden und diese mit einem Faktor passend für eine sinusförmige Schwingung umgerechnet werden, müssen die gemessenen Amplituden bei einer Rauschmessung um 2.5 dB nach oben korrigiert werden.
Bei der Netzwerkanalyse nach Abbildung 4.259 wird ein von einer Quelle stammendes bekanntes *Signal* an den Eingang eines Netzwerkes gelegt. Das übertragene *Signal* und das reflektierte *Signal* werden gemessen.
Beim reflektierten Signal werden

- das Stehwellenverhältnis
- die S-Parameter $\underline{S}_{21}$ und $\underline{S}_{12}$
- der Reflexionskoeffizient
- die Impedanz $\underline{Z}$ und
- die Rückflussdämpfung

gemessen. Das übertragene *Signal* wird durch

- die Verstärkung oder Dämpfung,
- die S-Parameter $\underline{S}_{11}$ und $\underline{S}_{22}$,
- den Übertragungsfaktor,
- die Phasenverschiebung und
- die Gruppenlaufzeit

charakterisiert.
Abbildung 4.259 zeigt den prinzipiellen Aufbau eines *Netzwerkanalysator*s. Die notwendigen *Signale* werden in einem Signalgenerator erzeugt. Mit diesem Signal wird das zu untersuchende Netzwerk gespiesen. Ein kleiner Bruchteil der ursprünglichen Signalleistung wird in die Schaltung zur Signaltrennung eingespiesen. Dieser Signalanteil dient als Referenz. Das Rückreflektierte Signal wird mit einem Richtkoppler vom Eingangssignal getrennt und in die Signaltrennungsschaltung

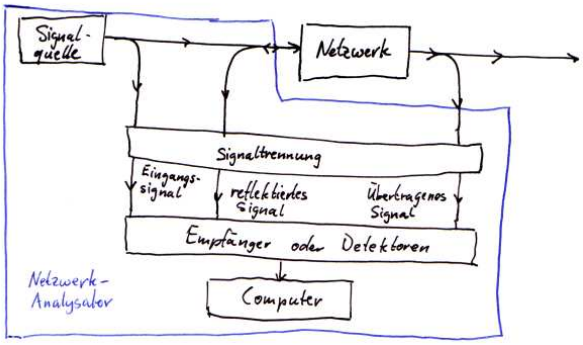



Abbildung 4.259.: Prinzipieller Aufbau eines Spektrumanalysators.

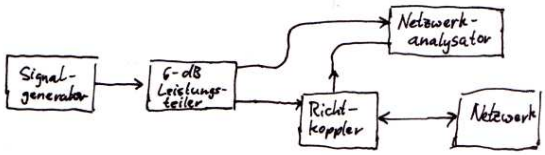



Abbildung 4.260.: Durchführung einer Reflexionsmessung mit einem Netzwerkanalysator.

eingespiesen. Schliesslich wird ein kleiner Bruchteil des übertragenen Signals weiterverarbeitet.

Die Empfänger/Detektorschaltung ist analog zu einem Lock-In-Verstärker aufgebaut und liefert Phase und Amplitude des übertragenen und des reflektierten Signals.

Bei einer Reflexionsmessung wird der Netzwerkanalysator wie in der Abbildung 4.260 gezeigt, verschaltet. Das von einem Signalgenerator kommende Signal wird in einem 6 dB-Teiler in zwei Signale mit je der halben Leistung aufgetrennt. Das eine Signal (oben) dient als Referenz. Das andere wird über einen Richtkoppler[?][?] in den Prüfling eingespiesen. Der Richtkoppler trennt das rückreflektierte Signal ab und speist es ebenfalls in den Netzwerkanalysator.

Die Reflexionsmessung kann verwendet werden, wenn man feststellen möchte, ob ein Kabel unterbrochen ist oder ob es einen Kurzschluss hat. In beiden Fällen ist die Impedanz an der Störstelle nicht gleich dem Wellenwiderstand des Kabel: Reflektierte Signale treten auf. Deren Phase zeigt an, ob es sich um einen Kurzschluss oder eine Unterbrechung handelt. Die Laufzeit hängt von der Entfernung der Störstelle ab.

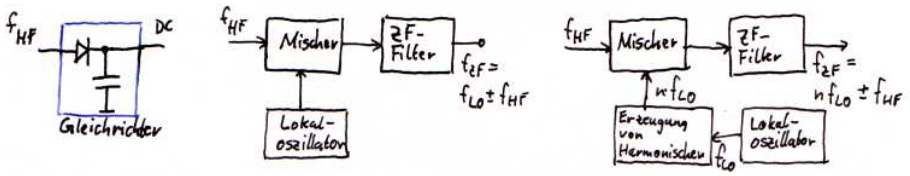



Abbildung 4.261.: Detektionsverfahren. a) ist ein Diodendetektor. b) ist ein Grundwellenmischer und c) ein Oberwellenmischer

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

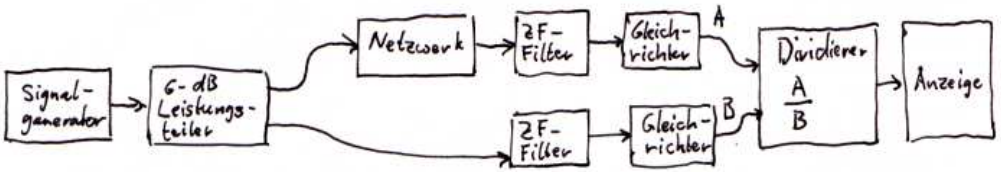



Abbildung 4.262.: Amplitudendetektion.

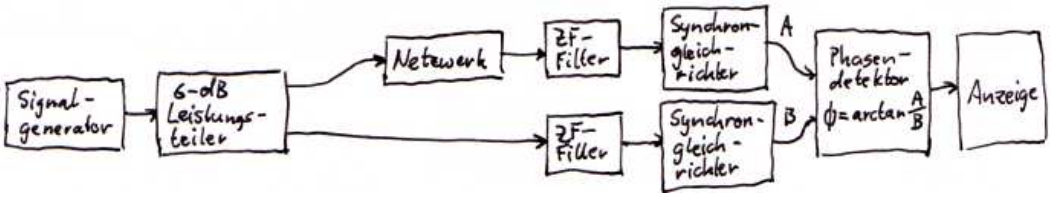



Abbildung 4.263.: Phasenmessung.

Typische Empfängerschaltungen werden in der Abbildung 4.261 gezeigt. Links ist eine einfache Spitzenwertdetektionsschaltung mit einer Diode als Gleichrichter und mit einem Kondensator als Filter angegeben. Damit lässt sich die Einhüllende der detektierten Signale bestimmen. Das mittlere und das rechte Bild zeigen Misch-Verstärker, wie sie auch in Radioempfängerschaltungen eingesetzt werden. Das Eingangssignal wird mit einem um die Zwischenfrequenz versetzten Signal multipliziert und in einem schmalbandigen Filter mit einem festen Durchlassbereich verstärkt und anschliessend demoduliert. Ist die Frequenz des Eingangssignals zu hoch, dann kann der Lokaloszilator einen Hilfsoszillator auf einem Vielfachen seiner Frequenz steuern. Dessen Signal wird mit dem Eingangssignal gemischt. Der verbleibende Teil der Schaltung ist analog zur ursprünglichen Schaltung.
Die Detektion des Zwischenfrequenzsignals kann entweder skalar oder, wie oben angesprochen, über einen Lock-In Verstärker geschehen. Im ersten Falle nennt man das Gerät einen *skalaren Netzwerkanalysator*, im zweiten Falle handelt es sich um einen *vektoriellen Netzwerkanalysator*. Abbildung 4.262 zeigt das Funktionsschema eines skalaren Netzwerkanalysators. Das Ausgangssignal eines *Signalgenerators* wird in einem 6-dB-Teiler auf den referenzzweig (unten) und den Signalzweig (oben) aufgeteilt. Das Signal wird in einer Zwischenfrequenzstufe (ZF-Filter) entweder direkt nach dem Teiler (Referenzzweig) oder nach Durchlaufen des Prüflings verarbeitet. In beiden Zweigen wird das Signal durch einen Gleichrichter demoduliert. Ein Dividierer bildet den Quotienten aus dem Signal am Ausgang des Prüflings und dem Referenzsignal. Das Resultat wird angezeigt, oder in Rechnern weiterverarbeitet.
Zur *Phasenmessung* wird der in der Abbildung 4.263 gezeigte Aufbau verwendet. Anstelle eines Dividierers wird ein *phasenempfindlicher Detektor* verwendet. Die Darstellung von Amplitude und Phase wird vielfach mit einem *Smith-Chart* durchgeführt. Eine ausführliche Diskussion dieser Darstellungsform findet sich auf der Webseite von Spread Spectrum Scene[25].
Bei einer linearen Übertragungskette ist das Verhältnis zwischen der Ausgangsspannung $\underline{U}_{aus}$ und der Eingangsspannung $\underline{U}_{ein}$ der komplexe Übertragungskoeffizient $\underline{v}$. Um eine gültige Übertragungsmessung durchführen zu können, muss der

[25] http://sss-mag.com/smith.html

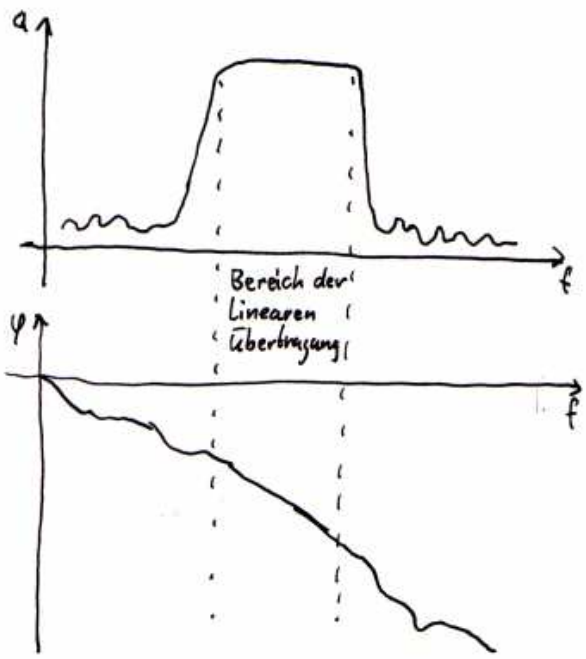



Abbildung 4.264.: Amplituden- und Phasengang bei einer linearen Übertragung.

Netzwerkanalysator mit einer bekannten Teststrecke kalibriert werden.
Bei der Untersuchung einer Teststrecke ist von besonderem Interesse, das Verzerrungsverhalten zu bestimmen. Wie man durch Anwendung der Übertragungstheorie (siehe auch den Abschnitt 2.3 und folgende) berechnen kann, ist eine verzerrungsfreie Übertragung nur dann möglich, wenn, wie in Abbildung 4.264 die Amplitude konstant und die Phase linear ist[**?**]. Der erste Teil dieses Messproblems kann sowohl mit einem skalaren Netzwerkanalysator wie auch mit einem vektoriellen Netzwerkanalysator gelöst werden.
Die Phasenmessung jedoch ist nur mit einem vektoriellen Netzwerkanalysator durchführbar. Um die Wirkung einer linear zunehmenden Phase zu verstehen, betrachten wir die *Gruppenlaufzeit*

$$t_g = -\frac{d\phi}{d\omega} = -\frac{1}{2\pi}\frac{d\phi}{df} \approx -\frac{1}{2\pi}\frac{\Delta\phi}{\Delta f} \tag{4.9.4}$$

Eine linear ansteigende oder abnehmende Phase bedeutet eine konstante Gruppenlaufzeit. Das heisst, dass jede Signalkomponente bei einer Frequenz im Bereich der linearen Phase um die gleiche zeit verzögert wird: Das Signal wird also nicht verzerrt.
Bei einer Messung des Reflexionsverhaltens wird die aus einem Messobjekt in die Quelle zurückreflektierte Spannung $\underline{U}_R$ bestimmt. Sie hängt mit der Eingangsspannung $\underline{U}_{ein}$ über

$$\underline{U}_R = \frac{\underline{Z}_L - \underline{Z}_0}{\underline{Z}_L + \underline{Z}_0}\underline{U}_{ein} \tag{4.9.5}$$

zusammen. $\underline{Z}_0$ ist die (konstante?) Impedanz einer Übertragungsstrecke und $\underline{Z}_L$ ist die Impedanz des Abschlusses der Übertragungsstrecke. Es ist üblich, einen

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

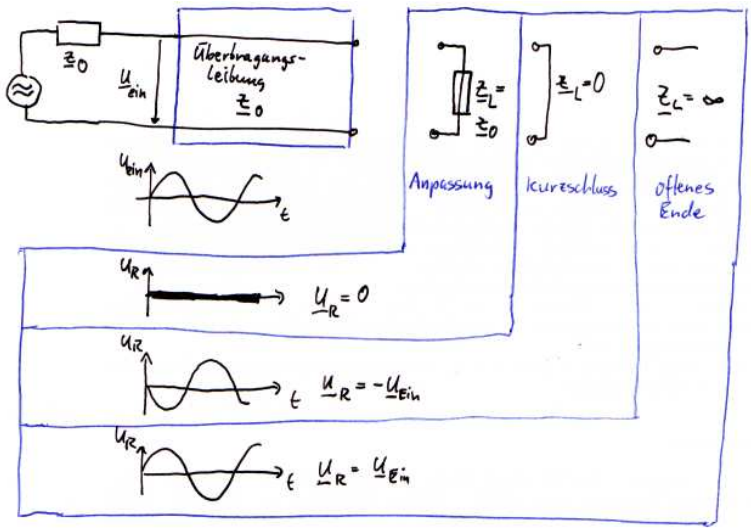



Abbildung 4.265.: Reflektierte *Signale* an einer Leitung abhängig vom Abschlusswiderstand.

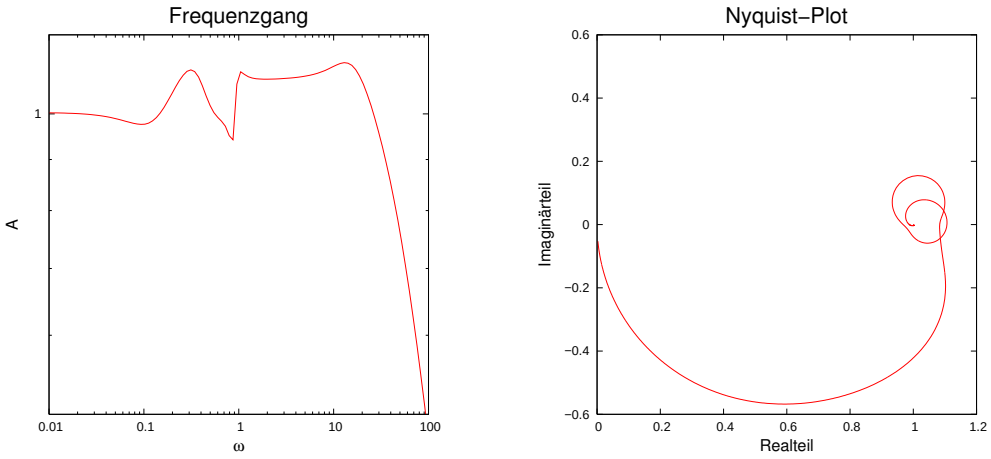



Abbildung 4.266.: Links: Frequenzgang des Abschlusswiderstandes für Abbildung 4.267. Rechts ist die entsprechende Nyquist-Darstellung gezeigt.

komplexen *Reflexionsfaktor*

$$\underline{r} = \frac{\underline{U}_R}{\underline{U}_{ein}} = \frac{\underline{Z}_L - \underline{Z}_0}{\underline{Z}_L + \underline{Z}_0} \tag{4.9.6}$$

zu definieren. Für $\underline{r}$ gilt, dass $0 \leq |\underline{r}| = r \leq 1$ ist. Abbildung 4.265 zeigt die Grösse der reflektierten Signale für drei Fälle: Kurzschluss ($\underline{Z}_L = 0$), Leerlauf ($\underline{Z}_L = \infty$) und Anpassung ($\underline{Z}_L = \underline{Z}_0$).
Die rückreflektierte Welle interferiert mit der eingespeisten Welle. Das dabei entstehende Muster von Wellenbergen und Tälern hängt vom Reflexionsfaktor $r$ ab. In der Antennentechnik ist es üblich das Stehwellenverhältnis

$$SWR = \frac{1+r}{1-r} = \frac{U_{max}}{U_{min}} \tag{4.9.7}$$

zu verwenden. Es gilt, dass $1 \leq SWR \leq \infty$ ist.

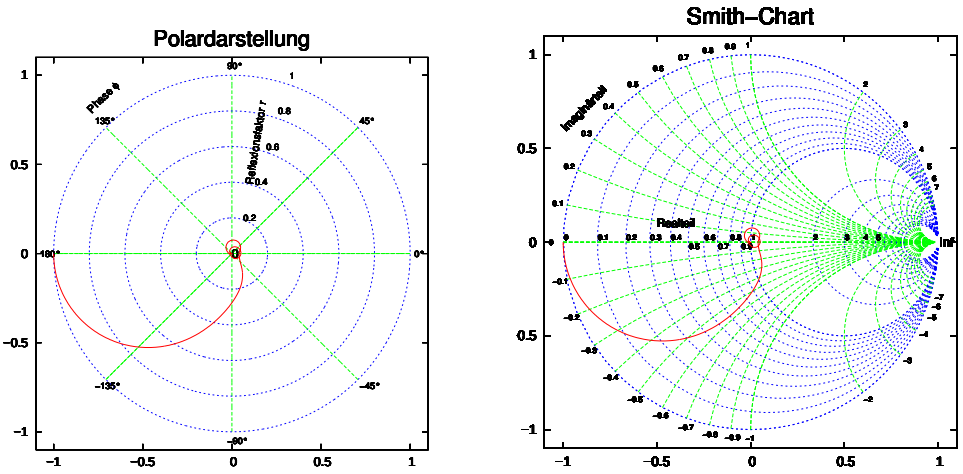



Abbildung 4.267.: Darstellung der reflektierten *Signale* in der *Polardarstellung* und im *Smith-Diagramm*. Die Polardarstellung zeigtden Reflexionsfaktor $r$ und die Phase $\phi$, während die Smith-Darstellung Real- und Imaginärteil zeigt.

Die Transformation des Abschlusswiderstandes (4.9.6) ist äquivalent zu einer Abbildung der komplexen Ebene auf sich. Wenn alle Impedanzen auf den Wert $\underline{Z}_0$ bezogen werden, bekommt man für $r$

$$\underline{r} = \frac{\frac{\underline{Z}_L}{\underline{Z}_0} - 1}{\frac{\underline{Z}_L}{\underline{Z}_0} + 1} = \frac{\underline{z} - 1}{\underline{z} + 1} \tag{4.9.8}$$

Dabei wurde $z = \underline{Z}_L/\underline{Z}$ gesetzt. Die Transformation von $\underline{z}$ nach $\underline{r}$ ist eine konforme Abbildung auf sich selber. Mit dieser Abbildung nach Gleichung (4.9.8) kann im Polardiagramm (siehe Abbildung 4.267, links) direkt der Betrag des Reflexionsfaktors und seine Phase abgelesen werden.
Andererseits kann man mit der gleichen Abbildung Gleichung (4.9.8) auch die Koordinatenachsen und das Gitternetz der $\underline{Z}$-Ebene abbilden. Man tranformiert also die Geraden mit konstantem Real- (Widerstands-) und Imaginärteil. Konforme Abbildungen bilden Geraden auf Kreise ab. Die entstehenden Kurven (siehe Abbildung 4.267, rechts) heissen dann *Smith-Chart*.
Die Transformation der Linien mit konstantem Real- oder Imaginärteil ist im Anhang G.4 gezeigt. Die entstehenden Ortskurven sind Kreise, ihre Anordnung heisst *Smith-Chart*.
Wir wollen nun einen Abschlusswiderstand, der sich wie in der Abbildung 4.266 verhält, in einen *Smith-Chart* eintragen. Dazu werden der Real- und der Imaginärteil auf den Gitternetzlinien im Bild 4.267 abgetragen. Die entstehende rote Ortskurve zeigt das reflektierte Signal (abzulesen an den äusseren Achsenbeschriftungen) als Funktion der Abschlussimpedanz.
Polardiagramm und Smith-Chart können kombiniert werden. Dies ist in Abbildung 4.268 gezeigt.
Bei der Berechnung von Schaltungen mit Smith-Charts hilft eine einfache graphische Regel. Nehmen wir an, $\underline{Z}$ sei die Impedanz und $\underline{Y}/\underline{Y}_0 = \underline{Z}_0/\underline{Z}$ die dazugehörige Admittanz (Dualismus Widerstand - Leitwert). $\underline{Z}_0$ ist die Referenzimpedanz,

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

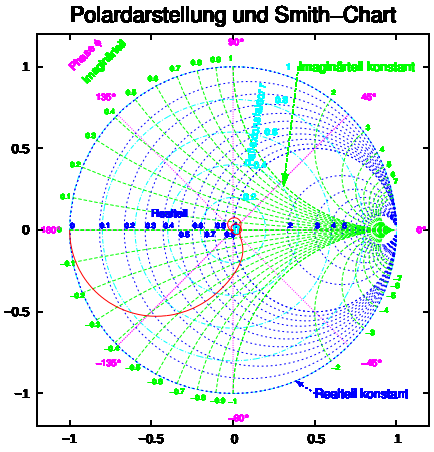



Abbildung 4.268.: Darstellung der reflektierten *Signale* kombiniert in der *Polardarstellung* und im *Smith-Diagramm*.

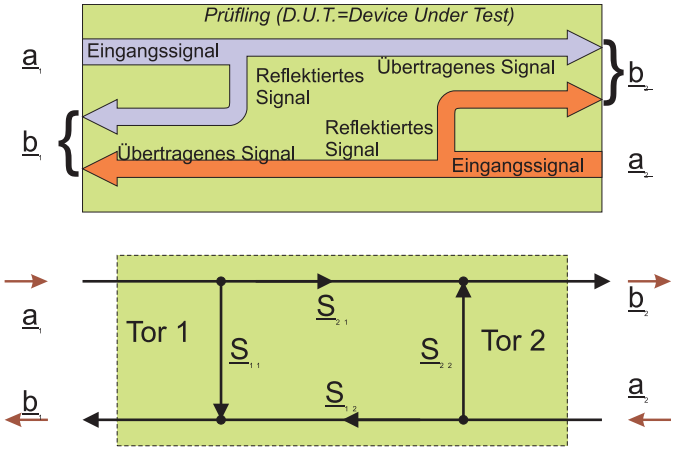



Abbildung 4.269.: S-Parameter. Oben ist die Definition gezeigt, unten das dazugehörige Flussdiagramm

$\underline{Y}_0 = 1/\underline{Z}_0$ die Referenzadmittanz. Dann liegt der zu $\underline{Y}/\underline{Y}_0$ gehörige Punkt im Smith-Diagramm bei

$$\underline{r}_Y = \frac{\frac{\underline{Y}}{\underline{Y}_0} - 1}{\frac{\underline{Y}}{\underline{Y}_0} + 1} = \frac{\frac{\underline{Z}_0}{\underline{Z}} - 1}{\frac{\underline{Z}_0}{\underline{Z}} + 1} = \frac{1 - \frac{\underline{Z}}{\underline{Z}_0}}{1 + \frac{\underline{Z}}{\underline{Z}_0}} = -\frac{\frac{\underline{Z}}{\underline{Z}_0} - 1}{\frac{\underline{Z}}{\underline{Z}_0} + 1} = -\underline{r}_Z$$

Dies bedeutet, dass die zu $\underline{Z}/\underline{Z}_0$ gehörige Admittanz $\underline{Y}/\underline{Y}_0$ einfach am Punkt $\underline{r} = 0$ gespiegelt ist. Damit könnenin einem Smith-Chart sehr leicht Reihen- und Parallelschaltungen berechnet werden. Bei Reihenschaltungen von Impedanzen werden die Real- und Imaginärteile addiert. Bei Parallelschaltungen von Impedanzen werden die zugehörigen Admittanzen addiert. Damit kann sehr leicht aus Impedanzkurven im Smith-Chart der resultierende Frequenzverlauf konstruiert werden.

Zur Untersuchung von komplexeren Objekten wie *Transistoren* oder Verstärkern führt man sie auf Vierpole (siehe auch Abschnitt 2.5) zurück. Im Hoch- und Höchstfrequenzbereich werden bevorzugt die *S-Parameter* verwendet. Die Definition der *S-Parameter* ist in der Abbildung 4.269 gezeigt. In Gleichungen gefasst ergibt sich

$$\begin{aligned} \underline{b}_1 &= \underline{S}_{11}\underline{a}_1 + \underline{S}_{12}\underline{a}_2 \\ \underline{b}_2 &= \underline{S}_{21}\underline{a}_1 + \underline{S}_{22}\underline{a}_2 \end{aligned} \tag{4.9.9}$$

Die Grössen $\underline{a}_i$ und $\underline{b}_i$ können mit einem vektoriellen Netzwerkanalysator direkt gemessen werden. Durch Messung in Vorwärtsrichtung lassen sich mit $\underline{a}_2 = 0$

$$\begin{aligned} \underline{S}_{11} &= \frac{\underline{b}_1}{\underline{a}_1} \\ \underline{S}_{21} &= \frac{\underline{b}_2}{\underline{a}_1} \end{aligned} \tag{4.9.10}$$

bestimmen. In Rückwärtsrichtung erhält man mit $\underline{a}_1 = 0$

$$\begin{aligned} \underline{S}_{12} &= \frac{\underline{b}_1}{\underline{a}_2} \\ \underline{S}_{22} &= \frac{\underline{b}_2}{\underline{a}_2} \end{aligned} \tag{4.9.11}$$

Damit sind die Kleinsignalparameter eines Prüflings bestimmt.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

# 4.10. Messung mit Elektronen

All Scanning Probe Microscopes (SXM) are based on similar principles. The aim of this article is to work out the common aspects, to elucidate the differences, and to point to possible applications in the field of biology. The Scanning Tunneling Microscope (*STM*), invented by Binnig and Rohrer[**?**] serves as a model system. A summary of the theory of the *STM* points out the different operating modes and techniques, deals with the problem of imaging and gives resolution criteria.
A detailed introduction to the mechanical end electronic design of the *STM* is presented. Design rules are worked out to help the builders of an *STM* and to allow the users to judge their instruments. An important part of any *STM* experiment is the data acquisition and the image processing. Critical points in the data acquisition systems and common image processing techniques are worked out. All the technical issues of the *STM* are equally valid for other SXM techniques.
The section on the *STM* concludes with the description of a few experiments. The application of the *STM* to the imaging of biological and organic matter is treated in depth by other papers in this book.
The Scanning Force Microscope (SFM) is the most successful offspring of the *STM*. The design principles worked out for the *STM* are equally valid for the SFM. The additional critical points of an SFM are treated. Special emphasis is given to the description of the various interaction forces and the force sensing techniques, including the Scanning Force and Friction Microscope (SFFM). The section on the SFM is closed by presenting a few representative experiments.
The interested reader might also want to consult review articles on scanning probe microscopy[**?**, **?**, **?**, **?**, **?**, **?**, **?**].

## 4.10.1. Tunneleffekt, Bänderstruktur

Some basic knowledge of the physics of *STM* is necessary to judge the relevance of experiments. The tunneling junction of a *STM* is a quantum mechanical system; hence a basic knowledge of quantum mechanics is required to understand the physics. An overview of methods and approximations used to model the tunneling process in *STM* can be found in Baratoff[**?**].

### 4.10.1.1. The Tunneling Current - A Simple Theory

To get a first intuitive view about electron tunneling between the tip and the sample of an *STM* we will consider the textbook case of quantum mechanical electron tunneling between to infinite, parallel, plane metal electrodes. We only treat the simplest case with no time dependent potentials. Excellent articles on tunneling between infinite plane parallel metal plates can be found in the literature[**?**, **?**, **?**, **?**, **?**, **?**, **?**]. The axis perpendicular to the plane parallel electrodes is the $z$-axis, with its zero on the left side of the tunnel gap (see figure 4.270).
The electron motion is governed by the Schrödinger equation

$$i\frac{\partial}{\partial t}\Psi\left(\boldsymbol{z},t\right)=\mathrm{H}\Psi\left(\boldsymbol{z},t\right) \tag{4.10.1}$$

where $H$ is the Hamiltonian of the system. The Hamiltonian for a simple tunnel junction consists of a kinetic energy part $-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2}\Psi\left(\boldsymbol{z},t\right)$ and a potential energy

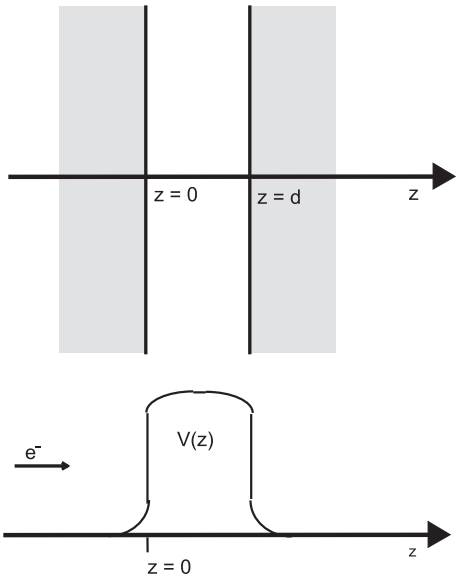



Abbildung 4.270.: Coordinate system for calculating the transmissivity of a one dimensional tunneling barrier. The electron plane wave is incident on the barrier from the left (negative $z$-axis). The two electrodes are separated by the distance $d$.

part $V(\boldsymbol{z})\,\Psi(\boldsymbol{z},t)$. The potential energy is equal to zero everywhere except in the barrier between the electrodes, from 0 to $d$ where $d$ is the thickness of the barrier. The wave function $\Psi(\boldsymbol{z},t)$ of the electrons is a solution of the equation

$$i\frac{\partial}{\partial t}\Psi(\boldsymbol{z},t) = -\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2}\Psi(\boldsymbol{z},t) + V(\boldsymbol{z})\,\Psi(\boldsymbol{z},t) = \tag{4.10.2}$$

$$\left(-\frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2} + V(\boldsymbol{z})\right)\Psi(\boldsymbol{z},t)$$

The probability to find a particle described by the wave function $\Psi(\boldsymbol{z},t)$ at the position $\boldsymbol{z}$ at the time $t$ is

$$P(\boldsymbol{z},t) = \Psi(\boldsymbol{z},t)\,\Psi^*(\boldsymbol{z},t) = |\Psi(\boldsymbol{z},t)|^2 \tag{4.10.3}$$

To simplify the calculation we will consider the one-dimensional case of a tunneling barrier with a potential independent of time. The wave function $\Psi(z,t)$ is written as the product $\Psi_z(z)\,\Psi_t(t)$. Equation (4.10.2) can then be separated and written as

$$0 = \frac{\hbar^2}{2m}\frac{\partial^2}{\partial z^2}\Psi_z(z) + (E-V)\,\Psi_z(z) \tag{4.10.4}$$

We assume that the electrons are incident on the barrier from the left. There are three solutions to the Schrödinger equation, at the left of the barrier, in the barrier and at the right. Our ansatz for this problem is

$$\Psi_z(z) = \begin{cases} Ae^{ipz/\hbar} + Be^{-ipz/\hbar}, & z < 0 \\ Ce^{-kz} + De^{kz}, & 0 \le z \le d \\ AS(E)e^{ip(z-d)/\hbar}, & z > d \end{cases} \tag{4.10.5}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

where $p = \sqrt{2mE}, \hbar k = \sqrt{2m(V-E)}$ (See for instance Baym[?] for a detailed treatment of the problem). At the boundaries of the three regions, these functions and their first derivative must be continuous. The function $S(E)$ is called the tunneling matrix element. It is a measure for the probability to tunnel from left to right for a particle being present at the left side of the junction.
Satisfying the boundary conditions in equation (4.10.5) leads to four simultaneous equations for the five parameters $A$, $B$, $C$, $D$, $S(E)$. We can chose an arbitrary value for the amplitude of the incoming electron wave, hence we set $A = 1$. The tunneling matrix element is for $E < V$:

$$S(E) = \frac{2i\hbar kp}{2i\hbar kp\cosh(kd) + (p^2 - \hbar^2k^2)\sinh(kd)} \tag{4.10.6}$$

The tunneling barrier has both a transmissivity and a reflectivity. In a *measurement* of the tunneling current, we can only detect the transmissivity $R$, which is given by

$$T(E) = |S(E)|^2 = \left[1 + \frac{\sinh^2(kd)}{4(E/V)(1-E/V)}\right]^{-1} \tag{4.10.7}$$

This equation can be simplified for electrons with a de Broglie-wave length much smaller than the barrier width $d$, or $kd \gg 1$. Equation (4.10.7) becomes:

$$\begin{aligned} T(E) &\approx 16\frac{E}{V}\left(1-\frac{E}{V}\right)\exp\left(-2kd\right) \\ &= 16\frac{E}{V}\left(1-\frac{E}{V}\right)\exp(-\frac{2}{\hbar}\sqrt{2m(V-E)}d) \end{aligned} \tag{4.10.8}$$

To get a feeling for the magnitude of the transmission coefficient $T(E)$ we use values for $V$ and $E$ typical for a metal. The zero point of the energy scale is at the bottom of the conduction band for a metal, which is typically 12 eV below the Fermi energy. All electron states between the bottom of the conduction band and the Fermi energy are filled, at zero temperature. The barrier height is, for a clean metal surface, about 4 eV above the Fermi energy, hence $V = 16$ eV and $E = 12$ eV. We further assume that the tunneling barrier width is $d = 1$ nm. Using these values we will get $T(E) \approx 10^{-9}$ for electrons at the Fermi energy. Figure 4.271 shows the transmission of the tunnel barrier as a function of the electron energy $E$ for the above values of $V$ and $d$.

## 4.10.2. Tunneltheorie von Simmons

We can extend this simple picture by including the Fermi distribution to model the electron energy density. This allows us to calculate a tunneling current density for infinite, plane parallel plates. We will assume that the temperature is 0 K. The calculation does not provide a description of all phenomena occurring in tunneling, but will elucidate some basic aspects. For the tunneling process only the velocities perpendicular to the sample surface matter. Assuming a tunneling barrier with a position dependent barrier height $V(z)$ we obtain in the WKB approximation

$$T(E) = \exp\left\{-\frac{2}{\hbar}\int_{s_1}^{s_2}\left[2m\left(V(z)-E_z\right)\right]^{1/2}dz\right\} \tag{4.10.9}$$

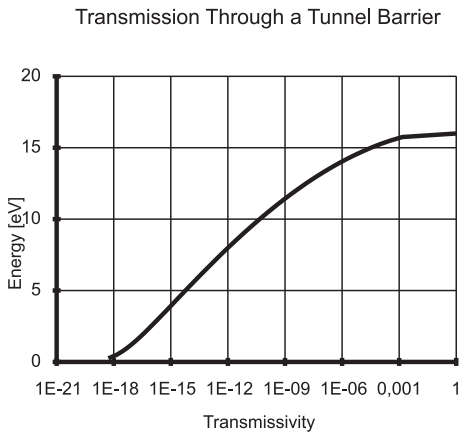



Abbildung 4.271.: The transmission coefficient as a function of the electron energy. The zero energy corresponds to an electron at the bottom of the conduction band. The Fermi energy for this calculation is set to $E = 12$ eV and the work function is 16 eV.

Using the formalism of Simmons[**?**] we can calculate the number of electrons tunneling from left to right

$$N_{L\to R} = \frac{1}{m_e}\int_0^{E_M} n(v_z)T(E_z)dE_z \tag{4.10.10}$$

where the $z$-coordinate is perpendicular to the tunneling junction, $E_M$ is some maximum energy of the electrons, $n(v_z)dv_z$ the number of electrons per unit volume with velocities between $v_z$ and $v_z+dv_z$, and $m_e$ the electron mass. The transmission coefficient $T(E)$ is given by equation (4.10.9).
Next we assume that the electrons in the solids are distributed according to the Fermi-statistics $f(E)$.

$$n(v)dv_xdv_ydv_z = \frac{m_e^3}{4\pi^3\hbar^3}f(E)dv_xdv_ydv_z \tag{4.10.11}$$

We are only interested in the number of electrons in $z$-direction $n(v_z)dv_z$. We get this number by integrating over $v_x$ and $v_y$.

$$n(v_z) = \frac{m_e^3}{4\pi^3\hbar^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(E)dv_xdv_y \tag{4.10.12}$$

To facilitate the integration we change the integration from $dv_xdv_y$ to polar coordinates $v_rdv_rd\phi$ and then to energy using the relation $E_r = m_ev_r^2/2$.

$$n(v_z) = \frac{m_e^2}{2\pi^2\hbar^3}\int_0^{\infty} f(E)dE_r \tag{4.10.13}$$

Combining equations (4.10.10) and (4.10.13) we get for the number of electrons tunneling from left to right:

$$N_{L\to R} = \frac{m_e}{2\pi^2\hbar^3}\int_0^{E_M} T(E_z)dE_z\int_0^{\infty} f(E)dE_r \tag{4.10.14}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

The number of electrons tunneling from right to left, $N_{R\to L}$ is given by an analogous formula, only the energy $E$ in the Fermi distribution is replaced by $E+eV_t$, where $V_t$ is the bias voltage for the tunneling junction.

The current density for the tunneling current is then given by

$$\begin{aligned} J &= Ne \\ &= (N_{L\to R} - N_{R\to L})e \\ &= \frac{m_e e}{2\pi^2\hbar^3}\int_0^{E_M} T(E_z)dE_z \times \\ &\qquad \int_0^{\infty} f(E) - f(E+eV_t)dE_r \end{aligned} \tag{4.10.15}$$

The average potential $V(z)$ in the barrier is expressed as a sum of the Fermi-energy $E_F$ and a potential $\Phi(z)$. For a single electrode $\Phi(\infty)$ is commonly referred to as the work function of the metal. Simmons[**?**] obtains analytical expressions for the tunneling current at a temperature of 0 K using an averaged barrier height $\bar{\Phi} = 1/(s_2 - s_1)\int_{s_1}^{s_2}\Phi(z)dz$.

$$\begin{aligned} J &= \frac{e}{4\pi^2\hbar(\beta\Delta s)^2} \\ &\qquad \left(\bar{\Phi}\exp\left(-\frac{2\beta\Delta s}{\hbar}\left[2m_e\bar{\Phi}\right]^{\frac{1}{2}}\right) - \right. \\ &\qquad \left.\left(\bar{\Phi}+eV\right)\exp\left(-\frac{2\beta\Delta s}{\hbar}\left[2m_e\bar{\Phi}+eV\right]^{\frac{1}{2}}\right)\right) \end{aligned} \tag{4.10.16}$$

where $\Delta s = s_2 - s_1$ and $\beta$ a factor of order 1 to correct the substitution of the integral over the barrier by the average barrier height $\bar{\Phi}$. Equation (4.10.16) can be used to evaluate the current through a rectangular barrier (at zero bias voltage) of width $s$ and height $\Phi_0$

$$J = \frac{(2m_e\Phi_0)^{\frac{1}{2}}}{4\pi^2 s}\left(\frac{e}{\hbar}\right)^2 V \exp\left(-\frac{2s}{\hbar}(2m_e\Phi_0)^{\frac{1}{2}}\right) \quad \text{for} \quad V \simeq 0 \tag{4.10.17}$$

$$J = \left(\frac{e}{4\pi^2\hbar s^2}\right) \left\{\left(\Phi_0 - \frac{eV}{2}\right)\exp\left[-\frac{2s}{\hbar}\left(2m_e\left(\Phi_0 - \frac{eV}{2}\right)\right)^{\frac{1}{2}}\right] - \left(\Phi_0 + \frac{eV}{2}\right)\exp\left[-\frac{2s}{\hbar}\left(2m_e\left(\Phi_0 + \frac{eV}{2}\right)\right)^{\frac{1}{2}}\right]\right\} \quad \text{for} \quad V < \frac{\Phi_0}{e} \tag{4.10.18}$$

$$J = \left(\frac{e^3\left(F/\beta\right)^2}{8\pi^2\hbar\Phi_0}\right) \left\{\exp\left[-\frac{2\beta}{\hbar eF}m_e^{\frac{1}{2}}\Phi_0^{\frac{3}{2}}\right] - \left(1 + \frac{2eV}{\Phi_0}\right)\exp\left[-\frac{2\beta}{\hbar eF}m_e^{\frac{1}{2}}\left(1 + \frac{2eV}{\Phi_0}\right)^{\frac{3}{2}}\right]\right\} \quad \text{for} \quad V > \frac{\Phi_0}{e} \tag{4.10.19}$$

where for $V > \Phi_0/e$ the field strength is $F = V/s$ and the correction factor $\beta = 23/24$[**?**]. The tunneling current through the rectangular barrier around zero bias voltage follows Ohm's law, whereas the dependence on the barrier width is exponential. The current depends on the square of the voltage in the very high voltage range. At intermediate voltages, the tunneling resistance is highly nonlinear.

The theory outlined above does not account for the effect of the image potential. An electron in the barrier will create image charges in the two metal electrodes. These image charges will change the properties of the tunneling barrier. They will round off the barrier shape and lower its average value. The rounding of the tunneling barrier will increase the non-linearity of the tunneling resistance. Simmons[**?**] includes the image potential effects into his theory by calculating an effective tunneling distance which turns out to be smaller than the distance between the two electrodes defined the surfaces of constant charge density.

The use of the WKB approximation further obscures the existence of resonances in the tunnel barrier. Gundlach[**?**] calculates a better approximation than WKB and obtains an oscillatory behavior of the tunneling current. These resonances have been observed by the *STM*[**?**, **?**]. Gundlach[**?**] finds, that the energy and the periodicity of the resonances depends critically on the shape of the tunneling barrier.

The simple model outlined in this section does describe quite accurately the tunneling current between two metal surfaces. It can account for resonances due to

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

image states and due to field states. It does not, however, provide any information on the spatial resolution of an *STM*. An estimate of the resolution based on the assumption of nearly plane parallel electrodes of the tunnel junction will be given in section 4.10.5. The theory of Simmons[?, ?] also does not account for energy dependent density of states, which is characteristic for semiconductors, superconductors and semimetals. The transfer Hamiltonian method outlined in the next chapter will provide a more accurate approach.

## 4.10.3. The Transfer Hamiltonian Method

The transfer Hamiltonian method takes into account the detailed electronic states of the electrodes. It also allows the calculation of effects resulting from the tip geometry. Tersoff and Hamann[?] and Baratoff[?] first applied this formalism to *STM* related problems. The transfer Hamiltonian formalism originally was used by Bardeen[?] to explain the first tunneling spectra obtained by Giaever[?]. Introductions to the formalism can be found in Kirtley[?] and Wolf[?]. The short outline given here follows these two books.
The transfer Hamiltonian formalism is a perturbation theory. The electron waves in the two metal electrodes are considered to be independent. The coupling of the electron waves in the gap is treated as a perturbation, leading to a total Hamiltonian

$$H = H_L + H_R + H_T \tag{4.10.20}$$

where $H_L$ and $H_R$ are the unperturbed Hamiltonians of the two electrodes. The initial wave functions in WKB approximation are (outside and inside the barrier)

$$\begin{aligned} \psi_i^{out} &\propto k_z^{-\frac{1}{2}} \exp\Big[i(k_x x + k_y y)\Big] \sin\left(k_z z + \gamma\right) \\ \psi_i^{in} &\propto |k_z|^{-\frac{1}{2}} \exp\Big[i(k_x x + k_y y)\Big] \exp\Big[-\int_0^d |k_z| dz\Big] \end{aligned} \tag{4.10.21}$$

where $k_z = (1/\hbar)\Big[2m(V(z) - E_z)\Big]^{\frac{1}{2}}$. $d$ is the width of the barrier and $V(z)$ the potential in the barrier. To get the elastic tunneling current, we write a linear combination of an initial state and of a sum of final states:

$$\psi(t) = a(t)\psi_i \exp\left(-i\omega_0 t\right) + \sum_f b_f(t)\psi_f \exp\left(-i\omega_f t\right) \tag{4.10.22}$$

.
$\psi(t)$ is inserted into the time dependent Schrödinger equation $H\psi = i\hbar(\partial/\partial t)\psi$ and solved to first order in $b_f(t)$. One obtains for the transition rate per unit time:

$$\omega_{if} = \frac{2\pi}{\hbar}\Big|M_{if}\Big|^2 \delta(\omega_0 - \omega_f) \tag{4.10.23}$$

The transfer matrix element $M_{if}$ is given by

$$M_{if} = -\frac{\hbar^2}{2m}\int\int\Big[\psi_i^* \frac{\partial}{\partial z}\psi_f - \psi_f \frac{\partial}{\partial z}\psi_i^*\Big] dx dy \;|_{z=constant} \tag{4.10.24}$$

The integrals in equation (4.10.24) have to be evaluated on a plane anywhere inside the barrier. The total tunneling current is the sum over all initial and final states

$$J = \frac{4\pi e}{\hbar} \sum_{k_i} \sum_{k_f} \left| M_{if} \right|^2 \Big[ f(E_i) - f(E_f - eV) \Big] N_i(E_i) N_f(E_f + eV) \delta(E_i - E_f) \quad (4.10.25)$$

where $f(E)$ is the Fermi distribution. Equation (4.10.25) contains the densities of states $N(E)$, unlike the equations derived in the previous chapter. Since we take the unperturbed electron wave functions for the left and the right side, we are calculating the tunneling current in a set of basis functions which are not orthonormal.
Tersoff and Hamann[**?**] use surface electron wave functions with an exponential decay into the vacuum (z-direction). They expand these wave functions in the form

$$\psi_i = \Omega_s^{-\frac{1}{2}} \sum_G a_G \exp \Big[ -(k^2 + |\boldsymbol{k}_{\|} + \boldsymbol{G}|^2)^{\frac{1}{2}} z \Big] \exp \Big[ i(\boldsymbol{k}_{\|} + \boldsymbol{G})\ \boldsymbol{z} \Big] \quad (4.10.26)$$

where $\boldsymbol{G}$ is the reciprocal-lattice vector at the surface, $\boldsymbol{k}_{\|}$ the surface Bloch wave vector. The factor $\Omega^{-1/2}$ ensures a proper normalization of the wave function. The inverse decay length $k$ is given by $k = (2m\Phi)^{\frac{1}{2}}\hbar^{-1}$ where $\Phi$ is the work function, as defined in the previous chapter. Tersoff and Hamann[**?**] find, that only the first few $a_G$ are of order unity.
They further show, that for many situations the tip can be approximated as an electron s-wave. The wave function modeling the tip is

$$\psi_f = \Omega_t^{-1} c_t k r \exp\left[kR\right] \Big[ k|\boldsymbol{r} - \boldsymbol{r_0}| \Big]^{-1} \exp\left[-k|\boldsymbol{r} - \boldsymbol{r_0}|\right] \quad (4.10.27)$$

where $\Omega_t$ is the tip volume $\boldsymbol{r}_0$ the center of curvature of the tip and $R$ the radius of curvature of the tip. The factor $c_t$ depends on the exact tip geometry, its surface conditions and the tip electronic structure and is of order 1. This tip model is excellent for metal tips. The calculation of the tunneling current to or from semiconductor tips with non-vanishing higher order electron orbitals in the conduction band might have to include the dependence on angular and spin momentum.
The tip wave function is expanded in terms of the sample wave functions. The expanded wave functions are used to evaluate the transfer matrix element. Substituting this matrix element in the equation for the tunneling current yields:

$$J = 32\pi^3 \hbar^{-1} e^2 V \Phi^2 D_t(E_F) R^2 k^{-4} \exp\left[2kR\right] \sum_{\nu} |\psi_\nu(\boldsymbol{r_0})|^2 \delta(E_\nu - E_F), \quad (4.10.28)$$

where $D_t$ is the density of states per unit volume of the sample at the center of curvature $\boldsymbol{r}_0$ of the tip. This result was derived for small bias voltages. Again it is found that the tunneling junction is ohmic at low voltages and that the tunneling current depends exponentially on the distance. The approximations used by Tersoff and Hamann[**?**] make the model perform better for small tips (small $R$) than for large ones.
The local density of states can be accommodated by this model. Using this model Tersoff[**?**] calculated the topography of *graphite* as imaged by *STM*.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

The model of Tersoff and Hamann[**?**] was extended by Lang[**?**, **?**, **?**, **?**] to model the tunneling from a chemisorbed atom on a uniform, featureless metal, called *jellium*, to another chemisorbed atom also on a *jellium*. Lang derives a formalism analogous to the transfer Hamiltonian formalism which not only yields the total tunneling current but also a spatial distribution of the tunneling current. The spatial distribution of the tunneling current is expected to be important in imaging materials with non-vanishing electron wave functions of p- or higher order. Lang[**?**] calculates the tunneling current for a sodium, a calcium, and a *sulfur adatom* on the *jellium surface*. The electron levels of *sodium adatom* are predominantly s-like. The diameter of the current distribution of the sodium adatom is narrower than the diameter of the current distribution of the calcium atom with its filled 4s shell. The p-resonance of the calcium lies above the Fermi level. The contribution of p-wave functions (or maybe in other atoms even higher order wave functions) tends to spread out the tunneling current, because these wave functions have a node on the *adatom*. Sulfur on the other hand has its p-resonance below the Fermi level. Its current distribution resembles closer that of sodium than that of calcium. Lang[**?**] states, that most valence-p states away from the Fermi level are not visible in a *STM* experiment.

## 4.10.4. Rastertunnelmikroskopie (STM)

In this section we will focus on the critical points in the design of an *STM* system. We will treat both the mechanics and the electronics of an *STM*. Most findings of this section are also applicable to other Scanning Probe Microscopes (SXM).

### 4.10.4.1. Mechanical Design

Scanning Probe Microscopes (SXM) have to fulfill specifications which are contradictory. First, the scan range of an SXM should exceed the size of the largest features on the sample. This usually implies the use of large and thin walled piezo electric translators. For many applications, it is desirable to have the possibility to coarse position laterally the sample versus the tip. This can be done by adding some hardware, thus increasing the size of the microscope. On the other hand, any SXM is, first of all, a system of mechanical resonators. This system has many resonances with varying quality factors. These resonances can be excited either by the surroundings or by the rapid movement of the tip or the sample. It is of paramount importance to optimize the design of the SXM for high resonance frequencies. This usually means to decrease the size of the microscope[**?**].
By using cube-like or ball-like structures for the microscope, one can considerably increase the lowest eigen-frequency. The eigen-frequency of any spring is given by

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m_{eff}}} \tag{4.10.29}$$

where $k$ is the spring constant and $m_{eff}$ is the effective mass. The spring constant $k$ of a cantilevered beam with uniform cross section is given by[**?**]

$$k = \frac{3EI}{l^3} \tag{4.10.30}$$

where $E$ is the Young's modulus of the material, $l$ the length of the beam and $I$ the moment of inertia. For a rectangular cross section with a width $b$ (perpendicular to the deflection) and a height $h$ one obtains for $I$

$$I = \frac{bh^3}{12} \tag{4.10.31}$$

Combining equations (4.10.29), (4.10.30) and (4.10.31) we get the final result for $f$:

$$f = \frac{1}{2\pi}\sqrt{\frac{3EI}{l^3 m_{eff}}} = \frac{1}{2\pi}\sqrt{\frac{Ebh^3}{4l^3 m_{eff}}} \tag{4.10.32}$$

The effective mass can be calculated using Raleigh's method. The general formula for the kinetic energy T using Raleigh's method is

$$T = \frac{1}{2} m_{eff} \left(\frac{\partial y}{\partial t}\right)^2 \tag{4.10.33}$$

We divide our bar of uniform cross section in infinitesimally small segments and calculate the kinetic energy $T$

$$T = \int_0^\ell \frac{m}{2\ell} \left(\frac{\partial y(x)}{\partial t}\right)^2 dx \tag{4.10.34}$$

For the case of a uniform beam with a constant cross section and length $\ell$ one obtains for the static deflection $y(x) = y_{max}\,(1 - (3x)/(2\ell) + (x^3)/(2\ell^3))$. Inserting $y_{max}$ into equation (4.10.33) and solving the integral yields

$$\begin{aligned} T &= \int_0^\ell \frac{m}{\ell}\left[\frac{\partial y_{max}}{\partial t}\left(1 - \frac{3x}{2\ell} + \frac{x^3}{2\ell^3}\right)\right]^2 dx \\ &= \frac{1}{2} m_{eff} \left(\frac{\partial y_{max}}{\partial t}\right)^2 \qquad (4.10.35) \\ m_{eff} &= \frac{33}{140} m \qquad (4.10.36) \end{aligned}$$

for the effective mass.
Combining equations (4.10.32) and (4.10.35) and noting that $m = \rho l b h$, where $\rho$ is the density of mass, one obtains for the eigen-frequency

$$f = \left(\frac{1}{2\pi}\sqrt{\frac{35}{33}}\sqrt{\frac{E}{\rho}}\right)\frac{h}{l^2} \tag{4.10.37}$$

A better estimate can be obtained by using the shape equation for an oscillating beam.

$$y(x) = a_1 \cos(kx) + a_2 \cosh(kx) + a_3 \sin(kx) + a_4 \sinh(kx) \tag{4.10.38}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Using the boundary conditions that $y(0) = 0$, $y'(0) = 0$, $y''(\ell) = 0$ and $y'''(\ell) = 0$ one obtains the relations

$$\begin{aligned} a_1 &= -a_2 \\ a_3 &= -a_4 \\ a_1 &= a_3 \frac{\sin(k\ell) + \sinh(k\ell)}{\cos(k\ell) + \cosh(k\ell)} \\ \cos(k\ell) &= -\frac{1}{\cosh(k\ell)} \quad \text{or} \\ k_0\ell &= 0.596864\pi \end{aligned} \tag{4.10.39}$$

or

$$\begin{aligned} y(x) = y_{max}(t) \Bigg[ & 0.1896058494 \cos\left(\frac{1.875103558x}{l}\right) \\ & -0.1896058494 \cosh\left(\frac{1.875103558x}{l}\right) \\ & +0.1391888034 \sin\left(\frac{1.875103558x}{l}\right) \\ & -0.1391888034 \sinh\left(\frac{1.875103558x}{l}\right)\Bigg] \end{aligned} \tag{4.10.40}$$

Plugging this into the integral (Gleichung (4.10.35) ) we obtain

$$T = 0.08796733941m \left(\frac{\partial y_{max}(t)}{\partial t}\right)^2 \tag{4.10.41}$$

and hence

$$m_{eff} = 0.1759346788m \tag{4.10.42}$$

This estimate is smaller than the one using the static deflection shape.
Further reading on how to derive this equation can be found in Thomson[**?**]. It is evident from this equation, that one way to increase the eigen-frequency is to choose a material with as high a ratio $E/\rho$. Table 4.14 gives an overview over some materials. This table shows, that aluminum and steel are the best construction materials in terms of eigen-frequencies.
Another way to increase the lowest eigen frequency is also evident in equation (4.10.29). By optimizing the ratio $h/l^2$ one can increase the resonance frequency. However it does not help to make the length of the structure smaller than the width or height. Their roles will just be exchanged. Hence the optimum structure is a cube. This leads to the design rule, that long, thin structures like sheet metal should be avoided. If a given resonance frequency can not be changed any more, its quality factor should be as low as possible. This means that an inelastic medium such as rubber should be present to convert kinetic energy into heat.
A typical SXM consists of many structures with coupled resonance frequencies. Hence there might be low frequency beats in the amplitude of an oscillation. Let us investigate the consequences of such beats for an SXM with linear response (such as certain force microscopes) and for an SXM with a nonlinear response

| Material | E in $10^{10}$ $N/m^2$ | $\rho$ in $10^3 kg/m^3$ | $E/\rho$ in $10^7 m^2/s^2$ |
|---|---|---|---|
| Aluminum | 7.0 | 2.7 | 2.6 |
| Brass | 9.0 | 8.5 | 1.1 |
| Bronze | 10.5 | 8.8 | 1.2 |
| Copper | 11.0 | 9.0 | 1.2 |
| Crown Glass | 6.0 | 2.6 | 2.3 |
| Nickel | 21.0 | 8.88 | 2.4 |
| Steel | 20.0 | 7.8 | 2.6 |
| Tungsten | 39.0 | 19.3 | 2.0 |

Tabelle 4.14.: Material constants used for calculating resonance frequencies (adapted from Anderson[**?**])

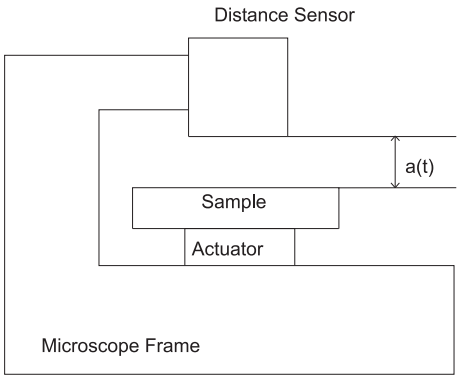



Abbildung 4.272.: Influence of beat frequencies in SXM. Two resonating bodies act on the distance between the sample and the tip $a(t)$. For a system with a nonlinear response to $a(t)$ the response also be present at beat frequencies.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

such as an *STM*. Figure 4.272 shows the set-up for this calculation. The sample is supposed to be mounted on an oscillating block with an amplitude of

$$a(t) = a_0 + a_1 \sin(\omega_1 t) + a_2 \sin(\omega_2 t) \tag{4.10.43}$$

opposite to a fixed probe. This probe is sensitive to variations in the distance between the sample and the tip. We assume that the frequencies $\omega_1$ and $\omega_2$ are larger than the frequency response of the probe whereas the difference frequency $\omega_1 - \omega_2$ is well within. The probe will only respond to the average *signal* given by

$$s = \frac{1}{\Delta T} \int_{t-\Delta T}^{t} r(a(t'))dt' \tag{4.10.44}$$

where $r(x)$ is the response function of the probe and $\Delta T$ is the integration time. For a linear response function $r(x) = x$ and an integration time $\Delta T$ large compared with $1/\omega_1$ and $1/\omega_2$ the integral in equation (4.10.44) gives 0. No effect of the two vibration modes can be seen in the output *signal* for a linear system.
If we assume a nonlinear response function $r(x) = x + \beta x^2$ then $r(a(t))$ becomes

$$\begin{aligned} r(a(t')) &= a_1 \sin(\omega_1 t) + a_2 \sin(\omega_2 t) + \beta \left(a_1 \sin(\omega_1 t) + a_2 \sin(\omega_2 t)\right)^2 \\ &= a_1 \sin(\omega_1 t) + a_2 \sin(\omega_2 t) + \beta \left(a_1^2 + a_2^2\right) - \\ &\quad \frac{1}{2}\beta a_1^2 \cos(2\omega_1 t) - \frac{1}{2}\beta a_2^2 \cos(2\omega_2 t) + \\ &\quad \beta a_0 a_1 \cos\left((\omega_1 - \omega_2)t\right) - \beta a_0 a_1 \cos\left((\omega_1 + \omega_2)t\right) \end{aligned} \tag{4.10.45}$$

Inserting equation (4.10.45) into equation (4.10.44) and solving the integral gives non-zero results only for the term with constant amplitude $\beta\left(a_1^2 + a_2^2\right)$ and the term $\beta a_0 a_1 \cos\left((\omega_1 - \omega_2)t\right)$ oscillating at $\omega_1 - \omega_2$. The first non-zero term represents a DC offset to probe response, whereas the second term is a low frequency modulation which may well be within the control bandwidth of the feedback loop. Such an interference *signal* will appear as an apparent surface structure.
As a consequence, one has to be careful not to design two parts of an SXM with nearly the same frequency. One case which is especially bad is if the tip has resonance close to an eigen frequency in some other part of the structure.

### 4.10.4.2. Vibration Isolation

We have seen in section 4.10.4.1 that a SXM viewed as a mechanical system has a variety of resonance frequencies, which may be even coupled. These resonant frequencies are excited both by the surroundings and by the scanning piezo within the SXM. The effect of the scanning piezo can be minimized through careful design, but it cannot be eliminated.
The microscope, however, can be isolated from the vibrations present in the surrounding. First, the microscope is affected by building vibrations and by the sounds people create when moving around. The frequency spectrum of these noises peaks between 10 and 100 Hz. The amplitudes of the ground movement are of the order of several $\mu$m. It is a wise practice to avoid any resonating body within the SXM with a resonance frequency $\omega_0$ below a few 100 Hz to 1 kHz. Since the coupling to a mechanical harmonic oscillator scales like $1 - 2Q^2((\omega - \omega_0)/\omega_0)^2$ a larger difference

between the driving frequency and the resonance frequency decreases the coupling. This scaling law is correct for small deviations of $\omega$ from $\omega_0$. The reference amplitude is the amplitude on resonance. The amplitude of the building vibrations reaching the microscope can be further decreased by isolating the microscope from these frequencies.

The first tunneling microscopes used a spring system with eddy current damping for vibration isolation. A description of this system can be found in Binnig and Rohrer[**?**]. This vibration isolation system, tough effective, has its disadvantages. The fundamental resonance frequency of the spring mass system has to be of order 1 Hz to be effective in suppressing the building vibrations. If this resonance frequency were aperiodically damped, then the increase in damping with frequency above the resonant frequency would be too slow. Furthermore it is difficult to get aperiodic damping with an eddy current brake at room temperature. For an under-critical damping of the spring-mass system the vibration amplitudes at resonance are amplified! Therefore using a vibration isolation system with a resonant frequency between 10 and 100 Hz, where the building vibrations peak, has to be avoided.

A second disadvantage of the spring stage vibration isolation system is the lack of a well defined position of the microscope with respect to the surrounding. This deficiency becomes a problem when using mechanical approach systems, in situ sample transfer and tip exchange. All these tasks demand spatially well defined locations and the ability to exert forces on the microscope. One solution to this problem is to temporarily bridge the vibration isolation and to grip the microscope with a special tool. Another problem one might have is the very high Q of the springs on which the microscope is suspended. To reach the necessary low resonant frequency, their spring constant has to be low, which also means that their resonance frequency can be fairly low. Special problems might occur if the springs come into resonance with some other part of the microscope or the surroundings. Lastly the spring-mass vibration isolation systems tend to be very bulky.

To decrease the size of the vibration isolation system, Gerber *et. al.*[**?**] proposed to use stacks of metal plates, separated by viton spacers. This set-up is still a mass-spring damping system, but it is most effective at high frequencies, where its multiple stages act like a higher order low pass filter. Combined with an SXM of rigid design, such a viton-metal stack can provide superior performance. Such designs are widely used in *UHV-STM*-systems and are known as pocket size STMs.

Outstanding results can be obtained using optical tables or at least the vibration isolation supports designed for these tables. Entire UHV-chambers can be mounted on such vibration isolated tables. An additional decrease of the vibration amplitudes can be achieved by placing the microscope in a room with low ground movements, such as a basement room or on the separate basement of a microscopy room.

A cost effective, though not very efficient way to isolate the microscope from the building vibrations is to mount it on a platform suspended by bungee cords from the ceiling[**?**]. If the microscope is very rigidly designed then it can be used with no further vibration isolation on these platforms or on optical tables. For an improved performance, the viton-metal stack can be combined with the external vibration isolation methods.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

#### 4.10.4.3. Sound Isolation

Environmental noise not only consists of building vibrations, but also of sounds. Sources of sound may be the voices or actions of the experimenters, air conditioning, motors built into equipment like computers, or street noise, to name a few. SXMs operating in vacuum are perfectly isolated against this noise. Microscopes working in air need added protection. A good way to isolate an ambient pressure microscope from sound is to build it into an airtight box with heavy walls, similar to the way high class loudspeaker boxes are built. The closed room inside the box has its own resonances, which must be damped by using suitable materials. Good are the foams used on the walls of sound recording studios or in soundless rooms.

When using metal coatings on the sound isolation box or even metal as the construction material of the sound isolation box, it can be used in addition as an electrical shield, protecting the sensitive electronic circuits from electronic noise.

#### 4.10.4.4. Thermal Drift

Scanning Probe Microscopes operate in environments where the temperature is not constant. Any change in temperature causes a contraction or expansion of the various parts of the SXM. Since these parts are manufactured from different materials, their thermal expansion coefficients will be different. This causes the probe tip to be displaced both laterally and vertically from the old position with respect to the sample. To estimate the thermal drift, we will assume that an SXM is built of a steel body, with a piezo tube holding the probe. We further assume that the size of the microscope, i.e. the distance from the probe tip to the point where it is joined with the sample holder, is 5 cm. We further assume that the difference in the thermal expansion coefficients between the steel and the piezo ceramic material is a very good $10^{-6}$ $1/K$. The probe tip will move 50 nm for every K change in temperature. To perform experiments such as local tunneling spectroscopy, where the gap should be kept constant to better than 0.01 nm a thermal environment with a temperature variation of less than 0.2 mK over the *measurement* time is required! For *measurement* times of a few seconds such a temperature stability is easily achieved. Problems arise for temperature controlled samples. Only the best of these temperature controllers, together with sophisticated shielding techniques will provide the necessary stability.

The sensitivity to temperature variations can be minimized by using a symmetric and balanced design[**?**, **?**]. As an example we consider an SXM with a cylindrical symmetry. We will use a piezo tube as the tip translator. The tip is mounted in the center of the piezo tube, which, in turn, is mounted in the center of the microscope body. Because of symmetry, the tip will remain at the center of the microscope body even if the temperature varies. It will move up or down with respect to the sample. The differential thermal expansion can be further minimized by employing the same materials for, in our example, the piezo tube and the microscope body. This will give the best match for the thermal expansion coefficients and would for our example minimize the vertical movement of the tip with respect to the sample.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

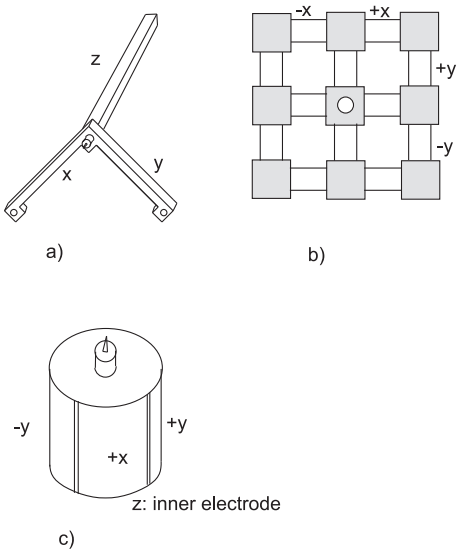



Abbildung 4.273.: Types of piezo scanners: a) the tripod; b) the thermally compensated scanner and c) the piezo tube.

#### 4.10.4.5. Piezo Scanners

Almost all STMs use piezo translators to scan the tip, or seldom, to scan the sample. Even the first *STM*[**?**, **?**] and some of the predecessor instruments[**?**, **?**] used piezo translators for scanning. Microscopes using magnetostrictive materials[**?**] or electromagnetic drives[**?**] have been proposed. We will concentrate on piezo electric materials.

An electric field applied across a piezo electric material causes a change in the crystal structure, with expansion in some directions and contraction in others. Also, a net change in volume occurs. Detailed descriptions of the piezo electric effect can be found in solid state physics textbooks[**?**]. The transverse piezo electric effect is by far the most important for scanning probe microscopes. The expansion perpendicular to the applied electric field $\boldsymbol{E}$ for a long slab of material with the field applied across the small sides is

$$\Delta l = l|\boldsymbol{E}|d_{31} = l\frac{V}{t}d_{31} \tag{4.10.46}$$

where $d_{31}$ is the piezoelectric constant, $V$ the applied voltage and $t$ the thickness of the piezo slab or the distance between the electrodes where the voltage is applied. This allows to choose the sensitivity of a piezo actuator within the limits of its mechanical stability.

The first STMs all used piezo tripods for scanning (see for instance Binnig and Rohrer[**?**]). The piezo tripod (figure 4.273a)) is an intuitive way to generate the three dimensional movement of a tip attached to its center. However, to get a suitable stability and scanning range, the tripod needs to be fairly large (about 5 cm). Its size and its asymmetric shape make it very susceptible to thermal drift. The design of van Kempen and van de Walle[**?**] (figure 4.273b)) tries to circumvent this problem by using a symmetrical design. Its thermal drift performance is much better than the simple tripod. However a complicated assembly of many piezo pieces is required. The tube scanner (figure 4.273c)) is now widely used in scanning

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

tunneling and scanning probe microscopy for its simplicity and its small size[**?**]. The outer electrode is segmented in four equal sectors of 90 degrees. Opposite sectors are driven by signals of the same magnitude, but opposite sign. This gives, through bending, a two dimensional movement on, approximately, a sphere. The inner electrode is normally driven by the $z$ *signal*. It is possible, however, to use only the outer electrodes for scanning and for the $z$-movement. The main drawback of applying the $z$-*signal* to the outer electrodes is, that the applied voltage is the sum of both the $x$- or $y$-movement and the z-movement. Hence a larger scan size effectively reduces the available range for the $z$-control.

Piezo scanners, tubes and tripods, are made of piezo ceramic material. Piezo materials with a high conversion ratio, *i.e.,* a large $d_{31}$ or small distances between the electrodes, allowing large scan ranges with low driving voltages, do have substantial hysteresis resulting in a deviation from linearity by more than 10 %. The sensitivity of the piezo ceramic material (mechanical displacement divided by driving voltage) increases with reduced scanning range, whereas the hysteresis is reduced. A careful selection of the material for the piezo scanners, the design of the scanners, and of the operating conditions is necessary to get optimum performance.

#### 4.10.4.6. Control Electronics: Basics

An important role for an optimum performance of a Scanning Probe Microscope plays the control electronics and software. Control electronics and software are supplied nowadays with commercial SXMs. Some manufacturers sell their control electronics and software without a microscope. Control electronic systems can use either analog or digital feedback. While digital feedback offers greater flexibility and the ease of configuration, analog feedback circuits might be better suited for ultra low noise operation.

We will describe here the basic electronic setups for scanning tunneling microscopy and spectroscopy. The next section is devoted to the discussion of spectroscopy instrumentation. The concepts worked out in the following are also applicable to the control electronics for other scanning probe microscopes.

The main task for the control electronics for a *STM* is to maintain a distance of order 1 nm between the tip and the sample with an accuracy of $< 0.01$ nm. Figure 4.274 shows a block schematic of the a typical *STM* feedback loop. The sample is connected to a pipolar, adjustable tunnel voltage source. The other side of the tunnel junction, the tip, is connected to a current to voltage converter (I/V-converter). Alternatively, the tunnel voltage could have been connected to the tip and the I/V-converter connected to the sample. In any case the I/V-converter should be connected to the side which is less affected by ambient noise and has less stray capacitance to earth. The tunnel voltage side is connected to a low impedance voltage source, and hence not very much affected by electrical interference. The I/V-converter on the other hand may have a considerable input impedance and is connected to its voltage supply via the tunneling resistance and the feedback resistance (both of the order of MΩ to GΩ) and is therefore very much affected by stray electrical fields. The unavoidable stray capacitance to earth will severely degrade the frequency response of the I/V-converter and hence of the *STM*. Therefore the I/V-converter should be located as close to the sample or to the tip as possible. In the case of additional mechanical or electrical equipment connected

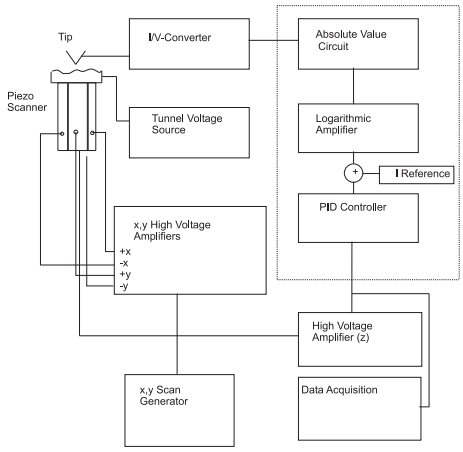



Abbildung 4.274.: Block schematics of the electronics for *STM*. The tunneling current at the tip is converted to a voltage and processed by feedback electronics. The drive voltages for the $x$-, $y$-, and $z$-electrodes are provided by high voltage amplifiers. The scan generator and the data acquisition system can be analog or digital.

to the sample (heater, sample exchange mechanism, electrochemical cell, ...) the I/V-converter has to be connected to the tip.

The output voltage of the I/V-converter is fed into an absolute value circuit. The absolute value circuit simplifies the use of both polarities of the bias voltage. The distance to current characteristics of the tunnel junction is exponential. Often it is advantageous to linearize the feedback loop after the rectifier. A set voltage "I Reference" is subtracted from the output of the linearizing unit or the rectifier. The resulting error voltage is then integrated and, optionally, amplified ("PID Controller"). The gain of the integrator (high gain corresponds to short integration times) and of the proportional amplifier is set as high as possible without generating more than 1% overshoot. High gain minimizes the error margin of the current and forces the tip to follow the contours of constant density of states as good as possible. This operating mode is known as "Constant Current Mode". The outputs of the integrator and the amplifier are summed and amplified by a high voltage amplifier. STMs using piezo tubes usually require ±150 V at the output. Other designs might require amplifiers delivering more than ±1000 V or as little as ±15 V.

To scan the sample, additional voltages at high tension driving the scanning piezos are required. For example, with a tube scanner, four scanning voltages are required, namely $+V_x$, $-V_x$, $+V_y$ and $-V_y$. The $x$- and $y$-scanning voltages are generated in a scan generator (analog or computer controlled). Both voltages are input to the two respective power amplifiers. Two inverting amplifiers generate the input voltages for the other two power amplifiers.

The topography of the sample surface is determined by recording the input-voltage to the $z$-high voltage amplifier as a function of $x$ and $y$ ("Constant Current Mode"). Recording devices like chart recorders, analog storage oscilloscopes, video frame grabbers, or computer data acquisition systems are used. The height variations on the sample surface can be determined quite accurately from the known piezo calibration.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

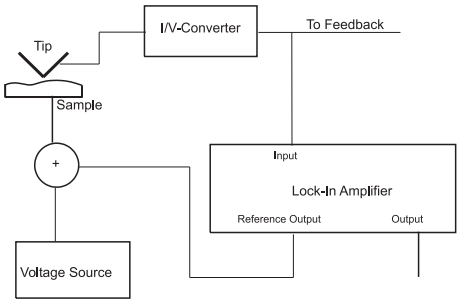



Abbildung 4.275.: The block schematic for measuring $\frac{\partial I}{\partial V}$ with the *STM*. A small ($\approx$ 1mV) modulation voltage is added to the tunnel voltage. The resulting tunneling current is converted to a voltage. The feedback circuit operates on the DC-component of the tunneling current, whereas the AC-component is demodulated by the lock-in amplifier.

Another operating mode is the "Constant Height Mode". It is feasible only on flat surfaces like *graphite*. The gain in the feedback loop is lowered and the scanning speed increased such that the tip does not follow any more the surface contours. It is held at constant height. Here the tunneling current is recorded as a function of $x$ and $y$. This mode usually has a better *signal* to noise ratio than the "Constant Current Mode", mainly because the surface data appears in higher frequency bands in the tunneling current.

### 4.10.4.7. Control Electronics: Spectroscopy

The basic electronic circuits outlined in section 4.10.4.6 can be expanded to allow spectroscopic imaging of the sample. Four modes of spectroscopic imaging are in common use: measuring $\frac{\partial I}{\partial V}$, $\frac{\partial I}{\partial s}$ spatially resolved, alternating the tunneling voltage between two values between scans and measuring $I(V)$ curves at constant height for some or all points on the sample.
Figure 4.275 shows an electronics set up for a $\frac{\partial I}{\partial V}$ *measurement*. The output voltage of an oscillator ("Reference Output") is added to the bias voltage. The sum is connected to the sample. The resulting tunneling current is converted to a voltage by the I/V-converter, as usual. The low frequency part is fed into the feedback electronics and controls the position of the tip. The modulation on the bias voltage, whose frequency should be well above the cut off frequency of the feedback loop, is connected to the *signal* input of a lock-in amplifier. The reference for demodulating the $\frac{\partial I}{\partial V}$-*signal* comes from the oscillator. The output of the lock-in amplifier can be recorded as a function of position alone or together with the $z$-*signal*.
Figure 4.276 shows the necessary connections to measure $\frac{\partial I}{\partial s}$. This time the output of the oscillator ("Reference Osc.") is summed with the output of the feedback control electronics. The amplitude of the modulation voltage is chosen such as to move the tip rapidly with an amplitude of about 1 pm to 10 pm. The resulting modulation on the tunneling current is fed into the lock-in amplifier and detected. The lock-in amplifier output can be recorded in the data acquisition system, alone or together with the $z$-position. The frequency of the modulation should be smaller by a factor of 3 than the lowest resonance frequency of the piezo tube. If the

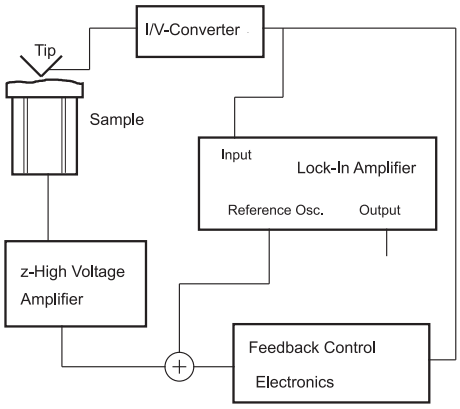



Abbildung 4.276.: The block diagram for measuring $\frac{\partial I}{\partial s}$ with the *STM*. The spacing between the tip and the sample is modulated by adding a small AC-voltage onto the output of the feedback amplifier. This results in a modulation of the tunneling current which is detected by the lock-in amplifier.

modulation frequency is too close to a resonance frequency, peaking will occur and the mechanical movement $\partial s$ will no longer be given by the static sensitivity of the piezo tube. The lower limit of the modulation frequency is given by the cutoff frequency of the feedback loop. The gain of the feedback loop will decrease the current modulation due to the mechanical modulation and it will also introduce an additional gain dependent phase shift. Since the gain depends on the local barrier height, it is hard to give a quantitative interpretation for too low a modulation frequency. For microscopes with large scan piezos it is sometimes impossible to fulfill both requirements. An interesting implementation of this kind of spectroscopy using photothermal modulation of the tunnel gap has been published by Amer[**?**].

Figure 4.277 shows the set-up for alternating two bias voltages $V_1$ and $V_2$ on alternating scans. The basic feedback loop is the same as in previous experiments. This time we also need to consider the electronics generating the x-scanning *signal*. The triangular output of the scan generator (function generator or computer output) is connected to the x and, through an inverting amplifier, to the -x electrodes of the scanning piezo. Trigger pulses marking the begin or end of each x-scan are used to toggle a divider by 2[**?**]. Its output in turn toggles the electronic switch conntecting the sample to the voltage sources $V_1$ and $V_2$. Odd and even numbered scan lines are stored in 2 separate images showing the topography at the two different voltages. Why does one alternate the voltages between the scan lines and not from frame to frame? The thermal drift always present would make it difficult to determine a registry between the two images. By interlacing the two voltages, the time from a scan line in one image to the corresponding scan line in the other is smaller by the number of scan lines. Hence the drift is smaller by the same factor.

The set-up for the last mode of spectroscopy, Current Imaging Tunneling Spectroscopy (CITS) is depicted in figure 4.278. To do CITS, we need to modify the basic feedback loop. The bias voltage source is a function generator which is triggered from a timing control. A trigger pulse starts the A/D conversion for the voltage sweep and also disconnects the integrator from the error *signal*. The integrator holds the output voltage and, hence, the tip stays at a fixed height above the

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

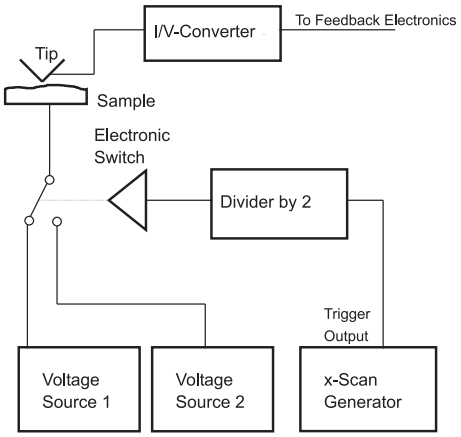



Abbildung 4.277.: The block diagram for measuring with alternating bias voltages on alternating scan lines. The $x$-scan generator toggles at the end of a scan line the bias voltage between two preset values. The feedback electronics keeps the tunneling current constant. The resulting topographs are a function of the bias voltage. This setup allows a direct comparison of two energies in the local density of states.

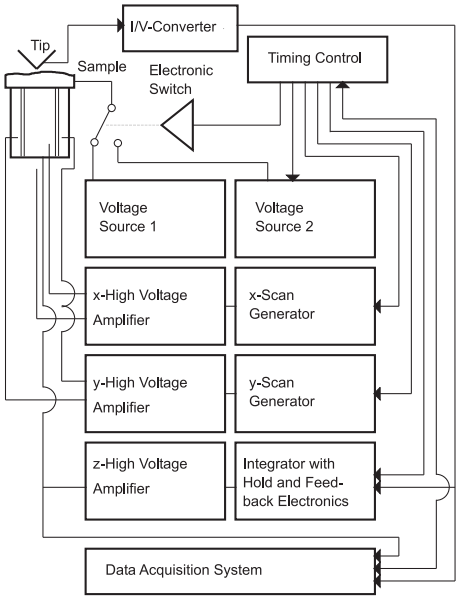



Abbildung 4.278.: Current Imaging Tunneling Spectroscopy (CITS). During the raster scan of the surface the tip is stopped at predetermined locations controlled by the timing control. At each such points the bias voltage is switched from voltage source 1 to voltage source 2. While the feedback loop is disabled, voltage source 2 ramps the voltage over a predefined range. The current variations are recorded as a function of the bias voltage. Then voltage source 1 is turned on again and the feedback loop is enabled.

sample. The feedback loop is disabled until the end of the bias voltage sweep. The error signal $V_E$ was zero at the begin of the voltage sweep. If there were no drift, it would be again zero after the voltage sweep. Since thermal drift is always present, it is necessary to optimize the microscope for low thermal drift and to disable the feedback loop only for short times. Even if there is a non-vanishing error, the integrator will ramp smoothly to the new position. The proportional amplifier in the feedback loop should be omitted so that the output voltage does not jump to the new position.

#### 4.10.4.8. Tip Preparation for Scanning Tunneling Microscopes

An important point in the successful operation of a *STM* is the tip preparation. The tips should have a minimal radius of curvature at the end, they should have a narrow diameter to penetrate into holes, but they should also be thick enough to keep the resonance frequency high. The tip material should be stable in high electric fields and, when operated in air or corrosive liquids, chemically stable. Tips have been made out of wires of tungsten, platinum-iridium, platinum, gold, and nickel.
A variety of tip preparation methods was used by *STM* researchers. In the first days of *STM* tips were prepared by grinding tungsten wire. However this resulted in tips of minor quality. A simple preparation method is to cut a thin wire (0.1 mm to 0.3 mm) of the desired material by a cutter or by scissors. Low quality tools work best, because they tear the material apart, forming sharp tips. Cutting is fast, but the success rate for working tips is not too high. Moreover, the shape of the tips is not well defined, making it harder to use them on strongly corrugated samples. The shape of the tip is less important when imaging atomically flat samples. The material suited best for this method seems to be platinum-iridium wires.

#### 4.10.4.9. Coarse Sample Positioning

For many experiments it is important to position the tip over a well defined area of the sample or to move the tip from one location to another. Biological samples, integrated circuits and samples with surfaces showing localized phenomena demand a coarse positioning. On other samples, one has to compare images from a range of locations to ensure that the observed structures are indeed representative. In this chapter we will discuss coarse positioning under optical control, whereas the next chapter gives a short introduction to positioning the tip under Scanning Electron Microscope control.
The coarse positioning devices in an *STM* have to fulfill contradicting requirements. First, they should be as small and rigid as possible to preserve the high resonance frequencies of the original design. On systems having a vibration isolation, the drive for the coarse positioning device needs to be on the microscope or, alternatively, its drive connection should be detachable.
The first STMs[**?**, **?**] employed a “louse” for coarse positioning and for the approach. It is a two-dimensional piezo motor, moving on a steel plate. In addition to linear movements, it also allows to turn the sample. A detailed discussion of this device can be found in the literature[**?**, **?**]. The main disadvantage of a “louse” is that it has a rather large size. Newer designs use commercially available inchworms. The position of the tip with respect to the sample is observed by optical

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

telescopes with a position resolution of a few 10 $\mu$m.
Another type of coarse positioning is used in the microscopes of the "Besocke"-type[**?**]. There the sample rests on three piezo tubes. The tip is mounted on a fourth piezo tube in the center of the other three. By moving the outer three piezo tubes rapidly in one direction and contracting them slightly during the movement, their contact points can be moved by a small amount. The three tubes return then slowly to their starting position. By using a sample holder with a tilted border, movements in x, y and z as well as a sample rotation can be accomplished. A high power optical microscope can be used to position the tip, if the sample is transparent[**?**].
Other coarse positioning devices employ micrometer screws and manual or micromotor drives. The manual drives require a decoupling during *measurement* to avoid the transmission of vibrations to the microscopes. Motorized micrometers can be used on vibration isolation platforms, without short-circuiting the isolation.
It is important to view the movement of the tip over the sample surface. A medium to high magnification optical microscope is usually incorporated into the design of the SXM. The scanning probe microscope can be remotely operated if the optical microscope is connected to a video camera. This allows to mount the microscope in a sound isolated box, under inert gas and on vibration isolation tables and still having control.

#### 4.10.4.10. Integrating a Scanning Tunneling Microscope into a Scanning Electron Microscope

The obtainable resolution of optical positioning is of order 1 $\mu$m. An order of magnitude in resolution can be gained for conducting samples when using an Scanning Electron Microscope (SEM)[**?**, **?**] (See figure 4.279). There are two crucial points to be observed: first, the sample and tip have to be arranged such that they are well visible from the electron gun of the SEM. The *STM* has to be connected rigidly to the SEM to ensure a good resolution of the SEM. The vibration isolation has to be located outside the SEM vacuum system. Secondly, the SEM has to have an ultra high vacuum chamber. The rest gas in ordinary high vacuum contains hydrocarbons which are cracked by the electron beam at the sample surface. The resulting carbon film is poorly conducting and can render an *STM* inoperable (A hydrocarbon rich rest gas together with an electron beam is used to write patterns on integrated circuits. The hydrocarbon film is known as "contamination resist").
The tunneling current in the *STM* is affected by the electron beam. If the electron beam current becomes comparable to the tunneling current it might cause the feedback loop of the *STM* to become unstable. The SEM should always be operated at currents lower than the tunneling current, unless the *STM* is not in the tunneling regime.
One advantage of the SEM is its huge depth of view. The tip and the sample can be seen sharp simultaneously. This allows a very precise positioning of the tip, which must have a small opening angle to not obscure the sample. If the SEM is equipped with Auger electron analysis it can determine the chemical composition of the sample surface on submicrometer scale.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

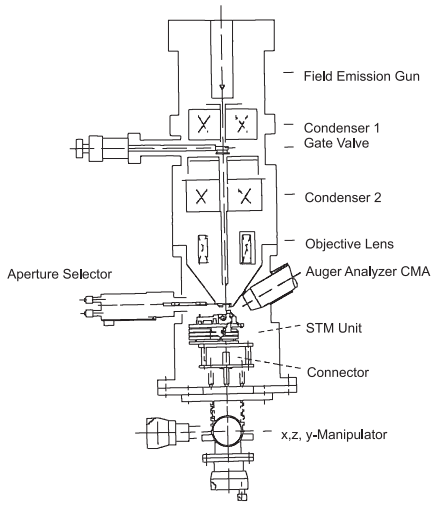



Abbildung 4.279.: A schematic drawing showing the integration of a *STM* into a scanning electron microscope. This figure is taken from Ch. Gerber *et al.*[**?**] and reproduced with permission of the American Institute of Physics.

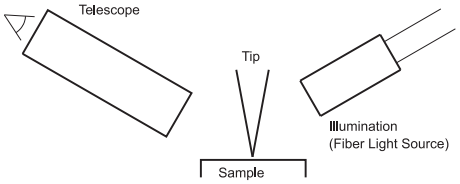



Abbildung 4.280.: Typical setup of a microscope for approach. The tip and the sample are vied through a telescope or a long distance microscope as a stereo microscope. The tip and its mirror image form an X-shaped pattern, which is separated at the center. The approach is finished when the tip and its mirror image seem to be touching. Typically the tip is then a few $\mu$m from the sample apart.

### 4.10.4.11. Approaching the Tip to the Sample

One of the most important steps in operating an *STM* is the approach of the tip to the sample. A carefully prepared tip is of no great use if it is damaged the moment it reaches the sample. The task is, to bring the tip from a distance of 1 to a few millimeters down to a distance of about 1 nm and to establish a tunneling current of about 1 nA (A scaled up task would be to start a car 1 km away from the wall and bring it to a halt at a distance of 1 mm from the wall. The additional difficulty would be, that the driver does not see the wall until 2 to 3 mm away). If we assume that the tip approaches the sample surface with 1 $\mu$m/s and that it is to stop within 1 nm, then the acceleration is 100 times the earth's acceleration for 1 ms.
If we were to use the speed of 1 $\mu$m/s for the whole distance of several mm, the whole approach could easily take an hour or even more. Most STMs are equipped with a two stage approach: first a coarse approach under optical control and then the fine approach under electronic control. The coarse approach is mostly done

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

with fine pitched screws. A binocular with a long working distance is set up as depicted in figure 4.280. On flat surfaces, one can see at the same time the tip and its mirror image. Depending on the magnification of the binocular, the tip can be approached to 2-4 $\mu$m of the surface.
One way to make the fine approach is to use similar fine pitched screws, but with a mechanical disadvantage between the screw displacement and the tip displacement[**?**, **?**]. This fine approach screw can be driven by a stepper motor, by a synchronous AC-motor or by a DC-motor. Stepper motors have the advantage, that they can be stopped very rapidly, whereas synchronous motors an DC-motors run more smoothly. The effect of the jumps from step to step can be minimized by selecting a stepper motor with gear reduction and by running it with a reduced driving voltage (e.g. 5 V instead of 12 V). In addition, stepper motors are easily interfaced with digital logic or with computers.

#### 4.10.4.12. The Operating Medium

The first STMs were operated in vacuum[**?**, **?**]. But one soon learned to operate the *STM* in other media. Park and Quate[**?**] demonstrated the operation of the *STM* in air. Sonnenfeld and Hansma[**?**] showed that atomic resolution was possible in water. Elrod[**?**] operated their *STM* at cryogenic temperatures. *Cryogenic STM*s may be operated either immersed in liquid helium or in small cooled vacuum chambers.
STMs operating in fluids are particularly appealing for the investigation of biological samples. The possibility to image molecules in their native environment is unparalleled at the resolutions the *STM* obtains. The chemical and electrochemical environment can be accurately controlled. Chemical reactions[**?**, **?**] can be triggered while imaging and their time evolution can be observed, provided it is not too fast. The operation of the *STM* in polar fluids like water requires the use of isolated tips to minimize the ionic current between tip and sample. Such tips are readily available from commercial sources.

### 4.10.5. Resolution of a Scanning Tunneling Microscope

The resolution of an *STM* depends on the geometry of the tip and the sample and on their respective electronic structure. For large ($\mu$m) objects, the geometry of the tip on a $\mu$m length scale plays an important role on how the image is modified by the tip shape. On those length scales, it is usually possible to determine the tip shape and to partially correct the resulting image.
On an atomic scale, there are no general resolution criteria in *STM* like the well-known diffraction limit in optical microscopy. But even there the vertical resolution of phase-contrast microscopy has nothing to do with the diffraction limit except that the lateral extension of the object must be comparable to or larger than the wavelength of the light. In *STM*, a measure of the vertical resolution is the stability of the tunnel barrier width, since height variations of the sample surface smaller than the tip-sample vibration amplitudes are usually masked. The lateral resolution of an *STM* is governed by the width of the tunneling current filament, by the physical properties the tip and sample and by the electronic states on the surface and on the tip. Various resolution criteria have been published in the literature:

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

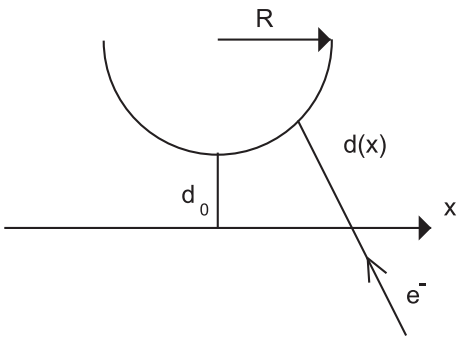



Abbildung 4.281.: Semiclassical calculation of the resolution of the *STM* as a function of the tip radius and the tip-sample separation. The electrons are assumed to tunnel from the sample to the tip on straight lines crossing the center of curvature of the tip. $x$ denotes the distance from the center of the current filament to the point where the electrons leave the sample.

1. A first criterion is based on the effective diameter $L_{eff}$ of the tunnel current filament. It is applicable for theoretical approaches which provide a tunnel current density profile from which $L_{eff}$ can be derived[**?**, **?**].

2. For periodic structures equicurrent lines traced by the tip are calculated as a function of the tip radius, surface corrugation periodicity, amplitude and electronic structure, and tip-sample separation. The resolution limit is reached when the equicurrent lines become flatter than the *STM* system noise[**?**, **?**]. This approach connects lateral and vertical resolution in a nontrivial way.

3. Consider an incident plane wave perpendicular to the average sample surface. If the tip is able to sample the directly transmitted wave $\boldsymbol{G} = 0$ but excludes the first order waves, the periodic structure can not be resolved. The inverse of the largest detectable surface reciprocal lattice vector is taken as the resolution of the *STM*[**?**, **?**, **?**].

4. The approach of Lang[**?**] who investigates two parallel metal planes with an adsorbed atom on each side also yields a characteristic current variation when two atoms, each adsorbed on an infinite size electrode, are scanned past each other. Though not directly providing a resolution criterion, the resulting current variations give a good idea of the peculiar effects arising with adsorbates. It also gives indications on how to resolve them.

We will first derive a semiclassical argument[**?**] on how the tip radius affects atomic resolution images. Then we will discuss the results of Tersoff and Hamann[**?**] and of Lang[**?**]. In a last paragraph we will consider a $\mu$m-size structure and investigate the effect of the tip shape.
In the following, we give a simple, instructive derivation of the resolution of the *STM*, by calculating $L_{eff}$ (criterion 1) from the Simmons model[**?**] and compare it to the more rigorous treatments. For the calculation we assume that electrons tunnel from the sample to the tip at an angle $\alpha$ from the surface normal and that they behave classically as far as the direction of their movement is concerned (see figure 4.281). Hence all tunneling paths aim at the center of curvature of the tip

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

and the tunneling probability of the electron depends on the energy of its motion perpendicular to the surface and on the projection, $d(x)$, of the tunneling path onto the surface normal.
The kinetic energy of the electron motion perpendicular to the surface is determined by the magnitude of the wave vector $\boldsymbol{k}$

$$\boldsymbol{k} = \boldsymbol{K}_{\parallel} + \boldsymbol{k_z} \tag{4.10.47}$$

Using

$$k^2 = K_{\parallel}^2 + k_z^2 \tag{4.10.48}$$

and

$$E_{kin} = \frac{\hbar^2 k^2}{2m_e} \tag{4.10.49}$$

for the relation between the electron energy and the magnitude of its $\boldsymbol{k}$ we can calculate the effective decay constant which determines the tunneling probability

$$\kappa_{0,eff} = \left(\kappa_0^2 + K_{\parallel}^2\right)^{1/2}, \tag{4.10.50}$$

where $\kappa_0$ is the decay constant for electrons with $K_{\parallel} = 0$. Next we approximate $K_{\parallel}$ by

$$K_{\parallel}(x) = k\frac{x}{d+R}, \tag{4.10.51}$$

where $d$ is the distance from the surface to the apex of the tip of radius $R$ and $x$ is the lateral distance from the center of the tip. $\kappa_{0,eff}$ can then be expanded for small $x$, small bias voltages and zero temperature:

$$\kappa_{0,eff} = \kappa_0\left(1 + \frac{1}{2}\frac{k^2}{\kappa_0^2}\frac{x^2}{(d+R)^2}\right) = \kappa_0\left(1 + \frac{1}{2}\frac{E_f}{\Phi_0}\frac{x^2}{(d+R)^2}\right), \tag{4.10.52}$$

were $E_f$ is the Fermi energy measured from the bottom of the conduction band and $\Phi_0$ is the barrier height.
The tunneling distance as a function of the lateral position $x$ is

$$d(x) = (d+R)\left\{1 - \frac{R}{[(d+R)^2 + x^2]^{\frac{1}{2}}}\right\} \tag{4.10.53}$$

For small $x$, it can also be expanded. Neglecting higher order terms we get

$$d(x) = d + \frac{R}{2}\frac{x^2}{(d+R)^2}. \tag{4.10.54}$$

The total current flowing in a filament of radius $r$ is then

$$I(r) = A\int_0^r \exp\left[-2\kappa_0\left(1 + \frac{1}{2}\frac{E_f}{\Phi_0}\frac{x^2}{(d+R)^2}\right)\left(d + \frac{R}{2}\frac{x^2}{(d+R)^2}\right)\right] 2\pi x dx \tag{4.10.55}$$

All prefactors of the exponential are contained in $A$ and set constant to ease the integration.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

Keeping only terms up to order $x^2$ in the exponential slightly underestimates the tunneling current and the filament diameter. The total tunneling current after integration is then given by

$$I(r) = I_0 \left(1 - \exp\left[-\kappa_0 \frac{r^2}{(d+R)^2}\left(d\frac{E_f}{\Phi_0} + R\right)\right]\right), \tag{4.10.56}$$

where $I_0$ is the total current. The effective diameter $L_{eff}$ of the tunneling current filament carrying a fraction $\delta$ of the total current $I_0$ is

$$L_{eff} = 2\left[-\frac{(d+R)^2}{\kappa_0\left(d\frac{E_f}{\Phi_0} + R\right)}\ln(\varepsilon)\right]^{\frac{1}{2}}, \tag{4.10.57}$$

where $\varepsilon = 1 - \delta < 1$.
The Fermi energy $E_f$ and the barrier height $\Phi_0$ have approximately the same magnitude in metals. Hence equation (4.10.57) reduces to

$$L_{eff} = 2\left[-\frac{d+R}{\kappa_0}\ln(\varepsilon)\right]^{\frac{1}{2}}, \tag{4.10.58}$$

The resolution of an *STM* on the atomic level is critically dependent on the radius of curvature of the apex of the tip. This suggests to use very narrow tips with single atom apexes. These tips, however, are not stable during scanning. Landman and Luedtke[**?**] calculated in a molecular dynamics calculation that the forces between the tip and the sample will tear apart single atom tips. Hence a high resolution tip has to be a compromise between stability and high curvature. It is found that tip radii of a few nm to a few 10 nm give best results.
Tersoff[**?**] calculates the resolution of the *STM* under the assumption of small corrugation. Under this assumption, the inherently nonlinear behavior of an *STM* can be treated like the linear imaging process of an optical microscope, where the measured image is a convolution of the ideal image and an instrument response function. The calculation[**?**] yields a resolution function which is a Gaussian with an RMS width of

$$L_{eff} = \left(\frac{z_0}{2\kappa_0}\right)^{0.5}, \tag{4.10.59}$$

where $z_0 = d + R$ is the distance between the sample and the center of curvature of the tip and $\kappa_0$ is defined as above. Tersoff's solution for metals shows, that our crude semiclassical calculation gives the correct functional behavior and will agree with Tersoffs result, if we set $\varepsilon = 3.3 \times 10^{-4}$. Tersoff[**?**] notes, that for semiconductors or semimetals like *graphite* the resolution of an *STM* is higher than for metals imaged with the same tip.
The resolution of an *STM* for larger features with sizes of $\approx 1\mu$m or more is mainly determined by the tip shape[**?**, **?**]. The variation of the tunnel barrier width is of the order of 1 nm or less and hence negligible. To demonstrate the influence of the tip shape we are considering a conical tip with a tip angle of $\alpha$. Figure 4.282 shows calculated traces for tip angles of 70° and 30°. The effect of a larger tip angle on features extending from the surface is to broaden the width of the feature. On ditches, however, a larger tip angle can prevent the imaging of the bottom of the ditch (see figure 4.282a).

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

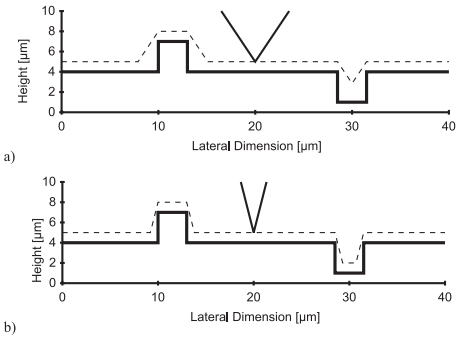



Abbildung 4.282.: Tip shape dependency of the *STM* resolution for large features. a): A tip angle of 70°; b) a tip angle of 30°. The figure shows the trajectory of the apex of the tip.

If we assume that the tip opening angle is $\alpha$ we can calculate the maximum depth $d_{max}$ for a given width $w$ of a rectangular ditch which can be imaged.

$$d_{max} = \frac{w}{2\tan(\alpha/2)} \tag{4.10.60}$$

Using the same simple geometrical arguments we can also calculate the apparent width $W$ at half height for a rectangular feature of width $w$ extending from the sample surface

$$W = w + h * \tan(\alpha/2) \tag{4.10.61}$$

| Tip Angle | Ditches $d_{max}$/w | Broadening 0.1 | 0.2 | 0.5 | 1 | 2 | 5 | 10 |
|---|---|---|---|---|---|---|---|---|
| 120° | 0.29 | 17% | 35% | 87% | 173% | 346% | 866% | 1732% |
| 110° | 0.35 | 14% | 29% | 71% | 143% | 286% | 714% | 1428% |
| 100° | 0.42 | 12% | 24% | 60% | 119% | 238% | 596% | 1192% |
| 90° | 0.50 | 10% | 20% | 50% | 100% | 200% | 500% | 1000% |
| 80° | 0.60 | 8% | 17% | 42% | 84% | 168% | 420% | 839% |
| 70° | 0.71 | 7% | 14% | 35% | 70% | 140% | 350% | 700% |
| 60° | 0.87 | 6% | 12% | 29% | 58% | 115% | 289% | 577% |
| 50° | 1.07 | 5% | 9% | 23% | 47% | 93% | 233% | 466% |
| 40° | 1.37 | 4% | 7% | 18% | 36% | 73% | 182% | 364% |
| 30° | 1.87 | 3% | 5% | 13% | 27% | 54% | 134% | 268% |
| 20° | 2.84 | 2% | 4% | 9% | 18% | 35% | 88% | 176% |
| 10° | 5.72 | 1% | 2% | 4% | 9% | 17% | 44% | 87% |

Tabelle 4.15.: Large scale resolution of a *STM*. The tip angle is the angle between the surfaces in a cross section. The broadening is referenced to the Full Width at Half Maximum and given as a function of $h/w$, where $w$ is the real width.

Table 4.15 summarizes the results for a number of tip angles $\alpha$. The first column gives the maximum ratio $h/w$ with which the tip just reaches the bottom of the ditch. The remaining columns give, as a function of $h/w$ the broadening at half

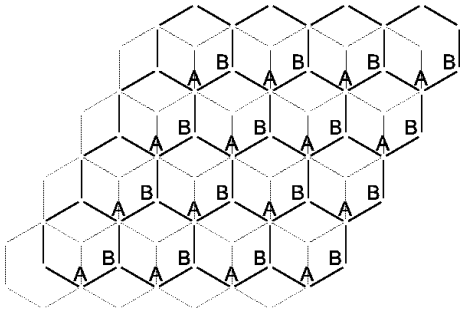



Abbildung 4.283.: Structure of the *graphite* crystal. The carbon atoms in the hexagonal rings at the surface occupy two inequivalent sites: At the A-site the carbon atom has a bond with the next lower layer, whereas at the B-site, there is no out-of-plane bond.

height of the structure due to the tip shape. Please note that $h$ and $w$ are the real widths and heights.

## 4.10.6. Tunnelspektroskopie

Tunneling spectroscopy on a local length scale is one of the strengths of *STM*. Images of semiconductor surfaces depend critically on the bias voltage, at which the image is acquired. Approximations and recipes to analyze spectroscopic data exist, but a full theory of tunneling spectroscopy based on the electron wave functions in metals is under way[**?**, **?**]). Selloni[**?**] proposed a generalization to the transfer Hamiltonian method by noting that for small voltages

$$J(V) \approx \int_{E_F}^{E_F+V} \rho(E)T(E,V)dE \tag{4.10.62}$$

is a qualitative approximation. $\rho(E)$ is the local density of states near the sample surface. Tersoff[**?**] criticizes that this expression does not explain the relationship of $dJ/dV$ with the spectrum of the electron density of states as determined for example by photo emission and inverse photo emission spectroscopy.
Stroscio *et al.*[**?**, **?**] proposed the use of $d\ln J/d\ln V = (dJ/dV)/(J/V)$ as a representation of the spectrum. The authors demonstrated a good agreement with known spectra of electron densities. The success of this representation of tunneling spectra depends on the cancellation of the exponential behavior of $T(E,V)$. Theoretical calculations[**?**] confirm the validity of this data representation and of equation (4.10.62).

## 4.10.7. *Graphite*

*Graphite* was the first substance to be imaged in air[**?**] and under liquids[**?**] at atomic resolution. Investigations under UHV conditions[**?**] as well as low temperature experiments[**?**, **?**] revealed the atomic scale structure of this surface. The relative ease of the imaging of the *graphite* surface under very different conditions has promoted its use as a calibration and test surface. Experiments going beyond the determination of the lateral scales revealed, however, that many unexpected effects play a role in the determination of the final appearance of the images.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

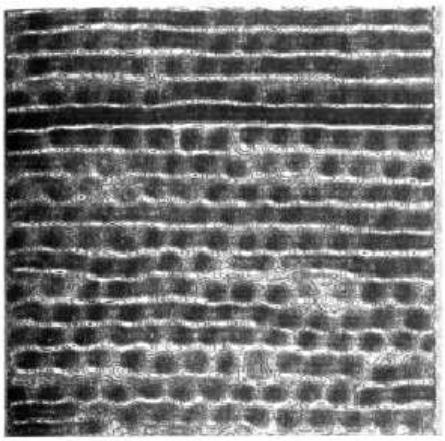


Abbildung 4.284.: Low temperature image of *graphite*. The sample is held at 6.8 K. The size of the image is 3.3 nm. The height varies by 0.54 nm, from black to white.

Figure 4.283 shows a schematic view of the structure of the *graphite* crystal. The carbon atoms are organized in layers of hexagons, with only weak bonding between the layers. This structure permits an easy cleavage of the surface, for instance by taping an adhesive tape to the surface and removing it carefully. The layered structure assures, too, that there are large, atomically flat terraces. As an example we show the topography of the *graphite* surface imaged at 6.8 K (figure 4.284). The image is similar to those obtained under other conditions. The hexagonal pattern of a *graphite* sheet is not resolved. Instead a mound-like pattern with the repeat distance of the unit cell of the *graphite* surface 0.54 nm is observed. At other times, hexagonal like structures or other structures are observed, but all with the periodicity of the *graphite* surface. Moreover, the height of the observed corrugation is often much larger than the height determined by He-scattering[**?**] or the calculated heights[**?**].

This behavior can be understood by noting that the *STM* images of the *graphite* surface are determined by the density of states of the surface. The Fermi surface of *graphite*[**?**] is confined entirely to the edges of the hexagonal Brillouin zone[**?**] and the references therein). This means that, to a good approximation, the local density of states of the *graphite* surface at the Fermi energy can be represented by three standing waves[**?**]. Tersoff[**?**] showed, that the particular form of the Fermi surface leads to a point with vanishing tunneling current in the unit cell. Since the *STM* traces curves of constant density of states, this would lead to large corrugations, limited by the actual tip shape and by the timing of the scanning and the feedback loop. Another theory put forward to explain the giant corrugations by Soler[**?**] first noted that the forces between the tip and the sample can not be ignored. They argued, that because of the electronic structure of the *graphite*, the tip was so close to the surface that it would introduce a deformation of the surface. The different dependencies on the distance of the force and the tunneling current would explain the observed corrugations. Another explanation[**?**, **?**] involved contamination layers between the tip and the sample. These contamination layers in air consist partly of water and of other substances present in the air.

Mizes[**?**] also noted, that the different possible images of the *graphite* surface can

be explained by multiple tip effects. If two or more tips coherently sample the amplitudes of the three standing waves describing the *graphite* surface a wide variety images can result. It is also possible to imagine entire leaflets of *graphite* are dragged across the *graphite* surface. Many tips in registry with the *graphite* surface sample the current. These images are, in principle, not distinguishable from single tip images of a large ordered area. This mode of imaging is more likely to occur at higher tunneling currents, where the tip is closer to the *graphite* surface. Steps, however, can only be imaged with a single or few atom tip and not with a leaflet of *graphite*. Normally, steps are only seen at low tunneling currents and high bias voltages, *i.e.,* at larger separations between tip and *graphite*.
In summary, *graphite* is an excellent sample to check the operation of a microscope and to calibrate the $x$- and $y$-deflections. However, the details of the electronic structure and the intricacies of the surface condition of *graphite* make SXM-images of this surface all but easy to understand.

## 4.10.8. Low Temperature Experiments

For structural investigations of proteins and biological membranes it may be desirable to image the samples at low temperatures to fix their structure of the sample. STMs have been used to image metal surfaces and superconductor surfaces at low temperatures from the start. Elrod[**?**] reported the first *measurements* of a superconducting tunneling gap using a *STM*. They soon thereafter were able to image the spatial variation of the superconducting tunneling gap[**?**]. The *measurement* of superconducting tunneling gaps by *STM*-like instruments has become an essential tool in characterizing High-TC-superconductors[**?**, **?**].
Figure 4.284 discussed in the previous section shows an atomic resolution image of *graphite* obtained at 6.8 K[**?**].
The design of a low temperature *STM* is much more demanding than the design of a room temperature *STM*[**?**, **?**, **?**, **?**, **?**, **?**] . The microscope has to be shielded to keep the consumption of coolants like liquid helium low enough. The encapsulation and the cool-down and warm-up times greatly increase the turnaround time. The mechanical systems have to be designed such as to work at 4.2 K without freezing. The sample has to be kept warmer than the rest of the microscope to prevent the condensation of the rest gas on it. The microscope normally is mounted in a tiny vacuum chamber, which is evacuated before the introduction to the cryostat. This pre-evacuation removes adsorbed water layers, which could mask the sample structure.
The control electronics must be modified, since normal electronic components do not work below 200 K. in addition the temperature gradients will cause thermovoltages at all connections of wires. This thermovoltage is added to the tunneling voltage, which could move the energy position of features like the superconducting band gap.

## 4.10.9. Related Techniques

There are a variety of related techniques to the *STM*. We will discuss a few selected techniques operating an *STM* in an unusual environment or an unusual way. We will begin our discussion with the oldest member, the Scanning Tunneling

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

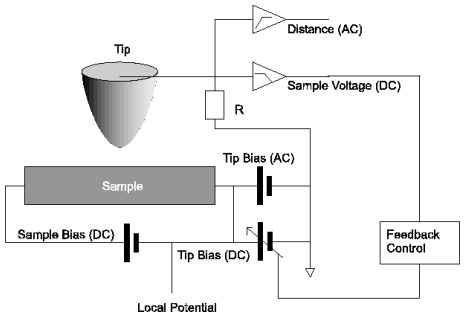



Abbildung 4.285.: Experimental setup for the Scanning Tunneling Potentiometry. A voltage is applied along the sample surface by the voltage source (Sample Bias). The distance between the tip and the sample is measured with a small AC voltage. The DC-current is nulled by the feedback loop controlling the DC-portion of the tip bias voltage. The voltage at the tip bias (DC)-voltage source is equal to the local potential on the sample.

Potentiometer.

#### 4.10.9.1. Scanning Tunneling Potentiometry

On samples with a gradient of the electrical potential along the surface it is important to know the potential as a function of position. The surface of an integrated chip with its diodes and transistors is a typical example. It is of utmost importance for the designer of a chip that the potential gradients are not too steep. On exceeding a certain limit, avalanche breakdown can occur and destroy the respective junction. Local variations due to a imperfect processing of the chip might give high local gradients in certain junctions, even though the design would be adequate otherwise.

Figure 4.285 shows an experimental set-up[**?**]. The voltage drop between the two lateral contacts is 5 V. The sample surface consists basically of two ohmic leads connected by a pn-junction, where most of the voltage drop occurs. How does one to measure this voltage drop? Since the surface is not atomically flat and the junction covers an area of several 1000 $nm^2$ a constant height experiment, similar to those discussed in previous chapters for *graphite* is not possible. We have to measure two quantities simultaneously, the local potential and the surface topography. Muralt and Pohl[**?**] solved the problem by measuring an AC-voltage to control the $z$-position of the tip and the DC-Voltage to measure the local potential. Since the resistance of the local potential is high and since it should not be loaded to give an accurate *measurement*, they used two feedback loops. The first feedback loop controls the $z$-position, as in any other *STM*. An AC-voltage is applied between the sample and the tip. The AC-current on the tip is high pass filtered and demodulated by a lock-in-amplifier. The output of this amplifier is further processed by the normal *STM* feedback loop. The upper limit of this AC-voltage is given by the DC-Resistance $R_t$ of the tunnel junction and the stray capacitances $C_s$ of the system tip-sample including the connecting wires. For a successful operation the operating frequency should be smaller than the cutoff frequency of the tunnel

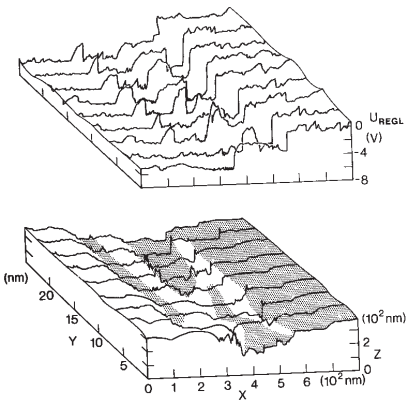



Abbildung 4.286.: Scanning Tunneling Potentiometry: the top part shows the local potential on the sample, a MIM structure. The bottom part of the figure is the corresponding topography. The image size is 800 nm x 25 nm. This figure is Copyright 1986 by International Business Machines Corporation; reprinted with permission.

junction and the connecting wires, given by

$$f_{3dB} = \frac{1}{2\pi R_t C_s} \tag{4.10.63}$$

Using typical values of 1 MΩ for the resistance of the tunnel junction $R_t$ and of 5 pF for the stray capacitance $C_s$ we get a cutoff frequency of 32 kHz. We have to bear in mind that at this frequency, the phase shift is 45° and very sensitive to the actual tunneling resistance (we can assume that the stray capacitance is a constant in this context). Hence a practical operating frequency might be 3 kHz. The lower limit of the operating frequency is given by the scanning speed and the bandwidth of the second feedback loop which we will discuss in a moment. If the typical distance between surface features is and the scanning speed $v_s$,then the operating frequency should be large compared with $v_s/a$. To maximize the scanning speed (and to minimize the waiting time for the operator!) the frequency of the AC-voltage of the Scanning Tunneling Potentiometer is set near the upper limit.

To measure the local surface potential with no loading, a potentiometer set-up is chosen. In this set-up, a second potential ("Tip Bias (DC)") is varied such as to null the DC-current. This potential is controlled by a feedback loop which keeps the DC-component of the tunnel current zero. The time constants of the two feedback loops have to be sufficiently different (one, better two orders of magnitude). Otherwise they would interfere with each other.

Figure 4.286 shows an example of a simultaneous *measurement* of the local potential and the surface topography. Theoretical considerations suggest that the ultimate resolution of the Scanning Tunneling Potentiometer would be the same as that of a *STM*. The resolution of the potential map in figure 4.286 is far worse. Assuming a minimal detectable potential change of 1 mV over 0.5 nm would give an electrical field of 20000 V/cm. Such fields only rarely exist in semiconductor devices. Beside the errors introduced by the dual feedback loop concept at least one other error source exists: thermoelectric potentials. These potentials are in

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

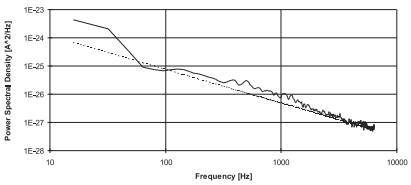



Abbildung 4.287.: The spectral noise density in a tunneling junction between the tip and the sample in an *STM*. The noise curve has been measured on a *GaAs*-sample by B. Koslowski. Plotted is the spectral power density as a function of the frequency. The dotted line is a $1/f^1.2$ power law. Used with permission.

series with the sought potential at the surface and are indistinguishable from it. High current densities in a pn-junction might cause considerable heating of the surface, which in turn could increase the error of the measured potential.

#### 4.10.9.2. Scanning Noise Microscopy and Potentiometry

We have seen in the previous chapters, that noise is present in any *STM measurement*. The origin of this noise is the tunneling current consisting of discrete charges, the electrons, and, additionally, the electronic components of the feedback system. The noise in the electronics can be minimized by a careful selection of the individual components and by the use of appropriate circuits. The noise on the tunneling current on the other hand is of far more fundamental origin and not yet fully understood. Möller *et al.*[**?**, **?**] have built and operated a scanning probe microscope based on noise measurements.
Figure 4.287 shows a *measurement* of the spectral noise density in the tunneling current between a tungsten tip and a *GaAs*-surface, as a function of the frequency. The spectral noise density and the frequency are plotted on a logarithmic scale. The electronic noise of the I/V-converter had been carefully minimized and the bandwidth optimized for this *measurement*. At high frequencies, above a few kHz and outside the range of figure 4.287, the spectral noise density is flat. The constant spectral noise density above a few kHz is characteristic of white noise. In all cases except for a vanishing average tunnel current, the spectral noise density increases with decreasing frequency below a few kHz. The exact nature of this noise is not yet known. It is present in all electronic devices, diodes and transistors, and it limits the performance of these devices. There are suggestions, that the $1/f$ noise (named for its dependence on the frequency $f$) might be caused in the tunneling current by adsorbate atoms or molecules passing through the tunnel junction. These adsorbate atoms or molecules will diffuse through any tunnel junction, whether there is a bias voltage or not. However at zero bias voltage, there is no electric field to modify the trajectory of the particles under the tip. The high field gradients between the *STM* tip and the sample will induce dipoles in atoms and molecules diffusing around near it. The induced dipole will then be attracted by the high field region and be present for longer times under the tip than in the case of no field. This mechanism is one possibility to generate a $1/f$ dependence of the low frequency spectral noise density. The $1/f$ noise component distinguishes the tunnel junction from an common resistor.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

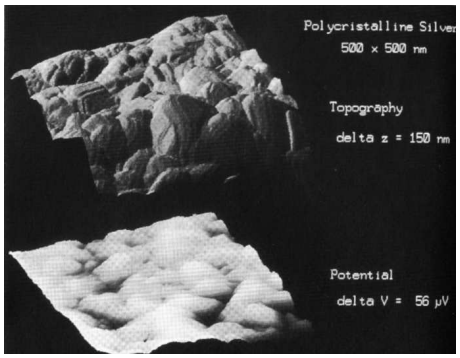



Abbildung 4.288.: Topography (upper part) and local potential (lower part) on a silver film. The scan size is 500 nm × 500 nm. The corrugation on the topography is 150 nm. The total potential variation is 56 $\mu$V. This image was provided by R. Möller and is used with permission.

Möller *et al.*[**?**] have found that the zero bias white noise in the *STM* tunnel junction varies with the spacing of the tip and sample surface, as varies the tunneling resistance. Using low noise electronics, they measured the noise density in a narrow frequency band and adjusted the relative spacing between tip and sample such that the total noise in this bandwidth was constant. The noise is also measured in a second narrow frequency band centered around a second frequency. If the noise is white then the ratio of the amplitudes of these two bands is equal to 1. However if the $1/f$-noise is present then the amplitude of the lower frequency band is larger than that of the higher frequency band. Using a feedback circuit a small bias voltage is applied to the sample to obtain white noise. The bias voltage is then exactly opposite to the potential at the sample surface. This method, called noise potentiometry, is capable to measure potentials down to the $\mu$V level, as can be seen in figure 4.288. The upper part of this figure shows the topography of a silver film consisting of a few grains and measured with noise microscopy. The lower part is a potential image of this same surface.
Scanning noise microscopy and potentiometry might be useful to image the activity of biological molecules. The transmission coefficient for electrons in all tunneling junctions depends critically on the exact arrangement of atoms and bonds present in the junction. A change of the conformation of a molecule hence changes the transmissivity of the tunnel junction. It is foreseen that displacements of a few picometers can show up in the noise spectrum.

#### 4.10.9.3. Ballistic Electron Emission Microscopy

On most samples, a *STM* provides the experimenter with information on the surface properties of the electron states near the Fermi Energy. The *STM* images are little affected by the underlying structure of the sample.
Figure 4.289 shows a sketch of the electron energy in a cross section of tip and sample. Here the sample is a Schottky diode. As we have seen in the theoretical section (figure 4.271), mainly electrons near the Fermi Energy of the negatively biased (higher energy) electrode tunnel through the tunnel barrier. Most of the electrons emerge from the barrier with the same energy they had before ente-

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

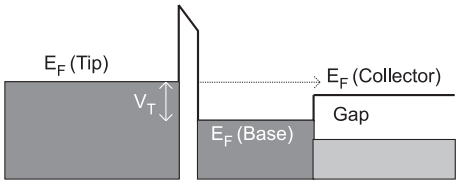



Abbildung 4.289.: Ballistic Electron Emission Microscopy. The electrons in the tip at the Fermi energy are injected in the base of a Schottky barrier. Some electrons move ballistically through the base and are collected by the collector. By varying $E_F$ of the collector with respect to $E_F$ of the tip transmission properties of the Schottky barrier can be measured.

ring the tunnel region. Only a minor fraction (1 in 1000 or less) loose energy in the process of tunneling. These inelastic tunneling processes are neglected in this chapter. The main fraction of the electrons entering the positively biased electrode have an excess energy equal to $eV_t$, the electron charge times the bias voltage. The electrons do not loose their excess energy instantly. They move ballistically through the sample with little interaction with the crystal structure. On the average electrons travel a distance called the "mean free pathïn a crystal between two collisions. This mean free path is a characteristic number of the sample depending on its crystal structure, its chemical composition, the abundance of crystal defects, and on the temperature. On the average electrons collide after one mean free path with an ion core of the sample and loose their excess kinetic energy.
We now assume that the sample is very thin, smaller than the mean free path of the electrons, and that we have an energy analyzer at the back. We could then determine the number of electrons as a function of their energy loss in the sample. If we further had a point source of these hot electrons, we could characterize a material's properties to within a few nanometers diameter on the sample surface. Bell and Kaiser[**?**] used the *STM* as their source of hot electrons. They used a Schottky diode with a distance of only a few 10 nanometers between the Schottky barrier and the surface. By biasing the collector electrode negatively, one is able to measure a spectrum of the hot electrons arriving at this electrode. By comparing this current with the injected tunneling current, one gets a measure of the mean free path between the tip and the back electrode. By scanning the tip over the sample surface, this mean free path can be mapped on the surface. The presence of defects like dislocations, grain boundaries or a different species of atoms is likely to reduce the mean free path. This change manifests itself as a reduction of the current measured at the back electrode.
This technique promises to give additional information on technologically important devices on semiconductor surfaces in a non-destructive way.

## 4.10.10. Serieschaltung von Tunneldioden

Wenn man, wie in Abbildung 4.290 einen Tunnelübergang mit einer dazwischengeschalteten Nanoinsel hat, treten neue physikalische Effekte auf. Da die Tunnelübergänge sehr kurz sind, kann man annehmen, dass in jedem Tunnelkontakt maximal ein Elektron ist[**?**]. Damit Einzelelektroneneffekte auftreten, müssen die

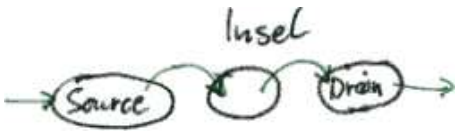



Abbildung 4.290.: Schematische Darstellung eines Tunnelübergangs mit einer dazwischen geschalteten nanoskopischen Insel[?].

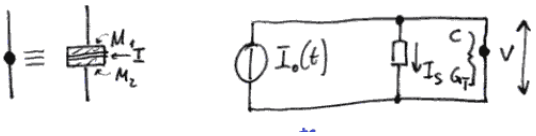



Abbildung 4.291.: Ersatzschaltschema für einen Tunnelübergang[?]. Nach Likharev ist nur der blaue Prozess erlaubt.

Tunnelwiderstände die Bedingung

$$R_T \gg R_K = \frac{h}{e^2} = 25.8k\Omega \tag{4.10.64}$$

erfüllen. Der Tunnelwiderstand ist eine praktisch definierte Grösse. Wenn $V$ die an einem Tunnelübergang anliegende Spannung und $I_T = \Gamma e$ der Tunnelstrom ist, dann ist

$$R_T = \frac{V}{I_T} = \frac{V}{e\Gamma} \tag{4.10.65}$$

Hier ist $\Gamma$ die Tunnelrate. Wenn man mit $\mathcal{T}$ die Transmissionswahrscheinlichkeit durch die Tunnelbarriere bezeichnet, dann gilt auch

$$\frac{1}{R_T} = 4\pi N \mathcal{T} \frac{1}{R_K} \tag{4.10.66}$$

Die Energieunschärfe herrührend von der durch Tunnelprozesse begrenzten Lebensdauer eines Elektron auf der Nanoinsel $\tau_r = R_T C$, ist

$$E_r \cdot \tau_r \geq h \tag{4.10.67}$$

wobei $C$ die Kapazität der Insel ist. Ist ein zusätzliches Elektron auf der Insel gespeichert, dann entspricht dies einer Energieänderung von

$$E_C = \frac{e^2}{2C} \tag{4.10.68}$$

Die Energie $E_C$ wird auch Coulomb-Energie genannt. Fordert man, dass $E_C \gg E_r$ ist, folgt daraus (4.10.64). Damit der Prozess beobachtbar ist, muss die Coulombenergie sehr viel grösser als die thermischen Energien sein.

$$E_C \gg E_{thermisch} = k_B T \tag{4.10.69}$$

Wenn (4.10.64) und (4.10.69) erfüllt sind, wird der Tunnelprozess und damit der Stromfluss durch die *Coulombblockade* gesteuert.
Die Abbildung 4.291 zeigt das für Tunnelübergänge verwendete Ersatzschaltbild. Neben dem Tunnelstrom $I_T$ und der Leitfähigkeit des Tunnelüberganges $G_T = 1/R_T$ ist die angelegte Spannung $V$ und die Kapazität $C$ von Bedeutung.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

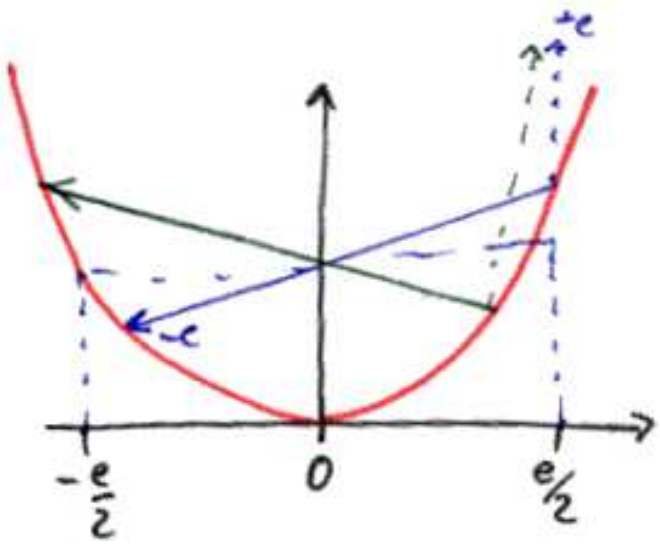



Abbildung 4.292.: Potentialverlauf auf der zentralen Insel im klassischen Bild.

Die Abbildung 4.292 zeigt eine graphische Darstellung der Energieverhältnisse bei der Coulomb-Blockade. Die Energie der Insel soll vor dem Tunneln eines Elektrons

$$E_i = \frac{Q^2}{2C} \tag{4.10.70}$$

sein. Nachdem eine Elementarladung $-e$ dazugekommen ist, ist die Energie

$$E_f = \frac{(Q-e)^2}{2C} \tag{4.10.71}$$

Der Prozess läuft nur dann von selber ab, wenn

$$0 < E_f - E_i = \frac{(Q-e)^2}{2C} - \frac{Q^2}{2C} = \frac{Q^2 - 2eQ + e^2 - Q^2}{2C} = -\frac{e\cdot(Q-e/2)}{C} \tag{4.10.72}$$

ist. Wenn $Q = 0$ ist, dann ist (4.10.72) nicht erfüllt. Wird ein nicht verschwindender Strom von aussen aufgeprägt, dann muss die Insel zuerst über Influenz auf $Q = e/2$ aufgeladen werden. Dann wechselt ein Elektron von der Insel und das Spiel beginnt von neuem.
Wenn $\Delta E = E_f - E_i$ positiv ist, wenn also kein Strom durch die Insel fliessen kann, dann gilt auch

$$-\frac{e}{2C} < V < \frac{e}{2C} \tag{4.10.73}$$

Die Spannung $V$ ist durch die Vorgeschichte gegeben

$$C\cdot V = \left(\int_{-\infty}^{t} i(t')dt'\right) \text{ modulo } |e| \tag{4.10.74}$$

Dies es Verhalten ist auch in der Abbildung 4.292 ersichtlich. Es wird angenommen, dass die Ausgangsladung $Q$ auf der Insel irgendwo im Bereich $-\frac{e}{2} < Q < \frac{e}{2}$ ist. Wann immer ein Elektron auf die Insel oder von der Insel tunnelt, erhöht sich die Gesamtenergie, da die Ladung sich nur in Schritten von $e$ ändern kann. Nur

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

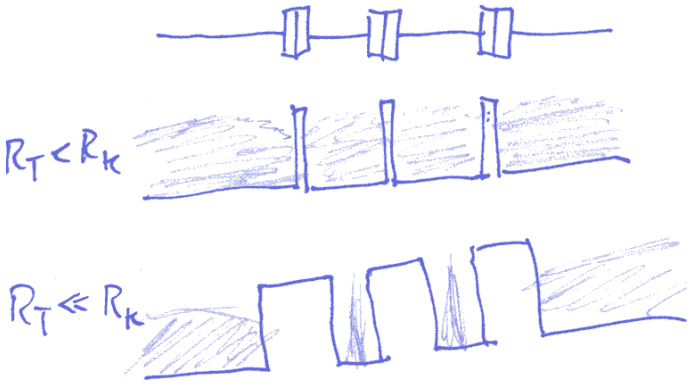



Abbildung 4.293.: Wahrscheinlichkeitsverteilung der Elektronen in Abhängigkeit von $R_T$

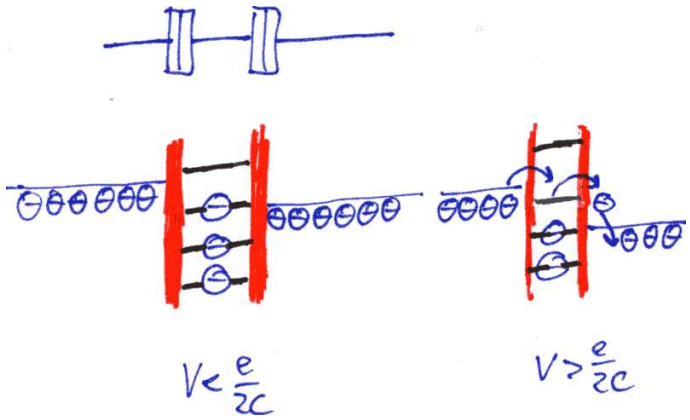



Abbildung 4.294.: Potentialbild einer Serieschaltung von zwei Inseln

wenn sich das Wellenpaket des Elektrons in der eigentlich verbotenen Zone rechts von $e/2$ befindet, kann mit durch den Tunnelprozess die Gesamtenergie gesenkt werden.
Wenn die Tunnelbarriere breit ist, das heisst wenn die Widerstandsbedingung (4.10.64) erfüllt ist, sind die Elektronen wie in der Abbildung 4.293, unten skizziert, lokalisiert. Die Kopplung zwischen den verschiedenen Leiterstücken ist schwach. Wird die Kopplung vergrössert, das heisst, werden die Tunnelbarrieren dünner, nähert sich die räumliche Verteilung der Elektronen immer mehr der von freien Elektronen an.
Die *Coulomb-Blockade* arbeitet qualitativ gesprochen wie folgt.

- Ein Elektron tunnelt auf die Insel.
- Die Energie der Insel wird so stark erhöht, dass keine weiteren Elektronen tunneln können.
- Das Elektron tunnelt aus der Insel, wobei die Richtung statistisch ist. Wenn der Tunnelwiderstand in eine Richtung kleiner ist, tritt der Tunnelprozess in die Richtung bevorzugter auf.

Dieses Verhalten wird sehr schön durch das Potentialbild in Abbildung 4.294 gezeigt. Da die Insel zwischen den zwei Tunnelkontakten sehr klein sein soll, gibt es

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

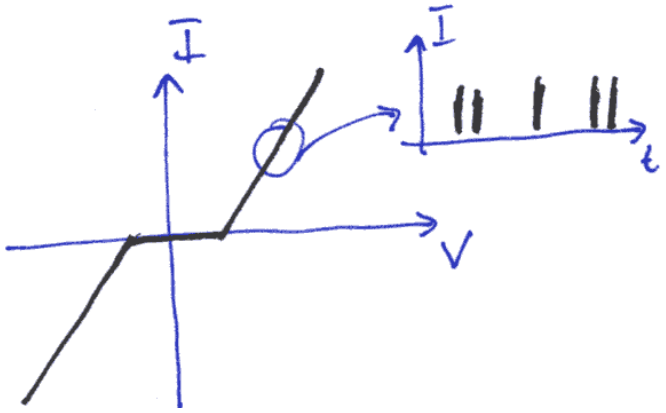



Abbildung 4.295.: Strom-Spannungscharakteristik für einen doppelten Tunnelkontakt

für die Elektronen nur diskrete Energieniveaus im Abstand der Coulombenergie $E_C$. Je kleiner die Insel, desto weiter sind die möglichen Elektronenzustände separiert. Wenn die angelegte Spannung kleiner als $e/(2C)$ ist, dann ist nach erreichen des Gleichgewichtszustandes ein unbesetztes Elektronenniveau der Insel über der linken *Fermi-Energie* und ein besetztes Niveau unter der *Fermi-Energie* des rechten Leiters. Für die Elektronen, die von links nach rechts gelangen wollen, besteht eine effektive Tunnelbarriere bestehend aus zweimal der Breite der ursprünglichen Barriere und der Breite der Insel.
Wird die angelegte Spannung über $e/2C$ angehoben, muss ein Elektronenzustand zwischen den Fermi-Energien der beiden Elektroden liegen. Elektronen können sich also von links nach rechts bewegen. Ist jedoch der Elektronenzustand unbesetzt, ändert sich die Energie der Insel: weitere Elektronen können erst nach einer gewissen Zeit wieder tunneln.
Dieses Verhalten bewirkt, dass bei der Temperatur T=0 die Leitfähigkeit null ist. Die Kennlinie wird wie in Abbildung 4.295 aussehen. Fliesst ein Strom, könnte dieser bei genügender Nachweisempfindlichkeit als eine Folge einzelner Elektronen aufgelöst werden. Im Mittel ist die Folgefrequenz

$$f_{SET} = I/e \tag{4.10.75}$$

## 4.10.11. Single Elektron Transistor

Wird nun bei einem doppelten Tunnelübergang die Insel, also die mittlere Elektrode über eine Spannung angesteuert, dann ergibt sich ein Potentialbild wie in der Abbildung 4.296. Die Anschlüsse dieses *Single-Electron-Transistor*s (*SET*)werden analog zu denen eines *FET* bezeichnet: *Source*, *Drain* und *Gate*. Das linke Bild zeigt die Situation mit einer angelegten Spannung $|V| < e/(2C)$: es kann kein Strom fliessen. Durch das Anlegen einer Spannung werden die Energieniveaus der Insel so verschoben, dass ein Strom fliessen kann.
Abbildung 4.297 zeigt das gebräuchliche Ersatzschema eines SET. Eingezeichnet sind die relevanten Kapazitäten $C_1$, $C_2$ und $C_g$ und die Widerstände $R_1$ und $R_2$. Eine Analyse der Gleichungen[?] zeigt, dass der SET in den in der Abbildung 4.298 eingeschlossenen Flächen nicht leitet.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

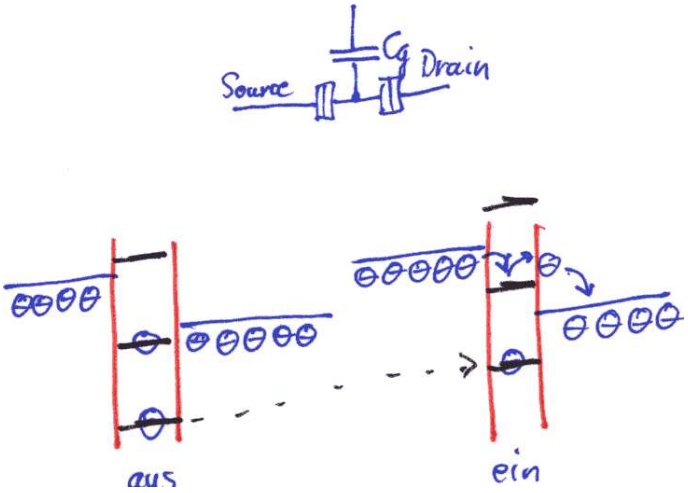



Abbildung 4.296.: Potentialbild eines Single-Elektron-Transistors.

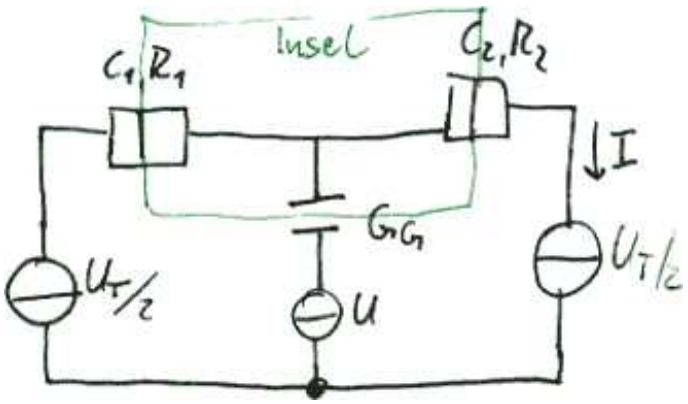



Abbildung 4.297.: Ersatzschaltschema für den *Single Electron Transistor*.

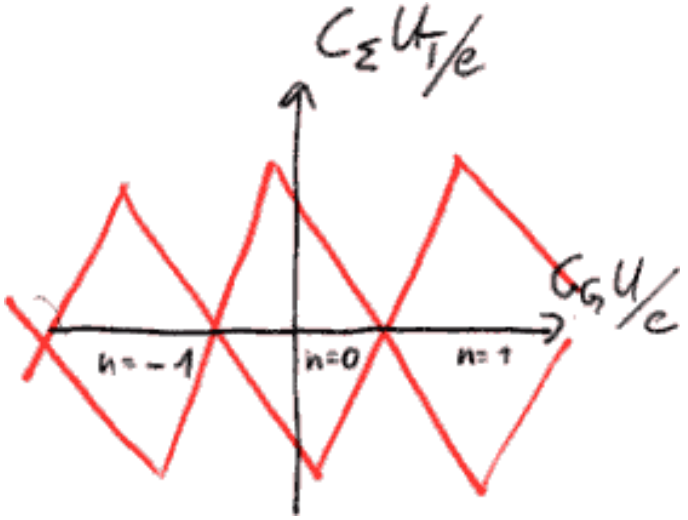



Abbildung 4.298.: Diagramm der Blockade in einem *Single Electron Transistor*.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Das Charge-Konsortium[26] stellt auf seiner Website mehrere Applets zur Verfügung, mit denen das Verhalten von SET's erforscht werden kann.

[26] http://qt.tn.tudelft.nl/CHARGE/

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

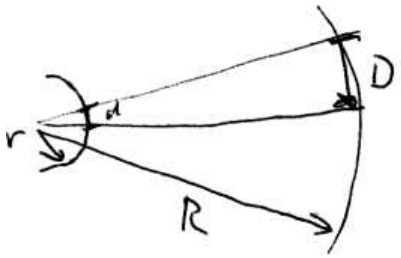



Abbildung 4.299.: Geometrie der Abbildung in einem Feldionenmikroskop

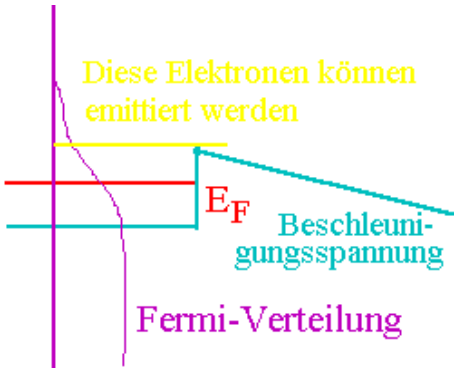



Abbildung 4.300.: Feldemissionsprozess

## 4.10.12. Feldemissionsmikroskopie, Feldionenmikroskopie

Bei einem Feldemissionsmikroskop wie auch bei einem Feldionenmikroskop werden Teilchen von einer stark gekrümmten Oberfläche radial emittiert. Die geometrische Situation ist in der Abbildung 4.299 gezeigt. Wenn die Spitze den Krümmungsradius $r$ und der Beobachtungsschirm den Radius $R$ haben, dann ist die Vergrösserung nach dem Strahlensatz durch

$$A = \frac{D}{d} = \frac{R}{r} \tag{4.10.76}$$

Wenn wir für den Beobachtungsschirm einen Radius von $R = 0.1m = 10^{-1}m$ annehmen und für die Spitze einen Krümmungsradius von $r = 100nm = 10^{-7}m$, dann ist die Vergrösserung $A = 10^6$. Zwei Kohlenstoffatome in einem Graphitgitter mit einem Abstand von $0.14nm$ würden in einem Abstand von $0.14mm$ abgebildet. Kann der Krümmungsradius der Spitze auf $10nm$ gesenkt werden, dann würden die zwei Kohlenstoffatome im Abstand von $1.4mm$ abgebildet.
Bei der Diskussion der Abbildung in einem Feldemissionsmikroskop oder in einem Feldionenmikroskop ist nicht berücksichtigt worden, dass die lokale Topographie den Austrittsort der Elektronen oder den Ort der Feldemission beeinflussen kann. Diese Topographie der *Potentiallinien* in der Nähe der Spitze mitteln sich jedoch über wenige nm aus, so dass die Trajektorie und damit die Abbildung nicht beeinflusst wird.

### 4.10.12.1. Feldemissionsmikroskop

Wie in der Abbildung 4.299 diskutiert, bewegen sich die Elektronen auf radialen Trajektorien von der Spitze zum Schirm. Das elektrische Feld in einem Feldemis-

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

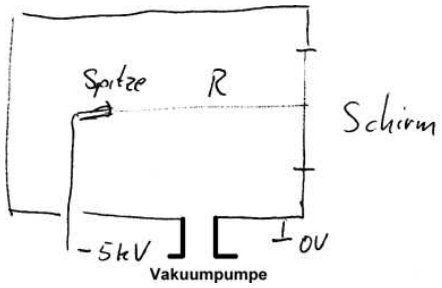



Abbildung 4.301.: Prinzipieller Aufbau eines Felemissionsmikroskopes

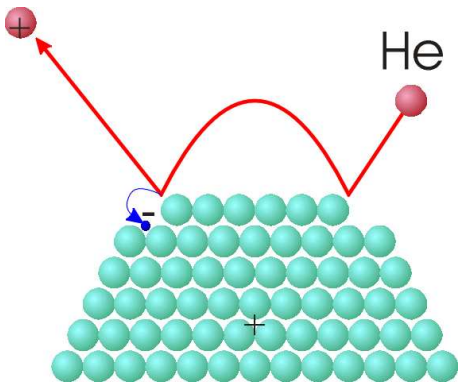



Abbildung 4.302.: Feldionisation an der Spitze eines Feldionenmikroskopes

sionsmikroskop entspricht dem eines Kugelkondensators. Die im obigen Abschnitt angesprochene Topographie der Emitterspitze bewirkt eine tangentiale Geschwindigkeit der Elektronen. Diese Tangentialgeschwindigkeit der Elektronen entspricht einer Energie $\approx 0.1$ eV. Diese muss mit der gesamten kinetischen Energie von 5 keV verglichen werden, ist also vernachlässigbar.
Die Potentialverhältnisse beim Emissionsprozess von Elektronen werden in der Abbildung 4.300 gezeigt. Die Berechnung kann mit den in den Abschnitten zur Theorie des Tunnelmikroskopes angegebenen Gleichung (4.10.19) für den Feldemissionsstrom berechnet werden.
Die Abbildung 4.301 zeigt den prinzipiellen Aufbau eines Feldemissionsmikroskopes. Die Emitterspitze ist in einem Gefäss mit einem Hochvakuum- oder Ultrahochvakuumdruck eingeschlossen. Der Phosphorschirm rechts wandelt die Elektronen entweder direkt oder mit einer vorgeschalteten Mikrokanalplatte in für das Auge sichtbare Bilder um.
Durch die geringe Masse der Elektronen ist deren Brownsche Bewegung so gross, dass eine Abbildung einzelner Atome nicht möglich ist. Deshalb wurde durch Müller und Tsong[**?**] das Feldionenmikroskop entwickelt.

### 4.10.12.2. Feld-Ionenmikroskopie (FIM)

Beim Feldionenmikroskop wird eine positive Spannung an die Spitze angelegt. Ein Abbildungsgas, in der Regel $He$, $H_2$ oder $Ne$ wird mit niedrigem Druck zur Probenkammer hinzugegeben. Diese ist ähnlich aufgebaut wie die Kammer der Feldemissionsmikroskope (siehe Abbildung 4.301). Die *Auflösung* bei dieser Art Mikroskopie ist $0.1nm$. Sie ist besser als bei der Flemissionsmikroskopie, da die

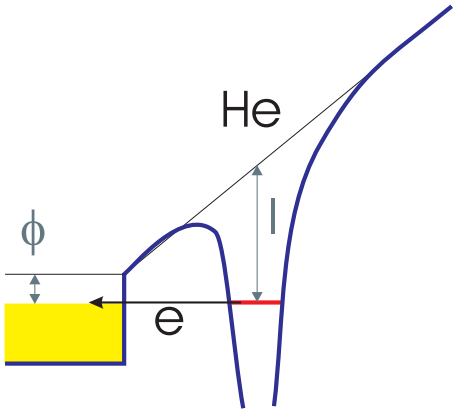



Abbildung 4.303.: Potentialverlauf bei der Feldionenmikroskopie

Masse der abbildenden Teilchen sehr viel grösser ist, die Brownsche Bewegung deshalb entsprechend geringer ist.
Die Abbildung 4.303 zeigt den Potentialverlauf an der Spitze eines Feldionenmikroskopes. Wenn ein Atom sich der Oberfläche nähert, verliert es durch Stösse kinetische Energie. Durch die Polarisierbarkeit wird das Atom an der Spitzenoberfläche gehalten. Zwischen dem Atom und der Spitze bildet sich eine Tunnelbarriere. Die Tunnelwahrscheinlichkeit ist desto grösser, je näher das Atom an der Probenoberfläche ist.
Wenn die *Austrittsarbeit* der Oberfläche $\phi$ und die Ionisationsenergie $I$ des obersten Energieniveaus ist, dann ist bei einer Feldstärke $E$ an der Spitze die Ionisation durch den Tunneleffekt möglich, wenn

$$x_c = \frac{I - \phi}{e \cdot E} \tag{4.10.77}$$

Dann liegt das oberste Energieniveau des Atoms über der *Fermi-Energie* $E_F$. Für Helium liegt das oberste besetzte Niveau , das $He_I$-Niveau, bei 24.5 eV. Mit der typischen *Austrittsarbeit* von Wolfram von $\phi = 4.5$ eV und einem Feld von $50V/nm$ ergibt Gleichung (4.10.77) für die Ionisationsdistanz den Wert $x_c = 0.4nm$. Ein Heliumatom aus dem Füllgas der Probenkammer, das durch seine zufällige Bewegung in die Nähe der Spitze kommt, wird durch den Feldgradienten polarisiert. Der entstehende Dipol ist so gerichtet, dass die negative Ladung zur positiven Spitze hin zeigt. Das polarisierte Heliumatom wird deshalb von der positiv geladenen Feldionenmikroskopspitze angezogen und gewinnt an Geschwindigkeit. Wie Bild 4.302 zeigt, stösst das He-Atom mit der Spitze. Durch den teilweise inelastischen Stoss verliert es an Energie. Es wird jedoch, mit verminderter Geschwindigkeit von der Spitze zurückreflektiert. Die Bewegung folgt den Feldlinien des elektrischen Feldes. Aus der Elektrodynamik ist bekannt, dass sich die Feldlinien des elektrischen Feldes an Orten mit kleinen Krümmungsradien konzentrieren. Dieser Effekt, der auch bei Blitzableitern ausgenutzt wird, bewirkt, dass an Ecken und Kanten die Feldstärke besonders hoch ist. Deshalb sind Ecken und kanten die Orte, Die Feldionisation geschieht bevorzugt an Orten hoher Feldstärke (Ecken, Kanten). Das heisst, dass der nächste und die folgenden Stösse des polarisierten He-Atoms bevorzugt an den Kanten und Ecken geschehen. Nach wenigen Annäherungsbewegungen bleibt das Atom solange an der Oberfläche der Spitze, dass das Valenzelektron (das elek-

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

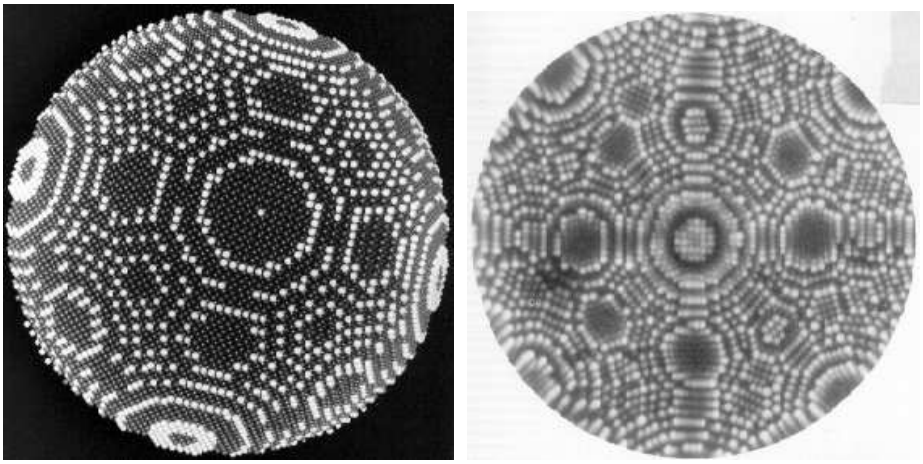


Abbildung 4.304.: Vergleich eines feldionenmikroskopischen Bildes mit einem Kugelmodell. Links ist das Kugelmodell für eine $Ni_4Mo$-Verbindung gezeigt. Rechts ist das entsprechende feldionenmikroskopische Bild.

trische Feld bricht die Symmetrie, so dass die Energieniveaux aufspalten) vom Restion getrennt wird. Sobald das He-Atom ionisiert ist, fliegt das Ion entlang der Feldlinien von der Spitze weg und wird auf dem Schirm detektiert.
Abbildung 4.304 zeigt auf der linken Seite Kugelkalottenmodell, bei dem die Randatome eingefärbt sind und ein ein gemessenes Feldionenmikroskopbild[27].
Damit die Abbildung gelingt, ist es notwendig, dass sowohl die Spitze wie auch die Atome des Füllgases gekühlt werden. Als Kühlmittel werden typischerweise flüssiger Stickstoff ($LN_2$) oder flüssiges Helium ($He$) verwendet. Damit werden die Tangentialgeschwindigkeiten der He-Ionen klein gehalten.
Die Feldionenmikroskopie war die erste experimentelle Methode, die individuelle Atome abbilden konnte. Da die Spitzen leitfähig sein müssen, ist die Anwendung dieser höchstauflösenden mikroskopischen Methode auf Metalle oder leitfähige Materialien beschränkt.
Erhöht man die an der Spitze angelegte Spannung zu stark, dann ist die Polarisation der Atome an ihrer Oberfläche so gross, dass auch Atome der Spitze im eigenen Feld ionisiert werden. Dieser Effekt wird *Felddesorption* genannt. Bei erhöhten Feldern wird $x_c$ kleiner. Wenn $x_c$ in den Bereich von $0.1nm$ kommt, werden auch Atome in der Spitzenoberfläche ionisiert.
Spitzen für die Feldionenmikroskopie werden analog zu den Spitzen für die Rastertunnelmikroskopie präpariert (Siehe den Abschnitt 4.10.4.8). Zusätzlich kann der Effekt der Felddesorption zur Formung der Spitze und zur Reduktion der Spitzengrösse verwendet werden. Durch Adsorption aus der Gasphase kann die Spitze auch gezielt vergrössert werden.
Damit ein Material für eine *feldionenmikroskopische Abbildung* tauglich ist, muss die Felddesorptionsschwelle für die Spitze höher liegen als Feldionisationsschwelle für das Füllgas. Typische in der Feldionenmikroskopie verwendete Materialien sind $W$, $Re$, $Mo$, $Fe$ und $Cu$. Dazu kommen Mischkristalle und andere, hier nicht genannte Materialien.

[27] http://www.ornl.gov/ORNLReview/rev28-4/text/contents.htm

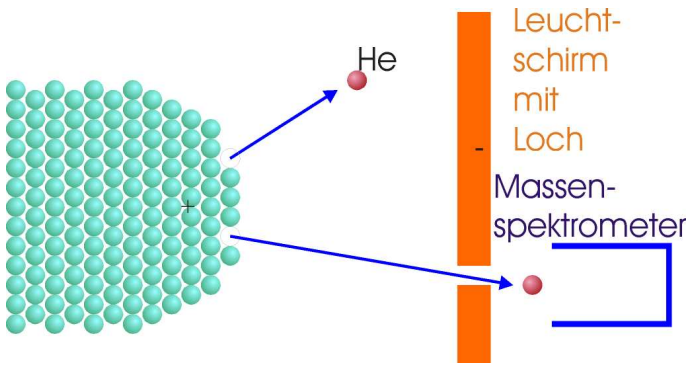



Abbildung 4.305.: Prinzipbild der Atom-Probe- Feldionenmikroskopie

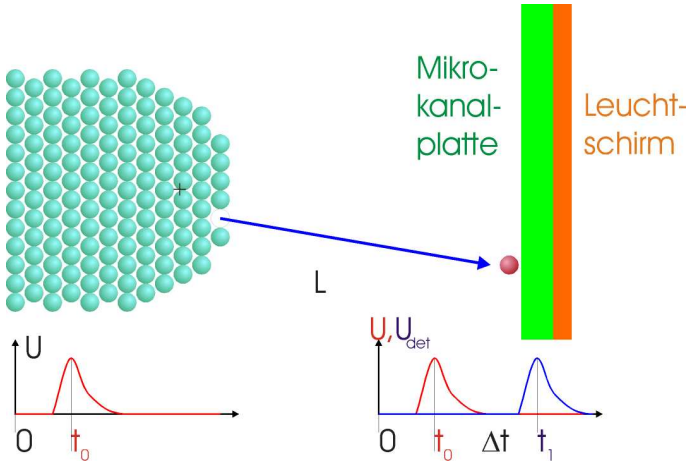



Abbildung 4.306.: Messung der Atommasse über die Laufzeit der felddesorbierten Atome.

Die Abbildung der Geometrie der Probe steht, zumindest bei reinen Oberflächen, heute nicht mehr im Vordergrund. Die Beobachtung der Bewegung einzelner Atome auf Terrassen von 20 bis 30 Substratatomen ermöglicht eine Bestimmung der Sprungwahrscheinlichkeit. Aus dieser lässt sich der *Diffusionskoeffizient* für einzelne, markierte Atome bestimmen. Wenn der Strom der ionisierten Bildgasatome nicht auf einem Fluoreszenzschirm mit oder ohne vorgeschaltete Mikrokanalplatte, sondern mit einem, die Fläche einer Atomposition abdeckenden *Faradaybecher* detektiert wird, kann man über die Stromschwankungen auf den Diffusionskoeffizienten rückschliessen.

### 4.10.12.3. Atom-Probe Feldionenmikroskopie

Die *Atom-Probe Feldionenmikroskopie* (Abbildung 4.305) ist eine Weiterentwicklung der Feldionenmikroskopie. Die Feldionenmikroskopie erlaubt eine Abbildung der Position von an Kanten oder Ecken sitzenden Atomen. dabei ist jedoch die Atomsorte nicht feststellbar. Die einfachste Art der Atom-Probe-Feldionenmikroskopie beruht darauf, dass bei der Felddesorption ein Atom aus der Probenoberfläche entfernt wird. Die kinetische Energie dieser Ionen hängt von ihrer Masse ab. In einem Massenspektrometer kann, da die Ionen einfach geladen sind, die Masse bestimmt werden.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

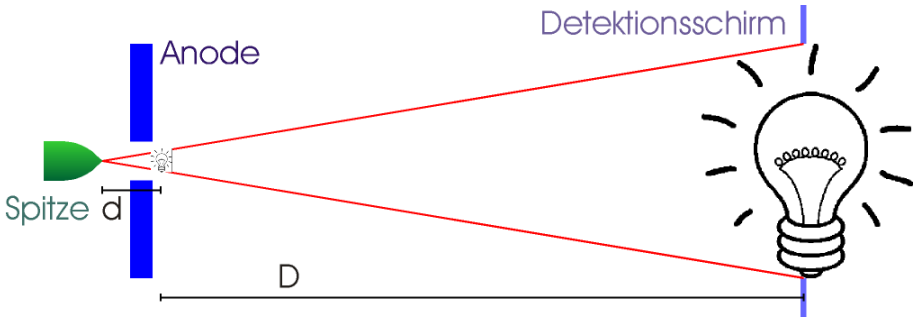



Abbildung 4.307.: Prinzip des Projektionselektronenmikroskops

Wenn die Felddesorption mit Spannungspulsen durchgeführt wird, kann über die Flugzeit der Ionen wie in der Abbildung 4.306 gezeigt die Masse bestimmt werden. Zur Zeit $t_0$ soll ein Spannungspuls an die Spitze angelegt werden. Die Ionen sollen mit der Beschleunigungsspannung $U$ beschleunigt werden. Da die Potentialgradienten des elektrischen Potentials nur in der Nähe der Spitze wesentlich von null verschieden sind, kann man das beschleunigende Feld aus der angelegten Desorptionsspannung berechnen. Zwischen der als konstant angenommenen Geschwindigkeit $v$, der Masse des Ions $m_I$ und der Beschleunigungsspannung gilt die folgende Beziehung.

$$\frac{1}{2}m_I v^2 = eU \tag{4.10.78}$$

Mit der Flugstrecke $L$, der Distanz zwischen der Spitze und dem Detektor kann die Flugzeit $\Delta t$ gefunden werden.

$$\Delta t = \frac{L}{v} = L\sqrt{\frac{m_I}{2eU}} \tag{4.10.79}$$

Damit ist die Ionenmasse

$$m_I = \frac{2eU}{v^2} = \frac{2eU\Delta t^2}{L^2} \tag{4.10.80}$$

Unter der Voraussetzung, dass jedes einzelne Ion detektiert werden kann, kann mit dieser Technik die Oberfläche der Spitze Atom für Atom abgetragen werden. Indem die Spannung des Desorptionspulses so klein gewählt wird, dass im Mittel wesentlich weniger als ein Atom desorbiert wird und indem man nach jedem Desorptionspuls eine Abbildung der Oberfläche aufnimmt, kann man die Zusammensetzung einer Spitze dreidimensional aufgelöst messen.

## 4.10.13. Projektionselektronenmikroskopie

Wenn man ein Feldemissionsmikroskop als punktförmige Quelle von Elektronen betrachtet kommt man zum Konzept eines Projektionselektronenmikroskopes[?]. Es ist möglich, eine Feldemissionsspitze so zu präparieren, dass sie in einem einzelnen Atom endet. Durch dieses einzelne Atom können Ströme von einigen $\mu A$ fliessen. Durch das kleine Volumen des Atoms, kann sich immer nur ein Elektron in diesem Gebiet aufhalten. Es ist zu erwarten und auch beobachtet worden, dass Elektronen, die aus einem einzelnen Atom emittiert werden, aussergewöhnliche Kohärenzeigenschaften haben.

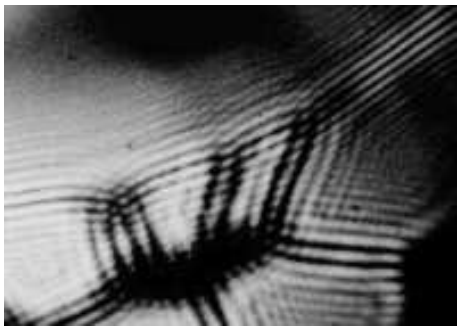


Abbildung 4.308.: Hologramm von Kohlenstofffasern

Abbildung 4.307 zeigt das Prinzip eines Projektionsmikroskopes. Eine feine Spitze wird gegenüber einer Anode mit einem Mikrometer grossen Loch positioniert. Die Spannungsdifferenz zwischen der Spitze und der Anode bestimmt die kinetische Energie der Elektronen. Typischerweise werden Spannungen von einigen 10 Volt angelegt. Dadurch ist die kinetische Energie der Elektronen so gering, dass die Wechselwirkung mit Kohlenstoff zu einem deutlich sichtbaren Kontrast führt[**?**]. Die Vergrösserung eines Projektionselektronenmikroskopes ist durch das Verhältnis der Distanz von der Spitze zum Beobachtungsschirm zur Distanz von der Spitze zum Objekt gegeben.

$$a = \frac{d+D}{d} = 1 + \frac{D}{d} \tag{4.10.81}$$

Wenn das Objekt einen oder einige Mikrometer von der Spitze entfernt ist und die Distanz zum Schirm 10 cm beträgt, wäre die Vergrösserung zwischen $10^4$ und $10^5$. Die Abbildung mit einem Projektionselektronenmikroskop ist prinzipbedingt verzerrungsfrei und schädigt die Proben kaum.
Abbildung 4.308 zeigt ein Hologramm von Kohlenstofffasern. Die Tatsache, dass Interferenzerscheinungen sichtbar sind, deutet auf die für Elektronen aussergewöhnlich hohe Kohärenzlänge hin.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

## 4.10.14. Rasterelektronenmikroskopie

Ein Elektronenmikroskop ist prinzipiell wie ein klassisches Mikroskop aufgebaut. Linsen bilden eine Quelle auf die Probe und die Probe auf den Bildschirm ab. In dieser Vorlesung sollen nun die Prinzipien der Rasterelektronenmikroskopie besprochen werden. Eine Rasterelektronenmikroskop ist ähnlich aufgebaut wie ein konfokales Mikroskop. Der Elektronenstrahl wird über die Probe gerastert. Anders als im konfokalen Mikroskop wird jedoch die gestreuten Elektronen ohne Rückwirkung gesammelt.

### 4.10.14.1. Linsen für Elektronen

**4.10.14.1.1. Elektrostatische Linsen** Ein Elektron, das die Potentialdifferenz $\Delta U$ durchläuft ändert seine kinetische Energie um.

$$e\Delta U = \frac{1}{2}mv_B^2 - \frac{1}{2}mv_A^2 \tag{4.10.82}$$

Ist $v_A = 0$ so gilt

$$v = \sqrt{\frac{2eU}{m}} = \sqrt{\frac{2e}{m}}\sqrt{U} \tag{4.10.83}$$

Wir können dieses Resultat mit dem für Licht vergleichen.
Dort ist

$$c^{'} = {}^{c_0}\!/_{n} \tag{4.10.84}$$

oder

$$n = \frac{c_0}{c'} \tag{4.10.85}$$

Dabei ist $c_0$ die Vakuumlichtgeschwindigkeit und $c'$ die Lichtgeschwindigkeit im Medium.
Für Elektronen gilt:

$$n \propto v \propto \sqrt{\frac{2e}{m}}\sqrt{U} \tag{4.10.86}$$

Nach Abb. 4.309 gilt, da

$$\frac{1}{2}m({v'_z}^2 - v_z^2) = eU \tag{4.10.87}$$

d.h. im Gebiet wo das Elektron schneller ist wird es zur Grenzflächennormale gebeugt. Das Brechungsgesetz ist dann

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{n_2}{n_1} = \frac{v_2}{v_1} = \sqrt{\frac{U_2}{U_1}} \tag{4.10.88}$$

Eine elektrostatische Linse wird nach dem in Abb. 4.310 gezeigten Schema aufgebaut. Die Äquipotentialflächen bilden eine linsenförmige Oberfläche. Die Trajektorien der Elektronen sind in Abb. 4.311 gezeigt. Bei jeder konkaven Äquipotentialfläche wird das Elektron weg von der Achse abgelenkt, bei jeder konvexen zur Achse hin.

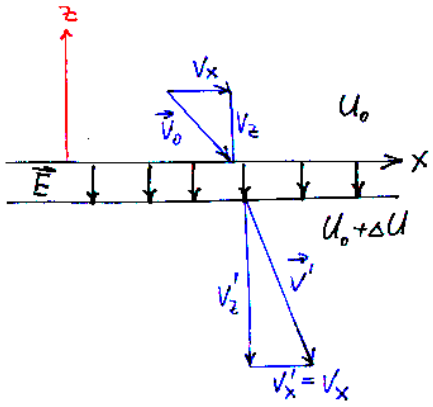



Abbildung 4.309.: Durchgang von Elektronen durch eine elektrostatische Potentialdifferenz

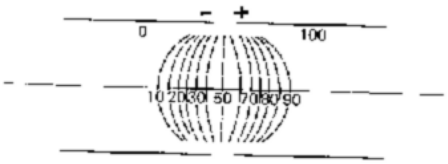



Abbildung 4.310.: Aufbau einer elektrostatischen Linse

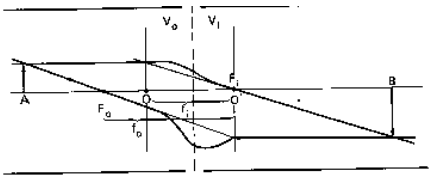



Abbildung 4.311.: Schematischer Aufbau einer elektrostatischen Zylinderlinse

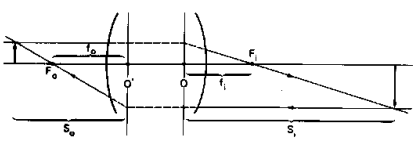


Abbildung 4.312.: Hauptebenen einer elektrostatischen Linse

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

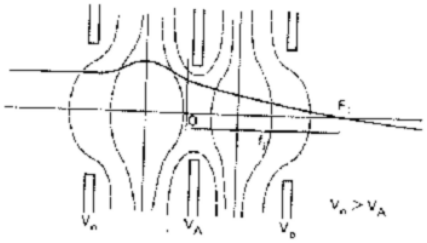


Abbildung 4.313.: Elektrostatische Einzellinse

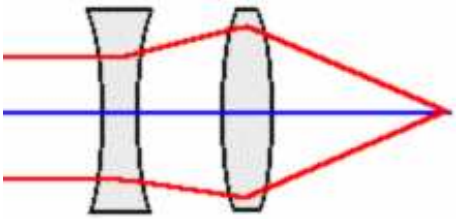


Abbildung 4.314.: Zur Einzellinse äquivalenter Aufbau mit optischen Linsen

Wie bei optischen Linsen können Hauptebenen und Brennweiten bestimmt werden (siehe Abb. 4.312). Da die Geschwindigkeit von links nach rechts zunimmt ist die Linse nicht umkehrbar.
Es gilt: $\frac{f_0}{S_0} + \frac{f_i}{S_i} = 1$. Damit wird die Vergrösserung:

$$\frac{y_i}{Y_0} = -\frac{f_0}{f_i}\frac{S_i}{S_0} \tag{4.10.89}$$

und

$$\frac{f_i}{f_0} = -\sqrt{\frac{U_i}{U_0}} = -\sqrt{\frac{v_i}{v_0}} \tag{4.10.90}$$

Die in Abb. 4.312 dargestellte Linse ist nicht praktisch, da sie das Potential ändert. Üblicher ist die Einzellinse nach Abb. 4.313. Diese Linse ist symmetrisch. Sie hat fokussierende Wirkung und ist formal äquivalent zu zwei optischen Linsen mit gleichen Brennweiten, aber unterschiedlichen Krümmungen (Siehe Abb. 4.314.
Elektrostatische Linsen werden üblicherweise nicht in Rasterelektronenmikroskopen verwendet, da sie die Geschwindigkeit ändern und eine gringe Brechkraft haben. In modernen Instrumenten (später) werden sie jedoch benutzt, um die Elektronen vor der Wechselwirkung mit der Probe abzubremsen.

**4.10.14.1.2. Magnetische Linsen** Elektronen werden in einem Magnetfeld durch die Lorentzkraft abgelenkt (Abb. 4.315).

$$\boldsymbol{F} = -e\left(\vec{v} \times \boldsymbol{B}\right) \tag{4.10.91}$$

Elektronen die parallel zum Magnetfeld fliegen, werden nicht abgelenkt. Elektronen, deren Bahnen geneigt sind, haben eine Komponente $v_\perp$ senkrecht zum Ma-

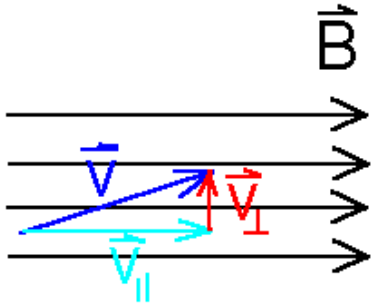



Abbildung 4.315.: Schema der Wirkungsweise der Lorentzkraft

gnetfeld. Deshalb gibt es eine Kraft

$$F_z = eBv_\perp = eBv\sin\alpha = m\frac{v_\perp^2}{r} \tag{4.10.92}$$

Diese Kraft ist senkrecht zu $v_\perp$. In der Projektion ergibt sich die Zyklotron-Kreisbewegung. $F_z$ ist auch eine Zentripetalkraft. Also ist

$$v_\perp = \frac{eBr}{m} \tag{4.10.93}$$

Die Umlaufzeit ist

$$T_{uml.} = \frac{2\pi r}{v_\perp} = \frac{2\pi m}{eB} \tag{4.10.94}$$

Wir erhalten das bemerkenswerte Resultat, dass die Umlaufzeit eines Elektrons unabhängig von seiner Anfangsgeschwindigkeit und von seiner Bahnneigung ist. Wenn wir ein Bündel Elektronen, das aus einem Punkt emittiert wird betrachten, dann sind alle nach $L = nvT_{uml}$ $n \in N$ wieder fokussiert d.h. wir können sagen, dass diese Linse die Brennweite

$$f = \frac{L}{4} = nvT_{uml} = \frac{2\pi nvm}{eB} \qquad n \in N \tag{4.10.95}$$

hat. Diese Linse hat einen Abbildungsmassstab von 1:1, (deshalb ist auch $f = L/4$). Der Krümmungsradius der Bahn ist dabei

$$\begin{aligned} r &= \frac{mv_\perp}{eB} = \frac{mv\sin\alpha}{eB} \\ &= \frac{1}{B}\sqrt{\frac{2eU\,m^2}{me^2}}\sin\alpha \\ &= \frac{1}{B}\sqrt{\frac{2Um}{e}}\sin\alpha \end{aligned} \tag{4.10.96}$$

wobei die folgenden Abkürzungen verwendet wurden:

$$\frac{1}{2}mv^2 = eU \tag{4.10.97}$$

$$v = \sqrt{\frac{2eU}{m}} \tag{4.10.98}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

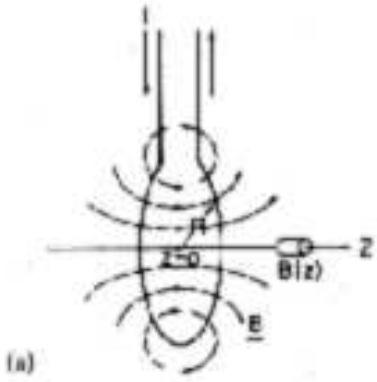



Abbildung 4.316.: Drahtschlaufe als Linse. Die magnetischen Feldlinien sind so, dass die Lorentzkraft in der Drahtschlaufe die Elektronen wie in einer Linse ablenkt.

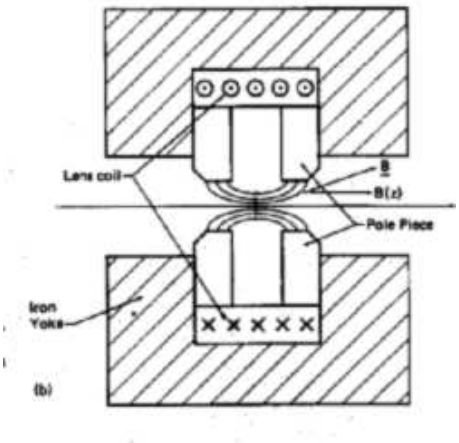



Abbildung 4.317.: Bild einer kurzen magnetischen Linse

Übliche magnetische Linsen verwenden inhomogene Magnetfelder Das einfachste Beispiel ist eine Drahtschlaufe (siehe Abb. 4.316).
Es gilt nach Biot-Savart

$$B(z) = \frac{\mu_0 I}{2R\left(1+\left(\frac{z}{R}\right)^2\right)^{3/2}} \tag{4.10.99}$$

Für eine Spule mit N Windungen ist

$$B\left(z\right) = \frac{N\mu_0 I}{2R\left(1+\left(\frac{z}{R}\right)^2\right)^{3/2}} \tag{4.10.100}$$

Übliche magnetische Linsen werden mit Eisenkernen aufgebaut. Damit kann die Stärke des Magnetfeldes um mehr als eine Grössenordnung gesteigert werden. Eine kurze Linse (Abb. 4.317) kann wie folgt berechnet werden:
Es gilt $\boldsymbol{F} = e\boldsymbol{v} \times \boldsymbol{B}$ und $\operatorname{div} \boldsymbol{B} = 0$.

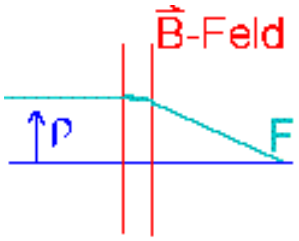



Abbildung 4.318.: Zylinderkoordinaten zur Berechnung einer magnetischen Linse

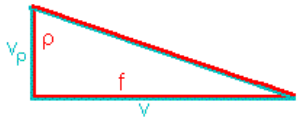



Abbildung 4.319.: Geometrische Zusammenhänge in magnetischen Linsen zwischen radialer Geschwindigkeit $v_\rho$ und der axialen Geschwindigkeit $v$ sowie dem Abstand von der optischen Achse $\rho$ und der Brennweite $f$.

In Zylinder-Koordinaten (Abb. 4.318) ist

$$\begin{aligned} \mathrm{div}\boldsymbol{B} &= \frac{1}{\rho}\frac{\partial}{\partial\rho}(\rho B\rho) + \frac{1}{\rho}\frac{\partial B\varphi}{\partial\varphi} + \frac{\partial B_z}{\partial z} \\ &= \frac{1}{\rho}\left(B_\rho + \rho\frac{\partial B_\rho}{\partial\rho}\right) + \frac{\partial B_z}{\partial z} \\ &= \left(\frac{B_\rho}{\rho} + \frac{\partial B_\rho}{\partial\rho}\right) + \frac{\partial B_z}{\partial z} \end{aligned} \tag{4.10.101}$$

Dabei ist $B\varphi = const$ da Zylindersymmetrie angenommen wurde´.
Für Felder, die sich räumlich nicht zu stark ändern, gilt

$$\frac{B_\rho}{\rho} = \frac{\partial B_\rho}{\partial\rho} \tag{4.10.102}$$

Dies ist die paraxiale Näherung.
Also gilt

$$\mathrm{div}\boldsymbol{B} = 0 = 2\frac{\partial B_\rho}{\partial\rho} + \frac{\partial B_z}{\partial z} \tag{4.10.103}$$

wieder mit der paraxialen Näherung gilt $v_z \approx v$

$$\left.\begin{aligned} \dot{v}_\varphi &= \tfrac{evB_\rho}{m} \\ \dot{v}_\rho &= \tfrac{ev_\varphi B_z}{m} \end{aligned}\right\} \Rightarrow \ddot{v}_\rho = -\frac{eB_z}{m}\frac{evB_\rho}{m} = \frac{-e^2B_\rho}{m^2}B_z v \tag{4.10.104}$$

Aus der Geometrie folgt:

$$\frac{v_\rho}{v} \approx \frac{\partial v_\rho}{\partial\rho} = \frac{v}{f} \tag{4.10.105}$$

und daraus

$$\frac{\ddot{v}_\varphi}{v} = \frac{\ddot{\rho}}{f} = \frac{\dot{v}_\rho}{f} = -\frac{e^2}{m^2}B_\rho B_z = \frac{\partial}{\partial\rho}v_\rho \tag{4.10.106}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Abgeleitet ergibt sich

$$\begin{aligned}\frac{\partial}{\partial\rho}\ddot{v}_\rho &= -e^2 v\left(B_z\frac{\partial B_\rho}{\partial\rho}+B_\rho\frac{\partial B_z}{\partial\rho}\right)\frac{1}{m^2}\\ &= \frac{-e^2 v}{m^2}B_z\frac{\partial B_\rho}{\partial\rho}\end{aligned} \tag{4.10.107}$$

mit

$$\frac{\partial B_z}{\partial\rho}\approx 0 \tag{4.10.108}$$

Verwenden wir die Beziehung $\dot{x}=f(z)\Rightarrow dx=f(z)\,dt=f(z)\,\frac{dz}{v}$ und $v=\frac{dz}{dt}$, so wird

$$\Rightarrow x=\frac{1}{v}\int f(z)dz \tag{4.10.109}$$

sowie mit $x=\frac{\partial\ddot{v}_\rho}{\partial\rho}$ wird

$$\frac{\partial\ddot{v}_\rho}{\partial\rho}=-\frac{1}{2}\frac{e^2}{m^2}\int B_z\frac{\partial B_z}{\partial z} \tag{4.10.110}$$

Mit div$\boldsymbol{B}=0$ folgt

$$\frac{\partial B_\rho}{\partial\rho}=-\frac{1}{2}\frac{\partial B_z}{\partial z} \tag{4.10.111}$$

und

$$B_z\frac{\partial B_z}{\partial z}=\frac{1}{2}\frac{\partial}{\partial z}\left(B_z^2\right) \tag{4.10.112}$$

Also ist

$$\frac{\partial\ddot{v}_\rho}{\partial\rho}=\frac{1}{4}\frac{e^2}{m^2}B_z^2 \tag{4.10.113}$$

Weiter gilt

$$\frac{\partial v_\rho}{\partial\rho}=\frac{1}{4}\frac{e^2}{m^2}\int_{-\infty}^{\infty}B_z^2dt=\frac{1}{4}\frac{e^2}{mv}\int_{-\infty}^{\infty}B_z^2dz \tag{4.10.114}$$

Mit den Beziehungen $\frac{1}{2}mv^2=eU$ und $\frac{\partial v_\rho}{\partial\rho}=\frac{v}{f}$ bekommt man

$$\frac{1}{f}=\frac{1}{8}\frac{e^2}{\frac{1}{2}m^2v^2}\int_{-\infty}^{\infty}B_z^2dz=\frac{1}{8}\frac{e}{mU}\int_{-\infty}^{\infty}B_z^2dz \tag{4.10.115}$$

Da die Brennweite von $B_z^2$ abhängt, gibt es nur fokussierende Linsen:

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

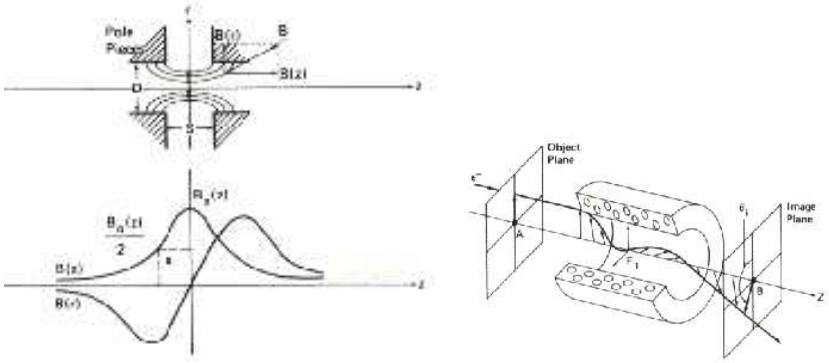



Abbildung 4.320.: Links die Feldverteilung $B(z)$ in axialer und $B(r)$ in transversaler Richtung in einer symmetrischen magnetischen Linse mit der Öffnung $D$ und dem Gap $S$). Rechts ist das Schema der Bildformung in einer magnetischen Linse gezeigt.

**4.10.14.1.3. Wie wird die Bewegung der Elektronen beschrieben?** Mit einer ähnlichen Umformung wie im vorherigen Absatz erhält man:

$$F_\rho = -\left(\frac{e^2}{4m}\right) B_z^2(z)\,\rho \tag{4.10.116}$$

Dies ist die Zentripetalkraft. Deshalb gilt

$$F_\rho = m\ddot{\rho} = -\frac{e^2}{4m} B^2(z)\rho \tag{4.10.117}$$

Mit der Elektronengeschwindigkeit $\frac{dz}{dt} = v = \left(\frac{2eU}{m}\right)^{\frac{1}{2}}$ wird

$$\frac{d^2\rho}{dt^2} + \frac{e^2}{4m^2} B_\rho^2 = O = \frac{\partial^2\rho}{\partial z^2} + \frac{e}{8mU} B_\rho^2 = 0 \tag{4.10.118}$$

d.h. wir erhalten eine Bewegungsgleichung für die Rotation der Elektronenbahn. Bei grösseren $\frac{B^2}{U}$ wird die Brennweite kleiner.
Für die Bahndrehung gilt:

$$\frac{d\rho}{dz} + \left(\frac{e}{8mU}\right)^{\frac{1}{2}} B = O \tag{4.10.119}$$

Aus der Beziehung

$$\rho(z_2) - \rho(z_1) = -\int_{z_1}^{z_2} \left(\frac{e}{8mU}\right)^{\frac{1}{2}} B(z)\,dz \tag{4.10.120}$$

ersieht man, dass die Bilddrehung von der Richtung von $\boldsymbol{B}$ abhängt. Das heisst durch alternierende Richtungen von $\boldsymbol{B}$ kann sie kompensiert werden.
Wir berechnen nun die Trajektoren für

$$B(z) = \frac{B_0}{1 + \left(\frac{z}{a}\right)^2} \tag{4.10.121}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

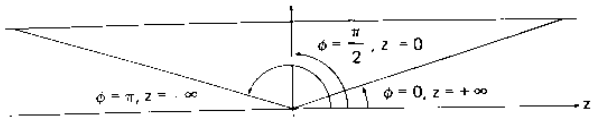



Abbildung 4.321.: Beziehung zwischen $\phi$ und $z$

Dabei ist $a$ die FWHM, das heisst die volle Breite bei halber Höhe, eine Funktion, die typisch für magnetische Linsen mit kleinem Luftspalt ist.

$$\frac{d^2\rho}{dz^2} + \frac{eB_0^2}{8mU} \frac{\rho}{\left(1 + \left(\frac{z}{a}\right)^2\right)^2} = 0 \tag{4.10.122}$$

mit $x = \frac{z}{a} = \cot\phi$ und $y = \frac{\rho}{a}$ wird

$$\frac{d^2y}{d\phi^2} + 2\cot\phi \frac{dy}{d\phi} + k^2 y = 0 \tag{4.10.123}$$

dabei ist $k^2 = \frac{eB_0^2}{8mU} a^2$ der Linsenparameter.
Eine mögliche Lösung ist $y = \frac{1}{w}\left(\sin w\phi / \sin\phi\right)$, wobei

$$w^2 = k^2 + 1 \tag{4.10.124}$$

ist. $\phi$ wandelt eine lineare Distanz, $z$, in einen Winkel im $(0 \leq \phi \leq \pi)$ um. Hier ist $y$ der skalierte Achsabstand. Wir können auch die Steigung berechnen.

$$y' = \frac{dy}{dz} = \frac{dy}{d\phi}\frac{d\phi}{dz} = \frac{1}{wa}(\sin w\phi - wsin\phi\cos w\phi) \tag{4.10.125}$$

Nun ist $y' = 0$ wenn $z' = +\infty$ ist. Dies ist äquivalent zu $\phi = 0$.Die Frage ist, wann $y' = 0$ ist wenn $\phi = \pi$ oder $z = -\infty$ ist. Wir erhalten

$$y' = -\frac{1}{wa}\sin w\pi = 0 \tag{4.10.126}$$

Daraus ergibt sich, dass $w = 1, 2, 3, \ldots$ ist. Was bedeutet nun $w$?

**w=1** heisst $k^2 = 0$ oder $B = 0$

**w=2** heisst $k^2 = 3$ eine Nullstelle, d.h. der Elektronenstrahl überquert $z$ einmal

**w=3** heisst $k^2 = 8$, also 2 Übergänge

Abb. 4.322 zeigt die möglichen Bahnen.
Für $k^2 < 3$ wirkt die Linse analog einer Glas-Sammellinse. Praktisch alle Magnetlinsen sind deshalb für $k^2 < 3$ konstruiert.
Was ist die Steigung?

$$y' = \frac{dy}{dx} = \frac{dy}{d\phi}\frac{d\phi}{dz} = \frac{1}{wa}\left(\sin w\phi - w\sin\phi\cos w\phi\right) \tag{4.10.127}$$

Wir erhalten $y' = 0$ für $z = +\infty\,(\phi = 0)$.
Wann ist für einen einfallender Strahl $y' = 0$ für $\phi = \pi\,(z = -\infty)$?

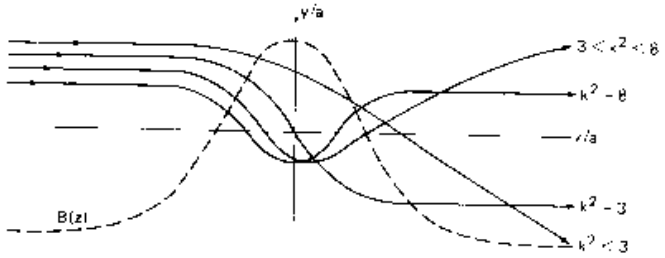



Abbildung 4.322.: Elektronenbahnen in einer Magnetlinse

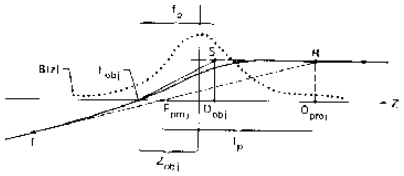



Abbildung 4.323.: Typische Elektronenbahnen in einer Magnetlinse mit einer Gauss-förmigen Magnetfeldverteilung

Antwort $y' = -\frac{1}{wa}\sin w\pi = 0 \Rightarrow \; w = 0, 1, 2, 3, 4$
In der Abbildung 4.323 gibt es zwei Fokuslängen

| | |
|---|---|
| Prinzipalebene $O_{proj}$<br>fp: Projektionsfokus | (gegeben durch RT) definiert<br>gebraucht für Vergrösserungslinsen |

**4.10.14.1.4. Kondensorlinsen** Wenn das Objekt im Magnetfeld plaziert wird, benötigt man andere Kardinalelemente. Ausgehend vom Brennpunkt $F_{obj}$ definiert man die Hauptebene $O_{obj}$ und $F_{obj}$ ist die Objektbrennweite $f_0$
Da $a$ als Achsabstand eines Strahls aufgefasst werden kann, ist

$$f_p = \frac{1}{y'(\infty)} = \frac{a}{r'(-\infty)} = \frac{aw}{\sin w\pi} = aw\csc(w\pi) \tag{4.10.128}$$

$f_p$ ist undefiniert für $w = 2, 3, \ldots$
$f_0$ ist durch $0 = y = \frac{1}{wa}\frac{\sin w\phi}{\sin\phi}$ d.h. $w = n, \; n \in N$ gegeben.
Die Lage bezüglich der Linsenmitte ist durch

$$\frac{z}{a} = \cot\phi = \cot\frac{n\pi}{w} \tag{4.10.129}$$

gegeben. Objekt- und Bildebene sind dann bei

$$z_{obj} = a\cot\frac{\pi}{w} \tag{4.10.130}$$

und

$$z_{im} = -a\cot\frac{\pi}{w} \tag{4.10.131}$$

Diese beiden Punkte heissen fokale Mittenpunkte. Bei $z_im$ erscheint das Beugungsmuster der Probe. Dies ist die Arbeitsebene, wenn man Elektronendiffraktion durchführen will.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

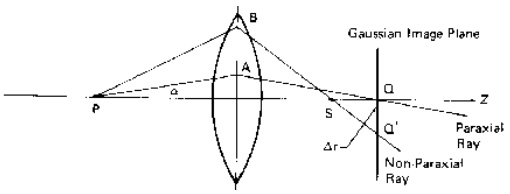



Abbildung 4.324.: Sphärische Aberration illustriert durch paraxiale Strahlen

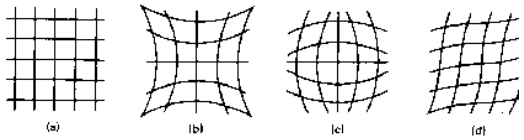



Abbildung 4.325.: Illustration der Kissen- und Trapezverzerrungen

Weiter gilt

$$\frac{1}{f_0} = -\frac{1}{y}\frac{dy}{dz} = -\frac{1}{a}\sin\frac{\pi}{w} \tag{4.10.132}$$

wenn $(n = 1)$ ist. Umgeformt ergibt sich

$$f_0 = -a\,\csc\frac{\pi}{w} = f_i \tag{4.10.133}$$

Wenn $l_i = s_i - f_i$ ist, wobei $s_i$ der Abstand Hauptebene - Bild sei, und $l_0 = s_0 - f_0$ ist, dann gilt $l_i l_0 = f_i f_0$ wie bei der optischen Linse. Die Vergrösserung wird dann

$$M = -\frac{f_0}{l_0} = -\frac{l_i}{f_i} \tag{4.10.134}$$

Dabei ist die Bilddrehung:

$$\Delta\theta = \frac{nk\pi}{(k^2+1)^{\frac{1}{2}}} = \frac{nk\pi}{w} \tag{4.10.135}$$

mit $n = 1$, $k^2 = 1$, $w^2 = 2$ folgt $\Delta\theta = \frac{\pi}{\sqrt{2}} \approx 127°$

#### 4.10.14.1.5. Linsenfehler

**Sphärische Aberration** Tritt für Strahlen mit $\sin\alpha \neq \alpha$ auf (Abb. 4.324

**Kissen- und Spiralverzerrungen** sind werden durch eine sich mit dem Achsabstand ändernde Vergrösserung oder durch eine Bildrotation hervorgerufen (siehe auch Abb. 4.325).

**Astigmatismus** Fehlende Zylindersymetrie

**Chromatische Abberation** Elektronenquellen können keine monochromatischen Strahlen aussenden

**Boersch-Effekt** Elektronen sind geladen und stossen sich ab. Der Effekt tritt nur bei hohen Elektronenströmen auf.

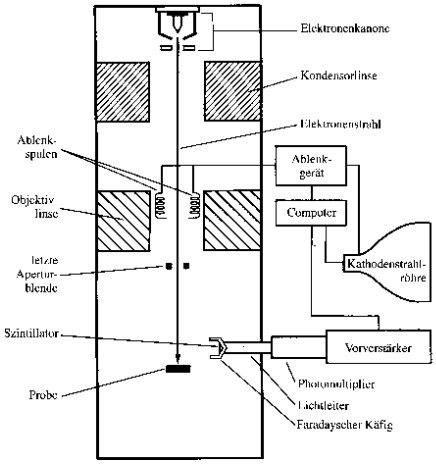



Abbildung 4.326.: Schematischer Aufbau eines Elektronenmikroskopes

Abbildung 4.327.: Aufbau einer Wolframdrahtkathode und rechts ein Detailbild der Kathode

### 4.10.14.2. Elemente eines Rasterelektronenmikroskopes

Ein Rasterelektronenmikroskop (Abb. 4.326) besteht aus einer Elektronenquelle, einer Kondensorlinse sowie einer Objektivlinse. Ablenkspulen rastern den Elektronenstrahl.
Die Kathode besteht entweder aus Wolframdraht (Abbildungen 4.327-4.329) mit Glühemission, einer $LaB_6$-Quelle (Abb. 4.330) oder einer Feldemissionsquelle (Abb. 4.331). Der Unterschied der Quellen besteht in ihren gegen die Feldemissionsquelle abnehmenden Energieunschärfe den erzeugten Elektronen. Die Feldemissionskathode benötigt, anders als die anderen Kathoden, Ultrahochvakuum zum Betrieb.

**4.10.14.2.1. Glühemission** Die Glühemission wird durch die Richardsson-Gleichung

$$j = CT^2 e^{-w/kT} \tag{4.10.136}$$

wobei $W$ die *Austrittsarbeit* ist. Dabei haben die Elektronen eine Energieunschärfe von einigen eV. Lanthanhexaborid hat eine *Austrittsarbeit* von 0,5 eV im Gegensatz zu W mit 5 eV. Entsprechend ist die Betriebstemperatur geringer und die Energieunschärfe liegt bei 1eV.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

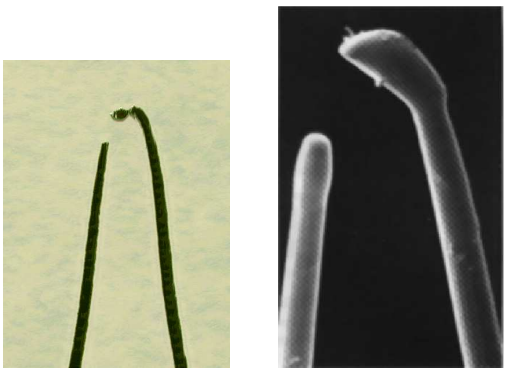


Abbildung 4.328.: Links eine normal durchgebrannte Wolframkathode und rechts eine übersättigte Wolframkathode (mit zu grossem Strom betrieben)

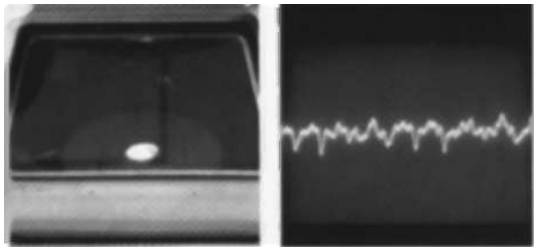


Abbildung 4.329.: Links die Überwachung der Fadensättigung in einem TEM mit Hilfe eines projizierten Fleckes auf einem Phosphorschirm und rechts ein Bild des Fadenstromes eines REM auf einem Oszilloskopschirm.

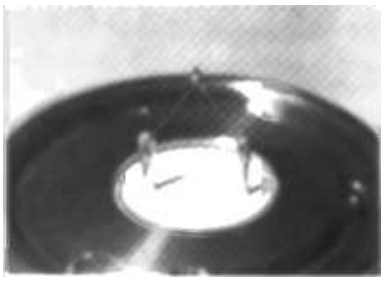


Abbildung 4.330.: Bild einer $LaB_6$-Quelle. Hier wird eine $LaB_6$-Kristall von Wolfram- oder Rhenium-Drähten gehalten.

Abbildung 4.331.: Bild einer Feldemissionskathode. Dabei wird, analog zu einem STM, eine zugespitzte Wolframnadel als Emitter verwendet.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

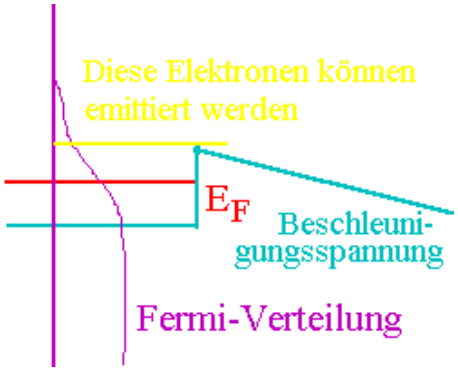



Abbildung 4.332.: Energieverhältnisse an der Oberfläche einer Elektrode

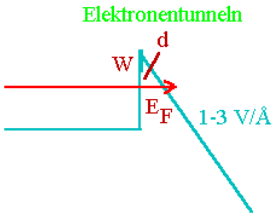



Abbildung 4.333.: Feldemission

**4.10.14.2.2. Feldemission** Die Feldemissionskathode arbeitet mit dem Tunneleffekt. An einer sehr scharfen Spitze fällt das elektrische Feld sehr schnell ab.
Die Energieunschärfe ist hier etwa $100meV$, gegeben durch $kT \approx 47meV$.

**4.10.14.2.3. Emissionsfläche** Ausser bei der Feldemissionskathode wird der Strom aus einer grossen Fläche emittiert. Bei jeder Abbildung ist das Produkt aus Fläche und Strahldivergenz konstant.
Bei Glühkathoden kann mit einem Wehneltzylinder eine kleinere virtuelle Emissionsfläche erzeugt werden.

**4.10.14.2.4. Detektoren** Als Detektor dient ein Faradaybecher mit Szintillator und Photovervielfacher.

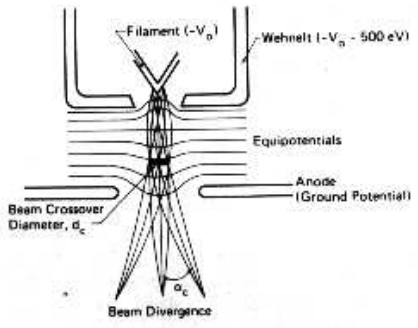



Abbildung 4.334.: Wehnelt-Zylinder zur Erzeugung einer kleineren Emissionsfläche

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Abbildung 4.335.: Everhart-Thornley-Detektor. Die ringförmige Struktur aussen ist der Faraday-Käfig, hinten befindet sich der Szintillator-Kristall

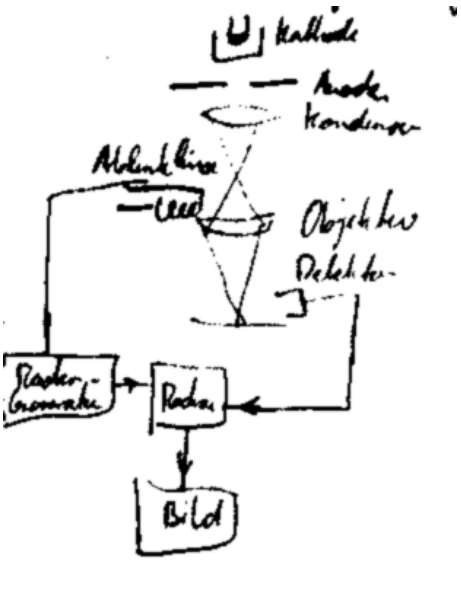



Abbildung 4.336.: Bilderrasterung im REM

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

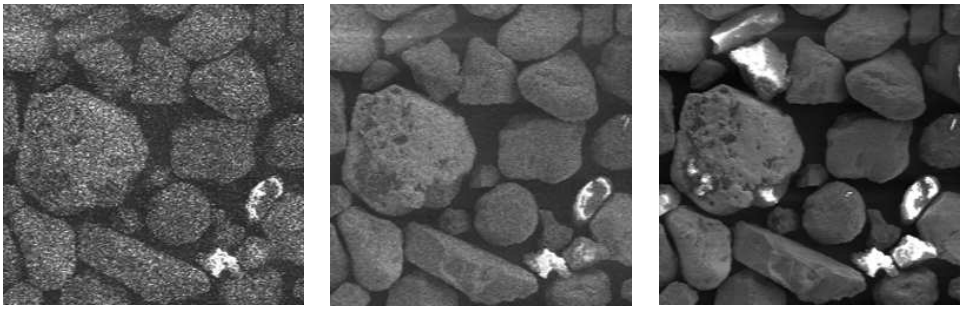


Abbildung 4.337.: Verbesserung der Bildqualität durch Mittelung. Links ist eines der Originalbilder. Die Mitte zeigt das Resultat der Mittelung über 15 Bilder. Rechts ist eine Messung mit optimaler Einstellung gezeigt.

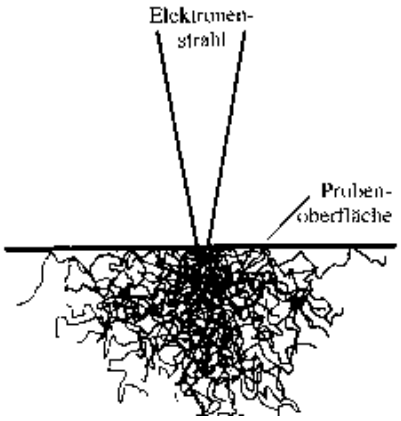



Abbildung 4.338.: Streuung von Elektronen aus dem Strahl eines REMs im Innern einer Probe

Das Bild wird digital erzeugt (siehe auch Abb 4.336). Üblicherweise werden 1024 x 1024 Punkte gemessen.
Bei verrauschten Bildern kann durch Mitteilung eine bessere *Auflösung* erreicht werden (Abb. 4.337).

#### 4.10.14.3. Strahl-Probe Wechselwirkung

Wenn Elektronen aus dem Elektronenstrahl auf die Probe treffen, werden sie gestreut und abgebremst (Abb. 4.338).
Dabei legt jedes Elektron eine andere Bahn zurück. Insgesamt ergibt sich eine Verteilung wie in Abbildung 4.339.
Schwach gebundene Elektronen im Leitungsband werden durch inelastische Prozesse aus der Probe emittiert und vom Detektor abgesaugt (Abb. 4.340).
Die Form der Wechselwirkungsbirne hängt von den Elektronenenergie ab.
Beim TEM (Transmissionselektronenmikroskop) wird die Probe so dünn geschnitten, dass nur der Hals der Wechselwirkungszone in der Probe ist.
Je grösser die Ordnungszahl der Probenatome ist, desto kleiner ist das abgetastete Probenvolumen.
Bei hohen Beschleunigungsspannungen können Elektronen besser fokussiert wer-

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

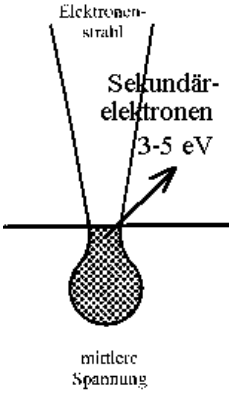



Abbildung 4.339.: Wechselwirkungszone für Elektronen im REM

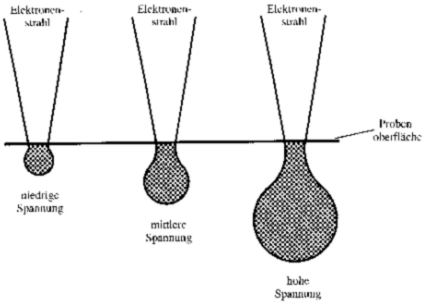



Abbildung 4.340.: Abhängigkeit des Wechselwirkungsvolumens von der Beschleunigungsspannung.

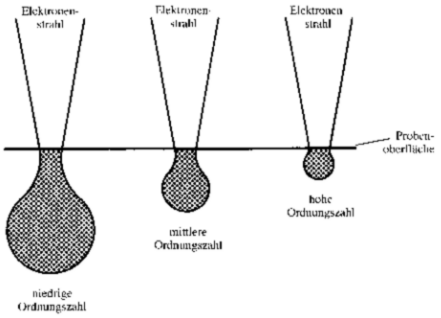



Abbildung 4.341.: Abhängigkeit der Wechselwirkungszone von der Ordnungszahl

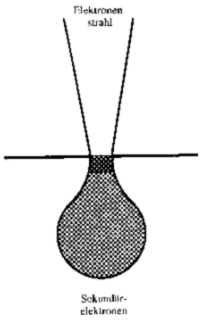



Abbildung 4.342.: Austrittstiefe von Sekundärelektronen

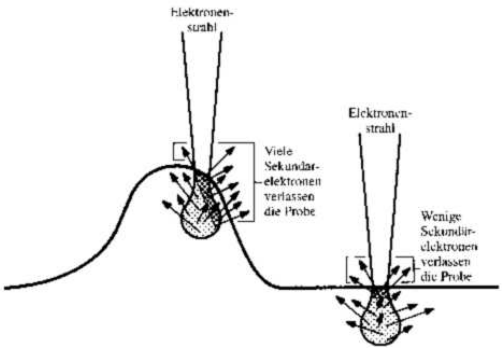



Abbildung 4.343.: Die vier an der Bildgebung beteiligten Prozesse

den. Deshalb ist tendenziell die *Auflösung* besser. Bei dicken Proben und hohen Elektronenenergien nimmt jedoch das Streuvolumen zu, so dass Rasterelektronenmikroskopie etwa bei 20 kV ihre maximale *Auflösung* haben.
Da die erzeugten Sekundärelektronen eine äusserst kleine Energie haben, können sie maximal einige Nanometer vom Ort der Thermalisierung diffundieren (siehe Abb. 4.342. (Bei Isolatoren aus Tiefen bis zu 50 nm). Deshalb trägt der grössere Teil des Streuvolumens nicht zur Abbildung bei. Da Sekundärelektronen nur bei einer kurzer Diffusionsdistanz zur Probenoberfläche austreten können, werden die Kanten und geneigten Flächen hervorgehoben. Deshalb sehen REM-Bilder so plastisch aus.
Zur Bildgebung (Abb. 4.343) tragen vier Prozesse bei:

1. Direkte Sekundärelektronen (Typ I)

2. Durch rückgestreute Elektronen ausgelöste Sekundärelektronen (Typ II (rückgestreute Elektronen haben etwa 60-80 % der ursprünglichen Elektronenenergie)

3. Rückgestreute Elektronen lösen Sekundärelektronen in den Polschuhen aus (Typ III) Diese Elektronen haben eine schlechte *Auflösung*.

4. Sekundärelektronen, die durch Primärelektronen in der letzten Blende ausgelöst werden, (Schlechte Auflöung) (Typ IV). Diese Elektronen erzeugen ein Hintergrundsrauschen.

Rückstreuelektronen werden wegen ihrer hohen Energie nicht in den Detektor abgelenkt.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

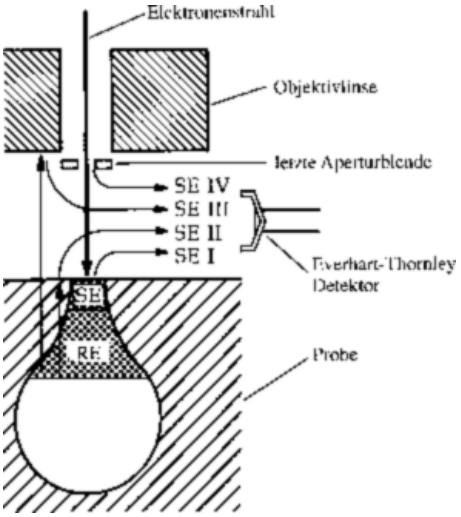



Abbildung 4.344.: Herkunft der Sekundärelektronen im REM

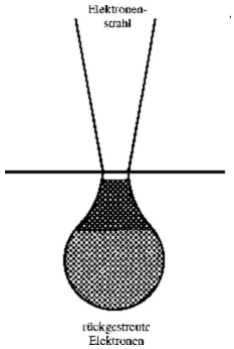



Abbildung 4.345.: Austrittstiefe rückgestreuter Elektronen

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

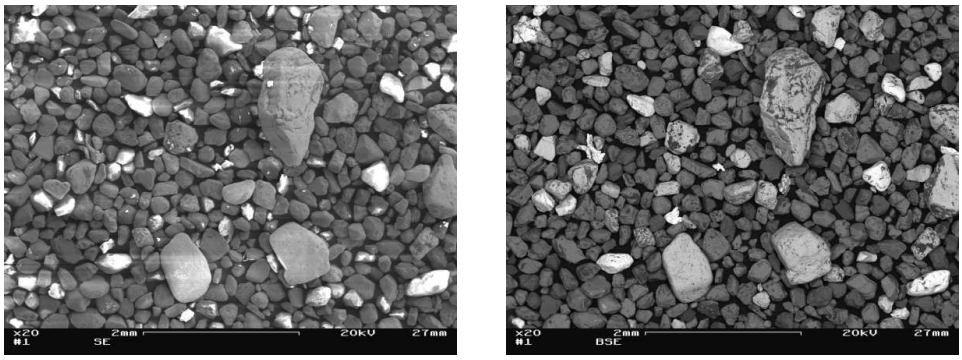



Abbildung 4.346.: Sekundärelektronenbild (links) und Rückstreubild (rechts)

Abbildung 4.347.: Vergleich der REM-Bilder als Funktion der Oberflächenbeschaffenheit. Links ist ein Sekundärelektronenbild der polierten Oberfläche eine antiken griechischen Münze. In der Mitte befindet sich das Rückstreubild, während rechts ein Sekundärelektronenbild einer Bruchfläche gezeigt ist.

Da die Sekundärelektronen (Abb. 4.344) vorwiegend aus oberflächennahen Zonen kommen und Typ II-Elektronen durch aus der Tiefe kommende Rückstreuelektronen (Abb. 4.345) ausgelöst wurden, erzeugen sie Bilder aus unterschiedlicher Tiefe.
Dabei hängt die Rückstreuung von der Ordnungszahl ab (5% bei N=7, 40 % bei N = 47 (Silber))
Bild 4.347 zeigt, dass die beiden Arten Sekundärelektronen bei polierten Proben (griechische Menge) die Legierungsbestandteile trennen können. Typ I-Elektronen zeigen zusätzlich Oberflächenkratzer. Gebrochene Proben werden durch Kanteneffekte dominiert.
Bei der Wechselwirkung von Licht mit der Probe treten weitere inelastische Prozesse auf:

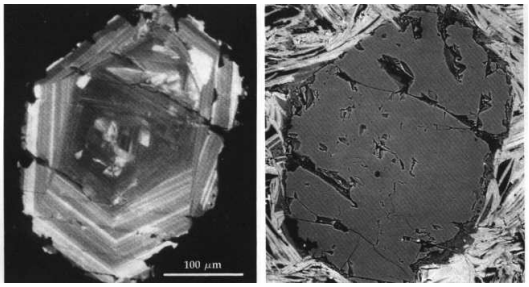



Abbildung 4.348.: Links ist ein Kathodolumineszenzbild einer geologischen Probe gezeigt. Rechts ist zum Vergleich dazu ein Rückstreubild gezeigt.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

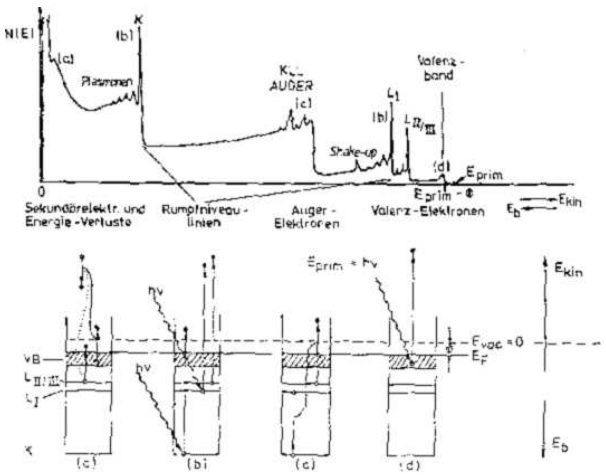



Abbildung 4.349.: Schematische Übersicht über die Zahl der photoemittierten Elektronen. Oben wird das Spektrum der Sekundärelektronenprozesse gezeigt. Unten sind, von links, die Sekundärelektronenanregung und die dabei auftretenden Energieverluste, Emission aus Rumpfniveaus, Augerprozesse und die Emission aus dem Valenzbandbereich dargestellt.

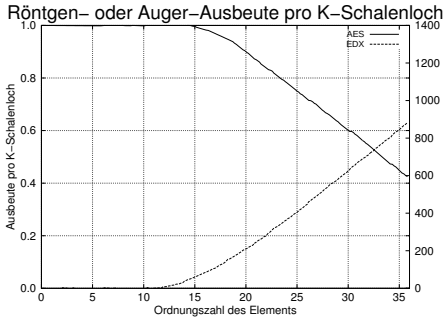



Abbildung 4.350.: Ausbeute von Augerelektronen als Funktion der Ordnungszahl

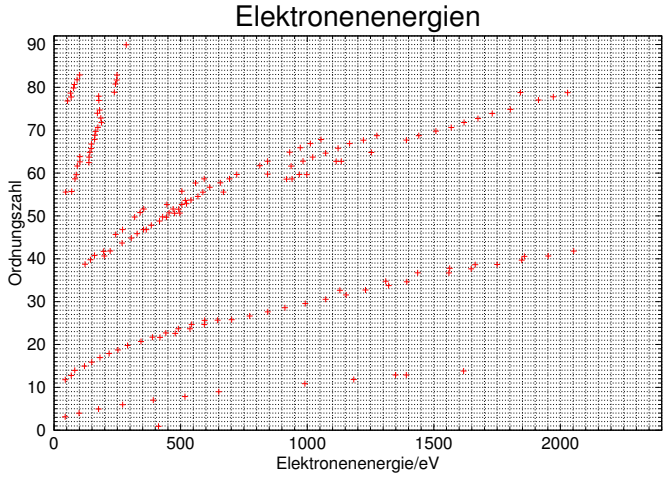



Abbildung 4.351.: Elektroneneenrgien beim Auger-Prozess als Funktion der Ordnungszahl

**Lumineszenz** Primärelektronen regen Elektronen ins Leitungsband an. Diese rekombinieren intern Aussendung von Licht. Dieser Abbildungsmodus ist vor allem bei Halbleiterproben beliebt (Abb. 4.348).

**Auger-Elektronen** Augerelektronen treten durch einen Folgeprozess neben photoemittierten Elektronen auf. Als Konkurrenzprozess zur Emission von Augerelektronen kann die durch den Elektronenübergang erzeugte Energie auch als charakteristische Röntgenstrahlung abgegeben werden. Analysiert man die Energie dieser Strahlung, so spricht man von EDX („Energy Dispersive X-Ray Analysis"). Abb. 4.350 zeigt, dass die Augerelektronenspektroskopie bei relativ niedrigen Ordnungszahlen, EDX bei relativ hohen Ordnungszahlen einen empfindlichen Nachweis von Elementen ermöglicht. Wegen der wesentlich grösseren Fluchttiefe von Photonen wird bei EDX jedoch über einen tiefen Bereich (ca. 1 $\mu$m) unter der Oberfläche gemittelt, so dass schon weitgehend Volumeneigenschaften erfasst werden. Der Augerelektronenprozess ist bestimmt durch drei Orbitalenergien (siehe Abb. 4.351). So lässt sich beispielsweise die kinetische Energie von $KL_IL_{II/III}$-Elektronen über

$$E(KL_IL_{II/III}) = E(K) - E(L_I) - E(L_{II/III})* \tag{4.10.137}$$

grob abschätzen. Darin ist E(K) die Bindungsenergie des unteren Lochzustandes, $E(L_I)$ die Bindungsenergie des Elektrons, das diesen Lochzustand auffüllt, und $E(L_{II/III})*$ die effektive Bindungsenergie des emittierten Augerelektrons. Letztere weicht signifikant von der Energie des neutralen Atoms ab, da starke Wechselwirkungen zwischen den beiden Endzustandslöchern im Atom auftreten. So wird in dem o.g. Beispiel nach Auffüllung der $K$-Schale durch das $L_I$-Elektron die Bindungsenergie des $L_{II/III}$-Elektrons erhöht durch das Erzeugen eines Lochs im $L_I$-Orbital. Die Loch/Loch-Wechselwirkung in der Endzustandskonfiguration hängt dabei davon ab, ob beide Löcher in den Rumpfniveaus, ein Loch im Rumpfniveau und ein anderes in schwächer gebundenen Bändern oder beide in Bändern auftreten. In guter Näherung lassen sich die Augerelektronenenergien abschätzen über:

$$\begin{aligned} E\left[KL_IL_{II/III}\right] &= E\left[K(Z)\right] \\ &\quad -\frac{1}{2}\left\{E\left[L_I(Z)\right] - E\left[L_I(Z+1)\right]\right. \\ &\quad \left. +E\left[L_{II/III}(Z)\right] + E\left[L_{II/III}(Z+1)\right]\right\} \end{aligned} \tag{4.10.138}$$

Auch gebräuchlich ist es, die Coulomb-Abstossung der Lochzustände über einen separaten Energieterm zu erfassen. Dabei wird angesetzt:

$$\begin{aligned} E[KL_IL_{II/III}] &= E[K(Z)] - E[L_I(Z)] \\ &\quad -E[L_{II/III}(Z)] - U[KL_ILII/III] \end{aligned} \tag{4.10.139}$$

Darin erfasst der Term $U[KL_IL_{II/III}]$ alle Korrelationseffekte. Bei hoher Korrelation der Bewegung der Löcher und grosser räumlicher Nähe erfolgt starke Coulomb-Abstossung. Diese qualitativen Beispiele machen deutlich,

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

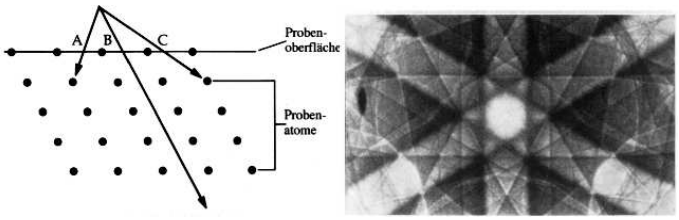



Abbildung 4.352.: Elektronenchanneling. Links ist das Prinzip gezeigt, rechts eine daraus resultierende Abbildung.

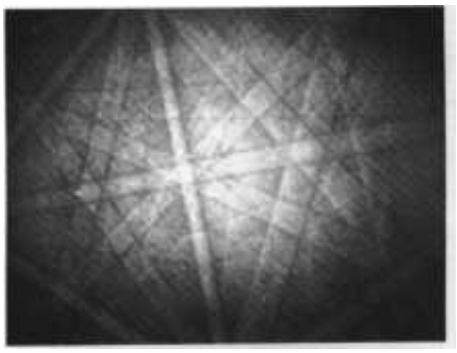


Abbildung 4.353.: Elektronenchannelingbild einer kubisch-raumzentrierten Aluminium/Eisen-Legierung.

dass die Augerelektronenspektroskopie neben dem überwiegenden Einsatz zur Elementcharakterisierung auch zur Charakterisierung lokaler Bindungsverhältnisse am Zentralatom herangezogen werden kann. Ebenso wie bei XPS sind Augerelektronenübergänge unter ausschliesslicher Beteiligung von Rumpfniveaus durch relativ scharfe Linien gekennzeichnet, deren Form in erster Näherung unabhängig von der chemischen Umgebung ist, die jedoch eine charakteristische chemische Verschiebung aufweisen können. Augerelektronen unter Beteiligung des Valenzhandes zeigen dagegen eine extreme Abhängigkeit der Linienform vom Zustand der Oberfläche. Eine quantitative Auswertung ist allgemein schwierig, da wegen der Beteiligung mehrerer Orbitale eine Entfaltung vorgenommen werden muss, um die Valenzbandstruktur aus Augerelektronenspektren zu ermitteln. Die grosse Oberfiächenempfindlichkeit der Augerelektronenspektroskopie ist durch die Fluchttiefe der Elektronen bei kinetischen Energien der Elektronen unter 1000 eV gegeben.

**Erzeugung von Röntgenquanten** Die Energie der Elektronen wird auf Photonen übertragen. (Deshalb muss die Strahlenschutzverordnung an Elektronenmikroskopen beachtet werden!).

**Channeling** In bestimmte Kristallrichtungen können Elektronen über lange Distanzen durch die Probe wandern (siehe Abb. 4.352 und 4.353). Deshalb verlassen an diesen Punkten weniger Sekundärelektronen die Probe. Dunkle Linien zeigen die Kristallstrukturen.

**Probenstrombild** Es werden vor allen innere Strukturen abgebildet. Dieses Verfahren wird in den Materialwissenschaften angewandt (Abb 4.354).

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

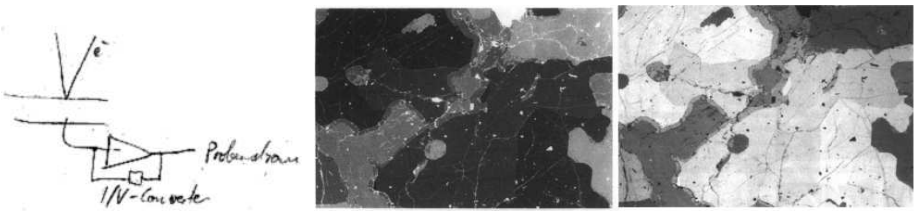


Abbildung 4.354.: Aufnahme von Probenstrombildern im REM. Links ist die dazugehörige Schaltung gezeichnet. In der Mitte befindet sich ein Probenstrombild von Felsit, rechts das dazugehörige Rückstreubild.

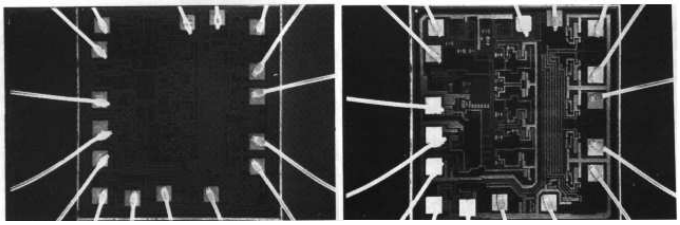


Abbildung 4.355.: Potentialkontrastbild. Links ist ein Sekundärelektronenbild gezeigt, rechts der gleiche Chip mit angelegter Spannung.

**Potentialabbildung** Wenn an einem kleinen Probenbereich eine Spannung angelegt wird, verändert dies die Anzahl Sekundärelektronen (siehe Abb. 4.355).

**EBIC Electron-Beam induced current** Die auf einem P-N-Übergang fokussierten Elektronen erzeugen Elektronen-Loch-Paare (Abb. 4.356). Diese sind als induzierten Strom messbar. Anwendung findet dieses Verfahren vorwiegend im Bereich der Halbleitercharakterisierung.

**Magnetfelder** - Letztlich beeinflussen lokale Magnetfelder die Emissionsrichtung der Elektronen.

### 4.10.14.4. Einfluss der Gerätevariablen

Wenn die Brennweite der Kondensorlinse verringert wird, verringert sich auch der Durchmesser des Brennfleckes auf der Probe. Die Divergenz der Elektronenstrahlen zwischen Kondensorlinse und Objektivlinse nimmt jedoch zu, so dass weniger

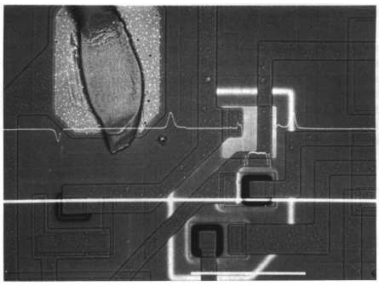


Abbildung 4.356.: Elektronenstrahlinduzierte Ströme (EBIC)

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

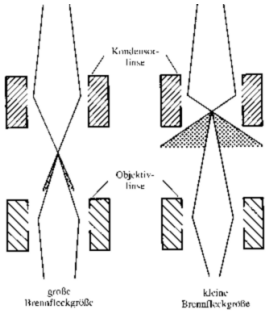



Abbildung 4.357.: Brennfleckgrösse und Strahlstrom

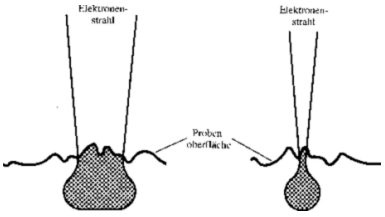



Abbildung 4.358.: Brennfleckgrösse und *Auflösung*

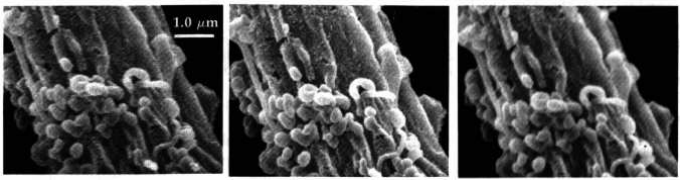



Abbildung 4.359.: Einfluss der Brennfleckgrösse auf *Auflösung* und Bildqualität. Links ist eine Abbildung mit dem kleinstmöglichen Brennfleck, in der Mitte mit einem etwas grösseren Brennfleck und links mit einem ganz grossen Brennfleck.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

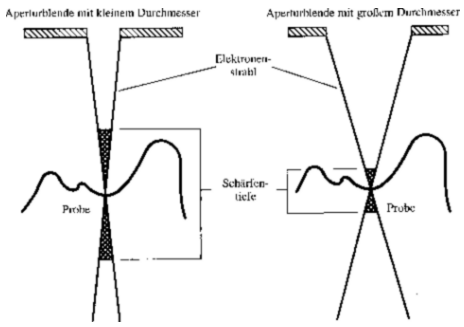



Abbildung 4.360.: Aperturgrösse und Schärfentiefe

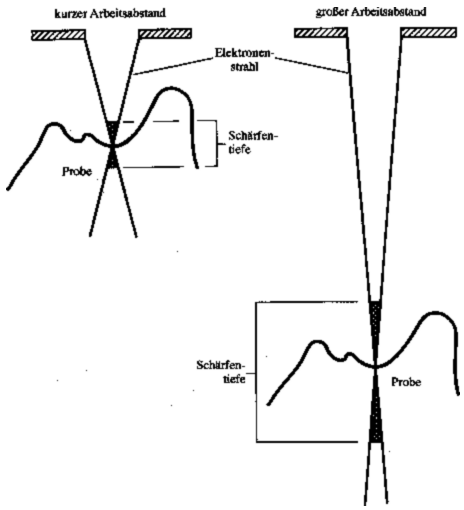



Abbildung 4.361.: Arbeitsabstand und Schärfentiefe

Elektronen die Proben treffen. Deshalb wird das Bild verrauschter (die Photonenstatistik kann nicht umgangen werden..) (Siehe Abbildungen 4.357 bis 4.359).

Wenn der Öffnungswinkel des Elektronenstrahles vergrössert wird (grössere Aperturblende) verringert sich die Tiefenschärfe. Diesen Effekt gibt es auch bei optischen Mikroskopen. Ebenso beeinflusst der Arbeitsabstand die Schärfentiefe. Je weiter dieser Arbeitsabstand ist, desto grösser die Schärfentiefe, da der Öffnungswinkel kleiner wird. Der Effekt ist in den Abbildungen 4.360 bis 4.362 gezeigt.

Die letzte Möglichkeit, die *Auflösung* zu ändern, bietet die Beschleunigungsspannung (Siehe auch die Abbildungen 4.363 und 4.364). Üblicherweise arbeiten REMs mit 5-30 keV. Metallische Proben werden dabei bei höheren Beschleunigungsspannungen besser abgebildet. Organische Materialien sind jedoch bei tieferen Spannungen durch die verringerte Eindringtiefe besser aufzulösen. Spezialisierte REM's arbeiten mit Elektronenenergien von 100 eV, so dass auch biologische sowie Polymerproben befriedigend abgebildet werden können.

Zusätzlich besteht heute auch die Möglichkeit, Rasterelektronenmikroskopie bei fast physiologischen Bedingungen durchzuführen. Dabei wird eine dünne Zone mit einigen mbar Druck und hoher Luftfeuchtigkeit geschaffen. Der Rest des Elektronenmikroskopes wird über differentielle Pumpstufen von der Probe entkoppelt.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

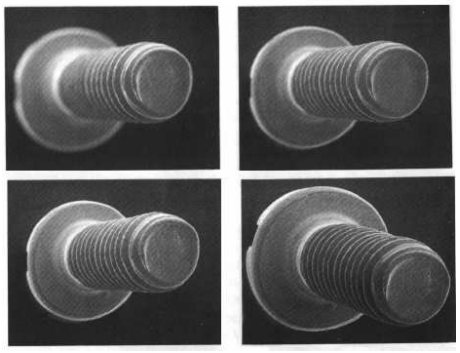


Abbildung 4.362.: Wirkung der Aperturblende und des Arbeitsabstandes. Die Einstellungen sind: oben links: Aperturblende 600 $\mu$m, Arbeitsabstand 15 mm, oben rechts: Aperturblende 200 $\mu$m, Arbeitsabstand 15 mm, unten links: Aperturblende 100 $\mu$m, Arbeitsabstand 15 mm, unten rechts: Aperturblende 100 $\mu$m, Arbeitsabstand 39 mm.

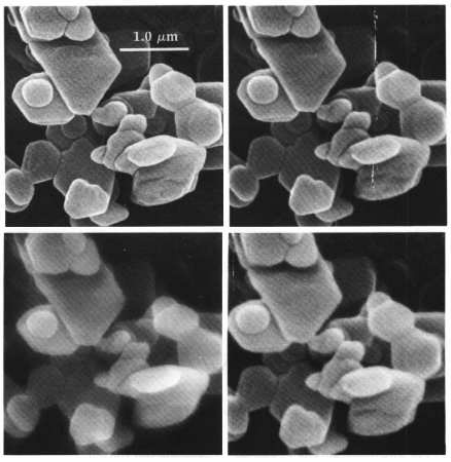



Abbildung 4.363.: Beeinflussung der *Auflösung* durch Aperturblende und Arbeitsabstand. Die Einstellungen sind: oben links: Aperturblende 100 $\mu$m, Arbeitsabstand 15 mm, oben rechts: Aperturblende 200 $\mu$m, Arbeitsabstand 15 mm, unten links: Aperturblende 600 $\mu$m, Arbeitsabstand 15 mm, unten rechts: Aperturblende 200 $\mu$m, Arbeitsabstand 39 mm.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

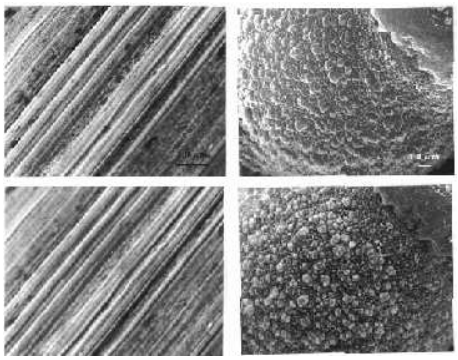


Abbildung 4.364.: Wirkung der Beschleunigungsspannung auf die Bildqualität. Oben links: Drähte aus reinem Gold mit 35 kV Spannung fotografiert. Unten links das gleiche, aber mit 10 kV Spannung. Oben rechts: Eichenpollen bei 35 kV, unten rechts: Eichenpollen bei 10 kV

Abbildung 4.365.: Probenhalter für REM

#### 4.10.14.5. Propenpräparation für REM

Proben müssen für die meisten Geräte frei von Wasser und Lösungsmitteln sein, da diese die Säule (mit den Linsen) verunreinigen. Proben werden auf gerätespezifischen Haltern montiert, wie sie in der Abb. 4.365 gezeigt sind.
Dabei soll der Kleber folgende Eigenschaften haben:

- Ziemlich hohe Viskosität, da sonst die Probe benetzt würde.
- Mittlere Trocknungszeit. Bei zu kurzer Trocknungszeit hat man nicht genügend Zeit die Probe zu positionieren. Bei zu langer Trocknungszeit treten während langer zeit Gase aus dem Kleber in das Mikroskop aus.
- Leitfähigkeit ist wünschenswert, da so eine getrennte Kontaktierung vermieden werden kann.

Weiter muss die Probe in den meisten Fällen elektrisch leitend sein. Bei nichtleitenden Proben bauen sich dabei Ladungen auf, die den Elektronenstrahl ablenken können, wie aus Abb. 4.366 ersichtlich. Um dies zu verhindern muss die nichtleitende Probe besputtert oder bedampft werden.
Eine Möglichkeit besteht darin, mit Argon-Ionen Gold oder anderen Atome zu zerstäuben und die Oberfläche der Probe mit einer wenigen nm dichten leitfähigen Schicht zu bedecken (Abb. 4.367).

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

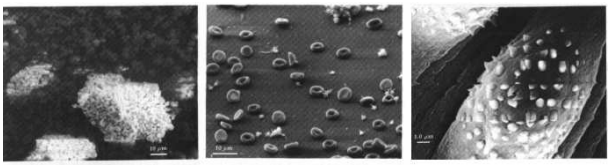


Abbildung 4.366.: Aufladen von Proben im REM. Links: Flugasche, bei der Aufladung zu anomalem Kontrast führt. Mitte: menschliche rote Blutkörper mit Linien erzeugender Aufladung. Rechts: Schistosom, die Aufladung führt zu einer Ablenkung des Strahls.

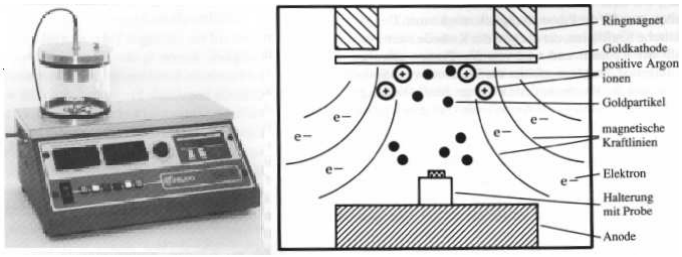



Abbildung 4.367.: Kathodenzerstäuber. Links ist das Bild eines typischen Zerstäubers und rechts eine Skizze der Arbeitsweise.

Soll eine Röntgen-Elementanalyse durchgeführt werden, dürfen keine schweren Atome zur Bedampfung verwendet werden. Diese würden sehr stark mit den Elektronen wechselwirken. Deshalb verwendet man in solchen Fällen Kohlenstoff.
Unter bestimmten Bedingungen kann auf eine Bedampfung verzichtet werden. Die Emissionswahrscheinlichkeit für Elektronen hängt von der Beschleunigungspannung ab.

### 4.10.14.6. Emission

Es geht ein Energiefenster zwischen $E_a$ und $E_b$, bei dem die Emissionswahrscheinlichkeit grösser als 1 ist (Abb. 4.368).
Bild 4.369 zeigt, dass Salz bei 1 kV und 15 kV Aufladungseffekte zeigt, nicht aber bei 5 kV.
Um die innere Struktur von Proben sichtbar zu machen, müssen diese zum Teil

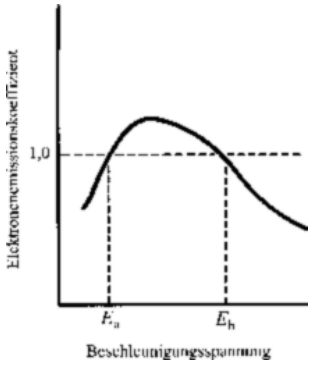



Abbildung 4.368.: Elektronenemissionskoeffizient als Funktion der Beschleunigungsspannung

Abbildung 4.369.: Abbildung von Salz bei verschiedenen Beschleunigungsspannungen (von links: 1 kV, 5kV, 15kV).

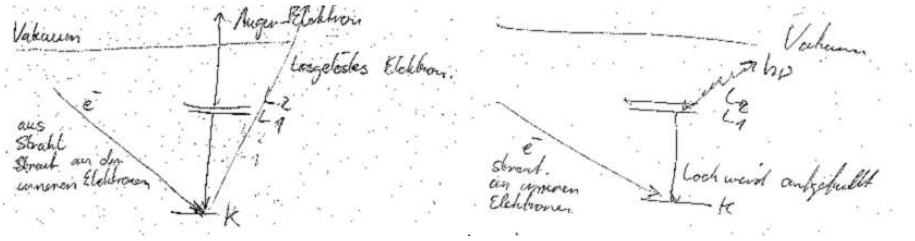



Abbildung 4.370.: Auger-Prozesse (links) und Röntgenprozesse (rechts)

angeschliffen oder angeätzt werden.

### 4.10.14.7. Röntgenmikroanalyse

Durch die inelastische Wechselwirkung von Elektronen mit der Probe ist es möglich materialspezifische *Signale* zu generieren. Neben den rückgestreuten Elektronen sind dies insbesondere die Auger-Elektronen und die Röntgen-Prozesse (Abb. 4.370) .
Auger-Elektronen haben eine Energie, die unabhängig von der Energie der einfallenden Elektronen ist. Ebenso sind die entstehenden Röntgenphotonen materialspezifisch. Da neben Auger-Elektronen auch andere Sekundärelektronen vorhanden sind und da der Anger-Prozess keine grosse Ausbeute hat, sind Auger-Elektronen schwer zu detektieren. Zudem verlieren Elektronen durch Streuung in der Probe Energie.
Diese Verluste können minimiert werden, wenn die Probe dünn geschnitten wird (Abb. 4.371).
Röntgen-Photonen wechselwirken wesentlich weniger mit der Probe als Elektronen. Sie geben deshalb ein besseres Bild der Probenzusammensetzung. Die *Auflösung*

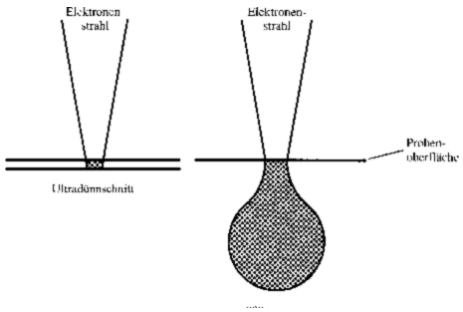



Abbildung 4.371.: Wechselwirkung zwischen Elektronenstrahl und Probe

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

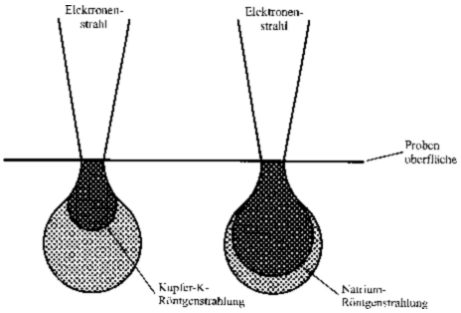



Abbildung 4.372.: Räumliche *Auflösung* bei der EDX-Abbildung

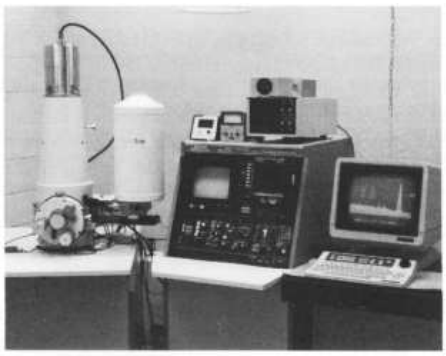


Abbildung 4.373.: EDS-System angeschlossen an ein REM

der Röntgenphotonen hängt vom Material ab (Abb. 4.372).
Die Energie der Strahlelektronen nimmt nach aussen in der Wechselwirkungsbirne ab. Deshalb werden Elemente mit höheren Anregungsenergien mit besserer Lokalisierung abgebildet.
Gemessen werden die Röntgenphotonen mit einer EnergieDispersiven Spektroskopie (EDS). Dies ist in den Abbildungen 4.373 und 4.374 gezeigt .
Das obige Bild zeigt einen EDS-Detektor. Er besteht aus Kollektor, Detektionskristall und Feldeffektransistor. Ein Röntgenphoton löst im Detektorkristall eine zu seiner Energie proportionale Anzahl Elektronen. Diese werden auf dem Gate des FET gesammelt und erhöhen die Kanalleitfähigkeit. Die Spannung am Arbeitswiderstand steigt (Abb. 4.375).
Aus der Rampe werden Impulse geformt und diese digitalisiert. So kann die Energieverteilung der Röntgenphotonen gemessen werden. Dazu wird ein Vielkanalanalysator verwendet.
Der Detektorkristall besteht üblicherweise aus Silizium dotiert mit Lithium. Um

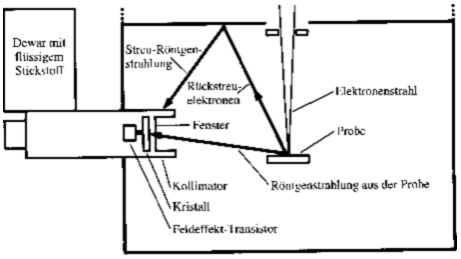



Abbildung 4.374.: Schematische Darstellung eines EDS-System

Abbildung 4.375.: Funktion eines EDS-Detektors

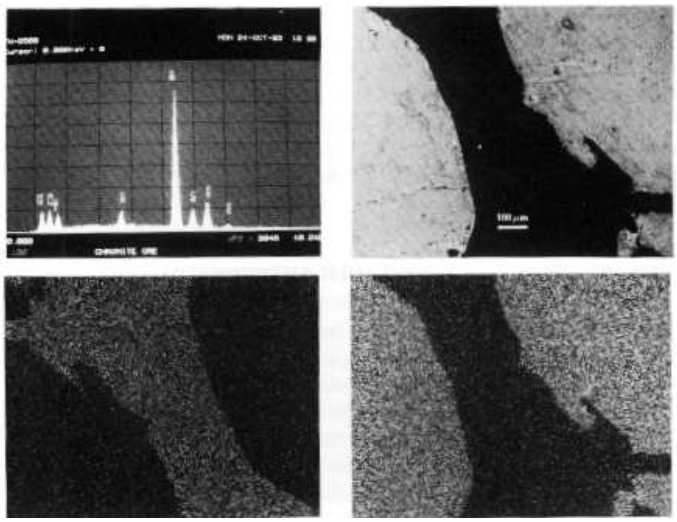


Abbildung 4.376.: EDS und REM-Analyse von Chromiterz. Oben links: EDS-Spektrum mit Peaks für Magnesium, Aluminium, Silizium, Kalzium, Chrom und Eisen. Oben rechts: Rückstreuelektronenbild. Unten links: Punktdichtebild von Silizium. Unten rechts: Punktdichtebild von Chrom.

die Diffusion von Li zu verhindern, muss der Kristall bei 77 k gehalten werden. Pro 3,8eV Energie wird ein Elektronen-Loch-Paar gebildet. Bei der Messung von Spektrum gibt es einige Optimierungsmöglichkeiten (Abb. 4.376).

**4.10.14.7.1. Optimierung der Röntgenmikroanalyse** Der Detektorabstand und das Aufnahmewinkel (Abb. 4.377) bestimmen ob stark oder weniger absorbierte Röntgen-Photonen gemessen werden.
Ebenso beeinflusst die Neigung der Probe, ob die Röntgenstrahlen mehr oder weniger absorbiert werden (Abb. 4.378).
Der Raumwinkel hängt vom Abstand des Detektors von der Probe sowie dem Abstand vom Elektronenstrahl ab (Abb. 4.379 und 4.380).

**4.10.14.7.2. Artefakte** Bei der Bestimmung der Röntgenspektren können Materialien im Mikroskop Beiträge leisten.
Die Fluoreszenz der Aperturblende sowie der Probenkammern bestimmt die Empfindlichkeitsgrenze (Abb. 4.381).
Weiter können Röntgenphotonen im Detektorkristall Röntgenquanten erzeugen. Diese werden postwendend wieder absorbiert und erzeugen sogenannte Escape-Peaks (Abb. 4.382 und 4.383) .

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

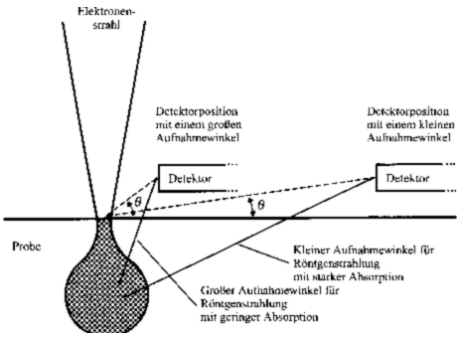



Abbildung 4.377.: Optimierung von Detektorabstand und Aufnahmewinkel für die Röntgen-Mikroanalyse

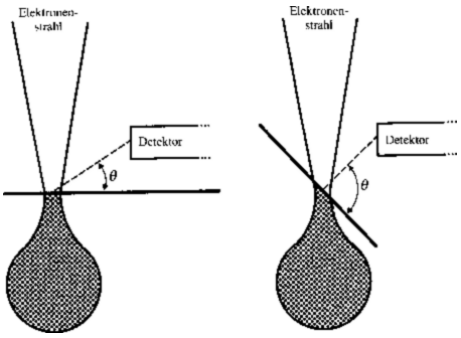



Abbildung 4.378.: Optimierung von Probenneigung und Aufnahmewinkel für die Röntgenmikroanalyse

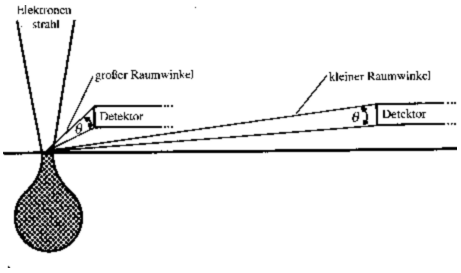



Abbildung 4.379.: Optimierung von Detektorabstand und Raumwinkel für die Röntgenmikroanalyse

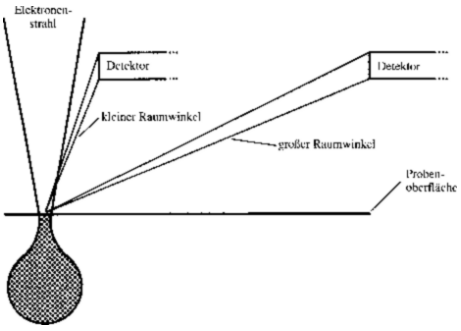



Abbildung 4.380.: Detektorabstand und Raumwinkel bei grossem Abstand

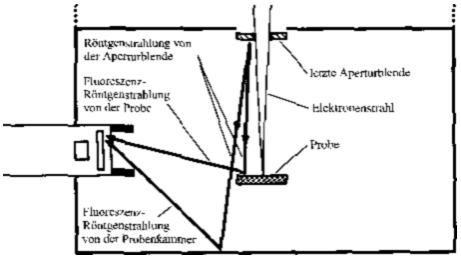



Abbildung 4.381.: Systempeaks aufgrund der Fluoreszenz der letzten Aperturblende

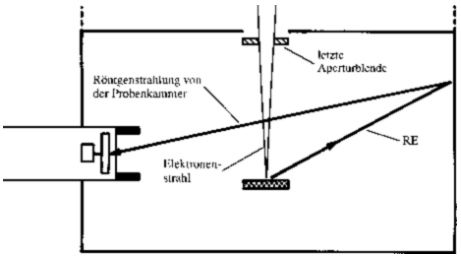



Abbildung 4.382.: Rückgestreute Elektronen und Systempeaks

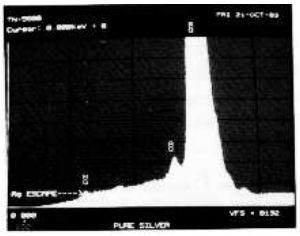



Abbildung 4.383.: Escape-Peak von Silber

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

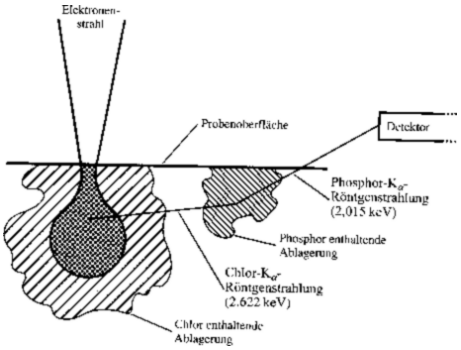



Abbildung 4.384.: Sekundärfluoreszenz

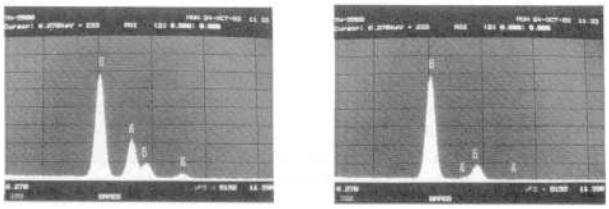


Abbildung 4.385.: Entfalten von Peak-Überlapps

Ein weiterer Artefakt ist die Sekundärfluoreszenz. Aus der Probe austretende Röntgenstrahlung kann auf dem Weg in anderen Materialien Fluoreszenz hervorrufen. Dann glaubt man, am Ort des Elektronenstrahls sei ein dort nicht vorhandenes Element präsent (Abb. 4.384).
Für eine quantitative Analyse müssen zusammenfallende Peaks entfaltet werden. Die Verfahren dazu stammen aus der Röntgenstrukturanalyse.
Im obigen Bild ist ein Beispiel gegeben.

## 4.10.15. Transmissionselektronenmikroskopie(TEM)

Die *Transmissionselektronenmikroskopie*, auch *TEM* genannt, ist eine genaue Umsetzung eines klassischen optischen Mikroskopes auf die Elektronenoptik. Wie bei einem Rasterelektronenmikroskop besteht die Elektronenquelle aus einer klassischen Kathode, einer Lanthan-Hexaborid-Kathode oder einer Feldemissionskathode. Bei den ersten beiden bildet ein Wehnelt-Zylinder eine virtuelle Punktquelle nach. Diese Punktquelle wird nun, anders als bei der Rasterelektronenmikroskopie zur Beleuchtung der gesamten abzubildenden Fläche verwendet. Die Beschleunigungsspannung der Elektronen beträgt zwischen einigen 10 Kilovolt bis zu über einem Megavolt. Die Wellenlänge der Elektronen ist demnach sehr viel kleiner als der typische interatomare Abstand. Da die numerische Apertur $NA$ von magnetischen Linsen, wie sie im Abschnitt 4.10.14 über die Rasterelektronenmiktroskopie beschrieben wurden, sehr viel kleiner als eins ist, benötigt man diese extrem kleinen Wellenlängen. Nur Mikroskope, deren Elektronenenergien grösser als etwa $200keV$ sind, können Gitternetzebenen abbilden.
Die von den Elektronen durchleuchtete Probe wird mit magnetischen Linsensystemen, die der Optik von Lichtmikroskopen nachempfunden sind, abgebildet. Als Detektoren verwendet man entweder Phosphorschirme, photographische Emulsionen oder aber CCD-Kameras.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

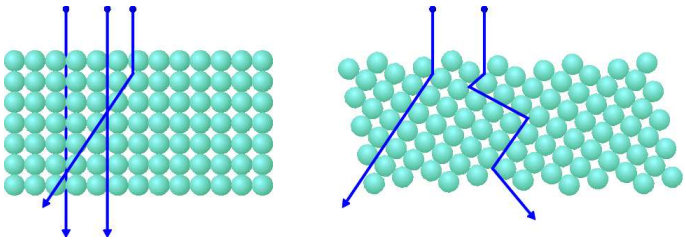


Abbildung 4.386.: Abbildung von Gitternetzebenen

Transmissionselektronenmikroskope sind im allgemeinen nicht in der Lage, einzelne Atome abzubilden. Eine *Auflösung* von Gitternetzebenen ist dann möglich, wenn eine niedrig indizierte Kristallfläche parallel zur Abbildungsachse steht. Abbildung 4.386 zeigt zwei typische Situationen bei der TEM-Abbildung. Links ist ein zur optischen Achse des Transmissionselektronenmikroskopes parallel ausgerichteter Kristall zu sehen. Diejenigen Elektronen, die in einer Lücke zwischen Atomen auftreffen, werden praktisch nicht gestreut. Dieses Verhalten nennt man auch *Channeling*. Nur diejenigen Elektronen, die ein gebundenes Elektron auf einer inneren Bahn treffen, werden gestreut. Deshalb werden alle ausgerichteten Säulen von Atomen schwarz abgebildet.
Die rechte Seite von Abbildung 4.386 zeigt die Situation bei einer fehlausgerichteten Probe. Hier gibt es keine in der Ausbreitungsrichtung der Elektronen liegenden Kanäle. Deshalb werden praktisch alle Elektronen gestreut und fallen deshalb für die Abbildung weg. Diese Stelle der Probe erscheint schwarz.
Die Abbildung im Transmissionselektronenmikroskop ist nur dann möglich, wenn die Proben so dünn sind, dass bei der Abbildung keine wesentliche Absorption der Elektronen und keine Vielfachstreuung auftritt.
Wenn die Detektion der Elektronen nicht in einer Objektebene sondern in einer Fokusebene erfolgt, misst man wie auch beim klassischen optischen Mikroskop die *Fouriertransformation* der Abbildung. Wenn die Atome der Probe entlang der optischen Achse des Mikroskopes niedrig indizierte Säulen bilden, dann tritt eine *Elektronenbeugung* ähnlich der *Laue-Abbildung* in einem Röntgenexperiment auf. Das Beugungsmuster erlaubt einen eindeutigen Rückschluss auf die Kristallstruktur. Abbildung 4.387 zeigt als Beispiel eine transmissionselektronenmikroskopische Abbildung von *Si(111)-(7x7)*. Auch wenn es mit einem transmissionselektronenmikroskop nicht möglich ist, Gitternetzebenen abzubilden, kann man immer noch mit den beschriebenen Beugungsexperimenten Anhaltspunkte zur Struktur der Probe gewinnen.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

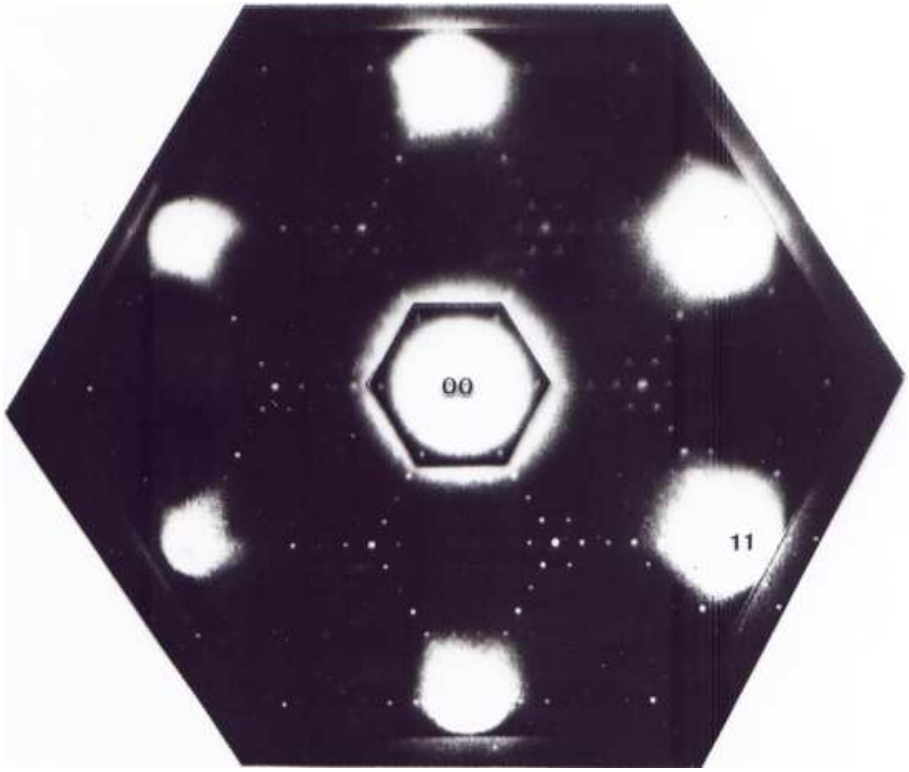



Abbildung 4.387.: Abbildung von Gitternetzebenen und der Oberflächenstruktur von *Si(111)-(7x7)*

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

## 4.10.16. Elektronenbeugung

Elektronenbeugung ist eine in der Oberflächenphysik[?] übliche Methode zur Untersuchung periodischer Probenoberflächen. In den nächsten beiden Abschnitten werden die Beugung niederenergetischer Elektronen sowie die Beugung von Elektronen mit mittlerer Energie besprochen.

### 4.10.16.1. Reziprokes Gitter

Periodische Anordnungen von Atomen werden Netze genannt. Oberflächennetze sind translationsinvariant. Es gilt also

$$f(\boldsymbol{r}+\boldsymbol{T}) = f(\boldsymbol{r}) \tag{4.10.140}$$

mit $\boldsymbol{T} = v\boldsymbol{a}_1 + w\boldsymbol{a}_2$ wobei $(v, w, \in \mathbb{Z})$. Dabei ist $f$ die funktionale Darstellung einer beliebigen Eigenschaft der Oberfläche. Die Entwicklung von $f(\boldsymbol{r})$ in eine Fourier-Reihe ergibt

$$f(\boldsymbol{r}) = \sum_{\boldsymbol{G}} f_{\boldsymbol{G}} e^{i\boldsymbol{G}\boldsymbol{r}} \tag{4.10.141}$$

Die Summe in Gleichung (4.10.141) geht über alle reziproken Gittervektoren. Dabei ist

$$\boldsymbol{G} = h\boldsymbol{A}_1 + k\boldsymbol{A}_2 \tag{4.10.142}$$

wobei $h$ und $k$ ganze Zahlen sind. $\boldsymbol{A}_1$ und $\boldsymbol{A}_2$ sind die erzeugenden Vektoren des primitiven Netzes.
Zwischen dem Netz im realen Raum aufgespannt durch $\boldsymbol{a}_1$ und $\boldsymbol{a}_2$ und dem Netz im reziproken Raum aufgespannt durch $\boldsymbol{A}_1$ und $\boldsymbol{A}_2$ muss die Beziehung

$$\boldsymbol{G}\cdot\boldsymbol{T} = n2\pi \; n \in \mathbb{Z} \text{ für alle } \boldsymbol{G}, \boldsymbol{T} \tag{4.10.143}$$

Aus den Beziehungen (4.10.140) bis (4.10.143) folgt:

$$\begin{aligned} \boldsymbol{A}_1\cdot\boldsymbol{a}_1 &= 2\pi \\ \boldsymbol{A}_1\cdot\boldsymbol{a}_1 &= 0 \\ \boldsymbol{A}_1\cdot\boldsymbol{a}_2 &= 0 \\ \boldsymbol{A}_2\cdot\boldsymbol{a}_2 &= 2\pi \end{aligned} \tag{4.10.144}$$

Diese Bedingungen werden erfüllt wenn $\boldsymbol{A}_1$ und $\boldsymbol{A}_2$ wie folgt konstruiert werden:

$$\boldsymbol{A}_1 = 2\pi\frac{\boldsymbol{a}_2\times\boldsymbol{n}}{\boldsymbol{a}_1(\boldsymbol{a}_2\times\boldsymbol{n})} \tag{4.10.145}$$

und

$$\boldsymbol{A}_2 = 2\pi\frac{\boldsymbol{n}\times\boldsymbol{a}_1}{\boldsymbol{a}_1(\boldsymbol{a}_2\times\boldsymbol{n})} \tag{4.10.146}$$

Dabei ist $\boldsymbol{n}$ ein beliebiger Vektor senkrecht zum Oberflächennetz

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

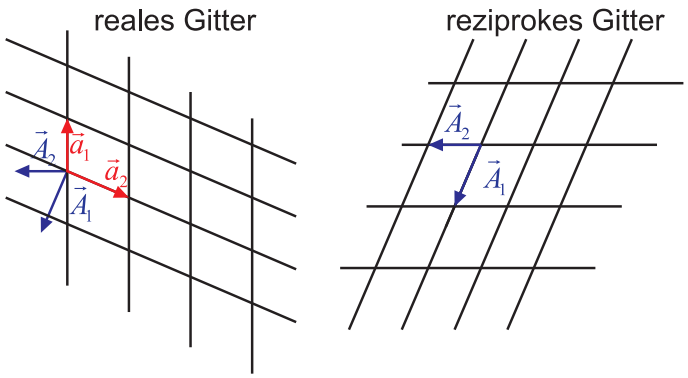



Abbildung 4.388.: Reales Gitter (links) und reziprokes Gitter (rechts).

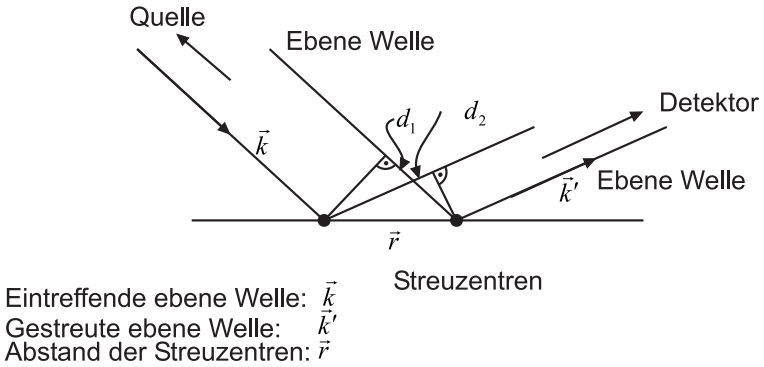



Abbildung 4.389.: Skizze zur Streuung an Oberflächenatomen

### 4.10.16.2. Streuung (Beugung) an Oberflächen

Abbildung 4.389 zeigt die Geometrie der Streuung. Die Einfallende ebene Welle wird mit $\boldsymbol{k}$ und die gestreute ebene Welle mit $\boldsymbol{k}'$ bezeichnet. Der Abstand der Streuzentren sei $\boldsymbol{r}$.
Die Wegdifferenzen der Wellenzüge zwischen zwei benachbarten Streuzentren sind

$$\begin{aligned} d_1 &= |\boldsymbol{r}| \cdot \cos(\boldsymbol{r},\boldsymbol{k}) = |\boldsymbol{r}| \cdot \frac{\boldsymbol{r}\cdot\boldsymbol{k}}{|\boldsymbol{r}|\,|\boldsymbol{k}|} = \frac{\boldsymbol{r}\cdot\boldsymbol{k}}{|\boldsymbol{k}|} \\ d_2 &= |\boldsymbol{r}| \cdot \cos(\boldsymbol{r},\boldsymbol{k}') = \frac{\boldsymbol{r}\cdot\boldsymbol{k}'}{|\boldsymbol{k}'|} \end{aligned} \tag{4.10.147}$$

Aus dem Wegunterschied berechnet man die Phasendifferenzen zu

$$\begin{aligned} \phi &= |\boldsymbol{k}| \cdot d_1 = \boldsymbol{r}\cdot\boldsymbol{k} \\ \phi' &= |\boldsymbol{k}'| \cdot d_2 = \boldsymbol{r}\cdot\boldsymbol{k}' \end{aligned} \tag{4.10.148}$$

Die Phasendifferenz ist

$$\Delta\phi = \phi - \phi' = \boldsymbol{r}\cdot\boldsymbol{k} - \boldsymbol{r}\cdot\boldsymbol{k}' = -\Delta\boldsymbol{k}\cdot\boldsymbol{r} \tag{4.10.149}$$

mit $\Delta\boldsymbol{k} = \boldsymbol{k}' - \boldsymbol{k}$. Für die Amplituden gilt $\psi' = \psi e^{i\Delta\boldsymbol{k}_i\cdot r}$ für das $i$-te Atom.

Für die Beträge der Wellenvektoren gilt

$$k = \frac{2\pi}{\lambda} = \frac{p}{h} \tag{4.10.150}$$

mit der de Broglie–Wellenlänge $\lambda$ und dem Impuls der Teilchen $p$.
Für die Streuamplitude eines Netzes mit monoatomarer Basis erhält man:

$$\psi = \sum_{\boldsymbol{T}} e^{i\Delta \boldsymbol{k}\boldsymbol{T}} \tag{4.10.151}$$

mit $\boldsymbol{T} = v \cdot \boldsymbol{a}_1 + w \cdot \boldsymbol{a}_2$. Für eine mehratomige Basis erhält man:

$$\psi = \left( \sum_{\boldsymbol{T}} e^{-i\Delta \boldsymbol{k} \cdot \boldsymbol{T}} \right) \cdot \left( \sum_j f_j e^{-i\Delta \boldsymbol{k} \cdot \boldsymbol{r}_j} \right) \tag{4.10.152}$$

$f_j$ ist der Streufaktor des $j$-ten Streuzentrums und $\boldsymbol{r}_j$ ist die Position dieses Streuzentrums in der Einheitszelle. Der erste Faktor in der Gleichung (4.10.152) hängt nur vom Oberflächennetz ab und nicht von der Struktur der Einheitszelle. Dieser Faktor wird Gittersumme

$$G_{\Delta \boldsymbol{k}} = \sum_{\boldsymbol{T}} e^{-i\Delta \boldsymbol{k} \cdot \boldsymbol{T}} \tag{4.10.153}$$

genannt. Der zweite Faktor in Gleichung (4.10.152) ist die geometrische Strukturamplitude

$$G_{\Delta \boldsymbol{k}} = \sum_j f_j e^{-i\Delta \boldsymbol{k} \cdot \boldsymbol{r}_j} \tag{4.10.154}$$

Da $\boldsymbol{T}$ in der Oberfläche liegt, ist

$$\Delta \boldsymbol{k} \cdot \boldsymbol{T} = \left( \Delta \boldsymbol{k}_\perp + \Delta \boldsymbol{k}_{||} \right) \cdot \boldsymbol{T} = \Delta \boldsymbol{k}_{||} \cdot \boldsymbol{T} \tag{4.10.155}$$

Also ist die Laue-Bedingung

$$\begin{aligned} \Delta \boldsymbol{k}_{||} \cdot \boldsymbol{a}_1 &= 2\pi h \\ \Delta \boldsymbol{k}_{||} \cdot \boldsymbol{a}_2 &= 2\pi k \end{aligned} \tag{4.10.156}$$

Bei elastischer Streuung gilt

$$E = E' \Rightarrow \boldsymbol{k}^2 = \boldsymbol{k}'2 \quad \text{oder} \quad |\boldsymbol{k}| = |\boldsymbol{k}'| \tag{4.10.157}$$

Aus dieser Bedingung kann man die in Abb. 4.390 gezeigte Ewald-Konstruktion für Oberflächennetze ableiten.

### 4.10.16.3. LEED

LEED[?] ist die am häufigsten angewandte Methode zur strukturellen Untersuchung periodischer Kristalloberflächen. Die Elektronen werden mit einer bestimmten, möglichst monochromatischen Energie aus einer wohldefinierten Richtung auf

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

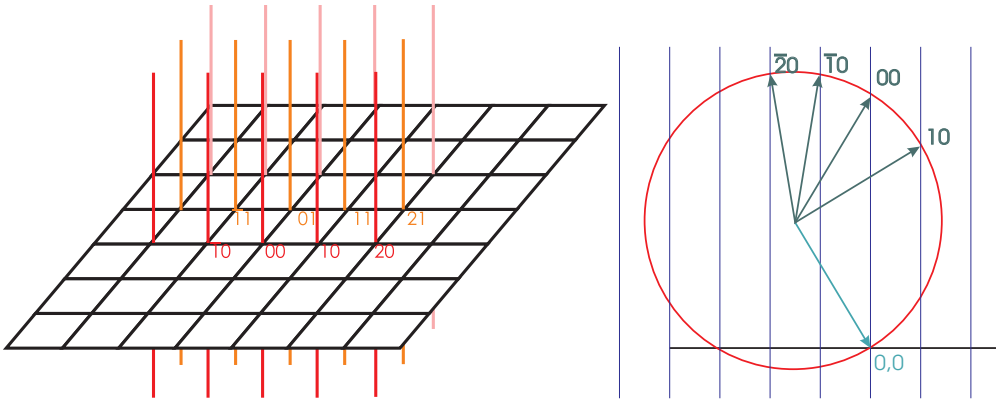



Abbildung 4.390.: Ewald-Konstruktion für Oberflächennetze. Rechts wird ein Schnitt dargestellt.

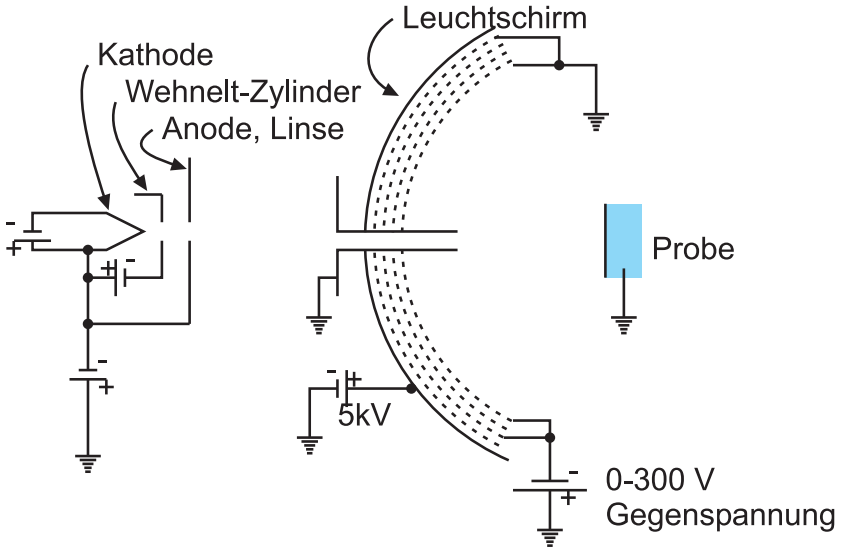



Abbildung 4.391.: Aufbau eines LEED. Links ist die Elektronenkanone gezeigt. Rechts ist der schematische Aufbau des LEED-Schirms gezeigt.

die Probe gesandt. Ihre de Broglie-Wellenlänge muss von der gleichen Grössenordnung wie die Gitterperiode an der Kristalloberfläche sein. Wenn man eine Periodizität von $0.1nm$ annimmt, so ergibt sich

$$0.1nm = \lambda = \frac{h}{\sqrt{2mE}} \tag{4.10.158}$$

Daraus folgt für die Energie

$$E = \frac{h^2}{2m\lambda^2} = \frac{(6.6 \cdot 10^{-34})^2}{2 \cdot 9.1 \cdot 10^{-31} \cdot 10^{-20}} \approx 100eV \tag{4.10.159}$$

Abbildung 4.391 zeigt den Aufbau eines LEED. Die Elektronen stammen in der Regel aus einer thermischen Kathode, wie sie in der Abb. 4.327 gezeigt ist. Der von den Elektronen bei der Glühemission durchquerte Potentialverlauf ist in der Abb. 4.332 gezeigt. Nach der Beschleunigungsphase bewegen sich die Elektronen in einem feldfreien Raum bis zur Probe. Die rückgestreuten Elektronen nähern sich dem mit einer phosphoreszierenden Substanz belegten kugelkalottenförmigen Schirm in

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

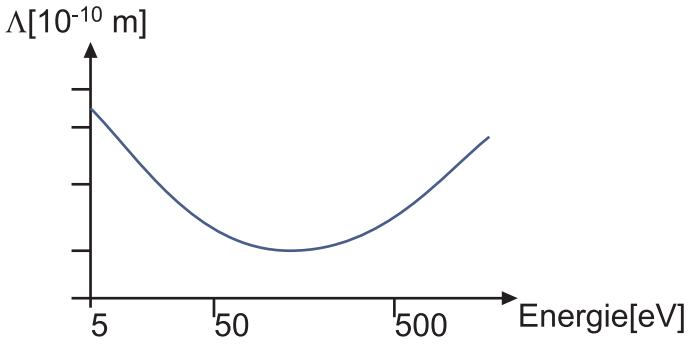



Abbildung 4.393.: Eindringtiefe der Elektronen als Funktion der Energie

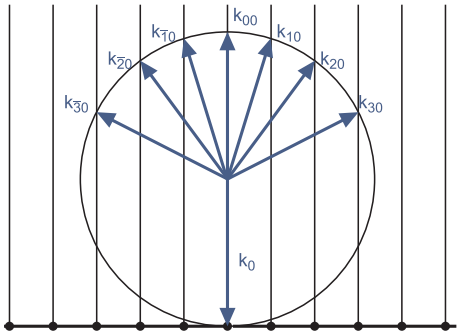



Abbildung 4.394.: Ewaldkonstruktion für LEED

einem feldfreien Raum. Der Energieverlauf im LEED-Detektor ist schliesslich in der Abb. 4.392 gezeigt.

Die Energieunschärfe bei der Emission muss mit der thermischen Energie bei Raumtemperatur verglichen werden. Diese ist $\Delta E \approx kT \approx \frac{1}{40}eV$. Die Glühemission bei $T = 2000K$ ist mit einer Energieunschärfe von $\Delta E \approx 0.2eV$ behaftet und damit etwa acht mal grösser als $kT$ bei Raumtemperatur. Die Energieunschärfe der Feldemission bei $T = 300K$ ist schliesslich gleich der thermischen Energie $kT$, also $\Delta E \approx 0.025eV$.

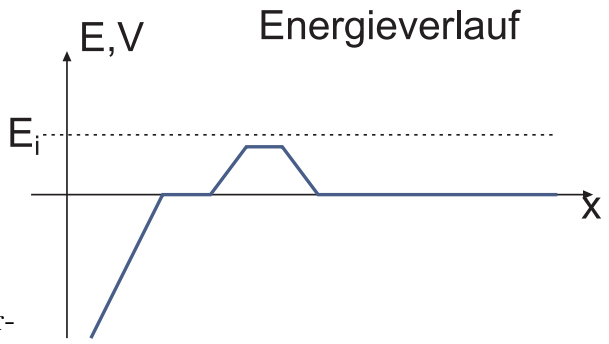



Abbildung 4.392.: Energieverlauf im LEED-Detektor. Rechts ist der Zwischenraum zwischen der Probe und dem Detektor.

Abbildung 4.393 zeigt die Eindringtiefe der Elektronen als Funktion ihrer kinetischen Energie. Die Eindringtiefe ist für Elektronen mit einer Energie von etwa $100eV$ minimal. Bei höheren Energien, wie sie zum Beispiel bei RHEED (siehe Abschnitt 4.10.16.4) oder bei der Elektronenmikroskopie (siehe Abschnitt 4.10.14) vorkommen ist die Eindringtiefe grösser. Sie nimmt über etwa $500eV$ monoton mit der kinetischen Energie der Elektronen zu.

Für LEED verwendet man Elektronen mit einer kinetischen Energie von $20 - 500eV$. Die Eindringtiefe der Elektronen ist entsprechend kleiner als einen Nanometer.

Das durch die Wechselwirkung der langsamen Elektronen mit der Probe entstehende Beugungsbild kann mit Hilfe der *Ewald-Konstruktion* nach Abb. 4.394 in-

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

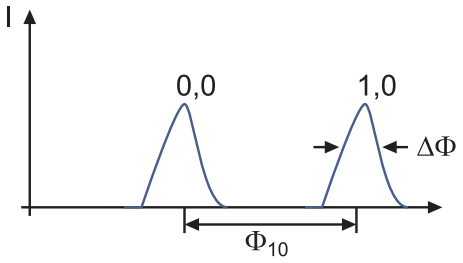



Abbildung 4.395.: Beugungsmuster und Definitionen zur Transferweite

terpretiert werden.
Zwischen der periodischen Struktur der Probenoberfläche oder einer eventuell vorhandenen Überstruktur und der Überstruktur im reziproken Raum besteht folgender Zusammenhang:

$$\text{reeller Raum} \quad \boldsymbol{b} = S \cdot \boldsymbol{a} \tag{4.10.160}$$

$$\text{reziproker Raum} \quad \boldsymbol{B} = \left(S^T\right)^{-1} \boldsymbol{A} = S_{rez} \cdot \boldsymbol{A}$$

$$\boldsymbol{A} = \left(S^T\right) \cdot \boldsymbol{B} \tag{4.10.161}$$

Hier ist nach dem Anhang L $S$ die die Struktur der Oberfläche charakterisierende Matrix. Nach der Gleichung (4.10.161) kennt man mit $S_{rez}$ auch $S$.
Damit Beugungseffekte in der Abbildung mit Elektronen beobachtet werden können, muss die Kohärenzlänge der Elektronen grösser als die maximal möglichen Wegunterschiede sein. Wie bei Licht müssen zwei Arten von Kohärenz unterschieden werden.

**Zeitliche Kohärenz** ist gegeben durch die Energieunschärfe.

**Räumliche Kohärenz** ist gegeben durch die Ausdehnung der Elektronenquelle (dominant)

Mit der Transferweite $t$ (für die Definitionen siehe Abb. 4.395) bezeichnet man die Breite des Elektronenstrahls, die bei perfekter Quelle und perfekter Abbildung die gleiche Breite der Leuchtflächen bewirkt wie der Elektronenstrahl im realen LEED. Sie ist gegeben durch

$$t = a\frac{\phi_{10}}{\Delta\phi} \tag{4.10.162}$$

Damit wird $t \approx 10nm$. Da Elektronen eine sehr kleine Kohärenzlänge haben und da sie als Fermionen nicht im gleichen Quantenzustand sein können[28] kann jedes Elektron nur mit sich selber interferieren.

[28] Das bedeutet für freie Elektronen, dass sich keine zwei Elektronen am gleichen Ort aufhalten können.

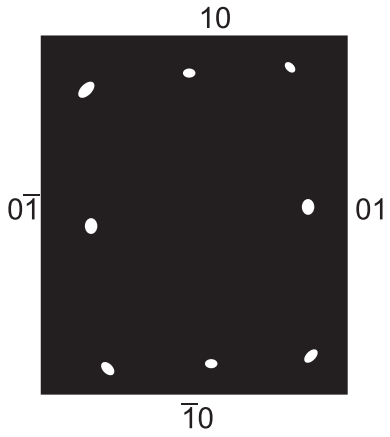



Abbildung 4.396.: Schematische Skizze eines LEED-Bildes von $Cu(110)$ (gezeichnet nach [**?**]). Dies ist eine FCC-Struktur.

Abbildung 4.397.: Schematische Skizze eines LEED-Bildes von Ni (111) bei einer Primärenergie von 205 eV (gezeichnet nach [**?**]).

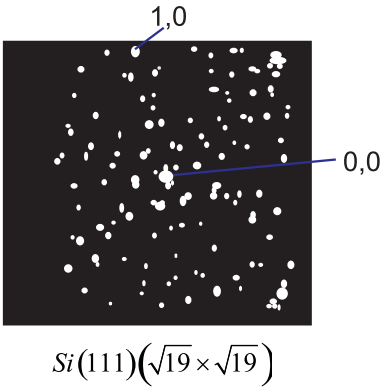



Abbildung 4.398.: LEED-Bild von *Si(111)* $\sqrt{19} \times \sqrt{19}$

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

**4.10.16.3.1. Nicht ideale Oberflächen** Bei einer endlich ausgedehnten Probe treten neue Effekte auf. Wir beschreiben die Struktur der Probenoberfläche mit der folgenden Gleichung

$$T = \boldsymbol{a}_1 v + \boldsymbol{a}_2 w \tag{4.10.163}$$

Hier sind $\boldsymbol{a}_1$ und $\boldsymbol{a}_2$ die Vektoren, die die Einheitszelle aufspannen. $v$ und $w$ sind ganze Zahlen, die die einzelnen Einheitszellen adressieren. Sie können auch als Vektor

$$\boldsymbol{v} = \begin{pmatrix} v \\ w \end{pmatrix} \tag{4.10.164}$$

geschrieben werden. Die Bewegung der Elektronen wird durch ihren k-Vektor beschrieben.

$$\Delta\boldsymbol{k} = \begin{pmatrix} \Delta k_1 \\ \Delta k_2 \end{pmatrix} \tag{4.10.165}$$

Die Wellenfunktion der gestreuten Elektronen ist gegeben durch

$$\psi_{\Delta k} \sum_{v=1}^{V} \sum_{w=1}^{W} e^{i(\Delta k_1 a_1 v + \Delta k_2 a_2 w)} = \sum_j f_j e^{i\Delta\boldsymbol{k}\boldsymbol{v}j} \tag{4.10.166}$$

Die obigen Gleichungen können für die $v$- und die $w$-Komponente einzeln umgeschrieben werden. Man erhält

$$\begin{aligned}
\sum_{v=1}^{V} e^{i\Delta k_1 a_1 v} &= \sum_{v=1}^{V} \left(e^{i\Delta k_1} a_1\right)^v \\
&= e^{i\Delta k_1 a_1} \frac{e^{i\Delta k_1 V a_1} - 1}{e^{i\Delta k_1 a_1} - 1} \\
&= e^{\frac{i\Delta k_1 a_1}{2}} e^{\frac{i\Delta k_1 a_1 V}{2}} \frac{e^{\frac{i\Delta k_1 a_1 V}{2}} - e^{-\frac{i\Delta k_1 a_1 V}{2}}}{e^{\frac{i\Delta k_1 a_1}{2}} - e^{\frac{-i\Delta k_1 a_1}{2}}} \\
&= e^{\frac{i\Delta k_1 a_1}{2}(V+1)} \frac{\sin\left(\frac{k_1 a_1 V}{2}\right)}{\sin\left(\frac{k_1 a_1}{2}\right)}
\end{aligned} \tag{4.10.167}$$

Diese Summe ist nur $\neq 0$ wenn $k_1 a_1$ ein Vielfaches von $\pi$ ist. Dann ist nach dem Satz von l'Hôspital $\sin\left(\frac{k_1 a_1 V}{2}\right) = V$. Der Grenzwert für $V \to \infty$ ist die $\delta$–Funktion. Daraus berechnet sich das Betragsquadrat der Elektronenwellenfunktion zu

$$|\psi_{\Delta k}|^2 \approx \frac{\sin^2 \Delta k_1 a_1 V}{\sin^2 \Delta k_1 a_1} \frac{\sin^2 \Delta k_2 a_2 W}{\sin^2 \Delta k_2 a_2} |F|^2 \tag{4.10.168}$$

wobei $|F|$ der Strukturfaktor aus $\sum_j f_j e^{i\Delta\boldsymbol{k}\boldsymbol{v}j}$ ist. Der Strukturfaktor beinhaltet die Information über die Struktur der Einheitszelle.
Die Breite der Beugungsreflexe ist proportional zu $\frac{1}{V}$ beziehungsweise zu $\frac{1}{W}$. Diese Aussage folgt aus der Tatsache, dass die erste Nullstelle von $\sin(k_1 a_1 V)$ neben $k_1 a_1 = n\pi$ bei $k_1 a_1 - \frac{k_1 a_1 \pm \frac{\pi}{2}}{V}$ liegt. Deshalb ist die Breite eines Reflexes $\frac{\pi}{V}$.

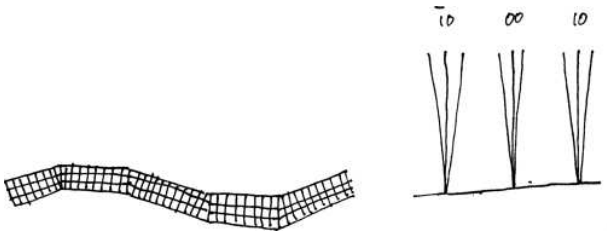



Abbildung 4.399.: Mosaikstruktur links im realen Raum und rechts im reziproken Raum

**4.10.16.3.2. Domänen** Bei vielen asymmetrischen Überstrukturen ist durch das Substrat keine Vorzugsrichtung vorgegeben. Deshalb können Domänen der Überstruktur mit unterschiedlicher Orientierung gleichzeitig auftreten.

**4.10.16.3.3. Mosaikstruktur** Man spricht von einer Mosaikstruktur, wenn gleichzeitig mehrere leicht gegeneinander verkippte Gitter vorliegen.

**4.10.16.3.4. Paarkorrelation und LEED-Bilder** Wir wandeln die Gittersumme in ein Integral um:

$$\psi_{\Delta k} = \sum_i e^{i\Delta \boldsymbol{k} \boldsymbol{r}_i} = \int_S P(\boldsymbol{r}) e^{i\Delta \boldsymbol{k} \boldsymbol{r}} d\boldsymbol{r} \tag{4.10.169}$$

mit

$$P(\boldsymbol{r}) = \sum_i \delta(\boldsymbol{r} - \boldsymbol{r}_i) \tag{4.10.170}$$

also $\psi_{\Delta \boldsymbol{k}}$ ist die Fouriertransformierte von $P(\boldsymbol{r})$
Relationen für Fouriertransformationen:

$$h(x) = \int_{-\infty}^{+\infty} f(t) g(x-t)\, dt \tag{4.10.171}$$

und

$$F(h(x)) = \quad F(f(x)) \bullet F(g(x)) \tag{4.10.172}$$

einer Faltung entspricht das Produkt der *Fouriertransformation*. Wenn wir in der obigen Gleichung $g(x) = f^*(x)$ setzen, dann wird die Autokorrelation

$$\phi(x) = \int_{-\infty}^{+\infty} f(t) f^*(x-f)\, df \tag{4.10.173}$$

Damit wird das Leistungsspektrum

$$F(\phi(x)) = |F(f(x))|^2 \tag{4.10.174}$$

Gemessen werden in einem LEED-Experiment oder, allgemeiner, in einem beliebigen Streuexperiment die Intensitäten

$$I = |\psi|^2 = |F(P(\boldsymbol{v}))|^2 = F(\phi(\boldsymbol{r})) \tag{4.10.175}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

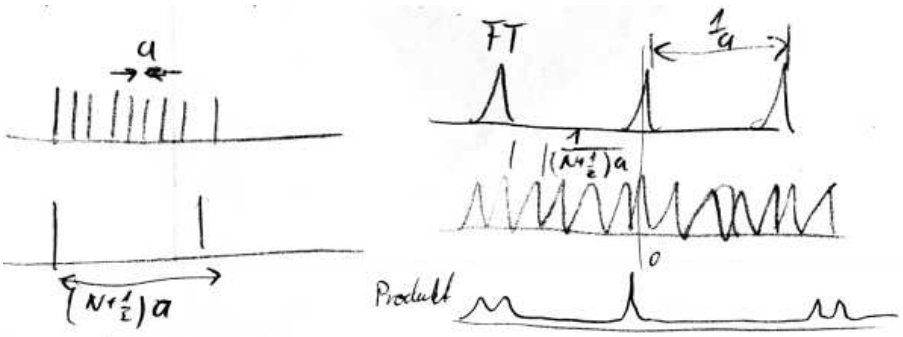



Abbildung 4.400.: Beispiel: Regelmässige Domänen einer Kristallstruktur mit der Periode $a$ jeweils im Abstand von $(N + 1/2)a$

Nach der Gleichung (4.10.175) ist die gemessene Intensität proportional zur Paarkorrelation der Streuzentren. Durch eine Theorie lässt sich aus der Paarkorrelationsfunktion die Intensitätsverteilung berechnen. Die Umkehrung ist aber nicht möglich, da bei der Intensitätsmessung die Phase verloren geht (Phasenproblem). Bei einer komplexen Abbildung ist die *Fouriertransformation* der Paarkorrelationsfunktion der gesamten Abbildung das Produkt der Fouriertransformationen der Paarkorrelationsfunktionen der einzelnen Schritte der Abbildung. Dies entspricht im realen Raum dies einer Faltung. Abbildung 4.400 zeigt als Beispiel die Abbildung mit mehreren Domänen. Dabei ist ersichtlich, dass jeweils jeder zweite Spot ist aufgespalten ist. Sind die Domänen nicht streng periodisch angeordnet, überlagern sich alle möglichen Verbreiterungen. Die LEED-Spots sind dann verbreitert. Mit dem gleichen Formalismus lassen sich die Beugungsmuster gestufter Oberflächen berechnen.

**4.10.16.3.5. Gittergase** Nichtperiodische Oberflächen können durch Gittergase beschrieben werden. Ein Gittergas ist ein Gas, dessen Teilchen sich statistisch verteilt auf Gitterplätzen aufhalten. Wenn nun eine Welle an dieser statistischen Oberfläche gestreut wird, dann ist die gestreute Intensität

$$I_{\Delta \boldsymbol{k}} = |\psi_{\Delta k}|^2 = \left| \sum_{n\ Atome} F_n e^{i\Delta \boldsymbol{k} \boldsymbol{r}} \right|^2 = \sum_{m,n} F_m\, F_n^* e^{i\Delta \boldsymbol{k}(\boldsymbol{r}_m - \boldsymbol{r}_n)} \tag{4.10.176}$$

Anders als bei periodischen Anordnungen muss jedem Atom eine eigene Streuamplitude angenommen werden. Wenn man statistische Unabhängigkeit annimmt, dann gilt für $m \neq n$

$$\langle F_m\, F_n^* \rangle \approx (\langle F \rangle)^2 \tag{4.10.177}$$

da keine Korrelation zwischen den einzelnen Faktoren herrscht. Für den Fall $m = n$ herrscht jedoch eine strenge Korrelation. Deshalb muss zuerst quadriert und dann erst der Mittelwert berechnet werden.

$$\langle F_m\, F_n^* \rangle = \langle F \cdot F \rangle = \left\langle F^2 \right\rangle \tag{4.10.178}$$

Damit wird

$$\langle F_m\, F_n^* \rangle = \langle F \rangle^2 + \delta_{m,n} \left( \left\langle F^2 \right\rangle - \langle F \rangle^2 \right) \tag{4.10.179}$$

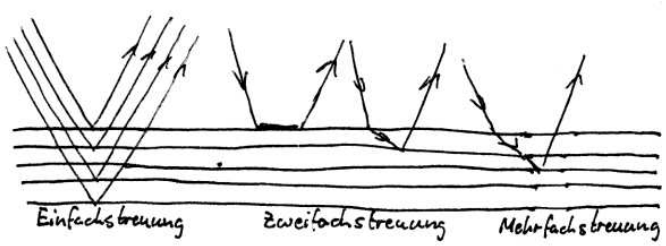



Abbildung 4.401.: Mögliche Wege der Elektronen bei der Streuung an Oberflächen.

Eingesetzt in Gleichung (4.10.176) erhält man

$$\begin{aligned} I_{\Delta \boldsymbol{k}} &= \sum_{m,n} \left( \langle F \rangle^2 + \delta_{m,n} \left( \left\langle F^2 \right\rangle - \langle F \rangle^2 \right) \right) e^{i \Delta \boldsymbol{k} (\boldsymbol{r}_m - \boldsymbol{r}_m)} \\ &= \underbrace{N \left( \left\langle F^2 \right\rangle - \langle F \rangle^2 \right)}_{\text{unabhängig von } \Delta \boldsymbol{k}} + \langle F \rangle^2 \underbrace{\sum_{m,n} e^{i \Delta \boldsymbol{k} (\boldsymbol{v}_m - \boldsymbol{v}_n)}}_{Gittersumme} \end{aligned} \tag{4.10.180}$$

Aus Gleichung (4.10.180) ist ersichtlich, dass statistisch verteilte Streuzentren einen konstanten Untergrund bilden. Sie verbreitern die Reflexe jedoch nicht.

**4.10.16.3.6. Abhängigkeit des LEED-Bildes von der Elektronenenergie** Für Elektronen sind die Intensitäten der Reflexe abhängig von der Elektronenenergie eV, vom Einfallswinkel $\varphi$, von d und von $\boldsymbol{a}_1$ und $\boldsymbol{a}_2$. Bei der Berechnung der Spannungsabhängigkeit muss der Einfluss des Strukturfaktors berücksichtigt werden. Bei einer Änderung der Spannung werden mehr oder weniger Netzebenen in der Tiefe beteiligt. Immer dann wenn die Elektronenenergie oder die Beschleunigungsspannung so ist, dass eine ganze Anzahl Netzebenen berücksichtigt werden, werden die Strukturfaktoren der beteiligten Atome in allen Tiefen phasenrichtig addiert. In allen anderen Fällen mittelt sich die gestreute Amplitude mehr oder weniger aus.
Wird die Intensität gegen die Beschleunigungsspannung aufgetragen und nicht gegen den Streuvektor, dann ergeben sich scharfe Maxima für alle Reflexe. Diese Maxima hängen vom Schichtabstand $d$ und, ausser beim $0,0$-Reflex, auch vom Einfallswinkel ab.
Für den $0,0$-Reflex gilt die Bragg-Bedingung für die folgenden Spannungen.

$$V = \frac{Mn^2}{4d^2 \cos^2 \varphi} \tag{4.10.181}$$

mit

$$M = \frac{h^2}{2em} = 1.5 V n m^2 = 150 V \text{Å}^2 = 1.5 \times 10^{-18} V m^2 \tag{4.10.182}$$

Für Vielfache dieser Spannungen treten Intensitätsmaxima auf. Durch Messung von $I_{\Delta \boldsymbol{k}}(V)$ kann die Wechselwirkung der Elektronen mit der Probe bestimmt werden. Dies wird auch I-V-Messung genannt.
Die oben vorgestellte einfache Rechnung kann mit realistischen Potentialen verbessert werden.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

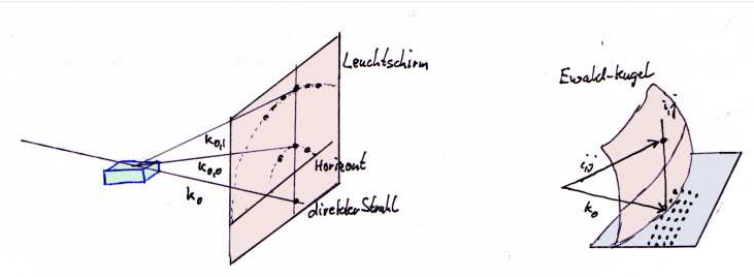



Abbildung 4.402.: Geometrie bei der RHEED-Abbildung. Links ist dargestellt, wie die Trajektorien der Elektronen angeordnet sind. Rechts ist gezeigt, wie die Ewald-Konstruktion zum Auffinden der RHEED-Reflexe angewandt werden.

#### 4.10.16.4. RHEED: Reflection high energy electron diffraction

Langsame Elektronen werden durch kleine elektrische und magnetische Streufelder abgelenkt. Elektronen höherer Energie zeigen wegen der kürzeren Wechselwirkungszeit mit den Störfeldern weniger Empfindlichkeit. Zudem ist bei höheren Elektronenenergien die relative Energieunschärfe kleiner.
Die Geschwindigkeitskomponente der Elektronen senkrecht zur Oberfläche muss im LEED-Bereich (20-500 eV) sein. Die Energien sind

$$E_{LEED} = \frac{1}{2} m v_{LEED}^2 \tag{4.10.183}$$

und

$$E_{RHEED} = \frac{1}{2} m v_{RHEED}^2 \tag{4.10.184}$$

Damit kann man die für eine Streuung benötigte Geschwindigkeitskomponente senkrecht zur Probenoberfläche berechnen.

$$V_{rhead}^2 = V_{||}^2 + V_{\perp}^2 \quad \text{mit} \quad V_{\perp} = V_{LEED} \tag{4.10.185}$$

Die möglichen Einfallswinkel sind also

$$\alpha \approx \frac{V_{\perp}}{V_{RHEED}} = \frac{\sqrt{\frac{2E_{LEED}}{m}}}{\sqrt{\frac{2E_{RHEED}}{m}}} = \sqrt{\frac{E_{LEED}}{E_{RHEED}}} \tag{4.10.186}$$

Wenn man typische Energien einsetzt wird der Einfallswinkel einer RHEED-Apparatur

$$\alpha \approx \sqrt{\frac{100eV}{10keV}} = \frac{1}{10} = 5.7^{\circ} \tag{4.10.187}$$

Die bei RHEED Reflexe liegen auf Kreislinien. Bei der RHEED-Abbildung erzeugen Defekte zigarrenförmige Reflexe. Der Spiegelreflex (das ist der $0,0$-Reflex) zeigt Intensitätsmodulationen abhängig von der Oberflächenrauhigkeit.
Ein grosser Vorteil der RHEED-Abbildung ist der flache Einfallswinkel der Elektronen. Deshalb ist fast der ganze Halbraum gegenüber der Probenoberfläche frei. Die Probenoberfläche ist für parallele laufende Experimente zugänglich. Typischerweise wird RHEED zur Prozesskontrolle bei Wachstumsprozessen verwendet. Abb. 4.404 zeigt links eine schematische Darstellung der Kristallstruktur und rechts den

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

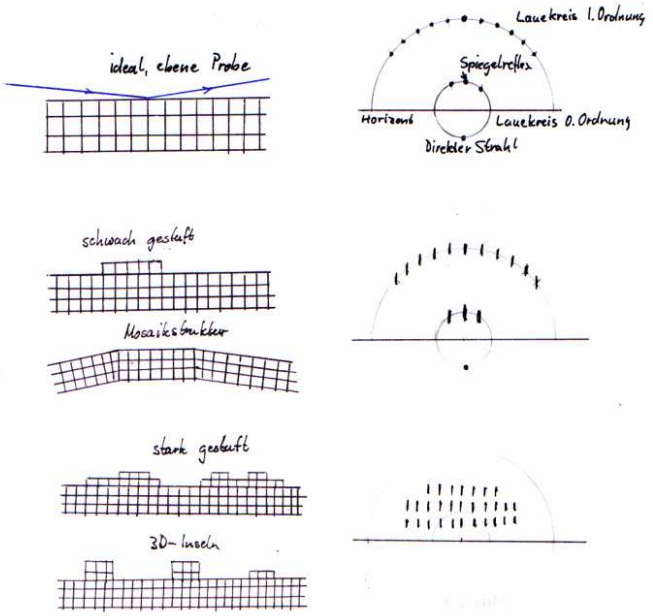



Abbildung 4.403.: Konstruktion der Reflexe bei RHEED. Links werden die Strukturen der Oberflächen gezeigt. Rechts ist das entsprechende RHEED-Bild. Von oben nach unten wird die Abbildung bei idealer Oberfläche, bei schwach gestuften Oberflächen, bei einer Mosaikstruktur, stark gestufte Oberflächen sowie 3D-Inseln

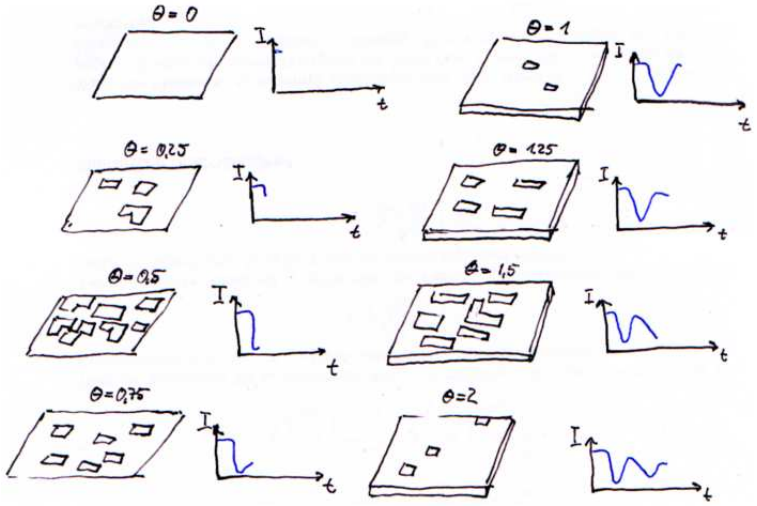



Abbildung 4.404.: RHEED-Kontrolle des Schichtwachstums zum Beispiel bei der Herstellung von Halbleiter-Quantenschichten.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

zeitlichen Verlauf der Intensität des Spiegelreflexes. Die Intensität bei einem perfekten Kristall ist maximal. Wenn das Schichtwachstum startet, dann sinkt die Intensität bis sie bei der halben Bedeckung minimal wird. Wenn eine Monolage voollständig abgeschieden ist, wird wieder ein Maximum erreicht. Dieses ist aber weniger hoch, da mit dem Kristallwachstum immer auch Defekte eingebaut werden.
Im Gegensatz zu LEED ist das reziproke Gitter aus RHEED-Bildern sehr viel schwieriger zu bestimmen.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

## 4.10.17. Scanning Force Microscopy

Scanning Force Microscopy (*SFM*)[**?**] was an early offspring of Scanning Tunneling Microscopy. The force between a tip and the sample was used to image the surface topography. The force between the tip and the sample, also called the tracking force, was lowered by several orders of magnitude compared to the profilometer[**?**]. The contact area between the tip and the sample too was reduced considerably. The force resolution was similar to that achieved in the *Surface Force Apparatus*[**?**]. Soon thereafter atomic resolution in air was demonstrated[**?**]. Marti *et al.*[**?**] demonstrated a *SFM* capable of atomic resolution under liquids. Kirk *et al.*[**?**] operated an *SFM* successfully at 4.2 K, the temperature of liquid helium. The *SFM* measures either the contours of constant forces or force gradients or the variation of forces or force gradients with position, when the height of the sample is not adjusted by a feedback loop. These measurement modes are similar to the ones of the *STM*, where contours of constant tunneling current or the variation of the tunneling current with position at fixed sample height are recorded.
The invention of the *SFM* demonstrated that forces can play an important role in other scanning probe techniques.
The type of force interaction between the tip and the sample surface is used to characterize SFMs. The highest resolution is achieved when the tip is pressed against the sample surface, the so called repulsive or contact mode. The forces in this mode basically stem from the Pauli exclusion principle which prevents the spatial overlap of electrons. As in the *STM*, the force applied to the sample can be constant, the so called constant force mode. If the sample $z$-position is not adjusted to the varying force, we speak of the constant $z$-mode. However for weak cantilevers (0.01 N/m spring constant) and a static applied load of $10^{-8}$ N we get a static deflection of $10^{-6}$ m, which means that even structures of several nanometers height will be subject to an almost constant force, whether it is controlled or not. Hence for the contact mode with soft cantilevers the distinction between constant force mode and constant $z$-mode is rather arbitrary. Additional information on the sample surface can be gained by measuring lateral forces (friction mode) or modulating the force to get $\frac{dF}{dz}$ which is nothing else than the stiffness of the surfaces. When using attractive forces one normally measures also $\frac{dF}{dz}$ with a modulation technique. In the attractive mode the lateral resolution is at least one order of magnitude worse than for the contact mode. The attractive mode is also referred to as the non-contact mode. Of widespread use is also the magnetic force microscope, another non-contact microscope.
We will first give an introduction to the theory of force microscopy. A discussion of the techniques to measure small forces follows. We will conclude this chapter with the discussion of selected experiments with inorganic samples.

### 4.10.17.1. Theory of Force Microscopy

There have been but few theoretical papers on the interaction between a tip and a sample surface. Investigations of the interaction of the tip of a *SFM* with graphite[**?**, **?**, **?**, **?**, **?**], of the imaging of stepped surfaces[**?**], of attractive mode *SFM*[**?**] and of Magnetic Force Microscopy[**?**] are published in the literature.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

### 4.10.17.2. The Interaction of a Tip and the Sample

The forces in *SFM* in the absence of added magnetic or electrostatic potentials are governed by the interaction potentials between atoms. The interaction potential between two atoms usually has the form outlined in figure 4.405. The interaction is attractive at large distances due to the van-der-Waals interaction. At short distances the repulsive forces have their origin in the quantum mechanical exclusion principle, which states that no two fermions can be in exactly the same state, that is to say have the same spin, angular momentum, z-component of the angular momentum and location.

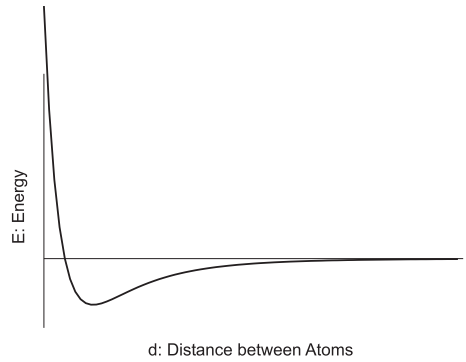



Abbildung 4.405.: Qualitative curve for the interaction potential between two atoms. This curve has been calculated using a Lennard-Jones potential.

The theoretical treatment of the force interaction between a sample and a tip is usually very complicated and requires large parallel computers. As an example we outline a simple model for the imaging of graphite by repulsive forces. Graphite is used because its potentials are well known and its structrue is sufficiently simple. The theoretical treatment of the imaging of graphite using repulsive forces follows the treatment of Gould *et al.*[**?**]. The tip is assumed to consist of one to a few atoms. Its atoms are assumed to be rigid with respect to one another. The force between the tip atoms and the surface atoms can be repulsive or attractive, depending on the distance between two atoms. The total potential is assumed to be the sum of two-body potentials. We first start with a single atom tip. The interaction is then given by

$$U(\boldsymbol{r}, \boldsymbol{r}_1, \ldots, \boldsymbol{r}_N) = \sum_i V(\boldsymbol{r} - \boldsymbol{r}_i) + \bar{V}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N) \tag{4.10.188}$$

where $V(\boldsymbol{r} - \boldsymbol{r}_i)$ is the interaction potential between the tip atom and the i-th atom of the surface and $\bar{V}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$ is the many-body potential in the absence of the tip. The vectors $\boldsymbol{r}_i$ denote the position of the i-th atom in the sample. $\boldsymbol{r}$ is the position of the tip.
Gould *et al.*[**?**] considered only interactions between an atom and its nearest four neighbors. They furthermore assumed, that the distortion of the lattice under the influence of the tip were small so that the harmonic approximation was valid:

$$\bar{V}(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N) = \frac{1}{2} \sum_{i,j,\mu,\nu} u_\mu^{(i)} D_{\mu\nu}^{(i,j)} u_\nu^{(j)}, \tag{4.10.189}$$

where $u_\mu^{(i)}$ is the $\mu$th cartesian component of the displacement from equilibrium. The $D_{\mu\nu}^{(i,j)}$ is the matrix of force constants in a solid. The interaction of the tip atom with the graphite is modeled with a Lennard-Jones potential

$$V(|\boldsymbol{r}_{tip} - \boldsymbol{r}_{atom}|) = V_0 \left[\frac{1}{2}(r_0/r)^{12} - (r_0/r)^6\right] \tag{4.10.190}$$

The interaction potential between atoms does not necessarily have the form of the Lennard-Jones potential. Other potentials could be used as well. Gould *et al.*[**?**] have chosen this interaction potential because of the computational simplicity. They selected the constants $V_0 = 2.8 \times 10^{-21}$ J and $r_0 = 0.28$ nm to reproduce the measured corrugation and to obtain the correct tip-sample spacing, as calculated with the theories of Soler *et al.*[**?**] and Batra and Çiraci[**?**].
To calculate the measured topography, they positioned the tip over various locations in the unit cell and let the surface relax according to

$$\frac{\partial U}{\partial \boldsymbol{r}_i} = 0, i = 1, \ldots, N, \tag{4.10.191}$$

$$-\frac{\partial U}{\partial z} = F_z \tag{4.10.192}$$

Here $z$ is the $z$-component of the tip position $\boldsymbol{r}$ and $F_z$ is the total force between tip and sample. The resulting values of z were then compared with the experiment. They found practically no height difference between the inequivalent A and B sites in graphite (See figure 4.283 for a sketch of the graphite crystal structure).
This theory is very simple and easy to calculate. Gould *et al.*[**?**]) used a desktop computer to solve the relaxation equations. Theories of this kind do not give the full physical description of the processes between tip and sample. But they will give a first idea on what to expect, on graphite and maybe with some modifications, on other surfaces.
To include many-body interactions Abraham and Batra[**?**] use the effective potential of Stilinger and Weber[**?**] developed for silicon. This potential gives a stable silicon lattice whereas the Lennard-Jones potential used by Gould *et al.*[**?**] does not yield a stable silicon lattice. The potential of Stilinger and Weber[**?**] includes doublet and triplet interactions. Abraham and Batra[**?**] modified this potential to accurately describe the interaction of the carbon atoms in the graphite layer. For the weak bonding between layers, they still use the Lennard-Jones potential.
Abraham and Batra[**?**] obtain a total corrugation of 10 pm or less, consistent with the calculation of the total charge density[**?**]. This corrugation is, however, smaller than the measured corrugations. Like the theory of Gould *et al.*[**?**], Abraham and Batra do not find a significant height difference between the A and B-sites.

### 4.10.17.3. Nonlocal Forces

Experiments using a single atom at a time to interact with the sample surface are routinely done in atom scattering experiments. As an example we will briefly outline the Helium scattering experiments. A well collimated beam of Helium atoms with a narrow distribution of velocities around a center velocity $v_0$ is aimed at the surface. The individual atoms hit the surface and in the case of Helium bounce back elastically from it. The angle between the line of incidence and the

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

line of emergence and the orientation of the plane defined by these two lines tells the experimenter about the structure and periodicity of the sample surface. These experiments a analogous to x-ray diffraction experiments. The scattering amplitudes and angles have to be transformed back to reveal the real surface structure. However since it is a scattering experiment, little information can be found on the defects of the periodic arrangement of the atoms. The area of interaction is of the order of 1 $mm^2$.

To get a local probe which needs not to be scattered off a sample surface, the probe atom is held on the apex of a tip. Figure 4.406 shows a once widely used example of such a tip, the cleavage planes of diamond. Diamond under stress is most likely to cleave along the (111) planes of its crystal structure.

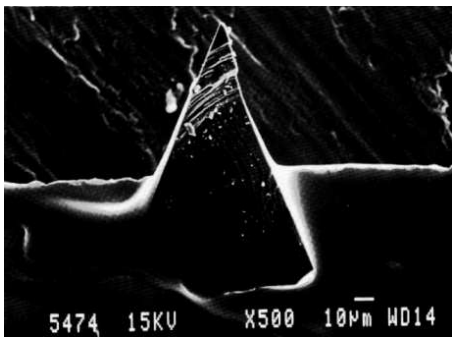



Abbildung 4.406.: Cleaved diamond used as a *SFM* tip. The diamond image was taken by Charles Bracker, Purdue University.

For real *SFM* tips the assumption of a single interacting atom is not justified. Attractive forces like van der Waals forces reach out for several nanometers. The attractive forces are compensated by the repulsion of the electrons when one atom tries to penetrate another. The decay length of the interaction and its magnitude depend critically on the type of atoms and the crystal lattice they are bound in. The shorter the decay length the smaller is the number of atoms which contribute a sizable amount to the total force. The decay length of the potential on the other hand is directly related to the type of force. Repulsive forces between atoms at small distances are governed either by an exponential law (like the tunneling current in the *STM*), by an inverse power law with large exponents, or by even more complicated forms. Hence the highest resolution images are obtained using the repulsive forces between atoms in contact. The high inverse power exponent or even exponential decay of this distance dependence guarantees that the other atoms beside the apex atom do not significantly interact with the sample surface. Attractive van-der-Waals interactions on the other hand are reaching far out into space. Hence a larger number of tip atoms takes part in this interaction and hence the resolution can not be as good. The same is true for magnetic potentials and for the electrostatic interaction between charged bodies.

A crude estimation of the forces between atoms can be obtained in the following way: assume that two atoms with mass $m$ are bound in molecule. The potential at the equilibrium distance can be approximated by a harmonic potential or, equivalently, by a spring constant. The frequency of the vibration $f$ of the atom around

its equilibrium point is then a measure for the spring constant $k$:

$$k = (\omega)^2 \frac{m}{2} \tag{4.10.193}$$

where we have to use the reduced atomic mass. The vibration frequency can be obtained from optical vibration spectra or from the vibration quanta $\hbar\omega$;

$$k = \left(\frac{(\hbar\omega)}{\hbar}\right)^2 \frac{m}{2} \tag{4.10.194}$$

As a model system we take the hydrogen molecule $H_2$. The mass of the hydrogen atom is $m = 1.673 \times 10^{-27}$ kg and its vibration quantum is $\hbar\omega = 8.75 \times 10^{-20}$ J. Hence the equivalent spring constant is $k = 560 \frac{\mathrm{N}}{\mathrm{m}}$. Typical forces for small deflections (1 % of the bond length) from the equilibrium position are $\approx 5 \times 10^{-10}$ N. The force calculated this way is an order of magnitude estimation of the forces between two atoms.
An atom in a crystal lattice on the surface is more rigidly attached since it is bound to more than one other atom. Hence the effective spring constant for small deflections is larger. The limiting force is reached when the bond length changes by 10 % or more, which indicates that the forces used to image surfaces must be of the order of $10^{-8}$ N or less. The sustainable force before damage is dependent on the type of surfaces. layered materials like mica or graphite are more resistant to damage than soft materials like biological samples. Experiments have shown that on selected inorganic surfaces like mica one can apply up to $10^{-7}$ N. On the other hand, some biological samples are destroyed by forces of the order of $10^{-9}$ N.

### 4.10.17.4. How to Measure Small Forces

The key to the successful operation of a *SFM* is the measurement of the interaction forces between a small probing structure, the tip, and the sample surface. The probing structure would ideally consist of only one atom which is brought in the vicinity of the sample surface. In the following chapters we will deal with how to approximate the single atom probing structure and how to detect the minute forces acting on this structure. Of the potential methods to detect minute distance changes we will discuss electron tunneling, interferometry, the optical lever method. We will not discuss capacitance measurements.

### 4.10.17.5. *Cantilever* Springs

The interaction forces between the *SFM* tip and the sample surface must be smaller than about $10^{-7}$ N for bulk materials and preferably well below $10^{-8}$ N for organic macromolecules. In order to obtain a measurable deflection larger than the inevitable thermal drifts and noise the *cantilever* deflection for static measurements should be at least 10 nm. Hence the spring constants should be less than 10 N/m for bulk materials and less than 1 N/m for organic macromolecules. Experience shows that cantilevers with spring constants of about 0.01 N/m work best.
Building vibrations usually have frequencies in the range from 10 Hz to 100 Hz. These vibrations are coupled to the *cantilever*. To get an estimate of the magnitude

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

we note that the resonance frequency of a structure in terms of its spring constant $k$ and a lumped effective mass $m_{eff}$ is

$$f_{res} = \frac{1}{2\pi}\sqrt{\frac{k}{m_{eff}}} \tag{4.10.195}$$

Inserting 100 Hz for the resonance frequency and a spring constant of 0.1 N/m we obtain an upper limit of the lumped effective mass $m_{eff}$ of 0.25 mg. The quality factor of this resonance in air is typically between 10 and 100. To get a reasonable suppression of the excitation of *cantilever* oscillations, the *cantilever*'s resonance frequency has to be at least a factor of 10 higher than the highest of the building vibration frequencies. This means, that $m_{eff}$ has to be under any circumstances no larger than $\frac{0.25\text{mg}}{100} = 2.5\mu\text{g}$. It would be preferably to limit the mass to 0.1 $\mu$g. A tungsten wire with 20 $\mu$m diameter must be shorter than 1.6 mm to have a mass of less then 0.1 $\mu$g. This lumped mass $m_{eff}$, however, is smaller than the real mass m, by a factor which depends on the geometry of the *cantilever*. A good rule of thumb says that the effective mass $m_{eff}$ is 1/3 of the real mass. Gluing tips or mirrors on cantilevers adds their mass to the effective mass. Since these additional gadgets are attached to the free end of the *cantilever*, they do not benefit from the factor 1/3 in calculating the effective mass.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

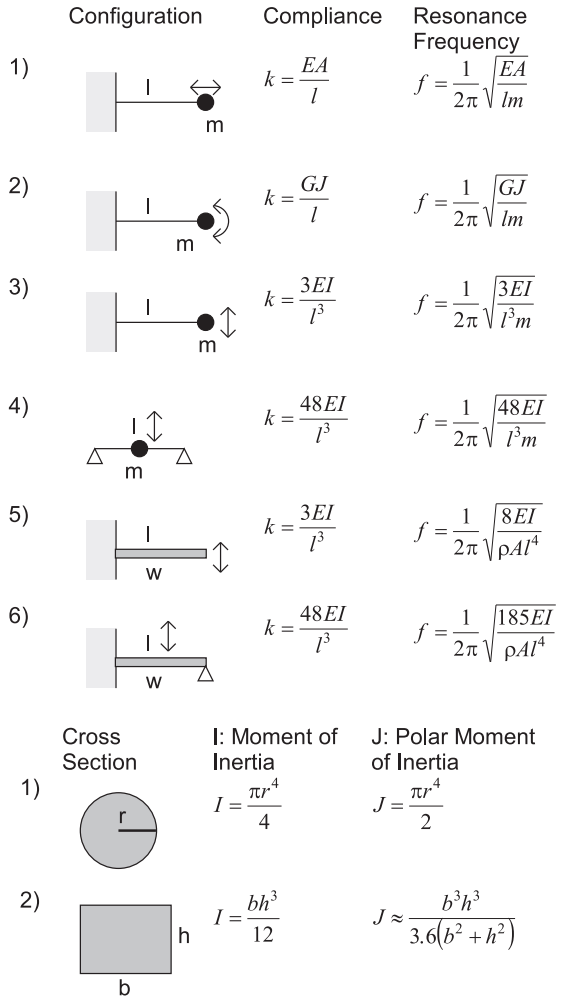



Abbildung 4.407.: Selected configurations of levers and their resonance frequency[**?**]. Part a) shows the compliance of levers and their resonance frequency. $k$ is the compliance, $f$ the resonance frequency, $l$ the length of the lever, $A$ the cross section, $E$ Young's modulus, $G$ the shear modulus, $\rho$ the density of the lever material, and $m$ a concentrated mass at the end of the lever. The moment of inertia $I$ and the polar moment of inertia $J$ are given in part b). $r$ is the radius of a circular cross section and $b$ and $h$ the width and height of a rectangular cross section, respectively. The following configurations are shown in part a): 1): a cantilevered massless beam in the compression mode with a weight concentrated at the end; 2) the torsion of a cantilevered massless beam with a concentrated weight; 3) the deflection of a cantilevered massless beam with a concentrated weight at the end; 4) a massless beam supported at the ends with a concentrated weight in the center; 5) a massive cantilevered beam and 6) a massive cantilevered beam supported at the end.

Figure 4.407a) gives approximate values for the spring constant and the resonance

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

frequency of selected configurations. Of particular importance for the understanding of the performance of a *SFM* are configurations 5) (the free *cantilever*), 6) (the *cantilever* in repulsive contact with the sample) and 2) (lateral force measurement). Comparing 5) and 6) we see that a *cantilever* in repulsive contact with a sample has a resonance frequency which 4.8 times that of the free *cantilever*. Part b) of figure 4.407 shows the moments of inertia for selected cross sections.
Today micromachined cantilevers are commercially available and are used almost exclusively. The manufacturing process of cantilevers has been published in the Literature[?, ?, ?, ?].

### 4.10.17.6. Detecting the Spring Deflection by Tunneling

The first *SFM* published by Binnig, Quate and Gerber[?] employed tunneling to detect the bending of the force sensing *cantilever*. Figure 4.408 shows a sketch of the arrangement. The authors sandwiched a *cantilever* between the tip of an *STM* and the sample. The sensitivity of the tunneling detector in the *SFM* is the best of all possible detectors. On clean surfaces the change in tunneling current might approach one order of magnitude for every 0.1 nm change in deflection. In air there are a few critical points which might degrade the performance of the *SFM*.

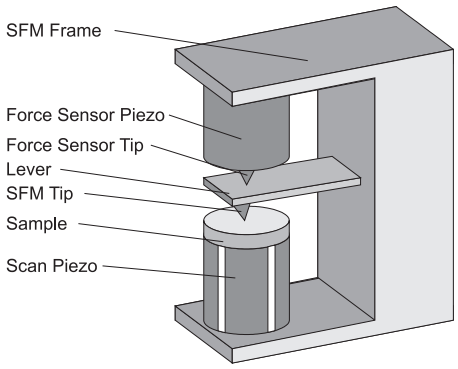



Abbildung 4.408.: The principle of the *SFM* proposed by Binnig *et. al.*[?]. A lever with a spring constant of $\approx$ 1 N/m is pressed into a sample mounted on a piezo tube. The deflection of the lever is measured by a tunnel junction. The tunnel gap is adjusted by a force sensor piezo. Alternatively the *SFM* tip on the lever can be operated in a non-contact mode via attractive forces.

An ideal deflection detector for an *SFM* should not have any sensitivity to the local surface structure of the force sensing *cantilever*. The tunneling current between the back of the *cantilever* and the sensing electrode, however, is confined to a narrow region whose width is mainly determined by the local curvatures of the sensing electrode and the *cantilever*. If the sensing electrode is a sharp tip, as used for *STM* experiments, the *SFM* can be very susceptible to the lateral bending of the *cantilever*.
A solution of the resolution problem is to use as smooth a *cantilever* back as possible and to increase the area of the tunneling current. However this aggravates a second problem which might occur in a *tunneling SFM*. Adsorbate layers are present on both the sensing electrode and on the back of the *cantilever*. Typical distances between the two electrodes in a tunneling junction are of the order of

1 nm. If two monolayers are present both on the sample and on the *cantilever* tip, the distance between the sensing electrode and the back of the *cantilever* might be too small to allow a tunneling current to pass. Therefore the sensing electrode has to be pressed against the back of the *cantilever* which will yield due to its low spring constant. It is possible that no tunneling current can be established. Furthermore the filled gap between the *cantilever* and the sensing electrode rigidizes the *cantilever*. The effective spring constant of the *cantilever* is then a function of the stiffness of the adsorbate layers. Whereas tunneling as a deflection detector has its deficiencies in air, it might become the method of choice in vacuum SFMs. We have seen that it is advantageous to use microfabricated cantilevers in an *SFM*. These cantilevers have a Q of about 100 in air and a Q of $> 10^4$ in vacuum. Any sudden change in the surface topography starts a damped oscillation of the *cantilever*. The amplitude of the oscillation will decay to 1/e after Q oscillations. If the resonance frequency were 10 kHz the time constant would be 1 second. Such a time constant would impose such a small scanning speed that the whole microscope would become impractical. Here the additional, highly nonlinear force between the sensing electrode and the back of the *cantilever* could help in damping the oscillations of the *cantilever*.

### 4.10.17.7. Homodyne and Heterodyne Interferometry

Soon after the first papers on the *SFM*[**?**] McClelland[**?**] published a *SFM* using interferometry. The sensitivity of the interferometer depends on the wavelength of the light employed in the apparatus. Figure 4.409 shows the principle of such an interferometeric design. The light incident from the left is focused by a lens on the *cantilever*. The reflected light is collimated by the same lens and interferes with the light reflected at the flat. To separate the reflected light from the incident light a $\lambda/4$- plate converts the linearly polarized incident light into circular polarized light. The reflected light is made again linear polarized by the $\lambda/4$- plate, but with a polarization orthogonal to that of the incident light. The polarizing beam splitter then deflects the reflected light to the photo diode.

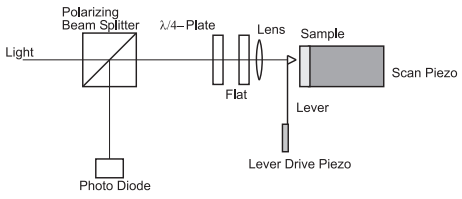



Abbildung 4.409.: Principle of an interferometer *SFM*. The light of the laser light source on the left is polarized by the polarizing beam splitter and focused on the back of the force measuring *cantilever* on the right. The *cantilever* oscillates at or near its resonance frequency. The light passes twice through a $\lambda/4$-plate. The returning light is therefore polarized orthogonally to the incident light and therefore reflected to the photo diode. The light reflected from the flat serves as the reference of the interferometer. The interference pattern is modulated at the oscillation frequency of the *cantilever*.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

To improve the *signal to noise ratio* of the *interferometer* the lever is driven by a piezo near its resonance frequency. The amplitude $\Delta z$ of the lever is

$$\Delta z = \Delta z_0 \frac{1}{\sqrt{(\Omega^2 - \Omega_0^2) + \Omega^2/Q^2}}. \tag{4.10.196}$$

where $\Delta z_0$ is the constant drive amplitude, $\Omega_0$ the resonance frequency of the lever, $Q$ the quality of the resonance, and $\Omega$ the drive frequency. The resonance frequency of the lever is given by the effective potential

$$\Omega_0 = \sqrt{(k + \frac{\partial^2}{\partial z^2} U)/m_{eff}}. \tag{4.10.197}$$

where $k$ is the spring constant of the free lever, $U$ the interaction potential between the tip and the sample, and $m_{eff}$ the effective mass of the *cantilever*. Equation (4.10.197) shows, that an attractive potential decreases the resonance frequency $\Omega_0$. The change in the resonance frequency $\Omega_0$ in turn results in a change of the lever amplitude $\Delta z$ (see equation (4.10.196)).

The movement of the *cantilever* changes the path difference in the interferometer. The light reflected from the lever with the amplitude $A_{l,0}$ and the reference light with the amplitude $A_{r,0}$ interfere on the detector. The detected intensity $I(t) = (A_l(t) + A_r(t))^2$ consists of two constant terms and a fluctuating term

$$\overline{2A_l(t)A_r(t)} = A_{l,0}A_{r,0} \sin(\omega t + \frac{4\pi\delta}{\lambda} + \frac{4\pi\Delta z}{\lambda} \sin(\Omega t)) \tag{4.10.198}$$

Here $\omega$ is the frequency of the light and $\Delta z$ is the instantaneous amplitude of the lever, given according to equations (4.10.196) and (4.10.197) by the driving frequency $\Omega$, the spring constant $k$ and the interaction potential $U$. The time average of equation (4.10.198) then becomes

$$\begin{aligned} \overline{2A_l(t)A_r(t)} &\propto \cos(\frac{4\pi\delta}{\lambda} + \frac{4\pi\Delta z}{\lambda} \sin(\Omega t)) && (4.10.199) \\ &\approx \cos(\frac{4\pi\delta}{\lambda}) - \sin(\frac{4\pi\Delta z}{\lambda} \sin(\Omega t)) \\ &\approx \cos(\frac{4\pi\delta}{\lambda}) - \frac{4\pi\Delta z}{\lambda} \sin(\Omega t) && (4.10.200) \end{aligned}$$

Here all small quantities have been omitted and functions with small arguments have been linearized. The amplitude of the lever oscillation $\Delta z$ can be recovered with a lock-in technique. However, equation (4.10.199) shows that the measured amplitude is also a function of the path difference $\delta$ in the interferometer. Hence this path difference $\delta$ must be very stable. The best sensitivity is obtained when $\sin(\frac{4\delta}{\lambda}) \approx 0$.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

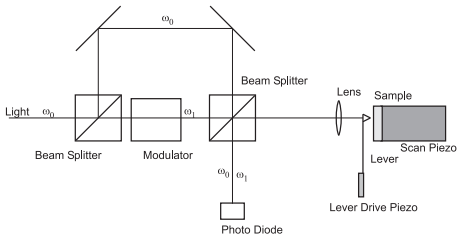



Abbildung 4.410.: Heterodyne interferometer *SFM*. Light with a frequency of $\omega_0$ is split into a reference path (upper light path) and a measurement path. The measurement light is shifted in frequency to $\omega_1$ by a modulator. This light is reflected by the *cantilever* oscillating at or near its resonance frequency and interferes with the reference beam at $\omega_0$ on the photo diode.

This influence is not present in the heterodyne detection scheme shown in figure 4.410. Light incident from the left with a frequency $\omega$ is split in a reference path (upper path in figure 4.410) and a measurement path. Light in the measurement path is shifted in frequency to $\omega_1 = \omega + \Delta\omega$ and focused on the *cantilever*. The *cantilever* oscillates at the frequency $\Omega$, as in the homodyne detection scheme. The reflected light $A_l(t)$ is collimated by the same lens and interferes on the photo diode with the reference light $A_r(t)$. The fluctuating term of the intensity is given by

$$2A_l(t)A_r(t) = A_{l,0}A_{r,0}\sin((\omega+\Delta\omega)t + \frac{4\delta}{\lambda} + \frac{4\Delta z}{\lambda}\sin(\Omega t))\sin(\omega t) \qquad (4.10.201)$$

where the variables are defined as in equation (4.10.198). Setting the path difference $\sin(\frac{4\delta}{\lambda}) \approx 0$ and taking the time average, omitting small quantities and linearizing functions with small arguments we get

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

$$
\begin{aligned}
\overline{2A_l(t)A_r(t)} &\propto \cos(\Delta\omega t + \frac{4\pi\delta}{\lambda} + \frac{4\pi\Delta z}{\lambda}\sin(\Omega t)) \\
&= \cos(\Delta\omega t + \frac{4\pi\delta}{\lambda})\cos(\frac{4\pi\Delta z}{\lambda}\sin(\Omega t)) - \\
&\quad \sin(\Delta\omega t + \frac{4\pi\delta}{\lambda})\sin(\frac{4\pi\Delta z}{\lambda}\sin(\Omega t)) \\
&\approx \cos(\Delta\omega t + \frac{4\pi\delta}{\lambda})\left(1 - \frac{8\pi^2\Delta z^2}{\lambda^2}\sin(\Omega t)\right) \\
&\quad -\frac{4\pi\Delta z}{\lambda}\sin(\Delta\omega t + \frac{4\pi\delta}{\lambda})\sin(\Omega t) \\
&= \cos(\Delta\omega t + \frac{4\pi\delta}{\lambda}) - \frac{8\pi^2\Delta z^2}{\lambda^2}\cos(\Delta\omega t + \frac{4\pi\delta}{\lambda})\sin(\Omega t) - \\
&- \frac{4\pi\Delta z}{\lambda}\sin(\Delta\omega t + \frac{4\pi\delta}{\lambda})\sin(\Omega t) \\
&= \cos(\Delta\omega t + \frac{4\pi\delta}{\lambda}) - \frac{4\pi^2\Delta z^2}{\lambda^2}\cos(\Delta\omega t + \frac{4\pi\delta}{\lambda}) \\
&\quad +\frac{4\pi^2\Delta z^2}{\lambda^2}\cos(\Delta\omega t + \frac{4\pi\delta}{\lambda})\cos(2\Omega t) \\
&- \frac{4\pi\Delta z}{\lambda}\sin(\Delta\omega t + \frac{4\pi\delta}{\lambda})\sin(\Omega t) \\
&= \cos(\Delta\omega t + \frac{4\pi\delta}{\lambda})\left(1 - \frac{4\pi^2\Delta z^2}{\lambda^2}\right) \\
&\quad +\frac{2\pi^2\Delta z^2}{\lambda^2} \\
&\quad \left(\cos((\Delta\omega + 2\Omega)t + \frac{4\pi\delta}{\lambda}) + \cos((\Delta\omega - 2\Omega)t + \frac{4\pi\delta}{\lambda})\right) \\
&\quad +\frac{2\pi\Delta z}{\lambda} \\
&\quad \left(\cos((\Delta\omega + \Omega)t + \frac{4\pi\delta}{\lambda}) + \cos((\Delta\omega - \Omega)t + \frac{4\pi\delta}{\lambda})\right) \qquad (4.10.202)
\end{aligned}
$$

Multiplying electronically the components oscillating at $\Delta\omega$ and $\Delta\omega + \Omega$ and rejecting any product except the one oscillating at $\Omega$ we obtain

$$
\begin{aligned}
A &= \left(1 - \frac{4\pi^2\Delta z^2}{\lambda^2}\right) 2\Delta z\lambda\cos((\Delta\omega + \Omega)t + \frac{4\pi\delta}{\lambda})\cos(\Delta\omega t + \frac{4\pi\delta}{\lambda}) \\
&= \left(1 - \frac{4\pi^2\Delta z^2}{\lambda^2}\right) \Delta z\lambda\left(\cos((2\Delta\omega + \Omega)t + 8\pi\delta\lambda) + \cos(\Omega t)\right) \\
&\approx \frac{\pi\Delta z}{\lambda}\cos(\Omega t) \qquad (4.10.203)
\end{aligned}
$$

Equation (4.10.203) shows that the amplitude $\Delta z$ of the *cantilever* motion can be recovered with a Unlike in the homodyne detection scheme the recovered *signal* is independent from the path difference $\delta$ of the interferometer. Furthermore a lock-in amplifier with the reference set $\sin(\Delta\omega t)$ can measure the path difference $\delta$ independent of the *cantilever* oscillation. If necessary, a feedback circuit can keep $\delta = 0$.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

### 4.10.17.8. Fiberoptic Interferometer

The first solution[?] is to use an fiberoptic interferometer. Its principle is sketched in figure 4.411. The light of a laser is fed into an optical fiber. Laser diodes with integrated fiber pigtails are convenient light sources. The light is split in a fiberoptic beam splitter into two fibers. One fiber is terminated by index matching grease to avoid any reflections back into the fiber. The end of the other fiber is brought close to the *cantilever* in the *SFM*. The emerging light is partially reflected back into the fiber by the *cantilever*. Most of the light, however, is lost. This is not a big problem since only 4% of the light is reflected at the end of the fibre, at the glass-air interface. The two reflected light waves interfere with each other. The product is guided back into the fiber coupler and again split into two parts. One half is analyzed by the photo diode. The other half is fed back into the laser. Communications grade laser diodes are sufficiently resistant against feedback to be operated in this environment. They have, however, a bad coherence length, which in this case does not matter, since the optical path difference is in any case no larger than 5 $\mu$m. Again the end of the fiber has to be positioned on a piezo drive to set the distance between the fiber and the *cantilever* to $\lambda$(n+1/4).

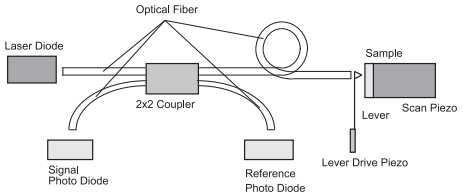



Abbildung 4.411.: Principle of the fiberoptic interferometer *SFM*: The light of the laser diode is coupled into a fiber. 50 % of the light is transmitted to the back of the *cantilever*. The other half serves as an intensity reference. Light reflected from the *cantilever* and light reflected from the end of the fiber interfere. The fiber coupler again splits the light traveling back evenly between the *signal photo diode* and the *laser diode*. The small path difference of $\approx$ 10 $\mu$m accounts for the excellent stability of the microscope.

### 4.10.17.9. Nomarsky-Interferometer

A third solution to minimize the optical path difference uses the Nomarski principle[?]. Figure 4.412 depicts a sketch of the microscope of these authors. The light of a laser is focused on the *cantilever* by lens A. A birefringent crystal B between the *cantilever* and the lens with its optical axis 45° off the polarization direction of the light splits the light beam into two paths, offset by a distance given by the length of the birefringent crystal. Birefringent crystals have varying indexes of refraction. In calcite, one crystal axis has a lower index than the other two. This means, that certain light rays will propagate at a different speed through the crystal than the others. By choosing a correct polarization, one can select the ordinary ray, the extraordinary ray or one can get any distribution of the intensity amongst those two rays. A detailed description of birefringence can be found in textbooks[?]. A

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

calcite crystal deflects the extraordinary ray at an angle of 6° within the crystal. By choosing a suitable length of the calcite crystal, any separation can be selected.

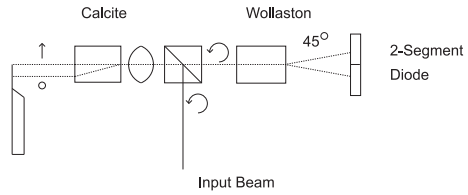



Abbildung 4.412.: Principle of the Nomarski *SFM*. After Schönenberger and Alvarado[**?**]. The circular polarized input beam is deflected to the left by a non-polarizing beam splitter. The light is focused onto a *cantilever*. The calcite crystal between the lens and the *cantilever* splits the circular polarized light into two spatially separated beams with orthogonal polarizations. The two light beams reflected from the lever are superimposed by the calcite crystal and collected by the lens. The resulting beam is again circular polarized. A Wollaston prism produces two interfering beams with a $\pi/2$ phase shift between them. The minimal path difference accounts for the excellent stability of this microscope.

The focus of one light ray is positioned near the free end of the *cantilever* while the other is placed close to the clamped end. Both arms of the interferometer pass through the same space, except for the distance between the calcite crystal and the lever. The closer the calcite crystal is placed to the lever, the less influence disturbances like air currents have.

#### 4.10.17.10. Detecting Spring Deflection by the Optical Lever Method

Still another spring detection system is the optical lever method[**?**, **?**]. This method, depicted in figure 4.413 employs the same technique as light beam deflection galvanometers used to have and still have. A fairly well collimated light beam is reflected off a mirror and projected to a receiving target. Any change in the angular position of the mirror will change the location, where the light ray hits the target. Galvanometers use optical path lengths of several meters and scales projected to the target wall as a read-out help.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

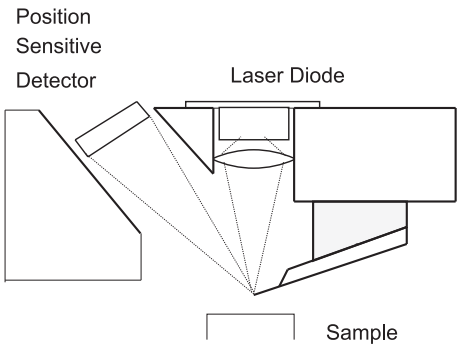



Abbildung 4.413.: The principle of the lever *SFM*. Light from a laser diode is focused on the back of a *cantilever*. The reflected light is deflected when the *cantilever* bends under an applied force. The deflection angle is measured by a position sensitive detector.

For the *SFM* using the optical lever method a photo diode segmented into two closely spaced devices is used. Initially, the light ray is set to hit the photo diodes in the middle of the two sub-diodes. Any deflection of the *cantilever* will cause an imbalance of the number of photons reaching the two halves. Hence the electrical currents in the photo diodes will be unbalanced too. The difference *signal* is further amplified and is the input *signal* to the feedback loop. Unlike the interferometeric SFMs, where often a modulation technique is necessary to get a sufficient *signal to noise ratio*, most SFMs employing the optical lever method are operated in a static mode. The domain of optical lever SFMs are the measurements in the repulsive regime. It is the simplest method to construct an optical readout and it can be confined in volumes smaller than 5 cm on the side. To evaluate the proper design parameters, let us calculate the sensitivity of the microscope. Figure 4.414 shows a cross section of the reflected light beam. For the sake of simplicity we assume that the light beam is of uniform intensity with its cross section increasing proportional to the square of the distance between the *cantilever* and the quadrant detector. The movement of the center of the light beam is then given by

$$\Delta x = \Delta z \frac{d}{l} \tag{4.10.204}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

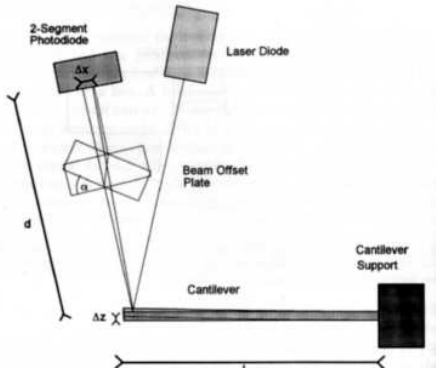



Abbildung 4.414.: Sensitivity and calibration of a lever *SFM*. The deflection $\Delta z$ of the *cantilever* translates into a displacement $\Delta x$ on the 2-segment photo diode. The photo diode is located at a distance $d$ from the *cantilever* of length $l$. The deflection $\Delta x$ can be calibrated by tilting the beam offset plate by an angle $\alpha$.

The photo current generated in a photo diode is proportional to the number of incoming photons hitting it. If the light beam contains a total number of $N_0$ photons then the change in difference current becomes

$$\Delta(I_R - I_L) = \Delta I = const\ \Delta x d N_0 \tag{4.10.205}$$

Combining equations (4.10.204) and (4.10.205) one obtains that the difference current ΔI is independent of the separation of the quadrant detector and the *cantilever*. This relation is true, if the light spot is smaller than the quadrant detector. If it is greater, the difference current ΔI becomes smaller with increasing distance. The light beam in reality has a Gaussian intensity profile. For small movements Δx (compared to the diameter of the light spot at the quadrant detector), equation (4.10.205) still holds. Larger movements Δx, however, will introduce a nonlinear response. If the *SFM* is operated in a constant force mode, only small movements Δx of the light spot will occur. The feedback loop will cancel out all other movements.

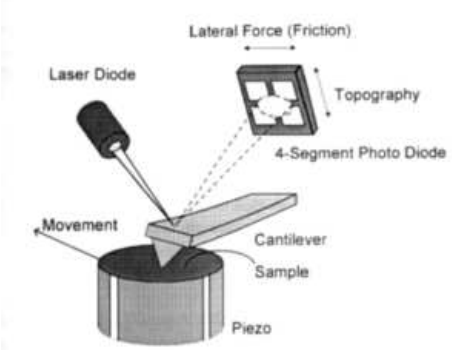



Abbildung 4.415.: Scanning Force and Friction Microscope (SFFM). The lateral forces exerted on the tip by the moving sample causes a torsion of the lever. The light reflected from the lever is deflected orthogonally to the deflection caused by normal forces.

The scanning of a sample with an *SFM* can twist the microfabricated cantilevers

because of lateral forces[**?**, **?**, **?**] and affect the images[**?**]. When the tip is subjected to lateral forces, it will twist the lever and the light beam reflected from the end of the lever will be deflected perpendicular to the ordinary deflection direction. For many investigations this influence of lateral forces is unwanted. The design of the triangular cantilevers stems from the desire, to minimize the torsion effects. However, lateral forces open up a new dimension in force measurements. They allow, for instance, a distinction of two materials because of the different friction coefficient, or the determination of adhesion energies. To measure lateral forces the original optical lever *SFM* has to be modified: figure 4.415 shows a sketch of the instrument. The only modification compared with figure 4.413 is the use of a quadrant detector photo diode instead of a two segment photo diode and the necessary readout electronics. The electronics calculates the following signals:

$$U_{Force} = \alpha \left( (I_{UpperLeft} + I_{UpperRight}) - (I_{LowerLeft} + I_{LowerRight}) \right) \tag{4.10.206}$$

$$U_{Friction} = \beta \left( (I_{UpperLeft} + I_{LowerLeft}) - (I_{UpperRight} + I_{LowerRight}) \right)$$

The calculation of the lateral force as a function of the deflection angle does not have a simple solution for cross-sections other than circles. Baumeister and Marks[**?**] give an approximate formula for the angle of twist for rectangular beams:

$$\Theta = \frac{M_t l}{\beta G b^3 h} \tag{4.10.207}$$

where $M_t = Fa$ is the external twisting moment due to friction, $l$ is the length of the beam, $b$ and $h$ the sides of the cross section, $G$ the shear modulus and $\beta$ a constant determined by the value of $\frac{h}{b}$. For the equation to hold $h$ has to be larger than $b$.
Inserting the values for a typical microfabricated lever with integrated tips

$$\begin{aligned} b &= 6 \times 10^{-7}\text{m} \\ h &= 10^{-5}\text{m} \\ l &= 10^{-4}\text{m} \\ a &= 3.3 \times 10^{-6}\text{m} \\ G &= 5 \times 10^{10}\text{Pa} \\ \beta &= 0.333 \end{aligned}$$

into equation (4.10.207) we obtain the relation

$$F = 1.1 \times 10^{-4}\Theta \tag{4.10.208}$$

Typical lateral forces are of order $10^{-10}$ N.

### 4.10.17.11. The Force Microscope

In this chapter we will first focus on the design of a *SFM* and give some hints how to build such a device. In the second part we will give recipes on how to adjust a *SFM*.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

### 4.10.17.12. Special Design Considerations

A *SFM* is very similar in design to a *STM*. The requirement of a small, rigid design is even more important for a *SFM* than for an *STM*. The construction of the force sensing unit imposes some changes to the arrangement of the scanning piezo and the sample location. The resonance frequency of the scanning piezo is decreased by loading it with an additional mass. In the case of the *STM*, this mass consists of the tip holder and the tip itself. It is smaller than the mass of the sample in most cases. In the previous chapters we have seen that the detection systems for the deflection of the force sensor can be quite bulky. Except for the tunneling detector, all are too big in size to be mounted on a piezo tube. Even the tunneling detector requires additional distance adjustment, which would lower the scanning piezo's resonance frequency too much. As a consequence, most published SFMs mount the sample on the scanning piezo. The force sensing unit is stationary, with the sample being scanned past the immobile tip. The structure of the force sensing unit has to be as rigid as possible to minimize errors due to thermal drift. Especially the tunneling detection method and the fiberoptic detection method are prone to this error. The problem is most severe in the tunneling deflection detector, since a rigid electrode, the sensing electrode, is at a distance of about 1 nm by the *cantilever*.
To illustrate the problem of thermal drift, we calculate the requirements on the temperature stability for a microscope working with repulsive forces and which does not employ heterodyne detection. If we assume a *cantilever* spring with a 1 N/m spring constant and if we set the force to $10^{-8}$ N, then the static deflection of the *cantilever* is 10 nm. The typical size of a tunneling force detector is 1 cm length from the tunnel junction to the common attachment plane. A design with well compensated thermal expansion coefficients will have a remaining thermal expansion coefficient of $10^{-6}\frac{\mathrm{m}}{\mathrm{Km}}$. This means that keeping the force within 10% requires a thermal stability of the microscope of 0.1 K. In less well compensated design, the allowable temperature fluctuations might be as low as 0.01 K.
If the temperature stability of the setup is not sufficient, one can either use larger static deflections, which means larger forces, or a softer *cantilever* spring, which means degraded frequency response. For measurements with the smallest possible forces, a careful design of the force sensor with respect to thermal drift is a prerequisite.
The interferometeric force sensors show the same drift problems. The classical Michelson or Mach- Zehnder interferometers are the worst, since their relevant distances for differential thermal expansions may be more than 10 cm long. The fiberoptic interferometer is comparable to the tunneling detector in its thermal performance, since the distances needed to position the end of the fiber are of order 1 cm. Much better is the Nomarski detector for the *cantilever* deflection. This detector is only sensitive to a thermally induced rotation of the *cantilever* spring. A crude estimate gives relevant distances for the thermal expansion of a few 10 mm. This increases the allowable temperature variations to more than 1 K.
Equally well suited is the optical lever method. This method is, to first order, only sensitive to the tilt of the reflecting mirror. For small angles between the incident and the reflected light beam, the change in distance between the plane defined by the quadrant detector and the light source is negligible. Any distance change between the light source and the quadrant detector affects directly the *output signal*. However the deflection of the *cantilever* is amplified by a factor of

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

up to 1000 due to the geometrical amplification. Hence the optical lever method is, to first order, insensitive to thermal drift.

### 4.10.17.13. How to Adjust a Force Microscope

Compared with an *STM* a *SFM* needs some additional adjustments. This chapter will describe some procedures to facilitate the adjustment.
The adjustment of a *SFM* can be divided in the adjustment of the force sensor and the approach of the force sensor to the sample surface. The latter is similar to the approach of an *STM*-tip to the sample and discussed in the section on *STM*.
When mounting a new *cantilever*, first one has to position it to the correct position respective to the deflection measurement sensor. The size of the cantilevers varies from 0.1 mm to 2 mm. The best sensitivity is obtained when the deflection sensor points to the end of the *cantilever*. This adjustment is best done under a microscope.

**Tunneling Sensor:** The sensing electrode in the tunneling sensor has to be brought to about 1 nm to the back of the *cantilever*. This requires an approach mechanism for the sensing electrode similar to that of a *STM*. One can use, for instance, the action of a differential spring system[**?**] or lever reduction systems.

The surfaces of the sensing electrode and the *cantilever* should be as clean as possible to avoid a stiffening of the *cantilever* by the sandwiched adsorbates between the lever and the sensing electrodes.

To keep the force accurately, the force sensor has to be constructed to minimize thermal drift. Alternatively, one could periodically readjust the force[**?**].

**Interferometric Force Sensor:** It is important to have the light reflected from the end of the *cantilever*. This adjustment can be done by using a microscope or by analyzing the diffraction patterns of the *cantilever*. **The experimenter should make sure, that no visible or invisible laser radiation can reach his eye. A good protection is the use of a small TV-camera mounted on the microscope.** The TV-monitor can be placed at any convenient location.

In addition to the location one has also to adjust the phase of the reflected light to get the best sensitivity. This usually means to move the *cantilever* by fractions of a $\mu$m towards or away from the fiber or the interferometer flat, or to shift the phase of one polarization in the Nomarski-interferometer.

**Optical Lever Sensor:** In order to adjust the optical lever sensor, one has first to focus the laser diode light in the plane of the *cantilever*. Next, the laser diode is moved to bring the focal point on the end of the *cantilever*. Both the focus adjustment and the positioning of the focal point have to be done using a microscope. **To avoid eye damage, it is best to use a small CCD-TV-camera to transmit the image from the microscope to a monitor. The adjustment can be observed on the monitor.**

The last adjustment is the positioning of the quadrant detector diode. The correct position is found, when the currents from all four segments are equal.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

This position guarantees also that the amplitude fluctuations do not influence the measurement (to first order) in the constant force mode.

The force sensor should be treated with utmost care after its adjustment. Strong accelerations should be avoided.

#### 4.10.17.14. Selected Experiments

In the following chapters we will give a few examples on experiments by *SFM*. We will focus on the imaging of inorganic surfaces, since all items related to the imaging of biological samples will be treated in chapter 8 in this book by Apell *et al.*

#### 4.10.17.15. Atomic Resolution Imaging

The first surface to be imaged with atomic resolution by a *SFM* was the graphite surface[**?**, **?**, **?**]. The graphite surface is of great importance to scanning probe microscopy as a reference sample and a substrate with flat terraces of several 100 nm lengths. Calculations show that for the *SFM*, the surface consists of hexagons of carbon atoms, each 0.146 nm apart. The centers of the rings are separated by 0.246 nm. There is a great variation in the appearance of the unit cell. The interpretation of the unit cell structure and the corrugation however is very complex. Abraham and Batra[**?**] and Gould *et al.*[**?**] explained the puzzling structures by multiple tips. These multiple tips create a superposition of several locations within the unit cell. They can produce an almost unlimited variation of the graphite unit cell appearance.

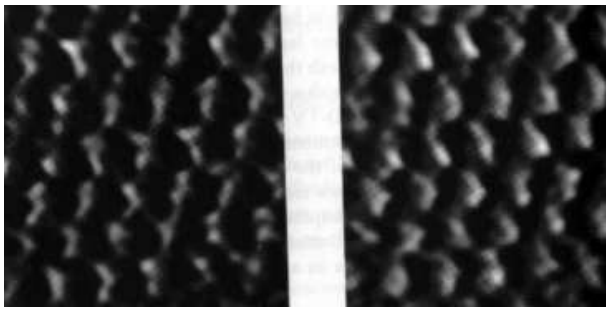


Abbildung 4.416.: a) Force image of mica. b) Friction image of mica. The size of the image is 1.8 nm by 2.4 nm. The corrugation is 0.2 nm for the topography.

Another layered material important for the biologist is mica. Like graphite, it has long flat terraces suitable for sample deposition. Since mica is an insulator, its binding properties with biological macromoleculs are different. By comparing the appearance, the adhesion and other properties of one sort of biological macromolecule bound to different substrates one can learn about the molecule itself and its binding properties. Figure 4.416a) shows as an example a measurement by *SFM*. The same instrument as in the case of graphite was used.

### 4.10.17.16. Lateral Forces and Friction

The imaging of surfaces by the *SFM* in the repulsive mode is based on the dragging of a fine tip across the sample. There are lateral forces between the tip and the sample. At the beginning of a scan the tip sticks to the surface. Later, it will move, but the lever will always feel a force parallel to the surface in addition to the normal force. If the lever is a simple wire, it will bend parallel to the surface. Mate *et al.*[**?**] and Erlandsson *et al.*[**?**] measured the sideways deflection by interferometry. They detected a variation of the lateral force with the periodicity of the graphite surface.

Figure 4.416b) is a lateral picture of the mica surface. This data was measured by an optical lever *SFM*, detecting the torsion of a microfabricated *cantilever* under the influence of friction. The mica periodicity is resolved with a lateral force modulation of $10^{-}9$ N. The band at the left side with no visible structure is due to the change in the scanning direction. The lateral force changes its sign, hence the width of the band is twice as large as the steady state friction.

### 4.10.17.17. Surface Profiles

For many industrial applications one needs to know the exact profile of a surface. Several methods are possible; a scanning electron microscope will give the desired information, provided the structures under investigation are not too shallow and the sample is conducting. The second method is the use of a profilometer, a widely used, proven instrument in industrial applications. Its lateral resolution is limited to a few 100 nm, which may be insufficient. A third method is the *STM*. It has the desired sensitivity, but requires conducting surfaces. The most versatile tool is the *SFM*. Figure 4.417a) shows a top view of an optical grating. Figure 4.417b) is the lateral force image measured at the same concurrently. The profile of the topography and the lateral forces shown in part c).

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

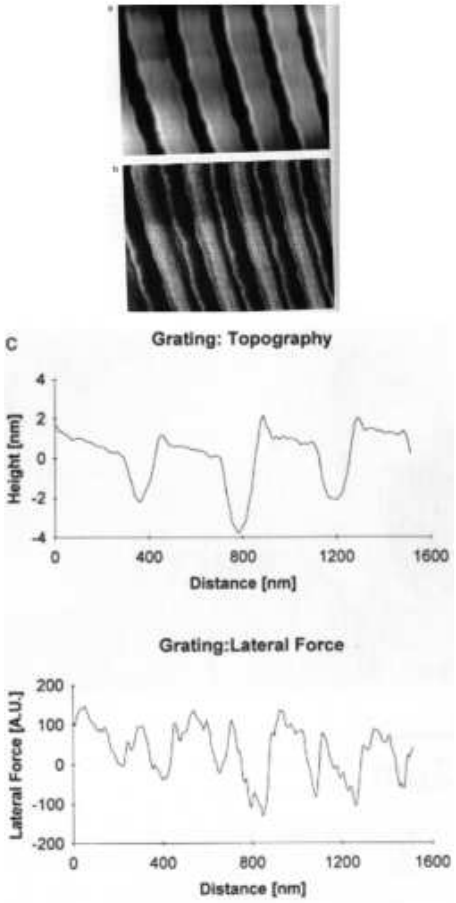



Abbildung 4.417.: Image of an optical grating. a) is the topograph of the image, b) a lateral force image. The size of the image is 1600 nm by 1200 nm. The corrugation of the topograph is ≈ 4 nm. c) shows a cross section through the topograph and the lateral force image.

### 4.10.17.18. Electrostatic Forces

The *SFM* is not only sensitive to the interaction between uncharged bodies, but to any other force. Martin *et al.*[**?**] demonstrated the use of a *SFM* to measure capacitances and local potentials. The ability to measure potentials parallels the Scanning Tunneling Potentiometer[**?**], but its theoretical resolution is inferior. The degraded resolution, however, is of no concern today, since the smallest electronic devices available are still larger than the resolution limit.
The capacitance $C$ between the tip and the sample is measured by applying a voltage between the tip and the sample. The stored charge on the capacitor plates causes an attractive force between the two electrodes, which is dependent on the dielectric constants of the materials within. Martin *et al.*[**?**] measure the force by a heterodyne detection technique. The voltage they apply is a combination of a DC bias voltage and an AC modulation voltage. The modulation frequency is set to a few kHz. In addition the *cantilever* is mechanically excited near its resonance frequency by a small piezo actuator. The two modulation frequencies are demodulated in two lock-in amplifiers. The *signal* near the resonance frequency of

the *cantilever* is a measure of the distance between the *cantilever* and the surface. The amplitude of the voltage modulation induced *signal*, however, is determined by the capacitance between the tip and the sample and by the charges present in between. A change in the force gradient will change the amplitude of the oscillation of the *cantilever*.

The force gradient $f(z)$ is given by

$$f(z) = \frac{1}{2}V^2\frac{\partial C}{\partial z} \tag{4.10.209}$$

where $V$ is the applied voltage, $C$ the capacitance and $z$ the separation between the tip and the sample.

Martin *et al.*[?] calculate, that the minimum detectable capacitance is $C_min = 8 \times 10^{-22}$ F in a 1 Hz bandwidth, measured with a silicon *cantilever* with a spring constant of 2.5 N/m having a resonance frequency of 33 MHz and a $Q$ of 200.

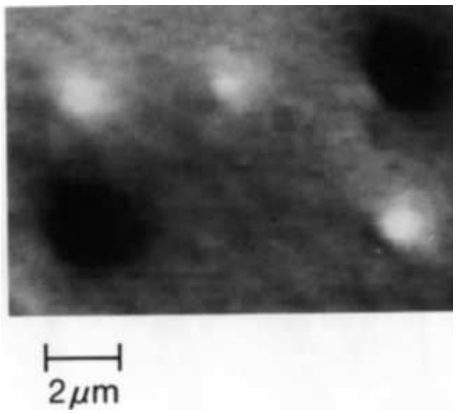



Abbildung 4.418.: Images of single electric charges by force microscopy. Taken from Terris *et al.*[?]. Used with permission from the American Physical Society.

An example of such a measurement is shown in figure 4.418. Single charges were deposited triboelectrically by shooting small isolating spheres on the surface[?].

### 4.10.17.19. Magnetic Forces

The first application of the *SFM* to forces other than the interatomic forces was the magnetic force microscope[?]. The magnetic domain structure of the sample was measured by scanning it past a tip made of a ferromagnetic material. The interaction between the tip and the sample can be selected such that the magnetic moments of the tip and the sample dominate the force. Sáenz *et al.*[?] have shown, that the force between the tip and the sample can be modelled as the magnetic dipole force. Both the tip and the sample are assumed to consist of microscopic magnetic domains with random orientation. The domains farther away from the nearest point between tip and sample tend to cancel their respective forces. Hence Sáenz *et al.*[?] were calculating the force $F(z)$ between the tip and the sample using the force $f_z(\boldsymbol{r})$ between two dipoles

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

$$F(z) = \int_{tip} d\boldsymbol{r}_1 \int_{sample} d\boldsymbol{r}_2 f_z(\boldsymbol{r}_1 - \boldsymbol{r}_2) \tag{4.10.210}$$

$$f_z(\boldsymbol{r}) = \left(\frac{\mu_0}{4\pi}\right) \frac{\partial}{\partial z} \left( \frac{3(\boldsymbol{r}\boldsymbol{\mu}_1)(\boldsymbol{r}\boldsymbol{\mu}_2)}{r^5} - \frac{(\boldsymbol{\mu}_1\boldsymbol{\mu}_2)}{r^3} \right). \tag{4.10.211}$$

$\mu_0$ is the permeability of the vacuum, $\boldsymbol{\mu_1}$ and $\boldsymbol{\mu_2}$ are the tip- and sample magnetic dipole, respectively. The two dipoles are assumed to be separated by the distance $\boldsymbol{r} = \boldsymbol{r}_1 - \boldsymbol{r}_2$.
The first consequence of equation (4.10.210) is, that a sample with a uniform magnetization will not exert a force on the tip. Only domain walls will be visible in the magnetic force microscope.
By assuming a spherical tip, Sáenz *et al.*[**?**] showed, that the force between the tip and the sample as a function of the distance $x$ between the tip and the domain wall was given by

$$F(x) = F_0 \frac{a x_r + b}{x_r^2 + 1} \tag{4.10.212}$$

where $x_r = x/L$ is the relative separation between the tip and the domain wall, $L \gg z$ is the radius of curvature of the tip, and $F_0 = 8/\pi\mu_0\mu_1\mu_2 L^2$. $a$ and $b$ are two constants depending on the spin states of the sample and the tip. The magnitude of the force is calculated to be of order $10^{-11}$ N to $10^{-10}$ N. This is one to two orders of magnitude smaller than the forces used in repulsive imaging.

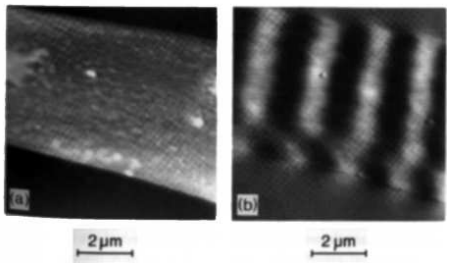



Abbildung 4.419.: Magnetic force microscopy of tracks written on a magnetic storage medium. The image was taken by Schönenberger and Alvarado[**?**]. Used with permission from Springer Verlag, Heidelberg.

A detailed account of the problems of magnetic force imaging can be found in the paper of Schönenberger and Alvarado[**?**]. Figure 4.419 gives an example of a magnetic force image.

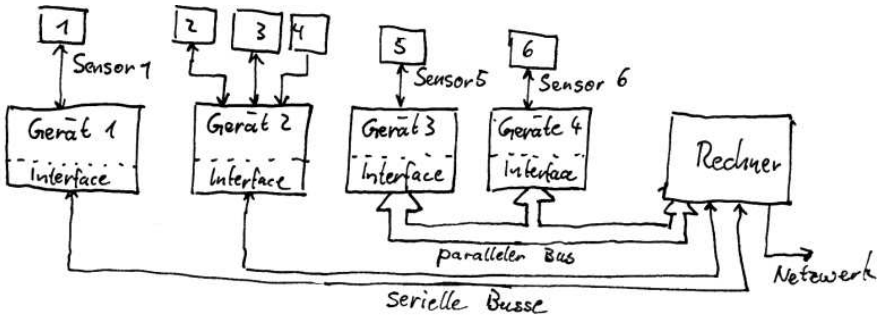



Abbildung 4.420.: Beispiel des prinzipiellen Aufbaus eines Messsystems mit externen Geräten

## 4.11. Rechnergestützte Messtechnik

Heutzutage werden komplizierte und langwierige Messreihen häufig mit Rechner kontrolliert oder durchgeführt. Wenn über mehrere Tage hinweg Daten erfasst werden müssen, wenn in kürzester Zeit grosse Datenmengen anfallen, wenn schwierige Steuerungsaufgaben gelöst werden müssen, dann ist es vorteilhaft, eine rechnergestützte Messtechnik einzusetzen. In den folgenden Abschnitten werden zuerst Messtechniken mit externen Geräten und dann der Einsatz von rechnerinternen Datenerfassungskarten besprochen.

### 4.11.1. Verwendung externer Geräte

Externe *Messgeräte* können meistens auf mehrere Arten betrieben werden:

- Als selbständige Geräte
- Als selbständige Geräte, deren Messdaten und Einstellungen an Rechner übertragen werden.
- Als Geräte, die vollständig durch die Rechner gesteuert werden.

#### 4.11.1.1. Bussysteme

Abbildung 4.420 zeigt den prinzipiellen Aufbau eines Messsystems. Vier *Messgeräte* sind an einen Rechner angeschlossen. Zwei davon über serielle Busse und zwei über einem gemeinsamen parallelen Bus. Beispiele für Bussysteme sind in den Tabellen 4.16 und 4.17 zusammengestellt. Die *Messgeräte* ihrerseits sind mit einem oder mehreren Sensoren verbunden. Anstelle der *Messgeräte* können mit Bussystemen auch Quellen von Rechnern aus gesteuert werden. Die in Abb. 4.420 gezeigten *Messgeräte* können mit einem oder mehreren Sensoren verbunden sein.
Die Leistungsfähigkeit der Datenübertragung hängt im wesentlichen von der Bustopologie und der Datenbreite in den Bussen ab. Die Tabelle 4.16 zeigt allgemein übliche Bussysteme. Diese sind in den meisten Rechnern zu finden. Parallele Systeme wie die Druckerschnittstelle dienen zur Datenkommunikation mit sehr preisgünstigen Geräten. Die Übertragungsgeschwindigkeit ist jedoch immer noch grösser als bei den seriellen Systemen, wie zum Beispiel der RS-232-Schnittstelle. Bei der seriellen Schnittstelle ist nur eine Punkt-zu-Punkt-Verbindung möglich.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

| **Schnittstelle** | **Datenrate** [Byte/$s$] | **Beschreibung** |
|---|---|---|
| Centronics-Schnittstelle | $> 4 \times 10^4$ | parallel (8 Bit und 4 Bit[29]) |
| ECP/EPP (IEEE 1284) | $> 6 \times 10^5$ | parallel (8 Bit) |
| RS-232-C | $< 1.4375 \times 10^4$ | seriell (1 Bit) |
| (V24 bzw V28) | | |
| PCMCIA (Rev. 2.1) | $> 2.4 \times 10^7$ | parallel (16 Bit) |
| PC-Card | $< 1.32 \times 10^8$ | parallel (32 Bit) |
| (PCMCIA Rev. 5) | | |
| SCSI | $< 4 \times 10^7$ | parallel (8, 16 oder 32 Bit) |
| USB | $1.5 \times 10^6$ | seriell (1 Bit) |
| FireWire | $1.25 \ldots 5 \times 10^7$ | 2× seriell (2 Bit) |
| IrDA | | |
| FC-AL Fiber Channel | $1 \ldots 4 \times 10^8$ | seriell (1 Bit) |

Tabelle 4.16.: Häufig vorkommende Bussysteme (nach [**?**])

Bei der parallelen Schnittstelle können mehrere Geräte an einen Bus geschaltet werden: es zeigt sich jedoch in der Praxis, dass mehr als ein Geräte ausser dem Rechner Kompatibilitätsprobleme geben kann.
Das SCSI-Bussystem ist eines der leistungsfähigsten, schon gut eingeführten Bussysteme für den Anschluss von mehreren Geräten. Die Datentransferrate ist ähnlich hoch wie bei internen Bussen.
Neuere Geräte sind alle mit dem USB-System[30], einem seriellen Bussystem, an dem mehrere Geräte angeschlossen werden können. Dieses Bussystem gilt als das zukünftige Low-Cost-Bussystem. Eine bessere Leistung hat das FireWire oder IEEE1394[31] System. Dieses wird gegenwärtig zur Datenübertragung im High-End-Bereich für Privathaushaltungen. Im Gegensatz zur SCSI-Technologie sind diese beiden Bussysteme preisgünstig. Beide haben relativ einfache Stecker. Die Treiber sind so angelegt, dass die Geräte im Betrieb ein- und ausgesteckt werden können. Die Fire-wire-Technologie sollte bis in den $10^8$ Bit/s- Bereich ausbaubar sein. Im Gegensatz zur USB-Technologie ist eine isochrone Übertragung neben der sonst üblichen asynchronen Datenübertragung vorgesehen. Damit können über eine FireWire-Schnittstelle auch zeitkritische Daten, wie zum Beispiel Videodaten übertragen werden.
Für Laboranwendungen werden meistens Geräte, die auf dem IEEE-488-Bus beruhen, verwendet. In Fabriken wird oft auch der VME-Bus angewandt. Wenn grosse Distanzen zu überbrücken sind, fällt die Wahl auf serielle Busse, wie den RS-232, den RS-422 oder den RS-485. Tabelle 4.17 gibt einen Überblick über diese Bussysteme.
Abbildung 4.421 zeigt die Steckerbelegung des IEEE-488-Busses. Dieser Bus ist, was die Kabel anbetrifft, identisch mit dem IEEE-488-Bus. Die Steckerbelegung ist jedoch unterschiedlich.
Die einzelnen Schnittstellenleitungen haben die folgenden Bedeutung:

**DAV (DAta Valid)** ein sendendes Gerät (Talker) hat gültige Daten.

[30] http://www.usb.org
[31] http://www.ti.com/sc/1394

| **Schnittstelle** | **Datenrate** [Byte/$s$] | **Beschreibung** |
|---|---|---|
| 4 . . . 20 mA, Current Loop, TTY | mehrere 10 | seriell (1 Bit) |
| RS-232-C (V24 bzw V28) | $< 1.4375 \times 10^4$ | seriell (1 Bit) |
| RS-422 (V.11) | $1.25 \times 10^4$ bis $1.25 \times 10^6$ | seriell (1 Bit) |
| RS-485 | $> 1.25 \times 10^5$ | seriell (1 Bit) |
| PCMCIA (Rev. 2.1) | $> 2.4 \times 10^7$ | parallel (16 Bit) |
| IEEE-488 (HPIB, IEC625,GPIB) | $< 1 \times 10^6$ | parallel (8 Bit) |
| VXI (MXI) | $> 4 \times 10^7$ | parallel (16 oder 32 Bit) |
| Feld-Busse (CAN, Profibus ...) | mehrere $10^2$ bis $1.25 \times 10^6$ | seriell (1 Bit) |

Tabelle 4.17.: Bevorzugte Bussysteme zur Messdatenerfassung (nach [**?**])

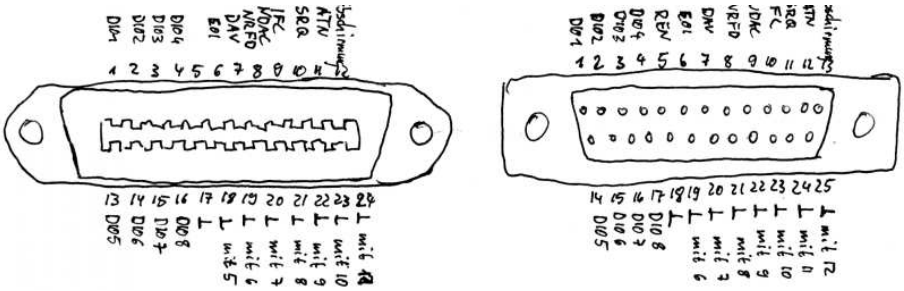


Abbildung 4.421.: Seckerbelegung beim IEEE-488-Bus beziehungsweise beim IEC-625-Bus

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

**NRFD (Not Ready For Data)** Nicht alle Geräte sind für den Empfang bereit. Über den IEEE-488-Bus können erst dann Daten übertragen werden, wenn alle Geräte bereit sind. Die NRFD-Ausgänge aller Geräte sind miteinander verknüpft. Das Parallelschalten von Open-Collector-Ausgängen bedeutet, dass eine Wired-AND-Verknüpfung vorliegt.

**NDAC (Not Data ACcepted)** Die Daten sind noch nicht von allen Geräten übernommen worden.

**ATN (ATtentioN)** Liegt an dieser Leitung ein Null-Signal (H-Niveau) an, dann sind Daten auf den Datenleitungen. Liegt eine eins an (L-Niveau), werden über die Datenleitungen Adressen oder Befehle übertragen.

**IFC (InterFace Clear)** Der Buscontroller steuert diese Leitung. Mit ihr werden die angeschlossenen Geräte nach dem Einschalten oder bei Netzausfall in einen definierten Zustand gebracht.

**SRQ (Service ReQuest)** Wenn eines oder mehrere Geräte eine Aktion des Controllers benötigen, wird diese Leitung aktiviert. Der Controller prüft einzeln (serial poll) oder parallel (parallel poll), welches Gerät sich gemeldet hat. Mit diesem Mechanismus kann jedes Gerät sich Zugriff auf den Computer verschaffen.

**REN (Remote ENable)** Dieses Signal bereitet die Angeschlossenen Geräte auf Fernbedienung vor. Damit werden zum Beispiel die Bedienungselemente an den Frontplatten deaktiviert.

**EOI (End Or Identify)** Wenn sich der Bus im Datenmodus befindet, kennzeichnet dieses Signal das letzte Byte des Übertragungsblockes. Befindet sich der Bus im Befehlsmodus, wird über diese Leitung eine parallele Abfrage eingeleitet.

Die Adressenleitung des IEEE-488-Busses liegt auf den gleichen Leitungen wie die 8 Datenleitungen. Davon werden 5 Leitungen zur Adressierung verwendet. Die sechste Leitung wird zur Kennzeichnung des Gerätes als Empfänger (Listener) verwendet. Ebenso kennzeichnet Bit 7 das Gerät als Talker. Von den zweiunddreissig möglichen Adressen ist eine für ein Kommando vordefiniert. Deshalb können an einem IEEE-488-Bus ausser dem Controller, das heisst dem Steuerrechner, noch dreissig Geräte angeschlossen werden.

#### 4.11.1.2. Geräte und Zusammenschaltung von Geräten

Abbildung 4.422 zeigt ein Beispiel einer Zusammenschaltung von Geräten. An einem IEEE-488-Bus können die folgenden Gerätetypen angeschlossen werden:

**Listener** Dies ist ein Gerät, das nur Befehle entgegennimmt und keine Daten zurücksendet. Ein typisches Beispiel für einen Listener ist ein Funktionsgenerator oder eine gesteuerte Quelle.

**Talker** Dies ist ein Gerät, das nur Daten liefert, aber keine Daten entgegennimmt. Ein Voltmeter ist ein Beispiel für einen Talker.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

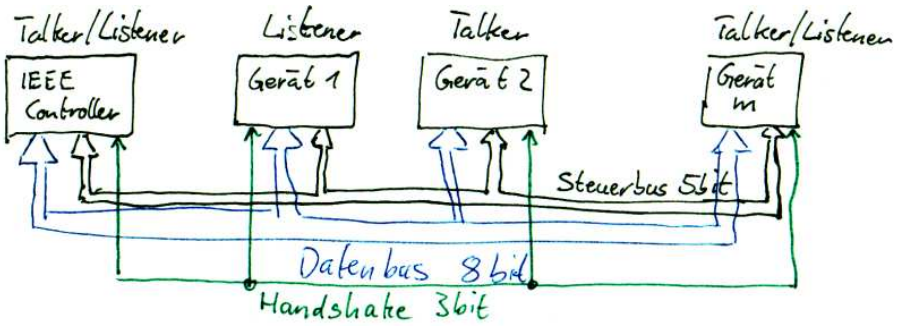



Abbildung 4.422.: Struktur eines Datenerfassungssystems mit dem IEEE-488-Bus

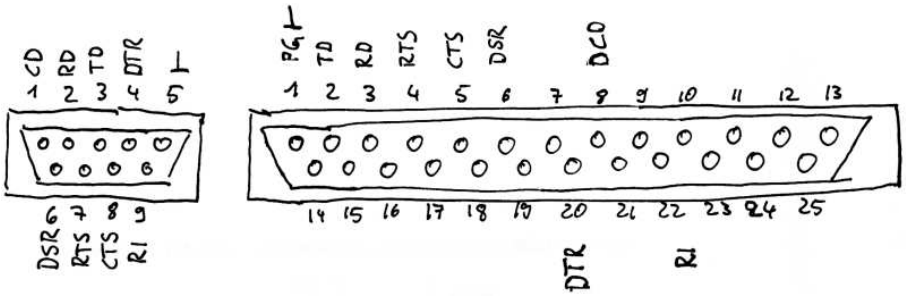



Abbildung 4.423.: 9-polige (links) und 25-polige Anschlüsse für die RS-232-Schnittstelle.

**Talker/Listener** Dies ist ein Gerät, das sowohl Befehle und Daten entgegennimmt und das Daten zurückliefert. Ein typisches Beispiel für ein solches Gerät ist ein Lock-in-Verstärker. Als Listener übernimmt das Gerät Einstellungen wie Frequenz und Amplitude der Referenzspannung. Es liefert die momentane Amplitude und Phase als Talker zurück. Genau besehen dürften auch die meisten Voltmeter sowohl Listener wie auch Talker sein.

**Controller** Ein Controller ist der Talker/Listener, der den Bus kontrolliert.

Die Abbildung 4.423 zeigt die beiden gebräuchlichen Anschlusstypen für die RS-232-Schnittstelle. Die beiden Geräte an den Enden einer RS-232-Schnittstelle sind die Datenendeinrichtung (DTE) und die Datenübertragungseinrichtung (DCE). Mit der ersten Bezeichnung ist der Rechner gemein, mit der zweiten die Modems oder das zu steuernde Gerät. Die Leitungen dieser Schnittstelle haben die folgende Bedeutung:

**D1=TD (Transmit Data)** Auf dieser Leitung gibt der Rechner die Daten an das zu steuernde Gerät aus.

**D2=RD (Receive Data)** Auf dieser Leitung empfängt der Rechner die Daten vom angeschlossenen Gerät.

**RTS (Ready To Send)** Auf dieser Leitung fordert der Rechner das angeschlossene Gerät zum senden auf.

**CTS (Clear To send)** Mit dieser Leitung zeigt der Rechner seine Sendebereitschaft an.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

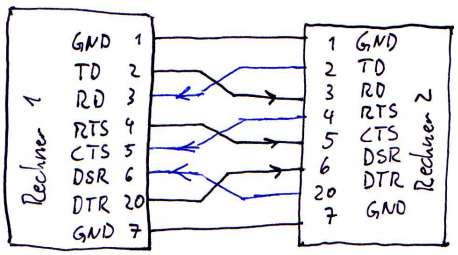



Abbildung 4.424.: Verbindung zweier Rechner

**M1=DSR (Data Set Ready)** Ein Signal auf dieser Leitung zeigt an, dass das angeschlossene Gerät bereit ist.

**M2=DCD (Data Carrier Device)** Mit diesem Signal zeigt das angeschlossene Gerät dem Rechner an, dass ein Signal im richtigen Spannungsbereich anliegt.

**T2=TC (Transmit Clock)** Das angeschlossene Gerät sendet auf dieser Leitung dem Rechner das Taktsignal.

**T4=RC (Receive Clock)** Auf dieser Leitung wird dem Rechner der Empfangsschrittakt mitgeteilt.

**S1.2=DTR (Data Terminal Ready)** Auf dieser Leitung teilt der Rechner mit, dass er sendebereit sei.

**M3=RI (Ring Indicator)** Auf dieser Leitung teilt das angeschlossene Gerät dem Rechner mit, dass es senden möchte.

Die Verbindung zweier Rechner mit einer RS-232-Schnittstelle wird in Abbildung 4.424 gezeigt.
Die Verwendung externer *Messgeräte* erlaubt einen flexiblen Aufbau des Messsystems. Als *Messgeräte* können die besten erhältlichen Geräte verwendet werden.
Die Messung eines Frequenzspektrums mit einer *Spannungsquelle* zum Einstellen einer Heizleistung, einem Voltmeter zur Messung der Temperatur über einen Pt100-Widerstand sowie einem Lock-in-Verstärker könnte so aussehen:

```
Bus initialisieren;
for Heizstrom = Startwert to Endwert do
  begin
    Heizstrom  setzen;
    Temperatur messen
    Repeat
      Temperatur messen
    until Temperatur aendert sich nicht mehr als 0.1 K
    Temperatur speichern
    Amplitude am Lock-In-Verstaerker auf Startwert setzen
    Frequenz am Lock-In-Verstaerker auf Startwert setzen
    Messverstaerkung setzen
    Phase abgleichen
```

| **Name** | **Leitungen** | **Datenrate** [Byte/$s$] | **Rechnertypen** |
|---|---|---|---|
| ISA (PC-Bus) | 8D,20A | $8 \times 10^6$ | IBM-PC |
| ISA (AT-Bus) | 16D,24A | $1.2 \ldots 2.4 \times 10^7$ | IBM-PC, IPC, DEC-Alpha |
| MCA | 16D oder 32D, 24A | $> 1.3 \times 10^8$ | PS/2, IBM-RS3000/4000 |
| PCI | 32D oder 64D, 32A | $> 2.5 \times 10^8$ | IBM-PC,MAC, DEC-Alpha,IPC |
| CompactPCI, IPCI | 32D oder 64D, 32A | $> 2.5 \times 10^8$ | IPC, VMEbus-Rechner |
| EISA | 32D, 32A | $> 1.3 \times 10^8$ | IBM-PC, DEC |
| VMEbus | 16D,32D oder 64D, 16A oder 32A | $< 8 \times 10^7$ | Workstations, 68K-Systeme |
| VXIbus | 16D oder 32D, 32A | $< 8 \times 10^7$ | Workstations |
| PC Card (PCMCIA) | 16D oder 32D, 26A oder 32A | $2.4 \ldots 13.2 \times 10^7$ | IBM-PC, IPC, MAC |

Tabelle 4.18.: Bussysteme in Rechnern (nach [**?**])

```
    for Frequenz = Startwert to Endwert do
      begin
        Frequenz setzen
        2 s warten
        Amplitude und Phase ablesen
        Messwerte speichern
        Messwerte darstellen
      end
  end
Messgeraete in Ausgangszustand setzen
Bus freigeben
```

## 4.11.2. In Rechner eingebaute Messdatenerfassung

Häufig werden heute auch in Rechner eingebaute Datenerfassungskarten verwendet. Bevorzugt werden Universalkarten mit einer gewissen Anzahl von analogen und digitalen Eingängen und Ausgängen sowie mit einigen Zeitgebern.

### 4.11.2.1. Bussysteme

Die Leistungsfähigkeit von rechnergestützten Datenerfassungssystemen hängt von der Qualität und der Geschwindigkeit der in den Rechnern verwendeten Bussysteme ab, und weniger von deren Prozessorleistung. Die Tabelle 4.18 gibt einen Überblick über die Bussysteme. Heute werden hauptsächlich

- PCI-Systeme in Rechnern, die ursprünglich für den Bürobedarf entwickelt wurden.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

- VME-Busse in dedizierten Steuerrechnern mit älterer Konzeption
- VXIbus-System in neueren dedizierten Steuerrechnern

verwendet. Neben der Hardware bestimmt das verwendete Betriebssystem wesentlich die Leistungsfähigkeit von Rechnern. Die Betriebssysteme können in die folgenden Kategorien eingeteilt werden:

- Single Task Betriebssysteme sind gut geeignet für einfache Aufgaben (Beispiel MSDOS)
- Multitasking-Betriebssysteme mit kooperativem Task-Scheduling sind ungeeignet. (Beispiel Windows 3.1, Windows 95, Windows 98)
- Multitasking-Betriebssysteme mit Zeitscheiben-Steuerung sind bedingt geeignet. (Beispiele: Windows NT, Linux)
- Multitasking-Betriebssysteme mit Real-Time Kernel sind hervorragend geeignet. Der Real-Time Kernel ist so aufgebaut, dass zeitkritische Applikationen eine garantierte Reaktionszeit haben. (Beispiele: RT-OS, QNX, OS9, Linux mit Real-Time-Kernel)

Heutige Desktoprechner sind oft schlecht für Messaufgaben geeignet, da die Hersteller aus Kostengründen nur noch wenige freie Steckplätze vorsehen.

### 4.11.3. Übersicht über Programme zur Datenerfassung

Traditionellerweise werden Messaufgaben durch einzelne in Hochsprachen geschriebene Programme durchgeführt. Dabei werden die einzelnen Einheiten, seien es Steuerkarten und Schnittstellen für externe Geräte oder interne Datenerfassungskarten entweder direkt oder über vom Hersteller gelieferte Bibliotheken angesprochen. Diese Art der Programmierung ist heute nur noch gerechtfertigt, wenn extrem hohe Datenströme oder selbstgebaute Instrumente bedient werden müssen.
In allen anderen Fällen sollte man Grafik-orientierte Enwicklungswerkzeuge verwenden. Die bekanntesten dieser Programme sind

- HP-Vee
- Labview
- Testpoint

Die Tabelle 4.19 gibt einen Überblick über einige der datenerfassungsprogramme. Bei diesen graphischen Entwicklungswerkzeugen werden Messprozesse grafisch mit vordefinierten Instrumenten oder, sollten diese nicht vorhanden sein, mit selbstgeschriebenen DLLs oder Active-X-Steuerelementen nachgebildet. Mit dieser Philosophie kann gewissermassen eine Verkabelung wie im Labor auf dem Rechner nachgebildet werden. Die Vorteile dieser Entwicklungsumgebungen sind:

- ein besserer Überblick über das Messsystem
- Datenflussplan

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

| **Name** | **Eigenschaften** | **Betriebssystem** | **Anbieter** |
|---|---|---|---|
| BEAM | Konfiguration von Funktionsmodulen | Dos/Win 3.x, Win 95, Win NT, MacOS | AMS, Flöha |
| DasyLab | Grafische Konfiguration mit Funktionsblöcken in Datenflussdiagrammen | Dos/Wn 3.x, Win 95, Win NT | Datalog, Mönchengladbach |
| DIAdem | Grafische Konfiguration mit Funktionsblöcken in getrennten, applikationsspezifischen Programmmodulen | Dos/Win 3.x, Win 95 | Gesellschaft für Strukturanalyse, Aachen |
| HP-VEE | Grafische Konfiguration und Programmierung mit Funktionsblöcken und virtuellen Instrumenten in Datenflussdiagrammen | DOS/Win 3.x, Win 95, Win NT, UNIX | Hewlett-Packard, München |
| LabView | Grafische Programmierung in 'G' mit Funktionsblöcken und virtuellen Instrumenten | DOS/Win 3.x, Win 95, Win NT, MacOS, UNIX | National Instruments, München |
| Lab-Windows / CVI | Textbasierte Programmierung in 'C' und mit virtuellen Instrumenten | DOS/Win 3.x, Win 95, Win NT, MacOS, UNIX | National Instruments, München |
| Testpoint | Grafische Konfiguration mit Funktionsblöcken, textbasierte Programmierung von Ablauf und Funktionsparametrierung | DOS/Win 3.x, Windows 95 | Keithley Instruments, Germering |
| Visual Designer | Grafische Konfiguration mit Funktionsblöcken in Datenflussdiagrammen | DOS/Win 3.x, Win 95, Win NT | Intelligent Instrumentation, Leinfelden |

Tabelle 4.19.: Beispiele grafischer Datenerfassungsprogramme (nach [**?**])

- eine implizite Dokumentation des Messaufbaus (was sonst leicht vergessen geht)
- die Möglichkeit, Nebenmessungen einzubinden und so katastrophales Versagen der Messung bei gewissen, nicht immer vorhersagbaren Bedingungen zu verhindern.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

# A. Physikalische Grundlagen

Zu den physikalischen Grundlagen der Physikalischen Elektronik und Messtechnik gehören die Maxwellschen Gleichungen, die Kirchhoffschen Gesetze sowie das rechnen mit komplexen Spannungen und Strömen.

## A.1. Maxwellsche Gesetze

Die Maxwellschen Gesetze in differentieller Formulierung lauten:

$$\mathbf{rot}\ \boldsymbol{H} = \boldsymbol{j} + \dot{\boldsymbol{D}} \tag{A.1.1}$$
$$\mathbf{rot}\ \boldsymbol{E} = -\dot{\boldsymbol{B}} \tag{A.1.2}$$
$$\mathbf{div}\ \boldsymbol{D} = \rho \tag{A.1.3}$$
$$\mathbf{div}\ \boldsymbol{B} = 0 \tag{A.1.4}$$

In integraler Form lauten die obigen Gleichungen:

$$\oint \boldsymbol{H} d\boldsymbol{s} = \int_A \left(\boldsymbol{j} + \dot{\boldsymbol{D}}\right) d\boldsymbol{A} \tag{A.1.5}$$
$$\oint \boldsymbol{E} d\boldsymbol{s} = -\int_A \dot{\boldsymbol{B}} d\boldsymbol{A} \tag{A.1.6}$$
$$\int_A \boldsymbol{D} d\boldsymbol{A} = Q \tag{A.1.7}$$
$$\int_A \boldsymbol{B} d\boldsymbol{A} = 0 \tag{A.1.8}$$

Zusätzlich benötigt man noch die Materialgleichungen

$$\boldsymbol{D} = \overline{\overline{\varepsilon}}\boldsymbol{E} \tag{A.1.9}$$
$$\boldsymbol{B} = \overline{\overline{\mu}}_r \mu_0 \boldsymbol{H} \tag{A.1.10}$$

## A.2. Kirchhoffsche Gesetze

Die Kirchhoffschen Gesetze gelten für **stationäre und quasistationäre** Anordnungen von Leitern und Bauelementen. Die Knotenregel kann auf die Ladungserhaltung zurückgeführt werden. Sie lautet:

$$\sum_{j=1}^{n} I_j = 0 \tag{A.2.1}$$

Die Maschenregel (entgegen anders lautenden Gerüchten hat sie nichts mit Nachbarschaftsstreitereien zu tun) kann im stationären und quasistationären Fall auf Eigenschaften konservativer Potentiale zurückgeführt werden.

$$\sum_{Masche} U_j = 0 \tag{A.2.2}$$

Hier muss beachtet werden, dass Quellen und Verbraucher vorzeichenrichtig eingesetzt werden!

## A.3. Komplexe Spannungen und Ströme

Wechselströme und -spannungen können einerseits mit Hilfe der Winkelfunktionen, andererseits aber auch mit komplexen Variablen dargestellt werden.

$$U(t) = \hat{U}\cos(\omega t + \varphi_U) \tag{A.3.1}$$

$$I(t) = \hat{I}\cos(\omega t + \varphi_I) \tag{A.3.2}$$

Man kann die Phase des Stromes oder der Spannung auf null setzen, ohne dass die Physik des Problems sich ändert. Wir setzen $\varphi = \varphi_U - \varphi_I$ und nehmen nachher an, dass die Phase der Spannung $\varphi_U = 0$ ist. Die Gleichungen heissen dann:

$$U(t) = \hat{U}\cos(\omega t) \tag{A.3.3}$$

$$I(t) = \hat{I}\cos(\omega t + \varphi) \tag{A.3.4}$$

Die komplexe Schreibweise ist:

$$\begin{aligned} \underline{U}e^{j\omega t} &= \hat{U}e^{j(\omega t+\varphi_u)} \\ &= \hat{U}\left[\cos(\omega t + \varphi_u) + j\sin(\omega t + \varphi_u)\right] \\ &= \underline{U}\left[\cos\omega t + j\sin\omega t\right] \end{aligned} \tag{A.3.5}$$

Wenn man die Amplitude reell schreibt, muss man eine Phase angeben. Alternativ kann mit komplexen Amplituden gerechnet werden.
**Warnung! Das Rechnen mit komplexen Grössen funktioniert nur bei linearen Systemen!**
Aus komplexem Strom und komplexer Amplitude kann ein komplexer Widerstand, auch Impedanz genannt, berechnet werden.

$$\underline{Z} = \frac{\underline{U}e^{j\omega t}}{\underline{I}e^{j\omega t}} = \frac{\underline{U}}{\underline{I}} = \frac{\hat{U}}{\hat{I}}e^{j\varphi} \tag{A.3.6}$$

Beim Ohmschen Gesetz ergibt sich so $\underline{Z} = R$. Für eine Spule gilt zum Beispiel

$$\begin{aligned} U &= L\frac{dI}{dt} \\ \underline{U}e^{j\omega t} &= L\frac{d}{dt}\left(\underline{I}e^{j\omega t}\right) = j\omega L\underline{I}e^{j\omega t} \\ \underline{Z}_L &= j\omega L \end{aligned} \tag{A.3.7}$$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

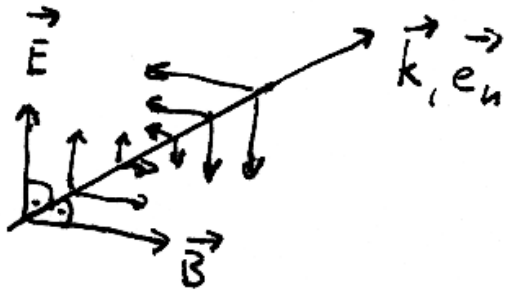



Abbildung A.1.: Schematische Darstellung einer elektromagnetischen Welle.

Analog gilt für den Kondensator

$$\begin{aligned} U &= \frac{Q}{C} & &= \frac{1}{C}\int I dt \\ \underline{U}e^{j\omega t} &= \frac{1}{C}\int \underline{I}e^{j\omega t}dt & &= \frac{1}{j\omega C}\underline{I}e^{j\omega t} \\ \underline{Z}_C &= \frac{1}{j\omega C} \end{aligned} \tag{A.3.8}$$

## A.4. Ebene Wellen

In einem Medium mit der relativen Dielektrizitätskonstanten $\varepsilon$ und der relativen Permeabilität $\mu$ ist eine der möglichen Lösungen der Maxwellgleichungen (A.1.2) die Wellenlösungen[?]. Wir setzen an:

$$\boldsymbol{E}(\boldsymbol{r},t) = \underline{\boldsymbol{E}}e^{-j\boldsymbol{k}\cdot\boldsymbol{r}+j\omega t} + c.c. \tag{A.4.1}$$

$$\boldsymbol{B}(\boldsymbol{r},t) = \underline{\boldsymbol{B}}e^{-j\boldsymbol{k}\cdot\boldsymbol{r}+j\omega t} + c.c. \tag{A.4.2}$$

und erhalten mit der ersten Maxwellgleichung $\boldsymbol{\nabla E} = -\dot{\boldsymbol{B}}$. Somit ist

$$-j\cdot\boldsymbol{k}\times\underline{\boldsymbol{E}} = -j\cdot k\boldsymbol{e}_n\times\underline{\boldsymbol{E}} = -j\omega\boldsymbol{B} \tag{A.4.3}$$

Als Folge der anderen Maxwellgleichungen haben wir für den Wellenvektor $k = \omega/c$ und für die Phasengeschwindigkeit im Medium $c = c_0/\sqrt{\mu\varepsilon}$ sowie für die Phasengeschwindigkeit im Vakuum $c_0 = 1/\sqrt{\mu_0\varepsilon_0}$. Also wird Gleichung (A.4.3)

$$\underline{\boldsymbol{B}} = \frac{\boldsymbol{e}_n\times\underline{\boldsymbol{E}}}{c} \tag{A.4.4}$$

Daraus können wir den Wellenwiderstand im Medium als das Verhältnis der Amplituden des elektrischen und des magnetischen Feldes bestimmen.

$$Z = \frac{\underline{E}}{\underline{H}} = \mu\mu_0\frac{\underline{E}}{\underline{B}} = \frac{\mu\mu_0}{\sqrt{\varepsilon\varepsilon_0\mu\mu_0}} = \sqrt{\frac{\mu}{\varepsilon}}\sqrt{\frac{\mu_0}{\varepsilon_0}} = \sqrt{\frac{\mu}{\varepsilon}}Z_0 \tag{A.4.5}$$

Dabei ist

$$Z_0 \equiv \sqrt{\frac{\mu_0}{\varepsilon_0}} = 120\pi\,\Omega \approx 377\,\Omega \tag{A.4.6}$$

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

Der Wellenwiderstand charakterisiert offensichtlich die Ausbreitungseigenschaften von **T**ransversalen **E**lektro**M**agnetischen Wellen oder **TEM**-Wellen.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

# B. Berechnung von Schaltungen

## B.1. Brückenschaltung mit Widerständen

Die unten folgende Berechnung bezieht sich auf Abb. 4.54. Wir beginnen mit den Bestimmungsgleichungen

$$\begin{aligned} U_1 &= R_4 I_4 && \text{(B.1.1)} \\ U - U_1 &= R_1 I_1 && \text{(B.1.2)} \\ U_2 &= R_3 I_3 && \text{(B.1.3)} \\ U - U_2 &= R_2 I_2 && \text{(B.1.4)} \\ I_1 &= I_4 + I_i && \text{(B.1.5)} \\ I_3 &= I_2 + I_i && \text{(B.1.6)} \\ U_1 - U_2 &= R_i I_i && \text{(B.1.7)} \end{aligned}$$

Aus den Gleichungen (B.1.1), (B.1.3) und (B.1.7) folgt:

$$\begin{aligned} R_i I_i &= R_4 I_4 - R_3 I_3 \\ I_4 &= \frac{R_i I_i + R_3 I_3}{R_4} \end{aligned} \qquad \text{(B.1.8)}$$

Aus den Gleichungen (B.1.1), (B.1.2) und (B.1.5)folgt:

$$U = I_4 (R_1 + R_4) + R_1 I_i \qquad \text{(B.1.9)}$$

Aus den Gleichungen (B.1.3), (B.1.4) und (B.1.6)folgt:

$$U = I_3 (R_2 + R_3) - R_2 I_i \qquad \text{(B.1.10)}$$

Endlich erhält man aus den Gleichungen (B.1.9) und (B.1.10) die folgende Beziehung:

$$U = I_3 \frac{R_3}{R_4} (R_1 + R_4) + I_i \left[ \frac{R_i}{R_4} (R_1 + R_4) + R_1 \right] \qquad \text{(B.1.11)}$$

Das Schlussresultat, die Gleichung (4.1.67), erhält man durch Kombination der Gleichungen (B.1.10) und (B.1.11).

$$I_i = U \frac{R_2 R_4 - R_1 R_3}{R_1 R_4 (R_2 + R_3) + R_2 R_3 (R_1 + R_4) + R_i (R_1 + R_4)(R_2 + R_3)}$$

Die Empfindlichkeit der Anordnung erhält man, indem man nach den variablen Widerständen, $R_1$ oder $R_4$ ableitet. Die anderen beiden Widerstände ergeben jeweils äquivalente Resultate.

$$\frac{\partial I_i}{dR_1} = \frac{UR_3}{R_i\left(R_1+R_4\right)\left(R_2+R_3\right)+R_1R_4\left(R_2+R_3\right)+R_2R_3\left(R_1+R_4\right)} - \frac{U\left(R_2R_4-R_3R_1\right)\left(\left(R_i+R_4\right)\left(R_2+R_3\right)+R_2R_3\right)}{\left(R_i\left(R_1+R_4\right)\left(R_2+R_3\right)+R_1R_4\left(R_2+R_3\right)+R_2R_3\left(R_1+R_4\right)\right)^2} \tag{B.1.12}$$

$$\frac{\partial I_i}{dR_4} = \frac{UR_2}{R_i\left(R_1+R_4\right)\left(R_2+R_3\right)+R_1R_4\left(R_2+R_3\right)+R_2R_3\left(R_1+R_4\right)} - \frac{U\left(R_2R_4-R_3R_1\right)\left(\left(R_i+R_1\right)\left(R_2+R_3\right)+R_2R_3\right)}{\left(R_i\left(R_1+R_4\right)\left(R_2+R_3\right)+R_1R_4\left(R_2+R_3\right)+R_2R_3\left(R_1+R_4\right)\right)^2} \tag{B.1.13}$$

Die Schlussresultate finden Sie im Abschnitte 4.1.7

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

# C. Tabellen

## C.1. Tabelle der Laplacetransformationen

| | $f(t)$, $(t > 0)$ | $F(p) = \int_0^\infty f(t)\, e^{-pt} dt$ |
|---|---|---|
| 1. | $\frac{t^n}{\Gamma(n+1)}$ | $\frac{1}{p^{n+1}}$ |
| 2. | $e^{-at}$ | $\frac{1}{p+a}$ |
| 3. | $\sin(kt)$ | $\frac{k}{p^2+k^2}$ |
| 4. | $\cos(kt)$ | $\frac{p}{p^2+k^2}$ |
| 5. | $e^{-at}\sin(kt)$ | $\frac{k}{(p+a)^2+k^2}$ |
| 6. | $e^{-at}\cos(kt)$ | $\frac{(p+a)}{(p+a)^2+k^2}$ |
| 7. | $t\sin(kt)$ | $\frac{2kp}{(p^2+k^2)^2}$ |
| 8. | $t\cos(kt)$ | $\frac{\left(p^2-k^2\right)}{(p^2+k^2)^2}$ |
| 9. | $e^{-at}\frac{t^n}{n!}$ | $\frac{1}{(p+a)^{n+1}}$ |
| 10. | $\frac{(2t)^n}{1\cdot 3\cdot 5\ldots(2n-1)\sqrt{\pi t}}$ | $\frac{\sqrt{p}}{p^{n+1}}$ (n ganzzahlig >0) |
| 11. | $\frac{1}{\sqrt{\pi t}} e^{-\frac{a^2}{4t}}$ | $\frac{1}{\sqrt{p}} e^{-a\sqrt{p}}$ (a>0) |
| 12. | $\frac{a}{2\sqrt{\pi t^3}} e^{-\frac{a^2}{4t}}$ | $e^{-a\sqrt{p}}$ |
| 13. | $J_0(t)$ (Besselfunktion) | $\frac{1}{\sqrt{1+p^2}}$ |
| 14. | $I_0(t)$ (Besselfunktion) | $\frac{1}{\sqrt{p^2-1}}$ |
| 15. | $\int_t^\infty \frac{e^{-x}}{x} dx$ | $\frac{\ln(1+p)}{p}$ |

Tabelle C.1.: Tabelle der Laplacetransformationen einiger ausgewählter Funktionen

# C.2. Tabelle der Carson-Heaviside-Transformation

| | $f(t)$, $(t>0)$ | $F(p)=p\int_0^\infty f(t)\,e^{-pt}dt$ |
|---|---|---|
| 1. | $\frac{t^n}{\Gamma(n+1)}$ | $\frac{1}{p^n}$ |
| 2. | $e^{-at}$ | $\frac{p}{p+a}$ |
| 3. | $\sin(kt)$ | $\frac{pk}{p^2+k^2}$ |
| 4. | $\cos(kt)$ | $\frac{p^2}{p^2+k^2}$ |
| 5. | $e^{-at}\sin(kt)$ | $\frac{pk}{(p+a)^2+k^2}$ |
| 6. | $e^{-at}\cos(kt)$ | $\frac{p(p+a)}{(p+a)^2+k^2}$ |
| 7. | $t\sin(kt)$ | $\frac{2kp^2}{(p^2+k^2)^2}$ |
| 8. | $t\cos(kt)$ | $\frac{p\left(p^2-k^2\right)}{(p^2+k^2)^2}$ |
| 9. | $e^{-at}\frac{t^n}{n!}$ | $\frac{p}{(p+a)^{n+1}}$ |
| 10. | $\frac{(2t)^n}{1\cdot3\cdot5\ldots(2n-1)\sqrt{\pi t}}$ | $\frac{\sqrt{p}}{p^n}$ (n ganzzahlig >0) |
| 11. | $\frac{1}{\sqrt{\pi t}}e^{-\frac{a^2}{4t}}$ | $\sqrt{p}e^{-a\sqrt{p}}$ (a>0) |
| 12. | $\frac{a}{2\sqrt{\pi t^3}}e^{-\frac{a^2}{4t}}$ | $pe^{-a\sqrt{p}}$ |
| 13. | $J_0(t)$ (Besselfunktion) | $\frac{p}{\sqrt{1+p^2}}$ |
| 14. | $I_0(t)$ (Besselfunktion) | $\frac{p}{\sqrt{p^2-1}}$ |
| 15. | $\int_t^\infty \frac{e^{-x}}{x}dx$ | $\ln(1+p)$ |

Tabelle C.2.: Tabelle der Carson-Heaviside-Transformationen einiger ausgewählter Funktionen

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

# C.3. Tabelle der z-Transformationen

| | $f_n$ Originalfolge | $F(z) = Z(f-n)$ | Konvergenzbereich |
|---|---|---|---|
| 1. | $1$ | $\frac{z}{z-1}$ | $\|z\| > 1$ |
| 2. | $(-1)^n$ | $\frac{z}{z+1}$ | $\|z\| > 1$ |
| 3. | $n$ | $\frac{z}{(z-1)^2}$ | $\|z\| > 1$ |
| 4. | $n^2$ | $\frac{z(z+1)}{(z-1)^3}$ | $\|z\| > 1$ |
| 5. | $e^{an}$ | $\frac{z}{z-e^a}$ | $\|z\| > \|e^a\|$ |
| 6. | $a^n$ | $\frac{z}{z-a}$ | $\|z\| > \|a\|$ |
| 7. | $\frac{a^n}{n!}$ | $e^{\frac{a}{z}}$ | $\|z\| > 0$ |
| 8. | $na^n$ | $\frac{za}{(z-a)^2}$ | $\|z\| > \|a\|$ |
| 9. | $n^2a^n$ | $\frac{az(z+a)}{(z-a)^3}$ | $\|z\| > \|a\|$ |
| 10. | $\binom{n}{k}$ | $\frac{z}{(z-1)^{k+1}}$ | $\|z\| > 1$ |
| 11. | $\binom{k}{n}$ | $\left(1+\frac{1}{z}\right)^k$ | $\|z\| > 0$ |
| 12. | $\sin(bn)$ | $\frac{z\sin b}{z^2-2z\cos b+1}$ | $\|z\| > 1$ |

Tabelle C.3.: Tabelle der z-Transformationen einiger ausgewählter Funktionen

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

## C.4. Einstellzeiten und Zeitkonstanten

| $t/\tau$ | Fehler | Dekaden Genauigkeit | Bruchteil des Endwertes |
|---|---|---|---|
| 0,1 | 0,904837 | 0,04 | 0,095163 |
| 0,2 | 0,818731 | 0,087 | 0,181269 |
| 0,5 | 0,606531 | 0,22 | 0,393469 |
| 1 | 0,367879 | 0,43 | 0,632121 |
| 2 | 0,135335 | 0,87 | 0,864665 |
| 2,3 | 0,1 | 1 | 0,9 |
| 3 | 0,049787 | 1,3 | 0,950213 |
| 4 | 0,018316 | 1,7 | 0,981684 |
| 4,61 | 0,01 | 2 | 0,99 |
| 5 | 0,006738 | 2,2 | 0,993262 |
| 6 | 0,002479 | 2,6 | 0,997521 |
| 6,91 | 0,001 | 3 | 0,999 |
| 7 | 0,000912 | 3,0 | 0,999088 |
| 8 | 0,000335 | 3,5 | 0,999665 |
| 9 | 0,000123 | 3,9 | 0,999877 |
| 9,21 | 0,0001 | 4 | 0,9999 |
| 10 | 0,000045 | 4,3 | 0,999955 |
| 11,51 | $10^{-5}$ | 5 | 0,99999 |
| 13,82 | $10^{-6}$ | 6 | 0,999999 |
| 16,12 | $10^{-7}$ | 7 | 0,9999999 |
| 18,42 | $10^{-8}$ | 8 | 0,99999999 |
| 20,72 | $10^{-9}$ | 9 | 0,999999999 |
| 23,03 | $10^{-10}$ | 10 | 0,9999999999 |
| 25,33 | $10^{-11}$ | 11 | 0,99999999999 |
| 27,63 | $10^{-12}$ | 12 | 0,999999999999 |
| 29,93 | $10^{-13}$ | 13 | 0,9999999999999 |
| 32,24 | $10^{-14}$ | 14 | 0,99999999999999 |

Tabelle C.4.: Einstellzeiten zu einer vorgegeben Genauigkeit als Funktion der Zeitkonstante $\tau$

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

# D. Vergleich der Kenngrössen von Bauarten analoger Filter

| Ordnung | 2 | 4 | 6 | 8 | 10 |
|---|---|---|---|---|---|
| *Kritische Dämpfung* | | | | | |
| Normierte Anstiegszeit $t_a/T_g$ | 0,344 | 0,342 | 0,341 | 0,341 | 0,340 |
| Normierte Verzögerungszeit $t_v/T_g$ | 0,172 | 0,254 | 0,316 | 0,367 | 0,412 |
| Überschwingen % | 0 | 0 | 0 | 0 | 0 |
| *Besselfilter* | | | | | |
| Normierte Anstiegszeit $t_a/T_g$ | 0,344 | 0,352 | 0,350 | 0,347 | 0,345 |
| Normierte Verzögerungszeit $t_v/T_g$ | 0,195 | 0,329 | 0,428 | 0,505 | 0,574 |
| Überschwingen % | 0,43 | 0,84 | 0,64 | 0,34 | 0,06 |
| *Butterworthfilter* | | | | | |
| Normierte Anstiegszeit $t_a/T_g$ | 0,342 | 0,387 | 0,427 | 0,460 | 0,485 |
| Normierte Verzögerungszeit $t_v/T_g$ | 0,228 | 0,449 | 0,663 | 0,874 | 1,084 |
| Überschwingen % | 4,3 | 10,8 | 14,3 | 16,3 | 17,8 |
| *Tschebyschefffilter 0,5dB Welligkeit* | | | | | |
| Normierte Anstiegszeit $t_a/T_g$ | 0,338 | 0,421 | 0,487 | 0,540 | 0,584 |
| Normierte Verzögerungszeit $t_v/T_g$ | 0,251 | 0,556 | 0,875 | 1,196 | 1,518 |
| Überschwingen % | 10,7 | 18,1 | 21,2 | 22,9 | 24,1 |
| *Tschebyschefffilter 1dB Welligkeit* | | | | | |
| Normierte Anstiegszeit $t_a/T_g$ | 0,334 | 0,421 | 0,486 | 0,537 | 0,582 |
| Normierte Verzögerungszeit $t_v/T_g$ | 0,260 | 0,572 | 0,893 | 1,215 | 1,540 |
| Überschwingen % | 14,6 | 21,6 | 24,9 | 26,6 | 27,8 |
| *Tschebyschefffilter 2dB Welligkeit* | | | | | |
| Normierte Anstiegszeit $t_a/T_g$ | 0,326 | 0,414 | 0,491 | 0,529 | 0,570 |
| Normierte Verzögerungszeit $t_v/T_g$ | 0,267 | 0,584 | 0,912 | 1,231 | 1,555 |
| Überschwingen % | 21,1 | 28,9 | 32,0 | 33,5 | 34,7 |
| *Tschebyschefffilter 3dB Welligkeit* | | | | | |
| Normierte Anstiegszeit $t_a/T_g$ | 0,318 | 0,407 | 0,470 | 0,529 | 0,692 |
| Normierte Verzögerungszeit $t_v/T_g$ | 0,271 | 0,590 | 0,912 | 1,235 | 1,557 |
| Überschwingen % | 27,2 | 35,7 | 38,7 | 40,6 | 41,6 |

Tabelle D.1.: Kenngrössen von Tiefpassfiltern

# E. Diagramme der Filterübertragungsfunktionen

Die folgenden Darstellungen sind mit Maple[1] berechnet worden.

## E.1. Tiefpassfilter

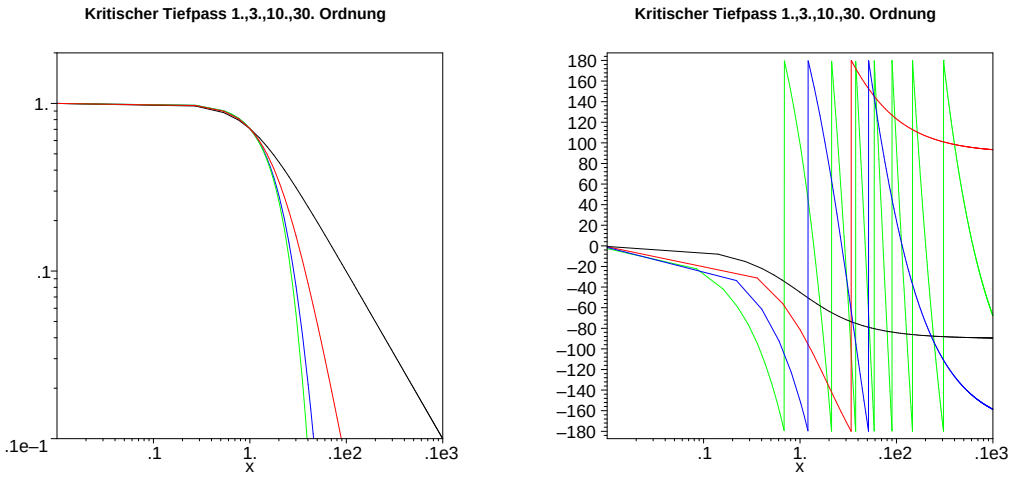



Abbildung E.1.: Amplituden- und Phasengang kritisch gedämpfter Tiefpassfilter. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung, blau 10. Ordnung und grün 30. Ordnung

[1] http://wwwex.physik.uni-ulm.de/lehre/PhysikalischeElektronik/Materialien/Gleichungen.mws

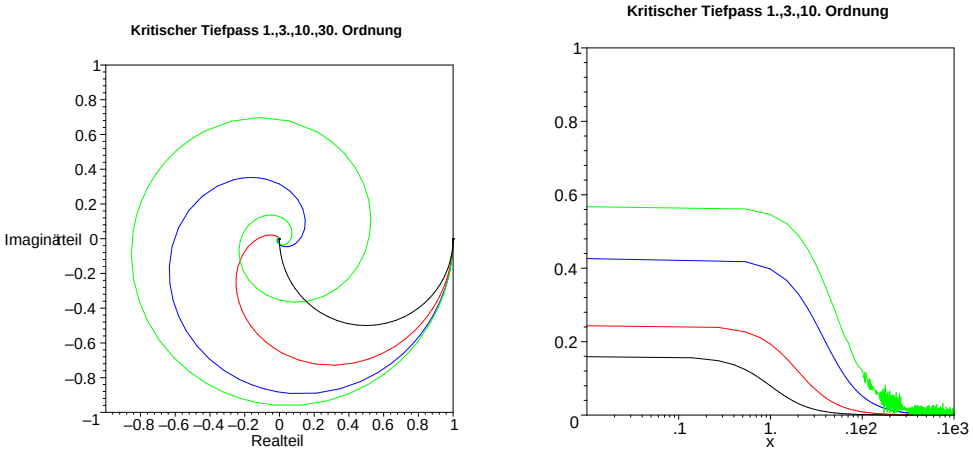



Abbildung E.2.: **Links:** Phasenbild von kritisch gedämpften Tiefpassfiltern. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung, blau 10. Ordnung und grün 30. Ordnung. **Rechts:** *Gruppenlaufzeit*en von kritisch gedämpften Tiefpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung, blau 10. Ordnung und grün 30. Ordnung. Das Rauschen auf der Grünen Kurve ist eine Rechen-Artefakt von Maple.

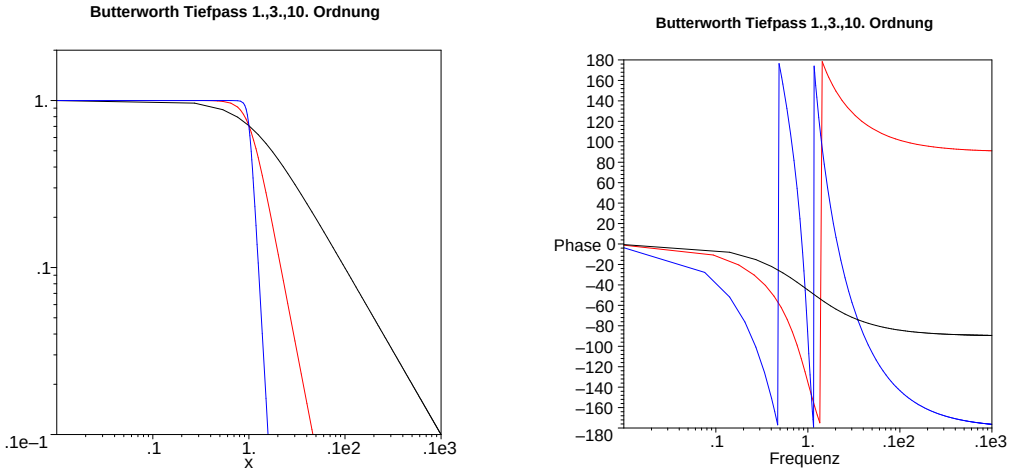



Abbildung E.3.: Amplituden- und Phasengang von Butterworth Tiefpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

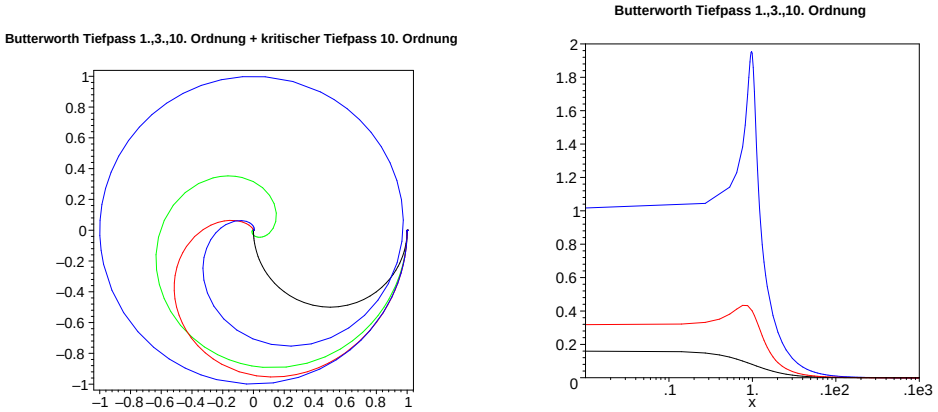



Abbildung E.4.: **Links:** Phasenbild von Butterworth Tiefpassfiltern. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Tiefpassfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en für Butterworth Tiefpassfilter. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

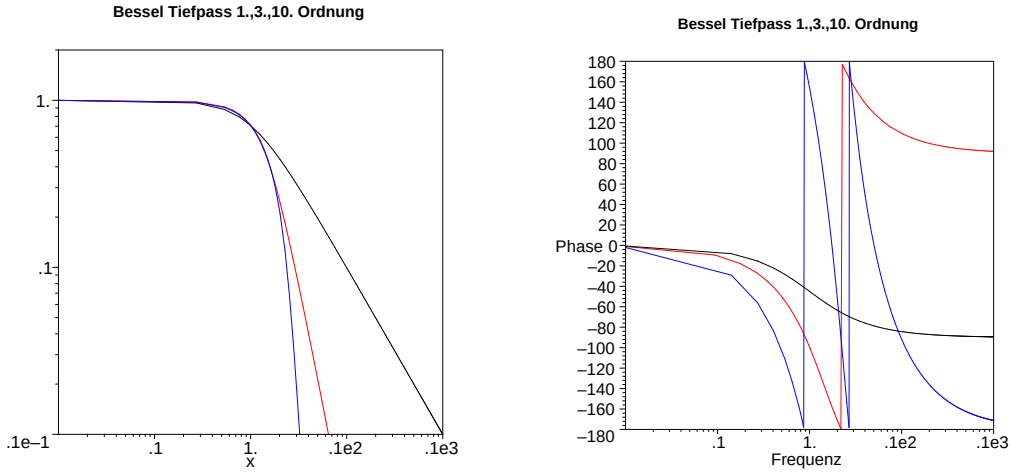



Abbildung E.5.: Amplituden- und Phasengang von Bessel-Tiefpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

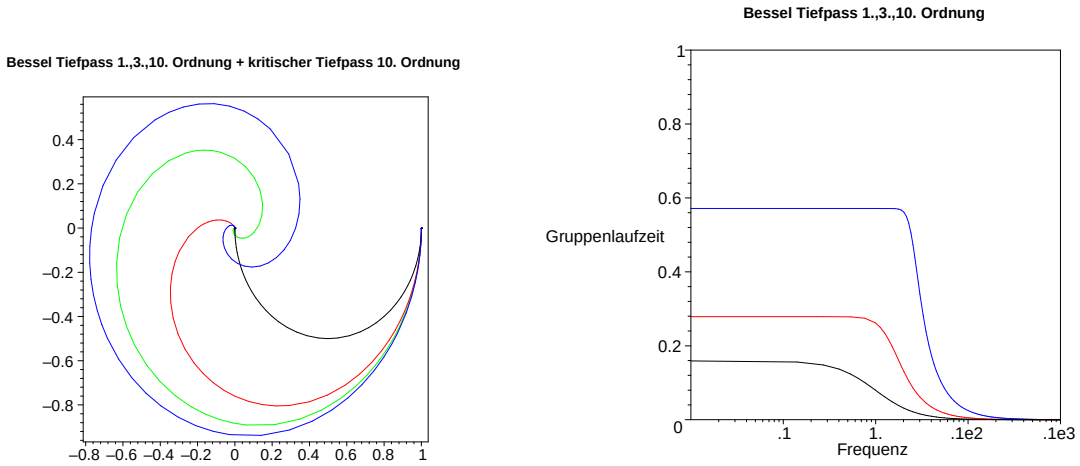



Abbildung E.6.: **Links:** Phasenbild von Bessel-Tiefpassfiltern. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Tiefpassfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en von Bessel-Tiefpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

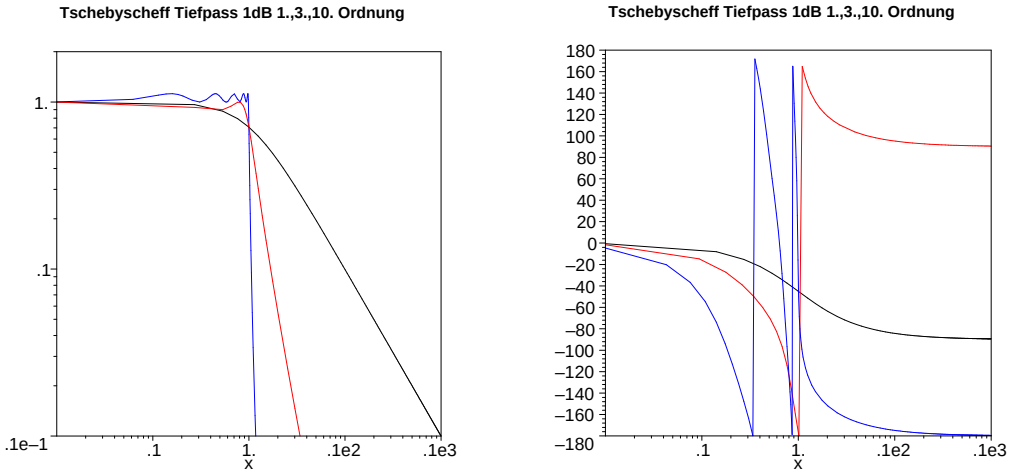



Abbildung E.7.: Amplituden- und Phasengang von Tschebyscheff-Tiefpassfiltern mit 1 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

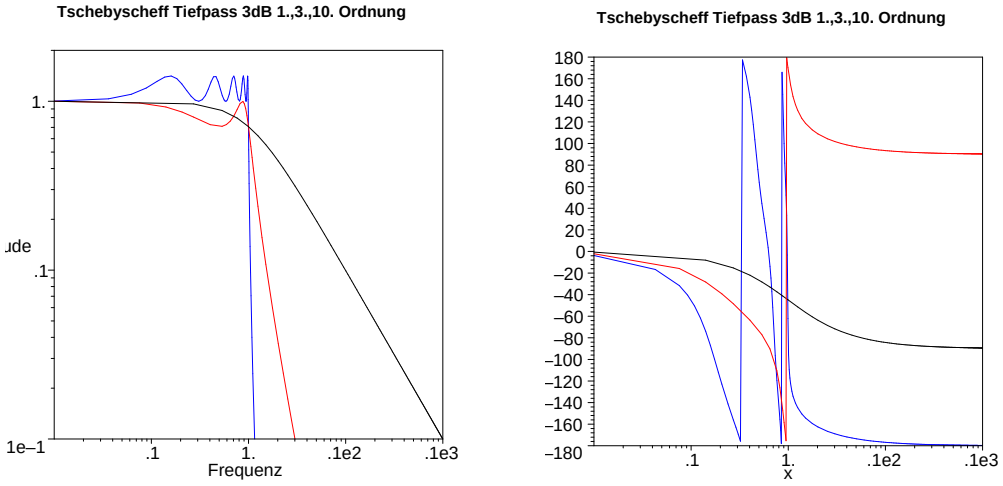



Abbildung E.8.: **Links:** Phasenbild von Tschebyscheff-Tiefpassfiltern mit 1dB Welligkeit. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Tiefpassfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en für Tschebyscheff-Tiefpassfilter mit 1 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

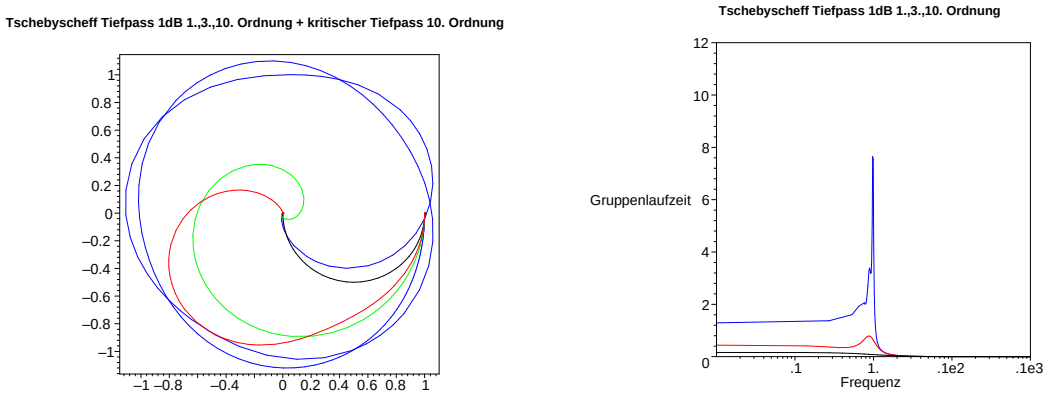



Abbildung E.9.: Amplituden- und Phasengang von Tschebyscheff-Tiefpassfiltern mit 3 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

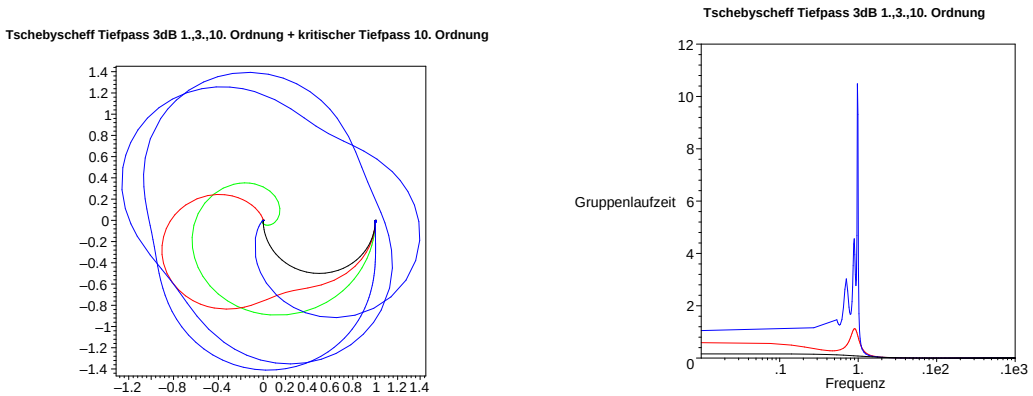



Abbildung E.10.: **Links:** Phasenbild für Tschebyscheff-Tiefpassfilter mit 3dB Welligkeit. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Tiefpassfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en für Tschebyscheff-Tiefpassfilter mit 3 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

# E.2. Hochpassfilter

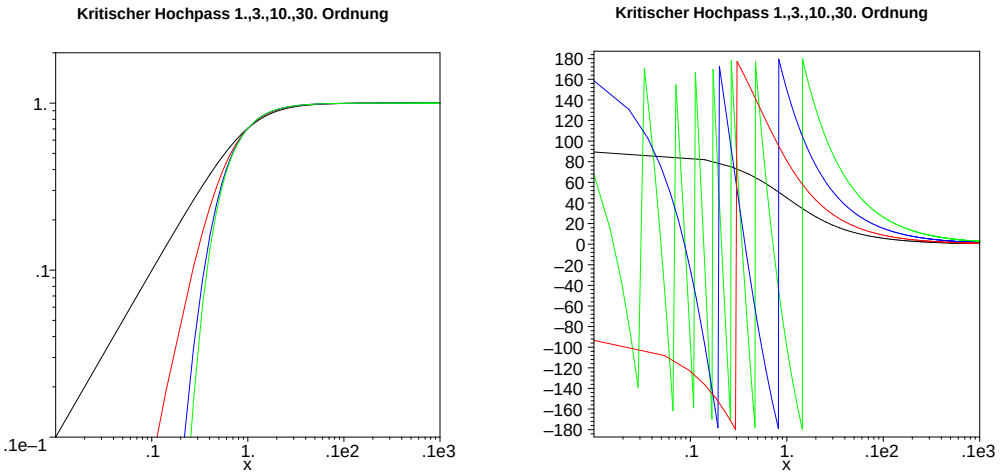



Abbildung E.11.: Amplituden- und Phasengang kritisch gedämpfter Hochpassfilter. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung, blau 10. Ordnung und grün 30. Ordnung

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

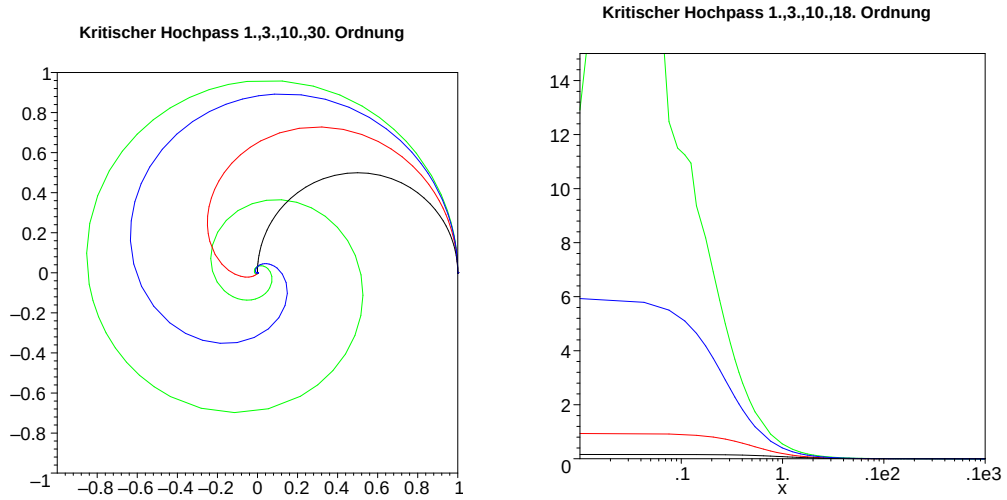



Abbildung E.12.: **Links:** Phasenbild für kritisch gedämpfte Hochpassfilter. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung, blau 10. Ordnung und grün 30. Ordnung. **Rechts:** *Gruppenlaufzeit*en von kritisch gedämpften Hochpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung, blau 10. Ordnung und grün 30. Ordnung. Der Verlauf der grünen Kurve links von 1 beruht auf einem Rechen-Artefakt von Maple.

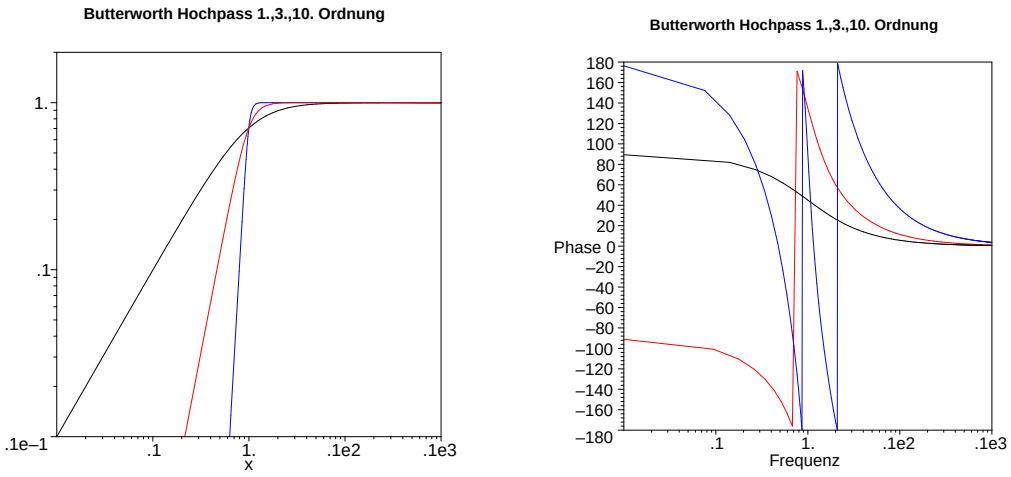



Abbildung E.13.: Amplituden- und Phasengang von Butterworth Hochpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

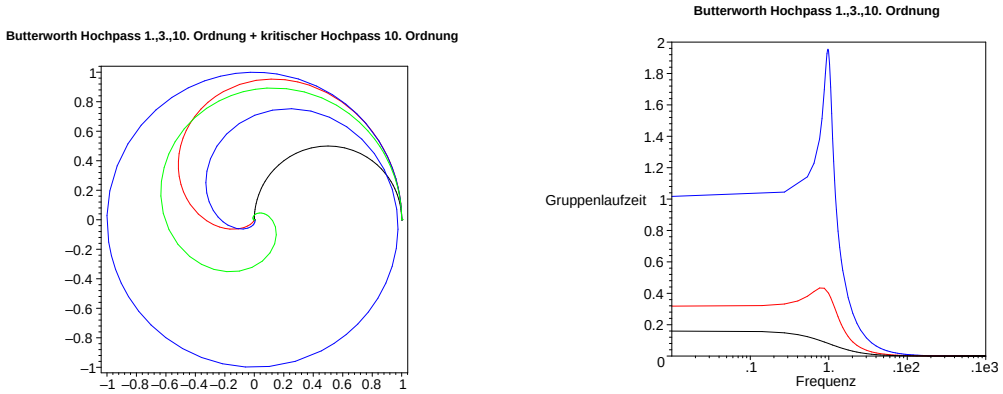



Abbildung E.14.: **Links:** Phasenbild für Butterworth Hochpassfilter. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Hochpassfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en für Butterworth Hochpassfilter. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

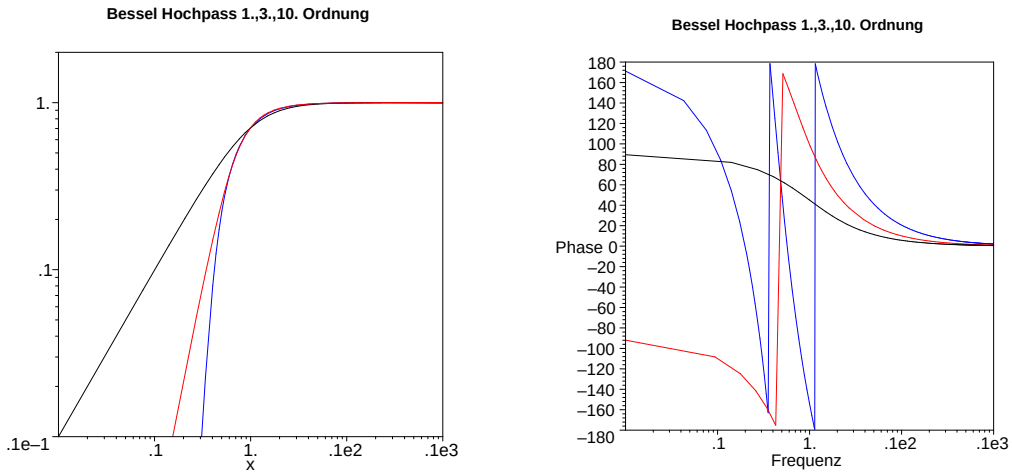



Abbildung E.15.: Amplituden- und Phasengang von Bessel-Hochpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

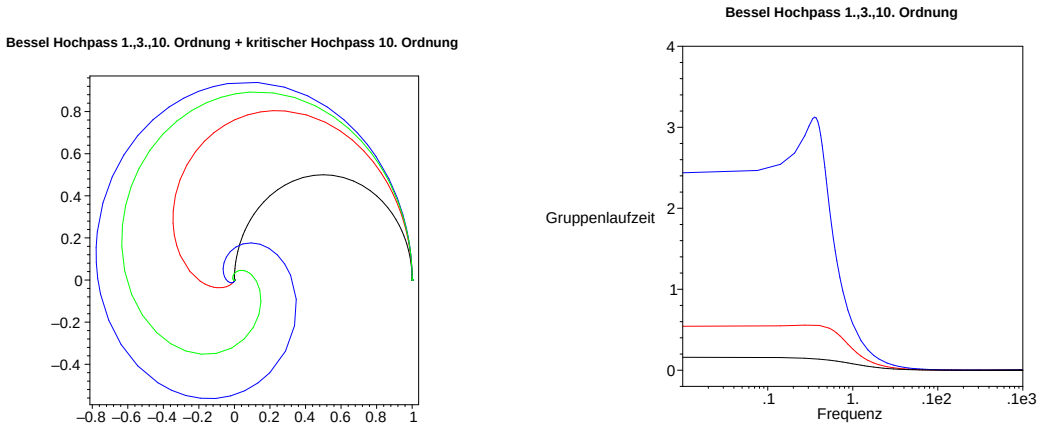



Abbildung E.16.: **Links:** Phasenbild für Bessel-Hochpassfilter. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung. Zum Vergleich ist grün ein kritisches Hochpassfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en von Bessel-Hochpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

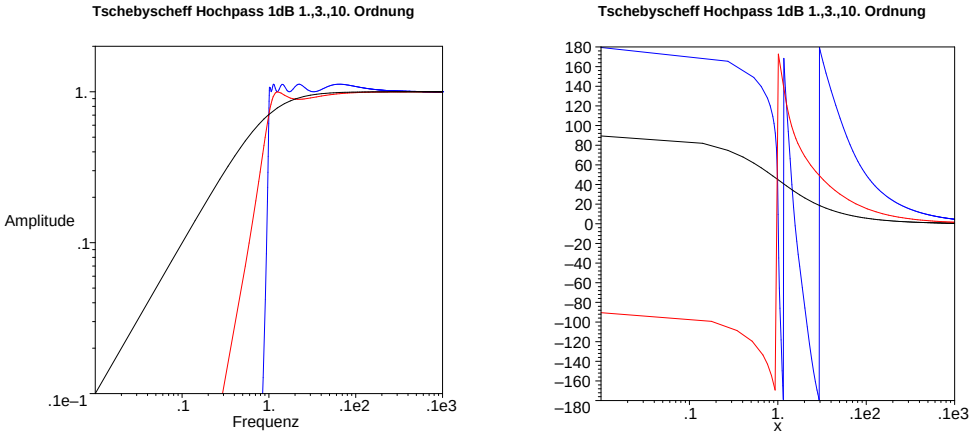



Abbildung E.17.: Amplituden- und Phasengang von Tschebyscheff-Hochpassfiltern mit 1 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

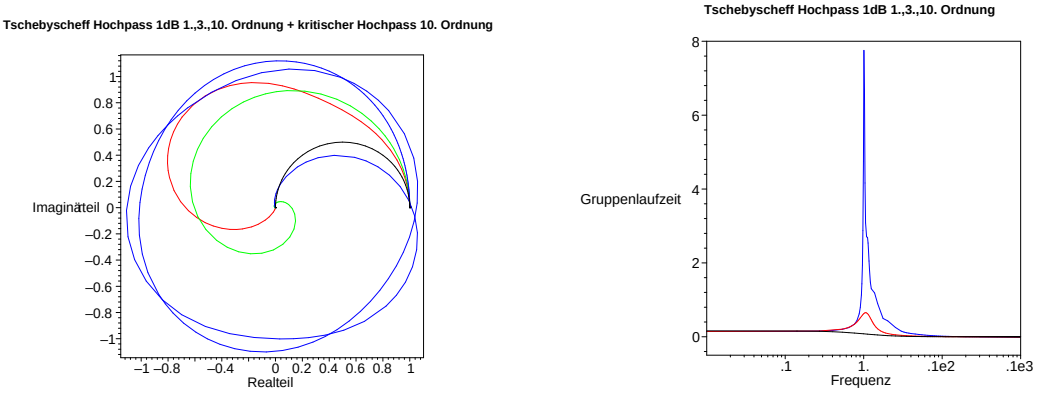



Abbildung E.18.: **Links:** Phasenbild von Tschebyscheff-Hochpassfiltern mit 1dB Welligkeit. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Hochpassfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en für Tschebyscheff-Hochpassfilter mit 1 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

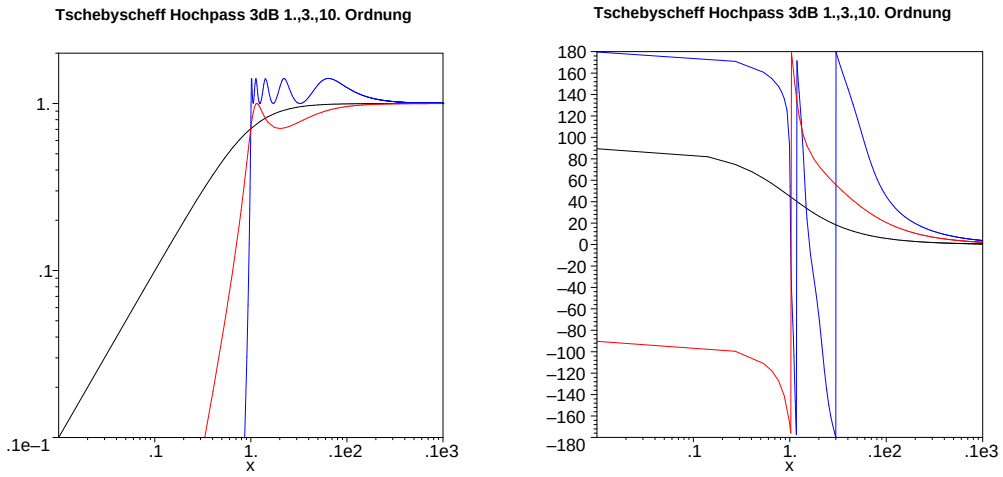



Abbildung E.19.: Amplituden- und Phasengang von Tschebyscheff-Hochpassfiltern mit 3 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

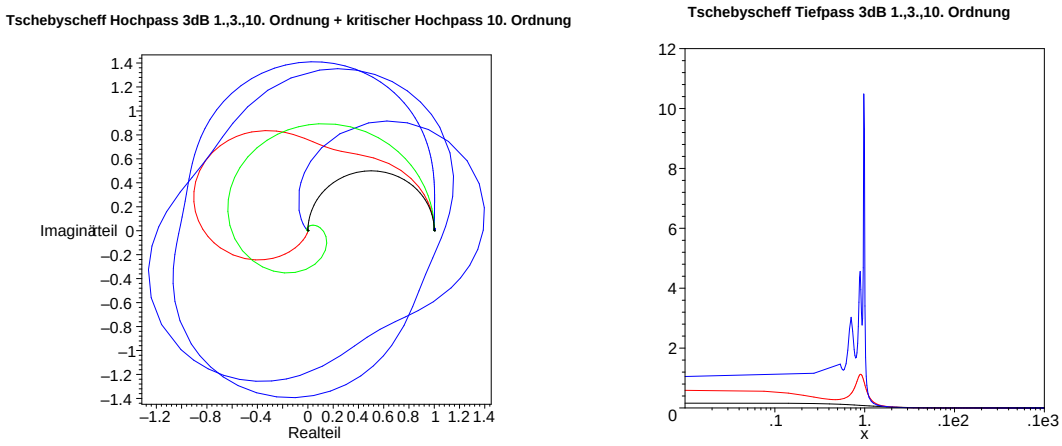



Abbildung E.20.: **Links:** Phasenbild von Tschebyscheff-Hochpassfiltern mit 3dB Welligkeit. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Hochpassfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en für Tschebyscheff-Hochpassfilter mit 3 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

# E.3. Bandpassfilter

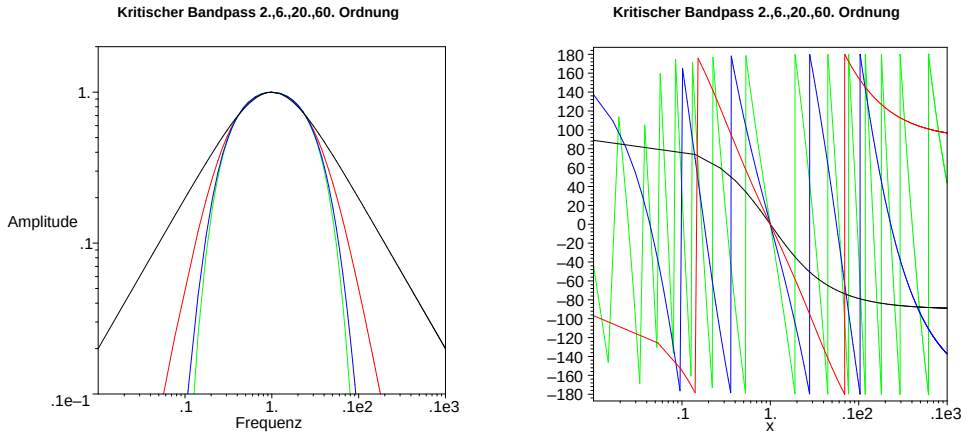



Abbildung E.21.: Amplituden- und Phasengang von kritisch gedämpften Bandpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung, blau 10. Ordnung und grün 30. Ordnung

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

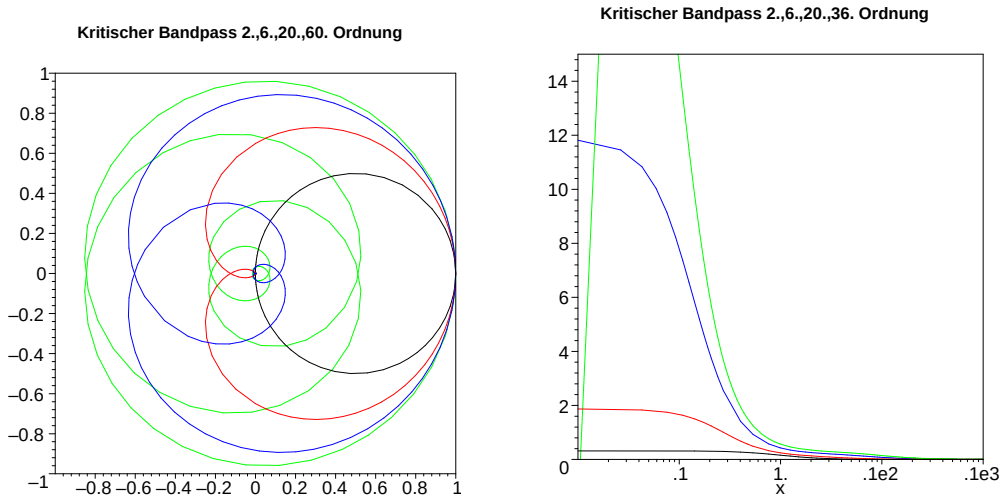



Abbildung E.22.: **Links:** Phasenbild von kritisch gedämpften Bandpassfiltern. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung, blau 10. Ordnung und grün 30. Ordnung. **Rechts:** *Gruppenlaufzeit*en für kritisch gedämpfte Bandpassfilter. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung, blau 10. Ordnung und grün 30. Ordnung. Der Verlauf der grünen Kurve links von 1 beruht auf einem Rechen-Artefakt von Maple.

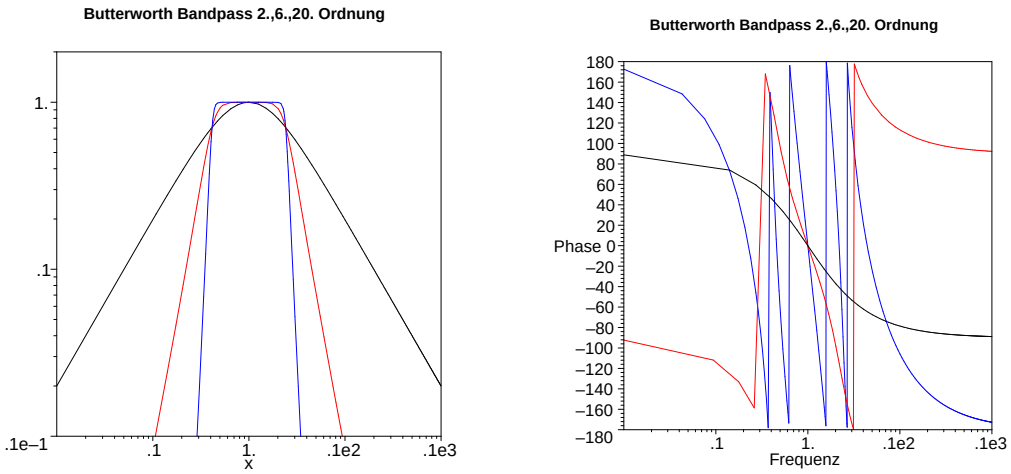



Abbildung E.23.: Amplituden- und Phasengang von Butterworth Bandpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

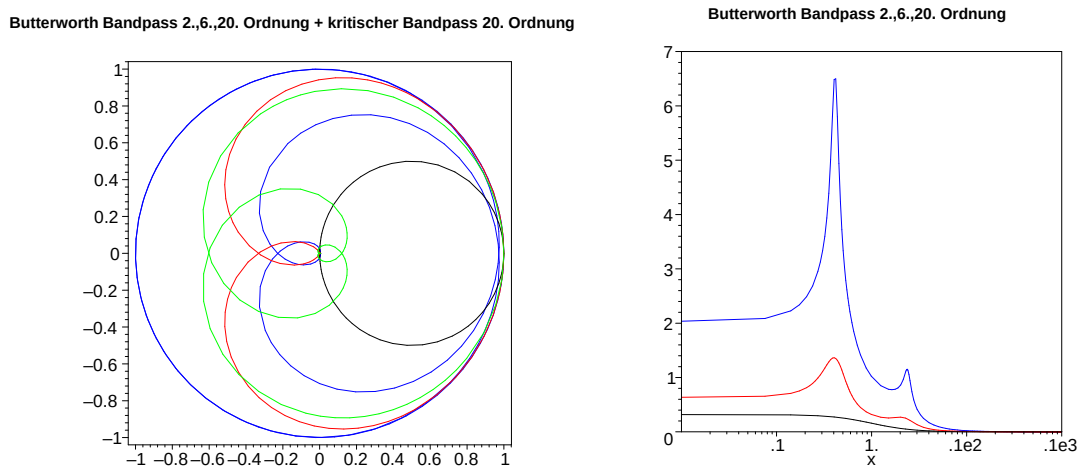



Abbildung E.24.: **Links:** Phasenbild von Butterworth Bandpassfiltern. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Bandpassfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en von Butterworth Bandpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

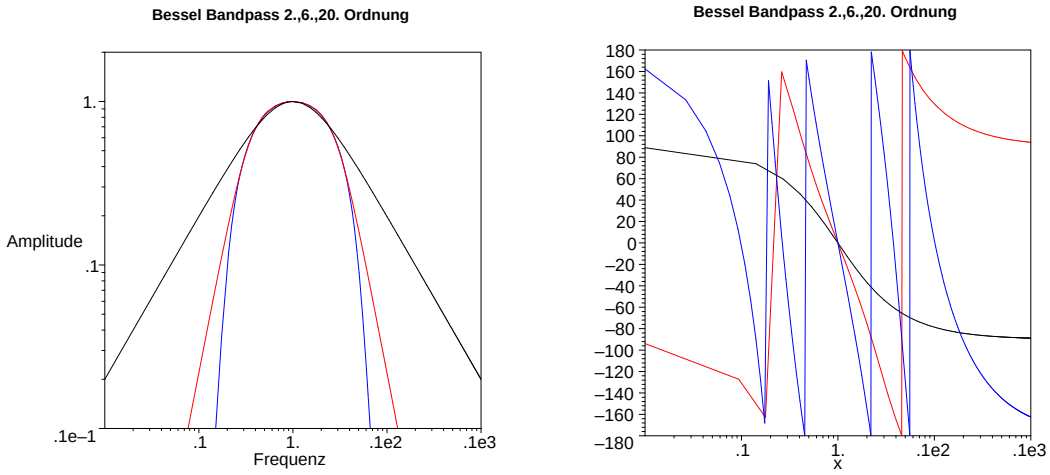



Abbildung E.25.: Amplituden- und Phasengang von Bessel-Bandpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

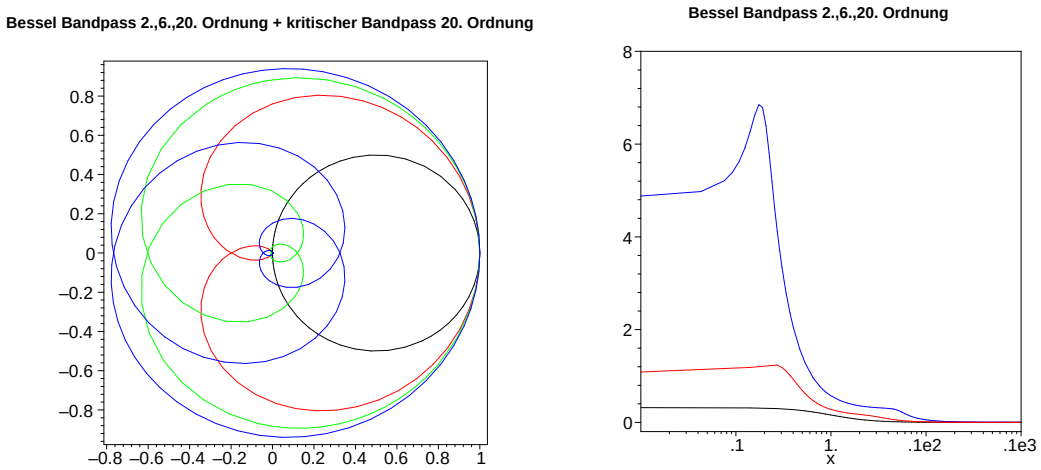



Abbildung E.26.: **Links:** Phasenbild von Bessel-Bandpassfiltern. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Bandpassfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en von Bessel-Bandpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

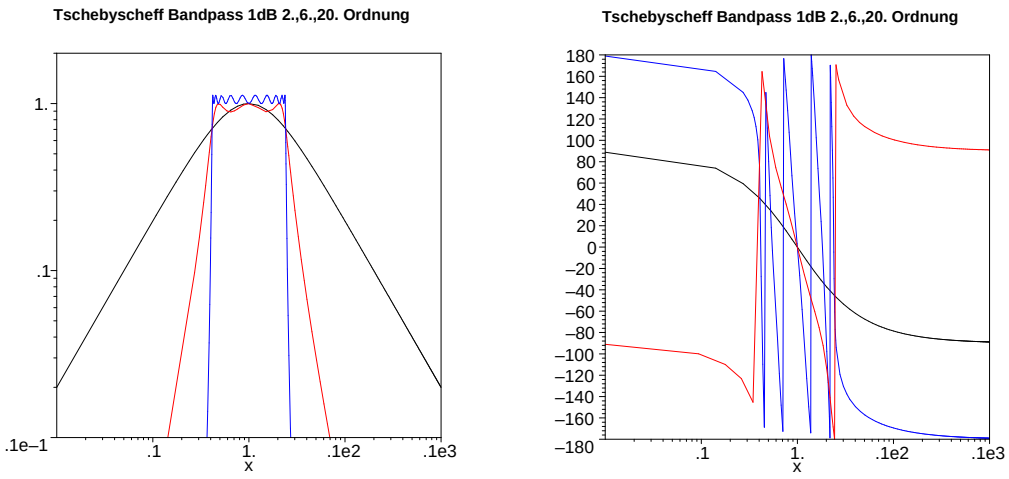



Abbildung E.27.: Amplituden- und Phasengang von Tschebyscheff-Bandpassfiltern mit 1 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

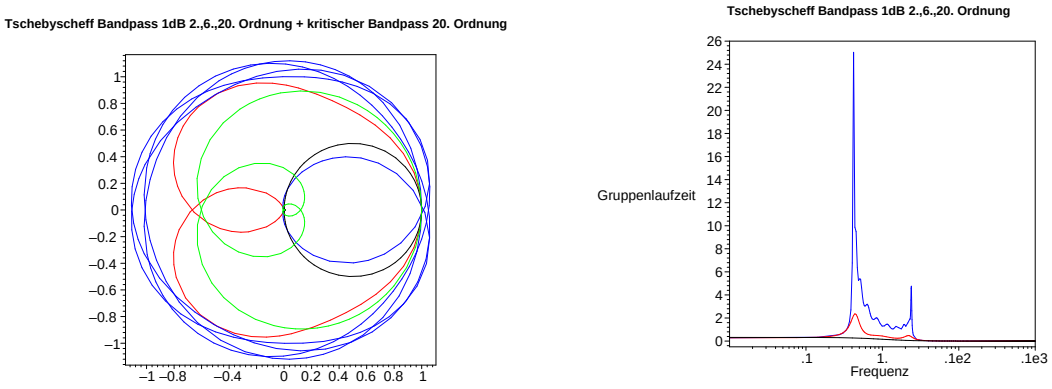



Abbildung E.28.: **Links:** Phasenbild von Tschebyscheff-Bandpassfiltern mit 1dB Welligkeit. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Bandpassfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en für Tschebyscheff-Bandpassfilter mit 1 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

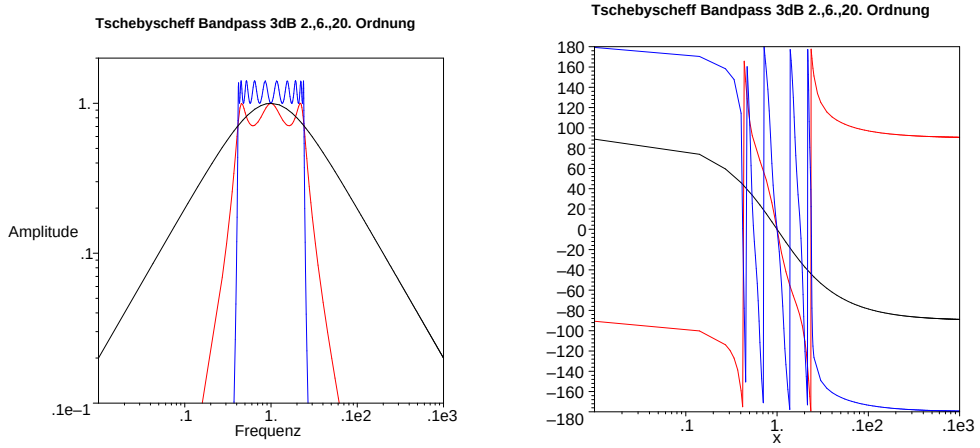



Abbildung E.29.: Amplituden- und Phasengang von Tschebyscheff-Bandpassfiltern mit 3 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

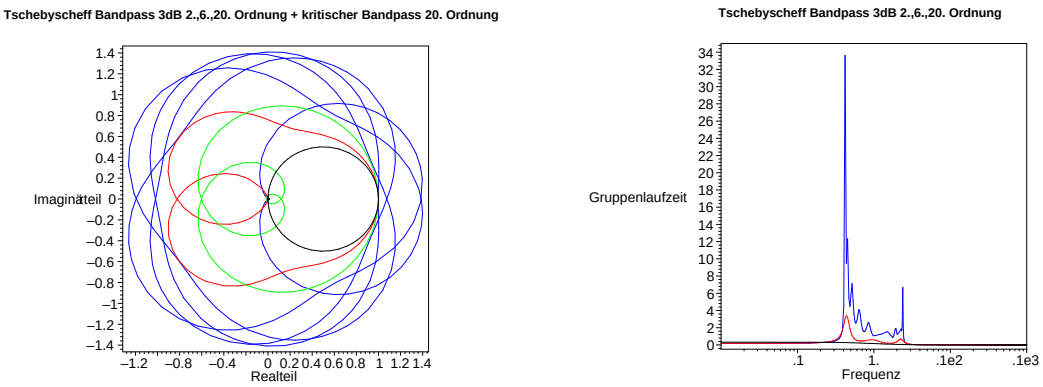



Abbildung E.30.: **Links:** Phasenbild von Tschebyscheff-Bandpassfiltern mit 3dB Welligkeit. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Bandpassfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en für Tschebyscheff-Bandpassfilter mit 3 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

# E.4. Bandsperrenfilter

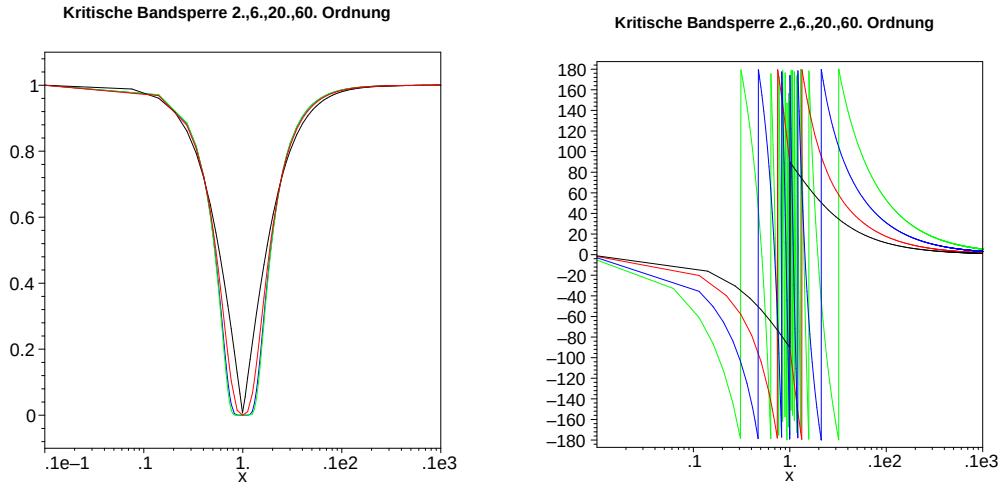



Abbildung E.31.: Amplituden- und Phasengang von kritisch gedämpften Bandsperrenfilten. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung, blau 10. Ordnung und grün 30. Ordnung

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

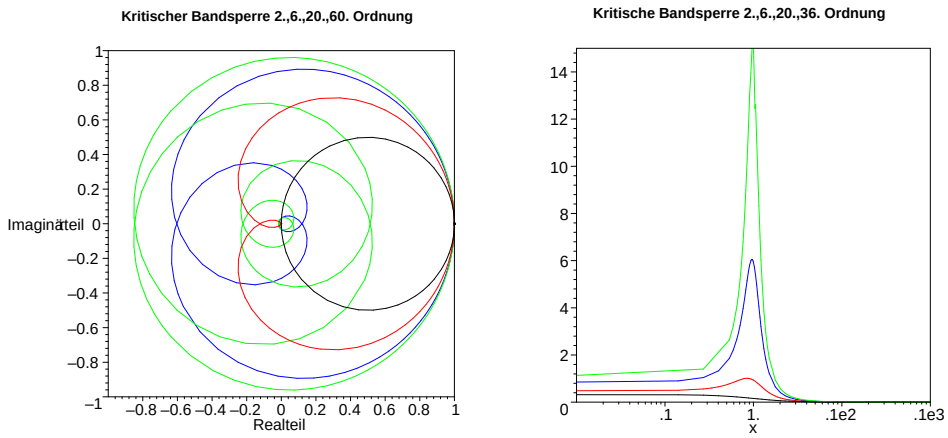



Abbildung E.32.: **Links:** Phasenbild von kritisch gedämpften Bandsperrenfiltern. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung, blau 10. Ordnung und grün 30. Ordnung. **Rechts:** *Gruppenlaufzeit*en für kritisch gedämpfte Bandsperrenfilter. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung, blau 10. Ordnung und grün 30. Ordnung. Der Verlauf der grünen Kurve links von 1 beruht auf einem Rechen-Artefakt von Maple.

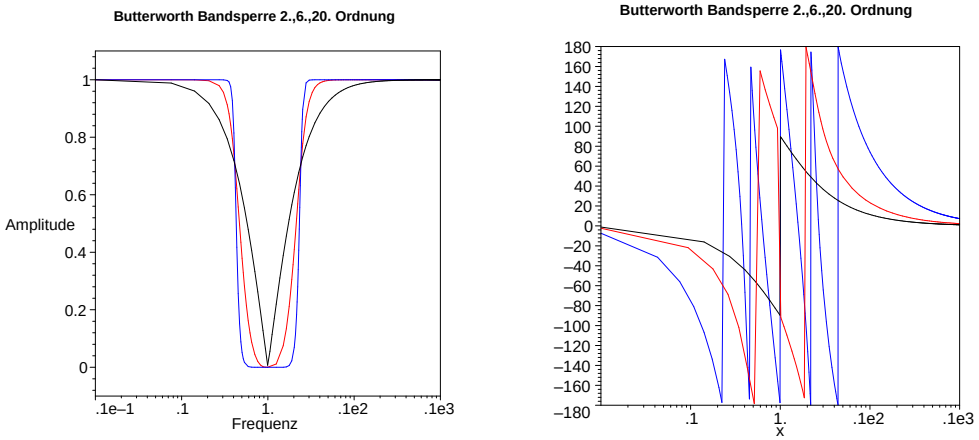



Abbildung E.33.: Amplituden- und Phasengang von Butterworth Bandsperrenfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

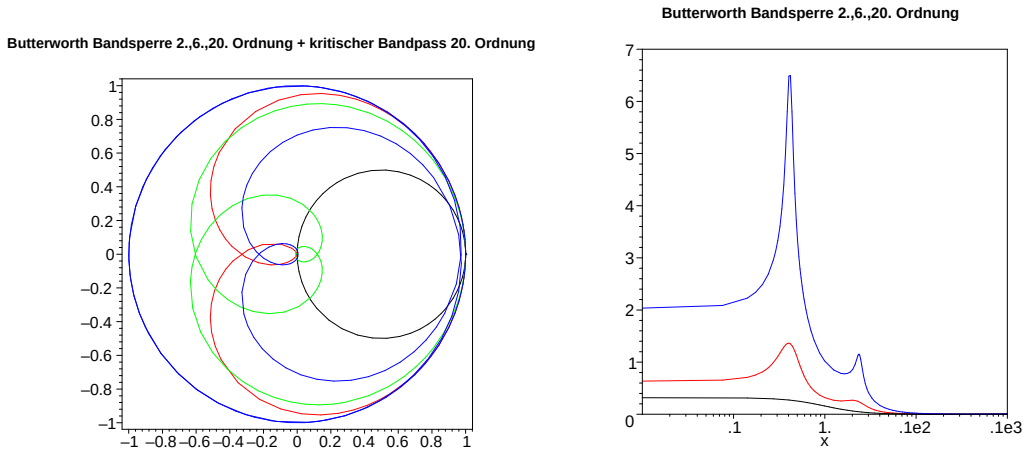



Abbildung E.34.: **Links:** Phasenbild von Butterworth Bandsperrenfiltern. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Bandsperrenfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en von Butterworth Bandsperrenfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

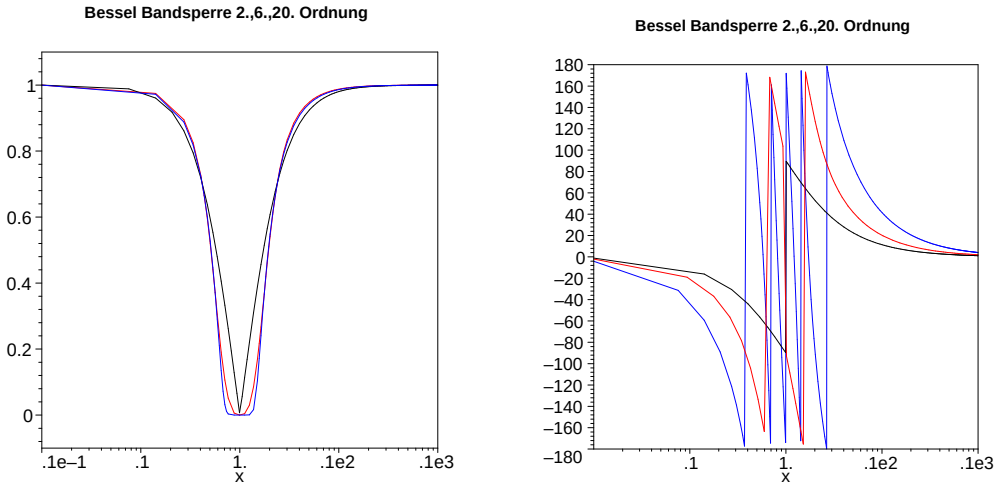



Abbildung E.35.: Amplituden- und Phasengang von Bessel-Bandsperrenfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

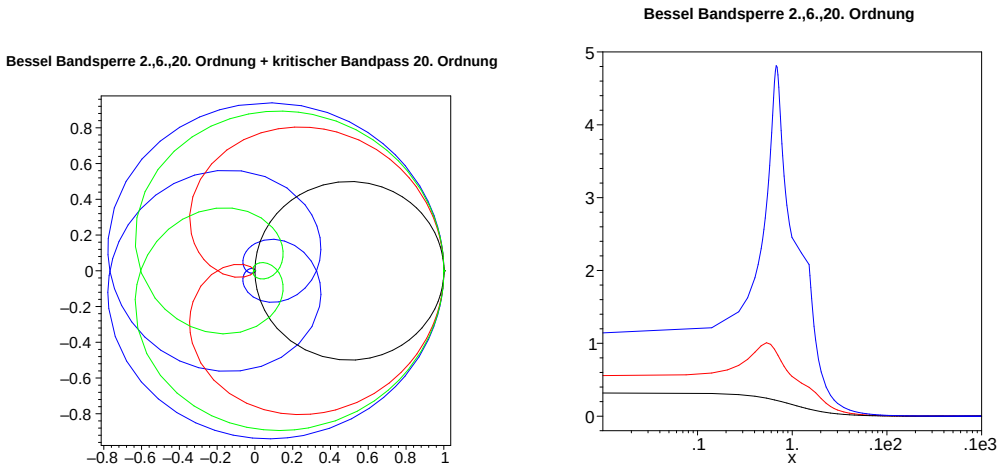



Abbildung E.36.: **Links:** Phasenbild von Bessel-Bandsperrenfiltern. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Bandsperrenfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en von Bessel-Bandsperrenfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

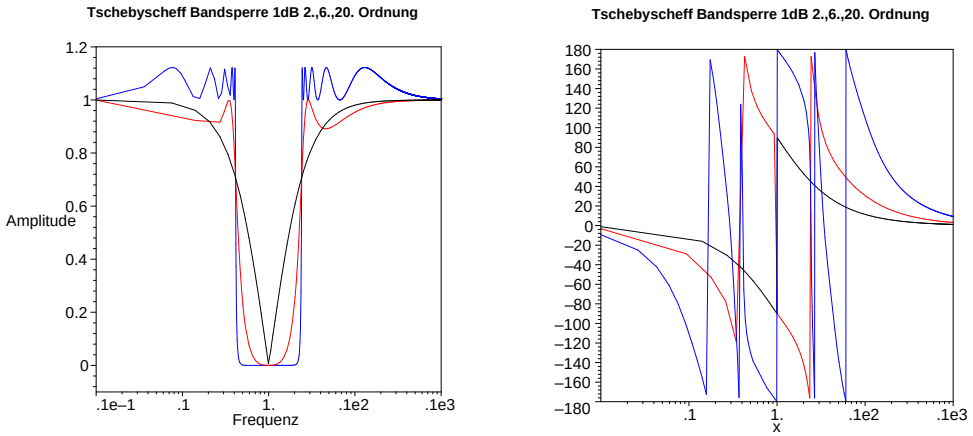



Abbildung E.37.: Amplituden- und Phasengang von Tschebyscheff-Bandsperrenfiltern mit 1 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

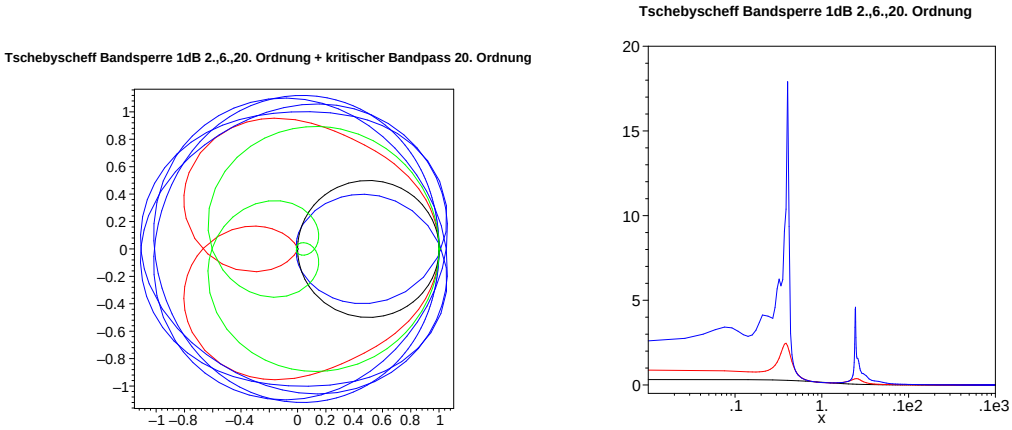



Abbildung E.38.: **Links:** Phasenbild von Tschebyscheff-Bandsperrenfiltern mit 1dB Welligkeit. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Bandsperrenfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en von Tschebyscheff-Bandsperrenfiltern mit 1 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

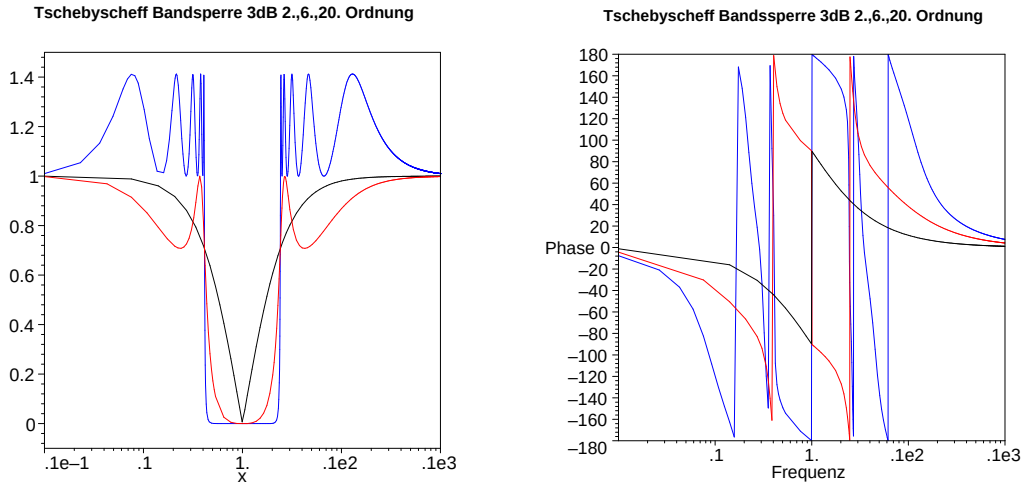



Abbildung E.39.: Amplituden- und Phasengang von Tschebyscheff-Bandsperrenfiltern mit 3 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung

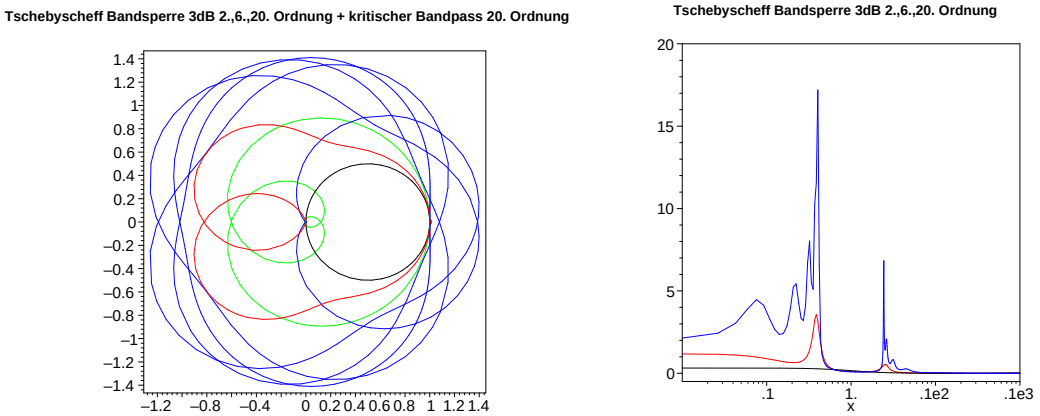



Abbildung E.40.: **Links:** Phasenbild von Tschebyscheff-Bandsperrenfiltern mit 3dB Welligkeit. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung Zum Vergleich ist grün ein kritisches Bandsperrenfilter 10. Ordnung aufgetragen. **Rechts:** *Gruppenlaufzeit*en von Tschebyscheff-Bandsperrenfiltern mit 3 dB Welligkeit. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

# E.5. Allpassfilter

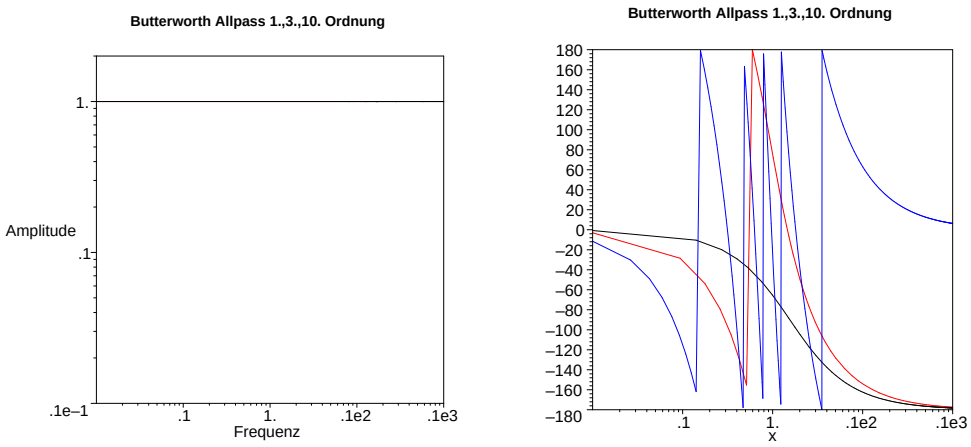



Abbildung E.41.: Amplituden- und Phasengang von maximal flachen Allpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

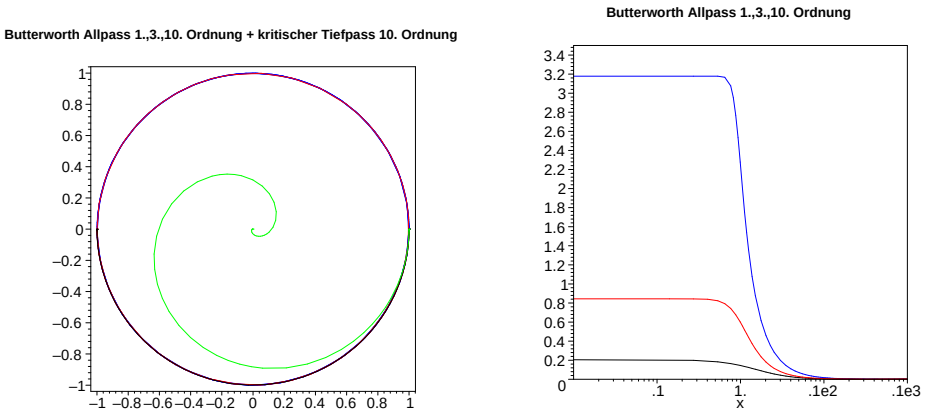



Abbildung E.42.: **Links:** Phasenbild von Allpassfiltern. Schwarz ist das Phasenbild eines Filters erster Ordnung, rot dritter Ordnung, blau 10. Ordnung. Grün ist zum Vergleich ein kritischer Tiefpass 10. Ordnung. **Rechts:** *Gruppenlaufzeit*en von Allpassfiltern. Schwarz ist der Frequenzgang eines Filters erster Ordnung, rot dritter Ordnung und blau 10. Ordnung.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

# E.6. Schwingkreis

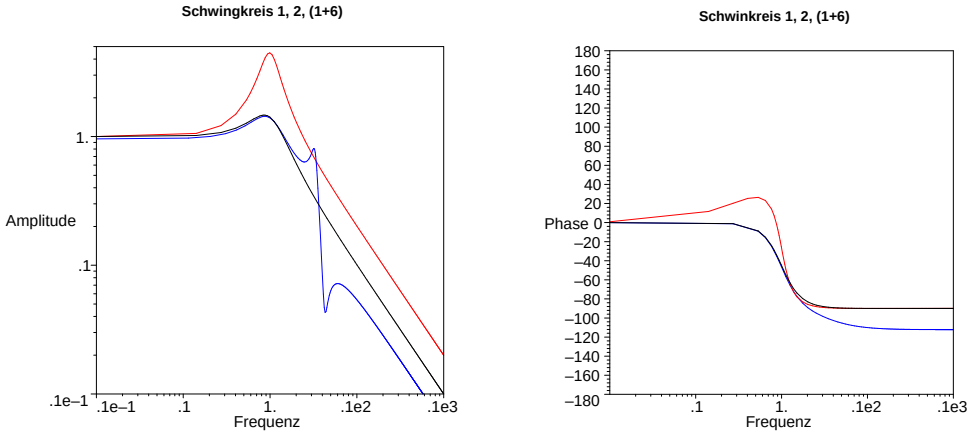



Abbildung E.43.: Amplituden- und Phasengang von Schwingkreisen. Schwarz ist der Frequenzgang eines Schwingkreises mit $Q = 1$, rot mit $Q = 2$ sowie Blau eine Kombination eines Schwingkreises mit $Q = 1$ und eines Schwingkreises mit $Q = 6$.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

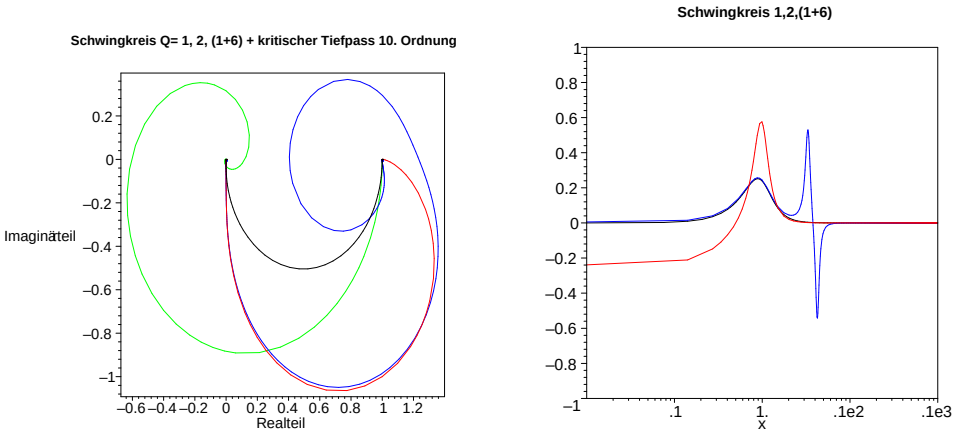



Abbildung E.44.: **Links:** Phasenbild eines Schwingkreises. Schwarz ist der Frequenzgang eines Schwingkreises mit $Q = 1$, rot mit $Q = 2$ sowie Blau eine Kombination eines Schwingkreises mit $Q = 1$ und eines Schwingkreises mit $Q = 6$ **Rechts:** die entsprechenden *Gruppenlaufzeit*en

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

# F. Filterkoeffizienten

| Ordnung | Filterstufe | $a_i$ | $b_i$ | $\frac{f_{g_i}}{f_g}$ | $Q_i$ |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 | - |
| 2 | 1 | 1,2872 | 0,4142 | 1 | 0,5 |
| 3 | 1 | 0,5098 | 0 | 1,961 | - |
| | 2 | 1,0197 | 0,2599 | 1,262 | 0,5 |
| 4 | 1 | 0,87 | 0,1892 | 1,48 | 0,5 |
| | 2 | 0,87 | 0,1892 | 1,48 | 0,5 |
| 5 | 1 | 0,3856 | 0 | 2,593 | - |
| | 2 | 0,7712 | 0,1487 | 1,669 | 0,5 |
| | 3 | 0,7712 | 0,1487 | 1,669 | 0,5 |
| 6 | 1 | 0,6999 | 0,1225 | 1,839 | 0,5 |
| | 2 | 0,6999 | 0,1225 | 1,839 | 0,5 |
| | 3 | 0,6999 | 0,1225 | 1,839 | 0,5 |
| 7 | 1 | 0,3226 | 0 | 3,1 | - |
| | 2 | 0,6453 | 0,1041 | 1,995 | 0,5 |
| | 3 | 0,6453 | 0,1041 | 1,995 | 0,5 |
| | 4 | 0,6453 | 0,1041 | 1,995 | 0,5 |
| 8 | 1 | 0,6017 | 0,0905 | 2,139 | 0,5 |
| | 2 | 0,6017 | 0,0905 | 2,139 | 0,5 |
| | 3 | 0,6017 | 0,0905 | 2,139 | 0,5 |
| | 4 | 0,6017 | 0,0905 | 2,139 | 0,5 |
| 9 | 1 | 0,2829 | 0 | 3,534 | - |
| | 2 | 0,5659 | 0,0801 | 2,275 | 0,5 |
| | 3 | 0,5659 | 0,0801 | 2,275 | 0,5 |
| | 4 | 0,5659 | 0,0801 | 2,275 | 0,5 |
| | 5 | 0,5659 | 0,0801 | 2,275 | 0,5 |
| 10 | 1 | 0,5358 | 0,0718 | 2,402 | 0,5 |
| | 2 | 0,5358 | 0,0718 | 2,402 | 0,5 |
| | 3 | 0,5358 | 0,0718 | 2,402 | 0,5 |
| | 4 | 0,5358 | 0,0718 | 2,402 | 0,5 |
| | 5 | 0,5358 | 0,0718 | 2,402 | 0,5 |

Tabelle F.1.: Filterkoeffizienten für kritisch gedämpfte Filter

| Ordnung | Filterstufe | $a_i$ | $b_i$ | $\frac{f_{g_i}}{f_g}$ | $Q_i$ |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 | - |
| 2 | 1 | 1,3617 | 0,618 | 1 | 0,58 |
| 3 | 1 | 0,756 | 0 | 1,323 | - |
| | 2 | 0,9996 | 0,4772 | 1,414 | 0,69 |
| 4 | 1 | 1,3397 | 0,4889 | 0,978 | 0,52 |
| | 2 | 0,7743 | 0,389 | 1,797 | 0,81 |
| 5 | 1 | 0,6656 | 0 | 1,502 | - |
| | 2 | 1,1402 | 0,4128 | 1,184 | 0,56 |
| | 3 | 0,6216 | 0,3245 | 2,138 | 0,92 |
| 6 | 1 | 1,2217 | 0,3887 | 1,063 | 0,51 |
| | 2 | 0,9686 | 0,3505 | 1,431 | 0,61 |
| | 3 | 0,5131 | 0,2756 | 2,447 | 1,02 |
| 7 | 1 | 0,5937 | 0 | 1,684 | - |
| | 2 | 1,0944 | 0,3395 | 1,207 | 0,53 |
| | 3 | 0,8304 | 0,3011 | 1,695 | 0,66 |
| | 4 | 0,4332 | 0,2381 | 2,731 | 1,13 |
| 8 | 1 | 1,1112 | 0,3162 | 1,164 | 0,51 |
| | 2 | 0,9754 | 0,2979 | 1,381 | 0,56 |
| | 3 | 0,7202 | 0,2621 | 1,963 | 0,71 |
| | 4 | 0,3728 | 0,2087 | 2,992 | 1,23 |
| 9 | 1 | 0,5386 | 0 | 1,857 | - |
| | 2 | 1,0244 | 0,2834 | 1,277 | 0,52 |
| | 3 | 0,871 | 0,2636 | 1,574 | 0,59 |
| | 4 | 0,632 | 0,2311 | 2,226 | 0,76 |
| | 5 | 0,3257 | 0,1854 | 3,237 | 1,32 |
| 10 | 1 | 1,0215 | 0,265 | 1,264 | 0,5 |
| | 2 | 0,9393 | 0,2549 | 1,412 | 0,54 |
| | 3 | 0,7815 | 0,2351 | 1,78 | 0,62 |
| | 4 | 0,5604 | 0,2059 | 2,479 | 0,81 |
| | 5 | 0,2883 | 0,1665 | 3,466 | 1,42 |

Tabelle F.2.: Filterkoeffizienten für Besselfilter

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

| Ordnung | Filterstufe | $a_i$ | $b_i$ | $\frac{f_{g_i}}{f_g}$ | $Q_i$ |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 | - |
| 2 | 1 | 1,4142 | 1 | 1 | 0,71 |
| 3 | 1 | 1 | 0 | 1 | - |
| | 2 | 1 | 1 | 1,272 | 1 |
| 4 | 1 | 1,8478 | 1 | 0,719 | 0,54 |
| | 2 | 0,7654 | 1 | 1,39 | 1,31 |
| 5 | 1 | 1 | 0 | 1 | - |
| | 2 | 1,618 | 1 | 0,859 | 0,62 |
| | 3 | 0,618 | 1 | 1,448 | 1,62 |
| 6 | 1 | 1,9319 | 1 | 0,676 | 0,52 |
| | 2 | 1,4142 | 1 | 1 | 0,71 |
| | 3 | 0,5176 | 1 | 1,479 | 1,93 |
| 7 | 1 | 1 | 0 | 1 | - |
| | 2 | 1,8019 | 1 | 0,745 | 0,55 |
| | 3 | 1,247 | 1 | 1,117 | 0,8 |
| | 4 | 0,445 | 1 | 1,499 | 2,25 |
| 8 | 1 | 1,9616 | 1 | 0,661 | 0,51 |
| | 2 | 1,6629 | 1 | 0,829 | 0,6 |
| | 3 | 1,1111 | 1 | 1,206 | 0,9 |
| | 4 | 0,3902 | 1 | 1,512 | 2,56 |
| 9 | 1 | 1 | 0 | 1 | - |
| | 2 | 1,8794 | 1 | 0,703 | 0,53 |
| | 3 | 1,5321 | 1 | 0,917 | 0,65 |
| | 4 | 1 | 1 | 1,272 | 1 |
| | 5 | 0,3473 | 1 | 1,521 | 2,88 |
| 10 | 1 | 1,9754 | 1 | 0,655 | 0,51 |
| | 2 | 1,782 | 1 | 0,756 | 0,56 |
| | 3 | 1,4142 | 1 | 1 | 0,71 |
| | 4 | 0,908 | 1 | 1,322 | 1,1 |
| | 5 | 0,3129 | 1 | 1,527 | 3,2 |

Tabelle F.3.: Filterkoeffizienten für Butterworth-Filter

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

| Ordnung | Filterstufe | $a_i$ | $b_i$ | $\frac{f_{g_i}}{f_g}$ | $Q_i$ |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 | - |
| 2 | 1 | 1,3614 | 1,3827 | 1 | 0,86 |
| 3 | 1 | 1,8636 | 0 | 0,537 | - |
| | 2 | 0,6402 | 1,1931 | 1,335 | 1,71 |
| 4 | 1 | 2,6282 | 3,4341 | 0,538 | 0,71 |
| | 2 | 0,3648 | 1,1509 | 1,419 | 2,94 |
| 5 | 1 | 2,9235 | 0 | 0,342 | - |
| | 2 | 1,3025 | 2,3534 | 0,881 | 1,18 |
| | 3 | 0,229 | 1,0833 | 1,48 | 4,54 |
| 6 | 1 | 3,8645 | 6,9797 | 0,366 | 0,68 |
| | 2 | 0,7528 | 1,8573 | 1,078 | 1,81 |
| | 3 | 0,1589 | 1,0711 | 1,495 | 6,51 |
| 7 | 1 | 4,0211 | 0 | 0,249 | - |
| | 2 | 1,8729 | 4,1795 | 0,645 | 1,09 |
| | 3 | 0,4861 | 1,5676 | 1,208 | 2,58 |
| | 4 | 0,1156 | 1,0443 | 1,517 | 8,84 |
| 8 | 1 | 5,1117 | 11,9607 | 0,276 | 0,68 |
| | 2 | 1,0639 | 2,9365 | 0,844 | 1,61 |
| | 3 | 0,3439 | 1,4206 | 1,284 | 3,47 |
| | 4 | 0,0885 | 1,0407 | 1,521 | 11,53 |
| 9 | 1 | 5,1318 | 0 | 0,195 | - |
| | 2 | 2,4283 | 6,6307 | 0,506 | 1,06 |
| | 3 | 0,6839 | 2,2908 | 0,989 | 2,21 |
| | 4 | 0,2559 | 1,3133 | 1,344 | 4,48 |
| | 5 | 0,0695 | 1,0272 | 1,532 | 14,58 |
| 10 | 1 | 6,3648 | 18,3695 | 0,222 | 0,67 |
| | 2 | 1,3582 | 4,3453 | 0,689 | 1,53 |
| | 3 | 0,4822 | 1,944 | 1,091 | 2,89 |
| | 4 | 0,1994 | 1,252 | 1,381 | 5,61 |
| | 5 | 0,0563 | 1,0263 | 1,533 | 17,99 |

Tabelle F.4.: Filterkoeffizienten für Tschebyscheff-Filter mit 0.5 dB Welligkeit

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

| Ordnung | Filterstufe | $a_i$ | $b_i$ | $\frac{f_{g_i}}{f_g}$ | $Q_i$ |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 | - |
| 2 | 1 | 1,3022 | 1,5515 | 1 | 0,96 |
| 3 | 1 | 2,2156 | 0 | 0,451 | - |
| | 2 | 0,5442 | 1,2057 | 1,353 | 2,02 |
| 4 | 1 | 2,5904 | 4,1301 | 0,54 | 0,78 |
| | 2 | 0,3039 | 1,1697 | 1,417 | 3,56 |
| 5 | 1 | 3,5711 | 0 | 0,28 | - |
| | 2 | 1,128 | 2,4896 | 0,894 | 1,4 |
| | 3 | 0,1872 | 1,0814 | 1,486 | 5,56 |
| 6 | 1 | 3,8437 | 8,5529 | 0,366 | 0,76 |
| | 2 | 0,6292 | 1,9124 | 1,082 | 2,2 |
| | 3 | 0,1296 | 1,0766 | 1,493 | 8 |
| 7 | 1 | 4,952 | 0 | 0,202 | - |
| | 2 | 1,6338 | 4,4899 | 0,655 | 1,3 |
| | 3 | 0,3987 | 1,5834 | 1,213 | 3,16 |
| | 4 | 0,0937 | 1,0423 | 1,52 | 10,9 |
| 8 | 1 | 5,1019 | 14,7608 | 0,276 | 0,75 |
| | 2 | 0,8916 | 3,0426 | 0,849 | 1,96 |
| | 3 | 0,2806 | 1,4334 | 1,285 | 4,27 |
| | 4 | 0,0717 | 1,0432 | 1,52 | 14,24 |
| 9 | 1 | 6,3415 | 0 | 0,158 | - |
| | 2 | 2,1252 | 7,1711 | 0,514 | 1,26 |
| | 3 | 0,5624 | 2,3278 | 0,994 | 2,71 |
| | 4 | 0,2076 | 1,3166 | 1,346 | 5,53 |
| | 5 | 0,0562 | 1,0258 | 1,533 | 18,03 |
| 10 | 1 | 6,3634 | 22,7468 | 0,221 | 0,75 |
| | 2 | 1,1399 | 4,5167 | 0,694 | 1,86 |
| | 3 | 0,3939 | 1,9665 | 1,093 | 3,56 |
| | 4 | 0,1616 | 1,2569 | 1,381 | 6,94 |
| | 5 | 0,0455 | 1,0277 | 1,532 | 22,26 |

Tabelle F.5.: Filterkoeffizienten für Tschebyscheff-Filter mit 1 dB Welligkeit

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

| Ordnung | Filterstufe | $a_i$ | $b_i$ | $\frac{f_{g_i}}{f_g}$ | $Q_i$ |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 | - |
| 2 | 1 | 1,1813 | 1,7775 | 1 | 1,13 |
| 3 | 1 | 2,7994 | 0 | 0,357 | - |
| | 2 | 0,43 | 1,2036 | 1,378 | 2,55 |
| 4 | 1 | 2,4025 | 4,9862 | 0,55 | 0,93 |
| | 2 | 0,2374 | 1,1896 | 1,413 | 4,59 |
| 5 | 1 | 4,6345 | 0 | 0,216 | - |
| | 2 | 0,909 | 2,6036 | 0,908 | 1,78 |
| | 3 | 0,1434 | 1,075 | 1,493 | 7,23 |
| 6 | 1 | 3,588 | 10,4648 | 0,373 | 0,9 |
| | 2 | 0,4925 | 1,9622 | 1,085 | 2,84 |
| | 3 | 0,0995 | 1,0826 | 1,491 | 10,46 |
| 7 | 1 | 6,476 | 0 | 0,154 | - |
| | 2 | 1,3258 | 4,7649 | 0,665 | 1,65 |
| | 3 | 0,3067 | 1,5927 | 1,218 | 4,12 |
| | 4 | 0,0714 | 1,0384 | 1,523 | 14,28 |
| 8 | 1 | 4,7743 | 18,151 | 0,282 | 0,89 |
| | 2 | 0,6991 | 3,1353 | 0,853 | 2,53 |
| | 3 | 0,2153 | 1,4449 | 1,285 | 5,58 |
| | 4 | 0,0547 | 1,0461 | 1,518 | 18,69 |
| 9 | 1 | 8,3198 | 0 | 0,12 | - |
| | 2 | 1,7299 | 7,658 | 0,522 | 1,6 |
| | 3 | 0,4337 | 2,3549 | 0,998 | 3,54 |
| | 4 | 0,1583 | 1,3174 | 1,349 | 7,25 |
| | 5 | 0,0427 | 1,0232 | 1,536 | 23,68 |
| 10 | 1 | 5,9618 | 28,0376 | 0,226 | 0,89 |
| | 2 | 0,8947 | 4,6644 | 0,697 | 2,41 |
| | 3 | 0,3023 | 1,9858 | 1,094 | 4,66 |
| | 4 | 0,1233 | 1,2614 | 1,38 | 9,11 |
| | 5 | 0,0347 | 1,0294 | 1,531 | 29,27 |

Tabelle F.6.: Filterkoeffizienten für Tschebyscheff-Filter mit 2 dB Welligkeit

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

| Ordnung | Filterstufe | $a_i$ | $b_i$ | $\frac{f_{g_i}}{f_g}$ | $Q_i$ |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 0 | 1 | - |
| 2 | 1 | 1,065 | 1,9305 | 1 | 1,3 |
| 3 | 1 | 3,3496 | 0 | 0,299 | - |
| | 2 | 0,3559 | 1,1923 | 1,396 | 3,07 |
| 4 | 1 | 2,1853 | 5,5339 | 0,557 | 1,08 |
| | 2 | 0,1964 | 1,2009 | 1,41 | 5,58 |
| 5 | 1 | 5,6334 | 0 | 0,178 | - |
| | 2 | 0,762 | 2,653 | 0,917 | 2,14 |
| | 3 | 0,1172 | 1,0686 | 1,5 | 8,82 |
| 6 | 1 | 3,2721 | 11,6773 | 0,379 | 1,04 |
| | 2 | 0,4077 | 1,9873 | 1,086 | 3,46 |
| | 3 | 0,0815 | 1,0861 | 1,489 | 12,78 |
| 7 | 1 | 7,9064 | 0 | 0,126 | - |
| | 2 | 1,1159 | 4,8963 | 0,67 | 1,98 |
| | 3 | 0,2515 | 1,5944 | 1,222 | 5,02 |
| | 4 | 0,0582 | 1,0348 | 1,527 | 17,46 |
| 8 | 1 | 4,3583 | 20,2948 | 0,286 | 1,03 |
| | 2 | 0,5791 | 3,1808 | 0,855 | 3,08 |
| | 3 | 0,1765 | 1,4507 | 1,285 | 6,83 |
| | 4 | 0,0448 | 1,0478 | 1,517 | 22,87 |
| 9 | 1 | 10,1759 | 0 | 0,098 | - |
| | 2 | 1,4585 | 7,8971 | 0,526 | 1,93 |
| | 3 | 0,3561 | 2,3651 | 1,001 | 4,32 |
| | 4 | 0,1294 | 1,3165 | 1,351 | 8,87 |
| | 5 | 0,0348 | 1,021 | 1,537 | 29 |
| 10 | 1 | 5,4449 | 31,3788 | 0,23 | 1,03 |
| | 2 | 0,7414 | 4,7363 | 0,699 | 2,94 |
| | 3 | 0,2479 | 1,9952 | 1,094 | 5,7 |
| | 4 | 0,1008 | 1,2638 | 1,38 | 11,15 |
| | 5 | 0,0283 | 1,0304 | 1,53 | 35,85 |

Tabelle F.7.: Filterkoeffizienten für Tschebyscheff-Filter mit 3 dB Welligkeit

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

| Ordnung | Filter-Stufe | $a_i$ | $b_i$ | $\frac{f_{g_i}}{f_g}$ | $Q_i$ | $T_{gr}$ |
|---|---|---|---|---|---|---|
| 1 | 1 | 0,6436 | 0 | 1,554 | - | 0,2049 |
| 2 | 1 | 1,6278 | 0,8823 | 1,064 | 0,58 | 0,5181 |
| 3 | 1 | 1,1415 | 0 | 0,876 | - | 0,8437 |
| | 2 | 1,5092 | 1,0877 | 0,959 | 0,69 | |
| 4 | 1 | 2,3370 | 1,4878 | 0,820 | 0,52 | 1,1738 |
| | 2 | 1,3506 | 1,1837 | 0,919 | 0,82 | |
| 5 | 1 | 1,2974 | 0 | 0,771 | - | 1,5060 |
| | 2 | 2,2224 | 1,5685 | 0,798 | 0,56 | |
| | 3 | 1,2116 | 1,2330 | 0,901 | 0,92 | |
| 6 | 1 | 2,6117 | 1,7763 | 0,750 | 0,51 | 1,8395 |
| | 2 | 2,0706 | 1,6015 | 0,790 | 0,61 | |
| | 3 | 1,0967 | 1,2596 | 0,891 | 1,02 | |
| 7 | 1 | 1,3735 | 0 | 0,728 | - | 2,1737 |
| | 2 | 2,5320 | 1,8169 | 0,742 | 0,53 | |
| | 3 | 1,9211 | 1,6116 | 0,788 | 0,66 | |
| | 4 | 1,0023 | 1,2743 | 0,886 | 1,13 | |
| 8 | 1 | 2,7541 | 1,9420 | 0,718 | 0,51 | 2,5084 |
| | 2 | 2,4174 | 1,8300 | 0,739 | 0,56 | |
| | 3 | 1,7850 | 1,6101 | 0,788 | 0,71 | |
| | 4 | 0,9239 | 1,2822 | 0,883 | 1,23 | |
| 9 | 1 | 1,4186 | 0 | 0,705 | - | 2,8434 |
| | 2 | 2,6979 | 1,9659 | 0,713 | 0,52 | |
| | 3 | 2,2940 | 1,8282 | 0,740 | 0,59 | |
| | 4 | 1,6644 | 1,6027 | 0,790 | 0,76 | |
| | 5 | 0,8579 | 1,2862 | 0,882 | 1,32 | |
| 10 | 1 | 2,8406 | 2,0490 | 0,699 | 0,50 | 3,1786 |
| | 2 | 2,6120 | 1,9714 | 0,712 | 0,59 | |
| | 3 | 2,1733 | 1,8184 | 0,742 | 0,62 | |
| | 4 | 1,5583 | 1,5923 | 0,792 | 0,81 | |
| | 5 | 0,8018 | 1,2877 | 0,881 | 1,42 | |

Tabelle F.8.: Filterkoeffizienten für Allpassfilter mit maximal flacher Gruppenlaufzeit

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

# G. Maple V Texte

Die verwendete Maple-Datei[1] finden Sie im Internet.

## G.1. Ortskurve in der komplexen Ebene

```
> ReF := (Omega,Q) -> 1/((1-Omega^2)^2+(Omega/Q)^2);
> ImF := (Omega,Q) -> Omega * ((1-Omega^2) * Q-1/Q) / ((1-Omega^2)^2+(Omega/Q)^2);
> ReFO := (Omega,Q) -> 1/((1-Omega^2)^2+(Omega/Q)^2)/Q^2;
> ImFO := (Omega,Q) -> Omega * ((1-Omega^2) * Q-1/Q) / ((1-Omega^2)^2+(Omega/Q)^2)/Q^
> plot({[ReFO(10^x,1), ImFO(10^x,1), x=-infinity..infinity], [ReFO(10^x,3),
ImFO(10^x,3), x=-infinity..infinity], [ReFO(10^x,10), ImFO(10^x,10),
x=-infinity..infinity]} ,0..2.5,-1.1..0.6,numpoints=100,title=‘Q=1,3,10‘);
> plot({[ReF(10^x,1), ImF(10^x,1), x=-100..100], [ReF(10^x,3), ImF(10^x,3),
x=-100..100], [ReF(10^x,10), ImF(10^x,10), x=-100..100]}, 0..100,-55..55,numpoints=10
```

## G.2. Definitionen der Filterfunktionen

### G.2.1. Kritische Filter

```
> f := (P)->1/(1+P);
> an := n->(2^(1/n)-1)^(1/2);
> fn := (P,n)->(1+an(n) * P)^(-n);
```

### G.2.2. Butterworth

```
> butterworth1 := (P)->fn(P,1);
> butterworth3 := P->1/((1+P) * (1+P+P^2));
> butterworth10 := P->1/((1+1.9754 * P+P^2) * (1+1.782 * P+P^2) * (1+1.4142
* P+P^2) * (1+0.9080 * P+P^ 2) * (1+0.3129 * P+P^2));
```

### G.2.3. Bessel

```
> bessel1 := (P)->fn(P,1);
> bessel3 := P->1/((1+0.756 * P) * (1+0.996 * P+0.4772 * P^2));
> bessel10 := P->1/((1+1.0215 * P+0.2650 * P^2) * (1+0.9393 * P+0.2549
* P^2) * (1+0.7815 * P+0.2351 * P^2) * (1+0.5604 * P+0.2059 * P^2) *
(1+0.2883 * P+0.1665 * P^2));
```

[1] http://wwwex.physik.uni-ulm.de/lehre/PhysikalischeElektronik/Materialien/Gleichungen.mws

### G.2.4. Tschebyscheff 1dB

```
> Tscheby1_1 := (P)->fn(P,1);
> Tscheby1_3 := P->1/((1+2.2156 * P) * (1+0.5442 * P+1.2057 * P^2));
> Tscheby1_10 := P->1/((1+6.3624 * P+22.7468 * P^2) * (1+1.1399 * P+4.5167
* P^2) * (1+0.3939 * P+1.9665 * P^2) * (1+0.1616 * P+1.2569 * P^2) *
(1+0.0455 * P+1.0277 * P^2));
```

### G.2.5. Tschebyscheff 3dB

```
> Tscheby3_1 := (P)->fn(P,1);
> Tscheby3_3 := P->1/((1+3.3496 * P) * (1+0.3559 * P+1.1923 * P^2));
> Tscheby3_10 := P->1/((1+5.4449 * P+31.3788 * P^2) * (1+0.7414 * P+4.7363
* P^2) * (1+0.2479 * P+1.9952 * P^2) * (1+0.1008 * P+1.2638 * P^2) *
(1+0.0283 * P+1.0304 * P^2));
```

### G.2.6. Allpass

```
> Allpass_1 := P-> (1-0.6436 * P)/(1+0.6436 * P);
> Allpass_3 := P -> ((1-1.1415 * P) * (1-1.5092 * P+1.0877 * P^2))/((1+1.1415
* P) * (1+1.5092 * P+1.087 7 * P^2));
> Allpass_10 := P->( (1-2.8406 * P+2.0490 * P^2) * (1-2.6120 * P+1.9714
* P^2) * (1-2.1733 * P+1.8184 * P^2) * (1-1.5583 * P+1.5923 * P^2) *
(1-0.8018 * P+1.2877 * P^2) )/ ( (1+2.8406 * P+2.0490 * P^2) * (1+2.6120
* P+1.9714 * P^2) * (1+2.1733 * P+1.8184 * P^2) * (1+1.5583 * P+1.5923
* P^2) * (1+0.8018 * P+1.2877 * P^2) ) ;
```

### G.2.7. Schwingkreis

```
> Schwink := (P,Q) -> (1+P * Q) / (1+P/Q+P^2);
> simplify(evalc(Re(Schwink(I * Omega,Q))));
> simplify(evalc(Im(Schwink(I * Omega,Q))));
```

## G.3. Darstellung der Filter

Sollen keine Tiefpässe dargestellt werden, muss das Argument `I*x` durch das entsprechende transformierte Element ersetzt werden.

### G.3.1. Tiefpass-Hochpasstransformation

```
> th := P -> 1/P;
```

### G.3.2. Tiefpass-Bandpasstransformation

```
> tb := (P) -> (P+1/P)/domega;
```

### G.3.3. Tiefpass-Bandsperrentransformation

```
> tbs := (P) -> P * domega/(P^2+1);
```

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

## G.3.4. Beispiel:Butterworth Tiefpässe

Die Funktion `butterworth3(I*10^x)` muss jeweils durch die entsprechenden Filterfunktionen für die anderen Filter ersetzt werden.

### G.3.4.1. Phasenbild

```
> p1 := complexplot(butterworth1(I * 10^x), x=-10..10,color=black, labels=[‘Realteil‘
title=‘Butterworth Tiefpass 1.,3.,10. Ordnung + kritischer Tiefpass 10.
Ordnung‘, titlefont=[HELVETICA,BOLD,9], numpoints=200, thickness=2, axes=boxed,
line style=1):
> p2 := complexplot(butterworth3(I * 10^x), x=-10..10,color=red, numpoints=200,thickn
axes=boxed,linestyle=1):
> p3 := complexplot(butterworth10(I * 10^x), x=-10..10,color=blue, numpoints=200,thic
axes=boxed,linestyle=1):
> p4 := complexplot(fn(I * 10^x,10), x=-10..10,color=black, numpoints=200,thickness=2
axes=boxed,linestyle=1,color=green):
> plots[display]({p1,p2,p3,p4});
```

### G.3.4.2. Bodeplot:Amplitude

```
> p1 := loglogplot(evalf(abs(butterworth1(I * x))), x=0.01..100,color=black,
labels=[‘Frequenz‘,‘Amplitude‘], title=‘Butterworth Tiefpass 1.,3.,10.
Ordnung‘, titlefont=[HELVETICA, BOLD,9], numpoints=200, thickness=2,
axes=boxed, linestyle=1, view=[0.01..100,10^(-2)..2]):
> p2 := loglogplot(evalf(abs(butterworth3(I * x))), x=0.01..100,color=red,
numpoints=200,thickness=2, axes=boxed,linestyle=1, view=[0.01..100,10^(-2)..2]):
> p3 := loglogplot(evalf(abs(butterworth10(I * x))), x=0.010..100,color=blue,
numpoints=200,thickness=2, axes=boxed,linestyle=1, view=[0.01..100,10^(-2)..2]):
> plots[display]({p1,p2,p3});
```

### G.3.4.3. Bodeplot:Phase

```
> p1 := semilogplot(evalf(180 * argument(butterworth1(I * x))/Pi), x=0.01..100,
color=black, labels=[‘Frequenz‘,‘Phase‘], title=‘Butterworth Tiefpass
1.,3.,10.Ordnung‘, titlefont=[HELVETICA,BOLD,9], numpoints=200,thickness=2,
axes=boxed,linestyle=1, view=[0.01..100,-180..180]):
> p2 := semilogplot(evalf(180 * argument(butterworth3(I * x))/Pi), x=0.01..100,color=
numpoints=40,thickness=2, axes=boxed,linestyle=1, view=[0.01..100,-180..180]):
> p3 := semilogplot(evalf(180 * argument(butterworth10(I * x))/Pi), x=0.01..100,color
numpoints=200,thickness=2, axes=boxed,linestyle=1, view=[0.01..100,-180..180]):
> plots[display]({p1,p2,p3});
```

### G.3.4.4. Gruppenlaufzeit

```
> p1 := semilogplot(eval( -diff(evalc(argument( butterworth1(I * y))),
y), y = x )/(2 * Pi), x=0.01..100,color=black, labels=[‘Frequenz‘,‘Gruppenlaufzeit‘],
title=‘Butterworth Tiefpass 1.,3.,10. Ordnung‘,titlefont=[HELVETICA,BOLD,9],
numpoints=200,thickness=2,axes=boxed,linestyle=1, view=[0.01..100,0.0..2]):
```

```
> p2 := semilogplot(eval( -diff(evalc(argument( butterworth3(I * y))),
y), y = x )/(2 * Pi), x=0.01..100,color=red, numpoints=200,thickness=2,axes=boxe
view=[0.01..100,0.0..2]):
> p3 := semilogplot(eval( -diff(evalc(argument( butterworth10(I * y))),
y), y=x )/(2 * Pi), x=0.01..100,color=blue, numpoints=200,thickness=2,axes=boxed
view=[0.01..100,0.0..2]):
> plots[display]({p1,p2,p3});
```

## G.4. Smith-Chart

***Smith Charts***

Dieses Maple-Programm beschreibt, wie Smith-Charts konstruiert werden. Smith-Charts sind bilineare Transformation der Impedanzebene (als Koordinate der Real- und der Imaginärteil) in die Ebene komplexer Reflektionskoeffizienten *gamma*.

```
> with(plots):
     setoptions(title=‘Smith Chart‘, axes=BOXED);

> mgamma:= z->(z-1)/(z+1);
```

$$mgamma := z \to \frac{z-1}{z+1}$$

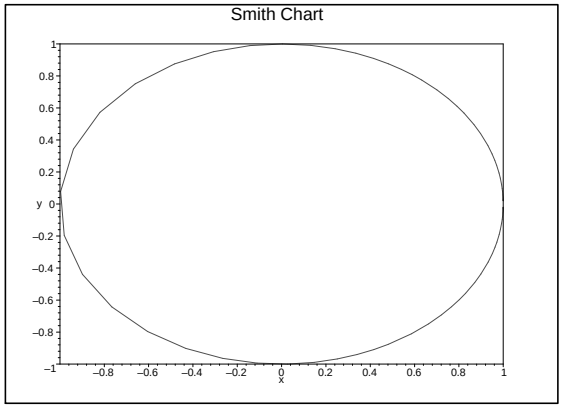
Smith Charts sind in der komplexen Ebene des Reflektionskoeffizienten *gamma*. Die Koordinaten sind jedoch die Impedanzen (Realteile) und die Admittanzen (Imaginärteile). Die Linien gleicher Realteile (reine Widerstände) und gleicher Imaginärteile (z.B. Kondensatoren) sind die neuen Achsen.

Als Beispiel wird der Realteil gleich null gesetzt und der Imaginärteil variiert.

Im folgenden werden normierte Grössen verwendet.

Damit bekommt man das Bild der imaginären Achse in der Gamma-Ebene

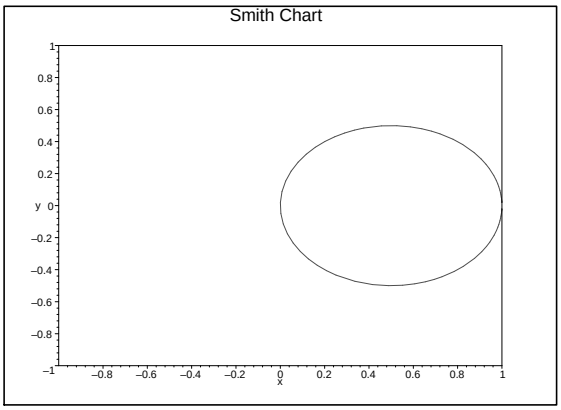
```
> plot([Re(mgamma(0 + I *x)), Im(mgamma(0 +I*x)),x=-100..100],
    x=-1..1,y=-1..1);
```




 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

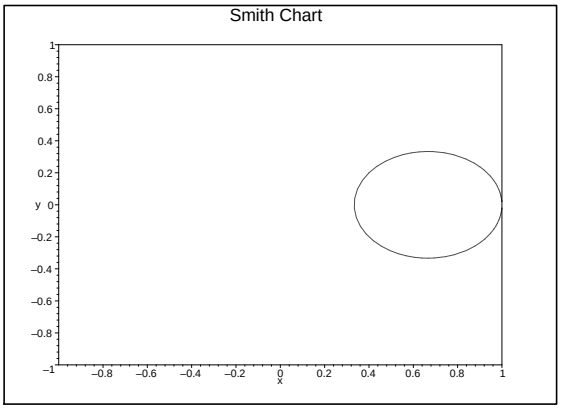
Als weiteres Beispiel setzen wir den Realteil auf 1.

```
> plot([Re(mgamma(1 + I *x)), Im(mgamma(1 + I*x)),
     x=-100..100],x=-1..1,y=-1..1);
```



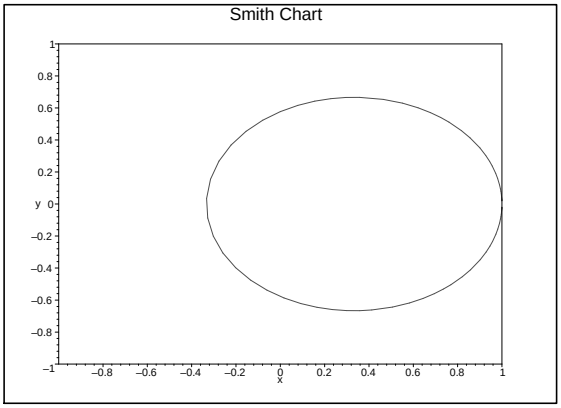

Und nun wird der Realteil auf 2 gesetzt.

```
> plot([Re(mgamma(2 + I *t)), Im(mgamma(2 + I*t)),
     t=-100..100],x=-1..1,y=-1..1);
```




Schliesslich setzen wir den konstanten Realteil auf 0.5.

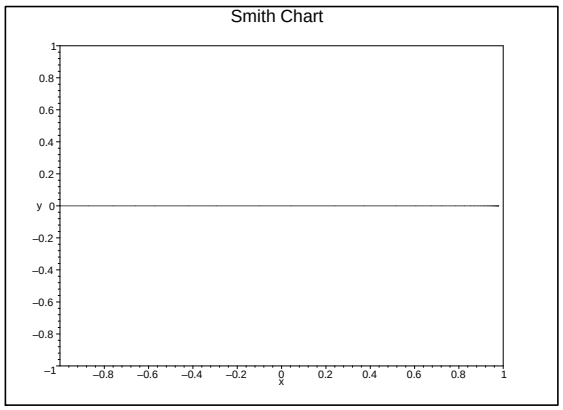
```
> plot([Re(mgamma(0.5 + I *t)), Im(mgamma(0.5 +I*t)),
     t=-100..100],x=-1..1,y=-1..1);
```




©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

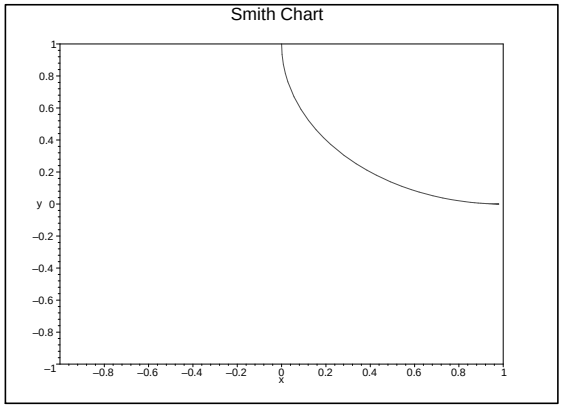
Nun werden wir den Imaginärteil festhalten. Ein reiner Widerstand (nur reale Grössen) wird wie im folgenden Bild abgebildet. Hier ist der Imaginärteil gleich null.
Der reale Widerstand R=1 wird auf gamma=0+I*0 abgebildet.

```
> plot([Re(mgamma(r + 0*I )), Im(mgamma(r +0*I)),r=0..100],
     x=-1..1,y=-1..1);
```




Nun setzen wir den Imaginärteil auf den konstanten Wert I.

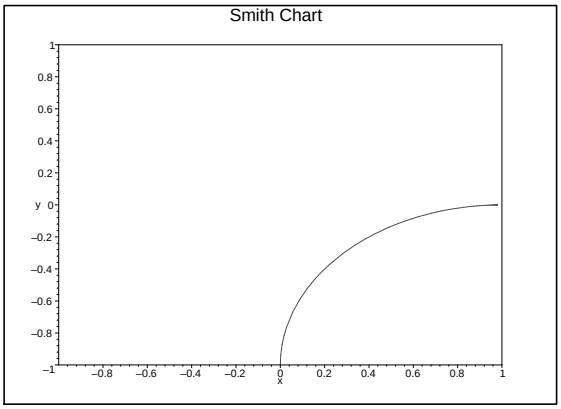
```
> plot([Re(mgamma(r + I )), Im(mgamma(r + I)),r=0..100],
     x=-1..1,y=-1..1);
```




Wenn der Imaginärteil -I ist, erhält man

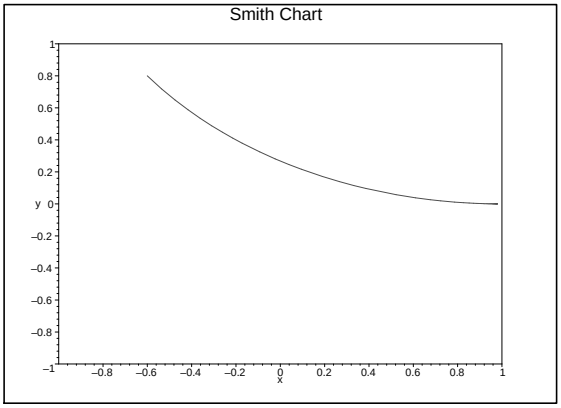
```
> plot([Re(mgamma(r - I )), Im(mgamma(r - I)),r=0..100],
     x=-1..1,y=-1..1);
```

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl



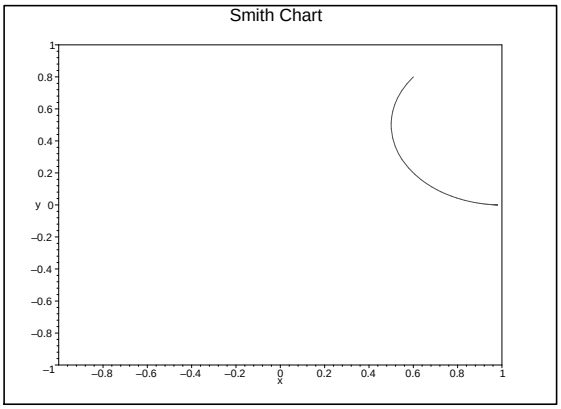

Für 0.5 I erhält man

```
> plot([Re(mgamma(r + 0.5*I )), Im(mgamma(r +0.5*I)),
     r=0..100],x=-1..1,y=-1..1);
```




Wenn der Imaginärteil 2 I ist, dann bekommt man

```
> plot([Re(mgamma(r + 2*I )), Im(mgamma(r +2*I)),
     r=0..100],x=-1..1,y=-1..1);
```




Für eine Smith-Chart werden nun die Abbildungen des Gitters des kartesischen Koordinatensystems übereinandergelegt.

Wir beginnen, indem wir eine Konstante definieren.

```
> maxi := 10;
```

$$maxi := 10$$

Damit erstellen wir einen Array der Grösse 4*maxi für die Bilder der Linien mit konstantem Imaginärteil.

```
> pp := array(1..4*maxi);
```

$$pp := \mathrm{array}(1..40, [])$$

Nun füllen wir den Array mit den Plots für die waagrechten (Imaginärteil = konstant) Linien aus der Widerstandsebene.

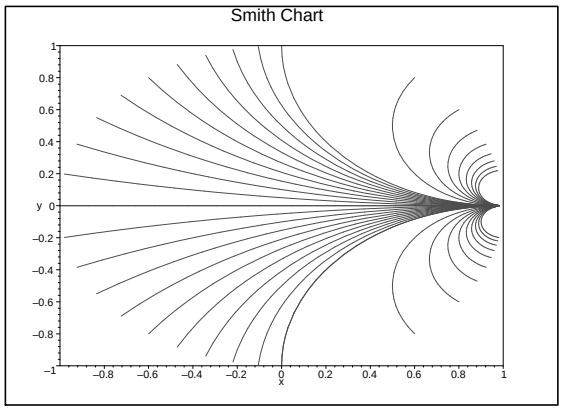
```
> for i from 1 to 2*maxi do
          pp[i] := plot([Re(mgamma(r - 10*(i-maxi)*I/maxi )),
           Im(mgamma(r -10*(i-maxi)*I/maxi)),
          r=0..100],x=-1..1,y=-1..1):
          pp[i+2*maxi] := plot([Re(mgamma(r - 1*(i-maxi)*I/maxi )),
          Im(mgamma(r - 1*(i-maxi)*I/maxi)),
          r=0..100],x=-1..1,y=-1..1):
     od:
```

Aus den Plots machen wir eine Liste

```
> L := seq(pp[i],i=1..4*maxi):
```

und stellen sie dar.

```
> plots[display](L);
```




Wir definieren einen Array der Grösse 4*maxi für die Bilder der Linien mit konstantem Realteil.

```
> rr := array(1..4*maxi);
```

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

$rr := \text{array}(1..40, [])$

Nun füllen wir den Array mit den Plots für die senkrechten (Realteil = konstant) Linien aus der Widerstandsebene.

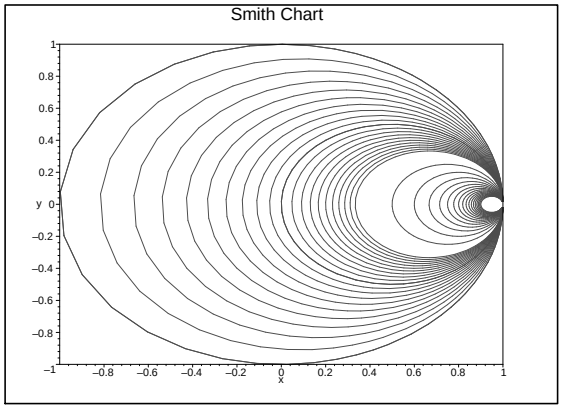
```
> for i from 1 to 2*maxi do
          rr[i] := plot([Re(mgamma(10*(i-1)/maxi+r*I )),
          Im(mgamma(10*(i-1)/maxi+r*I)),
          r=-100..100],x=-1..1,y=-1..1):
          rr[i+2*maxi] := plot([Re(mgamma(1*(i-1)/maxi+r*I )),
          Im(mgamma(1*(i-1)/maxi+r*I)),
          r=-100..100],x=-1..1,y=-1..1):
     od:
```

Aus den Plots machen wir eine Liste

```
> LL := seq(rr[i],i=1..4*maxi):
```

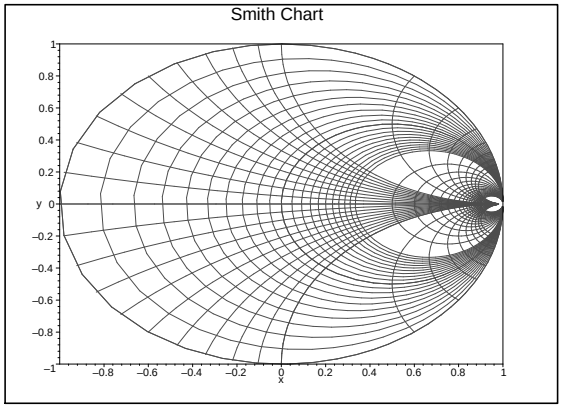
und stellen sie dar.

```
> plots[display](LL);
```




Wenn wir beide Listen übereinander zeichnen, erhalten wir eine Smith-Chart

```
> plots[display](L,LL);
```



Die obige Darstellung ist immer noch nach den gamma-Werten bezeichnet. Nun müsste das ganze noch mit den Real- und Imaginärteilen der Widerstände bezeichnet werden.

## G.5. Smith-Charts mit Gnuplot 4.2

Um eine *Smith-Chart* mit Gnuplot 4.2 herzustellen, braucht man zwei Dateien. Die erste, *smithgrid.plt* dient als Steuerdatei, während die zweite, *smithloop.plt* die einzelnen Kurven zeichnet.
Der Inhalt der ersten Datei („*smithgrid.plt*“) ist:

```
reset;
unset multiplot
set encoding iso_8859_1
cd 'ihr Verzeichnis';
set terminal win;
complex(x,y) = x*{1,0}+y*{0,1}
conformal(z) = (z-1)/(z+1)
pi = 4*atan(1)
lblIMRE(x) = 1.05*real(conformal(complex(0,x)))-0.025
lblIMIM(x) = 1.05*imag(conformal(complex(0,x)))
lblRERE(x) = real(conformal(complex(x,0)))+0.005

set nokey
set size square
set parametric
unset xlabel
unset ylabel
set title "Smith-Chart" font "Arial,24"
unset logscale x
unset logscale y
set xrange[-1.1:1.1]
set yrange[-1.1:1.1]
set samples 1500


set label 1 "0" at lblIMRE(0),lblIMIM(0) font "Arial,8"
set label 2 "0.1" at lblIMRE(0.1),lblIMIM(0.1) font "Arial,8"
set label 3 "0.2" at lblIMRE(0.2),lblIMIM(0.2) font "Arial,8"
set label 4 "0.3" at lblIMRE(0.3),lblIMIM(0.3) font "Arial,8"
set label 5 "0.4" at lblIMRE(0.4),lblIMIM(0.4) font "Arial,8"
set label 6 "0.5" at lblIMRE(0.5),lblIMIM(0.5) font "Arial,8"
set label 7 "0.6" at lblIMRE(0.6),lblIMIM(0.6) font "Arial,8"
set label 8 "0.7" at lblIMRE(0.7),lblIMIM(0.7) font "Arial,8"
set label 9 "0.8" at lblIMRE(0.8),lblIMIM(0.8) font "Arial,8"
set label 10 "0.9" at lblIMRE(0.9),lblIMIM(0.9) font "Arial,8"
set label 20 "1" at lblIMRE(1),lblIMIM(1) font "Arial,8"
set label 21 "2" at lblIMRE(2),lblIMIM(2) font "Arial,8"
set label 22 "3" at lblIMRE(3),lblIMIM(3) font "Arial,8"
set label 23 "4" at lblIMRE(4),lblIMIM(4) font "Arial,8"
set label 24 "5" at lblIMRE(5),lblIMIM(5) font "Arial,8"
set label 25 "6" at lblIMRE(6),lblIMIM(6) font "Arial,8"
set label 26 "7" at lblIMRE(7),lblIMIM(7) font "Arial,8"

set label 102 "-0.1" at lblIMRE(-0.1),lblIMIM(-0.1) font "Arial,8"
set label 103 "-0.2" at lblIMRE(-0.2),lblIMIM(-0.2) font "Arial,8"
set label 104 "-0.3" at lblIMRE(-0.3),lblIMIM(-0.3) font "Arial,8"
set label 105 "-0.4" at lblIMRE(-0.4),lblIMIM(-0.4) font "Arial,8"
set label 106 "-0.5" at lblIMRE(-0.5),lblIMIM(-0.5) font "Arial,8"
set label 107 "-0.6" at lblIMRE(-0.6),lblIMIM(-0.6) font "Arial,8"
set label 108 "-0.7" at lblIMRE(-0.7),lblIMIM(-0.7) font "Arial,8"
set label 109 "-0.8" at lblIMRE(-0.8),lblIMIM(-0.8) font "Arial,8"
set label 1010 "-0.9" at lblIMRE(-0.9),lblIMIM(-0.9) font "Arial,8"
set label 1020 "-1" at lblIMRE(-1),lblIMIM(-1) font "Arial,8"
set label 1021 "-2" at lblIMRE(-2),lblIMIM(-2) font "Arial,8"
set label 1022 "-3" at lblIMRE(-3),lblIMIM(-3) font "Arial,8"
set label 1023 "-4" at lblIMRE(-4),lblIMIM(-4) font "Arial,8"
set label 1024 "-5" at lblIMRE(-5),lblIMIM(-5) font "Arial,8"
set label 1025 "-6" at lblIMRE(-6),lblIMIM(-6) font "Arial,8"
set label 1026 "-7" at lblIMRE(-7),lblIMIM(-7) font "Arial,8"

set label 2000 "0" at lblRERE(0),0.025 font "Arial,8"
set label 2001 "0.1" at lblRERE(0.1),0.025 font "Arial,8"
set label 2002 "0.2" at lblRERE(0.2),0.025 font "Arial,8"
set label 2003 "0.3" at lblRERE(0.3),0.025 font "Arial,8"
set label 2004 "0.4" at lblRERE(0.4),0.025 font "Arial,8"
set label 2005 "0.5" at lblRERE(0.5),-0.025 font "Arial,8"
set label 2006 "0.6" at lblRERE(0.6),0.025 font "Arial,8"
set label 2007 "0.7" at lblRERE(0.7),-0.025 font "Arial,8"
set label 2008 "0.8" at lblRERE(0.8),0.025 font "Arial,8"
set label 2009 "0.9" at lblRERE(0.9),-0.025 font "Arial,8"
set label 2010 "1" at lblRERE(1),0.025 font "Arial,8"
set label 2011 "2" at lblRERE(2),0.025 font "Arial,8"
set label 2012 "3" at lblRERE(3),0.025 font "Arial,8"
set label 2013 "4" at lblRERE(4),0.025 font "Arial,8"
set label 2014 "5" at lblRERE(5),0.025 font "Arial,8"

set label 3000 "Imaginärteil" at -1,0.6 rotate by 45 font "Arial,12"
```

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

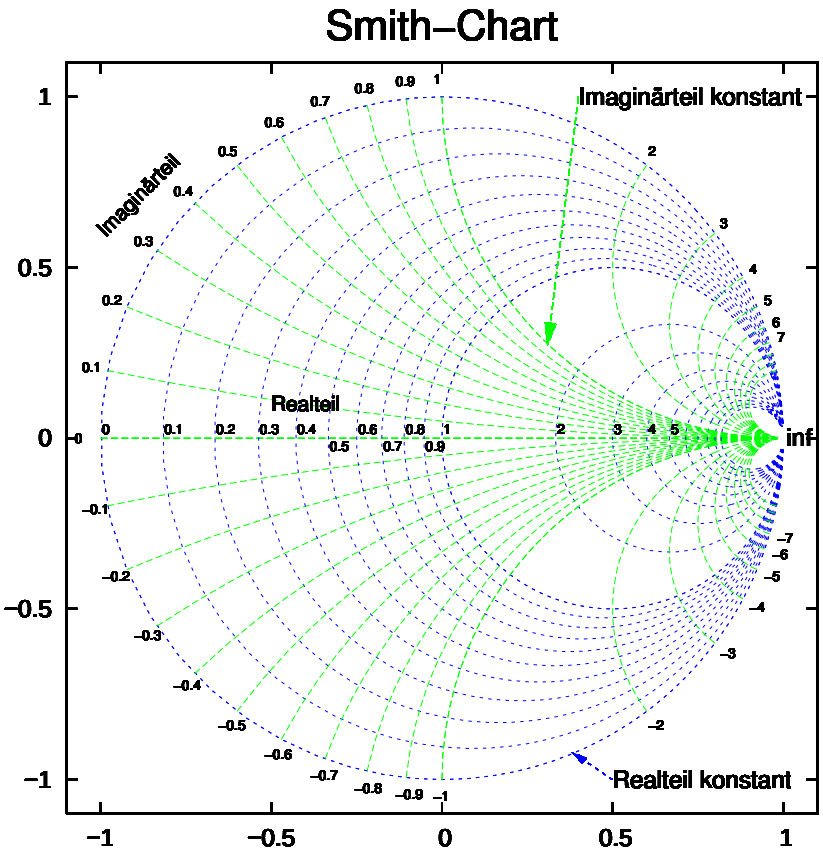



Abbildung G.1.: *Smith-Chart* mit Gnuplot erstellt

```
set label 3001 "Realteil" at -0.5,0.1 font "Arial,12"


set label 30 "inf" at 1.01,0

a=-1
set multiplot
load 'smithloop.plt'
unset multiplot
unset parametric
```

Die Schleifendatei („*smithloop.plt* “) lautet

```
a = a+1
set trange [-60:60]
plot real(conformal(complex(a,t))),imag(conformal(complex(a,t))) lt 3
plot real(conformal(complex(0.1*a,t))),imag(conformal(complex(0.1*a,t))) lt 3

set trange [0:120]
plot real(conformal(complex(t,a))),imag(conformal(complex(t,a))) lt 2
plot real(conformal(complex(t,a))),-imag(conformal(complex(t,a))) lt 2
plot real(conformal(complex(t,0.1*a))),imag(conformal(complex(t,0.1*a))) lt 2
plot real(conformal(complex(t,0.1*a))),-imag(conformal(complex(t,0.1*a))) lt 2


if(a<10) reread
```

Abbildung G.1 zeigt das Resultat, wenn *smithgrid.plt* mit Gnuplot 4.2 verarbeitet wird.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

## G.6. Daten auf Smith-Charts zeichnen mit Gnuplot 4.2

Die folgende Anweisung ist eine Erweiterung des Gnu-Plot-Verfahrens durch Helmut Sennewald (helmutsennewald (at) t-online.de). Neben der Schleifendatei („*smithloop.plt* “) verwenden wir eine Datei „*smithgrid-n.plt*“ um den Smith-Chart zu zeichnen und die Datei „*smith_plot_data.plt*“ um die Datendatei „*data1.txt*“ zu zeichnen.
Die Datei „*smithgrid-n.plt*“ ist eine modifizierte Version der Datei „*smithgrid.plt*“ aus dem vorherigen Abschnitt.

```
lblIMRE(x) = 1.05*real(conformal(complex(0,x)))-0.025
lblIMIM(x) = 1.05*imag(conformal(complex(0,x)))
lblRERE(x) = real(conformal(complex(x,0)))+0.005

unset key
set size square
set parametric
unset xlabel
unset ylabel
set title "Smith-Chart" font "Arial,24"
unset logscale x
unset logscale y
set xrange[-1.1:1.1]
set yrange[-1.1:1.1]
set samples 1500


set label 1 "0" at lblIMRE(0),lblIMIM(0) font "Arial,8"
set label 2 "0.1" at lblIMRE(0.1),lblIMIM(0.1) font "Arial,8"
set label 3 "0.2" at lblIMRE(0.2),lblIMIM(0.2) font "Arial,8"
set label 4 "0.3" at lblIMRE(0.3),lblIMIM(0.3) font "Arial,8"
set label 5 "0.4" at lblIMRE(0.4),lblIMIM(0.4) font "Arial,8"
set label 6 "0.5" at lblIMRE(0.5),lblIMIM(0.5) font "Arial,8"
set label 7 "0.6" at lblIMRE(0.6),lblIMIM(0.6) font "Arial,8"
set label 8 "0.7" at lblIMRE(0.7),lblIMIM(0.7) font "Arial,8"
set label 9 "0.8" at lblIMRE(0.8),lblIMIM(0.8) font "Arial,8"
set label 10 "0.9" at lblIMRE(0.9),lblIMIM(0.9) font "Arial,8"
set label 20 "1" at lblIMRE(1),lblIMIM(1) font "Arial,8"
set label 21 "2" at lblIMRE(2),lblIMIM(2) font "Arial,8"
set label 22 "3" at lblIMRE(3),lblIMIM(3) font "Arial,8"
set label 23 "4" at lblIMRE(4),lblIMIM(4) font "Arial,8"
set label 24 "5" at lblIMRE(5),lblIMIM(5) font "Arial,8"
set label 25 "6" at lblIMRE(6),lblIMIM(6) font "Arial,8"
set label 26 "7" at lblIMRE(7),lblIMIM(7) font "Arial,8"

set label 102 "-0.1" at lblIMRE(-0.1),lblIMIM(-0.1) font "Arial,8"
set label 103 "-0.2" at lblIMRE(-0.2),lblIMIM(-0.2) font "Arial,8"
set label 104 "-0.3" at lblIMRE(-0.3),lblIMIM(-0.3) font "Arial,8"
set label 105 "-0.4" at lblIMRE(-0.4),lblIMIM(-0.4) font "Arial,8"
set label 106 "-0.5" at lblIMRE(-0.5),lblIMIM(-0.5) font "Arial,8"
set label 107 "-0.6" at lblIMRE(-0.6),lblIMIM(-0.6) font "Arial,8"
set label 108 "-0.7" at lblIMRE(-0.7),lblIMIM(-0.7) font "Arial,8"
set label 109 "-0.8" at lblIMRE(-0.8),lblIMIM(-0.8) font "Arial,8"
set label 1010 "-0.9" at lblIMRE(-0.9),lblIMIM(-0.9) font "Arial,8"
set label 1020 "-1" at lblIMRE(-1),lblIMIM(-1) font "Arial,8"
set label 1021 "-2" at lblIMRE(-2),lblIMIM(-2) font "Arial,8"
set label 1022 "-3" at lblIMRE(-3),lblIMIM(-3) font "Arial,8"
set label 1023 "-4" at lblIMRE(-4),lblIMIM(-4) font "Arial,8"
set label 1024 "-5" at lblIMRE(-5),lblIMIM(-5) font "Arial,8"
set label 1025 "-6" at lblIMRE(-6),lblIMIM(-6) font "Arial,8"
set label 1026 "-7" at lblIMRE(-7),lblIMIM(-7) font "Arial,8"

set label 2000 "0" at lblRERE(0),0.025 font "Arial,8"
set label 2001 "0.1" at lblRERE(0.1),0.025 font "Arial,8"
set label 2002 "0.2" at lblRERE(0.2),0.025 font "Arial,8"
set label 2003 "0.3" at lblRERE(0.3),0.025 font "Arial,8"
set label 2004 "0.4" at lblRERE(0.4),0.025 font "Arial,8"
set label 2005 "0.5" at lblRERE(0.5),0.025 font "Arial,8"
set label 2006 "0.6" at lblRERE(0.6),0.025 font "Arial,8"
set label 2007 "0.7" at lblRERE(0.7),-0.025 font "Arial,8"
set label 2008 "0.8" at lblRERE(0.8),0.025 font "Arial,8"
set label 2009 "0.9" at lblRERE(0.9),-0.025 font "Arial,8"
set label 2010 "1" at lblRERE(1),0.025 font "Arial,8"
set label 2011 "2" at lblRERE(2),0.025 font "Arial,8"
set label 2012 "3" at lblRERE(3),0.025 font "Arial,8"
set label 2013 "4" at lblRERE(4),0.025 font "Arial,8"
set label 2014 "5" at lblRERE(5),0.025 font "Arial,8"

set label 3000 "Imag" at -1,0.6 rotate by 45 font "Arial,12"
set label 3001 "Real" at -0.5,0.1 font "Arial,12"


set label 30 "inf" at 1.01,0

a=-1
set multiplot
load 'smithloop.plt'
```

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Die aufzurufende Datei ist „*smith_plot_data.plt*“:

```
reset;
set encoding iso_8859_1
set terminal win;
complex(x,y) = x*{1,0}+y*{0,1}
conformal(z) = (z-1)/(z+1)
pi = 4*atan(1)

#this portion is by Helmut Sennewald (helmutsennewald (at) t-online.de)
load "smithgrid-n.plt"
set style data linespoints
pause 1
plot 'data1.txt' \
 using (real(conformal(complex($1,$2)))):\
 (imag(conformal(complex($1,$2)))) \
 linewidth 2
unset multiplot
unset parametric

pause -1 "Hit return to continue, jetzt auf smith-26.eps"

set output "smith-26.eps"
set terminal postscript eps enhanced color

set parametric
a=-1
load "smithgrid-n.plt"
set style data linespoints
pause 1
plot 'data1.txt' \
 using (real(conformal(complex($1,$2)))):\
 (imag(conformal(complex($1,$2)))) \
 linewidth 2
unset multiplot
unset parametric

pause -1 "Gespeichert"
set output
set terminal windows
```

Die Datendatei „*data1.txt*“ hat schliesslich den Inhalt

```
0, 0
0, 0.5
0.25, 0.1
0.5, -0.1
```

Abbildung G.2 zeigt das Resultat, wenn *smith_plot_dta.plt* mit Gnuplot 4.2 verarbeitet wird.

## G.7. Polar-Plots mit Gnuplot 4.2

Um einen *Polar-Plot* mit Gnuplot 4.2 herzustellen, braucht man zwei Dateien. Die erste, *polargrid.plt* dient als Steuerdatei, während die zweite, *polarloop.plt* die einzelnen Kurven zeichnet.
Der Inhalt der ersten Datei („*polargrid.plt*“) ist:

```
reset;
set encoding iso_8859_1
cd 'Ihr Verzeichnis';
set terminal win;
complex(x,y) = x*{1,0}+y*{0,1}
conformal(z) = (z-1)/(z+1)
pi = 4*atan(1)
lblIMRE(x) = 1.05*real(conformal(complex(0,x)))-0.025
lblIMIM(x) = 1.05*imag(conformal(complex(0,x)))
lblRERE(x) = real(conformal(complex(x,0)))+0.005

set nokey
set size square
set parametric
unset xlabel
unset ylabel
set title "Polardarstellung" font "Arial,24"
unset logscale x
unset logscale y
set xrange[-1.1:1.1]
set yrange[-1.1:1.1]
set samples 1500
```

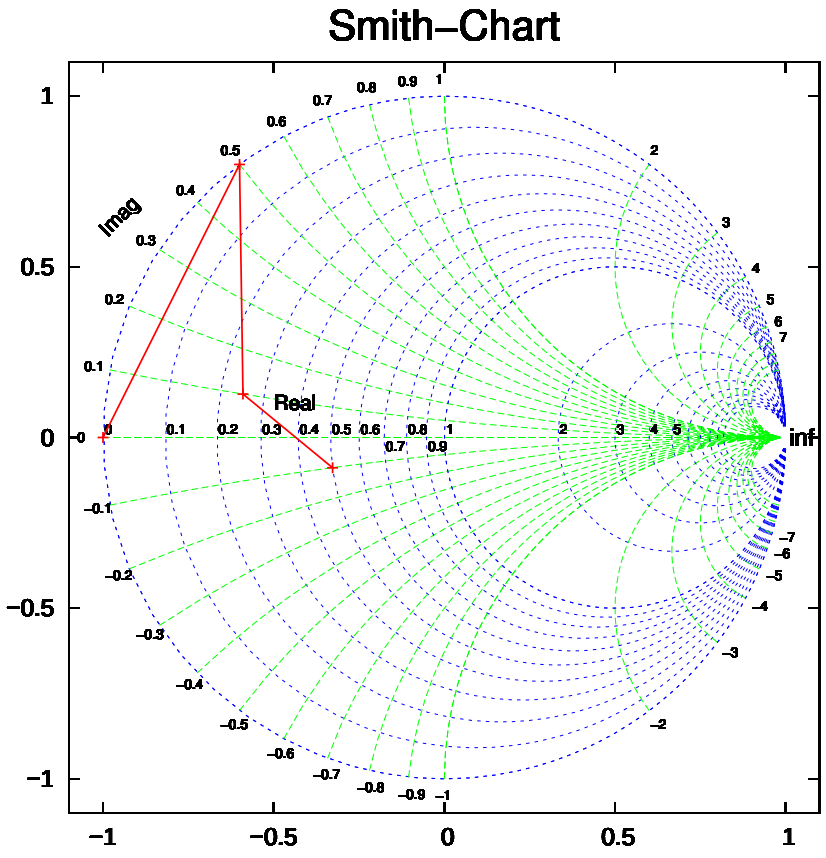



Abbildung G.2.: Plot einer Datendatei *Smith-Chart* mit Gnuplot 4.2

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

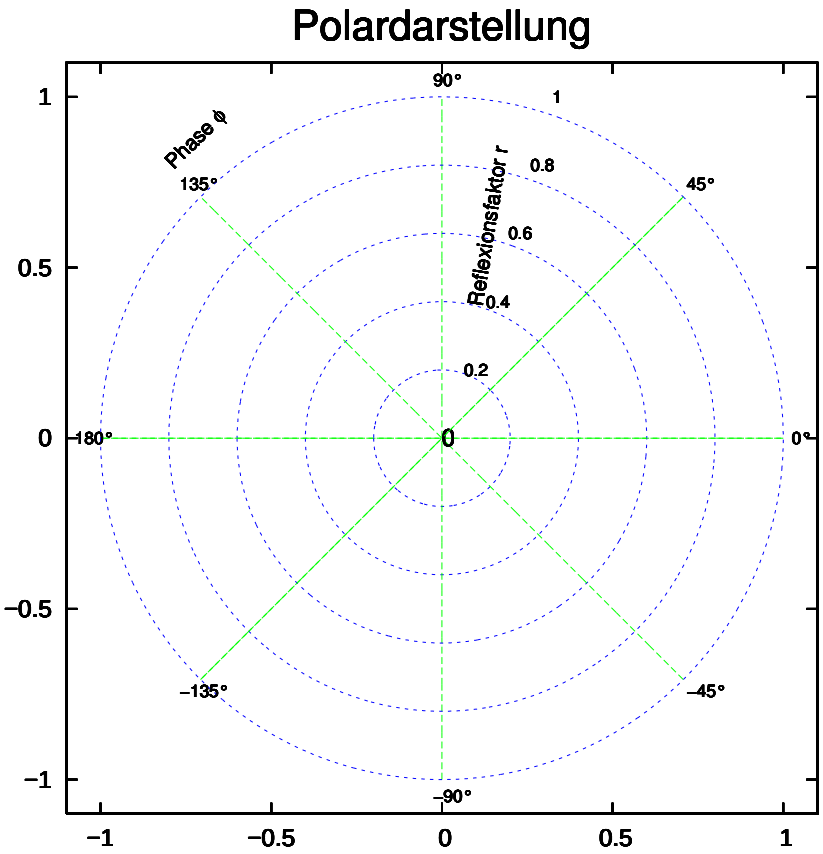



Abbildung G.3.: *Polar-Plot* mit Gnuplot erstellt

```
set label 1 "0°" at 1.05*cos(0*pi)-0.025,1.05*sin(0*pi) font "Arial,10"
set label 2 "45°" at 1.05*cos(0.25*pi)-0.025,1.05*sin(0.25*pi) font "Arial,10"
set label 3 "90°" at 1.05*cos(0.5*pi)-0.025,1.05*sin(0.5*pi) font "Arial,10"
set label 4 "135°" at 1.05*cos(0.75*pi)-0.025,1.05*sin(0.75*pi) font "Arial,10"
set label 5 "180°" at 1.05*cos(1*pi)-0.025,1.05*sin(1*pi) font "Arial,10"
set label 6 "-135°" at 1.05*cos(1.25*pi)-0.025,1.05*sin(1.25*pi) font "Arial,10"
set label 7 "-90°" at 1.05*cos(1.5*pi)-0.025,1.05*sin(1.5*pi) font "Arial,10"
set label 8 "-45°" at 1.05*cos(1.75*pi)-0.025,1.05*sin(1.75*pi) font "Arial,10"

set label 10 "0" at 1.05*0*cos(0.4*pi),1.05*0*sin(0.4*pi) font "Arial 10"
set label 11 "0.2" at 1.05*0.2*cos(0.4*pi),1.05*0.2*sin(0.4*pi) font "Arial,10"
set label 12 "0.4" at 1.05*0.4*cos(0.4*pi),1.05*0.4*sin(0.4*pi) font "Arial,10"
set label 13 "0.6" at 1.05*0.6*cos(0.4*pi),1.05*0.6*sin(0.4*pi) font "Arial,10"
set label 14 "0.8" at 1.05*0.8*cos(0.4*pi),1.05*0.8*sin(0.4*pi) font "Arial,10"
set label 15 "1" at 1.05*1*cos(0.4*pi),1.05*1*sin(0.4*pi) font "Arial,10"

set label 3000 "Phase {/Symbol=12 f}" at -0.8,0.8 rotate by 45 font "Arial,12"
set label 3001 "Reflexionsfaktor r" at 0.4*cos(0.42*pi),0.4*sin(0.42*pi) font "Arial,12" rotate by 0.45*180

a=-1
set multiplot
load 'polarloop.plt'
unset multiplot
unset parametric
```

Die Schleifendatei („*polarloop.plt*“) lautet

```
a = a+1
set trange [-pi:pi]
plot a*0.2*cos(t),a*0.2*sin(t) lt 5

set trange [-1:1]
plot t*cos(a*pi/4),t*sin(a*pi/4) lt 4

if(a<5) reread
```

Abbildung G.3 zeigt das Resultat, wenn *smithgrid.plt* mit Gnuplot 4.2 verarbeitet wird.

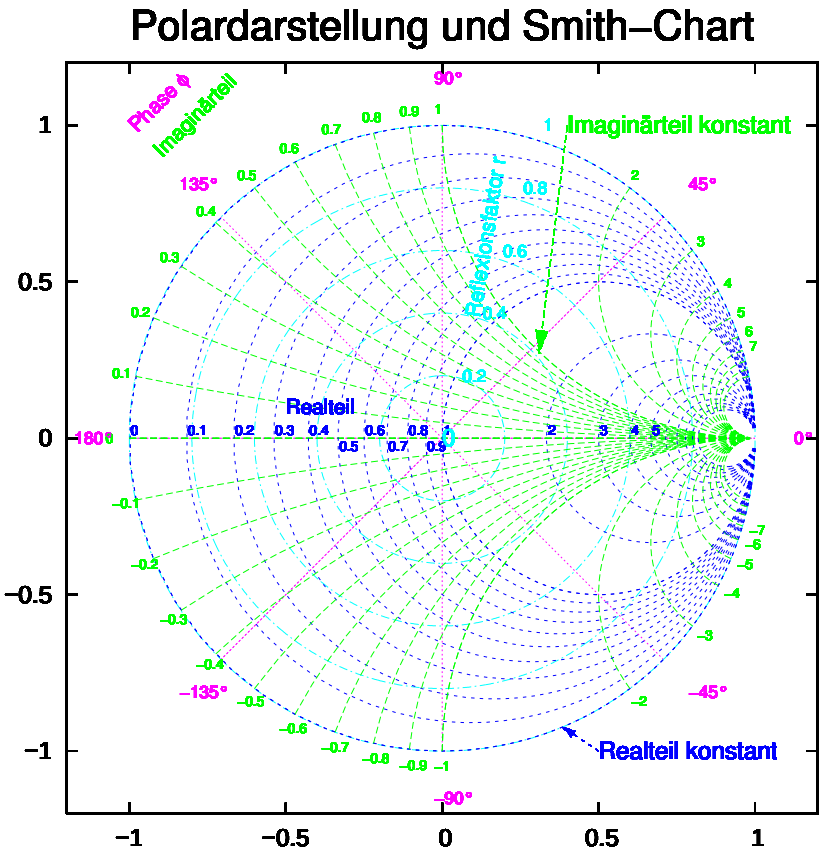



Abbildung G.4.: *Polar-Plot* und *Smith-Chart* in einer mit Gnuplot erstellten Darstellung.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Die beiden Darstellungen können verknüpft werden, wie in Abbildung G.4 gezeigt wird.

# H. Leistungen eines DSPs

| Benchmark Program | Sample Rate (Hz) or Execution Time | Memory Size (Words) | Number of Clock Cycles |
|---|---|---|---|
| 20-Tap FIR Filter | 500.0 kHz | 50 | 54 |
| 64-Tap FIR Filter | 190.1 kHz | 138 | 142 |
| 67-Tap FIR Filter | 182.4 kHz | 144 | 148 |
| 8-Pole Cascaded Canonic Biquad IIR Filter (4 x ) | 540.0 kHz | 40 | 50 |
| 8-Pole Cascaded Canonic Biquad IIR Filter (5x) | 465.5 kHz | 45 | 58 |
| 8-Pole Cascaded Transpose Biquad IIR Filter | 385.7 kHz | 48 | 70 |
| Dot Product | 444.4 ns | 10 | 12 |
| Matrix Multiply 2x2 times 2 x 2 | 1.556 $\mu s$ | 33 | 42 |
| Matrix Multiply 3x3 times 3 x 1 | 1.259 $\mu s$ | 29 | 34 |
| M-to-M FFT 64 Point | 98.33 $\mu s$ | 489 | 2655 |
| M-to-M FFT 256 Point | 489.8 $\mu s$ | 1641 | 13255 |
| M-to-M FFT 1024 Point | 2.453 ms | 6793 | 66240 |
| P-to-M FFT 64 Point | 92.56 $\mu s$ | 704 | 2499 |
| P-to-M FFT 256 Point | 347.9 $\mu s$ | 2048 | 9394 |
| P-to-M FFT 1024 Point | 1.489 ms | 7424 | 40144 |

Tabelle H.1.: Benchmark-Resultate für den 27 MHz-DSP-Prozessor DSP56001R27[**?**]

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

# I. Materialeigenschaften

## I.1. Eigenschaften von Isolationsmaterialien

## I.2. Thermolelektrische Koeffizienten

## I.3. Seebeck-Koeffizienten

## I.4. Debye-Temperatur und Temperaturkoeffizient des Widerstandes

| Material | Volumenwiderstand ($\Omega$cm) | wasserabweisend (hydrophob) | Minimale piezoelektrische Effekte | Minimale triboelektrische Effekte | Minimale dielektrische Absorption |
|---|---|---|---|---|---|
| Saphir | $10^{16} - 10^{18}$ | + | + | 0 | + |
| Teflon PTFE | $> 10^{18}$ | + | - | - | + |
| Polyethylen | $10^{16}$ | 0 | + | 0 | + |
| Polystyrol | $> 10^{16}$ | 0 | 0 | - | + |
| Kel-F | $> 10^{18}$ | + | 0 | - | 0 |
| Keramik | $10^{14} - 10^{15}$ | - | 0 | + | + |
| Nylon | $10^{13} - 10^{14}$ | - | 0 | - | - |
| Glasfaserverstärktes Epoxyharz | $10^{13}$ | - | 0 | - | - |
| PVC | $5 \times 10^{13}$ | + | 0 | 0 | - |
| Glasfaserverstärktes Phenolharz | $10^{13}$ | - | + | + | - |
| | | → | Eigenschaften | ← | ← |

Tabelle I.1.: Materialeigenschaften. Bedeutung der Zeichen: + sehr gut geeignet, - ungeeignet, 0 bescheidene, aber noch brauchbare Eigenschaften. Nach Keithley[**?**]

| Materialkombination | Thermoelektrisches Potential |
|---|---|
| Cu - Cu | $\leq 0.2\mu V/°C$ |
| Cu - Ag | $0.3\mu V/°C$ |
| Cu - Au | $0.3\mu V/°C$ |
| Cu - Pb/Sn | $1 - 3\mu V/°C$ |
| Cu - *Si* | $400\mu V/°C$ |
| Cu - Kovar | $40 - 75\mu V/°C$ |
| Cu - CuO | $1000\mu V/°C$ |

Tabelle I.2.: Thermoelektrische Koeffizienten nach Keithley[**?**]

| Metall | Seebeck-Koeffizient [$\mu$V/K] |
|---|---|
| Sb | 4.7 |
| Fe | 1.7 |
| Cd | 0.8 |
| Cu | 0.7 |
| Ag | 0.65 |
| Pb, Al | 0.4 |
| Hg, Pt | 0 |
| Ni | -1.5 |
| Bi | -7.3 |

Tabelle I.3.: Thermoelektrische Spannungsreihe und Seebeck-Koeffizienten

| $\Theta$ | 293 K | 350 K | 400 K |
|---|---|---|---|
| 50 K | $3.5 \times 10^{-3}$ | $2.91 \times 10^{-3}$ | $2.54 \times 10^{-3}$ |
| 100 K | $3.59 \times 10^{-3}$ | $2.98 \times 10^{-3}$ | $2.59 \times 10^{-3}$ |
| 200 K | $3.79 \times 10^{-3}$ | $3.12 \times 10^{-3}$ | $2.70 \times 10^{-3}$ |
| 400 K | $4.26 \times 10^{-3}$ | $3.42 \times 10^{-3}$ | $2.92 \times 10^{-3}$ |

Tabelle I.4.: Temperaturkoeffizient eines Widerstandes als Funktion der Temperatur und der Debye-Temperatur nach Debye.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

# J. Image Processing: an Introduction

The visualization and interpretation of images from Scanning Probe Microscopes is intimately connected to the processing of these images. This chapter discusses reasons for using image processing algorithms in order to remove distortions, filtering in fourier space and in real space. Finally we will show some methods to display data.

## J.1. Why Image Processing?

An ideal Scanning Probe Microscope is a noise free device that images a sample with perfect tips of known shape and has perfect linear scanning piezos. In reality, SXMs are not that ideal. The scanning device in SXM is affected by distortions. To do quantitative measurements like determining the unit cell size, these distortions have to be measured on test substances and have to be corrected for. The distortions are both linear and nonlinear. Linear distortions mainly result from imperfections in the machining of the piezo translators causing cross talk from the $z$-piezo to the $x$- and $y$-piezos and vice versa. Among the linear distortions there are two kinds which are very important: first, scanning piezos invariably have different sensitivities along the different scan axis due to the variation of the piezo material and uneven sizes of the electrode areas. Second, the same reasons might cause the scanning axis not be orthogonal. Furthermore the plane in which the piezo scanner moves for constant $z$ is hardly ever coincident with the sample plane. Hence a linear ramp is added to the sample data. This ramp is especially bothersome, when the height $z$ is displayed as an intensity map, also called top view display.
The nonlinear distortions are harder to deal with[**?**]. They can affect SXMs from a variety of reasons. First piezoelectric ceramics do have a hysteresis loop, much like ferromagnetic materials. The deviations of piezo ceramic materials from linearity increase with increasing amplitude of the driving voltage. The mechanical position for one voltage depends on the voltages applied to the piezo before. Hence to get the best position accuracy one should approach a point on the sample always from the same direction.
Another type of nonlinear distortion of the images occurs, when the scan frequency approaches the upper frequency limit of the $x$- and $y$-drive amplifiers or the upper frequency limit of the feedback loop ($z$-component). The distortion due to the feedback loop can only be minimized by reducing the scan frequency.
On the other hand there is a simple way to reduce distortions due to the $x$- and $y$-piezo drive amplifiers. To keep the system as simple as possible one uses normally a triangular wave form for driving the scanning piezos. However, triangular waves contain frequency components at multiples of the scan frequency. If the cut-off

frequency of the $x$- and $y$-drive electronics or of the feedback loop is too close to the scanning frequency (2 to 3 times the scanning frequency) the triangular drive voltage is rounded of at the turning points. This rounding error causes first a distortion of the scan linearity and second, through phase lags, the projection of part of the backward scan onto the forward scan. This type of distortion can be minimized by carefully selecting the scanning frequency and by using driving voltages for the $x$- and $y$-piezos with wave forms like trapezoidal waves which are closer to a sine wave.
The values measured for $x$, $y$ or $z$ are affected by noise. The origin of this noise can be either electronically, be disturbances through sound or be a property of the sample surface due to adsorbates. In addition to this incoherent noise, interference with mains and other equipment nearby might be present. Depending on the type of noise, one can filter it in the real space or in the fourier space.

## J.2. Correcting Distorted Images

To improve the usefulness of the SXM-data for measurements of distances and to enhance the visual appearance, the linear and nonlinear distortions have to be corrected. Normally, they will have well defined physical origins and can be determined by independent methods. A common linear correction is the removal of a background plane by fitting a plane to the data. A mathematical formulation of this background subtraction can be found in section J.6 in this paper.
Another common distortion removal is the correction for non orthogonal piezo scanners or for piezo scanners with an unequal sensitivity. An independent measurement of the distortion is usually required to get physically meaningful results. One excellent possibility is to correct the nonorthogonalities beforehand using an electronic matrix to mix the drive voltages of all three channels. Another method incorporates feedback control for all the movements of the piezo tube to reduce the nonlinearity[**?**].
If one does not have access to such electronics it is possible to use a measurement with the same piezo scanners on a test substance like graphite, a CD-disk stamper or an optical grating to determine the correction factors. Using this technique, it is only possible to get a measure of the distortion in the $xy$-plane. The mathematics of the distortion removal in the sample plane and in all three dimensions is rather complicated and treated in section K.
Nonlinear distortions pose far bigger problems to correct than the linear distortions. As said before, nonlinear distortions may stem from an insufficient analog bandwidth in the feedback system or from piezo hysteresis or creep. Data hampered by insufficient bandwidth could be corrected by fourier filtering and deconvolution in one dimension[**?**, **?**]. However both the forward and the backward scan must be recorded at equal time intervals. The data taken during the backward scan serves to determine the history of the feedback loop at the first few points of the forward scan.
The hysteretic behavior of the piezo ceramic is most annoying at large scan ranges. A test measurement could determine a look-up table, with which the measured data could be translated into undistorted data. The look-up-tables should be acquired at a few scan sizes. Intermediate scan sizes could then be interpolated. A very efficient method to linearize large scan ranges is based on the measurement

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

of the actual piezo deflection[?].

## J.3. Filtering and Data Analysis in Real Space

Real space filters are filters whose result for a point only depends on a few neighboring points. For large data sets they are more efficiently implemented than the corresponding filters in the spatial frequency domain (see section J.4). One of the most often occurring problems is to remove high frequency noise from data. The origin of this noise can be electronically, come from adsorbates on the sample surface or be digitizing noise. This high frequency noise can be reduced by replacing each point by the weighed average of its neighboring points. If we only consider the nearest neighbors, that is the points $z(x, y)$ with $x = [x_0 + 1|x_0|x_0 - 1]$ and $y = [y_0 + 1|y_0|y_0 - 1]$ then we can define the 3x3 convolution low pass filter by

| | | |
|---|---|---|
| $w_{-1,1}$ | $w_{0,1}$ | $w_{1,1}$ |
| $w_{-1,0}$ | $w_{0,0}$ | $w_{1,0}$ |
| $w_{-1,-1}$ | $w_{0,-1}$ | $w_{1,-1}$ |

The value at a point $x_0$, $y_0$ is the given by

$$z'(x,y) = \frac{\sum_{i=-1}^{1}\sum_{j=-1}^{1} w_{ij} z(x+i, y+j)}{\sum_{i=-1}^{1}\sum_{j=-1}^{1} w_{ij}} \tag{J.3.1}$$

It should be noted that if the denominator in equation (J.3.1) is zero, that we should use the modified equation

$$z'(x,y) = \sum_{i=-1}^{1}\sum_{j=-1}^{1} w_{ij} z(x+i, y+j) \tag{J.3.2}$$

By setting all the $w_{ij}$ to 1 we obtain the convolution averaging filter. Figure J.1a) shows the original, noisy data. Figure J.1b) is filtered by the convolution averaging filter. We can construct directional averaging filters, by setting selected $w_{ij}$ to 1 and the others to 0. For instance setting the $w_{ij=0}$ to 1 and all the other coefficients to 0, we obtain a filter along the x- axis. The coefficient need not necessarily be equal 1 or 0. By increasing the center value one can emphasize the value of the point to be filtered compared to its neighborhood. The reader can easily verify, that the filters in table J.1 do the respective tasks. Figure J.1c) shows the effect of the Laplacian filter and J.1d) the effect of a directional gradient filter.
The convolution averaging filter is well suited for removing high frequency, random noise. It has, however, the disadvantage that it also blurs steps and other well defined variations from one pixel to the other. A better filter to remove single pixel impulses is the median filter. We do consider the same points in the 3x3 neighborhood as for the convolution averaging filter. But this time we do order the points and we take the middle value, the median. Figure J.1e) finally shows the noise removal by the median filter. Park and Quate[?] give a summary of digital filtering procedures.

| Name | Filter |
|---|---|
| Low Pass Filter | 1 1 1<br>1 1 1<br>1 1 1 |
| Low Pass Filter in horizontal direction | 0 0 0<br>1 1 1<br>0 0 0 |
| Low Pass Filter in vertical direction | 0 1 0<br>0 1 0<br>0 1 0 |
| Laplacian Filter | 1 1 1<br>1 -n 1<br>1 1 1 |
| Edge detection (diagonal) | 0 1 0<br>1 -0 0<br>0 0 -2 |

Tabelle J.1.: 3x3 convolution kernels

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

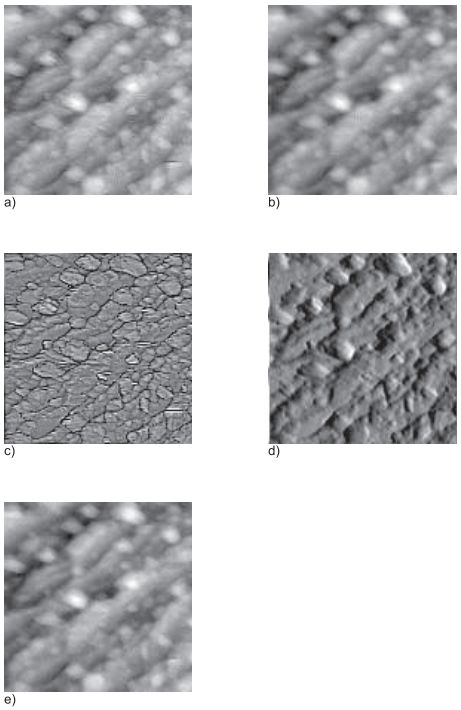



Abbildung J.1.: The effect of different real space filters. a) the original data; b) the convolution averaged data, c) Laplacian filter, d) directional gradient filter, and e) filtered by the median filter.

## J.4. Filtering and Data Analysis in the Spatial Frequency Domain

Filtering in the spatial frequency domain (fourier space) is a very powerful tool. All periodic surface data and many noise components like mains interference have well defined spatial frequencies, which show up as peaks in the fourier transformed data:

$$F(k_x, k_y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{i(xk_x + yk_y)} dx dy \tag{J.4.1}$$

is called the Fourier transform of $f(x, y)$. The inverse transform is given by

$$f(x, y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(k_x, k_y) e^{-i(xk_x + yk_y)} dk_x dk_y. \tag{J.4.2}$$

These two transforms are the basis of fourier filtering of images. An introduction to the topics of fourier transforms on computers and on filters can be found in Press *et al.*[**?**]. On computers, one uses normally the Fast Fourier Transform (FFT), an algorithm discovered by Danielson and Lanczos in 1942 and then rediscovered by Cooley and Tukey in the mid-1960s.
Fourier filtering and data analysis is especially powerful on periodic sample structures or on coherent noise. It is useful to display the fourier transformed data either as a power spectrum and a phase spectrum or as the real and imaginary part of the spectrum. The fast fourier transform maps the data into k-space such that $k_x$ and $k_y$ both run from 0 to $k_{max}$. The spectrum for $k > k_{max}/2$ is the mirror image of the spectrum for $k < k_{max}/2$. Zero spatial frequency is therefore at all

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

the four corners of the fourier transformed data. Symmetries are not obviously in this representation. It is advantageous to move zero spatial frequency to the center of the display. This can be done by swapping the area in the way represented in figure J.2. The fourier spectrum of surfaces in this representation is similar to the displays one gets from Low Energy Electron Diffraction (LEED)[**?**].

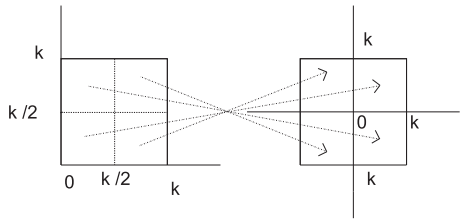



Abbildung J.2.: Fourier display: areas to be swapped to get a display like LEED displays. After the FFT the data points corresponding to low frequencies are located at the four edges of the data set. Fourier displays with 0 frequency at the center provide easy to get information on the symmetries of the surfaces.

Performing a filter, convolution, or deconvolution in the $\boldsymbol{k}$ space requires special attention: the FFT algorithm is based on periodic functions with the maximum period being the size of the data, or integral fractions thereof. Measured data will contain other frequency components which are of noninteger relation to the basic period of the FFT and which were truncated by the sampling process. Modifying the spectrum by a filter or convolution can introduce artifacts. Press *et al.*[**?**] describe the use of data windowing or padding to minimize the unwanted content in the spectrum. If these procedures are not followed, meaningless data might be created.
An important filter in the $\boldsymbol{k}$ space is the Wiener filter. It is assumed, that the scanning probe microscope has a response function $r(x, y)$ and a noise function $n(x, y)$. The real data $u(x, y)$ is first smeared out by $r(x, y)$ to

$$s(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} r(\hat{x}, \hat{y}) u(x - \hat{x}, y - \hat{y}) d\hat{x} d\hat{y} \tag{J.4.3}$$

The noise $n(x, y)$ is added to $s(x, y)$ to give

$$c(x, y) = s(x, y) + n(x, y) \tag{J.4.4}$$

The Wiener filter $\Psi(k_x, k_y)$ tries to reconstruct the original data $u(x, y)$ by taking into account the effect of the noise. The reconstructed spectrum is

$$\tilde{U}(k_x, k_y) = \frac{C(k_x, k_y)\Psi(k_x, k_y)}{R(k_x, k_y)} \tag{J.4.5}$$

where $C(k_x, k_y)$ and $R(k_x, k_y)$ are the fourier transforms of $c(x, y)$ and $r(x, y)$, respectively. The exact tip shape and the relevant interactions between the tip and the sample are not well known. Hence the assumption of a known response function $r(x, y)$ is usually not fulfilled in atomic resolution scanning probe microscopy. Therefore, one can not hope to deconvolute such data using a Wiener filter. A noise reduction, however, is all the same possible. For large scans (in the $\mu$m range), the tip shape can usually be determined by Scanning Electron Microscopy

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

or is known from the fabrication process (micro-fabricated cantilevers for scanning force microscopy) and the interaction details are of no concern on those length scales. In this situation, a successful noise reduction and deconvolution can be possible.
The filter function of the Wiener filter is

$$\Psi(k_x, k_y) = \frac{|S(k_x, k_y)|^2}{|S(k_x, k_y)|^2 + |N(k_x, k_y)|^2} \tag{J.4.6}$$

$\Psi(k_x, k_y)$ is determined by the power spectrum of the smeared data $s(x, y)$ and by the power spectrum of the noise function $n(x, y)$. The spectrum of the measured function $c(x, y)$ does not enter in the calculation of the filter function. One way to get the additional information is to guess the noise spectrum from suitable plots of the spectrum of $c(x, y)$. Another way is to record images with the scanning motion disabled. This produces the true noise spectrum, if there are no position dependent noise components[**?**, **?**, **?**].

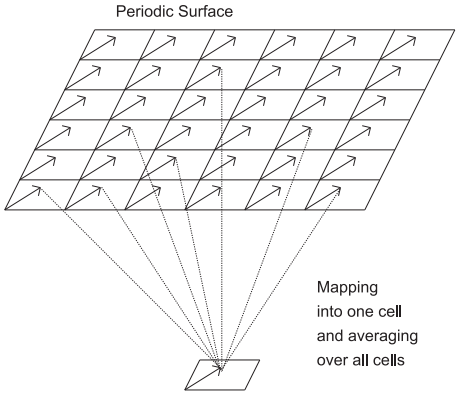



Abbildung J.3.: Unit Cell Filter. This figure depicts how an arbitrary point on a periodic surface is mapped to the unit cell.

Fourier transform filters are very powerful for periodic data. The computational effort, however, increases sharply with the number of data points. The periodicity of the data can be used to define a unit cell, which is repeated all over the surface. Any point at a specific location $x, y$ in this unit cell must have the same z-value as the corresponding points in the other unit cells. Figure J.3 is a sketch of a periodic surface. We can map all the points to one unit cell and average over them. The data in the one unit cell has a reduced noise background, since the coherent data, the structure of the unit cell, is passed unchanged, whereas the statistical noise, incoherent with the unit cell, is reduced by $\sqrt{n}$, where $n$ is the number of unit cells averaged over. The last step in this filter is the repetition of the filtered data over the original area. Correlation averaging filters in conjunction with scanning probe microscopy have been described by Soethout *et al.*[**?**].
Table J.2 is a summary of the various filter methods. It also gives hints, when to use which filter.

| Problem | Filter | Comments |
|---|---|---|
| Uneven Height | Background Subtraction | Fast, Reversible |
| Random Noise | Convolution Low Pass | Fast Processing, limited action |
| Interference with a fixed frequency | Fourier Filtering | fast for small data sets, Time consuming otherwise |
| Interference with a fixed frequency | Unit Cell filter | faster for large data sets, no complete suppression of the interference |

Tabelle J.2.: Usage of filters

## J.5. Viewing the Data

The most important part of image processing is to visualize the measured data. Typical SXM data sets can consist of many thousands to over a million points per plane. There may be more than one image plane present. The STM data represents a topography. The output of the first STMs was recorded on $xy$-chart recorders. Usually, the $z$-value or the height of the tip was plotted against the tip position in the fast scan direction. Often the position in the slow scan direction was not recorded, but assumed to be constant. A ramp added to the $y$-channel of the chart recorder helped to separate the scan lines. More sophisticated display systems added a fraction of the tip position in the slow scan direction to both the x and y- channels of the chart recorder. This way, a pseudo three- dimensional display was achieved. Figure J.4 shows the display of a sample surface using this technique. Chart recorders are slow devices, so people started using analog storage oscilloscopes, displaying the same line scan plots.

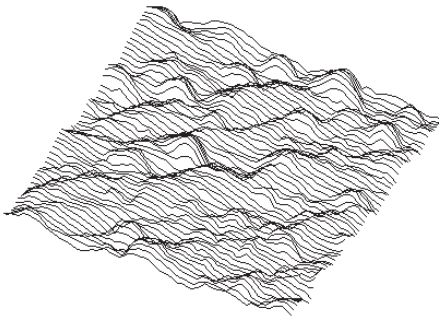


Abbildung J.4.: A typical example of an SXM output using chart recorders. The data displayed is a silicon surface with evaporated indium. The chart recorder is set up such that the horizontal axis display $x$ and a fraction of $y$, the vertical axis $z$ and a fraction of $y$.

A wire mesh display similar to the line scan display can be created on computer displays (Figure J.5). It is especially suitable for monochrome display systems with only two colors. The number of scan lines which can be displayed is usually well below one hundred and the display resolution along the fast scanning axis $x$ is

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

much better than along $y$.

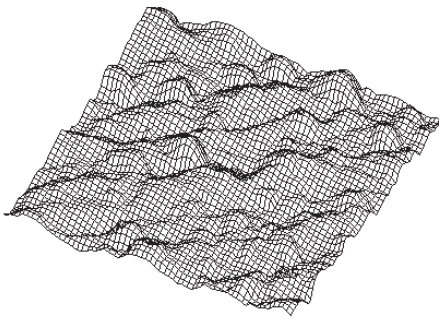


Abbildung J.5.: Wire mesh display of the data in figure J.4. The wire mesh display gives a quick look at the surface.

If the computer display is capable to display at least 64 shades of gray, then top view images can be created (Figure J.6). In these images, the position on the screen corresponds to the position on the sample and the height is coded as a shade of gray. Usually the convention is that the brighter a point, the higher it is. The number of points which can be displayed is only limited by the number of pixels available. This view of the data is excellent for measuring distances between surface features. Periodic structures show up particularly well on such a top view. The human eye is not capable to distinguish more than 64 shades of gray. If average $z$-height of the tip varies from one side of the image to the other, then the interesting features usually have too little contrast. Hence contrast equalization is needed. For data being affected by a large background slope, it is often possible to still detect some features in the line scan view. Some researchers prefer a simultaneous display of both line scan images and top view images to get the most information in the shortest time. Top views use much less calculation time than line scan images. Hence computerized fast data acquisition systems usually display the data as a top view first.

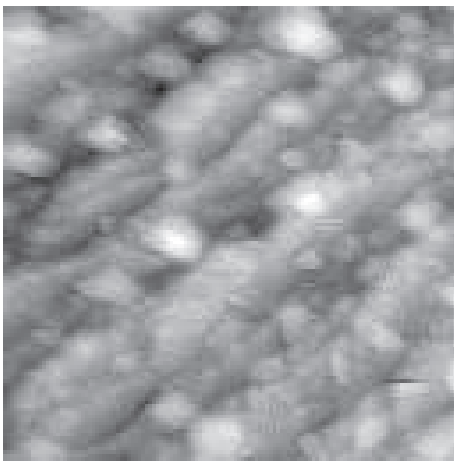


Abbildung J.6.: Top view display from a computer screen of the data in figure J.4. The top view display is the workhorse of the SXM data display methods. It allows a convenient judging of heights and sizes.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

The display can be made more illustrative by calculating the illuminated top view of the data, much like the way topographic maps are shaded. Figure J.7a) gives an example of a sample surface illuminated by a point light source at infinity. This technique is a powerful tool to enhance the appearance of a data set. But it can be abused! Changing the direction of the light source, as shown in J.7b), can obscure some undesired features. The effect of the illumination is similar to displaying the magnitude of the gradient of the sample surface along the direction to the light source. Features perpendicular to the illumination can not be seen. Multiple light sources or extended light sources diminish this effect, but the illumination is much more complicated to calculate. If not shown in conjunction with some other display method one is not able to judge the validity of such an image.

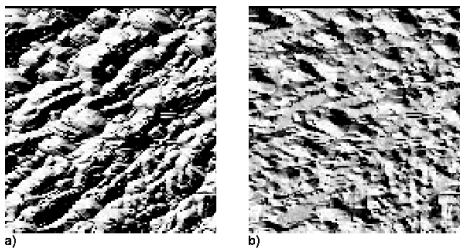



Abbildung J.7.: The same test surface as in figures J.4, J.5 and J.6 is illuminated from two different directions. Part a) is illuminated such that the structure seen in the previous figures is reproduced. Part b) shows that by changing the illumination direction information may be hidden. Data published with illumination, either flat or 3D-rendered, may have to be taken with a grain of salt.

One can combine top views or illuminated top views and wire mesh scan displays to form solid surface models of the sample surface. Such images are usually only generated in the final processing stage before publication because they need quite a lot of computing time. Figure J.8a) shows a combination of the top view display and the wire mesh display, a three dimensional model where the height is coded as a shade of gray. Figure J.8b) shows a combination of the illuminated top view and the wire mesh, a display much like a real landscape under the sun. Depending on the point of view, some features might be more prominent than others.

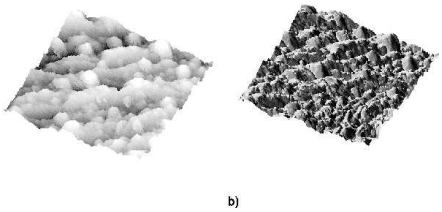



Abbildung J.8.: Two possible ways of a three dimensional surface rendering of the surface of figure J.4. In part a) the shade of gray is determined by the height of the data, as in figure J.6. Part b) show an illuminated 3D-rendering.

Additional information can be packed into an image by using color. Assume that an image has two planes of data. We can display the first plane with shades of

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

green and the second one with shades of red on top of each other. Where the magnitude of both planes is high, one gets an orange color, where both are low, one gets black. But if the magnitude of one plane is larger than that of the other plane on one pixel, one gets red or green colors. This way, one can display the registry of two different quantities in the same image.

## J.6. Background Plane Removal

We assume that the image affected by a background plane has n rows with m data points each.

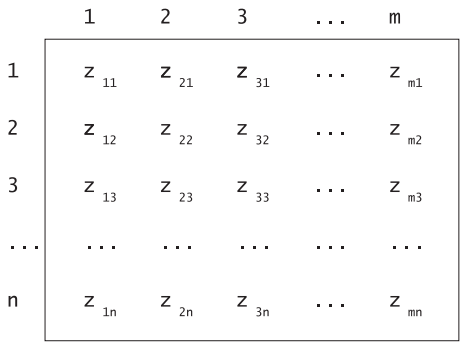



Abbildung J.9.: The labeling of the points in a data set for the background correction.

The background plane is defined by solving

$$ai + bj + c = z_{ij} \tag{J.6.1}$$

in a least squares approximation. $i$ and $j$ are the indices for the points and run from 1 to $m$ and 1 to $n$, respectively (See figure J.9). The $n \times m$ equations for all points can be combined to form the matrix equation

$$\mathbf{A}\boldsymbol{a} = \boldsymbol{z} \tag{J.6.2}$$

where

$$\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 1 & 1 \\ \vdots & \vdots & \vdots \\ m & 1 & 1 \\ 1 & 2 & 1 \\ 2 & 2 & 1 \\ \vdots & \vdots & \vdots \\ m & 2 & 1 \\ \vdots & \vdots & \vdots \\ m & n & 1 \end{pmatrix}, \qquad \boldsymbol{a} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \qquad \text{and} \qquad \boldsymbol{z} = \begin{pmatrix} z_{11} \\ z_{21} \\ \vdots \\ z_{m1} \\ z_{12} \\ z_{22} \\ \vdots \\ z_{m2} \\ \vdots \\ z_{mn} \end{pmatrix}$$

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

The overdetermined system of equations (J.6.2) can be solved in terms of a least square fit by multiplying it from the left by$\mathbf{A}^T$.

$$\mathbf{A}^T\mathbf{A}\boldsymbol{a} = \mathbf{A}^T\boldsymbol{z} \tag{J.6.3}$$

Written in components equation (J.6.3) becomes

$$\begin{pmatrix} n\sum_{i=1}^{m} i^2 & \sum_{i=1}^{m}\sum_{j=1}^{n} ij & n\sum_{i=1}^{m} i \\ \sum_{j=1}^{n} ij & m\sum_{j=1}^{n} j^2 & m\sum_{j=1}^{n} \\ n\sum_{i=1}^{m} i & m\sum_{j=1}^{n} & n^2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \tag{J.6.4}$$

$$\begin{pmatrix} n\frac{m(m+1)(2m+1)}{6} & \frac{m(m+1)n(n+1)}{4} & n\frac{m(m+1)}{2} \\ \frac{m(m+1)n(n+1)}{4} & m\frac{n(n+1)(2n+1)}{6} & m\frac{n(n+1)}{2} \\ n\frac{m(m+1)}{2} & m\frac{n(n+1)}{2} & n^2 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{m}\sum_{j=1}^{n} iz_{ij} \\ \sum_{i=1}^{m}\sum_{j=1}^{n} jz_{ij} \\ \sum_{i=1}^{m}\sum_{j=1}^{n} z_{ij} \end{pmatrix}$$

This system of three coupled equations can be solved using any standard numerical method[**?**] like the Gauss algorithm.

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

# K. Correction of Linear Distortions in Two and Three Dimensions

We first discuss how to remove distortions in a plane. Figure K.1 shows the coordinate system we use. The Cartesian coordinate system defined by the unit vectors $\hat{e}_x$ and $\hat{e}_y$ is the real undistorted coordinate system. The piezo moves in a system defined by $\hat{e}'_x$ and $\hat{e}'_y$, which is not necessarily orthogonal and in which the two unit vectors, seen from the unprimed system, do not have the same length.

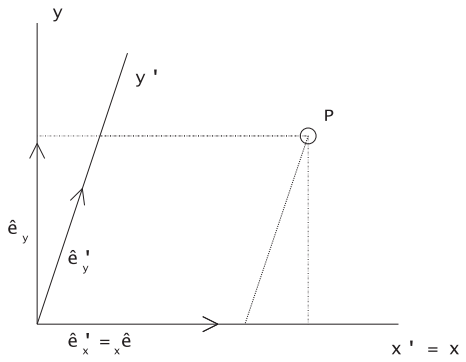



Abbildung K.1.: The coordinate system used for the linear distortion removal.

Without loss of generality we can assume that $\hat{e}_x = \hat{e}'_x$. $\hat{e}'_y$ is defined by the relation $\hat{e}'_y = a\hat{e}_x + b\hat{e}_y$. We now take a point $P(x, y)$ in the unprimed system and $P(x', y')$ in the piezos system and calculate a relation between the $(x, y)$ and the $(x', y')$.

$$x\hat{e}_x + y\hat{e}_y = x'\hat{e}'_x + y'\hat{e}'_y = x'\hat{e}_x + y'(a\hat{e}_x + b\hat{e}_y) \tag{K.0.1}$$

By equating the components we obtain

$$\begin{aligned} x &= x' + ay' \\ y &= by' \end{aligned} \tag{K.0.2}$$

This relation allows us to remove the distortion in the image. For computational purposes, however, it is not ideal. To get to the undistorted image, we would have to interpolate both the $x$- and the $y$-components. By defining the basis vectors differently, we can deduce a much more elegant equation in terms of computational efficiency. We set $\hat{e}'_x = a\hat{e}_x$ and $\hat{e}_{y'} = b\hat{e}_x + \hat{e}_y$. We then obtain

$$\begin{aligned} x &= ax' + by' \\ y &= y' \end{aligned} \tag{K.0.3}$$

Data is usually stored in computers in row order form, one row after the other. If we label the position of the points within a row by $x$ and number the rows by $y$, then we can treat each row separately. We can in addition speed up the process by noting that for every row y the same values ax' will occur. These values can be calculated once and can be stored in an array. The other term of the sum, $by' = by$ changes only from row to row. Hence the total computational effort per point is an array look-up and an addition, compared to an addition and an interpolation consisting of two multiplications and one addition in the first case. The three-dimensional correction is analogous.

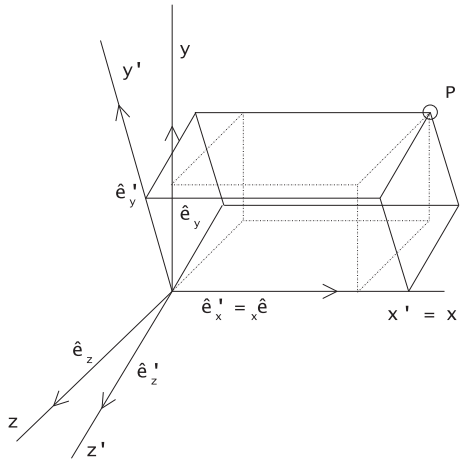



Abbildung K.2.: The coordinate system used for three-dimensional distortion removal.

Figure K.2 gives the definition of the coordinate system. This time we make the two basis vectors $\hat{e}_z$ and $\hat{e}_{z'}$ collinear. $\hat{e}_x$ and $\hat{e}_{x'}$ form a plane whose normal direction is parallel to $\hat{e}_z$. And as a last restriction we set the $\hat{e}_y$-component of $\hat{e}_{y'}$ to 1. Doing the same calculations as in the two dimensional case, we obtain

$$\begin{aligned} x &= Ax' + By' \\ y &= y' \\ z &= Cx' + Dy' + Ez' \end{aligned} \tag{K.0.4}$$

The computational effort for the xy plane is the same as for the two-dimensional case. In addition we have two array look-ups ($Cx'$ and $Dy'$), two additions and one multiplication. These last three equations allow a complete removal of any linear distortion. We have to warn the reader, that the coefficients $A$ and $B$ or $A$-$E$ have to be determined by independent means, because with some clever choice, one can for instance change a hexagonal pattern to a cubic one. Measurements of angles and distances in surface unit cells would not be reliable.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

# L. Beschreibung periodischer Oberflächen

Die Struktur der Oberfläche ist eine, wenn auch modifizierte Fortsetzung des 3-dimensionalen Kristalls. Die mathematische Beschreibung[**?**] lehnt sich an die Konventionen der Indizierung von Volumengittern an.

## L.1. Mathematische Beschreibung

Oberflächen werden durch Millersche Indizes $(hkl)$ beschrieben. Abbildung L.1 zeigt ein Beispiel einer Volumennetzebene. Die Millerschen Indizes werden wie folgt bestimmt:

1. Schnittpunkte der Ebene mit den Achsen $a$, $b$, $c$, bestimmen. Hier sind das $3a$, $6b$, $2c$.
2. Kehrwerte bilden unter Weglassung der des Faktors der jeweiligen Einheitslänge $a$, $b$ oder $c$. Hier erhält man $\left(\frac{1}{3}\ \frac{1}{6}\ \frac{1}{2}\right)$
3. Ganze Zahlen bilden. Hier muss $\times 6$ gerechnet werden. Die Millerschen Indizes sind dann $(2, 1, 3)$

Die in Abbildung L.1 eingezeichnete Ebene ist die $(2, 1, 3)$-Ebene. Negative Indizes werden mit $\bar{2} = -2$ angegeben.
Periodische Strukturen auf einer Fläche werden durch Bravais-Netze beschreiben. (Diese sind analog zu den Bravais-Gittern in drei Dimensionen). Der Ort einer Zelle im Netz einer periodischen Oberfläche ist durch

$$\boldsymbol{r} = m_1\boldsymbol{a}_1 + m_2\boldsymbol{a}_2 \tag{L.1.1}$$

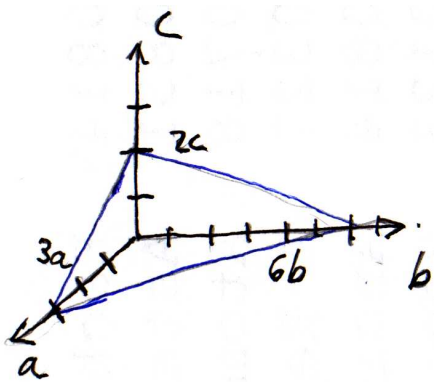



Abbildung L.1.: Skizze zur Bestimmung der Miller'schen Indizes

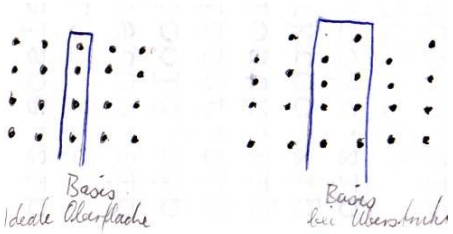
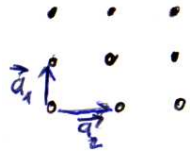



Abbildung L.2.: Basisatome einer Netzzelle eines Oberflächennetzes

gegeben. Die Vektoren $\boldsymbol{a}_1$ und $\boldsymbol{a}_2$ heissen Basisvektoren. Die Zuordnung zu den Indizes "1" und "2" wird durch zwei Konventionen bestimmt. Erstens muss $|a_1| < |a_2|$ sein und zweitens muss $\gamma = \angle(\boldsymbol{a}_1, \boldsymbol{a}_2) \geq 90°$ sein. Weiter sollte $\gamma - 90°$ minimal sein.

## L.1.1. Bravais-Netze

Die folgenden Bravais-Netze werden an Oberflächen unterschieden:

| | | |
|---|---|---|
| 1) quadratisch |  | $\|a_1\| = \|a_2\|$<br>$\gamma = 90°$ |
| 2) rechteckig | | $\|a_1\| < \|a_2\|$<br>$\gamma = 90°$ |
| 3) schiefwinklig | | $\|\boldsymbol{a}_1\| < \|\boldsymbol{a}_2\|$<br>$\gamma > 90°$ |
| 4) hexagonal | | $\|\boldsymbol{a}_1\| = \|\boldsymbol{a}_2\|$<br>$\gamma = 120°$<br>2 Atome in der Einheitszelle |
| 5) rechtwinklig, zentriert (nicht primitiv) | | $\|\boldsymbol{a}_1\| < \|\boldsymbol{a}_2\|$<br>$\gamma = 90°$<br>2 Atome in der Einheitszelle |

Beim schiefwinkligen Gitter (Nummer 3) ist grün eine zweite mögliche Wahl der Einheitsvektoren eingezeichnet. Dabei ist jedoch der Winkel $\gamma$ kleiner als 90°. Deshalb ist diese Wahl der Einheitsvektoren falsch.

Bei den Volumengittern werden aus den Bravais-Gittern die Raumgruppen, indem man die Motive der Einheitszellen betrachtet. Diese können zum Beispiel eine andere Symmetrie als das zugrunde liegende Gitter haben.Bei den Oberflächennetzen führt die Berücksichtigung der Motive in den Einheitszellen zu 17 Flächengruppen.

Anders als bei Netzen in einer Ebene müssen bei Oberflächennetzen alle unter einer Oberflächenzelle liegenden Atome als Basis berücksichtigt werden. Die Basisatome liegen also nicht in einer Ebene. Abbildung L.2 zeigt ein Beispiel.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

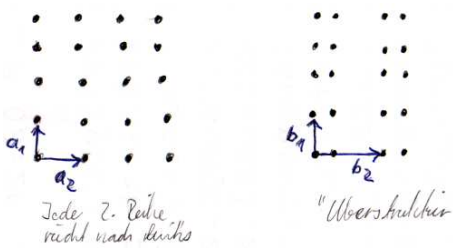



Abbildung L.3.: Atomare Anordnung bei Überstrukturen

## L.1.2. Überstrukturen, Rekonstruktionen

Als Grundlage zur Beschreibung von Oberflächen dient die hypothetische Oberfläche, die beim Entzweischneiden eines Kristalls entsteht. Die Oberflächenstruktur wird bezogen auf diese hypothetische Oberfläche spezifiziert.
Abbildung L.3 zeigt als Beispiel wie eine Überstruktur entsteht. Wenn man in der quadratischen Anordnung auf der linken Seite jeweils jede 2. Reihe nach links rückt, erhältm man die auf der rechten Seite gezeigte Überstruktur. Wenn wir die das Netz der Überstruktur aufspannenden Basisvektoren $\boldsymbol{b}_1$ und $\boldsymbol{b}_2$ nennen, dann können wir schreiben

$$\boldsymbol{b}_1 = \boldsymbol{a}_1 \tag{L.1.2}$$
$$\boldsymbol{b}_2 = 2\boldsymbol{a}_2 \tag{L.1.3}$$

Die Überstruktur in der Abbildung L.3 rechts ist demnach eine $(1 \times 2)$- Überstruktur oder auch eine $(1 \times 2)$-Rekonstruktion.
Im allgemeinen sind die Vektoren $\boldsymbol{a}_1$ und $\boldsymbol{b}_1$ sowie $\boldsymbol{a}_2$ und $\boldsymbol{b}_2$ nicht parallel. Der Winkel $\angle(\boldsymbol{a}_1, \boldsymbol{a}_2)$ kann vom Winkel $\angle(\boldsymbol{b}_1, \boldsymbol{b}_2)$ verschieden sein.
Gilt $\angle(\boldsymbol{a}_1, \boldsymbol{a}_2) = \angle(\boldsymbol{b}_1, \boldsymbol{b}_2)$ so gilt für die Beschreibung der Überstruktur folgendes Rezept:

1. Man bildet $\frac{b_1}{a_1}$ und $\frac{b_2}{a_2}$ und bestimmt den Winkel $\gamma$ zwischen den Netzen $a_i$ und $b_i$.

2. Die Rekonstruktion wird mit

| $Si$ | (111) | $(\sqrt{3}$ | $\times$ | $\sqrt{3})$ | $R30°$ | $-$(Adsorbat) |
|---|---|---|---|---|---|---|
| ↑ | ↑ | ↑ | | ↑ | ↑ | ↑ |
| Chem. Indizes | Millersche Indizes | $\frac{b_1}{a_1}$ | | $\frac{b_1}{a_1}$ | Rotiert um 30° | Adsorbat-bezeichnung |

   an. $p$ nach den Millerschen Indizes gibt eine primitive Struktur an, $c$ eine zentrierte Struktur.

Abbildung L.4 zeigt als Beispiel eine $Si(111)(\sqrt{3}\times\sqrt{3})R30°$-Rekonstruktion. Wird die Überstruktur durch Fremdatome erzeugt, so werden deren chemische Bezeichnungen hinten angefügt.
Es gibt Rekonstruktionen, die, wenn man auch nichtprimitive Einheitszellen zulässt, auf mehrere Arten bezeichnet werden können. Die Abbildung L.5 zeigt ein Beispiel. Die Rekonstruktion kann wie in der Abbildung links gezeigt, als eine

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

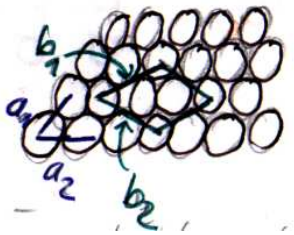



Abbildung L.4.: Struktur einer $Si(111)(\sqrt{3} \times \sqrt{3})R30°$-Rekonstruktion.

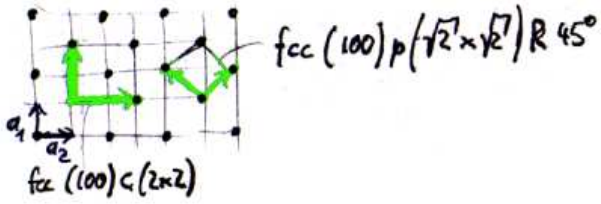



Abbildung L.5.: Beispiel einer Rekonstruktion, deren Bezeichnung nicht eindeutig festgelegt werden kann.

$fcc(100)c(2 \times 2)$-Rekonstruktion oder als $fcc(100)p(\sqrt{2} \times \sqrt{2})R45°$-rekonstruktion bezeichnet werden.
Es gibt Rekonstruktionen, die mit dem obigen Schema nicht zu beschreiben sind. In jedem Falle ist die Matrixschreibweise anwendbar.

$$\underline{b} = \underline{\underline{S}}\,\underline{a} \tag{L.1.4}$$

oder ausgeschrieben

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \tag{L.1.5}$$

Die Matrixschreibweise funktioniert auch bei transzendenten Koeffizienten. Die beiden Beispiele illustrieren die Matrixschreibweise.

- $Si(110)(2 \times 1) \Rightarrow \underline{\underline{S}} = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$
- $Si(111)\left(\sqrt{3} \times \sqrt{3}\right) R\,30° \Rightarrow \underline{\underline{S}} = \begin{pmatrix} 1 & 1 \\ -1 & 2 \end{pmatrix}$

### L.1.2.1. Klassifizierung

1. Alle $S_{ij}$ sind ganzzahlig. $\Rightarrow$ Die Überstrukturen sind einfache Strukturen. ($\forall(m_1, m_2)$ bezeichnen $m_1\boldsymbol{b}_1 + m_2\boldsymbol{b}_2$ gleichartige Gitterplätze bezüglich der Unterlage)

2. Die $S_{ij}$ sind rational. $\Rightarrow$ Es gibt Konzidenzstrukturen. Beispielsweise sei $b_1 = 1.5a_1$. Dann ist jede zweite Zelle bezüglich ihrer Lage zum Wirtsgitter auf einer äquivalenten Position.

3. Die $S_{ij}$ sind irrational. $\Rightarrow$ Es liegen inkohärente Strukturen vor. Das heisst dass die Oberflächenstruktur unabhängig von der Struktur der Unterlage ist.

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

#### L.1.2.1.1. Bemerkungen

- 1. und 2. heissen kommensurable Überstrukturen
- 3. ist eine inkommensurable Überstruktur
- Kommensurable Überstrukturen sind nur eindeutig identifizierbar, wenn die Kohärenzlänge der Untersuchungsmethode grösser als die Periode der Überstruktur ist.
- Kommensurable Überstrukturen mit grösseren Perioden sind von inkommersurablen Strukturen ununterscheidbar.
- Regelmässig gestufte Oberflächen können mit einer zur Beschreibung von Überstrukturen ähnlichen Notation beschrieben werden.

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

# M. Symbole

Im folgenden Abschnitt werden einige in diesem Skript gebrauchten Symbole definiert.

| Symbol | Grösse | Bedeutung |
|---|---|---|
| $\langle y(t) \rangle$ | | Ensemblemittel, Definition: $\langle y\,(t) \rangle \equiv \overline{y\,(t)} \equiv \frac{1}{N} \sum_{k=1}^{N} y^{(k)}\,(t)$, wobei $y^{(k)}\,(t)$ das k-te System ist. |
| $\left\{ y^{(k)}\,(t) \right\}$ | | Zeitmittel, Definition: $\left\{ y^{(k)}\,(t) \right\} \equiv \frac{1}{\Theta} \int_{-\Theta}^{\Theta} y^{(k)}\,(t+t)\,dt$ |
| A | | Ausgangssignal |
| $\alpha_S$ | | Seebeck-Koeffizient (Thermospannung) |
| $a_n$ | | Die Fourierkoeffizienten der *cos*-Funktion |
| B | | *Signal* in der Rückkoppelschleife vor dem Summationspunkt |
| $b_n$ | | Die Fourierkoeffizienten der *sin*-Funktion |
| $\boldsymbol{B}$ | | Magnetische Induktion |
| $c_n$ | | Die Fourierkoeffizienten der $e^{jn\omega t}$-Funktion |
| $\boldsymbol{D}$ | | Dielektrische Verschiebung |
| E | | Eingangssignal |
| $\boldsymbol{E}$ | | Elektrisches Feld |
| F | | Fehlersignal |
| $f(t)$ | | Funktion in der Zeitdomäne |
| $\underline{F}(\omega)$ | | Fouriertransformierte von $f(t)$ |
| G | | Verstärkung in der Vorwärtsrichtung |
| H | | Verstärkung im Rückkoppelungszweig |
| $\boldsymbol{H}$ | | Magnetische Feldstärke |
| $I$ | | Strom |
| $I_0\,(t)$ | | Besselfunktion |
| $\hat{I}$ | | Amplitude des Stromes |
| $i$ | | Stromdichte |
| $J_0\,(t)$ | | Besselfunktion |
| $j$ | | Imaginäre Einheit. Wir verwenden $j$, um eine Verwechslung mit der Stromdichte $i$ zu vermeiden |
| $p$ | | Parameter für Ortskurven in der komplexen Ebene |

| **Symbol** | **Grösse** | **Bedeutung** |
|---|---|---|
| $Q$ | | Ladung |
| $T$ | | Periodendauer |
| $U$ | | Spannung |
| $\underline{Y}$ | | Komplexer Leitwert |
| $\underline{Z}$ | | Komplexe Impedanz |
| $\underline{z}$ | | Normierte komplexe Impedanz |
| $\hat{U}$ | | Amplitude der Spannung |
| $\delta(t)$ | | Die Diracsche Deltafunktion |
| $\epsilon_0$ | $8,85 \times 10^{-12} \frac{(As)^2}{Nm^2}$ | Dielektrizitätszahl des Vakuums |
| $\epsilon_r$ | | Dielektrizitätszahl eines Materials |
| $\rho$ | | Ladungsdichte |
| $\varphi$ | | Phase |
| $\mu_0$ | $4\pi \times 10^{-7} \frac{Vs}{Am}$ | Induktionskonstante |
| $\mu_r$ | | Relative Permeabilität |
| $\omega$ | | Kreisfrequenz |
| $\omega_0$ | | Kreisfrequenz der freien, ungedämpften Schwingung eines Oszillators |
| $\Omega$ | | Normierte Kreisfrequenz |

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

# N. Hilfsprogramme

V.i.S.d.P.: Othmar Marti, Experimentelle Physik, Universität Ulm
**Für den Inhalt externer Links übernimmt weder die Universität Ulm noch das Universitätsrechenzentrum eine Verantwortung.**

- **Externe Datei**: Programm: Spektrum-Analysator für Soundkarte der **DazyWeb Labs**
- **Externe Datei**: Programm: Oszilloskop für Soundkarte der **DazyWeb Labs**
- **Externe Datei**: Programm: Funktionsgenerator für Soundkarte der **DazyWeb Labs**
- **Externe Datei**: Programm: Frequenzgenerator für Soundkarte der **DazyWeb Labs**
- **Externe Datei**: Programm: Audio-Testsystem für Soundkarte der **DazyWeb Labs**
- **Externe Datei**: Programm: Sinusgenerator für Soundkarte der **DazyWeb Labs**
- **Externe Datei**: Programm: Frequenzzsweep-Analysator für Soundkarte der **DazyWeb Labs**
- **Externe Datei**: Programm: IIR-Filterdesigner der **DazyWeb Labs**
- **Externe Datei**: Programm: FIR-Filterdesigner der **DazyWeb Labs**
- **Externe Datei**: Programm: Schaltplan-Editor der **DazyWeb Labs**

# Abbildungsverzeichnis


2.1 Ein grundlegender Systemblock . . . 11
2.2 Ein Differentialoperator in der Blockschaltbildschreibweise . . . 12
2.3 Ein Integraloperator in der Blockschaltbildschreibweise . . . 12
2.4 Ein Abnahmepunkt. Das Eingangssignal wird nach rechts auf drei Zweige aufgeteilt. . . . 12
2.5 Ein Abnahmepunkt. Hier wird ein Teil des Signals rückgekoppelt . . . 13
2.6 Ein Summationspunkt. Die Operatoren + und - geben an, ob addiert oder subtrahiert werden soll . . . 13
2.7 Das Blockschaltbild der Differentialgleichung (2.1.1) in der Form (2.1.3). . . . 13
2.8 Das Blockschaltbild der Differentialgleichung (2.1.1) in der Form (2.1.4). . . . 14
2.9 Kaskade von zwei Blöcken . . . 14
2.10 Das Kommutativgesetz für einen Block. . . . 14
2.11 Blockschaltbild eines rückgekoppeltes Systems . . . 16
2.12 Signalflussdiagramm. Oben ist ein allgemeiner Zweig dargestellt, bei dem gilt: $Y_i = A_{ij}X_j$. Als Beispiel ist unten das Ohm'sche Gesetz $U = RI$ gezeigt. . . . 18
2.13 Summation in einem Signalflussdiagramm: $Y_j = \sum_{i=1}^{n} A_{ji}X_i$. . . . 18
2.14 Übertragungsregeln in einem Signalflussdiagramm: $X_j = A_{jk}X_k$ für $(k = 1, 2, \ldots, n)$. . . . 18
2.15 Multiplikationsregeln für Signalflussdiagramme: $X_n = \left(\prod_{i=2}^{n} A_{i(i-1)}\right) X_1$. 19
2.16 Signalflussdiagramm zur Erklärung der Definitionen . . . 19
2.17 Zeigerdiagramm für die Spannung $\underline{U}$ . . . 21
2.18 Ortskurve für einen realen *Parallelschwingkreis*. Die Impedanz ist mit $R$ skaliert. Rechts befindet sich eine Skizze dieses Schwingkreises. . . . 21
2.19 Ortskurve für einen realen *Parallelschwingkreis*. Hier wurde sowohl durch R wie auch durch $Q^2$ geteilt. . . . 23
2.20 Digitale Übertragungskette im Zeitbereich . . . 32
2.21 Digitale Übertragungskette im Frequenzbereich . . . 32
2.22 Einschalten einer Spannung an einem RC-Glied . . . 37
2.23 Kombinierte Stoss- und Sprungfunktion $U_{st}(t, \Delta t, U_{st,0}, U_\infty)$, bestehend aus $U_{sp}(t, 0, U_{st,0})$ (blau) und $U_{sp}(t, \Delta t, U_\infty - U_{st,0})$ (rot) . . . 38
2.24 Anlegen einer Wechselspannung an eine Spule . . . 40
2.25 Resultierende Ströme beim Anlegen einer Wechselspannung an eine Spule. **Links:** Einschalten im Nulldurchgang, **Rechts:** Einschalten bei Maximalspannung. . . . 41
2.26 *Negation*, *Nicht-Gatter*, *Inverter* und *NOT* . . . 43

2.27 *Und*, Und-Gatter, *Konjunktion* und *AND* . . . . . 43
2.28 Oder, Oder-Gatter, Disjunktion, OR . . . . . 44
2.29 NAND, NAND-Gatter . . . . . 45
2.30 NOR, NOR-Gatter . . . . . 45
2.31 Äquivalenz, Exklusiv-NOR, XNOR . . . . . 46
2.32 Antivalenz, XOR-Gatter, . . . . . 47
2.33 Vergleich einer Wahrheitstabelle mit einem Karnaugh-Diagramm . . . . 49
2.34 Erweitertes Beispiel zum Vergleich einer Wahrheitstabelle mit einem Karnaugh-Diagramm . . . . . 51
2.35 Anschlüsse, Ströme und Spannungen bei einem Vierpol . . . . . 52
2.36 Serienschaltung zweier Vierpole . . . . . 53
2.37 Parallelschaltung zweier Vierpole . . . . . 55
2.38 Kettenschaltung zweier Vierpole . . . . . 55
2.39 Übertragungsfunktion eines Vierpols . . . . . 56
2.40 Ersatzschaltung eines Vierpols: T- Glied (Sternschaltung) . . . . . 58
2.41 Ersatzschaltung eines Vierpols . . . . . 59
2.42 Ersatzschaltung eines Vierpols: Kreuzpol . . . . . 59
2.43 Systemumgebung eines digitalen Filters . . . . . 67
2.44 Analoge und digitale Funktionen . . . . . 67
2.45 Spektrum der Eingangsspannung vor und nach dem Abtasten . . . . . 68
2.46 Oversampling oder Schwebung beim Abtasten . . . . . 69
2.47 Fensterfunktion beim Abtasten mit Rechteckpulsen . . . . . 69
2.48 Digitale Signalverarbeitungskette . . . . . 70
2.49 Ein Beispieltiefpass . . . . . 71
2.50 Frequenzgang eines digitalen Tiefpasses mit $\beta_1 = -0,85$ . . . . . 72
2.51 Sprungantwort eines digitalen Tiefpasses mit $\beta_1 = -0,85$ . . . . . 72
2.52 Frequenzgang eines digitalen Hochpasses mit $\beta_1 = 0,85$ . . . . . 72
2.53 Sprungantwort eines digitalen Hochpasses mit $\beta_1 = 0,85$ . . . . . 73
2.54 Der digitale Tiefpass als Integrator . . . . . 73
2.55 Digitales Filter mit verteilten Summierern . . . . . 74
2.56 Digitales Filter mit globalen Summierern an Eingang und Ausgang . . 74
2.57 Digitales Filter mit einem globalen Summierer am Ausgang . . . . . 74
2.58 Kaskadierung von digitalen Filtern . . . . . 75
2.59 FIR Tiefpass 1. Ordnung . . . . . 77
2.60 FIR Hochpass 1. Ordnung . . . . . 78
2.61 FIR Tiefpass 2. Ordnung . . . . . 79
2.62 FIR Hochpass 2. Ordnung . . . . . 80
2.63 FIR Bandsperre 2. Ordnung . . . . . 80
2.64 FIR Bandpass 2. Ordnung . . . . . 81
2.65 transformation eines Filters auf periodischen Frequenzgang . . . . . 82
2.66 Struktur eines IIR-Filters erster Ordnung . . . . . 83
2.67 Allgemeines IIR-Filter zweiter Ordnung . . . . . 84
2.68 Der RLC-Kreis für Rauschuntersuchungen . . . . . 90
2.69 Widerstand und allgemeine Impedanz im Gleichgewicht . . . . . 92
2.70 Berechnung der Rauschspannung nach Nyquist . . . . . 93
2.71 Signal- und *Rauschbandbreite* eines Tiefpassfilters . . . . . 96
2.72 von Neumannsche Struktur eines Computers . . . . . 98
2.73 Rechenoperationen bei von Neumannscher Archidektur . . . . . 98

2.74 Abarbeitung eines Maschinenbefehls . . . 99
2.75 Blockschaltbild eines Signalprozessors . . . 100
2.76 Ablaufs einer Rechenoperation in einem Signalprozessor . . . 101
2.77 Blockschaltbild der MAC-Einheit eines Signalprozessors . . . 103

3.1 Leitfähigkeit und spezifischer Widerstand von Metallen, Halbleitern und *Isolatoren* bei Zimmertemperatur. . . . 105
3.2 Energieschema für Metall, Halbleiter und Isolator . . . 106
3.3 Element-, binäre und ternäre Verbindungshalbleiter. . . . 107
3.4 Schematischer Verlauf der Bandaufspaltung . . . 108
3.5 Bindungsstruktur (kovalente Bindung) des Eigenhalbleiters in zweidimensionaler Darstellung mit zunehmender zeichnerischer Abstraktion. . . 108
3.6 Verschiedene Bandstrukturen . . . 109
3.7 Fermi–Funktion und Zustandsdichte . . . 112
3.8 Abhaengigkeit der *Fermi-Energie* . . . 112
3.9 Schematische Darstellung der Wirkung eines Donators . . . 114
3.10 Qualitative Lage der Grundzustandniveaus von Donatoren und Akzeptoren . . . 115
3.11 Schematisches Bändermodell, Zustandsdichte und Fermi–Verteilung sowie Ladungsträgerdichten für a) intrinsische, b) n– und c) p–Halbleiter. 116
3.12 Qualitative Abhängigkeit der Elektronenkonzentration . . . 118
3.13 Beweglichkeit als Funktion der Elektronen–Konzentration . . . 119
3.14 Schematische Abhängigkeit der Beweglichkeit . . . 119
3.15 Spezifischer Widerstand in Abhängigkeit von der Konzentration für verschiedene *Halbleiter*. . . . 120
3.16 Rekombinationsprozesse . . . 121
3.17 Elektronendiffusion mit Rekombinationsprozessen . . . 122
3.18 Qualitatives Schema eines p–n–Uebergangs . . . 126
3.19 Das Schottky–Modell der Raumladungszone . . . 127
3.20 Einige Herstellungsmethoden für p–n–Dioden . . . 128
3.21 Vergleich der Dotierprofile . . . 129
3.22 Schema eines p–n–Uebergangs . . . 130
3.23 Schematische Darstellung des Zustandekommens des Sperrstromes . . . 131
3.24 Schema der Strom–Spannungs Kennlinie . . . 132
3.25 Kennlinie einer p–n–Diode . . . 133
3.26 Sperrschichtkapazitaet, Diffusionskapazitaet und Gesamtkapazitaet . . . 133
3.27 Baenderschemata . . . 134
3.28 Baenderschema eines Halbleiter–Heterouebergangs . . . 135
3.29 Modulationsdotierter Heterouebergang . . . 136
3.30 Baenderschema eines modulationsdotierten Kompositions–Uebergitters 137
3.31 Energiezustände von Elektronen . . . 137
3.32 *Bänderschema* eines Metall–Metall–Kontakts. . . . 138
3.33 *Bänderschema* eines Kontakts zwischen Metall und n–Halbleiter. . . . 139
3.34 Verarmungsrandschicht für Löcher. . . . 139
3.35 Löcherinjektion im Zwei–Band–Modell. . . . 140
3.36 Elektronisches Bänderschema eines Metall – n–Halbleiter–Übergangs . 141
3.37 Aufbau einer PtSi–n–Si–Diode . . . 141
3.38 Ohmscher Metall – n–Halbleiter–Kontakt . . . 142

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

3.39 Kennlinie eines gleichrichtenden Metall–Halbleiter–Kontaktes. . . . . 142
3.40 Zum Vergleich: Beispiele gemessener Diodenkennlinien . . . . . 143
3.41 Aufwachsen von SiO2 . . . . . 143
3.42 Schematischer Aufbau einer MIS–Diode . . . . . 144
3.43 Baenderschema einer idealen MIS-Diode . . . . . 144
3.44 Baenderschema einer idealen MIS–Dioden mit aeusserer Spannung . . . 145
3.45 Ladungsverhältnisse in einer realen MOS–Diode . . . . . 146
3.46 Baenderschema und Raumladungen eines realen MOS–Kontaktes . . . 147
3.47 Baenderschema einer MOS–Diode . . . . . 149
3.48 Baenderschema einer Si-MOS-Diode unter Beleuchtung . . . . . 149
3.49 Schnitt durch eine Dreiphasen–CCD–Verschiebeeinheit . . . . . 150
3.50 Aufbau eines n–Kanal–CCD . . . . . 151
3.51 Potentialverlauf und Ladungsverteilung . . . . . 151
3.52 Trapping . . . . . 152
3.53 Aufbau eines n–Kanal–BCCD . . . . . 152
3.54 Aufbau eines BCCD . . . . . 152
3.55 Frame–Transfer–Konzept. . . . . 153
3.56 Schaltsymbole für den FET. . . . . 155
3.57 Prinzip des Sperrschicht–FET. . . . . 155
3.58 JFET mit negativer Gate–Source–Spannung. . . . . 156
3.59 Kennlinien des JFET. . . . . 156
3.60 Schematischer Aufbau eines planaren JFET . . . . . 157
3.61 Aufbau eines Double–Gate JFET . . . . . 158
3.62 Aufbau eines GaAs MESFET . . . . . 158
3.63 MESFET . . . . . 158
3.64 HEMT . . . . . 159
3.65 MOSFET . . . . . 160
3.66 Kennlinien von Kurz– und Langkanal–MOSFETs . . . . . 161
3.67 CMOS–Technik: Inverter im a) Schaltbild und b) im technologischen Aufbau. . . . . 162
3.68 CMOS–Inverter . . . . . 162
3.69 CMOS: p– und n–Kanal–E–Type MOSFETs . . . . . 164
3.70 Uebersicht ueber die verbreitesten Halbleiter–Speichertypen . . . . . 165
3.71 CMOS–Speicherzelle . . . . . 166
3.72 DRAM–Zelle . . . . . 166
3.73 FAMOS–Struktur . . . . . 168
3.74 Schaltsymbole eines a) pnp– und b) npn–Transistors . . . . . 169
3.75 Basisschaltung eines npn–Transistors . . . . . 170
3.76 npn–Transistor im Gleichgewicht . . . . . 170
3.77 Uebertragungs– und Ausgangskennlinie . . . . . 172
3.78 Emitterschaltung: Eingangs– und Ausgangskennlinien . . . . . 173
3.79 Emitterschaltung: Bipolartransistor als Schalter . . . . . 173
3.80 Querschnitt und Dotierprofil des vertikalen npn–Bipolartransistors . . . 174
3.81 AlGaAs–GaAs–Hetero–Bipolartransistor . . . . . 175
3.82 Schema der pnn-Fotodiode . . . . . 176
3.83 Prinzip des Fototransistors . . . . . 177
3.84 PN-Laserdiode mit angelegter Spannung . . . . . 178
3.85 Prinzip des Laserresonators . . . . . 178

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

3.86 Quantum Well Lasers. . . . . 179
3.87 *GaAs–VCSEL* für 750 nm, Arbeitsgruppe Ebeling, Universität Ulm. . . 180
3.88 Coulomb–Blockade und Single electron electronics. . . . . 180
3.89 Theveninsche Ersatzschaltung . . . . . 182
3.90 Unbelasteter und belasteter Spannungsteiler . . . . . 183
3.91 Beispiel fuer die Ersatzschaltung eines Spannungsteilers . . . . . 183
3.92 Diode: Schaltzeichen und Aufbau . . . . . 184
3.93 Kleinsignal–Diode . . . . . 185
3.94 Einweggleichrichter oben und zwei Klemmen unten . . . . . 186
3.95 Zweiweg–Brueckengleichrichter . . . . . 186
3.96 Schaltsymbol und Durchbruchkennlinien . . . . . 187
3.97 Schaltzeichen und spannungsabhängige Kapazitäten . . . . . 188
3.98 Spannungen und Ströme eines npn–Transistors . . . . . 189
3.99 Gross– und Kleinsignalverstärkung . . . . . 189
3.100Zur Bestimmung des Arbeitspunktes eines npn–Transistors . . . . . 190
3.101Zur Bestimmung des Arbeitspunktes im Eingangs– und Ausgangskennlinienfeld . . . . . 190
3.102Gleichstrom–Kleinsignalersatzschaltbild des Bipolartransistors . . . . . 191
3.103Emitterschaltung: Schaltbild und Ersatzschaltbild für den Normalbetrieb 193
3.104Emitterschaltung: Übertragungskennlinie . . . . . 193
3.105Emitterschaltung: Übertragungskennlinie und Verstärkung . . . . . 193
3.106Emitterschaltung: Kleinsignal–Ersatzschaltbild . . . . . 194
3.107Emitterschaltung mit Stromgegenkopplung . . . . . 195
3.108Emitterschaltung mit Stromgegenkopplung: Übertragungskennlinie . . 195
3.109Emitterschaltung mit Stromgegenkopplung: Kleinsignal–Ersatzschaltbild 195
3.110Emitterschaltung mit Spannungsgegenkopplung: Schaltung und Ersatzschaltbild (Normalbetrieb) . . . . . 196
3.111Emitterschaltung mit Spannungsgegenkopplung: Übertragungskennlinie und Verstärkung . . . . . 197
3.112Emitterschaltung mit Spannungsgegenkopplung: Kleinsignal–Ersatzschaltbild 197
3.113Kollektorschaltung: Schaltbild und Ersatzschaltbild . . . . . 197
3.114Kollektorschaltung: Kennlinien . . . . . 198
3.115Kollektorschaltung: Kleinsignal–Ersatzschaltbild . . . . . 198
3.116Kollektorschaltung: Widerstandstransformation . . . . . 198
3.117Basisschaltung: Schaltbild und Ersatzschaltbild für den Normalbetrieb 199
3.118Basisschaltung: Kennlinien . . . . . 200
3.119Basisschaltung: Kleinsignal–Ersatzschaltbild . . . . . 200
3.120npn–Darlington–Transistor: Schaltung und Schaltzeichen . . . . . 201
3.121Darlington–Transistor: zum Schaltverhalten . . . . . 201
3.122Zum Steilheitskoeffizient . . . . . 202
3.123Sourceschaltung: Schaltung und Ersatzschaltbild . . . . . 203
3.124Sourceschaltung: Übertragungskennlinie . . . . . 204
3.125Transimpedanzverstärker . . . . . 204
3.126Prinzip einer einfachen Stromquelle . . . . . 205
3.127Beispiele einfacher Stromquellen . . . . . 205
3.128Prinzip eines einfachen Stromspiegels . . . . . 206
3.129Prinzip der Kaskodenschaltung . . . . . 207
3.130Grundschaltung des Differenzverstärkers . . . . . 207

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

3.131Operationsverstärker . . . . . 208
3.132OPV–Typen . . . . . 210
3.133VV-OPV mit einfacher ohmscher Gegenkopplung . . . . . 210
3.134Aufbau eines einfachen VV–OPV . . . . . 212
3.135Einfacher Universalverstärker . . . . . 213
3.136Operationsverstärker der 741–Klasse . . . . . 213
3.137Bode–Diagramm eines unkorrigierten OPVs der 741–Klasse . . . . . 213
3.138Grenzfrequenzen der OPV der 741–Klasse . . . . . 214
3.139Zur Phasenreserve . . . . . 215
3.140Zur universellen Frequenzgangkorrektur . . . . . 215

4.1 Messung des Ausgangsstroms einer Stromquelle mit einem Strommesser 218
4.2 Strommessung mit Spannungsmesser . . . . . 218
4.3 Strommessung mit umschaltbarem Strommesser . . . . . 219
4.4 Strommesser mit umschaltbarem Innenwiderstand . . . . . 220
4.5 Strommesser mit umschaltbarem *Messwiderstand* . . . . . 220
4.6 Kompensationsmethode . . . . . 220
4.7 Strommessung mit Operationsverstärker . . . . . 221
4.8 Erdfreie Präzisionsmessung des Stromes . . . . . 222
4.9 Spannungsmessung mit Spannungsmesser . . . . . 222
4.10 Spannungsmessung mit Strommesser . . . . . 223
4.11 Spannungsmessung mit Teilerkette . . . . . 223
4.12 Spannungsmessung mit Strommesser und umschaltbaren Vorwiderständen . . . . . 224
4.13 Poggendorff'sche Kompensationsmethode . . . . . 224
4.14 Spannungsmessung mit Operationsverstärkern . . . . . 225
4.15 Spannungsmessung mit Instrumentenverstärker . . . . . 226
4.16 Frequenzgang der Gegentaktverstärkung eines Instrumentenverstärker . 226
4.17 Frequenzgang der Gleichtaktverstärkung eines Instrumentenverstärker . 227
4.18 Frequenzgang eines hochwertigen Instrumentenverstärkers . . . . . 227
4.19 Spannungsmessung an Koaxialkabel . . . . . 228
4.20 Spannungsmessung über Koaxialkabel mit nachgeführter Spannung am Schirm . . . . . 229
4.21 Spannungsmessung über Koaxialkabel: verbesserte methode . . . . . 229
4.22 Messung der Scheitelspannung . . . . . 230
4.23 Messung der Scheitelspannung mit Diode . . . . . 230
4.24 Halbwellengleichrichter mit Operationsverstärker . . . . . 232
4.25 Halbwellengleichrichter mit Operationsverstärker . . . . . 232
4.26 Halbwellengleichrichter mit TL074 . . . . . 233
4.27 Vollwellengleichrichter mit Operationsverstärker . . . . . 233
4.28 Vollwellengleichrichter mit LM741 . . . . . 234
4.29 Vollwellengleichrichter mit TL074 . . . . . 234
4.30 Vollwellengleichrichter mit Transistoren . . . . . 235
4.31 True-RMS Wandler . . . . . 235
4.32 Real-Time-Vektormesser . . . . . 236
4.33 Brückenschaltung (Graetz-Schaltung) . . . . . 237
4.34 Brückenschaltung mit Operationsverstärker . . . . . 237
4.35 Brückenschaltung mit Transformator als Wandler . . . . . 238

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

4.36 Ladungsmessung über Ladungstransfer . . . 239
4.37 Ladungsmessung über Ladungstransfer . . . 240
4.38 Ladungsmessung mit Integrator . . . 240
4.39 Ladungsmessung mit Integrator . . . 241
4.40 Stromrichtige Widerstandsmessung. . . . 241
4.41 Spannungsrichtige Widerstandsmessung. . . . 242
4.42 Widerstandsmessung mit bekannter Stromquelle. . . . 242
4.43 Widerstandsmessung mit bekannter *Spannungsquelle* und Spannungsmesser. . . . 243
4.44 Widerstandsmessung mit bekannter *Spannungsquelle* und Strommesser. 243
4.45 Widerstandsmessung durch Vergleich mit Referenz. . . . 244
4.46 Widerstandsmessung durch Vergleich mit Referenz . . . 245
4.47 Widerstandsmessung durch Stromvergleich . . . 245
4.48 Vierdraht-Widerstandsmessung für kleine Widerstände . . . 246
4.49 Vierdraht-Widerstandsmessung für kleine Widerstände . . . 246
4.50 Vierdraht-Widerstandsmessung für sehr kleine Widerstände . . . 248
4.51 Messung der Kapazität . . . 249
4.52 Messung der Induktivität . . . 250
4.53 Messung der Induktivität oder Kapazität . . . 250
4.54 Wheatstone-Brücke . . . 252
4.55 Unabgeglichene Wheatstone-Brücke . . . 252
4.56 Unabgeglichene Wheatstone-Brücke . . . 253
4.57 Unabgeglichene Wheatstone-Brücke . . . 253
4.58 Unabgeglichene Wheatstone-Brücke . . . 254
4.59 Impedanz des Strommessers im Nullpunktszweig der Wheatstone-Brücke 255
4.60 Empfindlichkeit der unabgeglichenen Wheatstone-Brücke . . . 256
4.61 Empfindlichkeit der unabgeglichenen Wheatstone-Brücke . . . 256
4.62 Empfindlichkeit der unabgeglichenen Wheatstone-Brücke . . . 257
4.63 Empfindlichkeit der unabgeglichenen Wheatstone-Brücke . . . 257
4.64 Empfindlichkeit der unabgeglichenen Wheatstone-Brücke . . . 258
4.65 Wheatstone-Brücke für allgemeine Impedanzen . . . 259
4.66 Digital-Analog-Wandlerschaltung . . . 260
4.67 Monotone und nicht-monotone Wandlerkennlinie . . . 261
4.68 Fehler von A/D- und D/A-Wandlern . . . 262
4.69 Direkter Digital/Analog-Wandler. . . . 263
4.70 Digital-Analog-Wandlerschaltung mit Analogschaltern . . . 264
4.71 Digital-Analog-Wandlerschaltung mit Stromwechselschaltern. . . . 265
4.72 Digital-Analog-Wandlerschaltung mit R-2R-Netzwerk. . . . 267
4.73 Digital-Analog-Wandlerschaltung mit R-2R-Netzwerk und bipolarem Ausgang. . . . 268
4.74 Digital-Analog-Wandlerschaltung mit Pulslängenmodulation. . . . 269
4.75 1-Bit D/A-Wandler . . . 270
4.76 Oversampling . . . 271
4.77 4-fach Oversampling. . . . 272
4.78 MASH (Multi-Stage Noise Shaping)-Wandlung . . . 272
4.79 Direkter Analog-Digital-Wandler. . . . 273
4.80 Analog-Digital-Wandler nach dem Nachlaufverfahren. . . . 275
4.81 *Analog-Digital-Wandler* nach dem Wägeverfahren . . . 275

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

4.82 *Analog-Digital-Wandler* nach dem Sägezahnverfahren . . . . . 276
4.83 *Analog-Digital-Wandler* nach dem Dual-Slope-Verfahren . . . . . 277
4.84 Spannungsverlauf beim Dual-Slope-Verfahren . . . . . 277
4.85 Sigma-Delta-Wandler. . . . . 279
4.86 Prinzipbild eines Lock-In-Verstärkers. . . . . 282
4.87 Signalformen des Lock-In-Verstärkers . . . . . 283
4.88 Lock-In-Verstärker mit dem AD630 . . . . . 284
4.89 Messung von Zeit oder Periodendauer. . . . . 285
4.90 Prinzip der Schaltung für eine Zeitmessung . . . . . 286
4.91 Prinzip der Schaltung für eine Periodendauermessung . . . . . 286
4.92 Schaltung für eine Periodendauermessung . . . . . 286
4.93 Messung der Frequenz . . . . . 287
4.94 Schaltung für eine Frequenzmessung . . . . . 287
4.95 Schaltung für eine Frequenzmessung. Der Unterschied zur Periodendauermessung liegt alleine in der Taktfrequenz des Referenzoszillators V1! . . . . . 288
4.96 Messung der Frequenz mit dem Verhältniszählverfahren. . . . . 288
4.97 Ersatzschaltbild eines Schwingquarzes . . . . . 289
4.98 Feineinstellung der Quarzfrequenz . . . . . 290
4.99 Quarzoszillator nach Colpitts . . . . . 291
4.100Quarzoszillator nach Hartley . . . . . 291
4.101Quarzoszillator mit Inverter-Gatter . . . . . 292
4.102Prinzipschaltbild eines Phasenregelkreises . . . . . 292
4.103Prinzipschaltbild eines Phasenregelkreises . . . . . 293
4.104Sample/Hold als Phasendetektor . . . . . 294
4.105Sample/Hold als Phasendetektor . . . . . 294
4.106Vorzeichenrichtiger Phasendetektor . . . . . 294
4.107Kennlinie des vorzeichenrichtigen Phasendetektor . . . . . 295
4.108Implementierung eines vorzeichenrichtigen Phasendetektor . . . . . 296
4.109Messung kurzer Zeiten mit definierten Rampen . . . . . 296
4.110Messung von Magnetfeldern mit Transformator . . . . . 297
4.111Messung von Magnetfeldern mit dem Halleffekt . . . . . 298
4.112Äquipotentialverlauf beim Halleffekt . . . . . 299
4.113Aufbau einer Feldplatte . . . . . 300
4.114Widerstand einer Feldplatte . . . . . 301
4.115Magnetoresistiver Effekt . . . . . 301
4.116Aufteilung der Felder und Ströme . . . . . 301
4.117Aufbau eines magnetoresistiven Sensors . . . . . 303
4.118Magnetoresistiver Sensor mit 'Barber Poles' . . . . . 303
4.119Kritischer Strom beim Josephson-Effekt . . . . . 304
4.120Aufbau eines SQUID . . . . . 304
4.121Dielektrische Funktion und Spektren . . . . . 305
4.122Temperaturabhängigkeit des elektrischen Widerstandes . . . . . 308
4.123Zusammenfassung Bändermodell . . . . . 308
4.124Temperaturabhängigkeit der Ladungsträgerkonzentration . . . . . 309
4.125Temperaturabhängigkeit der Spreading-Resistance . . . . . 310
4.126Kennlinie eines Heissleiters (NTC) . . . . . 311
4.127Kennlinie eines Kaltleiters (PTC) . . . . . 312

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

4.128Kennlinie eines integrierten Temperatursensors. Das Flächenverhältnis ist so eingestellt, dass die resultierende Spannung in $mV$ ein Zehntel der Temperatur in $K$ ist. Die Temperaturablesung mit einem digitalen Voltmeter wird dann sehr einfach. . . . . 314
4.129Thermospannung . . . . . 315
4.130Entstehung des pyroelektrischen Effektes . . . . . 315
4.131Äusserer Photoeffekt: Bändermodelle für Metalle und Halbleiter. . . . . 316
4.132Photozelle und Photovervielfacher . . . . . 317
4.133Innerer Photoeffekt für Halbleiter. . . . . 318
4.134Bändermodelle für den inneren Fotoeffekt . . . . . 319
4.135Spektrale Empfindlichkeit von verschiedenen Materialien im Vergleich zu Quanteneffizienzen. . . . . 320
4.136Kennlinie einer Photodiode . . . . . 320
4.137Photodiode und Photoelement . . . . . 321
4.138Phototransistor . . . . . 321
4.139Plattengeführte Elektromagnetische Welle . . . . . 322
4.140Wellenleiter und Ersatzschaltbild . . . . . 322
4.141Wellenleiter: Koordinaten und Vierpoldarstellung . . . . . 325
4.142Einfluss des Abschlusswiderstandes. . . . . 326
4.143Kurzgeschlossene Leitung . . . . . 327
4.144Stichleitung und Stützen für Lecherleitungen . . . . . 327
4.145Impedanzanpassung bei einer HF-Leitung . . . . . 328
4.146Offene Leitung . . . . . 328
4.147Impedanz-Transformation . . . . . 329
4.148Koordinatensystem für Zweidrahtleitung . . . . . 330
4.149Relative Lage von E- und B-Feldern . . . . . 332
4.150Integrationswege um Strom und Spannung zu erhalten. . . . . 332
4.151Berechnung von Kapazitätsbelag und Induktivitätsbelag . . . . . 333
4.152Bauformen von Wellenleitern. . . . . 335
4.153Feldlinienbild in rechteckförmigen Wellenleitern. . . . . 335
4.154Rechteckförmigen Wellenleiter . . . . . 336
4.155Dämpfung elektromagnetischer Wellen in Luft. . . . . 337
4.156Skintiefe elektromagnetischer Wellen. . . . . 338
4.157Dämpfungseigenschaften von Wellenleitern. . . . . 339
4.158Streifenleiter. . . . . 339
4.159Brechungsgesetz von Snellius. . . . . 339
4.160Dämpfung im Lichtwellenleiter . . . . . 340
4.161Spektrale Dämpfungswerte von verschiedenen Glasfasertypen. . . . . 341
4.162Glasfasertypen . . . . . 342
4.163Geometrie zur Berechnung von Stufenindexfasern. . . . . 343
4.164Geometrie zur Berechnung von Gradientenindexfasern. . . . . 344
4.165Gemessenes Brechzahlprofil einer Gradientenindexfaser. . . . . 346
4.166Kopplungswirkungsgrad in Lichtwellenleiter . . . . . 347
4.167Dämpfung bei Verbindung von Glasfasern . . . . . 348
4.168Feldradiusbestimmung bei einer Einmoden-Glasfaser. . . . . 349
4.169Felddurchmesser und Wellenlänge bei Einmoden-Glasfasern . . . . . 350
4.170Herstellung eines Bragg-Gitters mit Interferometrie. . . . . 351
4.171Bragg-Gitter mit Prismen- und Lloyd-Interferometer . . . . . 352

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

4.172Herstellung eines Bragg-Gitters mit einer Phasenmaske. . . . . 352
4.173Reflektionsspektrum eines Bragg-Gitters . . . . . 353
4.174Berechnetes Reflexionsspektrum eines Bragg-Gitters . . . . . 354
4.175Transmission durch starkes Bragg-Gitter . . . . . 355
4.176Bragg-Wellenlänge und mechanischen Spannung . . . . . 355
4.177Bragg-Wellenlänge und Temperatur . . . . . 356
4.178Temperaturabhaengigkeit der Reflektivitaet . . . . . 357
4.179Fabry-Perot-Resonator mit zwei Bragg-Gittern . . . . . 358
4.180Reflexionsspektrum für sieben Bragg-Gitter . . . . . 358
4.181Testfeld fuer eine *Spannungsquelle* . . . . . 359
4.182Beispiel fuer eine Testhuelle . . . . . 360
4.183Fehlerquellen bei der Spannungsmessung . . . . . 362
4.184Fehlerquellen bei der Strommessung . . . . . 362
4.185Fehlerquellen bei der Widerstandsmessung . . . . . 363
4.186Spannungsmessung: Einfluss des Kabelleckstromes . . . . . 368
4.187Messung kleiner Spannungen mit einer Guard-Konfiguration . . . . . 368
4.188Kompensation von ungewollten Thermospannungen . . . . . 369
4.189Gleichtaktstrom hervorgerufen durch die Ankopplung an die Netzspannung . . . . . 369
4.190Gleichtaktstrom hervorgerufen durch den falschen Anschluss der Messkabel . . . . . 370
4.191Durch Magnetfelder induzierte Spannungen . . . . . 370
4.192Erdschleifen hervorgerufen durch mehrfache Erdung . . . . . 371
4.193Verhinderung des Einflusses von Erdströmen . . . . . 371
4.194Elektrostatische Kopplung und elektrostatische Abschirmung . . . . . 372
4.195Ausgangswiderstandes und der kapazitiven Belastung einer Stromquelle 372
4.196Triboelektrische Effekte durch die Verbiegung von Kabeln . . . . . 373
4.197Elektrochemische Effekte und Leckströme an Oberflächen . . . . . 374
4.198Guarding-Techniken . . . . . 374
4.199Störspektrum in Schaltungen . . . . . 375
4.200Aufbau eines Lasers sowie Darstellung der Inversion . . . . . 377
4.201Verstärkung und der Verluste in einem Resonator . . . . . 377
4.202Laserpumpprozesse . . . . . 379
4.203Beugung einer ebenen Welle an einer Blende . . . . . 381
4.204Erklärung der Fresnelzahl . . . . . 381
4.205Optik eines Resonators . . . . . 382
4.206Feldamplitude und Fouriertransformation . . . . . 382
4.207Modenverteilungen in Laserresonatoren . . . . . 383
4.208Radiale Amplitudenverteilung in konfokalen Resonatoren . . . . . 384
4.209Beispiele für Laserresonatoren . . . . . 384
4.210Beugungsverluste von $TEM_{n,m}$-Moden . . . . . 385
4.211Stabilitätsdiagramm für optische Resonatoren . . . . . 385
4.212Beispiele von instabilen Resonatoren . . . . . 386
4.213Frequenzspektren von Resonatoren . . . . . 387
4.214Fabry-Perot-Resonanzen: oben ist ein Überblick gezeigt, unten die Vergrösserung um 1. Die Kurven sind auf einen Frequenzabstand von 1 normiert. . . . . 388

4.215Normierte Linienbreite als Funktion von $T = 1 - R$. Der Modenabstand im Fabry-Perot-Resonator ist 1. . . . . 389
4.216Verstärkungsprofil eines Laserüberganges . . . . . 390
4.217Modenprofil des aktiven Resonators in Abhängigkeit der Verstärkung. . 391
4.218Verstärkungsprofil des aktiven Mediums . . . . . 391
4.219Zeitliche Beziehung zwischen Pumppuls, Laserpuls und Besetzungsinversion . . . . . 392
4.220Spikes bei Kurzpulslasern . . . . . 392
4.221Güteschaltung . . . . . 393
4.222Pockelszelle . . . . . 393
4.223Kurvenverlauf bei Güteschaltung . . . . . 394
4.224Ablenkung von Licht durch Schallwellen . . . . . 394
4.225Cavity Dumper . . . . . 395
4.226Laserpuls mit Cavity Dumper . . . . . 396
4.227Aktive Modenkopplung mit AOM . . . . . 397
4.228Modengekoppelte Laserpulse . . . . . 397
4.229Modengekoppelter Laser mit sättigbarem Absorber . . . . . 399
4.230Modengekoppelter Puls . . . . . 400
4.231Synchron gepumpter Kurzpulslaser . . . . . 400
4.232Schematischer Aufbau eines CPM-Lasersystems. . . . . 401
4.233Pulsbreite und spektrale Breite des Lasermediums . . . . . 402
4.234Dispersionskompensation mit zwei Gittern. . . . . 403
4.235Aufbau eines Cr:LiSAF-Lasers . . . . . 404
4.236Schematischer Verlauf der Reflektivität in einem *sättigbaren Bragg-Spiegel* 405
4.237Aufbau zur Messung einer Absorption . . . . . 407
4.238Verbesserter Aufbau zur Messung einer Absorption mit Choppern . . . 408
4.239Aufbau zur Messung der Reflektivität . . . . . 408
4.240Aufbau zur Messung der Polarisation . . . . . 409
4.241Prismenspektrometer . . . . . 409
4.242Gitterspektrometer . . . . . 412
4.243Intensität des gebeugten Lichtes als Funktion der Anzahl beleuchteter Spalte. Links ist eine Übersicht, rechts die Detailansicht. Von unten nach oben sind $N = 3$,6 und 12 beleuchtete Spalte dargestellt. Die sonstigen Gitterparameter sind $a = 1$ und $\epsilon = 0.1$ . . . . . 413
4.244Aufbau eines Fabri-Perot-Spektrometer . . . . . 414
4.245Kennlinie eines Fabri-Perot-Inerferometers für Reflektivitäten von 0.6, 0.8, 0.9 und 0.95 beziehungsweise Werte von $M$ von 15, 80, 360 und 1520 (von oben) . . . . . 415
4.246Optischer Korrelator . . . . . 416
4.247Untergrundfreier Korrelator . . . . . 418
4.248Zweiphotonenfluoreszenz zur Messung von kurzen Pulsen . . . . . 418
4.249Autokorrelationsprofile . . . . . 419
4.250Aufbau einer Streak-Kamera . . . . . 420
4.251Schalten eines Wellenleiters mit optischen Pulsen . . . . . 420
4.252Zeitkorreliertes Einzelphotonenzählen. Die vertikale Achse zeigt die einzelnen Messungen. Horizontal ist jeweils ein roter Balken eingezeichnet, wenn bei einer bestimmten Verzögerungszeit ein Photon gemessen wurde. Die grüne Kurve ist die Summenhäufigkeit. . . . . 421

 ©*2001-2009 Ulm University, Othmar Marti and Alfred Plettl*

4.253Prinzip der Lumineszenzmessung . . . . . 421
4.254Messung einer Lumineszenzabklingkurve . . . . . 422
4.255Auswirkung des *Oversamplings* auf eine Messung. . . . . 423
4.256Vergleich einer Frequenz- und einer Zeitbereichsmessung. . . . . 423
4.257Blockschaltbild eines Spektralanalysators. . . . . 424
4.258Blockschaltbild eines Netzwerkanalysators. . . . . 425
4.259Prinzipieller Aufbau eines Spektrumanalysators. . . . . 426
4.260Durchführung einer Reflexionsmessung mit einem Netzwerkanalysator. 426
4.261Detektionsverfahren für Netzwerkanalysatoren . . . . . 426
4.262Amplitudendetektion. . . . . 427
4.263Phasenmessung. . . . . 427
4.264Amplituden- und Phasengang bei einer linearen Übertragung. . . . . 428
4.265Abschlusswiderstand und reflektierte Signale . . . . . 429
4.266Frequenzgang des Abschlusswiderstandes . . . . . 429
4.267Polardiagramm und Smith-Diagramm . . . . . 430
4.268Polardiagramm und Smith-Diagramm . . . . . 431
4.269S-Parameter. Oben ist die Definition gezeigt, unten das dazugehörige Flussdiagramm . . . . . 431
4.270Übertragungsverhalten einer Tunnelstrecke . . . . . 434
4.271Übertragungskoeffizient und Elektronenenergie . . . . . 436
4.272Schwingungsverhalten und Nichtlinearitäten . . . . . 444
4.273Bauarten von Piezo-Rastereinheiten . . . . . 448
4.274STM-Elektronik: *Blockschema* . . . . . 450
4.275Tunnelspektroskopie mit STM: *Blockschema* . . . . . 451
4.276Distanzspektroskopie im STM: *Blockschema* . . . . . 452
4.277Tunnelspektroskopie mit zwei Tunnelspannungen . . . . . 453
4.278CITS-Tunnelspektroskopie . . . . . 453
4.279Integration von STM und SEM . . . . . 456
4.280Annäherung der Tunnelspitze an die Probe im STM . . . . . 456
4.281*Auflösung* des STM: semiklassische Berechnung . . . . . 458
4.282Einfluss der Spitzenform auf Auflösung . . . . . 461
4.283Aufbau von Graphit . . . . . 462
4.284STM-Bild von Graphit bei 6.8 K . . . . . 463
4.285Potentiometrie mit STM:Aufbau . . . . . 465
4.286STM-Potientiometrie: Resultate . . . . . 466
4.287Rauschen im STM . . . . . 467
4.288STM Bilder mit Rauschmessung bestimmt . . . . . 468
4.289Ballistische Elektronenemission im STM . . . . . 469
4.290Schematische Darstellung eines Tunneluebergangs . . . . . 470
4.291Ersatzschaltschema fuer einen Tunnelübergang . . . . . 470
4.292Potentialverlauf auf der zentralen Insel im klassischen Bild. . . . . 471
4.293Wahrscheinlichkeitsverteilung der Elektronen in Abhängigkeit von $R_T$ . 472
4.294Potentialbild einer Serieschaltung von zwei Inseln . . . . . 472
4.295Strom-Spannungscharakteristik für einen doppelten Tunnelkontakt . . 473
4.296Potentialbild eines Single-Elektron-Transistors. . . . . 474
4.297Ersatzschaltschema für den *Single Electron Transistor*. . . . . 474
4.298Diagramm der Blockade in einem *Single Electron Transistor*. . . . . 474
4.299Geometrie der Abbildung in einem Feldionenmikroskop . . . . . 476

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

4.300Feldemissionsprozess . . . 476
4.301Prinzipieller Aufbau eines Felemissionsmikroskopes . . . 477
4.302Feldionisation an der Spitze eines Feldionenmikroskopes . . . 477
4.303Potentialverlauf bei der Feldionenmikroskopie . . . 478
4.304Vergleich eines feldionenmikroskopischen Bildes mit einem Kugelmodell. Links ist das Kugelmodell für eine $Ni_4Mo$-Verbindung gezeigt. Rechts ist das entsprechende feldionenmikroskopische Bild. . . . 479
4.305Prinzipbild der Atom-Probe- Feldionenmikroskopie . . . 480
4.306Messung der Atommasse über die Laufzeit der felddesorbierten Atome. 480
4.307Prinzip des Projektionselektronenmikroskops . . . 481
4.308Hologramm von Kohlenstofffasern . . . 482
4.309Durchgang von Elektronen durch eine elektrostatische Potentialdifferenz 484
4.310Aufbau einer elektrostatischen Linse . . . 484
4.311Schematischer Aufbau einer elektrostatischen Zylinderlinse . . . 484
4.312Hauptebenen einer elektrostatischen Linse . . . 484
4.313Elektrostatische Einzellinse . . . 485
4.314Zur Einzellinse äquivalenter Aufbau mit optischen Linsen . . . 485
4.315Schema der Wirkungsweise der Lorentzkraft . . . 486
4.316Drahtschlaufe als Linse. Die magnetischen Feldlinien sind so, dass die Lorentzkraft in der Drahtschlaufe die Elektronen wie in einer Linse ablenkt. . . . 487
4.317Bild einer kurzen magnetischen Linse . . . 487
4.318Zylinderkoordinaten zur Berechnung einer magnetischen Linse . . . 488
4.319Geometrische Zusammenhänge in magnetischen Linsen zwischen radialer Geschwindigkeit $v_\rho$ und der axialen Geschwindigkeit $v$ sowie dem Abstand von der optischen Achse $\rho$ und der Brennweite $f$. . . . 488
4.320Links die Feldverteilung $B(z)$ in axialer und $B(r)$ in transversaler Richtung in einer symmetrischen magnetischen Linse mit der Öffnung $D$ und dem Gap $S$). Rechts ist das Schema der Bildformung in einer magnetischen Linse gezeigt. . . . 490
4.321Beziehung zwischen $\phi$ und $z$ . . . 491
4.322Elektronenbahnen in einer Magnetlinse . . . 492
4.323Typische Elektronenbahnen in einer Magnetlinse mit einer Gauss-förmigen Magnetfeldverteilung . . . 492
4.324Sphärische Aberration illustriert durch paraxiale Strahlen . . . 493
4.325Illustration der Kissen- und Trapezverzerrungen . . . 493
4.326Schematischer Aufbau eines Elektronenmikroskopes . . . 494
4.327Aufbau einer Wolframdrahtkathode und rechts ein Detailbild der Kathode 494
4.328Links eine normal durchgebrannte Wolframkathode und rechts eine übersättigte Wolframkathode (mit zu grossem Strom betrieben) . . . 495
4.329Links die Überwachung der Fadensättigung in einem TEM mit Hilfe eines projizierten Fleckes auf einem Phosphorschirm und rechts ein Bild des Fadenstromes eines REM auf einem Oszilloskopschirm. . . . 495
4.330Bild einer $LaB_6$-Quelle. Hier wird eine $LaB_6$-Kristall von Wolfram- oder Rhenium-Drähten gehalten. . . . 495
4.331Bild einer Feldemissionskathode. Dabei wird, analog zu einem STM, eine zugespitzte Wolframnadel als Emitter verwendet. . . . 495
4.332Energieverhältnisse an der Oberfläche einer Elektrode . . . 496

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

4.333Feldemission . . . 496
4.334Wehnelt-Zylinder zur Erzeugung einer kleineren Emissionsfläche . . . 496
4.335Everhart-Thornley-Detektor. Die ringförmige Struktur aussen ist der Faraday-Käfig, hinten befindet sich der Szintillator-Kristall . . . 497
4.336Bilderrasterung im REM . . . 497
4.337Verbesserung der Bildqualität durch Mittelung. Links ist eines der Originalbilder. Die Mitte zeigt das Resultat der Mittelung über 15 Bilder. Rechts ist eine Messung mit optimaler Einstellung gezeigt. . . . 498
4.338Streuung von Elektronen aus dem Strahl eines REMs im Innern einer Probe . . . 498
4.339Wechselwirkungszone für Elektronen im REM . . . 499
4.340Abhängigkeit des Wechselwirkungsvolumens von der Beschleunigungsspannung. . . . 499
4.341Abhängigkeit der Wechselwirkungszone von der Ordnungszahl . . . 499
4.342Austrittstiefe von Sekundärelektronen . . . 500
4.343Die vier an der Bildgebung beteiligten Prozesse . . . 500
4.344Herkunft der Sekundärelektronen im REM . . . 501
4.345Austrittstiefe rückgestreuter Elektronen . . . 501
4.346Sekundärelektronenbild (links) und Rückstreubild (rechts) . . . 502
4.347Vergleich der REM-Bilder als Funktion der Oberflächenbeschaffenheit. Links ist ein Sekundärelektronenbild der polierten Oberfläche eine antiken griechischen Münze. In der Mitte befindet sich das Rückstreubild, während rechts ein Sekundärelektronenbild einer Bruchfläche gezeigt ist. 502
4.348Links ist ein Kathodolumineszenzbild einer geologischen Probe gezeigt. Rechts ist zum Vergleich dazu ein Rückstreubild gezeigt. . . . 502
4.349Schematische Übersicht über die Zahl der photoemittierten Elektronen. Oben wird das Spektrum der Sekundärelektronenprozesse gezeigt. Unten sind, von links, die Sekundärelektronenanregung und die dabei auftretenden Energieverluste, Emission aus Rumpfniveaus, Augerprozesse und die Emission aus dem Valenzbandbereich dargestellt. . . . 503
4.350Ausbeute von Augerelektronen als Funktion der Ordnungszahl . . . 503
4.351Elektroneneenrgien beim Auger-Prozess als Funktion der Ordnungszahl 503
4.352Elektronenchanneling. Links ist das Prinzip gezeigt, rechts eine daraus resultierende Abbildung. . . . 505
4.353Elektronenchannelingbild einer kubisch-raumzentrierten Aluminium/Eisen-Legierung. . . . 505
4.354Aufnahme von Probenstrombildern im REM. Links ist die dazugehörige Schaltung gezeichnet. In der Mitte befindet sich ein Probenstrombild von Felsit, rechts das dazugehörige Rückstreubild. . . . 506
4.355Potentialkontrastbild. Links ist ein Sekundärelektronenbild gezeigt, rechts der gleiche Chip mit angelegter Spannung. . . . 506
4.356Elektronenstrahlinduzierte Ströme (EBIC) . . . 506
4.357Brennfleckgrösse und Strahlstrom . . . 507
4.358Brennfleckgrösse und *Auflösung* . . . 507
4.359Einfluss der Brennfleckgrösse auf *Auflösung* und Bildqualität. Links ist eine Abbildung mit dem kleinstmöglichen Brennfleck, in der Mitte mit einem etwas grösseren Brennfleck und links mit einem ganz grossen Brennfleck. . . . 507

4.360Aperturgrösse und Schärfentiefe . . . 508
4.361Arbeitsabstand und Schärfentiefe . . . 508
4.362Wirkung der Aperturblende und des Arbeitsabstandes. Die Einstellungen sind: oben links: Aperturblende 600 $\mu$m, Arbeitsabstand 15 mm, oben rechts: Aperturblende 200 $\mu$m, Arbeitsabstand 15 mm, unten links: Aperturblende 100 $\mu$m, Arbeitsabstand 15 mm, unten rechts: Aperturblende 100 $\mu$m, Arbeitsabstand 39 mm. . . . 509
4.363Beeinflussung der *Auflösung* durch Aperturblende und Arbeitsabstand. Die Einstellungen sind: oben links: Aperturblende 100 $\mu$m, Arbeitsabstand 15 mm, oben rechts: Aperturblende 200 $\mu$m, Arbeitsabstand 15 mm, unten links: Aperturblende 600 $\mu$m, Arbeitsabstand 15 mm, unten rechts: Aperturblende 200 $\mu$m, Arbeitsabstand 39 mm. . . . 509
4.364Wirkung der Beschleunigungsspannung auf die Bildqualität. Oben links: Drähte aus reinem Gold mit 35 kV Spannung fotografiert. Unten links das gleiche, aber mit 10 kV Spannung. Oben rechts: Eichenpollen bei 35 kV, unten rechts: Eichenpollen bei 10 kV . . . 510
4.365Probenhalter für REM . . . 510
4.366Aufladen von Proben im REM. Links: Flugasche, bei der Aufladung zu anomalem Kontrast führt. Mitte: menschliche rote Blutkörper mit Linien erzeugender Aufladung. Rechts: Schistosom, die Aufladung führt zu einer Ablenkung des Strahls. . . . 511
4.367Kathodenzerstäuber. Links ist das Bild eines typischen Zerstäubers und rechts eine Skizze der Arbeitsweise. . . . 511
4.368Elektronenemissionskoeffizient als Funktion der Beschleunigungsspannung . . . 511
4.369Abbildung von Salz bei verschiedenen Beschleunigungsspannungen (von links: 1 kV, 5kV, 15kV). . . . 512
4.370Auger-Prozesse (links) und Röntgenprozesse (rechts) . . . 512
4.371Wechselwirkung zwischen Elektronenstrahl und Probe . . . 512
4.372Räumliche *Auflösung* bei der EDX-Abbildung . . . 513
4.373EDS-System angeschlossen an ein REM . . . 513
4.374Schematische Darstellung eines EDS-System . . . 513
4.375Funktion eines EDS-Detektors . . . 514
4.376EDS und REM-Analyse von Chromiterz. Oben links: EDS-Spektrum mit Peaks für Magnesium, Aluminium, Silizium, Kalzium, Chrom und Eisen. Oben rechts: Rückstreuelektronenbild. Unten links: Punktdichtebild von Silizium. Unten rechts: Punktdichtebild von Chrom. . . . 514
4.377Optimierung von Detektorabstand und Aufnahmewinkel für die Röntgen-Mikroanalyse . . . 515
4.378Optimierung von Probenneigung und Aufnahmewinkel für die Röntgenmikroanalyse . . . 515
4.379Optimierung von Detektorabstand und Raumwinkel für die Röntgenmikroanalyse . . . 515
4.380Detektorabstand und Raumwinkel bei grossem Abstand . . . 515
4.381Systempeaks aufgrund der Fluoreszenz der letzten Aperturblende . . . 516
4.382Rückgestreute Elektronen und Systempeaks . . . 516
4.383Escape-Peak von Silber . . . 516
4.384Sekundärfluoreszenz . . . 517

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

4.385Entfalten von Peak-Überlapps . . . 517
4.386Abbildung von Gitternetzebenen . . . 518
4.387Abbildung von Gitternetzebenen und der Oberflächenstruktur von *Si(111)-(7x7)* . . . 519
4.388Reales Gitter (links) und reziprokes Gitter (rechts). . . . 521
4.389Skizze zur Streuung an Oberflächenatomen . . . 521
4.390Ewald-Konstruktion für Oberflächennetze. Rechts wird ein Schnitt dargestellt. . . . 523
4.391Aufbau eines LEED. Links ist die Elektronenkanone gezeigt. Rechts ist der schematische Aufbau des LEED-Schirms gezeigt. . . . 523
4.393Eindringtiefe der Elektronen als Funktion der Energie . . . 524
4.394Ewaldkonstruktion für LEED . . . 524
4.392Energieverlauf im LEED-Detektor. Rechts ist der Zwischenraum zwischen der Probe und dem Detektor. . . . 524
4.395Beugungsmuster und Definitionen zur Transferweite . . . 525
4.396Schematische Skizze eines LEED-Bildes von $Cu(110)$ (gezeichnet nach [?]). Dies ist eine FCC-Struktur. . . . 526
4.397Schematische Skizze eines LEED-Bildes von Ni (111) bei einer Primärenergie von 205 eV (gezeichnet nach [?]). . . . 526
4.398LEED-Bild von *Si(111)* $\sqrt{19} \times \sqrt{19}$ . . . 526
4.399Mosaikstruktur links im realen Raum und rechts im reziproken Raum . 528
4.400Beispiel: Regelmässige Domänen einer Kristallstruktur mit der Periode $a$ jeweils im Abstand von $(N + 1/2)a$ . . . 529
4.401Mögliche Wege der Elektronen bei der Streuung an Oberflächen. . . . 530
4.402Geometrie bei der RHEED-Abbildung. Links ist dargestellt, wie die Trajektorien der Elektronen angeordnet sind. Rechts ist gezeigt, wie die Ewald-Konstruktion zum Auffinden der RHEED-Reflexe angewandt werden. . . . 531
4.403Konstruktion der Reflexe bei RHEED. Links werden die Strukturen der Oberflächen gezeigt. Rechts ist das entsprechende RHEED-Bild. Von oben nach unten wird die Abbildung bei idealer Oberfläche, bei schwach gestuften Oberflächen, bei einer Mosaikstruktur, stark gestufte Oberflächen sowie 3D-Inseln . . . 532
4.404RHEED-Kontrolle des Schichtwachstums zum Beispiel bei der Herstellung von Halbleiter-Quantenschichten. . . . 532
4.405Lennard-Jones-Potential . . . 535
4.406SFM-Spitze aus gespaltenem Diamant . . . 537
4.407Eigenschaften von Schwingungsbalken . . . 540
4.408Prinzip des Rasterkraftmikroskopes . . . 541
4.409Interferometrische Kraftdetektion . . . 542
4.410Heterodyn-interferometrische Kraftdetektion im SFM . . . 544
4.411Faseroptisch-interferometrische Kraftdetektion im SFM . . . 546
4.412Nomarsky-Interferometer zur Kraftdetektion . . . 547
4.413Aufbau eines Lichtzeiger-SFMs . . . 548
4.414Kalibration eines Lichtzeiger-SFMs . . . 549
4.415Kombiniertes Rasterkraft- und -reibungsmikroskop . . . 549
4.416Rasterkraftmikroskopische Bilder von Glimmer . . . 553
4.417Rasterkraftmikroskopische Bilder eines Gitters . . . 555

©2001-2009 Ulm University, Othmar Marti and Alfred Plettl 

4.418Kraftmikroskopische Abbildung einzelner Ladungen . . . . . 556
4.419Magnetic Force Microscopy . . . . . 557
4.420Beispiel des prinzipiellen Aufbaus eines Messsystems mit externen Geräten . . . . . 558
4.421Seckerbelegung beim IEEE-488-Bus beziehungsweise beim IEC-625-Bus 560
4.422Struktur eines Datenerfassungssystems mit dem IEEE-488-Bus . . . . . 562
4.4239-polige (links) und 25-polige Anschlüsse für die RS-232-Schnittstelle. . 562
4.424Verbindung zweier Rechner . . . . . 563

A.1 Schematische Darstellung einer elektromagnetischen Welle. . . . . 569

E.1 Amplituden- und Phasengang kritisch gedämpfter Tiefpassfilter . . . . 579
E.2 Phasenbild und Gruppenlaufzeit von kritisch gedaempften Tiefpassfiltern580
E.3 Amplituden- und Phasengang von Butterworth Tiefpassfiltern . . . . . 580
E.4 Phasenbild und Gruppenlaufzeit von Butterworth Tiefpassfiltern . . . . 581
E.5 Amplituden- und Phasengang von Bessel-Tiefpassfiltern . . . . . 581
E.6 Phasenbild und Gruppenlaufzeiten von Bessel-Tiefpassfiltern . . . . . 582
E.7 Amplituden- und Phasengang von 1dB-Tschebyscheff-Tiefpassfiltern . . 582
E.8 Phasenbild und Gruppenlaufzeit von 1dB-Tschebyscheff-Tiefpassfiltern 583
E.9 Amplituden- und Phasengang von 3dB-Tschebyscheff-Tiefpassfiltern . . 583
E.10 Phasenbild und Gruppenlaufzeit von 3dB-Tschebyscheff-Tiefpassfiltern 584
E.11 Amplituden- und Phasengang kritisch gedämpfter Hochpassfilter . . . . 585
E.12 Phasenbild und Gruppenlaufzeit von kritisch gedaempften Hochpassfiltern . . . . . 586
E.13 Amplituden- und Phasengang von Butterworth Hochpassfiltern. . . . . 586
E.14 Phasenbild und Gruppenlaufzeiten von Butterworth Hochpassfiltern. . 587
E.15 Amplituden- und Phasengang von Bessel-Hochpassfiltern. . . . . 587
E.16 Phasenbild und Gruppenlaufzeiten von Bessel-Hochpassfilter. . . . . 588
E.17 Amplituden- und Phasengang von 1dB-Tschebyscheff-Hochpassfiltern . 588
E.18 Phasenbild und Gruppenlaufzeiten von 1dB-Tschebyscheff-Hochpassfiltern589
E.19 Amplituden- und Phasengang von 3dB-Tschebyscheff-Hochpassfiltern . 589
E.20 Phasenbild und Gruppenlaufzeiten von 3dB-Tschebyscheff-Hochpassfiltern mit 3dB Welligkeit. . . . . 590
E.21 Amplituden- und Phasengang von kritisch gedämpften Bandpassfiltern. 591
E.22 Phasenbild und Gruppenlaufzeiten von kritisch gedämpften Bandpassfiltern. . . . . 592
E.23 Amplituden- und Phasengang von Butterworth Bandpassfiltern. . . . . 592
E.24 Phasenbild und Gruppenlaufzeiten von Butterworth Bandpassfiltern. . 593
E.25 Amplituden- und Phasengang von Bessel-Bandpassfiltern. . . . . 593
E.26 Phasenbild und Gruppenlaufzeiten von Bessel-Bandpassfiltern. . . . . 594
E.27 Amplituden- und Phasengang von 1dB-Tschebyscheff-Bandpassfiltern . 594
E.28 Phasenbild und Gruppenlaufzeiten von 1dB-Tschebyscheff-Bandpassfiltern . . . . . 595
E.29 Amplituden- und Phasengang von 3dB-Tschebyscheff-Bandpassfiltern . 595
E.30 Phasenbild und Gruppenlaufzeiten von 3dB-Tschebyscheff-Bandpassfiltern . . . . . 596
E.31 Amplituden- und Phasengang von kritisch gedämpften Bandsperrenfiltern597
E.32 Phasenbild und Gruppenlaufzeiten von kritisch gedämpften Bandsperrenfiltern . . . . . 598

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

E.33 Amplituden- und Phasengang von Butterworth Bandsperrenfiltern . . . 598
E.34 Phasenbild und Gruppenlaufzeiten von Butterworth Bandsperrenfiltern 599
E.35 Amplituden- und Phasengang von Bessel-Bandsperrenfiltern . . . . . 599
E.36 Phasenbild und Gruppenlaufzeiten von Bessel-Bandsperrenfiltern . . . 600
E.37 Amplituden- und Phasengang von 1dB-Tschebyscheff-Bandsperrenfiltern 600
E.38 Phasenbild und Gruppenlaufzeiten von 1dB-Tschebyscheff-Bandsperrenfiltern 601
E.39 Amplituden- und Phasengang von 3dB-Tschebyscheff-Bandsperrenfiltern 601
E.40 Phasenbild und Gruppenlaufzeiten von Tschebyscheff-Bandsperrenfiltern 602
E.41 Amplituden- und Phasengang von maximal flachen Allpassfiltern. . . . 603
E.42 Phasenbild und Gruppenlaufzeiten von Allpassfiltern . . . . . 604
E.43 Amplituden- und Phasengang von Schwingkreisen . . . . . 605
E.44 Phasenbild und Gruppenlaufzeiten eines Schwingkreises . . . . . 606

G.1 *Smith-Chart* mit Gnuplot erstellt . . . . . 625
G.2 Plot einer Datendatei *Smith-Chart* mit Gnuplot 4.2 . . . . . 628
G.3 *Polar-Plot* mit Gnuplot erstellt . . . . . 629
G.4 *Polar-Plot* und *Smith-Chart* in einer mit Gnuplot erstellten Darstellung. 630

J.1 Filter im Realraum . . . . . 641
J.2 Fourier-Transformation von Bildern . . . . . 642
J.3 Filterung über Einheitszellen . . . . . 643
J.4 Liniendarstellung der Topographie . . . . . 644
J.5 Gitterdarstellung der Topographie . . . . . 645
J.6 Grauskalendarstellung der Topographie . . . . . 645
J.7 Beleuchtete Darstellung der Topographie . . . . . 646
J.8 3D-Darstellung der Topographie . . . . . 646
J.9 Indizierung von Punkten für Filter . . . . . 647

K.1 Koordinatensystem zur Korrektur linearer Verzerrungen . . . . . 649
K.2 Korrektur von 3D-Verzerrungen . . . . . 650

L.1 Skizze zur Bestimmung der Miller'schen Indizes . . . . . 651
L.2 Basisatome einer Netzzelle eines Oberflächennetzes . . . . . 652
L.3 Atomare Anordnung bei Überstrukturen . . . . . 653
L.4 Struktur einer $Si(111)(\sqrt{3} \times \sqrt{3})R30°$-Rekonstruktion. . . . . 654
L.5 Beispiel einer Rekonstruktion, deren Bezeichnung nicht eindeutig festgelegt werden kann. . . . . 654

# Tabellenverzeichnis


2.1 Algebra mit Blockdiagrammen: Kombination von Blöcken . . . . . 15
2.2 Algebra mit Blockdiagrammen: Summationspunkte verschieben . . . . 16
2.3 Algebra mit Blockdiagrammen:Abnahmepunkte verschieben . . . . . 17
2.4 Vergleich der *Zahlensysteme* . . . . . 42
2.5 Logische Verknüpfungen mit einer Eingangsvariablen . . . . . 42
2.6 Wahrheitstabelle des Äquivalenzgatters . . . . . 46
2.7 Postulate der Schaltalgebra . . . . . 46
2.8 Theoreme der Schaltalgebra . . . . . 47
2.9 Wahrheitstabelle des ersten DeMorganschen Gesetzes . . . . . 48
2.10 Wahrheitstabelle des zweiten DeMorganschen Gesetzes . . . . . 48
2.11 Erzeugung einer DNF aus der Wahrheitstabelle . . . . . 49
2.12 Erzeugung einer KNF aus der Wahrheitstabelle . . . . . 50
2.13 Umrechnung der Vierpolparameter . . . . . 54
2.14 Butterworth-Polynome . . . . . 62
2.15 Tschebyscheff-Polynome . . . . . 62
2.16 Koeffizienten der Tschebyscheff-Polynome . . . . . 63
2.17 Tabelle der Besselpolynome . . . . . 64
2.18 Rausch- und Signalbandbreite von Filtern . . . . . 97

3.1 Inversionsdichten $n_i$ einiger Materialien bei Raumtemperatur. . . . . 113
3.2 Energetischer Abstand einiger Donatorenniveaus . . . . . 115
3.3 Energetischer Abstand einiger Akzeptorniveaus . . . . . 115

4.1 Vergleich von Wechselspannungsmessmethoden . . . . . 231
4.2 Ausgangswerte von Digital-Analog-Wandlern . . . . . 262
4.3 Erhöhung der Bitzahl durch oversampling . . . . . 271
4.4 Auflösung $\varepsilon$ der Periodendauermessung . . . . . 285
4.5 Genauigkeit der Frequenzmessung . . . . . 287
4.6 Verhältnismessung der Frequenz . . . . . 289
4.7 Eigenschaften homogener Leitungen . . . . . 323
4.8 Koppeldämpfung bei Faser-Faser-Kopplung . . . . . 348
4.9 Uebliche Fehlerquellen und Massnahmen . . . . . 361
4.10 Klassifizierung von Resonatoren . . . . . 386
4.11 Erzeugung kurzer Pulse . . . . . 399
4.12 Aufloesungsvermoegen eines Fabri-Perot . . . . . 416
4.13 Breiten von Autokorrelationsfunktion und Laserpuls . . . . . 419
4.14 Elastische Materialkonstanten . . . . . 444
4.15 Probentopographie und Auflösung . . . . . 461
4.16 Haeufig vorkommende Bussysteme . . . . . 559
4.17 Bevorzugte Bussysteme zur Messdatenerfassung . . . . . 560

4.18 Bussysteme in Rechnern . . . . . 564
4.19 Beispiele grafischer Datenerfassungsprogramme . . . . . 566

C.1 Tabelle der Laplacetransformationen einiger ausgewählter Funktionen . 573
C.2 Tabelle der Carson-Heaviside-Transformationen einiger ausgewählter Funktionen . . . . . 574
C.3 Tabelle der z-Transformationen einiger ausgewählter Funktionen . . . . 575
C.4 Einstellzeiten zu einer vorgegeben Genauigkeit als Funktion der Zeitkonstante $\tau$ . . . . . 576

D.1 Kenngrössen von Tiefpassfiltern . . . . . 577

F.1 Filterkoeffizienten für kritisch gedämpfte Filter . . . . . 607
F.2 Filterkoeffizienten für Besselfilter . . . . . 608
F.3 Filterkoeffizienten für Butterworth-Filter . . . . . 609
F.4 Filterkoeffizienten für Tschebyscheff-Filter mit 0.5 dB Welligkeit . . . . 610
F.5 Filterkoeffizienten für Tschebyscheff-Filter mit 1 dB Welligkeit . . . . . 611
F.6 Filterkoeffizienten für Tschebyscheff-Filter mit 2 dB Welligkeit . . . . . 612
F.7 Filterkoeffizienten für Tschebyscheff-Filter mit 3 dB Welligkeit . . . . . 613
F.8 Filterkoeffizienten für Allpassfilter mit maximal flacher Gruppenlaufzeit 614

H.1 Benchmark-Resultate für einen DSP . . . . . 634

I.1 Materialeigenschaften für Isolatoren . . . . . 636
I.2 Thermoelektrische Koeffizienten . . . . . 636
I.3 Thermoelektrische Spannungsreihe und Seebeck-Koeffizienten . . . . . 636
I.4 Temperaturkoeffizient des Widerstandes . . . . . 636

J.1 3x3 convolution kernels . . . . . 640
J.2 Usage of filters . . . . . 644

# Literaturverzeichnis

# Index


Äquivalenz, 46
Äquivalenz-Gatter, 45
Überabtastung, 422
Übergang
  hetero, 132–136
  p-n, 125–129
    vorgespannt, 129–132
Überstruktur
  inkommensurabel, 655
  kommensurabel, 655
Übertragungsfunktionen, 21–22
1-Bit Wandler, 270
1/f-Rauschen, 95, 374

105, 221, 225
111, 221
741, 221, 224
1436, 225
4040, 261

Abtast-Halte-Glied, 263
Abtast-Halteglied, 68, 294
Abtasttheorem, 68
AD630, 283, 284
Adatom, 441
  Sodium, 441
  Sulphur, 441
ADC, 272–281
AFM
  Cantilever, 252
AFPSA, 405
Akusto-optischer Modulator, 393
Akzeptanzwinkel, 343
Akzeptoren, 114, 115, 117, 119, 125, 135, 139, 140, 146, 176, 312
Algebra
  Boole, 46–49
Allpass, 66–67
ALU, 103
amorphes $SiO_2$, 143

Analog-Digital-Wandler, 272–278
Analog/Digital-Wandler, 272–281
Analogfilter, 60–67
Analogsignal, 67, 68
AND, 43
Anodenstrom, 95
Antivalenz, 47
Antivalenz-Gatter, 45
Arbeitspunkt, 53
Arithmetisch-logische Einheit, 103
Atom
  Si, 114
Atom-Probe Feldionenmikroskopie, 480
Atomlage, 134
Aufbau
  planar, 157
Auflösung, 268, 271, 272, 274, 285, 287, 376, 412, 414, 416, 498, 500, 507–509, 512, 513, 518
  spektral, 409, 414
  zeitlich, 417, 419, 420
Auflösung, 477
Ausgangskennlinie, 160
Ausgangssignal, 16
Ausgangswiderstandes, 362
Austrittsarbeit, 138, 140, 316, 317, 478, 494
Avalanche-Effekt, 321

Bänderschema, 125, 134, 137–139, 141, 144, 145, 147, 149
  elektronisch, 158, 169
  Heteroübergang, 135
  Heterostruktur, 134
Band, 134, 135, 137–139, 141, 144, 145, 147, 149, 158
Banddiskontinuitäten, 134
Bandpass, 65
Bandsperre, 66
Bandverbiegung, 134

Basis, 169
Basismessverfahren, 217–284
Basisschaltung, 169, 173, 199
Basisvektoren, 652
Baublock, 11
Bauelement
  Halbleiter, 120
  Si, 122
Bauelemente, 105–216
  Aufbau, 148–181
  bipolare, 154
  Dreitor–, 154
  Funktion, 148–181
  Grundschaltungen, 181–208
  ladungsgekoppelt, 148–150
  linear, 181–184
  optoelektronisch, 175–179
  passiv, 181–184
  Technologie, 148–181
  unipolare, 154
Baulelemente
  Halbleiter, 148–181
BCCD, 150
Bessel-Tiefpass, 63–64
Besselfilter, 63–64
Beugungsverluste, 381
Bias-Strom, 221
Bildaufnahmeeinheiten, 150
Bildsensor, 148
Bindung
  kovalente, 107
Bipolare Transistoren, 169–175
Bitstream-Verfahren, 272
Block
  Kaskadierung, 14
Blockschema, 11–16, 274
  Rechnen, 14–16
  Symbole, 11–14
Bode-Diagramm, 228
Bolometer, 316
Boolesche Algebra, 46
Boolsche Algebra, 46–49
Brückenschaltung, 251–259
Bragg-Gitter in Fasern, 349
Bragg-Spiegel
  sättigbar, 404, 405
Bravais-Gitter, 651
Bravais-Netz, 651
Brechungsindex, 329, 340, 342–344, 350, 351, 353, 356, 390, 395, 402, 403, 405
Brownsche Bewegung, 90
Buffer, 42
burried channel CCD, 150
Butterworth
  Polynom, 62
  Tiefpass, 61
Butterworth-Tiefpass, 61–62

Cantilever, 252, 538
cantilever, 538, 539, 541–549, 551, 552, 554–556
Cavity Dumping, 401
CCD, 148–150
CCD–Kamera, 148
CdS, 319
CdSe, 319
Channeling, 518
Chopper-Rad, 408, 409
CMOS, 161–168
Computerarchitektur, 97
Coulomb–Barriere, 181
Coulomb-Blockade, 472
Coulombblockade, 470
CPM-Laser, 401
CPM-Lasersystem, 401
Cryogenic STM, 457
Cs, 317
Curie-Temperatur, 313
Current Imaging Tunneling Spectroscopy, 453

Dämpfung, 341
DAC, 261–272
Darlington–Schaltung, 200
Darlington–Transistor, 201
Darlington-Transistor, 200, 201
Darstellung
  Blockschema, 11–16
  elektronische Messsysteme, 11–16
dB, 341
  deziBel, 341
Debye-Temperatur, 307
Deglitcher-Schaltung, 263
Dehnungsmessstreifen, 252
DeMorgansche Gesetze, 47
Demultiplexer, 263

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Detektor
phasenempfindlich, 427
deziBel, 341
dB, 341
Diagramm
Karnaugh, 49–51
Signalfluss, 16–20
Allgemeine Formel, 20
Begriffe, 19–20
Grundlagen, 17–19
Dichte
Ladungsträger, 114, 117, 121, 125, 140, 154, 160, 298, 299
injiziert, 123
Leitungsband, 114
Differenzverstärker, 207
Diffusionskoeffizient, 480
Diffusionskonstante, 122
Diffusionslängen, 122
Digital Signal Processor, 102
Digital-Analog-Wandler, 261, 263, 274, 276
digital-analog-Wandler, 68
Digital/Analog-Wandler, 261–272
Digitale Signale, 25, 41–51
Digitale Signalprozessoren, 97–103
Digitalfilter, 67–86
Abtastung, 67–69
Bausteine, 69–73
FIR, 76–81
Grundstrukturen, 73–76
IIR, 82–86
Digitalsignal, 67, 68
Diode, 130
1N4148, 228
1N914, 228
Anwendungen, 185
dynamisches Verhalten, 185
Esaki–Tunnel, 187
Foto
pin, 176
pn, 176
Kapazität, 187
MIS, 142–148
ideal, 144
MOS, 142–148
npn, 206
p–n
Herstellung, 127
PIN, 188
Schalt, 187
Schottky, 136–142, 187
Si, 135, 184
Zener, 187
Dioden, 129–132
Diodenmodelle, 185
Disjunktion, 44
Dispersion, 402
Dispersionskompensation, 402–405
Distributivgesetz, 47
disjunktiv, 47
konjunktiv, 47
Donatoren, 114, 115, 117, 119, 125, 135, 138, 140, 158, 176
Dotierkonzentration, 117–120
Dotierung, 114–116, 127
Drain, 154, 473
Drainschaltung, 204
Dreiecksschaltung, 59
DSP, 97–103
Digital Signal Processor, 102
Dual-Slope-Prinzip, 277

Ebene Welle, 322
ebene Welle, 381
effektive Masse, 111
Eingangsimpedanz, 57
Eingangssignal, 16
Eingangsstrom, 362
Eingangsverstärker, 11
Eingangswiderstand, 219, 326, 329, 360, 364
Einkristall
Si, 114
Einmoden-Gklasfaser, 350
Einschaltvorgang, 37–41
Einsteinbeziehung, 311
Electronics, 433, 441, 449–453, 464, 467, 468, 550, 638
elektrische Kontakte, 124–148
Grundlagen, 124–125
p-n-Übergang, 125–129
Elektromotorische Kraft, 319
Elektronenbeugung, 518
Elektronengas
2-dimensional, 159
Elektronenstrahllithographie, 159

Elektronik-Design-Labor, 217
Elementarladung, 95
EMF, 89, 90
  fluktuierend, 89, 90
Emitter, 169
Emittereffizienz, 171
Emitterfolger, 197
Emitterschaltung, 173, 192
  Spannungsgegenkopplung, 196
  Stromgegenkopplung, 194
Energie
  Fermi, 106, 117, 125, 144, 159, 307, 316, 317, 319, 473, 478
Energielücke, 106, 108, 111, 117, 134, 175
  Si, 148
Ensemblemittel, 657
Entropie, 90
Error signal, 452
Esaki–Tunnel–Diode, 187
Ewald-Konstruktion, 524
Exklusiv-NOR, 46

F, 340
Fabri-Perot-Interferometers, 415
Fabri-Perot-Spektrometer, 414
Fabry-Perot-Resonator, 358, 376
Faltung, 34
Faltungssatz, 34
Faradaybecher, 480
Fasern
  Bragg-Gitter, 349
Fehlersignal, 16
Felddesorption, 479
Feldeffekt-Transistor, 154
Feldeffekt-Transistoren, 154–160
Feldionenmikroskop, 477
feldionenmikroskopische Abbildung, 479
Feldplatte, 300
Fermi-Energie, 106, 117, 125, 144, 159, 307, 316, 317, 319, 473, 478
Fermi-Niveau, 125
Festkörperelektronik, 105
FET, 154, 473
  Double–Gate JFET, 157
  Grundschaltung, 203
  Sperrschicht–, 155
Filter, 60–86
  analog, 60–67
  Bessel, 63–64
  digital, 67–86
    Abtastung, 67–69
    Bausteine, 69–73
    FIR, 76–81
    Grundstrukturen, 73–76
    IIR, 82–86
  Finite Impulse Response, 76
  FIR, 76
  IIR, 76
  Infinite Impulse Response, 76
  Lineare Phase, 77
Finesse, 389, 416
Finite Impulse Response Filter, 76
FIR Filter, 103
FIR-Filter, 76–81
FireWire, 559
Flächengruppe, 652
Flash-Converter, 274
Flickerrauschen, 95
Flip–Flop, 161
Fluktuationen, 89
Fluktuations-Dissipations-Theorem, 90
FM-Übertragung, 88
Fotodiode
  pin, 176
  pn, 176
Fotodioden, 176
  Avalanche, 177
Fotoeffekt
  innerer, 175
Fotoelement, 177
Fotoleitfähigkeit, 175
Fotoleitung, 176
  Störstelle, 176
Fototransistoren, 177
Fourieranalysatoren, 423
Fourieroptik, 382
Fouriertransformation, 25–32, 34, 40, 69, 89, 400, 423, 518, 528, 529
Frame–Transfer–Konzept, 150
Frequenz, 285, 289–295
Frequenzgang–Kontrolle
  universell, 215
Frequenzgang–Korrektur, 215
Frequenzmessung, 285
  PLL, 292
  Quartz, 289–295

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Fresnel-Zahl, 384
Fresnelzahl, 381
Fuzzy-Logik, 25

Güteschaltung, 393
GaAa
  MESFET, 157
GaAs, 106, 109, 113, 119–122, 134, 148, 155, 157, 176–178, 300, 467
  i, 159
  MESFET, 157–159
  Substrat
    intrinsisch, 158
  VCSEL, 180
Gate, 154, 473
Gate–Steuerspannung, 155
Gateoxid, 105
Gateschaltung, 205
Gatter
  Äquivalenz, 45
  Antivalenz, 45
  NAND, 44
  Nicht, 43
  NOR, 44
  oder, 43
  Und, 43
  XNOR, 46
  XOR, 47
Ge, 321, 340
Generations-Rekombinationsrauschen, 95
Generationsraten, 123
Gitterspektrometer, 412
Glasfaser
  einmodig, 350
  Numerische Apertur, 343
Gleichgewicht
  thermodynamisch, 125
Gleichrichtung, 129–132, 185
Graded base band gap technique GBT, 175
Graetz-Schaltung, 236
Graphite, 440, 451, 460, 462–465, 534–536, 538, 553, 554, 638
Gray-Code, 50
Grenzflächenzustand, 140
GRIN-Linsen, 345
Grundlagen
  mathematisch, 11–103
Grundschaltung
  Bipolartransistor, 192
  FET, 203
  Transistor, 192
Grundschaltungen
  Bauelemente, 181–208
Gruppenlaufzeit, 63–67, 77, 402, 428, 580–584, 586–590, 592–596, 598–602, 604, 606
Gruuppe
  Fläche, 652

H-Matrix, 53
Halbleiter, 105, 106, 108, 110, 113, 116–122, 134, 136, 138–140, 144, 145, 155, 175, 179
  amorph, 105
  Bauelement, 120–124
  direkt, 109, 121, 148
  dotiert, 114–116
    schwach, 122
    stark, 122
  Einkristall, 105, 124
  Grundlagen, 105–124
  Heteroübergang, 135
  homogen, 116
  ideal, 110
  II-VI, 109
  III-V, 109, 158, 159
  indirekt, 109, 121
  intrinsisch, 110–113, 122, 175
  Ladungsträgerdichte, 116–117
  Material, 113, 155, 175, 179
    intrinsisch, 175
  Messtechnik, 139
  n-dotiert, 114, 123, 138, 140, 144, 174
  p-dotiert, 114, 140, 144
  quaternär, 106
  Schichten
    Epitaxie, 124
  Technologie, 161
  undotiert, 140
  Volumen, 123, 141
Halbleiter–Laser, 178
Halbleiter-Bauelemente, 148–181
Halbleiter-Fotodetektoren, 176–177
Halbleiter-Metall, 124

Halbleiter-Speicher, 161–168
Halbleiter-Strahlungsquellen, 177–179
Hall-Effekt, 298
Hamiltonian, 433, 439, 441, 462
Heaviside-Funktion, 39
Hermitsche Polynome, 383
Hetero-Übergang, 132–136
Heteroübergang
  isotyp, 135
  modulationsdotiert, 136
High-Bit-Verfahren, 272
Hochpass, 65
Hopping-Prozess, 310

IEC-625-Bus, 560
IEEE-488, 559
IEEE1394, 559
II–VI–Halbleiter, 109
III–V–Halbleiter, 109
IIR Filter, 103
IIR-Filter, 76, 82–86
Impedanz, 21, 22, 52
  Quelle, 57
Impedanzanpassung, 328
Impedanztransformation, 329
INA105, 221, 225
InAs, 300
Induktivität, 90, 91, 249–251
Induktivitätsbelag, 333
Infinite Impulse Response Filter, 76
Infrarot, 319
Injektion von Elektronen, 140
Injektion von Löchern, 140
inkommensurable Überstruktur, 655
innerer Fotoeffekt, 175
input signal, 548
InSb, 300, 319, 321
Integrator, 73
Interferometer
  Lloyd, 352
  Prisma, 352
interferometer, 543
Interferometrie, 351
Interline–Konzept, 150
Inversionsschicht, 140
Inverter, 43
Ir, 307
isochrone Übertragung, 559
Isolationswiderstand, 363
Isolatoren, 105, 106

Jellium, 441
  Metal, 441
  Surface, 441
Jitter, 423
Josephson-Effekt, 304

K, 317
Kanalbreite, 155
Kapazität, 249–251
Kapazitätsbelag, 333
Kapazitätsdiode, 187, 423
Karnaugh-Diagramm, 49
Karnaugh-Diagramme, 49–51
Kaskadierung
  Blöcke, 14
  Kommutativgesetz, 14–15
Kaskodenschaltung, 206
Khintchine, 28
Kirchhoff-Fresnel, 382
Knoten, 383
Koaxialleitung, 322
Kohärenz, 525
Kollektor, 169
Kollektorschaltung, 197
kommensurable Überstruktur, 655
Kommutativgesetz, 47
  Kaskadierung, 14–15
Kompositionsgitter, 136
Kondensator, 21, 91
konfokaler Resonator, 383
Konjunktion, 43
Konstruktion
  Ewald, 524
Kontakt
  elektrisch, 124–148
    Grundlagen, 124–125
    p-n-Übergang, 125–129
  injizierend, 140
  Metall und n-Halbleiter, 138
  Metall und p-Halbleiter, 139
  Metall-Halbleiter, 136–142
  Metall-Isolator-Halbleiter, 142–148
  ohmsch, 105, 136–142
    real, 142
  real, 141
Kontakte
  ohmsche, 141

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Kontaktpotential, 138
Kontinuitätsgleichung, 123
Kopplungsdämpfung, 346
Kopplungswirkungsgrad, 346
Korrelationsfunktion, 89
Korrelationszeit, 89
Kotankt
  Metall–Metall, 138
Kristall
  Si, 125
Kristallstrukturen, 106
kritischer Tiefpass, 60–61
Kurzpulslaser, 393

Ladung, 238–241
Ladungsgekoppelte Baulelemente, 148–150
Ladungsmessung, 238–241
Ladungsneutralität, 122
Ladungsträger
  Dichte, 116–117
  Transport, 121–124
Ladungsträgerdichte, 117, 121, 125, 140, 154, 160, 298, 299
  injiziert, 123
  Leitungsband, 114
  Valenzband, 114
Laplace-Transformation, 30–32
  diskret, 35
  Vergleich z-Transformation, 35
Laser
  Halbleiter, 178
Laserdiode, 346, 546
Laserresonator, 384
Last
  kapazitiv, 216
Laue-Abbildung, 518
Lawineneffekt, 321
Lecherleitung, 323, 327
LED, 346
  weiss, 178
LEED, 522
Leistungsanpassung, 57
Leistungsspektrum, 90
Leistungsspektrums, 91
Leiterbahn
  Si, 163
Leitfähigkeit, 117–120
Leitung
  koaxial, 322
  Streifen, 322
  Zweidraht, 322
Leitwert, 22
Leitwerts, 154
Lennard-Jones Potential, 535, 536
Lineare Phase, 77
Lithographie
  optisch, 163
Lloyd-Interferometer, 352
LM741, 221, 224, 232
Lock-In, 247
  AD630, 284
Lock-In Verstärker, 281–284
Lock-In Verstärker, 11
Logische Verknüpfungen, 42–46
Low Energy Electron Diffraction, 522

MAC, 103
  Multiplikator-Akkumulator, 103
Majoritätsladungsträger, 115
Maschinenbefehl, 99
MASH, 272
Masse
  effektive, 111
Material
  Halbleiter, 113, 155, 175, 179
    intrinsisch, 175
Mathematische Grundlagen, 11–103
MC1436, 225
Measurement, 435, 447, 451, 455, 464–467, 534, 538, 541, 544, 548, 550–553, 556, 637, 638, 650
Mechanische Vibrationen, 375
mechanische Vibrationen, 375
MESFET, 154
  GaAs, 157–159
Messgerät, 181, 222, 223, 287, 360, 362–366, 421, 558, 563
  analog, 223
Messgerät, 11
Messtechnik
  Halbleiter, 139
Messung
  AC-Grössen, 225–238
  Frequenz, 285
    PLL, 292
    Quartz, 289–295
  Ladung, 238–241

Spannung, 222–225
Strom, 217–222
Widerstand, 241–249
Messverfahren, 217–566
Messvorschrift, 52
Messwiderstand, 217–220, 360, 366, 367
Metall-Halbleiter, 124
Kontakt, 138
Metall-Halbleiter-Kontakt, 124, 136–142
Metall-Isolator-Halbleiterkontakt, 142–148
Millersche Indizes, 651
Minoritätsladungsträger, 115
MIS, 142
MIS-Diode, 142–148
MISFET, 154
Mo, 307
Modenverteilung, 383
Modulation, 86–89
Modulationsdotierung, 135
MOS, 143
MOS-Diode, 142–148
MOSFET
n–Kanal, 160
p–Kanal, 160
Multi-Stage Noise Shaping, 272
Multimoden-Wellenleiter, 340
Multiquantum Well, 179

n–Schicht
inonenimplantiert, 150
n–Wannen Silizium–Gate CMOS–Prozesses, 163
n-dotiert, 144
n-Halbleiter, 114, 123, 138, 140, 144, 174
Na, 317
NAND, 45
NAND-Gatter, 44
Negation, 43
Netzfrequenz, 374
Netzwerkanalysator, 57, 425
skalar, 427
vektoriell, 427
Neutralisationsbedingung, 111
Ni, 307
Nicht-Gatter, 43
NOR, 45
Exklusiv, 46
NOR-Gatter, 44
Norton, 181–184
NOT, 43
Numerische Apertur, 343, 412
Gradientenindexfaser, 345
Nyquist, 93
Abtasttheorem, 422
Rauschen, 89
Theorem, 90
Nyquist-Frequenz, 68
Nyquist-Theorem, 90, 92

Oberflächenrekombinations–Geschwindigkeit, 124
Oberflächenstörstellen, 123
Oberfrequenz, 374
Oder, 44
Oder-Gatter, 43
OP27, 227
OPA111, 221
Operationsverstärker, 221, 225, 228, 229
Bias-Strom, 221
INA105, 221, 225
LM741, 221, 224, 232
MC1436, 225
OP27, 227
OPA111, 221
Standard, 209
optische Faser, 340
Optische Messverfahren, 376
optischer Datenübertragung, 375
Optoelektronik
Bauelemente, 175–179
OPV
Standard, 209
OR, 44
OTA, 216
output signal, 551
Oversampling, 68, 271, 422
Oversamplings, 423
Oxidation
nass, 143
thermisch, 143
trocken, 143

p-Halbleiter, 114, 140, 144
p-n-Übergang, 125–129


 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

vorgespannt, 129–132
PAL, 48
parabolische Näherung, 110
Parallelschaltung
Vierpol, 55
Parallelschwingkreis, 21, 23, 326, 329
Parallelwiderstand, 362
PbS, 319
PbSe, 319
PbTe, 319
Phase
Messung, 427
Phase Locked Loop, 292
Phasenmaske, 351
Phasenmessung, 427
Phasenreserve, 214
Phasenschieber, 66
Photo Multiplier, 318
Photodiode, 225, 372
Phototransistor, 321
piezoresistiv, 252
PIN–Diode, 188
PLL, 292
PMOS Aluminium–Gate–Prozess, 161
Pockelszelle, 393
Poggendorff'sche Kompensationsmethode, 224
Poggendorff-Kompensator, 220
Poisson–Gleichung, 125
Poisson-Zahl, 356
Pol
reell, 61
Polar-Plot, 627, 629, 630
Polardarstellung, 430, 431
Polychromator, 409
Polynom
Butterworth, 62
Tschebyscheff, 62
Potentiallinie, 332, 476
Potentialtheorie, 332
Poynting-Vektor, 396
Prismen-Interferometer, 352
PROM, 167
Pt, 307
pyroelektrischer Effekt, 315

Quartz, 289–295

R-2R-Netzwerk, 266
RAM
dynamisch, 166
statisch, 165
Raumladung, 125, 126, 131, 140, 145
MIS-Diode, 146
negativ, 135
positiv, 135, 138
Raumladungen, 125
Raumladungszone, 125, 127, 129–132, 134, 135, 138, 140, 144–146, 157, 173, 176, 319, 320
Anreicherung, 135
BC, 174
pn-Übergang, 319
Rauschbandbreite, 96
Rauschen, 89–97
1/f, 95, 374
Flickerrauschen, 95
Generationsrauschen, 95
Nyquist, 89
rekombinationsrauschen, 95
Schroteffekt, 95
thermisch, 360, 362, 363
weiss, 90, 95, 360, 363, 374
Widerstand, 89
Rauschspektrum, 93, 95
Rauschstrom, 95
Rechenwerk, 99
Rechnen
Blockschema, 14–16
Referenzoszillator, 11
Reflection high energy electron diffraction, 531
Reflexionsfaktor, 429
Reflexionsspektrum
Bragg-Gitter, 354
Rekombinationsprozesse, 121–124
Rekombinationsraten, 123
Relation
Wiener-Khintchine, 28
Relaxationsschwingung, 393
Resonator, 376–378, 380–382, 396, 400
Fabry-Perot, 358, 376
konfokal, 383, 387
RHEED, 136, 531
RS-232, 559
RS-422, 559
RS-485, 559

Rückkoppelsignal, 16

S-Parameter, 431
Sägezahnverfahren, 276
Sample&Hold, 263
Sb, 317
Scanning Force Microscope, 433, 534
Scanning Tunneling Microscope, 225, 534
SCCD, 150
Schaltalgebra, 46
  Distributivgesetz, 47
  Kommutativgesetz, 47
  Postulate, 46
  Theoreme, 47
Schaltdiode, 187
Schalter, 173
Schaltungstechnik, 105–216
Schieberegister
  analog
    dynamisch, 148
Schottky–Barriere, 140
Schottky–Diode, 187
Schottky–Modell, 127
Schottky-Diode, 136–142
Schroteffekt, 95
Schwarzkörperstrahlung, 93
SCSI, 559
Seebeck-Koeffizient, 314, 636, 657
Selbst-Phasenmodulation, 403
SEM, 455
Sensoren, 217–566
Serienschaltung, 21
SET, 181, 473
SFM, 433, 534, 535, 537, 538, 541, 542, 544, 546–556
Shockley, 313
Shunt, 362
Si, 105–109, 113, 114, 117, 119, 121–123, 127, 129, 134, 143, 146, 149, 155, 159, 176, 177, 184, 300, 321, 636
  (111), 526
  (111)-(7x7), 519
  Atom, 114
  Bauelement, 122
  Diode, 135, 184
  dotiert, 146
  Einkristall, 114
  Kristall, 125
  Leiterbahn, 163
  n-dotiert, 142
  Oberfläche, 143
  p-dotiert, 142, 160
  Schicht, 163
  Substrat, 106, 148, 149
    n-dotiert, 160
  Wafer, 161, 175
    p-dotiert, 160
Si(111)-(7x7), 518
Signal, 20, 30, 66–69, 86–88, 96, 167, 197, 211, 224, 226, 228, 235, 239–241, 254, 263, 269–272, 274, 284, 287, 295, 375, 417, 423, 425, 429–431, 512, 657
  übertragen, 425
  analog, 67, 68
  digital, 67, 68
  oversampled, 272
  reflektiert, 425
signal, 445, 449, 451, 452, 545, 548, 555, 556
  difference, 548
  error, 452
  induced, 556
  output, 551
  x-scanning, 452
signal photo diode, 546
signal to noise ratio, 543, 548
Signale, 24–25
  Ausschaltvorgang, 24–25
  digital, 25, 41–51
  diskret, 25
  Einschaltvorgang, 24–25
  kontinuierlich, 24–51
  periodisch, 24
  wertdiskret, 24–51
  zeitdiskret, 24–51
Signalfluss, 11, 16–20
  Allgemeine Formel, 20
  Begriffe, 19–20
  Grundlagen, 17–19
  Richtung, 13
Signalflussdiagramm, 16–20
  Übertragung, 18
  Allgemeine Formel, 20
  Begriffe, 19–20

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

Definitionen, 19
Grundlagen, 17–19
Multiplikation, 19
Summation, 18
Signalgenerator, 427
Silicon, 536, 556, 644
Silizium, 106–109, 113, 114, 117, 119, 121–123, 127, 129, 134, 143, 146, 149, 155, 159, 176, 177, 184, 300, 321, 636
(111), 526
(111)-(7x7), 519
Atom, 114
Bauelement, 122
Diode, 135, 184
dotiert, 146
Einkristall, 114
einkristallin, 105
Kristall, 125
Leiterbahn, 163
n-dotiert, 142
Obergläche, 143
p-dotiert, 142, 160
Schicht, 163
Substrat, 106, 148, 149
n-dotiert, 160
p-dotiert, 157
Wafer, 161, 175
p-dotiert, 160
Silizium(111)-(7x7), 518
Single Electron Transistor, 474
Single electron transistor, 181
Single Quantum Well, 179
Single-Electron-Transistor, 473
Single-mode Glasfaser, 350
Sintern, 310
SiSubstrat
p-dotiert, 157
Skintiefe, 338
Slew–Rate, 215
Smith-Chart, 427, 430, 618, 624, 625, 628, 630
Smith-Diagramm, 430, 431
Software, 449
Solarzelle, 177
Solarzellen, 177
Source, 154, 473
Sourceschaltung, 203
Spannungsgegenkopplung, 204
Stromgegenkopplung, 204
Spannung, 222–225
Spannungsbegrenzung, 187
Spannungsgegenkopplung, 196
Spannungsmessung, 222–225, 360, 362
Spannungsquelle, 24, 181, 183, 184, 192, 209, 222–225, 228, 242–245, 247, 268, 359, 361, 365, 366, 374, 563
Spannungsstabilisierung, 187
Spectroscopy, 447, 449, 451–453, 462
Speicher
Halbleiter, 161–168
Spektralanalysator, 423
Spektrometer, 409
Spikes, 392
Spule, 21
SQUID, 304
SRAM, 165
Standard–OPV, 209
Sternschaltung, 58
Steuerung
bipolarer Transistor, 172
Stichleitung, 327
STM, 225, 372, 433, 438–441, 445, 448–452, 454–458, 460–469, 534, 537, 541, 551, 552, 554, 644
Cryogenic, 457
UHV, 446
Strahlung
ionisierend, 364
kosmisch, 364
Strahlungsquellen
Fotodetektoren, 176–177
Halbleiter, 177–179
Streakkamera, 420
Streifenleiter, 339
Streifenleitung, 322
Strom, 217–222
Strom–Spannungs–Wandler, 196
Strom–Verstärker, 209
Stromgegenkopplung, 194
Strommessung, 217–222, 360
Stromquelle, 205
Stromspiegel, 206
Successive Approximation, 274
Summationspunkt, 16

surface channel CCD, 150
Surface Force Apparatus, 534
Symbole
  Blockschema, 11–14
System
  linear, 39

Talbot-Effekt, 352
Technologie
  Halbleiter, 161
TEM, 517
TEM-Moden, 383
TEM-Welle, 322, 570
Temperatur
  leitfähigkeit, 117–120
Temperaturdrift, 375
Temperaturschwankung, 375
Theorie
  Modulation, 86–89
Theory, 433, 438, 439, 462, 463, 534, 536
thermische Oxidation, 143
Thermospannung, 362, 363
Thévenin, 181–184
Thyristoren, 376
Tiefpass, 61
  Bessel, 63–64
  Butterworth, 61–62
  kritisch, 60–61
  Tschebyscheff, 62–63
Tiefpass-Allpass-Transformation, 66–67
Tiefpass-Bandsperren-Transformation, 66
Tiefpass-Bandpass-Transformation, 65
Tiefpass-Bandsperre-Transformation, 66
Tiefpass-Hochpass-Transformation, 65
Tiefpassfilter, 236
Totzeitglied, 70
Transferweite, 525
Transformation
  Fourier, 25–30
  Laplace, 30–31
  Tiefpass-Bandpass, 65
  Tiefpass-Bandsperren, 66
  Tiefpass-Hochpass, 65
Transformationen, 15–16
Transimpedanz–Verstärker, 209
Transistor, 53, 142, 154, 155, 159, 160, 169, 173, 190, 196, 197, 200, 206, 207, 321, 431
  bipolar, 177, 365
    Grundschaltung, 192
    Modell, 192
    Si/SiGe, 175
  Darlington
    npn, 201
  Feldeffekt, 154
  Feldeffekt–, 154
  Grundschaltung, 192
  MOS, 180
  npn, 169, 170, 206
    bipolar,vertikal, 174
  PMOS, 161
  pnp, 169
    lateral, 174
  Vierpolparameter, 53
Transistoren
  bipolar, 169–175
  Feldeffekt, 154–160
  unipolar, 154–160
Transkonduktanz–Verstärker, 209
Transkonduktanzverstärker, 216
Transmissionselektronenmikroskopie, 517
Transport
  Ladungsträger, 121–124
True-RMS-Gleichrichter, 235
Tschebyscheff-Polynom, 81
Tschebyscheff-Tiefpass, 62–63

UHV-STM, 446
Und, 43
Und-Gatter, 43
Unipolartransistoren, 154–160
Universalverstärker, 212
Unschärferelation, 90
USB, 559

VC–OP, 209
Verknüpfungen
  logisch, 42–46
Verstärker
  Lock-In, 281–284, 363
Verzögerungsglied, 70
Vibration, 445, 446, 454, 455, 457, 537–539
Vierpol, 52–58

 ©2001-2009 Ulm University, Othmar Marti and Alfred Plettl

π-Glied, 59
Übertragungsfunktion, 56
Dtreiecksschaltung, 59
Kettenschaltung, 55
Kreuzglied, 59
Parallelschaltung, 55
Serienschaltung, 53
Sternschaltung, 58
T-Glied, 58
Theorie, 52–58
Vierpoldarstellung, 325
Vierpolparameter, 53
VME-Bus, 559
Voltmeter
digital, 365
Volumenstromdichte, 122
von Neumann-Architektur, 97
VV–OP, 209

Wägeverfahren, 274
Wafer
Si, 175
Wandler
Analog/Digital, 272–281
Digital/Analog, 261–272
Wandlerschaltungen, 259–281
Wechselspannung, 225–238
Wechselstrom, 225–238
weisses Rauschen, 90, 374
Widerstand, 21, 89, 90, 92–94, 217, 222–224, 228, 229, 235, 241–250, 261, 263, 568
gesteuert, 154
steuerbar, 202
Widerstandsmessung, 241–249, 363
Widerstandsrauschen, 89, 95
Wiener, 28
Wiener-Khintchine-Relation, 28

XNOR-Gatter, 46
XOR-Gatter, 47

Y-Matrix, 53

z-Transformation, 32–37
Rücktransformation, 35–37
Rechenregeln, 33–34
Vergleich Laplace-Transformation, 35
Zahlensystem, 41
Zahlensysteme, 42
Zeigerdiagramm, 21
Zeitmittel, 657
Zenerdiode, 187
Zustand
Grenzfläche, 140
Zweidrahtleitung, 322