

ing
elektrotechnik

Karl-Hermann Cordes
Andreas Waag
Nicolas Heuck

Integrierte Schaltungen

Grundlagen – Prozesse – Design – Layout

Integrierte Schaltungen

**Karl-Hermann Cordes
Andreas Waag
Nicolas Heuck**

Integrierte Schaltungen

Grundlagen – Prozesse – Design – Layout



ein Imprint von Pearson Education
München • Boston • San Francisco • Harlow, England
Don Mills, Ontario • Sydney • Mexico City
Madrid • Amsterdam

Bibliografische Information Der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.d-nb.de> abrufbar.

Die Informationen in diesem Buch werden ohne Rücksicht auf einen eventuellen Patentschutz veröffentlicht.

Warennamen werden ohne Gewährleistung der freien Verwendbarkeit benutzt.

Bei der Zusammenstellung von Texten und Abbildungen wurde mit größter Sorgfalt vorgegangen. Trotzdem können Fehler nicht ausgeschlossen werden.

Verlag, Herausgeber und Autoren können für fehlerhafte Angaben und deren Folgen weder eine juristische Verantwortung noch irgendeine Haftung übernehmen. Für Verbesserungsvorschläge und Hinweise auf Fehler sind Verlag und Autoren dankbar.

Alle Rechte vorbehalten, auch die der fotomechanischen Wiedergabe und der Speicherung in elektronischen Medien. Die gewerbliche Nutzung der in diesem Produkt gezeigten Modelle und Arbeiten ist nicht zulässig.

Es konnten nicht alle Rechteinhaber von Abbildungen ermittelt werden. Sollte dem Verlag gegenüber der Nachweis der Rechtsinhaberschaft geführt werden, wird das branchenübliche Honorar nachträglich gezahlt.

Fast alle Produktbezeichnungen und weitere Stichworte und sonstige Angaben, die in diesem Buch verwendet werden, sind als eingetragene Marken geschützt. Da es nicht möglich ist, in allen Fällen zeitnah zu ermitteln, ob ein Markenschutz besteht, wird das © Symbol in diesem Buch nicht verwendet.

10 9 8 7 6 5 4 3 2 1

13 12 11

ISBN 978-3-86894-011-4

© 2011 Pearson Studium

ein Imprint der Pearson Education Deutschland GmbH,
Martin-Kollar-Straße 10-12, D-81829 München/Germany

Alle Rechte vorbehalten

www.pearson-studium.de

Lektorat: Birger Peil, bpeil@pearson.de

Christian Schneider

Alice Kachnij, akachnij@pearson.de

Korrektorat: Katharina Pieper, Berlin

Einbandgestaltung: Thomas Arlt, tarlt@adesso21.net

Herstellung: Philipp Burkart, pburkart@pearson.de

Satz: mediaService, Siegen (www.media-service.tv)

Druck und Verarbeitung: Bercker Graphische Betriebe, Kevelaer

Printed in Germany

Inhaltsverzeichnis

Vorwort	15
Teil I Grundlagen integrierter Schaltungen	19
Kapitel 1 Technologien der Mikroelektronik	21
1.1 Leiterplattentechnik	22
1.2 Hybridtechnik	23
1.3 Halbleitertechnik	24
1.3.1 Integrierte Schaltungen	24
1.3.2 Aufbau einer integrierten Schaltung	28
1.3.3 Entwurfsprozess integrierter Schaltungen	30
Kapitel 2 Physikalische Grundlagen der Halbleitertechnik	35
2.1 Grundlagen der Halbleiterphysik	37
2.2 Vom einzelnen Quantentrog zur Bandstruktur von Halbleitern	38
2.2.1 Der Formalismus der Quantenmechanik	38
2.2.2 Bindungen und Austauschwechselwirkung	44
2.2.3 Der Spin des Elektrons	51
2.2.4 Die Entstehung von Bändern	53
2.3 Bandstruktur und Ladungsträgertransport	55
2.3.1 Fermi-Energie und die Ladungsträgerkonzentration	58
2.3.2 Akzeptoren und Donatoren	63
2.3.3 Driftstrom und Diffusionsstrom	64
2.4 Eigenschaften von Silizium	65
Kapitel 3 Integrierte Bauelemente	69
3.1 Der PN-Übergang	70
3.1.1 Aufbau und Funktionsweise	71
3.1.2 Der PN-Übergang im stromlosen Zustand	73
3.1.3 Der PN-Übergang mit äußerer Spannung	77
3.1.4 Die Kapazität eines PN-Übergangs	84
3.1.5 Durchbruchspannung einer Diode	89
3.1.6 Einfluss der Rekombination	89
3.2 Der Bipolar-Transistor	90
3.2.1 Funktionsweise des Bipolar-Transistors	91
3.2.2 Beschaltung eines Bipolar-Transistors	96
3.2.3 Kleinsignalgrößen	97
3.2.4 Kleinsignal-Ersatzschaltbild	99
3.2.5 Aufbau eines integrierten Bipolar-Transistors	100

3.2.6	Temperaturabhängigkeiten von Transistorparametern	101
3.2.7	Frequenzverhalten von Bipolar-Transistoren	103
3.2.8	Transistorsymbole	104
3.2.9	Ersatzschaltbild und SPICE-Parameter des Bipolar-Transistors	105
3.3	Der Junction-FET (JFET)	107
3.3.1	Funktionsweise	107
3.3.2	Kennliniengleichung	110
3.3.3	Aufbau eines integrierten JFETs	113
3.3.4	Kleinsignalgrößen	113
3.3.5	Symbole des JFETs	114
3.3.6	Frequenzverhalten	115
3.3.7	Ersatzschaltbild und SPICE-Parameter des JFETs	115
3.4	Der MOSFET	117
3.4.1	Aufbau und Funktionsweise	118
3.4.2	Kennlinien eines MOSFETs	122
3.4.3	Weitere MOSFET-Typen und ihre Schaltungssymbole	128
3.4.4	CMOS-Transistorpaar	129
3.4.5	Kleinsignalgrößen	130
3.4.6	Temperaturverhalten	135
3.4.7	Ersatzschaltbild und SPICE-Parameter des MOSFETs	135
3.5	Passive Bauelemente	140
3.5.1	Widerstände	140
3.5.2	Kondensatoren	143
3.6	Kurzkanaleffekte und Skalierung	144
3.7	Geschwindigkeit eines MOSETs: Optimierung der Taktfrequenzen	147
3.8	MOS-Speicher	149
3.8.1	Dynamic Random Access Memories (DRAM)	149
3.8.2	Static Random Access Memories (SRAM)	151
3.8.3	Floating-Gate-Speicher	152
Kapitel 4 Technologie integrierter Schaltungen		157
4.1	Wafer-Herstellung	159
4.1.1	Klassische Silizium-Technologie	159
4.1.2	SOI – Silicon On Insulator	163
4.2	Lithografie und Reinraumtechnik	166
4.2.1	Fotoresist	167
4.2.2	Maskentechnik und Belichtung	168
4.2.3	Reinräume	173
4.3	Dotierung	174
4.3.1	Ionenimplantation	174
4.3.2	Diffusion	175

4.4	Schichttechnik	179
4.4.1	Sputtern	179
4.4.2	CVD-Verfahren.	181
4.4.3	Aufdampfverfahren	183
4.4.4	Thermische Oxidation.	185
4.4.5	Silizierung	186
4.5	Ätztechnik	186
4.5.1	Nassätzen	187
4.5.2	Trockenätzen	188
4.6	Metallisierung, Planarisierung und Durchkontaktierung in integrierten Schaltungen.	188
4.6.1	Leiterbahnen	190
4.6.2	Planarisierung	191
4.6.3	Kontakte und Vias	193
 Kapitel 5 Aufbau- und Verbindungstechnik integrierter Schaltungen		 197
5.1	Vom Front-End zum Back-End.	199
5.2	Kontaktierung und Befestigung integrierter Schaltungen.	201
5.2.1	Drahtbonden: Chip-and-Wire	201
5.2.2	Flip-Chip-Kontaktierung.	207
5.2.3	Tape-Automated-Bonding.	209
5.2.4	Chipbefestigung.	211
5.3	Single-Chip-Packaging	216
5.3.1	Packages für die THT-Montage: DIP, SIP und PGA	216
5.3.2	Packages für die SMT-Montage: SOP, QFP und BGA	217
5.3.3	Eigenschaften von Vergussmasse und Lead-Frame in Kunststoffgehäusen	220
5.4	Direktmontage: Chip-On-Board	221
5.5	Multi-Chip-Packaging	221
 Kapitel 6 Defektmechanismen und Teststrategien		 227
6.1	Ausfallmechanismen und Defekte in integrierten Schaltungen	230
6.2	Teststrategien für integrierte Schaltungen	233
6.2.1	Funktionaler Test	233
6.2.2	Struktureller Test	234
6.2.3	Testabläufe	234
6.3	Test digitaler Schaltungen	235
6.3.1	Fehlermodelle und Testmustererzeugung	236
6.3.2	Testfreundlicher Entwurf digitaler ICs.	237
6.4	Test analoger Schaltungen	238
6.5	Test eines analogen ASICs	238
6.5.1	Testaufbauten und Electrical Test Specification (ETS)	238
6.5.2	Testprozedur	245

Teil II	Prozesse und Layout integrierter Schaltungen	251
Kapitel 7	Standardprozesse der IC-Fertigung	253
7.1	Layer und Masken	255
7.2	Prozessanbieter und Prozesse – ein Überblick	256
7.3	Bipolar-Prozesse	257
7.3.1	Standard-Bipolar-Prozess (C36).	258
7.3.2	Erweiterter Bipolar-Prozess (C14))	263
7.3.3	Widerstände in Bipolar-Prozessen	266
7.3.4	Kondensatoren in Bipolar-Prozessen	269
7.4	CMOS-Prozesse	272
7.4.1	CMOS-Standardprozess (CM5)	273
7.4.2	Fortschrittliche CMOS-Prozesse	281
7.4.3	Widerstände in CMOS-Prozessen	282
7.4.4	Kondensatoren in CMOS-Prozessen	284
7.5	Bipolar-CMOS-Technologien	285
Kapitel 8	Grundregeln für den Entwurf integrierter Schaltungen	287
8.1	Kurze Einführung in die Schaltungsberechnung	292
8.1.1	Aufgabenstellung (Spezifikation)	293
8.1.2	Kurze Erklärung der Schaltung	294
8.1.3	Lösung (Grobdimensionierung)	294
8.2	Hinweise zur Dimensionierung integrierter Schaltungen	309
8.2.1	Geometrische Abmessungen Passiver Elemente	309
8.2.2	Geometrische Abmessungen Aktiver Elemente	314
8.2.3	Absichern der Spezifikation durch eine Monte-Carlo-Simulation	320
8.3	Design-Rules	327
8.3.1	Abgeleitete Layer	328
8.3.2	Abstände und Weiten eines Prozess-Layers	329
8.3.3	Abstände zwischen verschiedenen Layern	332
8.3.4	Vergrößern und Verringern von Design-Maßen	335
8.3.5	Implementierung der Rules in das Layout-Programm	336
8.4	Layout-Synthese	342
8.4.1	Beispiel 1: Bipolar-Transistor	343
8.4.2	Beispiel 2: CMOS-Inverter	347
8.5	Richtlinien zur Layout-Erstellung	355
8.5.1	Paarung von Bauelementen (Matching)	355
8.5.2	Zusammenstellung wichtiger Layout-Regeln und Layout-Beispiele	360
8.5.3	Hinweise für Zellen einer Bibliothek	382
8.5.4	Spezielle Hinweise für CMOS-Zellen einer Bibliothek	384
8.5.5	Verbindungsleitungen	386

8.6	Parasitäre Effekte	389
8.6.1	Unerwünschte Kanalbildung, Vermeidung durch Channel-Stopper	390
8.6.2	Latch-Up in integrierten Schaltungen	394
8.7	Layout-Verifikation	403
8.7.1	Design-Rule-Check	404
8.7.2	Layout-Versus-Schematic (LVS)	404

Teil III Analoge integrierte Schaltungen: Design, Simulation und Layout 421

Kapitel 9 Stromspiegelschaltungen 425

9.1	Der einfache Stromspiegel	427
9.2	Die Widlar-Schaltung	436
9.3	Widerstandsbestimmter Stromteiler	438
9.4	Korrektur von Stromspiegelfehlern	440
9.4.1	Korrektur des Basis-Strom-Einflusses	440
9.4.2	Kompensation des Early-Effekt-Einflusses	441
9.4.3	Korrektur von Offset-Fehlern	446
9.4.4	Reduzieren der Sättigungsspannung	449
9.4.5	Stromspiegel mit geregelter Kaskoden-Schaltung	454
9.5	Praktisches Beispiel	455
9.5.1	Aufgabenstellung	455
9.5.2	Lösung	456
9.5.3	Überprüfung des Ergebnisses und erste Korrektur	458
9.5.4	Zweite Korrektur	460
9.5.5	Layout-Erstellung	461
9.6	Dynamisches Verhalten von Stromspiegelschaltungen	463
9.6.1	Bipolare Stromspiegel	464
9.6.2	MOS-Stromspiegel	467

Kapitel 10 Stromquellen 471

10.1	Stromeinstellung über einen Widerstand	472
10.2	Stromquelle mit Vorwärtsregelung	473
10.3	Stromquelle mit einem JFET	475
10.4	Verwendung von U_{BE} als Referenzspannung	477
10.5	PTAT-Stromquellen	478
10.5.1	Einfache PTAT-Stromquelle	478
10.5.2	PTAT-Stromquelle mit Early-Effekt-Kompensation nach Wilson	481
10.5.3	PTAT-Stromquelle mit Vorwärtsregelung	482
10.5.4	PTAT-Stromquelle mit Vorwärtsregelung und Kaskode	483
10.5.5	Erdi-Stromquelle	484

10.6	CMOS-Stromquellen	485
10.6.1	Beta-Multiplier	486
10.6.2	Praktisches Beispiel: Beta-Multiplier für kleine Ströme	487
10.6.3	CMOS-Stromquelle ohne Verwendung von Widerständen	491
10.7	Eine fast genaue Stromquelle	492
Kapitel 11 Spannungsreferenzen		495
11.1	Z-Diode als Spannungsreferenz	496
11.2	PTAT-Spannungsreferenzen	497
11.2.1	Einfache PTAT-Spannungsreferenz	498
11.2.2	Verwendung einer unsymmetrischen Differenzstufe	499
11.3	Bandgap-Spannungsreferenzen	504
11.3.1	Das Prinzip der Bandgap-Referenz	504
11.3.2	Die einfache Widlar-Bandgap-Referenz	508
11.3.3	Brokaw-Bandgap-Referenz	509
11.3.4	Widlar-Bandgap-Referenz	513
11.3.5	Beispiel: Design einer Widlar-Bandgap-Referenz	514
11.3.6	Second-Order-Temperaturkompensation	518
11.3.7	Zweipolige Vier-Transistor-Bandgap-Referenz	520
11.3.8	Referenzspannungen kleiner als 1 V	522
11.3.9	Bandgap-Referenzen in CMOS-Prozessen	527
Kapitel 12 Das Differenztransistorpaar		537
12.1	Das Emitter-gekoppelte Bipolar-Transistorpaar	538
12.1.1	Linearisierung durch Stromgegenkopplung	540
12.2	Das Source-gekoppelte MOSFET-Transistorpaar	542
Kapitel 13 Operationsverstärker		545
13.1	Allgemeines	547
13.2	Differenzeingangsstufe	549
13.2.1	Der Differenzverstärker	549
13.2.2	Eingangsstufe mit Widerständen	554
13.2.3	Differenzstufe mit Stromspiegelausgang	555
13.2.4	Eingangsstufe für Gleichtaktanteile außerhalb der Versorgungsspannung	559
13.2.5	PNP-Eingangsstufe	560
13.2.6	Kaskoden-Eingangsstufe mit Super-B-Transistoren	563
13.2.7	Kompensation des Eingangsstromes	564
13.2.8	Präzisionseingangsschaltung	565
13.2.9	Rail-to-Rail-Eingangsschaltung	566
13.3	Eingangs-Offset	571
13.3.1	Bipolarer Eingang	571
13.3.2	MOS-Eingang	573
13.3.3	Offset-Trimmen	574

13.4	Ausgangsstufe	576
13.4.1	Emitter- bzw. Source-Folger-Ausgang	576
13.4.2	Komplementär-Ausgangsstufe im AB-Betrieb	577
13.4.3	CMOS-Push-Pull-Ausgangsstufe mit Inverter-Ansteuerung	580
13.4.4	Push-Pull-Ansteuerung über Fehlerverstärker	582
13.4.5	Einfacher Buffer im Gegentakt-A-Betrieb	583
13.5	Dynamisches Verhalten und Stabilität von Operationsverstärkern	585
13.5.1	Frequenzgang, Übertragungsfunktion	586
13.5.2	Stabilität eines gegengekoppelten Systems	588
13.5.3	Frequenzgangkorrektur	591
13.5.4	Miller-Korrektur des Frequenzganges	596
13.5.5	Slew-Rate	601
13.6	Design-Beispiele von Operationsverstärkern	605
13.6.1	Einfacher Bipolar-Operationsverstärker mit PNP-Eingang	606
13.6.2	Einfacher Operationsverstärker in CMOS	616
13.6.3	Einfacher Transconductance-Verstärker (OTA) in CMOS	618
13.7	Operationsverstärker mit symmetrischem Ausgang	619
13.8	Komplettes Design eines CMOS-Operationsverstärkers: Berechnung, Simulation, Korrektur und Layout	625
13.8.1	Zusammenstellung der Formeln	625
13.8.2	Aufgabenstellung	631
13.8.3	Berechnung der Schaltung	632
13.8.4	Simulationsergebnisse	635
13.8.5	Layout-Erstellung	639
13.8.6	Layout-Verifikation	641
Kapitel 14 Einführung in GM-C-Schaltungen		647
14.1	Grundsaltungen	649
14.2	GM-C-Oszillator und GM-C-Filterschaltungen	651
14.2.1	GM-C-Oszillator	651
14.2.2	GM-C-Bandpass- und Tiefpass-Filter	655
14.3	Ausführung von GM-Zellen	659
Teil IV Digitale integrierte Schaltungen: Design, Simulation und Layout		667
Kapitel 15 Grundlagen digitaler integrierter Schaltungen		671
15.1	Grundbegriffe der digitalen Schaltungstechnik	672
15.2	Digitaltechnik	674
15.3	Digitale Schaltungsfamilien	678

Kapitel 16	Design und Layout digitaler Gatter in Emitter-gekoppelter Logik (ECL)	683
16.1	Typisches NOR-OR-Gatter	684
16.1.1	Dimensionierung	686
16.1.2	Referenzspannung	689
16.1.3	Anschluss externer Geräte über Leitungen	690
16.1.4	Layout	693
16.1.5	Layout-Verifikation	696
16.2	ECL-Gatter mit reduzierter Verlustleistung	698
16.3	EECL-Gatter mit geringer Verlustleistung	700
16.3.1	Layout	703
16.3.2	Layout-Verifikation	704
16.4	Andere ECL-Gatter	705
Kapitel 17	Design und Layout digitaler Gatter in Transistor-Transistor-Logik (TTL)	707
17.1	Die NAND-Funktion und der Multi-Emitter-Eingang	708
17.2	Layout	714
17.3	Weitere Vereinfachungen	715
17.4	Verbesserung einiger Eigenschaften	716
Kapitel 18	Design und Layout digitaler Gatter in CMOS	721
18.1	Der CMOS-Inverter als grundlegendes Schaltungselement	722
18.1.1	Schaltpunkt	723
18.1.2	Statischer Querstrom	724
18.1.3	Transiente Verlustleistung	725
18.1.4	Dynamische Verlustleistung	727
18.1.5	Gesamte Verlustleistung	728
18.1.6	Zeitverhalten	729
18.1.7	Festlegung der geometrischen Abmessungen	739
18.1.8	Inverter-Layout	740
18.1.9	Inverter mit höherer Treiberfähigkeit	741
18.1.10	CMOS-Inverter mit kontrolliertem Querstrom	748
18.2	CMOS-Schmitt-Trigger	749
18.3	TTL-CMOS-Interface	753
18.4	Transfer- oder Transmission-Gate	755
18.5	Inverter mit Tri-State-Ausgang	757
18.6	Das Exklusiv-ODER-Gatter (XOR)	758
18.7	NAND- und NOR-Gatter	760
18.7.1	Layout	764
18.7.2	Zeitverhalten	765

18.8	Verallgemeinerte CMOS-Gate-Strukturen	766
18.8.1	Dimensionierung	768
18.8.2	Layout der AOI-Schaltungen	771
18.8.3	Abstrahieren des Schaltplanes durch Zweige anstelle der Transistoren	774
18.9	Pseudo-NMOS-Logik	776
18.10	Dynamische logische Schaltungen (C ² MOS)	777
18.11	Domino-CMOS-Logik	781
18.12	Latches und Flip-Flops	784
18.12.1	Das einfache Latch und das RS-Flip-Flop	784
18.12.2	Getaktetes RS-Flip-Flop	788
18.12.3	Latch als Speicherzelle	789
18.12.4	Einfaches Daten-Flip-Flop (D-Flip-Flop)	796
18.12.5	Flanken-getriggertes D-Flip-Flop	798
18.12.6	Flanken-getriggertes D-Flip-Flop mit Set- und Reset-Eingang ..	803
Kapitel 19 Neue Entwicklungen		811
19.1	„More than Moore“	812
19.2	Verspanntes Silizium	815
19.3	„Low-k“- und „High-k“-Oxide als Dielektrika	816
19.4	Silizium-Photonik	819
19.5	Nano-FETs	820
19.6	Tri-Gate-Transistoren	821
19.7	Speichertechnologien	822
Anhang A		827
A.1	Frequenzgang eines einstufigen Verstärkers	828
A.2	Simulation mit SPICE	832
Register		835

Vorwort

Die Mikroelektronik ist wohl die wichtigste Basistechnologie unserer modernen Welt. Computer, Mobiltelefone, Kameras, das Internet, aber auch Automobile, Eisenbahnen und Flugzeuge sind ohne den Einsatz mikroelektronischer Systeme nicht mehr denkbar. Die Mikroelektronik begründet das Informationszeitalter, die moderne medizinische Diagnostik und unser gesamtes elektronisches Kommunikationswesen.

Herzstück all dieser Anwendungen sind *integrierte Schaltungen*, bei denen aktive und passive elektronische Halbleiterbauelemente auf einem einkristallinen Silizium-Chip integriert werden. Besonderes Merkmal ist dabei die enorme Miniaturisierung, die mittlerweile zu lateralen Abmessungen von weniger als 50 nm geführt hat. Damit konnte die Bauelement-Dichte und im selben Zuge die Taktfrequenz und damit die Leistungsfähigkeit mikroelektronischer Schaltungen massiv erhöht werden.

Die zur Herstellung integrierter Schaltungen verwendeten Entwicklungs- und Fertigungsverfahren erfordern heutzutage ein enorm hohes Maß an technischem Know-how, verbunden mit erheblichen Investitionskosten. Die Weiterentwicklung integrierter Schaltungen vollzieht sich deshalb auf unterschiedlichen, hoch spezialisierten Ebenen getrennt voneinander und mit einem jeweils sehr hohen Abstraktionsgrad: Bauelement-Physik, Halbleitertechnologie, Entwurf, Simulation, Layout. Gleichzeitig konnte gerade der Entwurf, die Simulation und Verifikation digitaler Schaltungen teilweise automatisiert werden. Die entsprechenden Software-Werkzeuge bewegen sich auf Abstraktionsebenen jenseits der elektrotechnischen Vorgänge auf Bauelementebene.

In den Curricula der Elektrotechnik, Physik und Informatik driften diese verschiedenen Spezialisierungsebenen immer weiter auseinander. Ziel dieses Buches ist es, dem entgegenzuwirken, umfassende Grundlagen zu legen und alle wesentlichen Aspekte der Bauelement-Physik, des Entwurfs, der Simulation und der Layout-Erstellung integrierter Schaltungen einheitlich zu vermitteln. Der Leser soll in die Lage versetzt werden, integrierte Schaltungen ausgehend vom einfachen Transistor und digitalen Gatter bis hin zu aufwendigen analogen Operationsverstärkerschaltungen selbst zu entwerfen und zu simulieren.

Als CAD-Werkzeuge für die Schaltungssimulation und Layout-Erstellung wurden ganz bewusst Softwareprogramme ausgewählt, die auf einem normalen PC unter Windows eingesetzt werden können und *kostenfrei* zur Verfügung stehen. So wird für die Schaltungsentwicklung die Vollversion des leistungsfähigen Simulators „LT-SPICE“ der Firma Linear Technology eingesetzt; für die Layout-Arbeiten stellt die Firma Tanner Research, Inc., Pasadena, CA, eine Studentenversion des hervorragenden und einfach zu bedienenden Werkzeuges „L-Edit“ zur Verfügung.

Der erfolgreiche Einstieg in das komplexe Gebiet des Chip-Designs steht und fällt mit dem Erlangen eigener praktischer Erfahrungen. Mittlerweile bieten namhafte Halbleiterhersteller ihre Technologie auch außerhalb der eigenen Produktion an. Über Multiprojekt-Wafer-Runs haben neben mittelständischen und kleinen Industriebetrieben auch Hochschulen den Zugang zu moderner Prozesstechnologie. Das vorliegende Buch soll deshalb vor allem auch zu Projektarbeiten anregen mit dem Ziel, dass einfache Schaltungen wirklich in Silizium realisiert, anschließend vermessen und die Resultate mit den Simulationsergebnissen verglichen werden können.

Inhalt

Das Buch ist in vier Teile gegliedert.

Nach einer Einführung in den Themenbereich der integrierten Elektronik stehen im ersten Teil des Buches die Beschreibung der Halbleiterphysik, integrierte Bauelemente sowie die grundlegenden Technologien bei der Fertigung im Mittelpunkt.

Im Teil II werden anhand einfacher, industriell verfügbarer Prozesse die Grundlagen für einen prozessspezifischen Schaltungsentwurf und die zugehörige Layout-Erstellung praxisnah erläutert. Dabei wird sowohl auf die erforderlichen schaltungstechnischen Grundlagen als auch auf die gängigen Design-Methoden und Verifikationsverfahren (Simulation, Design-Rule-Check, Layout-Versus-Schematic) eingegangen.

Die vorgestellten Prozesse und Grundregeln finden dann in den Teilen III und IV konkrete Anwendung: Anhand ausgewählter analoger und digitaler Grundsaltungen werden Rechen- und Layout-Beispiele ausführlich erläutert.

Der Leser wird mit diesem Buch in die Lage versetzt, integrierte Schaltungen vom einfachen digitalen Gatter bis hin zu aufwendigen analogen Operationsverstärkerschaltungen selbst zu entwerfen und zu simulieren. Die zugehörige Simulations- und Design-Software (LT-SPICE, L-Edit) steht im Downloadbereich des Verlages zur Verfügung.

Zielgruppe

Das Buch richtet sich an Studierende der Elektrotechnik, Informatik oder Physik und kann während des Studiums angefangen bei den Grundlagen der Elektronik und Schaltungstechnik über die Halbleiter- und Prozesstechnologie bis hin zum Design integrierter Schaltungen vorlesungsbegleitend verwendet werden. Die ausführlichen Rechen- und Simulationsbeispiele sollen zudem ein Selbststudium für Quereinsteiger ermöglichen.

Danksagung

Unser Dank gilt allen, die uns beim Verfassen des Buches unterstützt und zur Seite gestanden haben.

Dabei sei zunächst der Firma *ASIC* aus Emmering gedankt, die uns während des ganzen Buchprojektes sowohl bei den Prozessbeschreibungen als auch bei der Entwicklung der Layouts beratend zur Seite gestanden hat. Der Firma *Tanner Research* möchten wir für die Bereitstellung der kostenlosen Version der Layout-Software L-Edit unseren Dank aussprechen. Herrn Mike Engelhardt von der Firma *Linear Technology* danken wir für die Umsetzung zahlreicher LT-SPICE-Erweiterungen, durch die der Einsatz des Simulators erheblich vereinfacht wurde.

Den Herren Reiner Kaminski und PD Dr. Hergo-Heinrich Wehmann sei für die Durchsicht des Manuskripts, Frau Katharina Heuck für das Erstellen der zahlreichen Abbildungen gedankt.

Teile des Buches begleiten die Lehrveranstaltungen *Integrierte Schaltungen* und *Grundlagen der Elektronik*, die an der *TU Braunschweig* im Studiengang Elektrotechnik abgehalten werden. Die Grundzüge dieser Vorlesungen wurden in hohem Maße durch Herrn Prof. Andreas Schlachetzki, dem ehemaligen Leiter des *Instituts für Halbleitertechnik*, geprägt. Für diese ausgesprochen hilfreichen Vorarbeiten sind wir sehr dankbar.

Darüber hinaus danken wir dem Pearson-Verlag und insbesondere Herrn Birger Peil für die geduldige Begleitung beim Verfassen des Manuskripts.

Ein besonderer Dank gilt unseren Familien, durch deren Verständnis und Geduld dieses Buch erst möglich gemacht wurde.

Webseite

Die Simulations- und Design-Software (LT-SPICE und L-Edit) kann aus dem Internet heruntergeladen werden. Sie wird dem Leser darüber hinaus aber auch im Downloadbereich des Verlages zur Verfügung gestellt. Die Webseite des Verlages steht unter <http://www.pearson-studium.de>. Am schnellsten gelangen Sie von dort zur Buchseite, wenn Sie in das Feld „Schnellsuche“ die Buchnummer **4011** eingeben. In einer einzigen ZIP-Datei „IC.zip“ sind alle Unterlagen, die Sie zum Studium des Buches benötigen, zusammengefasst. So finden Sie neben den Programmen (Ordner „Programme“) auch die meisten Schaltungen und Layouts dieses Buches. Sie sind kapitelweise geordnet (Ordner „Kapitel 8“ bis „Kapitel 14“, „Kapitel 16“ bis „Kapitel 18“ sowie „Anhang“). Auch die Technologiedaten (Ordner „Technologie-Files“) und die SPICE-Parameter für die verwendeten Bauelemente (Ordner „Spicelib“) sind in der Datei IC.zip enthalten. Die Schaltungen sind mit dem Schematic-Editor von LT-SPICE erstellt worden (Dateiendung „asc“) und können deshalb direkt, eventuell nach einer Anpassung der Dateipfade, simuliert werden. Die Layout-Dateien (Dateiendung „tdb“) lassen sich mit L-Edit öffnen. Im Ordner „Skripten“ sind kurze Einführungen in die Programme SPICE, LT-SPICE und L-Edit im PDF-Format enthalten.



TEIL I

Grundlagen integrierter Schaltungen

1	Technologien der Mikroelektronik	21
2	Physikalische Grundlagen der Halbleitertechnik	35
3	Integrierte Bauelemente	69
4	Technologie integrierter Schaltungen.....	157
5	Aufbau- und Verbindungstechnik integrierter Schaltungen	197
6	Defektmechanismen und Teststrategien	227

Der erste Teil dieses Buches widmet sich den physikalischen, technologischen und fertigungsspezifischen Grundlagen integrierter Bauelemente und Schaltungen.

In **Kapitel 1** wird zunächst eine Unterscheidung der Technologien innerhalb der Mikroelektronik vorgenommen: Leiterplattentechnik – Hybridtechnik – Halbleitertechnik. Im Sinne einer Einführung in die Thematik wird in *Kapitel 1* der Begriff „Integrierte Schaltung“ erklärt, der generelle Aufbau vorgestellt und der allgemeine Verlauf eines Entwurfsprozesses erläutert.

Kapitel 2 widmet sich den **physikalischen Grundlagen** der Halbleitertechnik. Um den Begriff „Bandstruktur von Halbleitern“ zu erklären, erfolgt zunächst eine knappe Einführung in eine quantenmechanische Beschreibung der Energieniveaus in Halbleitern. Von dieser Bandstruktur ausgehend werden die elektronischen Eigenschaften von Halbleitern kurz besprochen und wichtige Begriffe wie Fermi-Energie, elektrische Leitfähigkeit und Dotierung diskutiert. Für ein Verständnis der quantenmechanischen Betrachtungen sind einige Grundkenntnisse der Vektoranalysis notwendig. Da diese z. B. an Fachhochschulen teilweise nicht gelehrt werden, sind die hier vorgestellten Zusammenhänge keine Voraussetzung für das Verständnis der weiteren Kapitel.

In **Kapitel 3** werden die wichtigsten **Bauelemente** in ihrer integrierten Form vorgestellt. Ausgehend von der Diode erfolgen die Beschreibung der wichtigsten Eigenschaften von Bipolar- und Feldeffekt-Transistoren sowie die Herleitung und Erläuterung der für eine Schaltungsberechnung wichtigen Kennliniengleichungen. Des Weiteren werden verschiedene Ausführungen integrierter Speicher und die Grundzüge passiver Bauelemente eingeführt. Die in *Kapitel 3* vorgestellten Bauelementbeschreibungen bilden die Basis für die Erläuterungen zur Schaltdimensionierung in den *Teilen III* und *IV*.

Kapitel 4 behandelt die **technologischen Grundlagen** der Herstellung integrierter Bauelemente in Silizium. Dabei werden die gängigen Prozeduren und Anlagen der IC-Fabrikation beschrieben; diese bilden wiederum die Basis für die in *Teil II* erläuterten Herstellungsabläufe bei den in diesem Buch verwendeten CMOS- und Bipolar-Prozessen.

Um eine in Silizium integrierte Schaltungsstruktur als elektronisches Bauelement nutzen zu können, müssen sowohl eine geeignete **Kontaktierung** als auch ein der Anwendung entsprechendes **Gehäuse** gewährleistet werden. Die dabei gängigen Verfahren zur Weiterverarbeitung und Kontaktierung des „reinen“ Silizium-Chips sowie die unterschiedlichen Gehäuseausführungen werden in **Kapitel 5** vorgestellt.

In **Kapitel 6** werden mögliche **Defekte** in integrierten Schaltungen sowie allgemeine **Teststrategien** im Überblick eingeführt. Das Kapitel schließt mit der Beschreibung einer praxisnahen Testprozedur an einer analogen Schaltung.

Technologien der Mikroelektronik

1.1 Leiterplattentechnik	22
1.2 Hybridtechnik	23
1.3 Halbleitertechnik	24

Einleitung

» Dieses Kapitel dient der Einführung des Lesers in den Themenbereich der integrierten Schaltungen. Dafür erfolgt zunächst eine kurze Beschreibung der gängigen Realisierungsmöglichkeiten für elektronische Schaltungen im Allgemeinen. Anschließend werden die Merkmale unterschiedlicher Typen von integrierten Schaltungen hinsichtlich der Anwendung und des Entwurfsprozesses kurz erläutert, um so eine erste Vorstellung für die Verfahrensabläufe beim Design und Layout kundenspezifischer Schaltungen (ASICs) zu vermitteln. «

LERNZIELE

- Unterscheidung von Leiterplattentechnik, Hybridtechnik und Halbleitertechnik
- Was ist eine integrierte Schaltung?
- Aufbau von integrierten Schaltungen
- Entwurfsprozess integrierter Schaltungen

In der Mikroelektronik gibt es eine Vielzahl von Möglichkeiten, elektronische Schaltungen zu realisieren. Um den Anwendungsbereich und die charakteristischen Eigenschaften integrierter Schaltungen einordnen zu können, werden an dieser Stelle die gängigen Technologien der Mikroelektronik zusammenfassend beschrieben. Eine übliche Einteilung unterscheidet drei Hauptgruppen: die **Leiterplattentechnik**, die **Hybridtechnik** und die **Halbleitertechnik**.

1.1 Leiterplattentechnik

Die **Leiterplattentechnik** stellt das Standardaufbaukonzept elektronischer Baugruppen dar. Es gibt viele verschiedene Ausführungsmöglichkeiten elektronischer Leiterplattenschaltungen, die abhängig von Einsatzgebiet und Anforderungen anwendungsspezifisch genutzt werden. Generell unterscheidet man Leiterplattenbaugruppen mit steckbaren Bauteilen (**Through Hole Technology, THT**) von Aufbauten mit aufsetzbaren Bauelementen (**Surface Mount Technology, SMT**; das Bauteil wird als **Surface Mounted Device**, kurz **SMD** bezeichnet). Während sich die Durchsteckmontage durch eine sehr hohe Robustheit und flexible Anwendung auszeichnet, sind als Vorteile des SMT-Prozesses vor allem eine bessere Automatisierbarkeit, ein höherer Integrationsgrad durch den Wegfall der Bohrungen und in den meisten Fällen auch die niedrigeren Kosten zu nennen. Zusätzlich bietet die Oberflächenmontage die Möglichkeit, eine Leiterplatte von beiden Seiten zu kontaktieren. Dies ist bei der Durchsteckmontage prinzipiell zwar auch möglich, birgt verfahrenstechnisch aber zu viele Schwierigkeiten.

In vielen Anwendungen werden die beiden Montagetechniken SMT und THT auch kombiniert (**Mixed Print**), was immer dann nötig wird, wenn die elektrische Belastung aufsetzbarer Bauteile überschritten wird oder aber bestimmte Komponenten nur in einer Bauart zur Verfügung stehen [1].

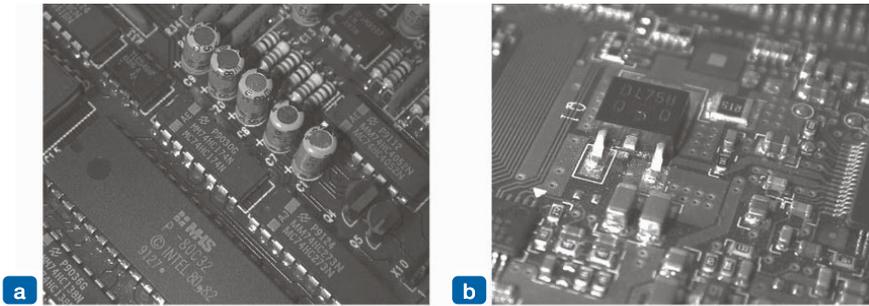


Abbildung 1.1: (a) Through Hole Technology (THT) und (b) Surface Mount Technology (SMT).

Das Einsatzfeld der Leiterplattentechnik reicht von einfachen Anwendungen als Schaltungsträger für Schaltungen aus diskreten Einzelbauelementen bis hin zur Plattform für komplexe Flachbaugruppen; bei diesen werden komplexe integrierte Bauelemente in Form von **SCPs (Single-Chip-Packages)** zusammen mit äußeren Beschaltungen durch aktive und passive Bauelemente zu einer Einheit zusammengefasst. In diesem Fall kann man nicht mehr von einem Leiterplattenschaltungskonzept sprechen; vielmehr stellt die Leiterplatte eine Ebene des Packaging (siehe *Kapitel 5*) dar.

Sowohl bei der SMT- als auch bei der THT-Realisierung von Leiterplattenbaugruppen werden die Bauelemente über metallische Leiterbahnen – zumeist Kupfer – miteinander verschaltet. Grundmaterialien der Leiterplatten sind für Low-Cost-Anwendungen vor allem Phenolharze mit Papier, für anspruchsvollere Strukturen kommen glasfaserverstärkte Epoxidharze zum Einsatz. Für komplexere Leiterplattenaufbauten, wie beispielsweise in Handys oder Computern, werden als Metallisierungsebenen nicht nur die Ober- und Unterseite genutzt, sondern mehrlagige Leiterplatten verwendet.

Die Leiterplattentechnik ist im Allgemeinen vor allem für Strukturen mit einer begrenzten Anzahl von Bauelementen sowie als Grundlage von Anwendungen mit integrierten Schaltungen geeignet. Der Entwicklungsaufwand von Schaltungen mit Leiterplattentechnik ist dabei im Vergleich zu anderen Technologien der Mikroelektronik geringer, Schaltungen können schnell und flexibel realisiert werden.

1.2 Hybridtechnik

In der **Hybridtechnik** werden im Gegensatz zur normalen Leiterplattentechnik viele passive Bauelemente mit speziellen Schichttechniken direkt auf der Leiterplatte realisiert und anschließend mit diskreten Funktionselementen und integrierten Bausteinen auf dem Trägersubstrat kombiniert. Für die Herstellung der Leiterbahnen und Funktionselemente auf dem Schaltungsträger kommen im Wesentlichen zwei Technologietypen zum Einsatz. Man unterscheidet **Dünnschichttechnik** und **Dickschichttechnik**. Generell gilt, dass die Hybridtechnik – wie die Leiterplattentechnik – sowohl ein separates Schaltungskonzept als auch eine Schaltungsträgervariante für verschiedene ICs darstellen kann.

Bei der Dickschichttechnik werden die Leiterbahn-, Widerstands-, Dielektrikums- und Abdeckstrukturen über einen Siebdruckprozess hergestellt. Hierbei müssen die siebdruckfähigen Pasten durch eine strukturierte Druckform auf dem Substratmaterial aufgebracht und anschließend in einem Wärmeschritt eingebrannt werden. Vor allem Mehrlagenstrukturen in

Dickschichttechnik haben eine sehr hohe Verbreitung. Dazu zählen auch Kernbestandteile von Computeranlagen und Array-Anordnungen für die Montage von Chip-Carriern. In Abbildung 1.2 ist ein solches Dickschichtmodul dargestellt. Man beachte dabei die Pins an den Seiten der Schaltung, die eine Weiterverarbeitung als THT-Bauteil auf einer höheren Hierarchieebene ermöglichen.

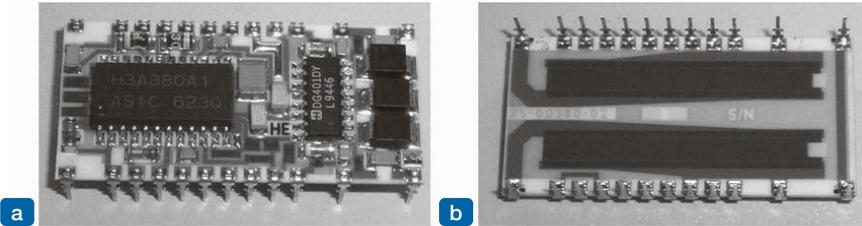


Abbildung 1.2: Vorder- und Rückseite eines Hybridträgers in Dickschichttechnik. Auf der Rückseite in (Teilbild b) sind zwei Widerstände zu erkennen, die getrimmt werden können (ASIC GmbH, D-83550 Emmering).

In der Dünnschichttechnik lassen sich zwei Grundprinzipien der Schichtabscheidung unterscheiden: das thermische Verdampfen eines Materials mittels Elektronenstrahl- oder Widerstandsverdampfung und das Sputtern (siehe *Abschnitt 4.4.1*). Die Strukturierung der Schichten erfolgt mithilfe fotolithografischer Techniken (siehe unten). Mit der Dünnschichttechnik lassen sich daher wesentlich kleinere Abmessungen realisieren als mit der Dickschichttechnik.

Als Substratmaterial werden für Hybridträger meist Keramiken wie z. B. Al_2O_3 oder AlN verwendet. In aufwendigeren Anordnungen kommen auch polymer- oder keramikbeschichtete Metallsubstrate zum Einsatz.

Der Anteil von Dickschichthybridanordnungen am gesamten Hybridträgermarkt liegt bei über 90 % [3]. Dies ist vor allem auf die niedrigeren Kosten, eine einfachere Prozessführung und die hohe Belastbarkeit von Dickschichtaufbauten zurückzuführen. Baugruppen in Schichttechnik können höher integriert werden als Standardmodule der Leiterplattentechnik, außerdem ist die thermische Belastbarkeit meist besser.

1.3 Halbleitertechnik

Unter dem Begriff der **Halbleitertechnik** fasst man allgemein die Realisierung von elektronischen Schaltungen oder optischen Bauelementen mithilfe von Halbleitermaterialien zusammen. Eine besondere Bedeutung kommt dabei den integrierten Schaltungen zu, die als Bauelemente der Halbleitertechnik die Kernbestandteile der modernen Elektronik darstellen.

1.3.1 Integrierte Schaltungen

Eine integrierte Schaltung stellt die Kombination von passiven und aktiven Bauelementen auf einem einzigen monolithischen Halbleiterkristall dar. Auf diesem Wege können sehr zuverlässige und energiesparende Schaltungen auf kleinster Fläche realisiert werden (Abbildung 1.3). Während die Entwicklung und das Design von ICs verhältnismäßig kostspielig sind, werden die Stückkosten bei einer Massenfertigung vergleichsweise niedrig. So stellen hochintegrierte Schaltungen auf Silizium-Basis die Kernbestandteile von Computern, Kameras und allen

modernen Steuerungs- und Kontrolleinheiten dar. Zudem finden ICs aller Art in Kombination mit der Leiterplatten- und Hybridtechnik als Speicher, logische Gatter oder Verstärkerbauelement vielfach Verwendung. Allgemein können integrierte Schaltungen als die Schlüsselkomponenten der modernen Mikroelektronik betrachtet werden.

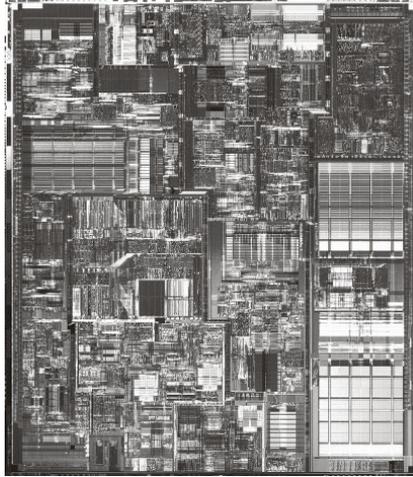


Abbildung 1.3: Intel Pentium IV Mikroprozessor (Foto: Intel Inc.).

Die Fertigung von modernen ICs stellt höchste Anforderungen an die Halbleitertechnologie. Die Verwendung immer neuer Materialien und besserer Verfahren der Fotolithografie und Dünnschichttechnik lässt Bauelementgrößen von unter 30 nm zu. Durch diese hohe Integration werden bezüglich Zuverlässigkeit, Komplexität, Preis pro Bauelement sowie Platz- und Energieverbrauch Werte erreicht, die mit anderen Verfahren der Mikroelektronik nicht annähernd realisiert werden können.

Im praktischen Einsatz benötigen integrierte Bauelemente immer einen Schaltungsträger und werden direkt (**Chip On Board**, siehe *Abschnitt 5.4*), einzeln verpackt (**Single-Chip-Package (SCP)**, *Abschnitt 5.3*) oder in Form eines **Multi Chip Module (MCM)** (siehe *Abschnitt 5.5*) auf einem Hybrid- oder Leiterplattenträger befestigt.

Hinweis

Dieses Buch behandelt im Schwerpunkt die Grundlagen, das Design, das Layout und die Technologie integrierter Bauelemente und Schaltungen. Die Leiterplatten- und Hybridtechnik finden nur in *Kapitel 5* im Rahmen der Aufbau- und Verbindungstechnik integrierter Schaltungen am Rande Erwähnung.

Die Herstellung integrierter Schaltungen basiert auf der **Fotolithografie**; hierbei erfolgt die Strukturierung der unterschiedlichen Bereiche auf dem Chip über eine Belichtung von vorstrukturierten Masken. Ein beispielhafter Schritt zur Strukturierung einer Leiterbahn ist in *Abbildung 1.4* dargestellt. Zunächst wird in *Abbildung 1.4a* die durch eine Oxidschicht passivierte Struktur ganzflächig mit der gewünschten Metallisierung (z. B. AlSiTi) bedeckt. Anschließend erfolgt in *Abbildung 1.4b* der Auftrag mit **Fotoresist**. Dieser ändert bei Ein-

strahlung mit UV-Licht seine chemischen Eigenschaften. In dem hier vorgestellten Beispiel resultiert aus der Bestrahlung mit UV-Licht eine höhere Vernetzung von Kohlenwasserstoffmolekülen innerhalb des Fotoresists, wodurch eine hohe Beständigkeit gegen den **Entwickler** erreicht wird. Durch die sich anschließende Entwicklung lösen sich so nur die unbelichteten Bereiche des Fotoresists. Danach kann der eigentliche strukturgebende Prozessschritt erfolgen, im Beispiel aus Abbildung 1.4 ist dies das Ätzen der Metallisierungsschicht in Abbildung 1.4 e). Im letzten Schritt wird der ausgehärtete Fotoresist wieder entfernt; es bleibt eine strukturierte Leiterbahnebene zurück.

Das Grundprinzip, über die Belichtung von strukturgebenden Masken die einzelnen Ebenen für die Bauelemente auf dem Silizium-Substrat erzeugen zu können, bestimmt die Herstellung von integrierten Schaltungen. Die technologischen Grundlagen für die IC-Fertigung werden in *Kapitel 4* eingehender erläutert; die Beschreibung vollständiger Herstellungsprozesse für grundlegende integrierte Schaltungskonzepte erfolgt in *Kapitel 7*.

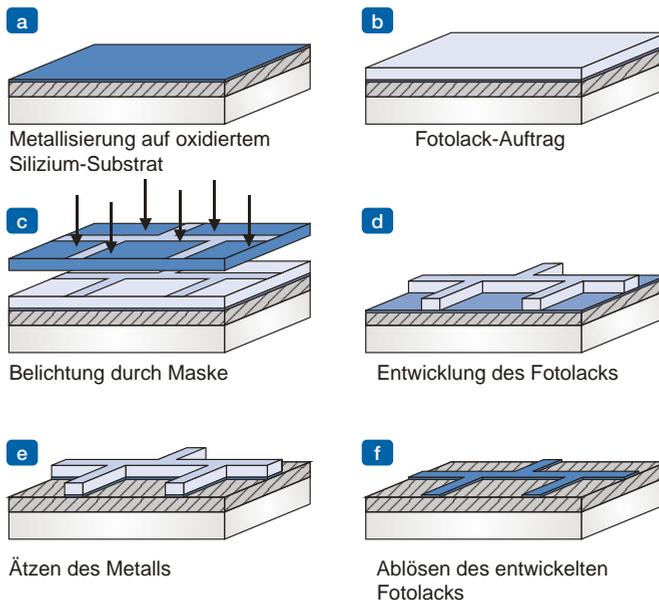


Abbildung 1.4: Herstellung einer Metallisierungsstruktur mittels Fotolithografie unter Verwendung von Negativ-Fotolack.

Die für den Herstellungsprozess notwendigen Masken werden aus dem **Layout** generiert; dieses besteht aus der grafisch darstellbaren Geometrie der für den Schaltungsdesigner relevanten Schaltungsstrukturen auf dem Chip. In Abbildung 1.5 ist das Layout einer einfachen Zwei-Transistor-Struktur (Inverter, siehe *Abschnitt 7.4* und *18.1*) mit dem zugehörigen Schaltplan dargestellt.

Aus Anwendersicht lassen sich vier Gruppen von integrierten Schaltungen unterscheiden. Der IC-Markt wird durch digitale Hochleistungsmikroprozessoren und kostengünstige programmierbare Mikrocontroller einerseits und ASICs sowie Standard-ICs andererseits bestimmt.

Die Grenzen zwischen Mikrocontrollern und Mikroprozessoren sind fließend; Unterschiede sind vor allem durch die Leistungsfähigkeit gegeben. Es besteht aber die Gemeinsamkeit, dass die konkrete Anwendung in gewissen Grenzen variabel ist und durch eine externe Programmierung vorgegeben wird.

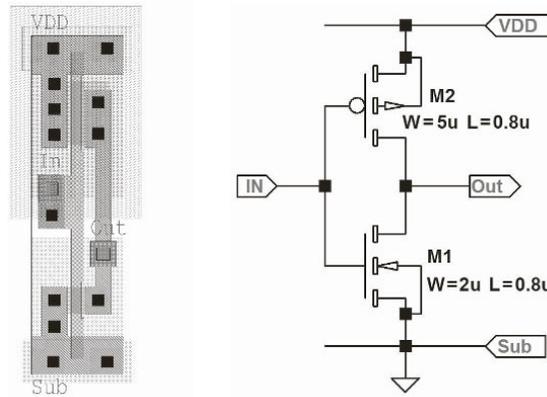


Abbildung 1.5: Layout und Schaltplan einer Inverter-Schaltung in CMOS-Technologie.

Standard-ICs werden als Grundbausteine für die Bestückung von Schaltungsträgern verkauft; die Funktionen sind vorher festgelegt. So können digitale Gatter oder Flip-Flops und analoge Verstärkerschaltungen erworben und auf einem Schaltungsträger mit Einzelbauelementen zu der gewünschten Schaltung zusammengefasst werden (z. B. auf einer Leiterplatte oder einem Hybridträger, siehe oben).

Bei den sogenannten ASICs (**Application Specific Integrated Circuit**) erfolgen der Entwurf und die Fertigung einer integrierter Schaltung nach den Erfordernissen eines konkreten Kundenwunsches (Abbildung 1.6). Auf diese Weise kann die Schaltung hinsichtlich der Performance, des Leistungsverbrauchs oder des Flächenbedarfs für die gewünschte Funktion optimiert werden.

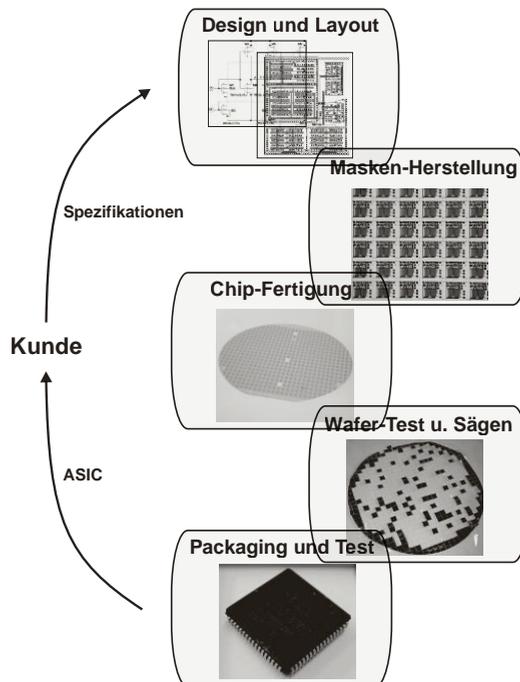


Abbildung 1.6: Der Weg vom Kundenwunsch zur integrierten Schaltung.

Auf Anwenderseite war man in der Vergangenheit – speziell im Low-Cost-Sektor – oft bemüht, nach Möglichkeit auf die Entwicklungskosten produktspezifischer ASICs zu verzichten und stattdessen Lösungen durch Verschalten von Standard-ICs zu finden. Doch gegenwärtig zeichnet sich in vielen Unternehmen die Tendenz ab, Kernfunktionen auch ohne technische Notwendigkeit in ASICs zusammenzufassen, um dadurch einen möglichen „Technologieklaue“ zu erschweren. Im Zuge dessen gehen Firmen dazu über, für Design und Layout der benötigten ICs eigene Abteilungen aufzubauen. Das fertige Layout eines ASIC kann dann an eine Auftragsfabrik (**Foundry**) weitergegeben werden, die die eigentliche Fertigung übernimmt.

Mikroprozessoren, Standard-ICs und ASICs lassen sich in Abhängigkeit der Anforderungen und des Anwendungsgebietes durch unterschiedlich aufwendige Prozesstechnologien realisieren (siehe *Kapitel 4* und *Kapitel 7*). Ein wichtiges Merkmal für einen IC-Prozess ist der jeweilige **Integrationsgrad**. Durch immer weiter optimierte Lithografie- und Fertigungsverfahren, verbesserte Materialien und neue Technologiekonzepte sind die minimalen Bauteilabmessungen stetig verringert worden; die Anzahl der Schaltungskomponenten auf den Chips hat sich zeitgleich massiv erhöht. In Tabelle 1.1 sind die gängigen Integrationsstufen und das jeweilige Jahr der Entwicklung zusammenfassend dargestellt.

Es soll an dieser Stelle nicht unerwähnt bleiben, dass die sehr hohen Integrationsgrade hauptsächlich für digitale Mikroprozessoren von maßgebender Bedeutung sind. Für analoge integrierte Schaltungen sowie die meisten Standard-ICs und ASICs sind niedrigere Integrationsgrade meist technisch ausreichend und in der Herstellung wesentlich kostengünstiger.

Tabelle 1.1

Integrationsgrade integrierter Schaltungen [4].			
Scale	Bedeutung	Transistoren/Chip	Jahr
SSI	Small Scale Integration	1–10	1964
MSI	Medium Scale Integration	10–100	1967
LSI	Large Scale Integration	100–1000	1972
VLSI	Very Large Scale Integration	10^3 – 10^4	1978
ULSI	Ultra Large Scale Integration	10^4 – 10^7	1989
GSI	Giant Scale Integration	$>10^7$	2000

Auf Basis der in *Kapitel 7* vorgestellten praxisrelevanten Prozesstechnologien beschäftigt sich dieses Buch in den *Kapiteln 8–18* vor allem mit dem Design und dem Layout von ASICs und Standardschaltungen.

1.3.2 Aufbau einer integrierten Schaltung

Aus Abbildung 1.7 geht hervor, wie eine integrierte Schaltung im Prinzip aufgebaut ist. Die Schaltungsteile, d. h. Standardzellen, spezielle Zellen, analoge Zellen usw. befinden sich im inneren Teil des Chips, dem sogenannten **Core**.

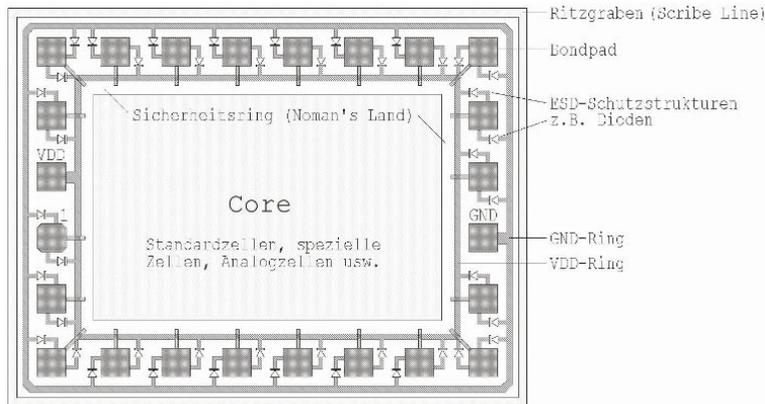


Abbildung 1.7: Prinzipieller Aufbau einer integrierten Schaltung.

Um den Chip-Core ist der **Pad-Ring** angeordnet. Über die Bondpads wird die Verbindung zur Außenwelt hergestellt. Die Pads sind in der Regel mit ESD-Schutzdioden (**Electro Static Discharge**) ausgestattet, damit von außen einwirkende elektrostatische Entladungen im Innern des Chips, d. h. im Core, keinen Schaden anrichten können [9]. Die Dioden sorgen dafür, dass Signale an den Bondpads die Potentiale der Versorgungsleitungen nicht um mehr als eine Diodenspannung über- bzw. unterschreiten können. Die Versorgungsleitungen (**Power-Rails**) VCC (bipolare Schaltungen) bzw. VDD (MOS-Schaltungen) und GND (Ground) werden meist ringförmig in die Chip-Peripherie gelegt. Bei einem Standard-Bipolar- oder einem CMOS-Prozess (siehe *Kapitel 7*) stellt GND als äußerster Ring die Verbindung zum Substrat her. Der innere Ring führt dagegen meist die positive Versorgungsspannung VCC bzw. VDD. Oft wird auch **innen ein zweiter GND-Ring** vorgesehen. Dieser ist in *Abbildung 1.7* nicht dargestellt.

Zwischen dem Core und dem inneren Versorgungsring bleibt in der Regel ein **Sicherheitsring (Noman's Land)** frei von aktiven Elementen. Hier sind nur „Verdrahtungsleitungen“ erlaubt.

Ganz außen wird der Chip durch den sogenannten **Ritzgraben (Scribe Line** oder **Scribe Street**) begrenzt. Wie der Name schon andeutet, ist dieser Platz für das Auftrennen des Wafers in die einzelnen Chips vorgesehen. In früheren Zeiten erfolgte das Trennen mittels eines „Ritzdiamanten“, daher der Name Ritzgraben. Heute wird fast ausschließlich ein Trennschleifverfahren verwendet, siehe *Abbildung 5.2*.

Der gesamte Pad-Rahmen, einschließlich Bondpads mit ESD-Schutzstrukturen, Versorgungsringen und Ritzgraben wird in der Regel vom Halbleiterhersteller in Form einer Chip-Rahmencelle zur Verfügung gestellt. Diese kann dann an den eigenen Chipentwurf angepasst werden. Eventuell müssen einige Standard-Pads durch spezielle, für andere Spannungswerte zugelassene Pads ausgetauscht werden.

Abbildung 1.8 zeigt ein Beispiel, das die Arbeiten von Studierenden der FH Hannover enthält. Es besteht aus zwei Teilen und wurde im Rahmen eines Multi-Projekt-Wafer-Runs im „CM5-Prozess“ (siehe Prozessbeschreibungen in den *Kapiteln 7* und *8*) gefertigt. Die gesamte Fläche beträgt ungefähr 5 mm^2 .

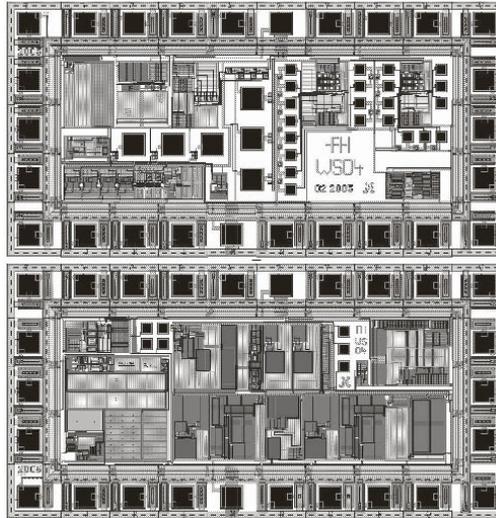


Abbildung 1.8: Beispiel eines Chip-Layouts mit einer Fläche von ca. 5 mm^2 .

1.3.3 Entwurfsprozess integrierter Schaltungen

Es ist das Ziel eines jeden Schaltungsdesigners, anhand der Spezifikationen und gewünschten Funktionen eines geplanten ICs mit möglichst geringem Aufwand das Layout als Grundlage für die Chipfertigung erstellen zu können. Zusätzlich sollte dieser Chip idealerweise platzsparend, verlustarm, schnell und kostengünstig in der Herstellung sein. Dabei ist es wünschenswert, möglichst viele Prozessschritte beim Entwurf durch automatisierte Software-Prozeduren bearbeiten zu können. Speziell bei hochintegrierten digitalen Mikroprozessoren liegen „Welten“ zwischen der abstrakten Beschreibung der Hardwarefunktionen, den elektrotechnischen Prinzipien der Schaltungstechnik und den physikalischen Abläufen auf Bauelementebene. Diese Unterschiede in den Zugängen zur Layout-Erstellung lassen sich am besten anhand von Abbildung 1.9 verdeutlichen: Ob über die direkte Ableitung aus dem Schaltplan oder über abstraktere Darstellung mittels Hardwarebeschreibungssprachen: Es bestehen in Abhängigkeit von der Schaltungsart und -anforderung prinzipiell mehrere Ansätze, um von einer Schaltungsspezifikation zu einem fertigen Layout zu gelangen.

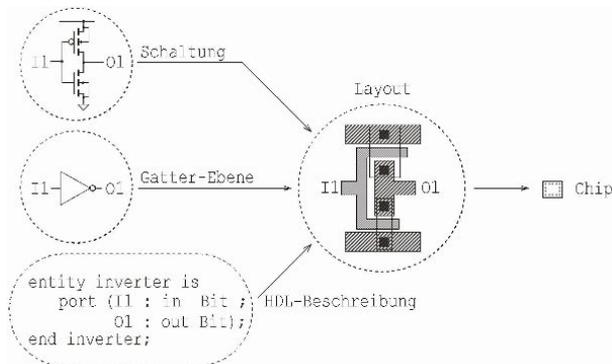


Abbildung 1.9: Der Weg zum Layout.

Dabei gibt es keine standardisierten Entwurfsabläufe, die für alle Arten von analogen und digitalen integrierten Schaltungen Gültigkeit haben. Allgemein ist aber die Einteilung aus Abbildung 1.10 für eine erste Betrachtung der unterschiedlichen Entwurfsebenen zweckmäßig.

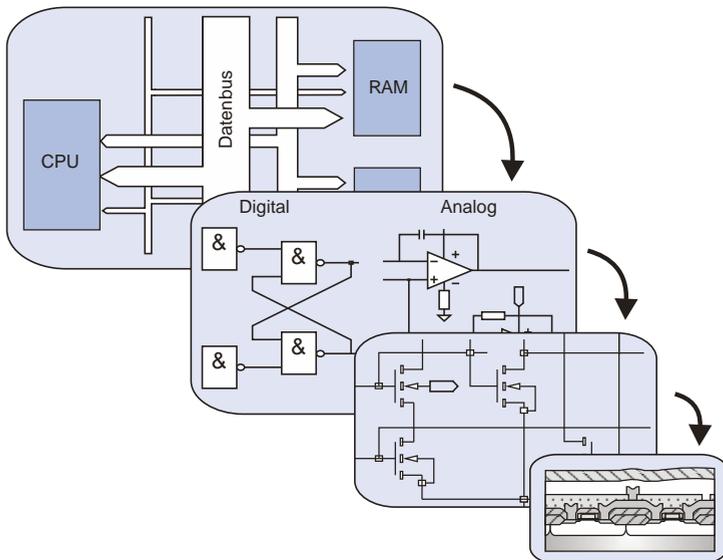


Abbildung 1.10: Schematische Darstellung der unterschiedlichen Ebenen des IC-Entwurfs: Ein Schaltkreis lässt sich entsprechend der Komplexität in eine unterschiedliche Anzahl von Blöcken unterteilen (die Darstellung oben links gilt nur für ein digitales System), diese wiederum können in Gatter- oder Baugruppenpläne zerlegt werden. In den darunterliegenden Betrachtungsebenen steht der Schaltplan der Untereinheiten bis hinab zum einzelnen Bauelement im Mittelpunkt.

Am Beginn des Schaltungsentwurfs steht der **Systementwurf**. Dabei werden die grundlegenden Merkmale der zu erstellenden Schaltung festgelegt. So ist neben der Bestimmung fundamentaler Rahmendaten wie Anforderungsprofil, genereller Abmessungen und Wahl der Prozess-technologie eine erste Strukturierung in Funktionsgruppen vorzunehmen. In der Folge findet – abhängig von der Schaltungsart und -komplexität – eine Unterteilung in weitere Entwurfsebenen statt, wobei sich die klar unterscheidbaren Entwurfsprozeduren für digitale und analoge Schaltungen in vielerlei Hinsicht unterscheiden. Der Weg vom Entwurf einer funktionalen Struktur bis hin zum vollständigen Layout lässt sich für digitale Chips in wesentlich stärkerem Maße automatisieren; die einzelnen Bauelemente und deren Verhalten innerhalb der Schaltung treten für den Digital-Designer dabei zunächst in den Hintergrund.

So erfolgt für den Entwurf hochintegrierter digitaler Chips in der Literatur eine Unterteilung in den **Architektur-, Verhaltens-, Registertransfer- und Logikentwurf**, bei dem das System jeweils in feinere Abstraktionsebenen unterteilt wird. Der Schaltungsentwurf und die -simulation finden zumeist mithilfe von Hardwarebeschreibungssprachen wie Verilog oder VHDL statt, wodurch ohne Rücksichtnahme auf herstellungsspezifische Bauelementeigenschaften eine Schaltung ereignisgesteuert entworfen und simuliert werden kann, die eigentlichen elektrotechnischen Abläufe sind für den Digital-Designer zunächst nur von geringer Bedeutung. Trotzdem greifen auch hierbei die Entwurfsprogramme an verschiedenen Punkten der Simulation und Verifikation (siehe unten) in Form von Standardzellen auf die relevanten technischen Daten der einzelnen logischen Baugruppen zurück. Um ein hochkomplexes digitales System mit einer Hardwarebeschreibungssprache zuverlässig erstellen zu können, müssen die Eigenschaften der einzelnen digitalen Grundgatter daher zuvor auf Transistorebene ermittelt werden (siehe *Teil IV*).

Bei analogen Schaltungen ist die Beschreibung der Zustände einzelner Bauelemente wesentlich komplexer, sodass beim Schaltungsentwurf nach einer groben Einteilung in einzelne Funktionsblöcke in den meisten Fällen direkt die Simulation auf **Transistorebene** mit geeigneten Simulationsprogrammen wie z. B. SPICE erfolgt. Diese Programme erlauben eine Funktionsüberprüfung der Schaltung über die aus dem Schaltplan (**Schematic Entry**) generierte Netzliste, wobei die Eigenschaften der einzelnen Transistoren, Widerstände oder Kondensatoren anhand von Bauelementmodellen direkt einfließen. Als Beispiel sei hier eine SPICE-Circuit-Datei für die Inverter-Schaltung aus Abbildung 1.5 angegeben. Sie besteht aus einer **Titelzeile**, der eigentlichen **Netzliste**, die die Schaltung beschreibt, der **Modellanweisung** für die nähere Beschreibung der Bauelemente, einer externen **Beschaltung**, der **Simulationsanweisung** und einer **End-Anweisung**:

```
*CMOS-Inverter
M1 Out IN VDD VDD MP7 W=5u L=0.8u
M2 Out IN 0 0 MN7 W=2u L=0.8u
.lib D:\Spicelib\CM5\CM5-N.phy
VDD VDD 0 1.5V
VIN IN 0 Pulse(0V 1.5 10ns 1ns 1ns 30ns 50ns)
.Tran 200ns
.end
```

Die Bezeichnungen der Einzelbauelemente (hier die Transistoren M1 und M2) stehen zu Beginn jeder Zeile, dahinter sind die Knoten zu finden, die den Einbau des Bauelementes in der Schaltung beschreiben (Knoten „IN“, „OUT“, „VDD“ und „0“). Zusätzlich werden bei MOS-Transistoren die geometrischen Daten (Kanallänge und -weite) angegeben.

Die Modelle für Bauelemente oder ganze Gatter sind Teil einer Bauelement-Bibliothek, die zumeist vom Anbieter des Herstellungsprozesses zur Verfügung gestellt wird (siehe *Teil II*).

Bei der **Layout-Synthese** wird die Netzliste schließlich in das für die eigentliche Chipfertigung erforderliche Layout übertragen. Für hochintegrierte digitale Systeme ist dieser Prozess wiederum hochautomatisiert und wird in Partitionierung, Floorplanning, Platzierung, Verdrahtung und Kompaktierung unterteilt. Im analogen Schaltungsdesign werden die Maskenlayouts mit sogenannten Layout-Generatoren (z. B. Cadence, L-Edit) „per Hand“ oder nur teilautomatisiert erzeugt. So ist es zwar möglich, sich das Maskenlayout von einzelnen elektrischen Bauelementen anhand der Modellparameter generieren zu lassen, doch sind eine Minimierung der genutzten Chipfläche und ein möglichst störungsfreier Entwurf nicht durch Computerprogramme systematisierbar. Um das Layout für die Herstellung einer integrierten Schaltung anfertigen zu können, muss zunächst die gewünschte Funktion und der Aufbau der Schaltung im Schaltplan (**Schematic**) bekannt sein. Der Übergang von einem für diskrete Bauelemente geplanten Schaltplan zu dem Layout einer ganzen integrierten Schaltung führt allerdings zu mehreren Problemstellungen, wie beispielsweise eng nebeneinanderliegende aktive Halbleiterbereiche, die unbeabsichtigt über parasitäre Bauelemente (wie z. B. Dioden oder Transistoren) miteinander verknüpft sein können – und so das eigentlich beabsichtigte Schaltungsverhalten verändern. Hier ist das Können des einzelnen Analog-Designers gefragt (*Kapitel 8–14*)!

Um die Richtigkeit des Maskendesigns überprüfen zu können, folgt dem Layout-Entwurf die **Layout-Verifikation**. Diese beginnt mit dem **DRC (Design-Rule-Check)**, wodurch sichergestellt wird, dass alle technologisch bedingten Entwurfsregeln erfüllt sind. Die Design-Rules sind prozessabhängige Größen und ändern sich für jede neu eingeführte Technologieform (siehe *Abschnitt 8.3*). Der **LVS-Test (Layout Versus Schematic)** wird mithilfe des Layout-Programms durchgeführt. So lässt sich aus dem Maskendesign wiederum eine Netzliste erstellen, die genaue Informationen über Ort und Parameter der einzelnen Bauelemente liefert. Wenn man diese Netzliste mit derjenigen aus der Schaltungssimulation auf Transistorebene vergleicht, dürfen keine Widersprüche auftauchen (*Abschnitt 8.7*). Vor allem bei manuellen Entwürfen analoger Schaltungen sind nach einzelnen Syntheseschritten (Transistorebene \Rightarrow Layout-Synthese) immer wieder Soll-Ist-Kontrollen vorzunehmen, sodass kein geradliniger Ablauf entsteht, sondern manche Schritte mehrfach wiederholt werden müssen.

Sofern das entworfene Design alle Anforderungen erfüllt, folgt die für eine Produktion notwendige **Maskenherstellung**. Hierbei wird mithilfe eines Elektronenstrahlenschreibers oder dem Pattern-Generator der eigentliche Maskensatz für die Fotolithografie hergestellt (*Abschnitt 4.2*).

Mit den Masken kann die parallele **Fertigung** einer großen Anzahl von Schaltungen auf einem Wafer vorgenommen werden (*Kapitel 4* und *7*).

Anschließend werden die ICs nach der mehrstufigen Fabrikation auf dem Wafer einem ersten elektrischen **Eignungstest** unterzogen (*Kapitel 6*).

Danach wird der IC mit einem anwendungsspezifischen Träger, Anschlüssen und einem Gehäuse versehen (**Packaging**, *Kapitel 5*).

Den Abschluss der Schaltungsherstellung bildet die **Endkontrolle** des fertigen Bauelements (*Kapitel 6*).

Zusammenfassung

Der inhaltliche Schwerpunkt dieses Buches liegt auf dem nicht automatisierten Design und Layout integrierter Schaltungen. Der Entwurf digitaler Systeme oberhalb der Bauelementebene wird in der Literatur vielfach behandelt und ist nicht Teil dieses Buches.

Literatur

- [1] W. Scheel (Hrsg.): *Baugruppentechologie der Elektronik – Montage*; Verlag Technik, Berlin, 2. Auflage, 1999.
- [2] R. R. Tummala, E. J. Rymaszewski, A. G. Klopfenstein: *Microelectronics Packaging Handbook – Semiconductor Packaging*; Kluwer Academic Publ., 2. Auflage, 1997.
- [3] H.-J. Hanke: *Baugruppentechologie der Elektronik – Leiterplatten*; Verlag Technik, Berlin, 1994.
- [4] S. Ramachandran: *Digital VLSI Systems Design*; Springer-Verlag, Dordrecht, 2007.
- [5] J. Lienig: *Layoutsynthese elektronischer Schaltungen – Grundlegende Algorithmen für die Entwurfsautomatisierung*; Springer-Verlag, Berlin, Heidelberg, 2006.

- [6] G. H. Schildt, D. Kahn, C. Kruegel, C. Moerz: *Einführung in die Technische Informatik*; Springer-Verlag, Wien, 2005.
- [7] H.-J. Wunderlich: *Hochintegrierte Schaltungen: Prüfgerechter Entwurf und Test*; Springer-Verlag, Berlin, Heidelberg, 1991.
- [8] J. P. Uyemura: *Physical Design of CMOS Integrated Circuits Using L-Edit*; PWS Publishing Company, Boston, 1995.
- [9] A. Hastings: *The Art of Analog Layout*; Prentice Hall, 2001.

Physikalische Grundlagen der Halbleitertechnik

2.1	Grundlagen der Halbleiterphysik	37
2.2	Vom einzelnen Quantentrog zur Bandstruktur von Halbleitern	38
2.3	Bandstruktur und Ladungsträgertransport	55
2.4	Eigenschaften von Silizium	65

Einleitung

» Im vorliegenden Kapitel werden die physikalischen Grundlagen moderner Halbleiterphysik kompakt dargestellt. Diese dienen als Grundlage für das Verständnis einer Bandstruktur in Halbleitern und damit der Funktionsweise von Halbleiterbauelementen.

Zur Bewertung neuer Technologien und Konzepte der Nanoelektronik – und es ist auch Ziel des vorliegenden Buches, diese Kompetenz zu vermitteln – muss notwendigerweise ein quantenmechanisches Grundverständnis vorhanden sein. Im Folgenden wird deshalb eine kompakte Einführung in den Formalismus der Quantenmechanik an den Anfang einer Beschreibung der Eigenschaften von Halbleitern gestellt.

Für das Verständnis der integrierten Schaltungstechnik selbst greift man später aber wieder auf eine parametrisierte Beschreibung des Verhaltens von Halbleiterbauelementen zurück. Die Kapitel zur integrierten Schaltungstechnik können deshalb auch ohne ein tieferes Eindringen in die physikalischen Grundlagen bearbeitet werden.

Die Eigenschaften von Halbleitern werden durch die atomaren Bindungen der Konstituenten im Halbleiterkristall bestimmt. Manche Halbleiter können Licht emittieren, andere sind gute elektrische Leiter. Wieder andere haben herausragende magnetische Eigenschaften, sind supraleitend oder können als Isolator verwendet werden. Nur eine quantenmechanische Beschreibung von Halbleitern kann derartige Unterschiede erklären.

In diesem Kapitel werden zunächst das Gedankengebäude sowie der Formalismus einer quantenmechanischen Beschreibung vorgestellt. Besonderes Augenmerk liegt auf einer knappen, prägnanten Darstellung, um möglichst zügig zu einer qualitativen Erklärung der Bandstruktur (das ist die Gesamtheit aller zur Verfügung stehenden Energieniveaus) in Halbleitern zu kommen. Dabei können viele Besonderheiten einer quantenmechanischen Beschreibung nicht ausreichend detailliert beschrieben werden. An diesen Stellen wird dann auf die weiterführende Literatur verwiesen.

Um Halbleiter in elektronischen Bauelementen zu verwenden, muss deren elektrische Leitfähigkeit kontrolliert eingestellt werden. Die Eigenschaften von Ladungsträgern wie Elektronen oder Löcher sowie deren Einfluss auf die Leitfähigkeit von Halbleitern werden vorgestellt.

Ströme können durch ein elektrisches Feld (Feldstrom) oder durch einen Konzentrationsgradienten (Diffusionsstrom) angetrieben werden. Beide Anteile sind für die Berechnung von Halbleiterbauelementen zu berücksichtigen. Neben der zugehörigen Driftdiffusionsgleichung müssen auch Kontinuitätsbedingungen erfüllt werden: Ladung kann im Laufe der Zeit nicht verloren gehen. Die Driftdiffusionsgleichung sowie die Kontinuitätsgleichung stellen die Grundlage für die Berechnung der Eigenschaften von Halbleiterbauelementen wie Dioden oder Transistoren dar. «

LERNZIELE

- Quantenmechanische Grundlagen
- Entstehung von Bandstruktur
- Leitungs- und Valenzbänder
- Direkte und indirekte Halbleiter
- Dotierung von Halbleitern: Ladungsträgerstatistik
- Fermi-Dirac-Verteilung, Fermi-Energie
- Effektive Masse
- Driftdiffusionsgleichung, Kontinuitätsgleichung
- Eigenschaften von Silizium

2.1 Grundlagen der Halbleiterphysik

Der Begriff Halbleiter bezieht sich auf Materialien, deren Leitfähigkeit in weiten Grenzen verändert werden kann, von „isolierend“ bis „metallisch“. Es ist gerade diese Variabilität, die Halbleiter für elektronische Bauelemente interessant macht. In Silizium kann z. B. durch eine von außen angelegte Spannung ein elektrisches Feld im Inneren des Halbleiters erzeugt und damit die Elektronenkonzentration verändert werden. Dieser „Feldeffekt“ schaltet einen Silizium-Kanal zwischen „isolierend,“ und „leitfähig“ hin und her. Elektronische Schalter sind die Grundlage integrierter digitaler Schaltungen ebenso wie für analoge integrierte Schaltungen wie z. B. Digital-Analog-Wandler oder Verstärker. Später werden wir diesen Feldeffekt und dessen physikalische Ursachen detailliert beschreiben.

Die elektronischen Eigenschaften von Halbleitern sind mit Methoden der klassischen Physik, d. h. im Rahmen der Mechanik und Elektrodynamik, grundsätzlich nicht umfassend zu verstehen. Es muss stattdessen eine quantenmechanische Beschreibungsweise gewählt werden. Dies liegt daran, dass Elektronen Wellencharakter haben und der quantenmechanischen Grundgleichung, der Schrödinger-Gleichung, gehorchen. Es sind Elektronenwellen, die sich durch den Atomverbund ausbreiten, den wir Kristallgitter nennen, und eben keine Punktladungen und schon gar keine kontinuierlichen Ströme. Erst die Lösung der Schrödinger-Gleichung liefert alle möglichen Energieniveaus, die die Elektronen in Halbleitern besetzen können. Diese Energieniveaus werden dann vom niedrigsten Energieniveau beginnend sukzessive von unten nach oben aufgefüllt. Manchmal wird dies verglichen mit dem Befüllen eines Eimers mit Wasser. Ganz analog können nämlich zwei Elektronen auch nicht in einem identischen Energieniveau sitzen. Die Gesamtheit aller Energieniveaus in Halbleitern wird Bandstruktur genannt. Die Bandstruktur ist die Grundlage für alle elektronischen, optischen und magnetischen Eigenschaften von Halbleitern.

Oft wird aus Gründen der Vereinfachung zur Beschreibung der elektronischen Eigenschaften von Halbleitern trotzdem ein klassischer Ansatz gewählt. Man betrachtet dann nicht mehr die Bewegung von Elektronenwellen, sondern beschreibt Ströme, die durch elektrische Felder (Feldstrom) oder Gradienten in der Ladungsträgerkonzentration (Diffusionsstrom) angetrieben werden. Hiermit kann die prinzipielle Funktionsweise von Dioden und Transistoren erstaunlich gut berechnet werden. Ein klassischer Ansatz zur Beschreibung von Halbleiterei-

enschaften ist einfacher und anschaulicher als ein quantenmechanischer. Hierzu müssen allerdings vereinfachende Modelle verwendet werden, die letztlich auch wieder in der richtigen, quantenmechanischen Beschreibung ihren Ursprung haben und durch diese begründet werden. Außerdem bleiben bei einer nur klassischen Beschreibung von Halbleitern viele grundlegende Aspekte zwangsläufig unverstanden. Elektronen sind eben doch Wellen, die sich durch ein periodisches Potential, den Halbleiterkristall, ausbreiten.

Eine vollständige quantenmechanische Beschreibung von Halbleitereigenschaften würde weit über die geplanten Inhalte dieses Buches hinausgehen. Trotzdem ist gerade für das Verständnis der neuesten Entwicklungen im Bereich der CMOS-Technologie und der Nanobaulemente quantenmechanisches Grundwissen wichtig, um zumindest gemeinsame Begriffe definieren zu können. Dieses Grundwissen kann aber insbesondere bei Studierenden der Elektrotechnik nicht von vornherein vorausgesetzt werden. Zunächst werden deshalb kurz die Grundlagen einer quantenmechanischen Beschreibung zusammengefasst, um zumindest im Prinzip und qualitativ im Rahmen des quantenmechanischen Formalismus die Entstehung von Bandstruktur und damit von Leitungs- und Valenzbändern nachvollziehen zu können. Erst wenn diese Grundlagen gelegt sind, werden wir Vereinfachungen treffen, um so die Eigenschaften von Halbleitern und die Arbeitsweise von realen elektronischen Bauelementen zu beschreiben.

Dabei bleibt die Natur von Elektronen aber immer wellenartig und letztlich ist es immer die Quantenmechanik, die die Arbeitsweise von elektronischen Bauelementen bestimmt. Wie stark elektronische Bauelemente heutzutage „quantisiert“ sind, lässt sich z. B. schon an der Zahl der Elektronen in den kleinsten heute hergestellten Transistoren ablesen. Der aktive Teil eines Transistors besteht heute nur noch aus wenigen 1000 Elektronen. Ein weiterer Aspekt, der die Beschränktheit einer klassischen Beschreibung demonstriert, ist die Tatsache, dass Silizium kein Licht emittieren kann, während aus GaN und GaAs sehr effiziente Lichtemitter (LEDs und Laserdioden) hergestellt werden können. Die Ursache liegt in Details der Bandstruktur von Silizium. Darüber hinaus werden heutzutage „Einzelelektronentransistoren“ erforscht, bei denen sich immer nur ein einziges Elektron im Transistor befindet. Einzelne Elektronen können wiederum einen quantenmechanischen Zustand repräsentieren, der dann die Grundlage für einen Quantencomputer sein kann. Auch dies ist Gegenstand aktueller Forschung.

Diese wenigen Beispiele sollen zeigen, dass zur Bewertung neuer Technologien und Konzepte – und es ist Ziel dieses Buches, gerade auch diese Kompetenz zu vermitteln – notwendigerweise ein quantenmechanisches Grundverständnis vorhanden sein muss. Im Folgenden wird deshalb der Formalismus der Quantenmechanik an den Anfang einer Beschreibung der Eigenschaften von Halbleitern gestellt.

2.2 Vom einzelnen Quantentrog zur Bandstruktur von Halbleitern

2.2.1 Der Formalismus der Quantenmechanik

Als Grundgleichung der Quantenmechanik betrachten wir zunächst die Schrödinger-Gleichung (genauer: die zeitunabhängige Schrödinger-Gleichung). Diese Gleichung erlaubt die Berechnung der möglichen quantisierten Gesamtenergien eines Systems.

$$H\Psi(\vec{r})=E\Psi(\vec{r}) \quad (2.1)$$

H ist der Hamilton-Operator

$$H = -\frac{\hbar^2}{2m} \Delta + V(\vec{r}) \quad (2.2)$$

Δ ist der Laplace-Operator

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad (2.3)$$

und $V(\vec{r})$ die potentielle Energie als Funktion des Orts $\vec{r} = (x, y, z)$. \hbar ist das Planck'sche Wirkungsquantum dividiert durch 2π , und m die Masse des Objektes, also z. B. die Masse eines Elektrons. E ist die Gesamtenergie des zu beschreibenden Systems.

Der Hamilton-Operator besteht aus einem Differentialoperator, dem Laplace-Operator und der potentiellen Energie $V(r)$. Es handelt sich bei der Schrödinger-Gleichung demnach um eine Eigenwertgleichung, deren Lösung aus einem Satz von Eigenfunktionen $\Psi_n(\vec{r})$ und Eigenwerten E_n besteht. Dabei bezeichnet der Index n einen Satz von Quantenzahlen. In einfachen Fällen läuft die Quantenzahl n als natürliche Zahl von 1 bis unendlich (wie z. B. bei Elektronen in eindimensionalen Potentialtöpfen, siehe später) und nummeriert damit die Lösungsfunktionen. Bei dreidimensionalen Problemen erhält man drei Quantenzahlen (das sind dann z. B. die Quantenzahlen n , l und m für ein Elektron, das an ein Proton gebunden ist, und die die Energieniveaus eines Wasserstoff-Atoms beschreiben). Wir werden auf den eindimensionalen Potentialtopf später noch genauer eingehen.

Die Lösungsfunktionen werden Wellenfunktionen genannt, Diese bestehen im Allgemeinen aus einem Real- und einem Imaginärteil. Das Betragsquadrat der Wellenfunktionen $\Psi_n(x)$ ergibt eine Wahrscheinlichkeitsdichte. Integriert über ein bestimmtes Volumen im Raum kann man daraus die Aufenthaltswahrscheinlichkeit berechnen, mit der man in genau diesem Volumen das Objekt (in diesem Fall ein Elektron) nachweisen kann. Die Integration über den gesamten Raum muss 1 ergeben.

Hinweis

Der Begriff „Aufenthaltswahrscheinlichkeit“ ist leicht irreführend. Man könnte annehmen, dass sich ein Objekt (hier das Elektron) mal an diesem und mal an jenem Ort befindet, und sich dadurch eine gewisse „Aufenthaltswahrscheinlichkeit“ ergibt. Diese Vorstellung ist vielfach verbreitet, aber grundsätzlich falsch. Das quantenmechanische Objekt (hier das Elektron) befindet sich nämlich an verschiedenen Orten gleichzeitig. Die „Aufenthaltswahrscheinlichkeit“ gibt an, welcher „Prozentanteil“ des Elektrons sich wo befindet. Diese Vorstellung eines im Raum verschmierten Elektrons kommt uns nur deshalb merkwürdig vor, weil wir das Elektron meist als Teilchen betrachten, was es aber nicht ist. Im Gegensatz dazu ist uns für ein Wellenobjekt die Tatsache, dass das Objekt an mehreren Orten gleichzeitig auftritt, sehr gut bekannt und dann auch völlig problemlos.

Allerdings darf man nun auch nicht annehmen, dass wir Bruchteile von Elektronen nachweisen könnten. Genau hier liegt die Eigenart einer quantenmechanischen Beschreibungsweise. Wollen wir das Elektron nachweisen, das über ein großes Volumen gleichzeitig verschmiert

ist, so können wir nur zwei Ergebnisse erhalten: Entweder wir weisen ein (ganzes) Elektron nach oder eben kein Elektron. Hier macht sich die inhärente Quantisierung der Natur bemerkbar. Diese tritt übrigens auch bei elektromagnetischen Wellen auf. Auch dort kann man ein (ganzes) Photon, also ein „Lichtteilchen“, entweder messen oder eben nicht.

Diese merkwürdige Mischung von teilchenartigen und wellenartigen Eigenschaften führt dazu, dass die uns in der Natur präsentierte Realität nicht durch eine nur wellenartige oder nur teilchenartige Beschreibung erfasst werden kann. Das gleichzeitige Auftreten von Wellen- und Teilcheneigenschaften wird oft als „Welle-Teilchen-Dualismus“ bezeichnet, auch dies ist allerdings eine eher unglückliche, historisch gewachsene Bezeichnung. Ein Elektron tritt eben nicht manchmal als Welle und manchmal als Teilchen auf. Es ist ein quantenmechanisches Objekt mit genau einer wohlbekanntem und in der Zeit konstant gültigen konsistenten Beschreibung.

Wenden wir uns wieder der Schrödinger-Gleichung zu. In der zeitunabhängigen Schrödinger-Gleichung kommt die Zeit offensichtlich nicht vor. Das bedeutet aber, dass durch diese Gleichung nur stationäre, also zeitunabhängige Situationen beschrieben werden können. Dynamische, also zeitabhängige Vorgänge werden damit grundsätzlich nicht erfasst. Hierzu müsste man dann die zeitabhängige Schrödinger-Gleichung bemühen, die wir allerdings an dieser Stelle nicht einführen wollen. Hierzu wird auf Bücher zur Einführung in die Quantenmechanik verwiesen (z.B. [7] und [8]). Dies stellt aber keine wesentliche Einschränkung dar, da zur Beschreibung quantenmechanischer Systeme meist die zeitunabhängige Schrödinger-Gleichung völlig ausreicht.

Zur Verdeutlichung des Formalismus wenden wir diesen nun auf ein einfaches, aber recht lehrreiches System an. Ein Elektron soll sich in einem rechteckigen Potentialtopf befinden. An diesem Beispiel kann man alle wesentlichen Aspekte des quantenmechanischen Formalismus verfolgen, und zwar gerade ohne eine komplexe mathematische Behandlung. Trotz der übersichtlichen Form der Schrödinger-Gleichung ist deren mathematische Behandlung nämlich meist kompliziert. Zur weiteren Vereinfachung berücksichtigen wir deshalb außerdem nur noch eine Koordinate x , statt der 3 Raumkoordinaten x , y und z , und nehmen darüber hinaus an, dass der Potentialtopf unendlich hohe Wände hat. Die Situation ist in Abbildung 2.1 skizziert.

Später werden wir dann diese Potentialtöpfe aneinanderreihen und dies als Modell für den atomaren Verbund von Atomen innerhalb eines Halbleiters diskutieren.

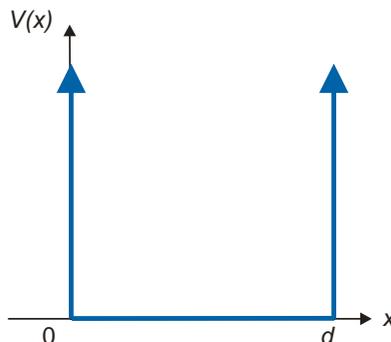


Abbildung 2.1: Potentialverlauf in einem rechteckigen Potentialtopf mit unendlich hohen Wänden.

Innerhalb des Potentialtopfs sei die potentielle Energie $V(x) = 0$, außerhalb des Potentialtopfs sei die potentielle Energie der Einfachheit halber unendlich hoch. Damit erhält man die Schrödinger-Gleichung für ein Elektron, das sich innerhalb des Potentialtopfs befindet:

$$-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \Psi(x) = E\Psi(x) \quad (2.4)$$

Da $V(x) = 0$ im Potentialtopf, fällt der Term der potentiellen Energie in der Schrödinger-Gleichung weg. Wir erhalten eine relativ übersichtliche Differentialgleichung zweiter Ordnung, deren Lösungen gut bekannt sind. Als Lösungsfunktionen ergeben sich Sinus- und Cosinusfunktionen:

$$\Psi(x) = A \sin(kx) + B \cos(kx) \quad (2.5)$$

Die erste und zweite Ableitung ergibt sich zu

$$\Psi'(x) = k[A \cos(kx) - B \sin(kx)] \quad (2.6)$$

$$\Psi''(x) = -k^2[A \sin(kx) + B \cos(kx)] = -k^2\Psi(x) \quad (2.7)$$

Wir setzen diese 2. Ableitung in die Schrödinger-Gleichung ein und erhalten folgenden Zusammenhang:

$$E = \frac{\hbar^2 k^2}{2m} \quad (2.8)$$

Der Parameter k hat die Bedeutung einer Wellenzahl, die über $k = 2\pi/\lambda$ in Wellenlänge umgerechnet werden kann. Es sei schon jetzt darauf hingewiesen, dass dies nicht als klassische kinetische Energie interpretiert werden darf, die eine Bewegung in einer Richtung anzeigt. Auf diese Problematik wird später noch genauer eingegangen.

Obwohl wir hiermit eine allgemeine Lösung der Differentialgleichung gefunden haben, ist unser Problem erst zur Hälfte gelöst. Bisher wurden in der Lösung nämlich die realen Eigenschaften des konkreten Potentialtopfes, wie Breite und Höhe, noch nicht berücksichtigt. Diese fließen erst jetzt, in einem zweiten Schritt, über die Randbedingungen in die Lösung der Schrödinger-Gleichung mit ein. Wir werden sehen, dass erst diese Randbedingungen – wie für Eigenwertprobleme üblich – zu einer Quantisierung der möglichen Energien von Elektronen im Potentialtopf führen.

Gerade beim unendlich hohen Potentialtopf ergeben sich recht einfache Randbedingungen (deshalb haben wir ihn auch unendlich hoch gewählt). Außerhalb des Potentialtopfs darf sich kein auch noch so kleiner Anteil des Elektrons aufhalten, da das Elektron ansonsten eine unendlich hohe Energie besitzen würde. Dies wäre unphysikalisch. Wir müssen deshalb folgende Randbedingungen an die Lösung der Schrödinger-Gleichung stellen:

$$\Psi(x=0) = A \sin(k \cdot 0) + B \cos(k \cdot 0) = 0 \quad \Rightarrow \quad B = 0 \quad (2.9)$$

$$\Psi(x=L) = A \sin(kL) + 0 \cdot \cos(kL) = A \sin(kL) = 0 \quad (2.10)$$

Der Wert $\sin(kL)$ muss also null sein und damit muss das Argument unter der Sinusfunktion ein ganzzahliges Vielfaches von π sein:

$$\begin{aligned} kL &= n\pi \quad n = 1, 2, 3, \dots \\ k_n &= n \frac{\pi}{L} \quad n = 1, 2, 3, \dots \end{aligned} \quad (2.11)$$

n ist die Quantenzahl, die diejenigen Wellenzahlen k indiziert, die zu Lösungen führen, die mit den Randbedingungen verträglich sind. Die Quantisierung hat sich damit ganz automatisch aus dem Anlegen der Randbedingungen an die allgemeinen Lösungen der Schrödinger-Gleichung ergeben. Damit ist dann k quantisiert und am Ende auch die Energie E .

Hinweis

Die Quantisierung der möglichen Energiewerte ergibt sich nicht aus der Schrödinger-Gleichung direkt, sondern erst in Verbindung mit den Randbedingungen, die für die Lösungen der Schrödinger-Gleichung gefordert werden müssen. Das Auffinden passender Randbedingungen erfordert physikalisches „Fingerspitzengefühl“.

$$E_n = \frac{\hbar^2 k_n^2}{2m} = \frac{\hbar^2 \pi^2}{2mL^2} n^2 \quad (2.12)$$

Mit der Quantisierung der k_n ergeben sich dann auch die zugehörigen Eigenfunktionen

$$\Psi_n(x) = A \sin(k_n x) \quad (2.13)$$

Damit das Quadrat der Wellenfunktionen tatsächlich als Wahrscheinlichkeitsdichte interpretiert werden kann, müssen diese noch normiert werden. Normierung bedeutet, dass das Integral der Wahrscheinlichkeitsdichte über den gesamten Raum 100 % ergeben muss. Damit wird dann der bis hierher noch nicht fixierte Vorfaktor A festgelegt. Das bedeutet, dass die Aufenthaltswahrscheinlichkeit des Objektes im gesamten Raum 100 % betragen muss. Anders ausgedrückt: Das Elektron muss sich auf jeden Fall innerhalb des Potentialtopfs befinden. Auf eine quantitative Darstellung des Normierungsfaktors wird hier aber verzichtet.

Damit haben wir nun das Problem „Elektron im eindimensionalen Potentialtopf“ gelöst. Als Ergebnis erhielten wir demnach einen Satz von Energien (auch Eigenenergien genannt) sowie einen Satz von Wellenfunktionen (auch Eigenfunktionen genannt). Zur selben ganzzahligen Quantenzahl n gehört immer auch die entsprechende Wellenfunktion $\Psi_n(x)$. Die zugehörigen Quantenzahlen n laufen von eins bis unendlich, dementsprechend erhält man auch unendlich viele mögliche k_n . Die Wellenfunktionen können entweder mit der Quantenzahl n indiziert werden oder alternativ mit dem Wert k_n . Die Indizierung über den Wert k_n wird später noch bei der Beschreibung der Bandstruktur von Halbleitern wichtig.

Dieses Ergebnis reflektiert die Tatsache, dass ein Elektron in einem Potentialtopf nur ganz bestimmte, diskrete Energiewerte annehmen kann. Elektronenenergien, die gerade zwischen diesen erlaubten Energiewerten liegen, existieren als Lösung der Schrödinger-Gleichung (inklusive der Randbedingungen) *nicht*, und kommen in der Natur deshalb auch *nicht* vor. Zu jedem dieser Energieeigenwerte existiert eine wohldefinierte Wellenfunktion nach Gleichung 2.13. Da $|\Psi(x)|^2$ die Aufenthaltswahrscheinlichkeitsdichte angibt und diese über den ganzen Potentialtopf verteilt ist, verteilt sich offensichtlich auch ein Elektron, das sich in die-

sem Energiezustand befindet, über den ganzen Potentialtopf. Die Aufenthaltswahrscheinlichkeit ist mal höher, mal niedriger, verteilt sich aber wie gesagt über den ganzen Potentialtopf.

Bisher sind wir von einem rechteckigen Potentialtopf (d. h. einem Potential $V(x)$ mit rechteckiger Form) ausgegangen. Die erhaltenen Ergebnisse gelten aber zumindest qualitativ unabhängig von der Potentialform. Andere Potentialformen ergeben sich z. B., wenn man ein Elektron betrachtet, das an einen Atomkern gebunden ist. Das attraktive Potential der Coulomb-Wechselwirkung erzeugt dann eine Potentialform, bei der V proportional zu $1/r$ ist (r = Abstand Elektron-Kern). Durch die Lösung der Schrödinger-Gleichung werden sich aber wiederum ganz bestimmte Eigenenergien und Eigenwellenfunktionen ergeben, dann aber für andere Energiewerte und andere Formen der Wellenfunktionen.

Eine weitere, häufig auch in der Natur auftretende Potentialform ist V proportional zu x^2 . Diese tritt bei allen Problemen auf, die mit Schwingungen zusammenhängen. Wenn die rücktreibende Kraft F proportional zur Auslenkung x ansteigt (Voraussetzung für einen „harmonischen Oszillator“), dann ändert sich die potentielle Energie quadratisch mit dem Abstand: $V \sim x^2$. Diese Situation trifft man nicht nur bei einer Masse an einem Federpendel an, sondern z. B. auch bei einem Atom, das in einem Molekül oder Atomverbund gebunden ist. Wenn man dieses Atom aus seiner eigentlichen Ruhelage auslenkt, dann wird eine rücktreibende Kraft auftreten, und auch hier wird $V \sim x^2$ gelten. Die potentielle Energie ist also quadratisch von der Auslenkung abhängig. Auch hierfür kann die Schrödinger-Gleichung gelöst werden. Es ergeben sich natürlich wieder bestimmte quantisierte Energieniveaus, die diesmal aber eine Besonderheit haben: Sie sind alle äquidistant. Äquidistante Energieniveaus sind damit ein Kennzeichen eines parabelförmigen Verlaufs der potentiellen Energie.

Die beiden besprochenen Potentialverläufe sind zusammen mit den Eigenenergien und Wellenfunktionen in Abbildung 2.2 dargestellt.

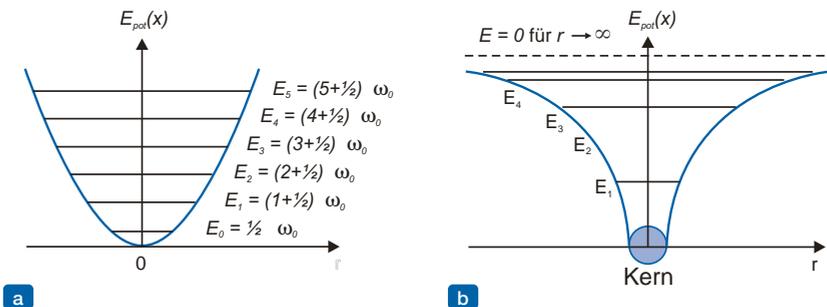


Abbildung 2.2: Potentialverläufe und Eigenenergien für ein Coulomb-Potential und ein parabolisches Potential. (a) Parabelförmiges Potential eines harmonischen Oszillators. (b) Potential eines Elektrons, das über die Coulomb-Anziehung an einen Atomkern gebunden ist (Energien nicht maßstabsgetreu eingezeichnet).

Die Eigenenergien hängen für beide Fälle wiederum von der Quantenzahl n ab. Die eigentliche Rechnung wird hier nicht nachvollzogen, diese führt aber zu folgendem Ergebnis:

$$E_n = Ry \cdot \frac{1}{n^2} = -13.6 \text{ eV} \cdot \frac{1}{n^2} \quad (\text{Elektron im Coulomb-Potential}) \quad (2.14)$$

$$E_n = \hbar \omega_0 \left(n + \frac{1}{2} \right) \quad (\text{parabelförmiger Potentialtopf}) \quad (2.15)$$

Ry ist dabei die berühmte Rydberg-Energie, die der Ionsationsenergie eines Wasserstoff-Atoms entspricht.

Hinweis

Eine charakteristische Eigenschaft tritt unabhängig von der Form des Potentials bei allen Potentialformen auf: Wenn die Breite des Potentialtopfs kleiner wird, so steigen unabhängig von der Form die Energiewerte generell an. Umgekehrt werden die Energiewerte generell abnehmen, wenn die Breite des Potentialtopfes größer wird.

Am Schluss dieses Abschnitts betrachten wir noch die Situation in einem sehr breiten Potentialtopf. Ein Elektron in einem Halbleiter kann – wie wir später sehen werden – oft als quasi-freies Elektron betrachtet werden, da es sich innerhalb des Festkörpers quasi frei bewegen kann, ohne Einschränkung durch die Atompotentiale (wir werden diese Vorstellung später noch präzisieren). Dies ist zunächst schwer einzusehen, ergibt sich aber aus einer quantenmechanischen Berechnung. Atome sind für die Elektronenwellen im Kristall quasi „durchsichtig“. Betrachten wir dann ein solches quasi-freies Elektron in einem Kristall, so kann sich dieses Elektron frei bewegen, aber eben nur innerhalb des Kristalls. Kommt das Elektron an die Oberfläche des Kristalls, so kann es aus diesem nicht austreten, es spürt eine Potentialbarriere nach draußen. Das niedrigste Energieniveau, das außerhalb des Halbleiterkristalls zur Verfügung steht, wird Vakuum-Niveau genannt. Die Situation entspricht demnach einem sehr breiten Potentialtopf, in dem Abmessungen von der Größenordnung der Kristallausdehnung in die Gleichungen eingesetzt werden können. Wir ersehen aus Gleichung (2.14), dass die Energieniveaus selbst und auch deren Abstände mit zunehmender Breite des Potentialtopfs stark abnehmen. Wir erhalten zwar selbst in diesem riesigen Potentialtopf wiederum quantisierte Energieniveaus, diese Energieniveaus liegen nun aber extrem eng beieinander und bilden damit ein „Quasi-Kontinuum“ an Energiezuständen. Die Quantisierung wird damit aufgrund der kleinen Energieabstände gar nicht mehr sichtbar, ist aber nach wie vor vorhanden. Dies kann man als klassischen Grenzfall ansehen.

2.2.2 Bindungen und Austauschwechselwirkung

Im Folgenden werden wir einen rechteckigen Potentialtopf als einfaches Modell für ein Atom betrachten. Während im Atom ein Elektron durch ein Coulomb-Potential gebunden ist, ist unser Potential rechteckig, was aber an den folgenden grundsätzlichen Aussagen nichts Wesentliches ändert. Man kann daraus trotz der Vereinfachungen einige wichtige Eigenschaften von Atombindungen und letztlich von Halbleiterbandstrukturen ableiten.

Wir fragen nun danach, was passiert, wenn sich zwei Atome nähern. Dies ist der erste Schritt hin zu einer Beschreibung eines Halbleiters, der allerdings dann nicht nur aus zwei, sondern aus vielen Atomen besteht (meist in der Größenordnung 10^{23} Atome pro cm^3). Wir betrachten zwei Potentialtöpfe, in jedem der beiden Potentialtöpfe soll jeweils ein Elektron in einem der Energieniveaus sitzen. Derartige Probleme nennt man Vielteilchenprobleme, da mehr als ein Elektron betrachtet werden muss. Im Folgenden soll gezeigt werden, wie mehrere Elektronen im quantenmechanischen Formalismus behandelt werden. Dies führt uns am Ende zum Begriff der Austauschwechselwirkung, die letztendlich für die Entstehung von Atombindungen und damit Bandstrukturen von Halbleitern verantwortlich ist.

Die zugehörige Schrödinger-Gleichung lautet für zwei Elektronen in zwei Potentialtöpfen

$$\begin{aligned} & \left\{ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_1^2} + V(x_1) \right\} \Psi_1(x_1) \Psi_2(x_2) \\ & + \left\{ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_2^2} + V(x_2) \right\} \Psi_1(x_1) \Psi_2(x_2) \\ & = (E_1 + E_2) \Psi_1(x_1) \Psi_2(x_2) \end{aligned} \quad (2.16)$$

Der Index 1 bzw. 2 der Wellenfunktionen bezieht sich auf Potentialtopf 1 bzw. 2, der Index 1 bzw. 2 an der Ortsvariablen x bezieht auf das Elektron 1 oder Elektron 2. Eine Ableitung nach x_1 berührt demnach nicht die Wellenfunktion $\Psi(x_2)$, diese wird für die Differentiation nach x_1 als konstanter Faktor angesehen.

Hierbei wurde zunächst angenommen, dass sich die beiden Elektronen nicht beeinflussen, d. h. nichts voneinander spüren. Nur dann ist die potentielle Energie $V(x)$ gleich der Summe der einzelnen Potentiale V_1 und V_2 . Wären die beiden Potentialtöpfe allerdings eng benachbart, so müsste man eine Wechselwirkung zwischen den beiden Elektronen mitberücksichtigen, die der Coulomb-Abstoßung entspricht ($V_{12} \sim 1/(x_1 - x_2)$). Dies werden wir allerdings erst in einem zweiten Schritt einbeziehen.

Wie wir aus Gleichung (2.16) ersehen, besteht die Gesamtwellenfunktion der beiden Elektronen aus dem Produkt der Wellenfunktionen der einzelnen Elektronen. Diese Produktwellenfunktion

$$\Psi(x_1, x_2) = \Psi_1(x_1) \Psi_2(x_2) \quad (2.17)$$

ist offensichtlich Lösung der gemeinsamen Schrödinger-Gleichung, da sich die Differentialoperatoren immer nur auf die eine oder andere Variable (also x_1 oder x_2) beziehen. Dies kann man durch Einsetzen in die Schrödinger-Gleichung direkt verifizieren.

An die so gefundene Lösung der gemeinsamen Schrödinger-Gleichung für zwei Potentialtöpfe muss allerdings innerhalb des quantenmechanischen Formalismus noch eine weitere Bedingung gestellt werden. Die Lösungen müssen nämlich eine Symmetriebedingung bezüglich der Vertauschung von Teilchenkoordinaten x_1 und x_2 erfüllen. Dies beruht auf dem Grundsatz der Ununterscheidbarkeit von identischen quantenmechanischen Objekten (hier zwei Elektronen). Überlappen die Wellenfunktionen zweier Elektronen aufgrund der Nähe der Potentialtöpfe (das sind später die Atome), so durchdringen die Wellenfunktionen einander und wir können das eine Elektron prinzipiell nicht mehr vom anderen Elektron unterscheiden. Wenn wir dann mit irgendeiner Messung ein Elektron nachweisen, können wir eben nicht sagen, ob dies das Elektron 1 oder das Elektron 2 war. Die beiden Objekte sind nicht unterscheidbar.

Diese Symmetrieanforderung spiegelt sich in einer Anforderung an die mathematische Beschreibung der gemeinsamen Wellenfunktion bezüglich der Vertauschung von Objektkoordinaten wider.

$$|\Psi(x_1, x_2)|^2 = |\Psi(x_2, x_1)|^2 \quad (2.18)$$

Die erste Koordinate zeigt an, welches Objekt in Potentialtopf 1 sitzt, die zweite entsprechend die Koordinate des Objekts in Potentialtopf 2. Diese Symmetrieforderung wird an das Quadrat der Wellenfunktion und nicht an die Wellenfunktion selbst gestellt, da nur diese messbar ist. Für die 2-Teilchen-Wellenfunktion selbst kommen damit zwei Möglichkeiten infrage:

$$\begin{aligned}\Psi(x_1, x_2) &= +\Psi(x_2, x_1) \\ \Psi(x_1, x_2) &= -\Psi(x_2, x_1)\end{aligned}\quad (2.19)$$

Es ergeben sich also zwei Wellenfunktionen mit völlig unterschiedlichem Symmetrieverhalten gegenüber Vertauschung der Objektkoordinaten. Der eine Typ von Wellenfunktion ist symmetrisch gegenüber der Vertauschung der Objektkoordinaten, der andere antisymmetrisch. Diese Unterscheidung hat tief greifende Konsequenzen.

Welche der beiden Typen von Wellenfunktionen für Elektronen verwendet werden muss, entscheidet letztlich das Experiment. Es stellt sich heraus, dass die Wellenfunktionen für Objekte mit einem halbzahligen Eigendrehimpuls, auch Spin genannt, mit antisymmetrischen Wellenfunktionen beschrieben werden müssen. Zu dieser Familie von Objekten gehören auch Elektronen, die bekanntermaßen einen Spin $S = 1/2$ aufweisen (genauer: die Spin-Quantenzahl beträgt $S = 1/2$). Derartige Objekte bezeichnet man als Fermionen. Auf eine exaktere Einführung und Beschreibung der Eigenschaften des Eigendrehimpulses wird an dieser Stelle verzichtet. Hier kann nur auf die einschlägige Literatur zur Einführung in die Quantenmechanik verwiesen werden (z.B. [7] und [8]).

Hinweis

Elektronen sind Fermionen, die Gesamtwellenfunktionen für mehrere Elektronen müssen antisymmetrisch gegenüber Teilchenvertauschung sein. Elektronen werden durch die Spin-Quantenzahl $S = 1/2$ beschrieben.

Hinweis

Objekte mit ganzzahligem Spin nennt man Bosonen. Hierzu gehören z. B. Helium-Atome oder auch Photonen.

Durch die bisherigen Überlegungen erhielten wir als Lösung der Schrödinger-Gleichung für zwei Potentialtöpfe eine Produktwellenfunktion

$$\Psi_1(x_1, x_2) = \Psi_1(x_1)\Psi_2(x_2) \quad (2.20)$$

Diese erfüllt allerdings im Allgemeinen nicht die notwendigen Symmetrieanforderungen bezüglich Vertauschung der Objektkoordinaten. Zur Verdeutlichung dieses Sachverhaltes nehmen wir an, dass sich das Elektron 1 in Zustand 1 mit Energie E_1 und zugehörigem Wellenvektor k_1 befindet. Entsprechendes gilt für Elektron 2. Die sich ergebende Wellenfunktion wäre dann

$$\Psi(x_1, x_2) = A \sin(k_1 x_1) \sin(k_2 x_2) \quad (2.21)$$

und die Wellenfunktion mit vertauschten Objektkoordinaten ergibt sich zu

$$\Psi(x_2, x_1) = A \sin(k_1 x_2) \sin(k_2 x_1) \quad (2.22)$$

Im Allgemeinen, d. h. für x_1 ungleich x_2 , sind diese beiden Funktion aber nicht identisch und damit weder symmetrisch noch antisymmetrisch gegen Vertauschung der Objektkoordinaten. Wir können aber aus diesen beiden Wellenfunktionen eine neue Wellenfunktion als Linearkombination bilden, die dann doch noch den Symmetrieanforderungen gerecht wird.

$$\Psi_{a,s}(x_1, x_2) = A \{ \Psi_1(x_1) \Psi_2(x_2) \pm \Psi_1(x_2) \Psi_2(x_1) \} \quad (2.23)$$

Ψ_a und Ψ_s stehen hierbei für symmetrische und antisymmetrische Wellenfunktionen. Es ist direkt durch Vertauschung der Koordinaten ersichtlich, warum diese Linearkombinationen die Symmetrieanforderungen gegen Vertauschung der Objektkoordinaten erfüllen. Weniger klar ist allerdings, warum diese Wellenfunktionen immer noch Lösungen der Schrödinger-Gleichung sind. Um dies zu beweisen, muss man die Gesamtwellenfunktion $\Psi_{a,s}$ in die Schrödinger-Gleichung einsetzen (siehe Beweis).

Beweis

Die (anti-)symmetrisierte Gesamtwellenfunktion ist Lösung der Schrödinger-Gleichung.

$$\begin{aligned} \Psi_{a,s} &= \Psi_1(x_1) \Psi_2(x_2) \pm \Psi_1(x_2) \Psi_2(x_1) \\ H_1 &= -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_1^2} + V(x_1) \quad H_2 = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_2^2} + V(x_2) \\ H_1 \Psi_1(x_1) &= E_1 \Psi_1(x_1) \quad \text{und} \quad H_2 \Psi_2(x_2) = E_2 \Psi_2(x_2) \\ H_1 \Psi_{a,s} &= \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_1^2} + V(x_1) \right) (\Psi_1(x_1) \Psi_2(x_2) \pm \Psi_1(x_2) \Psi_2(x_1)) \\ &= \Psi_2(x_2) E_1 \Psi_1(x_1) \pm \Psi_1(x_2) E_2 \Psi_2(x_1) \\ H_2 \Psi_{a,s} &= \left(-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_2^2} + V(x_2) \right) (\Psi_1(x_1) \Psi_2(x_2) \pm \Psi_1(x_2) \Psi_2(x_1)) \\ &= \Psi_1(x_1) E_2 \Psi_2(x_2) \pm \Psi_2(x_1) E_1 \Psi_1(x_2) \\ (H_1 + H_2) \Psi_{a,s} &= (E_1 + E_2) (\Psi_1(x_1) \Psi_2(x_2) \pm \Psi_1(x_2) \Psi_2(x_1)) \quad \text{q.e.d.} \end{aligned}$$

Terme mit vertauschten Koordinaten, in denen $\Psi_1(x_2) \cdot \Psi_2(x_1)$ vorkommt, fallen nach der Multiplikation mit dem durch einen Stern gekennzeichneten Konjugiert-Komplexen $(\Psi_2(x_2) \cdot \Psi_1(x_1))^*$ und Integration über den gesamten Raum grundsätzlich weg, während die Integration bei Termen ohne vertauschte Koordinaten grundsätzlich 1 ergibt. Diese Eigenschaft folgt aus der Orthonormierung der Wellenfunktionen (weitere Details siehe weiterführende Literatur).

Nun sind wir nicht mehr weit von der Erklärung der Austauschwechselwirkung entfernt, was ja erklärtes Ziel dieser kurzen Einführung war. Hierzu müssen wir jetzt „nur“ noch die Wechselwirkung zwischen den beiden Elektronen berücksichtigen. Die beiden Elektronen 'spüren' aneinander, sie wechselwirken über ihre Coulomb-Abstoßung. Die zugehörige potenzielle Energie

$$V(x_1, x_2) = \frac{1}{4\pi\epsilon_D} \frac{1}{|x_1 - x_2|} \quad (2.24)$$

muss zur Schrödinger-Gleichung addiert werden.

$$\left\{ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_1^2} + V(x_1) \right\} \Psi(x_1, x_2) + \left\{ -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x_2^2} + V(x_2) \right\} \Psi(x_1, x_2) + V(x_1, x_2) \Psi(x_1, x_2) = E \Psi(x_1, x_2) \quad (2.25)$$

Der erste Term der Gleichung ergibt – wie aus dem vorher Gesagten ersichtlich – die Energie E_1 , der zweite Term die Energie E_2 . Damit können wir schreiben:

$$E_1 \Psi(x_1, x_2) + E_2 \Psi(x_1, x_2) + V(x_1, x_2) \Psi(x_1, x_2) = E \Psi(x_1, x_2) \quad (2.26)$$

Wir multiplizieren diese Gleichung von links mit dem Konjugiert-Komplexen von $\Psi_{1,2}$ und integrieren anschließend jeden Summanden über den ganzen Raum. Dabei nutzen wir aus, dass die Wellenfunktionen normiert sein sollen, ein Integral des Quadrats einer Wellenfunktion über den ganzen Raum ergibt demnach 1. Dies gilt zumindest für die beiden ersten Terme. Im dritten Term ist der Integrand allerdings durch das Wechselwirkungspotential $V(x)$ modifiziert, der Integrand ist deshalb nicht mehr orthonormiert. Dies führt zu folgender Form der Schrödinger-Gleichung:

$$E_1 + E_2 + \int_{-\infty}^{+\infty} \Psi(x_1, x_2) \cdot V(x_1, x_2) \cdot \Psi(x_1, x_2) dx_1 dx_2 = E \quad (2.27)$$

$$E = E_1 + E_2 + V_{12}$$

Zur Erinnerung: Wir betrachten zwei Elektronen in zwei Potentialtöpfen als einfaches Modell zweier Atome. In großer Entfernung ist die Gesamtenergie direkt die Summe der Einzelenergien für Elektron 1 und Elektron 2. Kommen sich die Atome jedoch näher, so spüren die Elektronen die Coulomb-Wechselwirkung. In der Schrödinger-Gleichung ergibt sich dadurch ein zusätzlicher Term $V_{1,2}$, der auch Wechselwirkungsterm genannt wird.

Diesen Wechselwirkungsterm müssen wir nun noch genauer untersuchen. Aus einer rein klassischen Sicht erwarten wir, dass dieser Term zu einer Energieanhebung des Gesamtsystems führt, die beiden Elektronen stoßen sich schließlich aufgrund ihrer beiderseitigen negativen Ladung ab. Dies würde aber dann bedeuten, dass zwei Atome, deren Wellenfunktionen überlappen, niemals eine anziehende Wechselwirkung eingehen können und damit auch niemals eine Molekülbindung realisieren könnten. Die Realität widerspricht offensichtlich dieser klassischen Erwartung.

Wir wollen deshalb den Wechselwirkungsterm $V_{1,2}$ genauer analysieren. Hierzu setzen wir in das Integral der Gleichung (2.27) die Wellenfunktionen ein. An dieser Stelle müssen wir uns aber daran erinnern, dass die richtigen Wellenfunktionen Symmetrieanforderungen zu erfüllen haben. Wir müssen hier also die Wellenfunktion von Gleichung (2.23) einsetzen. Dies ergibt einen zunächst einigermaßen unübersichtlichen Term für die Wechselwirkungsenergie:

$$V_{12} = \int_{-\infty}^{+\infty} \Psi(x_1, x_2) \cdot V(x_1, x_2) \cdot \Psi(x_1, x_2) dx_1 dx_2$$

$$= \{ \Psi_1(x_1) \Psi_2(x_2) \pm \Psi_1(x_2) \Psi_2(x_1) \}^*$$

$$V(x_1, x_2) \{ \Psi_1(x_1) \Psi_2(x_2) \pm \Psi_1(x_2) \Psi_2(x_1) \} \quad (2.28)$$

deren Integrand nach dem Ausmultiplizieren aus den Termen (1–4) besteht:

$$\begin{aligned}
 (1) &= \{\Psi_1(x_1)\Psi_2(x_2)\}^* V(x_1, x_2) \{\Psi_1(x_1)\Psi_2(x_2)\} \\
 (2) &= \{\Psi_1(x_2)\Psi_2(x_1)\}^* V(x_1, x_2) \{\Psi_1(x_2)\Psi_2(x_1)\} \\
 (3) &= \{\Psi_1(x_1)\Psi_2(x_2)\}^* V(x_1, x_2) \{\Psi_1(x_2)\Psi_2(x_1)\} \\
 (4) &= \{\Psi_1(x_1)\Psi_2(x_1)\}^* V(x_1, x_2) \{\Psi_1(x_1)\Psi_2(x_2)\}
 \end{aligned}$$

Genauer betrachtet sind die beiden Terme (1) und (2) identisch, nur dass die Integrationsvariablen gerade vertauscht sind, was auf das Resultat nach der Integration aber keine Auswirkungen haben darf. Den Integranden können wir außerdem umsortieren:

$$(1) = (2) = \Psi_1^*(x_1)\Psi_1(x_1) \cdot V(x_1, x_2) \cdot \Psi_2^*(x_2)\Psi_2(x_2)$$

Diese ersten beiden Terme fassen wir zu einem einzigen zusammen, den wir Coulomb-Term C nennen

$$C = 2 \int_{-\infty}^{+\infty} \Psi_1^*(x_1)\Psi_1(x_1) \cdot V(x_1, x_2) \cdot \Psi_2^*(x_2)\Psi_2(x_2) dx_1 dx_2 \quad (2.29)$$

Wir erinnern uns wiederum daran, dass $\Psi^*\Psi$ das Quadrat der Wellenfunktion bedeutet und eine Wahrscheinlichkeitsdichte angibt. Diese Wahrscheinlichkeitsdichte ist aber einer Ladungsdichteverteilung proportional, da wir es hier ja mit Elektronen zu tun haben. Wir können deshalb schreiben:

$$C \sim \int_{-\infty}^{+\infty} n_1(x_1) \cdot V(x_1, x_2) \cdot n_2(x_2) dx_1 dx_2 \quad (2.30)$$

wobei n_1 und n_2 die Ladungsdichteverteilung des Elektrons 1 und 2 bedeuten sollen. Hieraus wird auch gleichzeitig die Bedeutung des Coulomb-Integrals deutlich. Es gibt an, inwieweit die beiden Wellenfunktionen von Objekt 1 und Objekt 2 überlappen, was dann in einer Coulomb-Abstoßung resultiert. Dies entspricht qualitativ auch dem, was wir durch eine klassische Überlegung erwartet hatten. Dementsprechend ist das Vorzeichen des Coulomb-Terms auch immer positiv, erhöht also die Gesamtenergie.

Genauso analysieren wir nun die Terme (3) und (4) des Integranden in Gleichung (2.28). Wiederum setzen wir die Wellenfunktionen ein, die die Symmetriebedingungen erfüllen, und sehen uns das Resultat genauer an. Auch hier unterscheiden sich Term (3) und (4) nur in der Benennung der Integrationsvariablen x_1 und x_2 . Diese können ebenso unter dem Integral vertauscht, d. h. umbenannt werden. Nach der Integration müssen beide Terme gleich sein. Wir nennen das erhaltene Integral Austauschintegral A .

$$A = \pm 2 \int_{-\infty}^{+\infty} \Psi_1^*(x_1)\Psi_2^*(x_2) \cdot V(x_1, x_2) \cdot \Psi_1(x_2)\Psi_2(x_1) dx_1 dx_2 \quad (2.31)$$

Dieses Austauschintegral kann nicht weiter vereinfacht und klassisch auch nicht interpretiert werden. Sein Auftauchen verdankt das Austauschintegral unserer strikten Anforderung an die Symmetrie der Wellenfunktionen gegenüber Vertauschung der Objektkoordinaten. Es hat damit keine klassische Entsprechung. Der Name resultiert daher, dass im hinteren Teil des Integranden die Integrationsvariablen gerade ausgetauscht sind. Eine weitere Vereinfachung wie im Falle des Coulomb-Integrals ist deshalb nicht möglich. Das Austauschintegral muss so stehen bleiben und muss im Bedarfsfall aus den Wellenfunktionen berechnet werden.

Hinweis

Die Tatsache, dass das Austauschintegral klassisch nicht erklärt werden kann, mutet zunächst eventuell merkwürdig an. An dieser Stelle sei deshalb daran erinnert, dass eine physikalische Theorie wie die Quantenmechanik sich nicht durch das Kriterium „Übersichtlichkeit“ oder „Anschaulichkeit“ als richtig erweist, sondern nur durch eine richtige Vorhersage aller in der Natur auftretenden Ereignisse. Genau dies leistet die Quantenmechanik, sofern der beschriebene Formalismus mit symmetrischen und antisymmetrischen Wellenfunktionen verwendet wird.

Trotz der etwas unübersichtlichen Rechnung ergibt sich ein recht anschauliches Endergebnis für die Gesamtenergie zweier Elektronen in getrennten Potentialtöpfen, die sich so nahe gekommen sind, dass deren Wellenfunktionen überlappen (zur Erinnerung: Dies ist unser einfaches Modell eines Moleküls aus zwei Atomen). Die Gesamtenergie beträgt

$$E = E_1 + E_2 + C \pm A \quad (2.32)$$

Sie setzt sich zusammen aus den Einzelenergien der beiden Elektronen, deren Coulomb-Abstoßung sowie der Austauschenergie. Während die Coulomb-Energie die Gesamtenergie wegen der gemeinsamen Abstoßung beider Elektronen immer erhöht und damit auf ein eventuell entstehendes Molekül destabilisierend wirkt, kann die Austauschenergie positiv oder negativ sein. Dies hängt einerseits vom Vorzeichen des Integrals selbst ab und davon, ob man symmetrische (+) oder antisymmetrische (-) Wellenfunktionen eingesetzt hat. Dabei hängen sowohl die Coulomb-Energie als auch die Austauschenergie vom Abstand der Atome bzw. vom Abstand der Potentialtöpfe ab. Ist der Abstand sehr groß, so ergibt sich gar kein Überlapp der Wellenfunktionen, im Integranden ist deshalb immer eine der Wellenfunktionen gleich null. Damit verschwindet dann sowohl das Coulomb-Integral als auch das Austauschintegral, was im Übrigen auch der Erwartung entspricht. Elektronen an weit entfernten Atomen „wissen“ schließlich nichts voneinander.

Hinweis

Die Austauschwechselwirkung entsteht, wenn die Wellenfunktionen von 2 (oder mehr) Elektronen überlappen. Für die Gesamtenergie ergibt sich dann

$$E = E_1 + E_2 + C \pm A$$

C = Coulomb-Energie

A = Austauschenergie

Durch den Überlapp der Wellenfunktionen entstehen zwei Energieniveaus, eines liegt energetisch ungünstiger, das andere energetisch günstiger als das Ausgangsniveau. Das Vorzeichen entspricht einer symmetrischen bzw. antisymmetrischen Gesamtwellenfunktion (gegen Vertauschung der Teilchenkoordinaten). Welches der beiden Vorzeichen in der Formel zu einer niedrigeren Gesamtenergie führt, hängt vom Vorzeichen des Austauschintegrals A ab. Das Austauschintegral kann positiv oder negativ sein.

2.2.3 Der Spin des Elektrons

Nun fehlt uns nur noch ein letzter Baustein zum grundlegenden Verständnis von atomaren Bindungen: die Berücksichtigung des Spins von Elektronen. Elektronen sind Elementar-,teilchen“, die neben der Eigenschaft „Ladung“ auch noch einen Eigendrehimpuls besitzen. Die quantenmechanischen Regeln im Umgang mit Eigendrehimpulsen sind kompliziert und sollen hier nicht genauer hergeleitet, sondern nur angegeben werden. Es wird ansonsten auf die Literatur zur Einführung in die Quantenmechanik verwiesen.

Der Eigendrehimpuls des Elektrons wird auch „Spin“ genannt. Mit diesem Spin ist ein magnetisches Moment der Größe μ_B verbunden. Ein magnetisches Moment der Größe μ_B wird als Bohr-Magneton bezeichnet. In diesem Sinne trägt das Elektron also sowohl die Elementarladung q ($=1,6 \cdot 10^{-19} \text{ C}$) als auch ein elementares magnetisches Moment von der Größe eines Bohr-Magnetons mit $\mu_B = 9,27 \cdot 10^{-24} \text{ J/T}$. Das magnetische Moment des Elektrons spielt für die konventionelle Elektronik keine Rolle, wird aber später bei der Beschreibung von magneto-elektronischen Halbleiterbauelementen wichtig.

Sowohl der Spin als auch das magnetische Moment sind vektorielle Größen. In der Quantenmechanik können dabei grundsätzlich nicht alle drei Komponenten dieser Vektoren gleichzeitig angegeben werden (eine Art Unschärferelation). Man gibt deshalb vom Spin des Elektrons nur seinen Betrag S und die zugehörige z-Komponente S_z an.

Sowohl der Betrag als auch die z-Komponente von Drehimpulsen gehorchen einer Richtungsquantisierung. Die Werte für ihren Betrag und den Betrag der z-Komponente werden von Quantenzahlen, die nur ganzzahlige oder halbzahlige Werte annehmen können, quantisiert.

Der Betrag des Spin-Elektrons ist

$$|\overline{S}| = S(S+1)\hbar \quad (2.33)$$

Dabei ist S die Spin-Quantenzahl, die für Elektronen $S = 1/2$ beträgt. Die z-Komponente S_z kann nach den Regeln der quantenmechanischen Richtungsquantisierung nur zwei Werte annehmen:

$$S_z = \pm \frac{1}{2} \hbar \quad (2.34)$$

Das zugehörige magnetische Moment des Elektrons beträgt ein Bohr-Magneton:

$$\mu_B = \frac{q\hbar}{2m} \quad (2.35)$$

Die weiter oben betrachteten Wellenfunktionen für Elektronen im Potentialtopf (als Modell für ein Elektron, das an ein Atom gebunden ist) waren deshalb noch nicht ganz vollständig. Diese hatten sich bisher nur auf den Ortsraum des Elektrons, also die Koordinaten x , y und z bezogen, nicht aber auf die weiteren „Freiheitsgrade“ eines Elektrons, den Spin. Die bisherigen Ortswellenfunktionen $\Psi(x,y,z)$ müssen deshalb durch eine Spin-Wellenfunktion ergänzt werden.

$$\Psi(n,s) = \Psi_n(x,y,z)\Psi_s(s) \quad (2.36)$$

Die Gesamtwellenfunktion $\Psi(n,s)$ ist gleich dem Produkt aus Ortswellenfunktion und Spin-Wellenfunktion. Die Differentialoperatoren innerhalb der Schrödinger-Gleichung aus Gleichung (2.25) wirken nur auf die Ortskoordinaten, die Spin-Wellenfunktion bleibt unberührt. Es gibt aber über die vorgestellten Beiträge zur Schrödinger-Gleichung hinaus Ergän-

zungen zur Schrödinger-Gleichung, die wir bisher noch nicht kennengelernt haben und die dann auch auf die Spin-Wellenfunktion wirken. Wir werden dieses Thema hier aber nicht weiter vertiefen.

Wir können nun ein Gesamtbild entwerfen, das das Auftreten von molekularen Bindungskräften plausibel macht. Wie oben erläutert muss die Wellenfunktion feste Symmetrieanforderungen bezüglich der Vertauschung von Teilchenkoordinaten erfüllen. Dies ist allerdings eine Anforderung an die Gesamtwellenfunktion und nicht nur an die Ortswellenfunktion. Elektronen werden durch antisymmetrische Gesamtwellenfunktionen beschrieben, da sie halbzahligen Spin haben und deshalb zur Klasse der Fermionen gehören.

Bei der Annäherung zweier Atome gibt es nun für jedes der beiden Elektronen zwei mögliche Spin-Orientierungen, die mit den zwei möglichen z -Komponenten des Spins und den zugehörigen Quantenzahlen $m_s = \pm 1/2$ einhergehen. Demnach können die beiden Spins der Elektronen entweder parallel oder antiparallel orientiert sein. Eine parallele Spin-Orientierung wird durch eine symmetrische Spin-Wellenfunktion wiedergegeben (was man auch mathematisch beweisen kann, siehe Literatur zur Einführung in die Quantenmechanik), die dann eine antisymmetrische Ortswellenfunktion nach sich ziehen muss. Nur so ergibt sich dann auch eine antisymmetrische Gesamtwellenfunktion. Umgekehrt folgt aus einer antisymmetrischen Spin-Wellenfunktion eine dann symmetrische Ortswellenfunktion. Der Symmetriecharakter der Ortswellenfunktion bestimmt aber, welches Vorzeichen der Term der Austauschwechselwirkung bekommt (siehe Gleichung (2.32)). Wir erhalten für eine symmetrische Ortswellenfunktion die Energie

$$E = E_1 + E_2 + C + A \quad (2.37)$$

und für eine antisymmetrische Ortswellenfunktion die Energie

$$E = E_1 + E_2 + C - A \quad (2.38)$$

wobei der Symmetriecharakter der Ortswellenfunktion von der Spin-Wellenfunktion „gesteuert“ wird. Es ist gerade diese Absenkung der Energie, die zur Bindung zweier Atome führt. Der Überlapp der Wellenfunktionen der beiden Elektronen ist die Ursache für das Entstehen des Austauschterms und damit die Grundlage für atomare Bindung. Neben dem energetisch günstigeren Energieniveau entsteht auch ein zweites, energetisch höher liegendes Niveau. Ersteres wird häufig „bindendes Molekülorbital“ genannt, Letzteres „antibindendes Molekülorbital“.

Welche der beiden Konfigurationen – Spins parallel oder Spins antiparallel – die energetisch günstige darstellt, entscheidet das Vorzeichen des Austauschintegrals A . Dieses ist jedoch oft (aber nicht immer) positiv, sodass in der Tat meist die Konfiguration mit antiparallelen Spins die energetisch günstigere ist, die dann zur Atombindung führt. Diese Elektronenpaare werden in der Chemie oft mit einem einfachen Strich zwischen zwei Atomen veranschaulicht und bindendes Elektronenpaar genannt.

Wir haben nun qualitativ erläutert, warum es beim Überlapp zweier Wellenfunktionen zu einer Aufspaltung der beiden eigentlich gleichen Energieniveaus in ein antibindendes und ein bindendes Niveau kommt. Ursache ist die Austauschwechselwirkung, deren Vorzeichen durch die relative Orientierung der beiden Spins gesteuert wird.

2.2.4 Die Entstehung von Bändern

Im nächsten Schritt können wir nun – ebenfalls qualitativ – die Entstehung von Energiebändern in Festkörpern erläutern. In einem Festkörper überlappen nicht nur die Wellenfunktionen von zwei Elektronen, sondern die Wellenfunktionen von sehr vielen Elektronen, gebunden an sehr viele Atome. Zur Erinnerung: In einem cm^3 Festkörper findet man etwa 10^{23} Elektronen. Damit spalten die Energieniveaus nicht in zwei, sondern eben in 10^{23} Niveaus auf. Diese Energieniveaus liegen eng benachbart und bilden die sogenannten Energiebänder.

Darüber hinaus gibt es nicht nur ein Energieniveau pro Atom, sondern mehrere, entsprechend den existierenden Atomorbitalen (1s, 2s, 2p, 3s, 3p, 3d etc.). Diese Energieniveaus überlappen mit denen der benachbarten Atome, davon die äußeren Atomorbitale mehr als die inneren. Demnach werden die äußeren Atomorbitale stärker verbreitern als die inneren. Besteht kein Überlapp der Atomorbitale, so gibt es auch keine Austauschwechselwirkung und damit auch keine Aufspaltung in Bänder. Die Situation ist in Abbildung 2.3 illustriert.

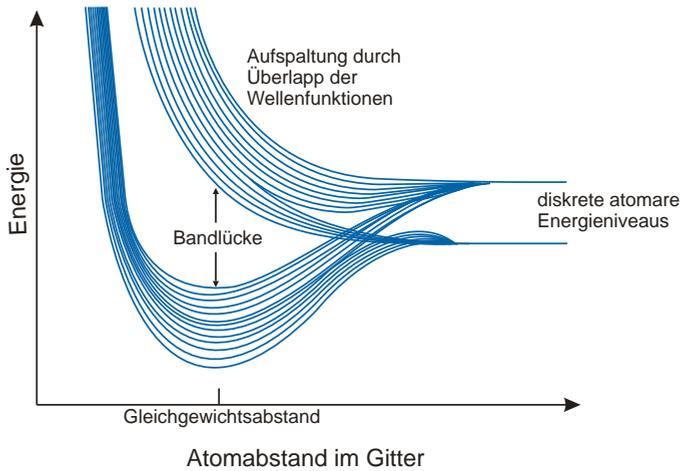


Abbildung 2.3: Entstehung von Bändern aus dem Überlapp von Elektronenwellenfunktionen in Halbleitern am Beispiel von Silizium. Die äußeren Zustände des Silizium-Atoms sind die 2s- und 2p-Zustände.

Will man die entstehenden Bänder quantitativ beschreiben, so muss man die Schrödinger-Gleichung lösen. Diese besteht aus den Energieniveaus aller Elektronen im Festkörper, die über die Coulomb-Anziehung von allen Atomen angezogen werden und über die Coulomb-Abstoßung von allen Elektronen abgestoßen werden. Zusätzlich muss man noch die Austauschwechselwirkung aller Elektronen untereinander berücksichtigen. Damit ergibt sich die Schrödinger-Gleichung zu

$$\left[\sum_i -\frac{\hbar^2}{2m} \left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \right) + \sum_{i,j} V(R_j - r_i) + \sum_{\substack{i,j \\ i \neq j}} V(r_i - r_j) \right] \Psi = E\Psi \quad (2.39)$$

R_j sind die Ortskoordinaten der Atome, r_i die Ortskoordinaten der Elektronen. Die erste Summe addiert alle kinetischen Energien aller Elektronen, die zweite Summe die potentiellen Energien der Elektronen im Coulomb-Potential der Atomkerne und die letzte Summe gibt alle Wechselwirkungen der Elektronen untereinander wieder.

Diese Gleichung lässt sich zwar einfach aufstellen, die Lösung ist allerdings sehr komplex und kann auch mit Hochleistungsrechnern nur näherungsweise durchgeführt werden. Grundsätzliches Ergebnis muss aber – wie bei jeder Lösung der Schrödinger-Gleichung – ein Satz von möglichen Energieniveaus sein, die durch mehrere Quantenzahlen indiziert werden. Die Auftragung dieser Energieniveaus nimmt man meist nicht als Funktion der Quantenzahl n vor (n müsste bis zu sehr hohen Zahlen laufen), sondern als Funktion der Quantenzahl k_n . Diese Auftragung wird Bandstruktur genannt.

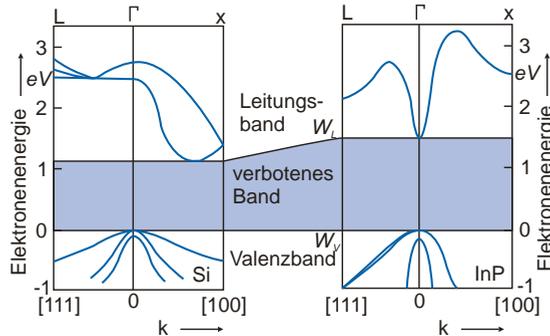


Abbildung 2.4: Bandstruktur von Silizium und Indiumphosphid (InP) in verschiedenen Kristallrichtungen [1].

Um einer verbreiteten Fehlinterpretation vorzubeugen wird noch einmal die Bedeutung von k hervorgehoben. k hat zwar die Dimension einer Wellenzahl, die hier aber nicht die Bedeutung eines Impulses entsprechend $p = \hbar \cdot k$ hat, sondern die Bedeutung einer Quantenzahl. Die zugehörigen Energieniveaus resultieren aus der Lösung der zeitunabhängigen Schrödinger-Gleichung. Ein Elektron im Zustand k_n bewegt sich also nicht mit dem Impuls $p = \hbar \cdot k$, sondern sitzt in einem zeitunabhängigen Zustand. $p = \hbar \cdot k$ wird deshalb auch als Quasi-Impuls bezeichnet. Will man Stromtransport berechnen, so muss man aus diesen zeitunabhängigen Zuständen Wellenpakete konstruieren, die sich dann mit einem bestimmten Impuls in eine bestimmte Richtung bewegen können. Die Wellenzahl k steht hier also für eine Quantenzahl, die bestimmte Energieniveaus indiziert.

Würde man alle Energieniveaus in den Bandstrukturgraphen eintragen, die sich aus den atomaren Energieniveaus der einzelnen Atome (1s, 2s, 2p, 3s, 3p, 3d etc.) ergeben, so wären dies sehr viele und der überstrichene Energiebereich wäre sehr groß. Für die Funktionalität von Halbleitern interessieren aber meist nur zwei dieser Bänder: das höchste noch mit Elektronen vollständig besetzte Band und das unterste schon vollständig unbesetzte Band. Das Auffüllen der Energieniveaus mit Elektronen, beginnend vom niedrigsten Energieniveau hin zu höheren Energien, ergibt sich aus dem Pauli-Verbot. Zwei Elektronen dürfen nicht in allen Quantenzahlen übereinstimmen. Deshalb passen in ein einzelnes Energieniveau nur zwei Elektronen, eines mit Spin-Up und eines mit Spin-Down. Will man demnach viele Elektronen, ca. 10^{23} pro cm^3 im Halbleiter, in die zur Verfügung stehenden Energieniveaus verteilen, so werden Elektronen diese Energieniveaus bis zu einer Maximalenergie besetzen.

Die beiden relevanten Bänder, das höchste besetzte und das niedrigste unbesetzte, bezeichnet man als Valenzband und Leitungsband, und normalerweise werden nur diese beiden in die Bandstruktur-Graphen eingetragen. Der interessierende Energiebereich liegt dann meist in der Größenordnung von 10 eV oder darunter.

In Abbildung 2.4 ist die Bandstruktur von Silizium und von Indiumphosphid aufgetragen. Die Wellenzahl k repräsentiert drei Quantenzahlen k_x , k_y und k_z , entsprechend der drei Raumrichtungen. Die zugehörigen Energiewerte müssten deshalb eigentlich 4-dimensional als Funktion von k_x , k_y und k_z aufgetragen werden, was natürlich nicht möglich ist. Deshalb wählt man bestimmte bevorzugte Richtungen im Raum, die einen Vektor (k_x, k_y, k_z) repräsentieren, und zeigt nur die Energiezustände als Funktion dieses k -Vektors in genau dieser Richtung. Es soll nochmals darauf hingewiesen werden, dass der Vektor k keine Bewegung in eine bestimmte Richtung anzeigt.

2.3 Bandstruktur und Ladungsträgertransport

Bisher wurde nur die zeitunabhängige Schrödinger-Gleichung analysiert. Das Phänomen **Strom** als zeitabhängiger Transport von Ladung kann demnach offensichtlich nicht durch die zeitunabhängige Schrödinger-Gleichung beschrieben werden. Eine zeitabhängige quantenmechanische Beschreibung ist möglich, aber auch kompliziert. Deshalb gehen wir nun – nachdem der Ursprung und die grundsätzliche Bedeutung einer Bandstruktur erläutert wurde – in ein klassisches Bild über. Eine quantenmechanische Betrachtung von Stromtransport würde weit über diese Einführung hinausgehen.

Das klassische Bild muss aber natürlich notwendigerweise inkonsistent bleiben. Wir betrachten Elektronen wieder als Teilchen, die einem elektrischen Feld folgen, das die Ursache für den Ladungstransport ist. Die Elektronen befinden sich aber innerhalb der Bandstruktur des Halbleiters Silizium und spüren dort den Einfluss aller auf sie wirkenden Kräfte und Einflüsse. Man sollte sich also nicht darüber wundern, dass diese vereinfachte Betrachtung regelmäßig an ihre Grenzen stößt und dass für eine Reihe von Eigenschaften von Halbleiterbauelementen doch wieder zu einer quantenmechanischen Betrachtung gewechselt werden muss.

Wir betrachten zunächst die „Bandstruktur“ eines freien Elektrons ohne den Einfluss des periodischen Potentials eines Festkörpers. Die Energie der Elektronen und deren Wellenvektor k sind gegeben durch

$$E(k) = \frac{\hbar^2 k^2}{2m} \quad (2.40)$$

wobei m die Masse (genauer: Ruhemasse) des Elektrons darstellt. Es gibt für ein freies Elektron keine Energiequantisierung, keine Bänder und demnach auch keine Bandlücken. Die $E(k)$ -Beziehung ist einfach parabelförmig und kontinuierlich.

Im Vergleich dazu ist die Bandstruktur eines realen Halbleiters wie Silizium nur parabelförmig in einem kleinen Bereich rund um $k = 0$. Beschränkt man sich auf diesen Bereich, so kann man den $E(k)$ -Zusammenhang schreiben als

$$E(k) = \frac{\hbar^2 k^2}{2m^*} \quad (2.41)$$

wobei \hbar^2 / m^* die zweifache Ableitung von $E(k)$ nach k und damit die Krümmung der $E(k)$ -Kurve angibt. Die Krümmung im parabelförmigen Bereich bestimmt also den Parameter m^* , der auch als effektive Masse bezeichnet wird. Man kann deshalb die Bewegung eines Elektrons im Halbleiter so betrachten, als ob sich das Elektron als freies Teilchen bewegt, allerdings mit der effektiven Masse m^* . Dies ist eine erhebliche Vereinfachung im Vergleich zu

einer Betrachtung der vollständigen Bandstruktur. Die komplizierten Eigenschaften der Bandstruktur spiegeln sich nun ausschließlich im Parameter m^* , der effektiven Masse, wider. Ansonsten betrachten wir das Elektron in der Bandstruktur einfach quasi als freies Teilchen. Dieser Sachverhalt wird durch den Begriff „quasi-freies Elektron“ beschrieben. Klar ist aber schon hier, dass die Beschreibung mit einer effektiven Masse nicht mehr möglich ist, wenn die Bänder nicht parabolisch sind, sich die Krümmung der Bänder also abhängig von der Energie ändert. Man spricht dann von „energieabhängigen effektiven Massen“.

Darüber hinaus sind derartige effektive Massen auch richtungsabhängig. Die Krümmung der Bänder ist abhängig von der Richtung im Kristall. Diese Richtung wird im kubischen Diamant-Gitter des Siliziums mit Kombinationen von Einheitsvektoren entlang der drei kubischen Symmetrierichtungen angegeben. So ist eine (100)-Richtung eine Richtung entlang einer Kante des Kubus, eine (111)-Richtung eine in Richtung der Raumdiagonalen des Kubus. Sind effektive Massen richtungsabhängig, so sind es auch die Ladungsträgerbeweglichkeiten, wie wir später sehen werden.

Das quasi-freie Elektron folgt nun einem von außen angelegten elektrischen Feld. Dadurch wird es beschleunigt, seine Geschwindigkeit nimmt mit konstanter Rate zu, so als ob es die Masse m^* hätte. Diese Geschwindigkeitszunahme kann nicht beliebig lange weitergehen. Nach einer gewissen Zeit wird die kinetische Energie des Elektrons so hoch, dass es seine zusätzliche Energie wieder abgeben wird, also abgebremst wird. Diesen Effekt bezeichnet man als Streuung oder Energie-Relaxation. Die zusätzliche Energie des Elektrons wird an andere Elektronen, Gitterschwingungen, Anregung von Atomen oder andere Freiheitsgrade des Festkörpers übertragen.

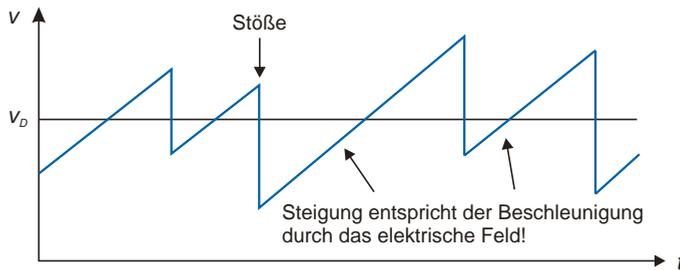


Abbildung 2.5: Beschleunigen und Abbremsen eines Elektrons in einem elektrischen Feld. Streuung führt zum Abbremsen und damit zu einer Reduktion der Energie (= Energie-Relaxation).

Wird diese statistische Bewegung des Beschleunigens und Abbremsens zeitlich gemittelt, so ergibt sich eine mittlere Geschwindigkeit für die Fortbewegung eines Elektrons in einem elektrischen Feld. Die mittlere Geschwindigkeit v_d wird als Driftgeschwindigkeit bezeichnet.

$$v_d = \mu \cdot E \quad (2.42)$$

Der Parameter μ wird Ladungsträgerbeweglichkeit genannt, v_d ist die Driftgeschwindigkeit.

Exkurs

Die Annahme eines linearen Zusammenhangs zwischen der Driftgeschwindigkeit v_d und dem elektrischen Feld E ist nicht unmittelbar klar, sie ergibt sich durch einen Vergleich mit dem Experiment und gilt eigentlich auch nur für einen begrenzten Bereich des elektrischen Feldes E . Diese lineare Abhängigkeit resultiert dann im wohlbekannten Ohm'schen Gesetz, wie wir gleich sehen werden. Insbesondere bei höheren Driftgeschwindigkeiten ergeben sich aber in jedem Material deutliche Abweichungen von diesem linearen Verhalten. Der Strom ist dann nicht mehr einfach nur proportional zur Spannung und damit zur elektrischen Feldstärke und die Beweglichkeit μ ist keine Konstante mehr, sondern hängt vom elektrischen Feld ab. Diese Situation findet man generell in Feldeffekt-Transistoren, wo in den sehr kurzen Kanälen sehr hohe elektrische Feldstärken auftreten. Auf diesen Sachverhalt werden wir später genauer eingehen.

Die Driftgeschwindigkeit kann in eine Stromdichte umgerechnet werden, sofern die Elektronendichte n bekannt ist.

$$J = q \cdot n \cdot v_d = \sigma \cdot E \quad (2.43)$$

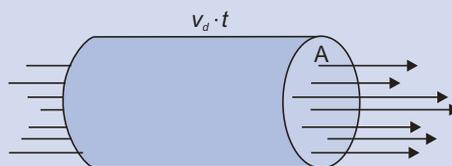
$$\sigma = q \cdot n \cdot \mu \quad (2.44)$$

σ ist die Leitfähigkeit und n die Ladungsträgerkonzentration, E die elektrische Feldstärke. Die angegebenen Formeln sind nichts anderes als das Ohm'sche Gesetz. Sie geben an, dass der Strom proportional zur elektrischen Feldstärke ist, die ja wiederum mit der von außen angelegten Spannung skaliert.

Hinweis

Zur Bestimmung des Stromes aus der Driftgeschwindigkeit v_d :

Alle der pro Zeit t durch eine Fläche A (am rechten Ende des Zylinders) hindurchtretenden Elektronen befinden sich in einem Zylinder mit dem Volumen $V = Av_d t$. Die Gesamtzahl der Elektronen beträgt $N = nV$, die Stromdichte $J = qN/At = qnv_d$.



$$Q = n \cdot \text{Vol.} = q \cdot n \cdot A \cdot v_d \cdot t$$

$$\frac{Q}{t} \cdot \frac{1}{A} = J = q \cdot n \cdot v_d$$

Abbildung 2.6: Zusammenhang von Stromdichte und Driftgeschwindigkeit.

Für die Leitfähigkeit und den Widerstand ergeben sich

$$\sigma = \frac{1}{\rho} \quad (2.45)$$

$$R = \rho \cdot \frac{l}{A} \quad (2.46)$$

ρ ist der spezifische Widerstand, l die Länge des leitfähigen Bereichs und A dessen Querschnitt. Damit können wir schreiben:

$$J \cdot A = I = \frac{1}{\rho} \cdot U \cdot \frac{A}{l} \quad (2.47)$$

Dies kann in die bekannte Form des Ohm'schen Gesetzes gebracht werden.

$$U = \rho \cdot \frac{l}{A} \cdot I = R \cdot I \quad (2.48)$$

Den Strom, der durch ein elektrisches Feld angetrieben wird, nennt man Feldstrom. Neben dem Feld kann aber auch ein Diffusionsstrom auftreten, bei dem die Bewegung von Ladungsträgern nicht durch ein Feld, sondern durch einen Konzentrationsgradienten verursacht wird. Die Ladungsträger werden grundsätzlich von einem Gebiet höherer Konzentration zu einem Gebiet niedrigerer Konzentration diffundieren. Dieser Diffusionsstrom tritt auch ohne elektrisches Feld auf. Da in Halbleiterbauelementen oft sehr große Konzentrationsgradienten auftreten, ist der Diffusionsstrom genauso wichtig wie der Feldstrom. Das Zusammenwirken von Diffusion und elektrischem Feld ist die Grundlage für die Beschreibung von Halbleiterbauelementen wie Dioden und Transistoren.

2.3.1 Fermi-Energie und die Ladungsträgerkonzentration

Betrachtet man die Besetzung der Bänder mit Elektronen, so ist das Valenzband definitionsgemäß voll besetzt. Es ist eventuell überraschend, dass Elektronen in einem voll besetzten Valenzband nicht zum Stromtransport beitragen können. Der Grund besteht darin, dass in einem vollständig gefüllten Valenzband keine freien Zustände zur Verfügung stehen, alle Zustände sind ja voll besetzt. Ein Elektron, das einem elektrischen Feld folgt, muss aber zunächst Energie aufnehmen, es muss sich also in etwas höhere Energieniveaus hinein bewegen können. Stehen solche Zustände nicht zur Verfügung, so findet eben keine Beschleunigung und damit auch kein Stromtransport statt. Die Elektronen sind quasi „eingefroren“. Uns begegnet an dieser Stelle wiederum die grundsätzlich quantenmechanische Natur der Dinge. Im Leitungsband findet man definitionsgemäß gar keine Elektronen, es ist völlig leer und damit unbesetzt. Auch im Leitungsband stehen so keine Ladungsträger für Stromtransport zur Verfügung. Ein Halbleiter wäre demnach ein Isolator. Diese Situation tritt aber nur bei sehr tiefen Temperaturen auf.

Betrachtet man einen Halbleiter dagegen bei Raumtemperatur, so ergibt sich eine Komplikation. Wir müssen nun die Besetzung der Bänder unter dem Einfluss der Temperatur betrachten. Die mittlere thermische Energie bei der Temperatur T beträgt ca. kT (k = Boltzmann-Konstante). Ein Energiebetrag von kT bei Raumtemperatur entspricht etwa 25 meV . Vergleicht man diesen Wert mit der Bandlücke, also dem minimalen Abstand zwischen Oberkante Valenzband und Unterkante Leitungsband (in Silizium $1,1 \text{ eV}$), so erkennt man, dass die mittlere thermische Energie von 25 meV bei Weitem nicht ausreicht, um Elektro-

nen vom Valenzband ins Leitungsband anzuheben. Silizium müsste also auch bei Raumtemperatur ein idealer Isolator sein – was aber nicht der Fall ist.

Ein kleiner Anteil der Elektronen hat nämlich selbst bei Raumtemperatur schon genügend thermische Energie, um aus dem Valenzband in das Leitungsband angehoben zu werden. Dies liegt daran, dass die thermische Energie kT nur einen Mittelwert darstellt, die momentane Energie von Elektronen kann zum Teil wesentlich über (oder unter) diesem Wert liegen.

Unter Berufung auf die Thermodynamik, deren Grundlagen wir hier natürlich nicht entwickeln können, stellen wir fest, dass die Besetzungswahrscheinlichkeit von Energieniveaus durch die Fermi-Dirac-Wahrscheinlichkeitsverteilung gegeben ist.

$$f_{FD}(E) = \frac{1}{\exp\left(\frac{E - E_F}{kT}\right) + 1} \quad (2.49)$$

Diese Funktion ist in Abbildung 2.7 dargestellt.

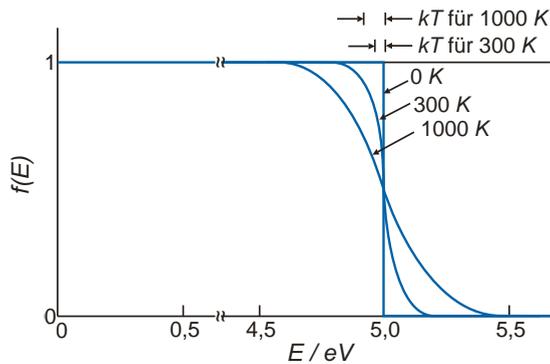


Abbildung 2.7: Fermi-Dirac-Verteilung: Besetzungswahrscheinlichkeit $f(E)$ als Funktion der Energie E , mit einer Fermi-Energie von $E_F = 5 \text{ eV}$.

Die Fermi-Dirac-Verteilung zeigt die Besetzungswahrscheinlichkeit als Funktion der Energie. Bei $T = 0 \text{ K}$ ist die Fermi-Dirac-Funktion eine Stufenfunktion. Alle Zustände unterhalb von E_F sind besetzt (Besetzungswahrscheinlichkeit $f(E) = 1$), alle darüber unbesetzt (Besetzungswahrscheinlichkeit $f(E) = 0$). Für höhere Temperaturen verschmiert die bei $T = 0 \text{ K}$ scharfe Stufe. Aus der Abbildung 2.7 ist ersichtlich, dass es dann doch einige Elektronen gibt – wenn auch nur anteilig sehr wenige –, die eine genügend hohe thermische Energie besitzen, um vom Valenzband in das Leitungsband angeregt zu werden. Erst hierdurch gelangen Elektronen ins Leitungsband, können einem elektrischen Feld folgen und damit Stromtransport übernehmen.

Hinweis

Der Parameter E_F in der Fermi-Dirac-Verteilungsfunktion wird Fermi-Energie genannt. Dieser Parameter spielt eine außerordentlich wichtige Rolle bei der späteren Beschreibung von Halbleiterbauelementen.

Existieren in einem Halbleiter ortsabhängig Unterschiede in der Lage der Fermi-Energie, so gibt es an einem Ort im Halbleiter Elektronen mit im Mittel höherer Energie, die in den Bereich mit niedriger Energie abfließen werden. Wir erkennen schon hier, dass ein Gradient in der Fermi-Energie zu einem Strom führen wird. Umgekehrt wird es keinen Gradienten der Fermi-Energie geben, wenn sich der Halbleiter im Gleichgewicht befindet, also stromlos ist. Dies ist eine der Grundüberlegungen zur späteren Beschreibung von Halbleiterbauelementen.

Die ins Leitungsband angeregten Elektronen fehlen im Valenzband, es ergeben sich im Valenzband nicht besetzte Zustände. Damit können auch einige Elektronen im Valenzband Stromtransport übernehmen, da deren Nachbarzustände unbesetzt sind und sie deshalb in einem elektrischen Feld Energie aufnehmen können. Zur Beschreibung dieser Leitfähigkeit im Valenzband hat sich das leistungsfähige Konzept des „fehlenden Elektrons“ oder „Lochs“ durchgesetzt. Der Stromtransport im Valenzband wird richtig beschrieben, indem jedem fehlenden Elektron (das als „Loch“ bezeichnet wird) eine positive Ladung $+q$ zugeschrieben wird und dann die Bewegung dieser „positiven Ladung“ als Funktion eines elektrischen Feldes betrachtet wird – ganz analog zur Bewegung von Elektronen im Leitungsband als Antwort auf ein elektrisches Feld.

Löcher sind positiv geladen und bewegen sich im Vergleich zu Elektronen bei gleicher elektrischer Feldstärke in die entgegengesetzte Richtung. Ein weiterer wichtiger Unterschied zwischen Löchern und Elektronen ist folgender: Elektronen wie auch Löcher begeben sich in der Bandstruktur grundsätzlich in Energieniveaus mit möglichst niedriger Energie. Dabei besetzen Elektronen Energieniveaus in der Bandstruktur, die in den üblichen Energiediagrammen möglichst weit „unten“ sind. Auch Löcher begeben sich in Energieniveaus, die möglichst günstig sind. Diese befinden sich für Löcher allerdings „oben“ in der Bandstruktur, da sie ja durch fehlende Elektronen entstehen, die weiter unten sitzen. Löcher benehmen sich so ähnlich wie Luftblasen in Wasser. Während das Wasser (Elektronen) nach unten fließt, begeben sich die Luftblasen (Löcher) nach oben. Es ist wichtig, diesen Unterschied zu verstehen.

Exkurs

Man könnte nun auf die Frage stoßen, warum sich nicht alle Elektronen gemeinsam im tiefsten Energieniveau aufhalten. Dies verbietet wie schon angesprochen eine Grundregel der Quantenmechanik, das Pauli-Verbot. Zwei Elektronen dürfen nie in allen Quantenzahlen übereinstimmen (= nie im gleichen Energiezustand sitzen). Dies ergibt sich ganz direkt aus der Tatsache, dass die Gesamtwellenfunktion von Elektronen antisymmetrisch gegen Teilchenvertauschung sein muss. Setzt man in einer antisymmetrischen Gesamtwellenfunktion zwei Zustände gleich, so ergibt sich identisch null, die Wellenfunktion verschwindet, eine solche Situation kommt in der Natur nicht vor. Deshalb füllen die Elektronen die Bandstruktur von unten nach oben auf, wie Wasser, das in einen Eimer gefüllt wird. Auch die Wassertropfen in einem Eimer dürfen sich nicht alle im selben „Energieniveau“ (gleicher Ort, gleiche Höhe) aufhalten.

Die durch die thermische Anregung entstehende Ladungsträgerkonzentration von Elektronen im Leitungsband und Löchern im Valenzband (beide sind gleich groß) nennt man intrinsische Ladungsträgerkonzentration. Diese hängt ab von der Bandlücke und der Temperatur. Je größer die Bandlücke, desto kleiner die intrinsische Ladungsträgerkonzentration bei Raumtemperatur (300 K).

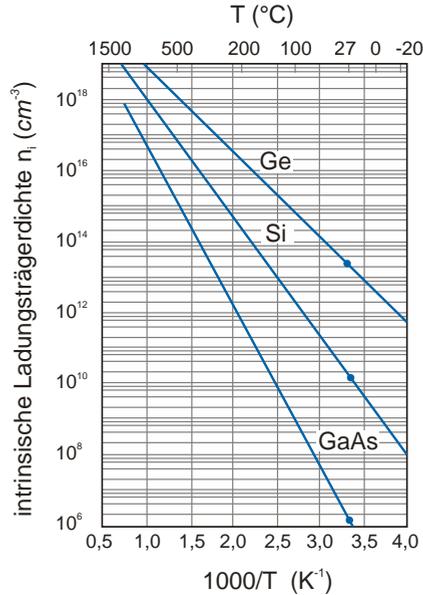


Abbildung 2.8: Intrinsische Ladungsträgerkonzentration n_i von Si, Ge und GaAs als Funktion der Temperatur [2].

Wir kommen nun zurück zur Diskussion der Bedeutung der Fermi-Energie.

Die Zahl der thermisch angeregten Elektronen im Leitungsband muss genauso groß sein wie die Zahl der Löcher im Valenzband, da ja jedes angeregte Elektron ein Loch hinterlässt. Zur Bestimmung der genauen Zahl der Elektronen im Leitungsband (bzw. analog der Zahl der Löcher im Valenzband) muss die Besetzungswahrscheinlichkeit noch mit der Zahl der besetzbaren Zustände multipliziert und danach über die Energie integriert werden. Die Zahl der besetzbaren Zustände pro Energieintervall wird als Zustandsdichte $D(E)$ bezeichnet. Sie hängt für quasi-freie Elektronen bzw. Löcher entsprechend Gleichung (2.50) von der Energie ab. Um die Energie E vom elektrischen Feld E unterscheiden zu können, wird in derartigen Rechnungen die Energie meist mit W beschrieben. Diese Nomenklatur wollen wir nun übernehmen.

$$D_L(W) = \frac{1}{(2\pi)^2} \cdot \frac{1}{\hbar^3} \cdot (2m_L)^{3/2} \sqrt{W - W_L} \quad (2.50)$$

$$D_V(W) = \frac{1}{(2\pi)^2} \cdot \frac{1}{\hbar^3} \cdot (2m_V)^{3/2} \sqrt{W_V - W}$$

Die Situation ist in Abbildung 2.9 veranschaulicht.

Die Herleitung der Zustandsdichte quasi-freier Elektronen soll hier nicht vorgestellt werden, sie kann entsprechenden Büchern zur Halbleiterphysik entnommen werden. Die Konzentration an Elektronen und Löchern muss aus dem Produkt aus Zustandsdichte und Besetzungswahrscheinlichkeit mit nachfolgender Integration berechnet werden:

$$n = 2 \int_{W_L}^{\infty} f(W) \cdot D(W) dW \quad (2.51)$$

Der Faktor 2 ergibt sich aus der Tatsache, dass sich zwei Elektronen mit entgegengesetztem Spin in einem Energiezustand aufhalten können. Jeder Zustand zählt deshalb doppelt. Eine entsprechende Gleichung gilt auch für Löcher, wobei allerdings die Wahrscheinlichkeit berücksichtigt werden muss, dass die Zustände unbesetzt sind:

$$p = 2 \int_{-\infty}^{W_V} [1 - f(W)] \cdot D(W) dW \quad (2.52)$$

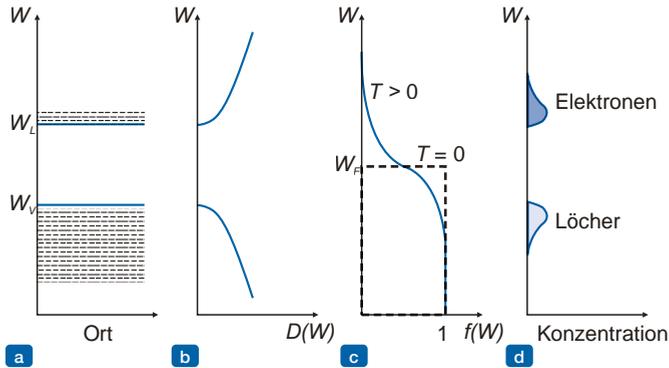


Abbildung 2.9: (a) Leitungs- und Valenzband. (b) Zustandsdichte $D(E) = D(W)$. (c) Fermi-Dirac-Verteilung $f(E) = f(W)$ und (d) Besetzung des Leitungs- und Valenzbandes.

Hinweis

Die Fermi-Energie eines undotierten Halbleiters liegt wegen der Forderung $n_i = p_i$ etwa in der Mitte der Bandlücke. Falls sich die Zustandsdichten von Elektronen und Löchern unterscheiden, so ist die Fermi-Energie leicht in Richtung niedrigere Zustandsdichte hin verschoben. Die Abweichung von der Mitte der Bandlücke ist allerdings normalerweise nicht sehr groß.

Wird die Fermi-Energie nach oben in Richtung Leitungsbandkante verschoben, so bedeutet dies eine Erhöhung der Zahl der Elektronen, während sich die Zahl der Löcher verringert. Umgekehrtes gilt für den Fall, dass die Fermi-Energie nach unten in Richtung Valenzbandkante verschoben wird.

Liegt die Fermi-Energie weit genug von beiden Bandkanten entfernt, so kann das Ergebnis des Integrals aus Gleichung (2.51) und (2.52) in vereinfachter Form angegeben werden. Für diesen Fall (d. h. für nicht zu hohe Elektronen- und Löcherkonzentrationen) ergibt sich für die Elektronen- und Löcherkonzentration:

$$\begin{aligned} n &= N_L \exp\left\{\frac{W_F - W_L}{kT}\right\} \\ p &= N_V \exp\left\{\frac{W_V - W_F}{kT}\right\} \end{aligned} \quad (2.53)$$

N_L und N_V enthalten die Konstanten aus dem Integral der Gleichung (2.51) und (2.52).

2.3.2 Akzeptoren und Donatoren

Die Elektronen- oder Löcherkonzentration kann einseitig durch die Zugabe von Dotieratomen zum Halbleiter erhöht werden. Derartige Dotieratome besitzen ein Elektron mehr als Silizium (fünf Valenzelektronen, Donatoren) oder ein Elektron weniger als Silizium (drei Valenzelektronen, Akzeptoren), die diese als Valenzelektronen zur Verfügung stellen können. Wird die Elektronenkonzentration durch die Zugabe von Donatoren erhöht, so wandert die Fermi-Energie nach oben in Richtung Unterkante des Leitungsbandes. In gleichem Maße sinkt dann aber auch die Löcherkonzentration. Dabei bleibt das Produkt aus Elektronenkonzentration und Löcherkonzentration konstant. Der zugrunde liegende physikalische Mechanismus ist die zunehmende Rekombination von Elektron-Loch-Paaren, die durch eine erhöhte Elektronenkonzentration zu einer niedrigeren Löcherkonzentration führt.

$$n \cdot p = \text{const} = n_i^2 \quad (2.54)$$

Dieses auf den ersten Blick erstaunliche Gesetz entspringt den Eigenschaften der Fermi-Dirac-Verteilungsfunktion, die symmetrisch zur Fermi-Energie verläuft.

Donatoren und Akzeptoren geben als neutrale Atome ein Elektron oder Loch an die Bandstruktur des Halbleiters ab. Dadurch werden die Donatoren positiv und die Akzeptoren negativ geladen. Die resultierende Coulomb-Anziehung zwischen Elektron und Donatorrumpf sowie Loch und Akzeptorrumpf führt zu einem gebundenen Zustand. Elektronen werden demnach an den Donator und Löcher an den Akzeptor gebunden. Die Bindungsenergien in Silizium sind mit wenigen meV allerdings so niedrig, dass Elektronen und Löcher bei Raumtemperatur durch ihre thermische Energie praktisch vollständig von Donatoren bzw. Akzeptoren entfernt werden, diese werden vollständig ionisiert. Die eigentlich an die Donatoren bzw. Akzeptoren gebundenen Elektronen und Löcher befinden sich dann im Leitungs- und Valenzband und sind deshalb frei beweglich.

Bisher konnte die Verteilung von Ladungsträgern auf die Bänder mit nur einer einzigen Fermi-Energie beschrieben werden. Hierzu muss der Halbleiter allerdings im thermodynamischen Gleichgewicht sein. Dies ist Grundvoraussetzung für die Anwendung der Fermi-Dirac-Verteilungsfunktion. Halbleiter, durch die ein Strom fließt oder die Licht absorbieren, befinden sich aber nicht mehr im thermodynamischen Gleichgewicht. Bei einer Solarzelle wird z. B. durch die Absorption von Licht ja sowohl die Elektronenkonzentration als auch die Löcherkonzentration erhöht. Diese Situation kann nicht mehr durch eine einzige Fermi-Energie in einer einzigen Verteilungsfunktion beschrieben werden. Es sind dann zwei Verteilungsfunktionen notwendig, eine für Elektronen und die andere für Löcher, die beide separat die jeweiligen Trägerkonzentrationen angeben. In den beiden Verteilungsfunktionen für Elektronen und Löchern tauchen dann auch zwei unterschiedliche Fermi-Energien auf. Diese Fermi-Energien werden als Quasi-Fermi-Niveaus bezeichnet (im Englischen oft auch IMREFs genannt).

Hinweis

Befindet sich ein Halbleiter nicht im thermodynamischen Gleichgewicht, so kann die Elektronen- und Löcherkonzentration nicht mehr durch eine einzige Fermi-Energie W_F beschrieben werden. Man benötigt zwei Fermi-Energien, eine für Löcher, die andere für Elektronen. Diese werden Quasi-Fermi-Niveaus W_{Fn} und W_{Fp} genannt.

2.3.3 Driftstrom und Diffusionsstrom

Das elektrische Feld als eine der treibenden Kräfte für elektrischen Strom wurde schon diskutiert. Es ergibt sich bei konstantem elektrischen Feld eine konstante Driftgeschwindigkeit, die in Stromdichte umgerechnet werden kann, sofern die Ladungsträgerdichte bekannt ist. Diese Drift ist der statistischen, thermischen Bewegung der Ladungsträger überlagert und führt zu einem Nettostrom, dem Driftstrom

$$J = q \cdot n \cdot v_d = \sigma \cdot E \quad (2.55)$$

Eine zweite treibende „Kraft“, die ebenso bei der Beschreibung von Halbleiterbauelementen unbedingt berücksichtigt werden muss, ist ein Gradient der Ladungsträgerkonzentration, der zur Diffusion führt. Teilchenkonzentrationen werden ausgeglichen, und dies führt zu einem Teilchentransport und damit Stromfluss

$$J = -q \cdot D \cdot \frac{\partial n}{\partial x} \quad (2.56)$$

q ist die Ladung der Teilchen, n die Teilchendichte, $grad(n)$ der Gradient (im Ort) der Teilchendichte, D der Diffusionskoeffizient und J die Stromdichte. Dieser Diffusionsstrom fließt in Richtung eines negativen Gradienten, d. h. in Richtung kleinerer Teilchendichten, was die Ursache für das Vorzeichen in Gleichung (2.56) ist.

Soll der Strom durch ein Halbleiterbauelement berechnet werden, so müssen der Diffusionsstrom und der Feldstrom addiert werden, und dies separat für Elektronen und Löcher.

$$\begin{aligned} J_{\text{gesamt}} &= J_n + J_p \\ J_n &= q \cdot n \cdot \mu_n \cdot E + q \cdot D_n \cdot \frac{\partial n}{\partial x} \\ J_p &= q \cdot n \cdot \mu_p \cdot E - q \cdot D_p \cdot \frac{\partial p}{\partial x} \end{aligned} \quad (2.57)$$

Die letzte wichtige Gleichung zur Beschreibung von Halbleiterbauelementen ist die Kontinuitätsgleichung. Diese besagt, dass eine Änderung der Ladungsträgerkonzentration an einem bestimmten Ort verursacht werden kann durch 1) Rekombination von Ladungsträgern, 2) Generation von Ladungsträgern und 3) einem Gradienten in der Stromdichte, sodass z. B. mehr Strom zugeführt als abgeführt wird (oder umgekehrt).

$$\begin{aligned} \frac{dn}{dt} &= G - R + \frac{1}{q \cdot A} \cdot \frac{\partial I_n}{\partial x} \\ \frac{dp}{dt} &= G - R - \frac{1}{q \cdot A} \cdot \frac{\partial I_p}{\partial x} \end{aligned} \quad (2.58)$$

Rekombination und Generation beschreiben Prozesse, bei denen Ladungsträger vernichtet bzw. erzeugt werden. Eine Vernichtung von Ladungsträgern findet statt, wenn Elektronen vom Leitungsband in freie Zustände des Valenzbandes zurückspringen (= rekombinieren). Dadurch gehen ein Elektron und ein Loch gleichzeitig verloren. Eine Generation von Ladungsträgern findet z. B. statt, wenn im Halbleiter ein Photon absorbiert wird und gleichzeitig ein Elektron vom Valenzband in das Leitungsband angehoben wird. Dadurch entstehen ein Elektron und ein Loch gleichzeitig.

Hinweis

Die Kontinuitätsgleichung ist eine direkte Konsequenz der Ladungserhaltung. Ladung kann insgesamt weder erzeugt noch vernichtet werden.

Die diskutierten Grundgleichungen werden im folgenden Kapitel 3 verwendet, um die Arbeitsweise von Halbleiterbauelementen quantitativ zu beschreiben.

2.4 Eigenschaften von Silizium

Integrierte Schaltungen werden heutzutage aus Silizium hergestellt. Silizium zeichnet sich durch seine hervorragende Bearbeitbarkeit aus. Mögliche Prozessierungsschritte wie z. B. das nasschemische Ätzen, Plasmaätzen, selektives Oxidieren oder Dotierung sind über viele Jahrzehnte entwickelt worden. Die Halbleiterprozesstechnik kann hier auf einen großen Erfahrungsschatz zurückgreifen. Andere Halbleiter wie z. B. GaAs wären für schnelle elektronische Schaltungen aufgrund ihrer höheren Ladungsträgerbeweglichkeit eigentlich besser geeignet. Die Prozessierung von Verbindungshalbleitern ist allerdings deutlich schwieriger als die von Silizium. Ein weiterer Vorzug von Silizium als Halbleitermaterial liegt in den Eigenschaften seines Oxids, dem SiO_2 . Dieses Oxid ist ein guter Isolator und kann direkt in MOSFET-Transistoren als Gate-Oxid verwendet werden. Die Ladungsträger im Kanal direkt unter dem Gate-Oxid haben zwar eine etwas niedrigere Beweglichkeit als im reinen Silizium, die Reduktion der Beweglichkeit durch den Einfluss des Isolators ist aber wesentlich geringer als bei Verbindungshalbleitern.

Aus Abbildung 2.4 geht hervor, dass Silizium ein indirekter Halbleiter ist. Elektronen im Leitungsband befinden sich in einem energetischen Grundzustand bei endlichem k -Wert, während Löcher im Valenzband einen energetischen Grundzustand bei $k = 0$ einnehmen. Die Konsequenz ist, dass Elektronen und Löcher nur sehr ineffizient miteinander rekombinieren, die Rekombinationszeiten sind im Vergleich zu direkten Halbleitern wie z. B. GaAs und GaN sehr lang. Silizium ist deshalb für optoelektronische Anwendungen wie LEDs und Laserdioden nicht geeignet. Komponenten der Optoelektronik beruhen deshalb auf Verbindungshalbleitern wie GaAs oder GaN.

In Tabelle 2.1 werden einige wichtige Eigenschaften von Silizium und SiO_2 aufgeführt.

Tabelle 2.1

Eigenschaften von Silizium und SiO_2 nach [4].

Eigenschaft	Silizium	SiO_2
Atomgewicht	28,09	60,08
Atome pro Volumen (cm^{-3})	$5,0 \cdot 10^{22}$	$2,3 \cdot 10^{22}$
Dichte (g/cm^{-3})	2,33	2,27
Kristallstruktur	Diamant	Amorph

Eigenschaften von Silizium und SiO_2 nach [4]. (Forts.)

Gitterkonstante (nm)	0,543	–
Bandlücke (eV)	1,12 (indirekt)	8-9
Relative Dielektrizitätskonstante ϵ_r	11,7	3,9
Intrinsische Ladungsträgerdichte (cm^{-3})	$1,4 \cdot 10^{10}$	–
Ladungsträgerbeweglichkeit μ (cm^2/Vs)	1430 (Elektronen) 470 (Löcher)	–
Effektive Zustandsdichte (cm^{-3})	$3,2 \cdot 10^{19}$ (Leitungsband) $1,8 \cdot 10^{19}$ (Valenzband)	–
Durchbruchfeldstärke ($V/\mu m$)	30	$> 10^3$
Schmelzpunkt ($^\circ C$)	1415	1600-1700
Thermische Leitfähigkeit (W/cmK)	1,5	0,014
Spezifische Wärme (J/gK)	0,7	1,0
Thermischer Ausdehnungskoeffizient (K^{-1})	$2,5 \cdot 10^{-6}$	$0,5 \cdot 10^{-6}$
Effektive Masse für das Minimum in (100)-Richtung (m^*/m_0)	0,92 (longitudinal) 0,19 (transversal)	

Die Position der Fermi-Energie W_F ist abhängig von der Dotierung und der Temperatur. Für niedrige Dotierung und hohe Temperaturen wird die Fermi-Energie in der Nähe der Mitte der Bandlücke liegen. Für hohe Dotierung wird die Fermi-Energie näher am Leitungsband (für n-Dotierung) oder näher am Valenzband (für p-Dotierung) liegen. Die Verhältnisse sind in Abbildung 2.10 gezeigt.

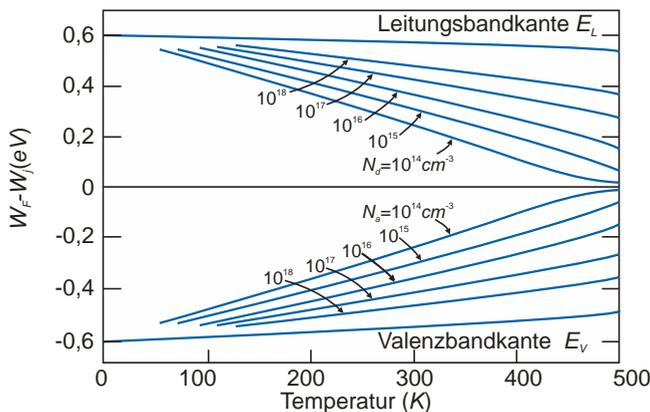


Abbildung 2.10: Die Position der Fermi-Energie W_F als Funktion der Dotierung und der absoluten Temperatur, nach [5].

Die Beweglichkeit der Ladungsträger in Silizium hängt von deren Streurrate ab. Bei Raumtemperatur und nicht zu hoher Dotierung ist die Streurrate dominiert durch Streuung an Gitterschwingungen, Phononen genannt. Bis zu einer Dotierung von ca. $10^{16} cm^{-3}$ ist die Beweglich-

keit praktisch unabhängig von der Ladungsträgerkonzentration. Bei höheren Dotierungen wird die Streuung an ionisierten Störstellen (Donatoren und Akzeptoren) immer wichtiger, die Beweglichkeit nimmt deshalb ab. Da Beweglichkeit und Diffusionskoeffizient korreliert sind, nimmt auch dieser dann ab. Die Situation ist in Abbildung 2.11 gezeigt.

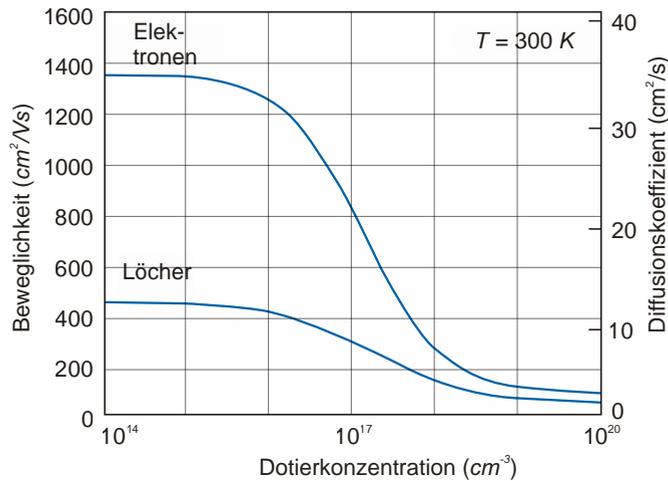


Abbildung 2.11: Beweglichkeit der Elektronen und Löcher als Funktion der Dotierung und bei einer Temperatur von 300 K (= Raumtemperatur) nach [4].

Beweglichkeit und Ladungsträgerkonzentration bestimmen den Widerstand. Es dominiert dabei die Abhängigkeit des spezifischen Widerstandes von der Ladungsträgerkonzentration, da sich diese im Gegensatz zur Beweglichkeit um viele Größenordnungen ändert. Die spezifischen Widerstände für Silizium bei Raumtemperatur sind in Abbildung 2.12 aufgetragen. Die intrinsische Ladungsträgerkonzentration bei Raumtemperatur beträgt ca. 10^{10} cm^{-3} , derart niedrige Dotierungskonzentrationen sind deshalb nicht mehr sinnvoll und technologisch auch nicht darstellbar.

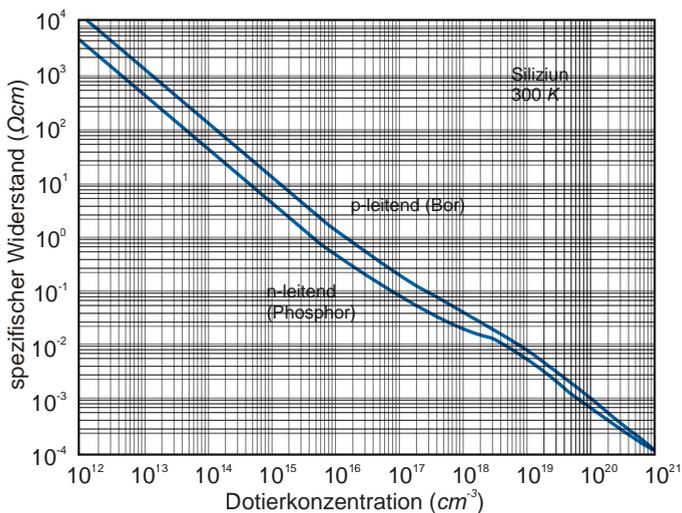


Abbildung 2.12: Spezifischer Widerstand von Silizium als Funktion der Dotierung bei Raumtemperatur, nach [6].

Zusammenfassung

In diesem Kapitel erfolgte mithilfe der Quantenmechanik eine Einführung in das Gebiet der Halbleiterphysik. Neben der Vorstellung des Bändermodells wurden der Ladungsträgertransport in Halbleitern, die Grundlagen der Dotierung und die Eigenschaften des Halbleitermaterials Silizium beschrieben.

Literatur

- [1] J. R. Chelikowsky, M. L. Cohen, *Phys. Rev. B* 14, 556, 1976.
- [2] C. D. Thurmond: *The Standard Thermodynamic Function of the Formation of Electrons and Holes in Ge, Si, GaAs and GaP*; *J. Electrochem. Society* 122, S. 1133(1975).
- [3] A. Schlachetzki: *Halbleiter-Elektronik*; Teubner Studienbücher, Stuttgart, 1990.
- [4] Y. Taur, T. H. Ning: *Fundamentals of Modern VLSI Design*; Cambridge University Press, New York, 1998.
- [5] A. S. Grove: *Physics and Technology of Semiconductor Devices*; Wiley, New York, 1967.
- [6] S. M. Sze: *Physics of Semiconductor Devices*; Wiley, 1981, S. 32.
- [7] M. Alonso, E. J. Finn.: *Quantenphysik und statistische Physik*; Oldenbourg Wissenschaftsverlag GmbH, München 2005.
- [8] V. F. Müller: *Quantenmechanik*; Oldenbourg Wissenschaftsverlag GmbH, München 2000.

Integrierte Bauelemente

3.1	Der PN-Übergang	70
3.2	Der Bipolar-Transistor	90
3.3	Der Junction-FET (JFET)	107
3.4	Der MOSFET	117
3.5	Passive Bauelemente	140
3.6	Kurzkanaleffekte und Skalierung	144
3.7	Geschwindigkeit eines MOSETs: Optimierung der Taktfrequenzen	147
3.8	MOS-Speicher	149

Einleitung

» Die herausragende Bedeutung der Halbleiterelektronik beruht auf der Integration von sehr vielen elektronischen Bauelementen auf einem einzigen Halbleiterchip. Im folgenden Kapitel wird der prinzipielle Aufbau und die Funktionsweise derartiger integrierter Bauelemente (wie z. B. Dioden, Bipolar-Transistoren und Feldeffekt-Transistoren) vorgestellt. Die Arbeitsweise von Transistoren kann an charakteristischen Kennlinien abgelesen werden, die Eingangs- und Ausgangsspannungen und -ströme miteinander in Verbindung setzen. Diese Kennlinien sind, anders als bei einfachen ohmschen Widerständen, immer nichtlinear.

Oft werden Bauelemente in einem festen Arbeitspunkt (d. h. fast konstante Ströme und Spannungen) betrieben, und die Ströme und Spannungen bewegen sich im Betrieb nur wenig von diesem Arbeitspunkt weg. In diesem Fall greift man häufig auf eine lineare Näherung zurück, die die Kennlinien in der Nähe des Arbeitspunktes gut wiedergibt. Die zugehörigen Größen wie Verstärkung, Eingangs- und Ausgangswiderstand etc. werden differentielle Größen oder auch Kleinsignalgrößen genannt. Ausgehend von der Beschreibung einer Diode werden die Kennlinien von Bipolar-Transistoren und Feldeffekt-Transistoren hergeleitet. Transistoren bilden die Basis jeder modernen elektronischen Schaltung. «

LERNZIELE

- Anwendung der Grundgleichungen (Drift-Diffusions-Gleichung und Kontinuitätsgleichung) zur Berechnung der Kennlinien von Halbleiterbauelementen
- Aufbau und Funktionsweise einer Diode
- Aufbau und Funktionsweise eines Bipolar-Transistors
- Aufbau und Funktionsweise eines MOSFETs
- Kleinsignal- und Großsignalparameter
- Grundlegende Transistorschaltungen: Emitter-, Kollektor- und Basis-Schaltung
- Passive Bauelemente: integrierte Kondensatoren und Widerstände

3.1 Der PN-Übergang

Hinweis

PN-Übergänge kommen in allen integrierten Bauelementen und Schaltungen vor und sind auch ein Grundelement bei Bipolar-Transistoren und MOSFETs. Ein Grundverständnis der Arbeitsweise eines PN-Übergangs ist deshalb wichtiger Ausgangspunkt für die dann anschließende Diskussion von Transistoren.

In integrierten Schaltungen kommen im Allgemeinen sowohl n-dotierte als auch p-dotierte Bereiche vor. Grenzt in einer integrierten Schaltung ein n-dotiertes an ein p-dotiertes Gebiet, so hat der entstehende PN-Übergang die gleichrichtenden Eigenschaften einer Diode. Ein derartiger PN-Übergang kann entweder als passives Bauelement (Diode) in Schaltungen ein-

gesetzt werden oder aber er dient als elektrische Isolation zwischen dem n- und dem p-Gebiet. Dann allerdings muss dafür gesorgt werden, dass die im Betrieb auftretenden Potentialdifferenzen immer in Sperrrichtung der Diode gepolt sind, sodass nur ein sehr kleiner und meist vernachlässigbarer Sperrstrom über einen solchen PN-Übergang fließt. Das Verständnis eines PN-Übergangs ist deshalb sehr wichtig für die Beschreibung der Arbeitsweise von Bauelementen in integrierten Schaltungen. Insbesondere werden die Erkenntnisse über Bandverbiegungen, Raumladungszonen und die damit verbundenen Kapazitäten für die Beschreibung von Metall-Oxid-Halbleiter-Feldeffekt-Transistoren (MOSFETs) benötigt.

Das Verständnis der Arbeitsweise einer einfachen Diode ist Voraussetzung und Ausgangspunkt für die Beschreibung von Standardbauelementen der Halbleitermikroelektronik, wie z. B. dem Bipolar-Transistor oder dem MOSFET. Es ist deshalb wichtig, eine gefestigte Vorstellung der physikalischen Abläufe in einer Diode zu haben. In diesem Kapitel werden deshalb die Eigenschaften einer PN-Diode detailliert beschrieben und weitestgehend hergeleitet.

3.1.1 Aufbau und Funktionsweise

Der Strom durch einen PN-Übergang ist nicht linear von der Spannung abhängig, obwohl auf beiden Seiten des PN-Übergangs ja leitfähige Halbleiterbereiche zur Verfügung stehen würden. An der Grenzfläche zwischen p- und n-dotiertem Bereich, dem metallurgischen PN-Übergang, entsteht eine Verarmungszone ohne freie Ladungsträger, die einem Stromtransport entgegensteht. Die Ausdehnung der Verarmungszone ist abhängig von der am PN-Übergang angelegten Potentialdifferenz und damit von der von außen angelegten Spannung. Wie später hergeleitet wird, führt dies dazu, dass der Strom durch die Diode exponentiell von der Spannung über dem PN-Übergang abhängt. Die $I(U)$ -Kennlinie ist gegeben durch

$$\frac{I(U)}{A} = J(U) = J_s \cdot \exp\left\{\frac{q \cdot U}{k \cdot T} - 1\right\} \quad \text{mit} \quad J_s = \frac{q \cdot D_n}{L_n} \cdot n_0 + \frac{q \cdot D_p}{L_p} \cdot p_0 \quad (3.1)$$

J_s ist die Sättigungsstromdichte oder auch Sperrstromdichte, D_n und D_p die jeweiligen Diffusionskonstanten für Elektronen (n) und Löcher (p), und n_0 und p_0 sind die jeweiligen Minoritätsladungsträgerkonzentrationen im p-Bereich (n_0) und im n-Bereich (p_0). L_n und L_p sind die Diffusionslängen für Elektronen und Löcher. Auf die Herleitung dieser Gleichung und die Bedeutung dieser Größen wird später genauer eingegangen.

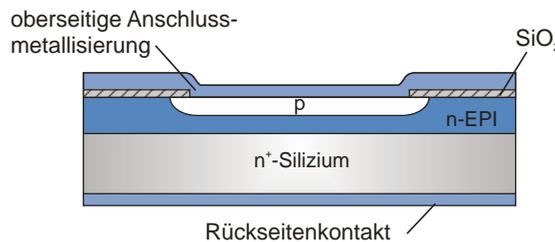


Abbildung 3.1: PN-Übergang in planarer Silizium-Technologie.

Gleichung (3.1) ist die berühmte Shockley-Gleichung. Wie an der Shockley-Gleichung abzulesen ist, überwiegt bei positiven Spannungen U der Exponentialterm in der Klammer, der Strom wird exponentiell mit der Spannung anwachsen. Diese Polarität wird als Durchlassrichtung bezeichnet. Ist die Spannung negativ, so wird der Exponentialterm sehr klein und der Strom ist gleich dem Sättigungsstrom. Diese Polarität wird als Sperrrichtung bezeichnet.

Hinweis

Positives Potential am p-dotierten Bereich und negatives Potential am n-dotierten Bereich bedeutet Durchlassrichtung. Bei zu hohen Spannungen in Sperrrichtung bricht die Diode durch.

Ein integrierter Aufbau einer PN-Diode ist in Abbildung 3.1 gezeigt, eine typische Strom-Spannungs-Kennlinie in Abbildung 3.2.

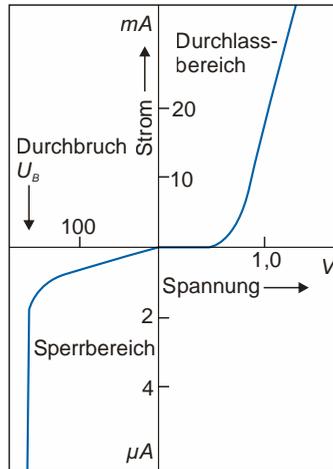


Abbildung 3.2: Typische Kennlinie einer PN-Diode, mit unterschiedlichen Maßstäben im Durchlass- und Sperrbereich. Bei hohen Spannungen in Sperrrichtung findet Lawinendurchbruch statt, nach [18].

Wird die exponentielle Abhängigkeit des Stroms von der Spannung in einem linearen Diagramm aufgetragen (rechter Bereich von Abbildung 3.2), so steigt der Strom ab einer bestimmten Spannung stark, nämlich exponentiell, an. Diese Spannung wird oft als Schwellspannung oder Knickspannung bezeichnet. Der offenbar relativ scharfe Knick in der Kurve hängt allerdings von der Skalierung der Stromachse ab. Streng genommen ist die Angabe einer Schwellenspannung physikalisch unsinnig, da der Anstieg des Stroms ja exponentiell erfolgt. Es existiert eben keine Schwelle, ab der erst ein Strom fließt. Stromfluss existiert schon ab dem ersten Millivolt. Müssen in Schaltungen aber typische Stromwerte erreicht werden, z. B. 10 mA, so kann die Kenntnis des dazugehörigen Spannungsabfalls an der Diode für die Grobplanung einer Schaltung durchaus sinnvoll sein. Auch wenn der Strom durch eine Diode in einer Schaltung um ein oder zwei Größenordnungen variiert, so ist wegen der exponentiellen Spannungsabhängigkeit die Variation der Spannung wesentlich kleiner. Man kann deshalb unter Vernachlässigung der realen Verhältnisse für eine erste, grobe Abschätzung annehmen, dass an einer Diode immer gerade die Schwellenspannung von ca. 0,7 V für Silizium-Dioden abfällt, sofern die Ströme nicht all zu sehr vom „normalen Bereich“ abweichen.

Um ein qualitatives Verständnis der Arbeitsweise einer Diode zu erhalten, muss zunächst die Situation im stromlosen Fall, also ohne äußere Spannung, betrachtet werden.

Bringen wir nun zwei Gebiete mit n- und p-Dotierung in Kontakt. Beide Zonen sind zunächst elektrisch neutral, allerdings sind die beiden Fermi-Energien in der p-Zone und der n-Zone nicht identisch. Während die Fermi-Energie W_F im n-Bereich in der Nähe des Leitungsbandes liegt, befindet sich W_F im p-Bereich in der Nähe des Valenzbandes. Deshalb werden Elektronen aus dem n-Bereich in den p-Bereich abfließen, da sie dort energetisch günstigere Energieniveaus vorfinden, dort allerdings mit Löchern rekombinieren. Für Löcher im p-Bereich gilt das ebenso. Sie werden in den n-Bereich abfließen und dort mit Elektronen rekombinieren. Durch das Abfließen von Ladungsträgern ergeben sich in unmittelbarer Umgebung des metallurgischen PN-Überganges Raumladungszonen. Die Ladungen entstehen durch die von den Ladungsträgern zurückgelassenen, nun ionisierten Donatoren und Akzeptoren. Die Raumladungszonen sind demnach negativ (bzw. positiv) auf der p-Seite (bzw. n-Seite) des PN-Überganges und erzeugen natürlich wie jede Raumladung ein inneres elektrisches Feld und damit eine Potentialdifferenz. Dadurch werden die Energieniveaus der n-Seite des PN-Überganges relativ zur p-Seite abgesenkt. Der Transfer von Ladungsträgern kommt zum Stillstand, wenn auf beiden Seiten des PN-Überganges die Fermi-Energien ausgeglichen wurden.

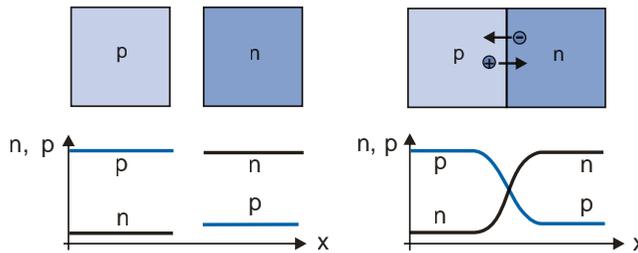


Abbildung 3.3: Ladungstransfer bei Kontakt eines n- und eines p-Gebietes. Die Ladungsträgerkonzentrationen werden dabei auf einer logarithmischen Skala aufgetragen, da die Minoritäts- und Majoritätsladungsträgerkonzentrationen sich um viele Größenordnungen unterscheiden.

Qualitativ wurde schon früher festgestellt, dass im thermodynamischen Gleichgewicht, d. h. ohne Stromtransport, die Fermi-Energie im gesamten Bauelement gleich sein muss, d. h. im Energiediagramm horizontal verläuft.

3.1.2 Der PN-Übergang im stromlosen Zustand

Die Situation kann man aber auch quantitativ betrachten. Der Gesamtstrom durch ein Bauelement setzt sich, wie in *Kapitel 2* beschrieben, zusammen aus dem Driftstrom und dem Diffusionsstrom. Dabei müssen sowohl die Elektronenströme als auch die Löcherströme berücksichtigt werden.

$$J_{\text{gesamt}} = J_n + J_p \quad (3.2)$$

$$J_n = q \cdot n \cdot \mu_n \cdot E + q \cdot D_n \cdot \frac{\partial n}{\partial x} \quad (3.3)$$

$$J_p = q \cdot p \cdot \mu_p \cdot E - q \cdot D_p \cdot \frac{\partial p}{\partial x} \quad (3.4)$$

Hinweis

Da wir hier die Gleichungen nur in einer Dimension betrachten, kommt die Ableitung nur nach der Ortskoordinate x vor. Im allgemeinen Fall steht hierfür der Gradient von n , damit müssten dann alle Größen **vektoriell** geschrieben werden.

Der Diffusionskoeffizient D hängt mit der Beweglichkeit über die Einstein-Gleichung zusammen.

$$D_{n,p} = \frac{k \cdot T}{q} \cdot \mu_{n,p} \quad (3.5)$$

Sowohl der Elektronen- als auch der Löcherstrom müssen null sein, da an der Diode zunächst keine äußere Spannung anliegen soll.

$$\begin{aligned} J_n &= q \cdot n \cdot \mu_n \cdot E + q \cdot D_n \cdot \frac{\partial n}{\partial x} = 0 \\ J_n = 0 &= q \cdot \mu_n \cdot \left(n \cdot E + \frac{k \cdot T}{q} \cdot \frac{\partial n}{\partial x} \right) = 0 \end{aligned} \quad (3.6)$$

Für den nicht entarteten Fall konnte die Ladungsträgerkonzentration n aus der Position der Fermi-Energie W_F relativ zur Leitungsbandkante bestimmt werden (siehe Gleichung (2.53)). Hieraus kann $grad(n)$ berechnet werden.

$$n = N_L \cdot \exp\left\{ \frac{W_F - W_L}{k \cdot T} \right\} \quad (3.7)$$

N_L ist dabei die Zustandsdichte, die bei extrinsischer Dotierung durch die Donatorkonzentration N_D ersetzt werden muss. Daraus ergibt sich

$$grad(n) = \frac{\partial n}{\partial x} = \frac{n}{k \cdot T} \cdot \left(\frac{\partial W_F}{\partial x} - \frac{\partial W_L}{\partial x} \right) \quad (3.8)$$

Der Gradient des Leitungsbandes ist aber gerade das elektrische Feld

$$grad(W_L) = q \cdot E \quad (3.9)$$

Eingesetzt in Gleichung (3.6) ergibt sich daraus

$$J_n = 0 = n \cdot \mu_n \cdot \frac{\partial W_F}{\partial x} \quad (3.10)$$

Dies bedeutet, dass ein Gradient in der Fermi-Energie grundsätzlich einen Strom nach sich ziehen muss. Umgekehrt gilt, dass bei Stromfluss ein Gradient der Fermi-Energie vorhanden sein muss.

Ohne Stromfluss muss die Fermi-Energie unabhängig vom Ort demnach konstant sein. Gleichzeitig muss W_F aber im n-dotierten Bereich in der Nähe des Leitungsbandes liegen, während sie im p-dotierten Bereich in der Nähe des Valenzbandes liegt. Dies ergibt dann den Bandverlauf über den PN-Übergang hinweg, wie er in Abbildung 3.4 skizziert ist. Die Potentialdifferenz zwischen Leitungsband im n-Bereich und Leitungsband im p-Bereich wird Diffusionspotential oder -spannung U_D genannt.

Schon an dieser Stelle soll festgestellt werden, dass eine äußere Spannung, die zusätzlich über den PN-Übergang angelegt wird, zur Diffusionsspannung addiert werden muss. Die wirksame Potentialdifferenz ist dann $(U_D - U)$. U ist positiv, sofern der p-Bereich positiv gegen den n-Bereich vorgespannt wird. Die Energieniveaus der p-Seite werden dann für Elektronen energetisch günstiger, d. h., Leitungs- und Valenzband werden dort in der Energie abgesenkt, also im Diagramm der Abbildung 3.4 nach unten verschoben. Leitungs- und Valenzband im n-Bereich dagegen werden energetisch ungünstiger und nach oben verschoben. Die Fermi-Niveaus im p- und n-Bereich unterscheiden sich dann ebenfalls durch die Potentialdifferenz U bzw. die Energiedifferenz $q \cdot U$.

Die Ladungsträgerkonzentrationen in dem Bereich, in dem W_F sowohl vom Leitungsband als auch vom Valenzband relativ weit entfernt ist, sind sehr klein gegenüber den jeweiligen Ladungsträgerkonzentrationen in den neutralen n- und p-Bereichen, da diese laut Gleichung (3. 7) exponentiell mit wachsendem Abstand $(W_F - W_{L/V})$ abnimmt.

$$n(x) = N_D \cdot \exp\left\{-\frac{W_L(x) - W_L(\infty)}{k \cdot T}\right\} \tag{3.11}$$

N_D ist dabei die Konzentration an Donatoren; die thermische Aktivierung von Ladungsträgern über die Bandlücke hinweg (intrinsische Ladungsträgerkonzentration) wird dabei vernachlässigt, da diese sehr klein im Vergleich zur extrinsischen Dotierung ist. Der Bereich, in dem das Leitungs- und das Valenzband variieren, ist demnach arm an Ladungsträgern, da gerade dort der Abstand zum Fermi-Niveau W_F groß ist. Dieser Bereich wird deshalb Verarmungszone genannt. In der Verarmungszone liegen noch die Donatoren und Akzeptoren, die positiv bzw. negativ geladen sind. Da die jeweils kompensierenden freien Ladungsträger hier größtenteils fehlen, treten innerhalb der Verarmungszone Raumladungen auf, die durch die (nicht mobilen) Ladungen der Donatoren und Akzeptoren verursacht werden. Die Verarmungszone wird deshalb auch **Raumladungszone** genannt.

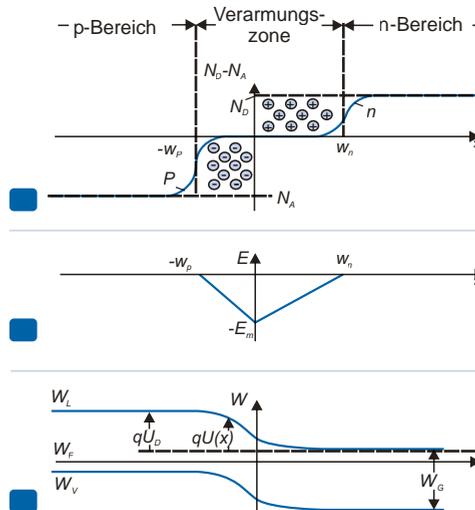


Abbildung 3.4: (a) Netto-Ladungsträgerkonzentrationen $N_D - N_A$ über die Raumladungszone hinweg. (b) Verlauf des elektrischen Feldes. (c) Verlauf des Leitungs- und des Valenzbandes. Die Bandverbiegung führt zur Diffusionsspannung U_D , nach [18].

Jedes Elektron, das die n-Seite verlässt, kompensiert auf der p-Seite ein Loch (und umgekehrt). Auf beiden Seiten des PN-Überganges müssen deshalb genau gleich viele Ladungen auftreten. Mit der Konzentration an Donatoren N_D und Akzeptoren N_A ergibt sich

$$N_A \cdot w_p = N_D \cdot w_n \quad (3.12)$$

w_p bzw. w_n ist die Ausdehnung der Raumladungszone ins p- bzw. n-Gebiet. Je höher die Dotierkonzentration, desto kleiner die Ausdehnung der Raumladungszone.

Innerhalb der Raumladungszone herrscht ein elektrisches Feld, das über die Poisson-Gleichung berechnet werden kann. Die Poisson-Gleichung für die n-Seite und die p-Seite des Überganges lautet:

$$\frac{\partial^2 \Phi}{\partial x^2} = \frac{\partial E}{\partial x} = \frac{q}{\epsilon_s} \cdot N_D^+ \quad (3.13)$$

$$\frac{\partial^2 \Phi}{\partial x^2} = \frac{\partial E}{\partial x} = \frac{q}{\epsilon_s} \cdot N_A^- \quad (3.14)$$

Hierin ist ϵ_s die gesamte Dielektrizitätskonstante im Halbleiter (einschließlich ϵ_0).

Im Prinzip hätte man auf der rechten Seite der Poisson-Gleichung noch die jeweils vorhandenen freien Ladungsträger für die Ladungsbilanz mitberücksichtigen müssen. Deren Konzentration ist aber innerhalb der Raumladungszone normalerweise so klein, dass diese vernachlässigt werden kann. Da auf der rechten Seite der Poisson-Gleichung eine konstante Größe steht, ergibt sich aus der direkten Integration damit ein linearer Verlauf des elektrischen Feldes und ein quadratischer Verlauf des elektrischen Potentials als Funktion des Orts x , siehe Abbildung 3.4.

Für das Feld $E(x)$ ergibt sich durch Integration der Gleichungen (3.13) und (3.14)

$$E(x) = \frac{q}{\epsilon_s} \cdot N_D^+ \cdot (x - w_n) \text{ für } 0 < x < w_n \quad (3.15)$$

$$E(x) = \frac{q}{\epsilon_s} \cdot N_A^- \cdot (x + w_p) \text{ für } -w_p < x < 0$$

Das maximale Feld E_m ergibt sich genau am metallurgischen PN-Übergang bei $x = 0$.

$$E_m = E(x=0) = \frac{q}{\epsilon_s} \cdot N_A^- \cdot w_p = \frac{q}{\epsilon_s} \cdot N_D^+ \cdot w_n \quad (3.16)$$

Außerhalb der Raumladungszone ist das elektrische Feld E gleich null. Wegen des linearen Zusammenhangs von $E(x)$ kann man (siehe Abbildung 3.4) direkt integrieren und es ergibt sich für die Diffusionsspannung U_D , d. h. die Bandverbiegung von Leitungs- und Valenzband über die Raumladungszone hinweg

$$U_D = \frac{1}{2} \cdot E_m \cdot (w_p + w_n); \quad w_p + w_n = w. \quad (3.17)$$

Zusammen mit Gleichung (3.16) kann man E_m eliminieren und erhält für die Ausdehnung der Raumladungszone

$$w = \sqrt{\frac{2 \cdot \epsilon_s \cdot U_D}{q} \cdot \left(\frac{1}{N_D} + \frac{1}{N_A} \right)} \quad (3.18)$$

und für die beiden Abschnitte in den p- und n-Gebieten

$$w_n = \sqrt{\frac{2 \cdot \epsilon_S \cdot U_D \cdot N_A}{q \cdot N_D \cdot (N_D + N_A)}} \quad (3.19)$$

$$w_p = \sqrt{\frac{2 \cdot \epsilon_S \cdot U_D \cdot N_D}{q \cdot N_A \cdot (N_A + N_D)}}$$

Hinweis

Ausdehnung der Raumladungszone:

Aus den Gleichungen ist zu entnehmen, dass die Breite der Raumladungszone bei höherer Dotierung immer kleiner wird. Bei asymmetrisch dotierten PN-Übergängen wird sich der größte Teil der Raumladungszone im niedriger dotierten Bereich befinden.

Beispiel

$N_A = 10^{19} \text{ cm}^{-3}$, $N_D = 10^{16} \text{ cm}^{-3}$, $\epsilon_S = 11,7 \cdot 8,854 \cdot 10^{-14} \text{ A s (V cm)}^{-1}$, $U_D = 900 \text{ mV}$ → $w_n = 0,34 \text{ } \mu\text{m}$; w_p ist entsprechend Gleichung (3.16) um drei Zehnerpotenzen kleiner.

Auch ein Metall-Halbleiter-Übergang, wie er für die Kontaktierung von Halbleiterbauelementen benötigt wird, erzeugt wie bei einer PN-Diode eine Bandverbiegung im Metall und im Halbleiter. Aufgrund der sehr viel höheren Elektronenkonzentration im Metall (ca. 10^{23} cm^{-3} gegenüber der Dotierung im Halbleiter ($10^{15} - 10^{17} \text{ cm}^{-3}$)) ist die Ausdehnung der Raumladungszone im Metall aber praktisch gleich null.

Betrachten wir die Situation an einem stromlosen PN-Übergang noch etwas genauer. Aufgrund des Ladungstransfers ergeben sich wie beschrieben Raumladungen, die zu einem elektrischen Feld, einem damit verbundenen Potentialverlauf und damit zu einer Bandverbiegung führen. Elektronen aus dem n-Bereich werden aufgrund des Konzentrationsgradienten auch ohne äußere Spannung in Richtung gegenüberliegende p-Seite diffundieren. Diese Diffusion wird aber durch das Potentialgefälle (also die Diffusionsspannung U_D) kompensiert. Diffusionsstrom und Feldstrom sind im Gleichgewicht und heben sich gegenseitig auf, der Nettostrom ist gleich null. Der PN-Übergang ist damit insgesamt stromlos.

3.1.3 Der PN-Übergang mit äußerer Spannung

Die Situation gerät allerdings aus dem Gleichgewicht, wenn eine äußere Spannung angelegt wird. Diese reduziert (in Durchlassrichtung) oder vergrößert (in Sperrrichtung) die Potentialdifferenz und damit die Bandverbiegung, je nach Polarität. Dabei fällt die äußere Spannung U über dem gesamten Bauelement ab, also über dem n-Bereich, dem eigentlichen PN-Übergang und dem p-Bereich. Ist allerdings die Dotierung auf beiden Seiten des PN-Überganges ausreichend hoch und fließt kein zu hoher Strom durch diese Zuleitungen, so kann man die Spannungsabfälle in den n- und p-Bahngebieten vernachlässigen. Die gesamte äußere Potentialdifferenz liegt dann direkt über der Raumladungszone an. Statt der Potentialdifferenz U_D , der

Diffusionsspannung, muss in die Gleichungen eine durch die äußere Spannung U modifizierte Potentialdifferenz ($U_D - U$) eingesetzt werden. Das Vorzeichen ergibt sich daraus, dass eine positive Spannung an der p-Seite relativ zur n-Seite die Potentialdifferenz erniedrigt. Diese Polarität wird als positive Spannung festgelegt.

In den bisher hergeleiteten Formeln muss damit U_D durch $(U_D - U)$ ersetzt werden. Es ergibt sich deshalb z. B. für die Breite der Raumladungszone:

$$w = \sqrt{\frac{2 \cdot \epsilon_S \cdot (U_D - U)}{q} \cdot \left(\frac{1}{N_D} + \frac{1}{N_A} \right)} \quad (3.20)$$

Ist U positiv (d. h. p-Bereich positiv gegenüber n-Bereich), so reduziert sich die Potentialdifferenz und die Diode wird in Durchlassrichtung betrieben. Obwohl die Potentialdifferenz die Elektronen immer noch von der p-Seite zur n-Seite treibt, ist dieser Feldstrom nun reduziert und kann den Diffusionsstrom nicht mehr kompensieren. In Summe ergibt sich damit ein Elektronenstrom vom n-Bereich in den p-Bereich. Analoges gilt für Löcher, die jetzt vom p-Bereich in den n-Bereich fließen.

Aufgabe ist es nun, diesen Diffusionsstrom von Elektronen vom n-Bereich hinein in den p-Bereich zu berechnen (und analog für Löcher). Beide Stromanteile müssen dann zum Gesamtstrom durch den PN-Übergang addiert werden.

Durch die Spannung in Durchlassrichtung wird auch die Breite der Raumladungszone entsprechend Gleichung (3.20) reduziert. Damit fällt die Ladungsträgerkonzentration durch die Raumladungszone nicht mehr auf die Gleichgewichtswerte n_{p0} und p_{n0} ab (siehe Abbildung 3.5). Die am Rande der reduzierten Raumladungszone erhöhte Minoritätsladungsträgerkonzentration n_p ist Ursache für einen Diffusionsstrom von Elektronen in den p-Bereich hinein. Während dieser Diffusion werden Ladungsträger rekombinieren, da diese ja auf eine hohe Konzentration an entgegengesetzt geladenen Majoritätsladungsträgern, hier Löcher im p-Bereich, treffen. Die Konzentration an Minoritätsladungsträgern (hier Elektronen im p-Bereich) wird deshalb in den p-Bereich hinein abnehmen. Für Löcher kann analog argumentiert werden.

Wir nehmen für das Weitere an, dass sich der Elektronenstrom innerhalb der relativ dünnen Raumladungszone nicht ändert, Ladungsträgerrekombination in der Raumladungszone selbst wird also vernachlässigt. Dies scheint gerechtfertigt, da aufgrund des hohen Feldes in der Raumladungszone Elektronen und Löcher beschleunigt werden und diese schnell durchqueren. Rekombination in der Raumladungszone wird erst später wichtig, z. B. bei LEDs, wo gerade durch die Rekombination in der Raumladungszone Licht erzeugt werden soll. Der Elektronenstrom am „Eingang“ der Raumladungszone (bei w_n) muss wegen der Vernachlässigung von Rekombination in der Raumladungszone damit gleich dem Elektronenstrom am „Ausgang“ der Raumladungszone (bei $-w_p$) sein.

Daraus folgt, dass der Diffusionsstrom von Elektronen in den p-Bereich hinein durch einen Elektronenstrom im n-Bereich nachgeliefert werden muss. Dieser Strom durch die hoch n-dotierten Zuleitungen ist ein Feldstrom. Wegen $J = q \cdot n \cdot \mu \cdot E$ ist die notwendige Feldstärke aufgrund der hohen Elektronenkonzentration im n-Bereich allerdings sehr klein und kann praktisch vernachlässigt werden. Die aufgrund dieser Feldstärke vorhandene Verkipfung der Bänder in den n-dotierten Zuleitungen ist in dem verwendeten Maßstab in Abbildung 3.5 normalerweise nicht sichtbar. Die Ausführungen entsprechen der Annahme, dass die gesamte Spannung am PN-Übergang selbst und nicht in den Zuleitungen abfällt, und gelten analog auch wieder für Löcher auf der p-Seite. Streng genommen muss natürlich eine (kleine) Spannung an den Zuleitungen abfallen, damit durch diese eine ausreichende

Stromdichte getrieben wird. Die Stromdichte durch die Zuleitungen muss genauso hoch sein wie die Stromdichte durch den eigentlichen PN-Übergang.

Im Folgenden wird die Situation nun quantitativ beschrieben. Stromfluss bedeutet grundsätzlich Nicht-Gleichgewicht. Zur Beschreibung der Konzentrationen an Elektronen und Löchern müssen daher die Quasi-Fermi-Niveaus herangezogen werden. Da wir Nicht-Entartung voraussetzen, die Ladungsträgerkonzentrationen also grundsätzlich nicht zu hoch sein sollen, können wir die Boltzmann-Näherung verwenden. Dies vereinfacht die Rechnung. Es gilt dann

$$n = n_i \cdot \exp\left\{\frac{W_{Fn} - W_i}{k \cdot T}\right\} \quad (3.21)$$

$$p = p_i \cdot \exp\left\{\frac{W_i - W_{Fp}}{k \cdot T}\right\} \quad (3.22)$$

Beide Gleichungen können als Definitionsgleichungen für die Quasi-Fermi-Energien angesehen werden. W_i ist die Position der Fermi-Energie im intrinsischen Fall, für den $n = p$ gilt. Für das Produkt $n \cdot p$ folgt daraus

$$n \cdot p = n_i^2 \cdot \exp\left\{\frac{W_{Fn} - W_{Fp}}{k \cdot T}\right\} \quad (3.23)$$

Die Differenz der Quasi-Fermi-Niveaus ist dabei gleich der außen angelegten Spannung U (Verlauf der Quasi-Fermi-Niveaus siehe Abbildung 3.5)

$$q \cdot U = W_{Fn} - W_{Fp} \quad (3.24)$$

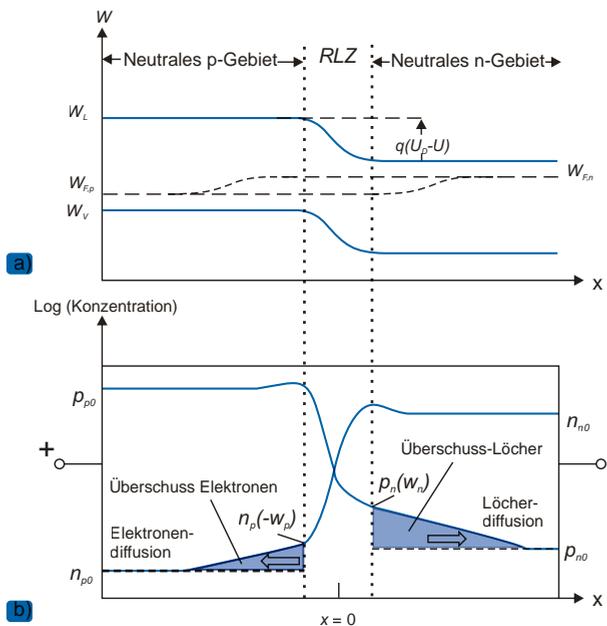


Abbildung 3.5: Verlauf der Quasi-Fermi-Niveaus (a) sowie der Elektronen- und Löcherkonzentrationen (b) an einem PN-Übergang mit angelegter äußerer Spannung U in Durchlassrichtung. Die Ladungsträgerkonzentrationen sind hier auf einer logarithmischen Skala gezeigt, der Anstieg der Majoritätsträgerkonzentrationen ist aus didaktischen Gründen übertrieben groß dargestellt.

Dies bedarf einer Begründung. Dazu betrachten wir den Elektronenstrom am „Eingang“ und am „Ausgang“ der Raumladungszone, der ja proportional zum Gradienten der Fermi-Energie ist.

$$J_n(w_n) = n \cdot \mu_n \left. \frac{\partial W_{Fn}}{\partial x} \right|_{w_n} = n \cdot \mu_n \left. \frac{\partial W_{Fn}}{\partial x} \right|_{-w_p} = J_n(-w_p) \quad (3.25)$$

Aus der Gleichung ist direkt ersichtlich, dass der Gradient der Fermi-Energie nur dort signifikant groß ist, wo die minimale Ladungsträgerkonzentration n vorliegt. Dies ist weder im n-Bereich noch in der Raumladungszone, sondern eben nur im p-Bereich der Fall. Nur dort sind Elektronen Minoritätsladungsträger. n nimmt kleinste Werte an und damit muss der Gradient der Fermi-Energie dort maximal sein. Die Quasi-Fermi-Energie ist deshalb im n-Bereich und durch die gesamte Raumladungszone hinweg konstant, da der Gradient hier vernachlässigbar ist. Analoge Überlegungen gelten für die Quasi-Fermi-Energien von Löchern. Daraus folgt der Verlauf der Quasi-Fermi-Energien aus Abbildung 3.5 und damit auch Gleichung (3.24).

Eingesetzt in die Gleichung (3.23) ergibt

$$n_n \cdot p_n = n_i^2 \cdot \exp\left\{\frac{q \cdot U}{k \cdot T}\right\} \quad (3.26)$$

und

$$p_n(w_n) = \frac{n_i^2}{n_n} \cdot \exp\left\{\frac{q \cdot U}{k \cdot T}\right\} = p_{n0} \cdot \exp\left\{\frac{q \cdot U}{k \cdot T}\right\} \quad (3.27)$$

Die entsprechende Gleichung gilt auch am anderen Ende der Raumladungszone für die Elektronen:

$$n_p(-w_p) = n_{p0} \cdot \exp\left\{\frac{q \cdot U}{k \cdot T}\right\} \quad (3.28)$$

Damit ist nun die Konzentration an Minoritätsladungsträgern an den beiden Enden der Raumladungszone als Funktion der von außen angelegten Spannung bekannt. Die Differenz zur jeweiligen Gleichgewichtskonzentration ist die Ursache für den Diffusionsstrom. Dieser Diffusionsstrom zeigt in Richtung niedrigerer Konzentration, für Elektronen also in den p-Bereich hinein, für Löcher in den n-Bereich hinein.

Zur Berechnung dieses Diffusionsstromes gehen wir von den Kontinuitätsgleichungen aus:

$$\begin{aligned} \frac{dn}{dt} &= \frac{1}{q} \cdot \frac{\partial J_n}{\partial x} - r_{net} = 0 \\ \frac{dp}{dt} &= -\frac{1}{q} \cdot \frac{\partial J_p}{\partial x} - r_{net} = 0 \end{aligned} \quad (3.29)$$

Hinweis

Da wir hier die Gleichungen nur in einer Dimension betrachten, kommt die Ableitung nur nach der Ortskoordinate x vor. Im allgemeinen Fall steht hierfür die Divergenz des Stroms.

Die Stromdichten ergeben sich als Summe von Feld- und Diffusionsstrom

$$\begin{aligned} J_n &= q \cdot n \cdot \mu_n \cdot E + q \cdot D_n \cdot \frac{dn}{dx} \\ J_p &= q \cdot p \cdot \mu_p \cdot E - q \cdot D_p \cdot \frac{dp}{dx} \end{aligned} \quad (3.30)$$

und eingesetzt in die Kontinuitätsgleichungen ergibt sich damit:

$$\begin{aligned} \mu_n \cdot E \cdot \frac{dn}{dx} + n \cdot \mu_n \cdot \frac{dE}{dx} + D_n \cdot \frac{d^2n}{dx^2} - r_{net} &= 0 \\ \mu_p \cdot E \cdot \frac{dp}{dx} + p \cdot \mu_p \cdot \frac{dE}{dx} + D_p \cdot \frac{d^2p}{dx^2} - r_{net} &= 0 \end{aligned} \quad (3.31)$$

Diese Differenzialgleichungen liefern als Ergebnis $n(x)$ im p-Bereich und $p(x)$ im n-Bereich, woraus sich dann die Stromdichten J_n und J_p ergeben. Zur Lösung muss allerdings zunächst die Rekombinationsrate r_{net} präzisiert werden.

Für die Rekombination von Ladungsträgern aufgrund der im Vergleich zum Gleichgewicht erhöhten Konzentration von Löchern und Elektronen wird hierzu ein einfacher, aber oft verwendeter Ansatz gemacht:

$$r_{net} = \frac{(p_n - p_{n0})}{\tau_p} \quad (3.32)$$

Dieser Ansatz für die Rekombinationsrate wird Relaxationszeitansatz genannt, eine analoge Gleichung gilt für Elektronen. Die Rekombinationsrate steigt proportional zur Abweichung der Ladungsträgerkonzentration von der Gleichgewichtskonzentration an. Nachdem die Störung wegfällt, findet das Ladungsträgersystem mit der Relaxationszeit τ wieder ins Gleichgewicht zurück.

Um die Differentialgleichung zu vereinfachen, müssen zunächst Gleichung (3.31) (erste Gleichung) mit $(p_n \cdot \mu_p)$ und die Gleichung (3.31) (zweite Gleichung) mit $(n_n \cdot \mu_n)$ multipliziert und dann beide Gleichungen addiert werden. Weiterhin nehmen wir Ladungsneutralität an. Dies bedeutet

$$(p_n - p_{n0}) = (n_n - n_{n0}) \quad (3.33)$$

Die örtlichen Ableitungen von p_n und n_n stimmen deshalb überein. Im gesamten n-Bereich sind die Löcher Minoritätsladungsträger, damit gilt

$$p_n \ll n_n \quad (3.34)$$

Weiterhin können Terme mit dem elektrischen Feld als Faktor vernachlässigt werden, da das Feld aufgrund der Ladungsneutralität verschwindend klein ist. Daneben wird noch die Einstein-Gleichung (3.5) verwendet.

Mit den genannten Näherungen ergibt sich eine neue Form der obigen Differentialgleichung:

$$\frac{d^2 p_n}{dx^2} - \frac{p_n - p_{n0}}{D_p \cdot \tau_p} = 0 \quad (3.35)$$

Die dazu gehörige Lösung lautet unter Berücksichtigung von Gleichung (3.27):

$$p_n(x) - p_{n0} = p_{n0} \cdot \left(\exp\left(\frac{q \cdot U}{k \cdot T}\right) - 1 \right) \cdot \exp\left\{-\frac{x - w_n}{L_p}\right\} \quad (3.36)$$

Die Größe L_p wird Diffusionslänge genannt, in diesem Fall ist dies die Diffusionslänge von Löchern im n-Bereich:

$$L_p = \sqrt{D_p \cdot \tau_p} \quad (3.37)$$

Geht man eine Diffusionslänge vom Rand der Raumladungszone bei w_n in den n-Bereich hinein, so ist die Löcherkonzentration auf $1/e$ der Ausgangskonzentration bei w_n durch Rekombination zurückgegangen. Daraus erschließt sich die Bedeutung des Begriffs Diffusionslänge.

Aus Gleichung (3.36) ergibt sich direkt die Diffusionsstromdichte von Löchern im n-Bereich am Ende der Raumladungszone bei w_n

$$\begin{aligned} J_p(w_n) &= q \cdot p \cdot \mu_p \cdot E - q \cdot D_p \cdot \frac{\partial p}{\partial x} \approx \left[-q \cdot D_p \cdot \frac{\partial p}{\partial x} \right]_{w_n} \\ &= \frac{q \cdot D_p \cdot p_{n0}}{L_p} \cdot \left(\exp \left\{ \frac{q \cdot U}{k \cdot T} \right\} - 1 \right) \end{aligned} \quad (3.38)$$

Dabei wurde die Feldstromdichte, die ja proportional zum elektrischen Feld ist, vernachlässigt. Das Feld im n-Bereich ist wie oben besprochen sehr klein und wird dann auch noch mit einer kleinen Minoritätsträgerkonzentration p multipliziert. Eine analoge Gleichung gilt für Elektronen im p-Bereich:

$$J_n(-w_p) = \frac{q \cdot D_n \cdot n_{p0}}{L_n} \cdot \left(\exp \left\{ \frac{q \cdot U}{k \cdot T} \right\} - 1 \right) \quad (3.39)$$

Die Gesamtstromdichte durch den PN-Übergang ergibt sich als Summe dieser beiden Anteile, da laut Voraussetzung keine Rekombination in der Raumladungszone stattfinden soll:

$$J = J_{\text{Gesamt}} = J_p(w_n) + J_n(-w_p) = J_S \cdot \left(\exp \left\{ \frac{q \cdot U}{k \cdot T} \right\} - 1 \right) \quad (3.40)$$

mit der Sättigungsstromdichte J_S

$$J_S = \frac{q \cdot D_p \cdot p_{n0}}{L_p} + \frac{q \cdot D_n \cdot n_{p0}}{L_n} \quad (3.41)$$

Dies ist die berühmte Shockley-Gleichung, die die Strom-Spannungs-Charakteristik eines PN-Übergangs beschreibt. Wesentliche Annahme war, dass keine Rekombination von Ladungsträgern in der Raumladungszone stattfindet und die äußere Spannung nur am PN-Übergang selbst und nicht an den Zuleitungsbereichen abfällt. Man bezeichnet dies oft als „ideale Diode“. Positive Spannungen U zeigen die Durchlassrichtung an. Der p-Bereich ist dann positiv gegenüber dem n-Bereich vorgespannt und der Strom steigt exponentiell mit der äußeren Spannung U an.

Mit der Einstein-Beziehung (Gleichung (3.5)) kann die Sättigungsstromdichte J_S in folgende Form gebracht werden:

$$J_S = \frac{k \cdot T \cdot \mu_p \cdot n_i^2}{L_p} \cdot \frac{1}{N_D} + \frac{k \cdot T \cdot \mu_n \cdot n_i^2}{L_n} \cdot \frac{1}{N_A} \quad (3.42)$$

In integrierten Schaltungen kommen oft asymmetrisch dotierte PN-Übergänge vor. Aus Gleichung (3.42) ergibt sich, dass in diesem Fall der Stromtransport vor allem von den Ladungsträgern des höher dotierten Bereichs getragen wird. In einem n⁺-p-Übergang fließt fast nur Elektronenstrom, der Löcherstrom ist meist vernachlässigbar. Beim Bipolar-Transistor hat er allerdings einen Einfluss auf die sogenannte Stromverstärkung.

Aus Gleichung (3.42) ergibt sich auch die Temperaturabhängigkeit des Sperrstroms. Mit

$$n_i^2 = N_L \cdot N_V \cdot \exp\left\{\frac{-W_G}{k \cdot T}\right\} \quad (3.43)$$

ergibt sich

$$J_s \sim T \cdot \exp\left\{\frac{-W_G}{k \cdot T}\right\} \quad (3.44)$$

Im Durchlassbereich kommt noch der exponentielle Anstieg des Stromes mit der Spannung dazu:

$$J \sim T \cdot \exp\left\{\frac{q \cdot U - W_G}{k \cdot T}\right\} \quad (3.45)$$

Da $q \cdot U$ normalerweise deutlich kleiner als W_G ist, folgt für den Durchlassbereich dieselbe Temperaturabhängigkeit wie für den Sperrbereich.

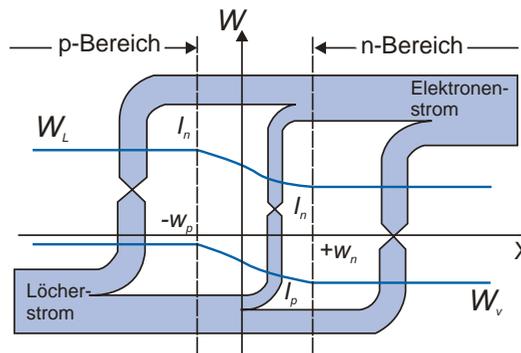


Abbildung 3.6: Zusammenfassung der Ergebnisse für den Stromtransport durch eine PN-Diode.

Zusammenfassend kann Folgendes festgehalten werden: Der Strom durch einen PN-Übergang wird im n-Gebiet ganz rechts in Abbildung 3.6 von Elektronen getragen. Es ist ein Feldstrom. Die hierfür notwendigen Felder sind aufgrund der hohen Majoritätskonzentrationen allerdings sehr klein. In diesem Bereich ist deshalb keine Verkippung der Bänder sichtbar. Im Übergangsbereich der n-Zuleitung (Diffusionszone) treffen Elektronen auf Löcher, die vom p-Bereich diffundieren. Dies führt zu einer zunehmenden Rekombination. Der Strom am „Elektronen-Eingang“ des PN-Bereichs bei w_n besteht also aus einem Elektronen-Feldstrom und einem Löcher-Diffusionsstrom. Elektronen durchqueren dann die Raumladungszone, in der ein großes Feld vorhanden ist. Durch Rekombination (z. B. bei einer LED) oder Generation (z. B. bei einer Solarzelle) von Ladungsträgern in der Raumladungszone könnte ein zusätzlicher Stromanteil entstehen, der aber in den vorherigen Überlegungen vernachlässigt wurde. Da analoge Überlegungen für Löcher gelten, besteht der Strom bei $-w_p$ aus einem Elektronen-Diffusionsstrom und einem Löcher-Feldstrom.

3.1.4 Die Kapazität eines PN-Übergangs

Die Sperrschichtkapazität

Hinweis

Ein PN-Übergang hat auch kapazitive Eigenschaften. Kapazitäten müssen z. B. bei der Untersuchung der maximalen Schaltgeschwindigkeiten von Transistoren berücksichtigt werden, da PN-Übergänge und die damit verbundenen Raumladungszonen sowohl in Bipolar-Transistoren als auch in MOSFETs auftreten. Die Kapazität eines PN-Übergangs ist abhängig von der angelegten Spannung.

Ein PN-Übergang hat kapazitive Eigenschaften. Eine Kapazität ergibt sich, wenn sich bei einer Änderung der Spannung an einem Bauelement die Ladung ändert:

$$c' = \frac{\Delta Q}{\Delta U} \cdot \frac{1}{A} \quad (3.46)$$

A ist die Fläche des Bauelements, c' die differentielle Kapazität (deshalb kleiner Buchstabe) pro Fläche. Durch den hochgestellten Strich (') soll angedeutet werden, dass es sich um eine auf die Fläche bezogene Größe handelt.

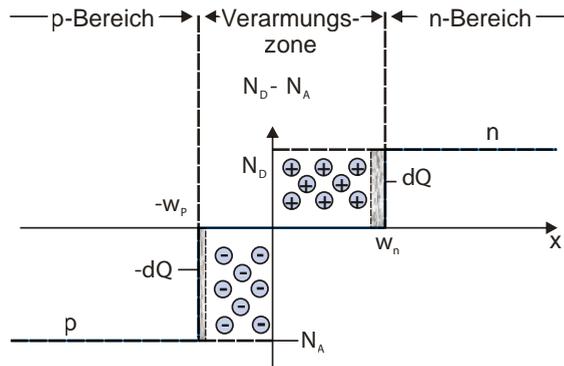


Abbildung 3.7: Raumladungszone an einer PN-Diode für zwei unterschiedliche Spannungen. Die Spannungsdifferenz zieht eine Umladung und damit eine Ladungsdifferenz nach sich.

Bei einer Änderung der Spannung an einem PN-Übergang ändert sich auch die Ausdehnung der Raumladungszone w (siehe Gleichung (3.20)):

$$w = \sqrt{\frac{2 \cdot \epsilon_s \cdot (U_D - U)}{q} \cdot \left(\frac{1}{N_D} + \frac{1}{N_A} \right)} \quad (3.47)$$

In Abbildung 3.7 ist die Situation bei Verkleinerung der Ausdehnung der Raumladungszone skizziert. Vom p-Bereich müssen Löcher mit der Ladung dQ in die Raumladungszone fließen, vom n-Bereich aus Elektronen mit der Ladung $-dQ$. Beide Ladungen sind gleich groß. In der