

Festkörperelektronik

Prof. Dr. Nils Weimann

Fachgebiet Bauelemente der Höchstfrequenzelektronik (BHE)

Fakultät Ingenieurwissenschaften

Universität Duisburg-Essen

<https://www.uni-due.de/bhe/>

nils.weimann@uni-due.de

9. April 2024

Inhaltsverzeichnis

Vorwort	x
Bücherliste	xi
1 Grundlagen der Quantentheorie	3
1.1 Quantennatur des Lichts	3
1.1.1 Energieerhaltung – Photoeffekt	5
1.1.2 Impulserhaltung – Compton-Effekt	7
1.2 Wellennatur der Materie	8
1.3 Heisenberg’sche Unschärferelation	10
1.4 Schrödinger-Gleichung	12
1.5 Freie Teilchen	15
1.5.1 Teilchen an einer Potentialstufe	16
1.5.2 Teilchen an einer Potentialbarriere	21
1.6 Gebundene Teilchen	26
1.6.1 Elektron im Potentialtopf	26
1.6.2 Teilchen im endlich tiefen Potentialtopf, Atommodell	30
1.6.3 Wasserstoffatom (H-Atom)	33
1.6.4 Bohr’sches Atommodell	36
1.6.5 Absorptionsspektroskopie	40
1.6.6 Emissionsspektroskopie	40
1.7 Kronig-Penny-Modell und Bandstruktur des Festkörpers	41
1.7.1 Bloch-Theorem	44
1.7.2 Fall 1: Elektronenenergie kleiner als Potentialbarriere	46
1.7.3 Fall 2: Elektronenenergie größer als Potentialbarriere	51
1.7.4 effektive Elektronenmasse; Elektronen und Löcher	51
1.7.5 Pauli-Prinzip im Festkörper	54

2	Elektronische Eigenschaften von Festkörpern	56
2.1	Einteilung in Isolatoren, Metalle und Halbleiter	56
2.1.1	Isolatoren	56
2.1.2	Halbleiter	58
2.1.3	Metalle	59
2.2	Konzentration von Elektronen und Löcher in Festkörpern	60
2.2.1	Fermi-Energie W_F für monovalentes Metall	67
2.2.2	Die Fermi-Dirac-Verteilungsfunktion $f(W)$ am Beispiel des mono- valenten Metalls)	69
2.2.3	Besetzungsstatistik in Halbleitern, Eigen- und Störstellenhalbleiter	70
2.2.4	Fermi-Energie W_{Fi} und Temperaturabhängigkeit	77
2.3	Störstellenhalbleiter	79
2.3.1	Punktförmige Defekte	79
2.3.2	Liniendefekte, Versetzungen	80
2.3.3	Zwei- und dreidimensionale Defekte	81
2.3.4	Dotierung von Halbleitern; Donatoren, Akzeptoren, tiefe Störstellen	81
2.3.5	Ladungsträgerstatistik im Störstellenhalbleiter	87
2.4	Ladungstransport in Festkörpern	94
2.4.1	Ladungsträgerbeweglichkeit und elektrische Leitfähigkeit	96
2.4.2	Ströme infolge eines elektrischen Feldes: Feldströme	105
2.4.3	Ströme infolge eines Konzentrationsgradienten: Diffusionsströme	106
2.4.4	Ladungsträgertransport bei Einwirkung eines magnetischen Feldes	109
2.4.5	Hall-Effekt	110
2.4.6	Kontinuitätsgleichungen	113
2.4.7	Poisson-Gleichung	116
2.4.8	Zusammenstellung der Halbleitergrundgleichungen für den La- dungstransport	117
3	Wechselwirkung zwischen elektromagnetischer Strahlung und Halbleitern	118
3.1	Absorption von Licht in Halbleitern	118
3.2	Einfluss der Bandstruktur auf die Absorption	120
3.2.1	Absorption durch freie Ladungsträger	122
3.2.2	Störstellenabsorption	122

Inhaltsverzeichnis

3.3	Photoleitung	122
3.3.1	Reine Elektron-Loch-Rekombination	124
3.3.2	Rekombination über Rekombinationszentren	127
3.3.3	Verstärkung in Photoleitern	128
3.4	Emission von Licht aus Halbleitern	130
3.4.1	Spontane Emission	130
3.4.2	Stimulierte Emission	132
3.4.3	Halbleiter-Laser	134
4	Halbleiteroberfläche	139
4.1	Bändermodell der Halbleiteroberfläche	140
4.2	Debye-Länge	143
4.3	Oberflächenleitfähigkeit	144
5	Halbleiter-Festkörper-Grenzflächen	147
5.1	Das Metall-Isolator-Halbleiter-System, MIS- und MOS-Kondensatoren	148
5.1.1	Bändermodell und Ladungsverteilung	148
5.1.2	Ideale MIS-Struktur	149
5.1.3	Nicht-ideale MIS-Struktur	154
5.1.4	Kapazitäts-Spannungs-Kennlinie der MIS-Diode	158
5.2	Metall-Halbleiter-Grenzflächen; Schottky-Dioden und Ohm-Kontakte	163
5.2.1	Bändermodell des Metall-Halbleiter-Kontaktes	163
5.2.2	Kapazitäts-Spannungskennlinie	167
5.2.3	Strom-Spannungs-Kennlinie des Metall-Halbleiter-Kontaktes	171
5.2.4	Schottky-Diode	171
5.2.5	Ohmsche Kontakte	179
5.3	Halbleiter-Halbleiter-Grenzflächen	182
5.3.1	pn -Übergang zwischen Halbleitern gleichen Materials (Homoübergang)	183
5.3.2	Stromloser pn -Übergang; Bänderdiagramm und Ladungsverteilung	183
5.3.3	Strom-Spannungscharakteristik des pn -Übergangs	195
5.3.4	Halbleiter-Halbleitergrenzflächen zwischen unterschiedlichen Halbleiterwerkstoffen; Heteroübergänge	204
5.3.5	Anwendungsbeispiele	208

Tabellenverzeichnis

2.1 Daten einiger Festkörper bei $T = 300\text{ K}$	75
2.2 Donator- und Akzeptor-Ionisierungsenergien verschiedener Halbleiter	85
3.1 verschiedene Anregungsarten der Lumineszenz	131
3.2 Halbleitersysteme und optische Anwendungen	137
4.1 Oberflächenleitfähigkeit	146
5.2 Vorgang an der MIS-Grenzfläche für unterschiedliche Differenzen der Austrittsarbeiten im n -Halbleiter	155
5.3 Berechnung des Bandverlaufs im pn -Übergang	190

Abbildungsverzeichnis

1.1	Energiespektrum der harmonischen Schwingung	4
1.2	Photoeffekt-Messanordnung	6
1.3	Auswertung des Photoeffekts	7
1.4	Compton-Effekt	7
1.5	Schwebung durch Addition zweier Wellenfunktionen	11
1.6	Aufenthaltswahrscheinlichkeit und Wellenfunktion eines freien Teilchens	16
1.7	einseitige Potentialstufe	17
1.8	endliche Potentialbarriere	21
1.9	Transmissionskoeffizient für endlich ausgedehnte Barriere	24
1.10	Potentialverlauf in Atomkernnähe	24
1.11	Tunneleffekt am Metall-Halbleiterkontakt	25
1.12	Elektron im Potentialtopf mit unendlich hohen Wänden	27
1.13	Energieeigenwerte und Wellenfunktionen im Potentialtopf	29
1.14	Erlaubte Energiewerte freier (Kurve) und gebundener (Punkte) Elektronen	31
1.15	Potentialtopf mit endlich hohen Wänden	31
1.16	Wellenfunktion (rot) im endlichen Potentialtopf	32
1.17	Graphische Lösung zum endlichen Potentialtopf	34
1.18	Coulombpotential des H-Atoms	35
1.19	Elektronenbahnen des Bohr'schen Atommodells	37
1.20	Orientierungsquantenzahl	38
1.21	Periodensystem	39
1.22	Absorptionslinie	40
1.23	Absorptionsspektrum	41
1.24	Emissionsspektrum	42
1.25	Potentialverlauf in einem Festkörper (eindimensional)	42
1.26	Kronig-Penney-Potentialverlauf	43
1.27	sinc-Funktion	48

Abbildungsverzeichnis

1.28 Grafische Lösung der Bestimmungsgleichung für das Kronig-Penney-Modell	49
1.29 Energiebänder des Festkörpers im Kronig-Penney-Modell	50
1.30 Energie, Geschwindigkeit und effektive Elektronenmasse	53
1.31 Aufspaltung zweier identischer Energieniveaus aufgrund des Pauli-Prinzips	54
1.32 Energieniveaus im Atom und Bänder im Festkörper	55
2.1 Aufspaltung der Niveaus als Funktion vom Atomabstand a	57
2.2 Bändermodell des Isolators	57
2.3 Bänderdiagramm für ein monovalentes Metall	59
2.4 Bänderdiagramm eines Metalls mit überlappenden Bändern	59
2.5 spezifischer Widerstand einiger Festkörper	61
2.6 Erlaubte k -Werte in einem eindimensionalen Potentialtopf der Breite L . . .	63
2.7 Erlaubte k -Werte im zweidimensionalen Potentialtopf mit den Abmessungen L^2	64
2.8 Flächen konstanter Energie für einen 3-dimensionalen Kristall	66
2.9 Zustandsdichte und effektive Masse	67
2.10 (a) Zustandsdichte und Elektronenverteilung in einem Metall ($T = 0$ K) und (b) daraus abgeleitete Verteilungsfunktion	69
2.11 Verteilungsfunktionen für das Leitungs- und das Valenzband im intrinsi- schen Halbleiter; exakte Form und Näherungen	74
2.12 Zusammenhang zwischen Verteilungsfunktion, Zustandsdichte und Ladungs- trägerkonzentration	75
2.13 Temperaturabhängigkeit der Eigenleitungskonzentration	76
2.14 Temperaturabhängigkeit der Fermi-Energie im intrinsischen Halbleiter . . .	77
2.15 Temperaturabhängigkeit der Fermi-Energie	78
2.16 Punktdefekte im Festkörper	79
2.17 Versetzung im Kristall	80
2.18 Donator Arsen in Silizium	82
2.19 Akzeptor Bor in Silizium	82
2.20 Dotierniveau im Banddiagramm: (a) Donator für n-Leitfähigkeit und (b) Akzeptor für p-Leitfähigkeit	83
2.21 Amphoteres Dotierverhalten von Si in GaAs	85
2.22 Aufspaltung des Donatorbiveaus bei steigender Dotierung	87
2.23 Besetzungsstatistik im n -Halbleiter	89

Abbildungsverzeichnis

2.24	Besetzungsstatistik im p -Halbleiter	89
2.25	temperaturabhängige Elektronenkonzentration in n -dotiertem Halbleiter, Er- schöpfungsbereich mit konstanter Konzentration	92
2.26	Besetzungswahrscheinlichkeit $f(W)$ in Abhängigkeit von der kinetischen Ener- gie ohne äußere Kraft	96
2.27	Quantentheoretisch erlaubte k -Werte in zwei Dimensionen ohne äussere Kraft	97
2.28	(a) Verschiebung der Fermi-Kugel (Fermi-Kreises) und (b) Veränderung der Geschwindigkeiten der einzelnen Elektronen durch Wirkung einer äußeren Kraft	97
2.29	Abhängigkeit der Elektronendriftbeweglichkeit von der Temperatur und der Dotierung	102
2.30	Feldstärkeabhängigkeit von Ladungsträgern in unterschiedlichen Halbleitern	104
2.31	$W(k)$ -Verlauf von GaAs (InP)	104
2.32	Verlauf der Bandkanten und des Potentials für einen Halbleiter im E-Feld .	107
2.33	Ablenkung von bewegten Elektronen in einem Magnetfeld	110
2.34	Richtung der Felder und Ströme im Hall-Effekt	110
2.35	Kontinuität in beliebigem quaderförmigen Volumen	114
3.1	Lichtabsorption	119
3.2	zum Absorptionsgesetz	119
3.3	Absorptionskoeffizient in Abhängigkeit von der Wellenlänge für verschiede- ne Halbleiter	120
3.4	Bandverlauf und Absorption für direkten Halbleiter	121
3.5	Bandverlauf und Absorption für indirekten Halbleiter	121
3.6	Absorption durch freie Ladungsträger	122
3.7	Störstellenabsorption im extrinsischen Halbleiter	123
3.8	Zeitliches Verhalten der Überschusskonzentration	129
3.9	Aufbau eines Photoleiters	129
3.10	Wirkungsweise (a) und Symbol der Leuchtdiode (b)	131
3.11	Die drei Wechselwirkungsprozesse zwischen Strahlungsfeld und atomarem System: a) spontane Emission, b) stimulierte Emission, c) (stimulierte) Ab- sorption	133
3.12	prinzipieller Aufbau eines Halbleiterlasers	135

3.13 AlGaAs/GaAs Streifenlaser mit vergrabener Doppelheterostruktur (Buried Heterostructure (BH)-Laser)	137
3.14 Doppel-Heterostruktur zur Erhöhung der lokalen Ladungsträgerkonzentration im aktiven Laserbereich	138
4.1 Bindungsverhältnisse an einer idealisierten Halbleiteroberfläche	139
4.2 Bändermodell und Ladungsverteilung an einer Halbleiteroberfläche: a) unmittelbar nach "Erzeugung", b) im Gleichgewicht	142
4.3 Debye-Länge in Silizium für verschiedene Dotierstoffkonzentrationen	144
4.4 Anreicherung einer p -Halbleiteroberfläche	145
5.1 Aufbau a) und Bändermodell der idealen MIS-Struktur bei äußerer Spannung $U = 0$ für b) n -Typ und c) p -Typ Halbleiter	151
5.2 Aufbau und Bändermodell einer idealen MIS-Struktur a) $U = 0$ (Flachband), b) $U > 0$ (Anreicherung)	152
5.3 Aufbau und Bändermodell einer idealen MIS-Struktur a) $U < 0$ (Verarmung), b) $U < 0$ (Inversion)	153
5.4 MIS-Struktur, Bändermodell und Ladungsverteilung für einen n -Halbleiter mit a) unterschiedliche Austrittsarbeiten Φ_M und Φ_{HL} , mit b) negative Oxidladungen oder mit c) negativen Grenzflächenladungen	156
5.5 Ladungsverteilung bei den äußeren Spannungen $U = 0$ (links) und $U = U_{FB}$ für die Fälle a) unterschiedliche Austrittsarbeiten Φ_M und Φ_{HL} , b) örtlich feste negative Oxidladungen oder c) örtlich feste negative Grenzflächenladungen	157
5.6 Anteile der Isolator- (c_{IS}) und der Raumladungskapazität (c_R) an der gesamten MIS-Kapazität	159
5.7 Spannungsabhängigkeit der MIS-Kapazität	159
5.8 Einfluss der Messfrequenz auf die Kapazitäts-Spannungsabhängigkeit (schematisch) für die Bereiche Verarmung und Anreicherung (1) ohne Umladungen (hohe Frequenzen) und (2) mit Umladungen (tiefe Frequenzen)	162
5.9 Spannungsabhängigkeit der Gesamtkapazität einer MIS-Struktur und Einfluss der Messfrequenz im Inversionsbereich	162
5.10 Zum Entstehen eines Metall-Halbleiter-Kontaktes: getrennte Systeme	164
5.11 Zum Entstehen des Metall-Halbleiter-Kontaktes: vereinte Systeme ($U = 0$)	165
5.12 Gemessene Barrierenhöhe auf n -Si bzw. n -GaAs für unterschiedliche Metalle	166

Abbildungsverzeichnis

5.13 Bändermodell und Ladungsträgerverteilung unter Berücksichtigung von Grenzflächenzuständen	168
5.14 Komplementäres Verhalten von Barrierenhöhen auf n - bzw. p -Halbleitern	169
5.15 Verlauf der Leitungsbandkante für unterschiedliche Spannungen	169
5.16 Spannungsabhängigkeit der Raumladungskapazität für zwei verschiedene Dotierungen	170
5.17 Thermionische Elektronenemission am Schottky-Kontakt ($U = 0$)	172
5.18 Normierte Strom-Spannungskennlinien einer Schottky-Diode bei $T = 300$ K	177
5.19 wie Abb.5.18, jedoch linearer Masstab	178
5.20 Einfluss des Bahnwiderstandes R_b auf die Kennlinie der Schottky-Diode	180
5.21 Bändermodell und Kennlinie eines Schottky- (a) und Ohm-Kontaktes (b)	181
5.22 Herstellung eines Ohm-Kontaktes durch Legieren	182
5.23 Typischer spezifischer Kontaktwiderstand in Abhängigkeit von der Dotierung	183
5.24 (a) Dotierstoff- und Ladungsträgerkonzentration sowie (b) Bändermodell eines abrupten pn -Überganges	185
5.25 Abrupter pn -Übergang im stromlosen Zustand, (a) Konzentration der Ladungsträger und Dotierstoffe, (b) Raumladung, (c) Feldverlauf, (d) Potentialverlauf, (e) Bandverlauf	189
5.26 Stromdurchflossener pn -Übergang ($U > 0$); Ladungsträgerverteilung und Bändermodell	197
5.27 Stromdurchflossener pn -Übergang ($U < 0$); Ladungsträgerverteilung und Bändermodell	198
5.28 Ladungstransport in (a) Homo- und (b) Heteroübergängen mit variierender Bandlücke W_g	205
5.29 Abrupte Typ I (a) und Typ II (b) Heteroübergänge (schematisch)	206
5.30 (a) nn -, (b) pp - und (c) pn -Heteroübergang; (zugrunde gelegt wurde der Typ I Heteroübergang) Eingezeichnet wurde zusätzlich W_0 , um die unterschiedlichen Elektronenaffinitäten beider Halbleiter und die beiden Diffusionsspannungen zu verdeutlichen	207
5.31 (a) Entstehung des Potentialtopfes und (b) Schichtaufbau sowie Bandstruktur eines HFET	209
5.32 Bänderdiagramm eines Heterostruktur-Bipolartransistors (HBT)	211
5.33 Doppel-Hetero-Laserstruktur	212
5.34 Bänderdiagramm einer QCL-Struktur	213

Vorwort

Das vorliegende Manuskript behandelt die Grundlagen der Festkörperelektronik, die als Vorlesung im 2. bzw. 4. Fachsemester im Bachelor-Studiengang EIT und Nanoengineering angeboten wird. In der Vorlesung werden Prinzipien der Quantenmechanik von Grund auf mit Ingenieurmathematik hergeleitet. Das Ziel der Vorlesung ist die mikroskopische Erklärung der makroskopisch erfassbaren Vorgänge im Halbleiter, besonders von fließendem Strom in Abhängigkeit von angelegter Spannung, sowie die Grundlagen der optoelektronischen Eigenschaften. Die in der Vorlesung behandelten Prinzipien bilden die Basis, um die Funktion von Halbleiterbauelementen aufgrund der elektronischen Bandstruktur in weiterführenden Vorlesungen behandeln zu können.

Das Skript stellt eine kompakte Zusammenfassung des im Rahmen der Vorlesung vorgestellten Stoffes dar. Da das Skript aber nicht unbedingt den gesamten, für die Klausur relevanten Inhalt der Vorlesung abdeckt, kann es diese nicht ersetzen. Besonders ist hierzu anzumerken, dass Lösungsansätze, mathematische Methoden und physikalische Zusammenhänge ausführlicher in der Vorlesung behandelt werden.

Der Text mit Abbildungen entstand aus vorigen Versionen von Prof. em. F. Tegude und Dr.-Ing. W. Brockerhoff.

Prof. Dr. Nils Weimann

Duisburg, 9. April 2024

Bücherliste

- [1] Ibach und Lüth. *Festkörperphysik*. de. Springer-Lehrbuch. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009. ISBN: 978-3-540-85794-5. DOI: [10.1007/978-3-540-85795-2](https://doi.org/10.1007/978-3-540-85795-2).
- [2] Harald Ibach und Hans Lüth. *Solid-State Physics: An Introduction to Principles of Materials Science*. Berlin, Heidelberg: Springer, 2009. ISBN: 978-3-540-93803-3. DOI: [10.1007/978-3-540-93804-0](https://doi.org/10.1007/978-3-540-93804-0).
- [3] Vimal Kumar Jain. *Solid State Physics*. Cham: Springer International Publishing, 2022. ISBN: 978-3-030-96016-2. DOI: [10.1007/978-3-030-96017-9](https://doi.org/10.1007/978-3-030-96017-9).
- [4] Frank Thueselt. *Physik der Halbleiterbauelemente*. de. Springer-Lehrbuch. Berlin/Heidelberg: Springer-Verlag, 2005. ISBN: 978-3-540-22316-0. DOI: [10.1007/b138004](https://doi.org/10.1007/b138004).

Einführung

Die Mikroelektronik ist heute überall – von Mobilfunk über Glasfasernetze und Cloud-server zu Internet of Things, Robotik, ... In jedem Auto sind heute *im Durchschnitt* (!) Halbleiterchips für über 600€ verbaut, Tendenz stark steigend aufgrund der Umstellung auf Elektromobilität. Viele gesellschaftlich wichtige Entwicklungen werden durch die Chiptechnologie erst ermöglicht: Photovoltaik-, Windkraft- und andere Anlagen zur erneuerbaren Energiegewinnung, Elektrifizierung des Verkehrs, oder zur Wärmeerzeugung.

Zentrale und funktionsbestimmende Elemente der Mikroelektronik, wie Feldeffekt- und Bipolartransistoren, MOS-Kondensatoren, Dioden, Laser und Photodetektoren etc. sind aus Halbleitermaterialien aufgebaut, die in die Klasse der Festkörper gehören. Die Elektronik nutzt die elektronischen Eigenschaften der Festkörper, und zwar Volumen- und Oberflächen bzw. Grenzflächeneffekte, die äußerst gezielt durch elektrische und magnetische Felder, durch Licht sowie durch moderne technologische Verfahren gesteuert und beeinflusst werden können.

Die Grundlage dieser Festkörperelektronik ist die Quantentheorie des Festkörpers. Das Verständnis dieser Grundlagen ermöglicht es, für die unterschiedlichsten Anwendungen maßgeschneiderte Bauelemente und Systeme zu entwickeln und deren Eigenschaften zu optimieren. Neben einer (auch in Deutschland wieder) wachsenden Halbleiterindustrie ist dieses Fachwissen für alle Ingenieure, die als Endanwender mit Halbleiterkomponenten in Berührung kommen, von Vorteil, und wird auch vermehrt von einschlägigen Firmen und Industrieunternehmen nachgefragt.

Gegenstand der Vorlesung "Festkörperelektronik" ist die Entwicklung der elektronischen Volumeneigenschaften der Festkörper auf quantentheoretischer Basis. Im Anschluss werden, darauf aufbauend, Halbleiteroberflächen und -grenzflächen sowie die Wechselwirkung mit elektromagnetischer Strahlung behandelt und erste einfache Bauelemente beschrieben.

Dieser einsemestrige Kurs "Festkörperelektronik" wendet sich an Studierende der

Bücherliste

Elektrotechnik, des NanoEngineering-Studiengangs sowie des ISE und legt die Grundlagen für Veranstaltungen, welche die elektronischen Bauelemente und Schaltungen zum Inhalt haben.

1 Grundlagen der Quantentheorie

1.1 Quantennatur des Lichts

Historisch entwickelt hat sich die Quantentheorie aus der Beobachtung der Strahlung eines schwarzen Körpers. Dieser Begriff deutet an, dass dieser Körper ein breites Spektrum unterschiedlicher Wellenlängen bzw. Frequenzen abstrahlt und nicht wie beispielsweise ein Helium-Neon-Laser eine einzelne Wellenlänge, die einem einfarbigen (monochromatischen) Rot entspricht. Die klassische Theorie, mit der man bis dahin die Strahlung beschrieb, ging von schwingenden Atomen und Molekülen aus, beschrieb aber das Verhalten nur bei längeren Wellenlängen richtig. Nach dieser Theorie wuchs der Energieinhalt für immer schneller schwingende Teilchen über alle Grenzen (Energiedichte $U \rightarrow \infty$, wenn Frequenz $f \rightarrow \infty$ bzw. Wellenlänge $\lambda \rightarrow 0$).

Einen Ausweg wies Max Planck durch seine Quantenhypothese, nach der sich die Energie einer Schwingung der Frequenz f nicht kontinuierlich steigern lässt, sondern die nur abzählbar viele diskontinuierliche Energiezustände zulässt, die Vielfache einer Grundeinheit sind (Abb. 2.1):

$$W_n = n \cdot hf$$

mit der Quantenenergie W , der Frequenz f , dem Planck'schen Wirkungsquantum $h = 6.624 \times 10^{-34}$ Js und der Quantenzahl $n = \{1, 2, 3, \dots\}$.

Diese Planck'sche Quantenhypothese führte zur Übereinstimmung der Theorie mit den Experimenten und ist als Ausgangspunkt der Quantentheorie anzusehen. Bemerkenswert ist, dass hiermit einer seinerzeit typischen Wellenerscheinung (Lichtstrahlung) mit kontinuierlichen Eigenschaften nun diskrete und damit Teilcheneigenschaften, "Energieportionen", zugeordnet wurden. Einstein (1905) hat diesen Aspekt konsequent weiterentwickelt, indem er diesen Energieportionen Teilchen zuordnete, die **Pho-**

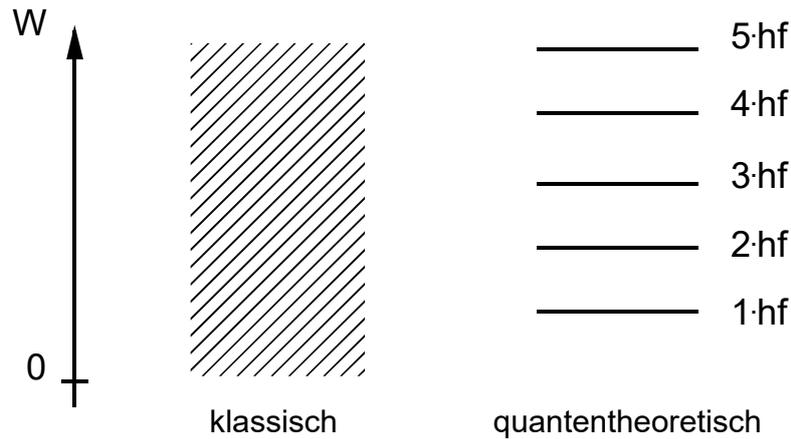


Abbildung 1.1: Energieverteilung einer harmonischen Schwingung in klassischer und quantenmechanischer Beschreibung

tonen, die jeweils einzeln eine Quantenenergie W_{ph} besitzen:

$$W_{ph} = h \cdot f = m_{ph} \cdot c^2 \quad (1.1)$$

mit der Photonenmasse m_{ph} und der Lichtgeschwindigkeit c .

Der letzte Teil der Gleichung (1.1) stellt den Bezug zur Relativitätstheorie her und führt auf die, neben der Energie, zweite charakteristische Größe für Teilchen, den Impuls p_{ph} des Photons:

$$p_{ph} = m_{ph} \cdot c = (h \cdot f)/c \quad (1.2)$$

Aus der Beziehung

$$\lambda \cdot f = c \quad (1.3)$$

folgt

$$p_{ph} = h/\lambda \quad (1.4)$$

und mit der Wellenzahl $k = 2\pi/\lambda$ und dem reduzierten Wirkungsquantum $\hbar = h/2\pi$ gilt auch

$$p = \hbar \cdot k \quad (1.5)$$

Der Impuls und die Wellenzahl sind Vektoren:

$$\vec{p}_p h = \hbar \cdot \vec{k} \quad (1.6)$$

Damit sind die wichtigsten Kenngrößen für Teilchen, die Energie und der Impuls, auch für die "Lichtteilchen", die Photonen, gefunden. Für diese Größen gelten bei der Wechselwirkung untereinander und mit anderen Partnern oder Mikrosystemen ganz allgemeine Erhaltungssätze:

$$\begin{aligned} h \cdot f_{ph} + W &= h \cdot f'_{ph} + W' && \text{Energieerhaltung} \\ \hbar \cdot \vec{k}_{ph} + \vec{p} &= \hbar \cdot \vec{k}'_{ph} + \vec{p}' && \text{Impulserhaltung} \end{aligned} \quad (1.7)$$

Dabei sind beispielsweise Atome, Kristalle oder allgemein Festkörper als Mikrosysteme aufzufassen. Mit diesen beiden Erhaltungssätzen sind beispielsweise die Prozesse der Absorption, Emission und Streuung von Licht zu beschreiben. Die Wechselwirkung findet besonders zwischen Photonen und den im Festkörper befindlichen Elektronen statt. Die Erhaltungssätze und damit die Quantennatur des Lichtes lassen sich mit relativ einfachen Experimenten überprüfen.

1.1.1 Energieerhaltung – Photoeffekt

Durch die Bestrahlung der Metallelektrode mit Photonen der Energie $h \cdot f$ (ultraviolettes Licht) wird den Elektronen Energie zugeführt, die diese zum Verlassen der Metalloberfläche veranlassen können. Je nach Beschaffenheit der Metalloberfläche und der Photonen besitzen diese zusätzlich noch eine kinetische Energie W_{kin} , die zu einer Bewegung der Elektronen auf die Gegenelektrode und damit zu einem elektrischen Strom führt, der im Außenkreis messbar ist. Zur Vereinfachung der Messung wird dieser Strom mittels eines Bremspotentials zu Null gemacht.

Dann gilt:

$$W_{kin} = \frac{1}{2} m_e v^2 = q \cdot \varphi = -q \cdot U \text{ für } I = 0 \quad (1.8)$$

Der Versuch besteht aus zwei Teilen:

1. Die Lichtintensität wird erhöht, aber die Frequenz f konstant gehalten:
 - a) Erwartung nach **klassischer Theorie**: Die Energie pro Zeiteinheit auf der Me-

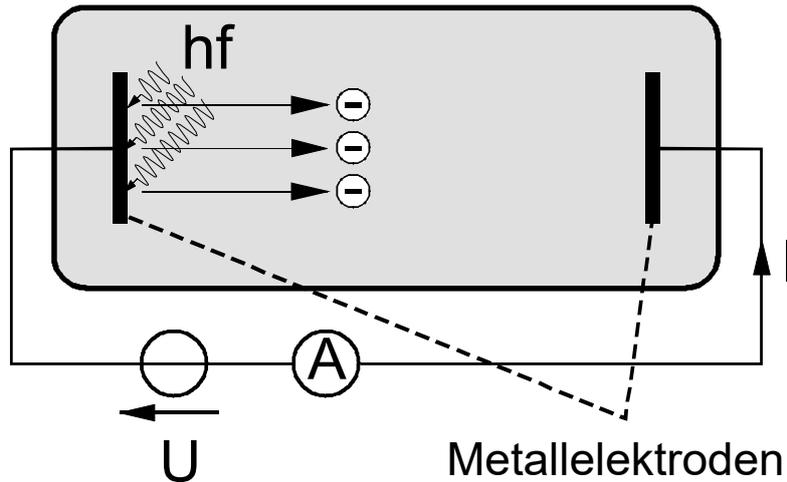


Abbildung 1.2: Photoeffekt-Messanordnung

taloberfläche nimmt zu und erhöht damit W_{kin} der Elektronen. Damit wird zur Einhaltung der Messrandbedingung $I = 0$ eine höhere Bremsspannung $-U$ notwendig.

- b) Erwartung nach **Quantentheorie**: Es gibt nur einzelne Lichtteilchen, Photonen, die einzeln in Elementarakt mit einzelnen Elektronen wechselwirken und Energie übertragen. Bei $f = \text{const.}$ bleibt damit $h \cdot f$ und somit auch W_{kin} der Elektronen unverändert, lediglich die Anzahl der Photonen wird erhöht. Die Bremsspannung wirkt aber auf jedes einzelne Elektron, somit bleibt trotz Erhöhung der Lichtintensität die für $I = 0$ notwendige Bremsspannung konstant.
2. Die Lichtintensität, d. h. die Energie pro Zeit und Flächeneinheit bleibt konstant, aber f wird erhöht:
- a) Erwartung nach klassischer Theorie: W_{kin} bleibt konstant und damit auch U
- b) Erwartung nach Quantentheorie: Mit steigender Photonenenergie $h \cdot f$ steigt auch die auf die Elektronen übertragene Energie W_{kin} und führt nach (1.8) zu einer höheren Bremsspannung $-U$.

In beiden Fällen bestätigt das Experiment die quantentheoretische Erwartung. Aus dem zweiten Teilversuch können zusätzlich weitere Erkenntnisse gewonnen werden: Trägt man die Bremsspannung $-U$ über der Lichtfrequenz f auf, so ergibt sich als Ach-

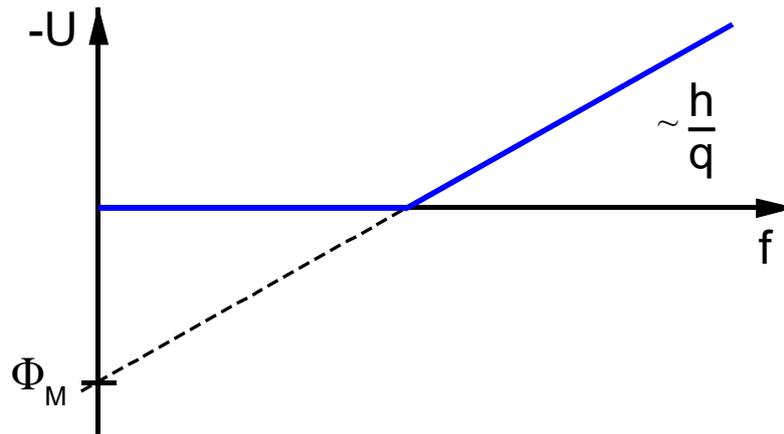


Abbildung 1.3: Auswertung des Photoeffekts

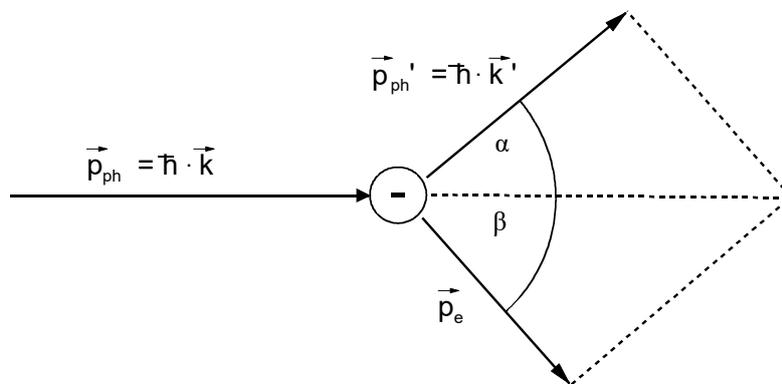


Abbildung 1.4: Compton-Effekt

senabschnitt die sogenannte Austrittsarbeit Φ_M der Elektronen aus der Metalloberfläche und aus der Steigerung der Messkurve der Quotient h/q (Abb. 1.3).

1.1.2 Impulserhaltung – Compton-Effekt

Photonen mit einem Impuls $\vec{p} = \hbar \vec{k}$ werden auf ruhende Elektronen geschossen. Mit zwei Detektoren (Abb. 1.4) werden sowohl Photonen- als auch Elektronenimpuls zum gleichen Zeitpunkt in der entsprechenden Winkelstellung gemessen. Dieser Anordnung liegt schon die quantentheoretische Vorstellung der Einzelwechselwirkung zwischen zwei Teilchen, Photon und Elektron, zugrunde. Sie wird durch die Auswertung der Messung gemäß (1.7) quantitativ bestätigt. Das Experiment wird mit hochenergetischen Röntgenquanten, die mit Elektronen in dünnen Metallfilmen wechselwirken, durchgeführt.

1.2 Wellennatur der Materie

Nachdem für die typische Wellenerscheinung Licht der Teilchencharakter nachgewiesen war, lag es nahe, typische Teilchen auf ihren Wellencharakter zu untersuchen. Beugung und Interferenz gilt als Nachweis für die Wellennatur, und Davisson und Germer (1927) konnten erstmals diese Phänomene an Kathodenstrahlen (Elektronenstrahlen) zweifelsfrei nachweisen. Hier sei angemerkt, dass nicht gemeint ist, Licht besitzt ausschließlich Teilchen- und Elektronen besitzen ausschließlich Welleneigenschaften. Vielmehr stellen diese beiden Beschreibungen, als Welle bzw. Teilchen, jeweils nur einen Aspekt der physikalischen Realität dar: jede Strahlung besitzt **simultan** Wellen- und Korpuskelnatur. Man nennt diesen Sachverhalt **Dualismus**.

Zurück zur Wellennatur von Korpuskeln: De Broglie formulierte in Analogie zur Einsteinschen Lichtquantenhypothese:

$$\begin{aligned} p &= m \cdot v = h/\lambda = \hbar \cdot k \\ W &= \frac{1}{2}mv^2 = h \cdot f = \hbar\omega \quad \text{mit} \quad \omega = 2\pi f \end{aligned} \quad (1.9)$$

Damit wird den massebehafteten Teilchen (m) über den Impuls eine Wellenzahl k und somit eine Wellenlänge λ und über die kinetische Energie eine Frequenz f bzw. ω zugeordnet. Die Beschreibung als Welle wird durch die Wellenfunktion Ψ vorgenommen:

$$\Psi(\vec{r}, t) = C \cdot \exp \left[j \left(\vec{k} \cdot \vec{r} - \omega \cdot t \right) \right] \quad (1.10)$$

In den Gleichungen für die Energie und den Impuls geht als wichtige Größe die Geschwindigkeit v des Teilchens ein. Wie findet sich diese Geschwindigkeit in der Wellenbeschreibung wieder? Die zunächst naheliegende Identifizierung der Teilchengeschwindigkeit ist die mit der Phasengeschwindigkeit der Welle, die man durch die Bedingung $\left| \vec{k} \cdot \vec{r} \right| - \omega \cdot t = 0$ gewinnt:

$$|\vec{v}_{ph}| = \frac{|\vec{r}|}{t} = \frac{\omega}{|\vec{k}|} \quad (1.11)$$

Dies ist die Geschwindigkeit, mit der sich ein Punkt konstanter Phase einer monochromatischen Welle mit der Kreisfrequenz ω bewegt. Die Phasengeschwindigkeit kann al-

lerdings in manchen Umständen Werte annehmen, die größer als die Lichtgeschwindigkeit sind (sog. anomale Dispersion). Ein Grundsatz der Physik besagt, dass Teilchen (und auch Information) nicht schneller als mit Lichtgeschwindigkeit bewegt werden können.

Stattdessen ist die Gruppengeschwindigkeit v_g das richtige Maß für z.B. die Ausbreitungsgeschwindigkeit eines Wellenpakets:

$$v_g = \frac{\partial \omega}{\partial k} \quad (1.12)$$

Die Gruppengeschwindigkeit ist der Differentialquotient von Kreisfrequenz ω und Wellenzahl k . Die Quantenmechanik ist eine statistische Theorie, die ihre Aussagen in Form von Wahrscheinlichkeiten trifft (vgl. unten: Unschärferelation). Insofern ist in der Quantentheorie der Ort als Wahrscheinlichkeit anzugeben, und es ist naheliegend, den Ort mit der höchsten Wellenamplitude als den wahrscheinlichsten Aufenthaltsort des Teilchens anzusehen.

Eine monochromatisch ebene Welle hat aber eine *ortsunabhängige* Amplitude und lässt in diesem Bild gar keine Ortsbestimmung zu. Erst die Überlagerung von Wellen unterschiedlicher Wellenlängen führt durch Interferenz zu örtlich unterschiedlichen Amplituden, oder auch Wellengruppen. Beispielsweise kann eine Rechteckwellengruppe durch die Überlagerung von Sinuswellen in Form der unendlichen Reihe

$$\Psi(x) = \frac{4a}{\pi} \sum_{n=0}^{\infty} \left[\frac{\sin(2n+1)x}{2n+1} \right] \quad (1.13)$$

dargestellt werden. Die Gruppengeschwindigkeit v_g ist gerade die Geschwindigkeit, mit der sich eine solche Wellengruppe oder auch Wellenpaket fortbewegt. Wir müssen jetzt noch den Begriff der Ortswahrscheinlichkeit exakt in Form der Wellenfunktion angeben. Da der Ort eine messbare Größe ist, muss die Ortswahrscheinlichkeit reell sein, auch wenn die Wellenfunktion Ψ komplex sein kann (1.10). Das Absolutquadrat der Wellenfunktion $|\Psi|^2$ ist aber immer reell:

$$|\Psi|^2 = \Psi \Psi^* \quad (1.14)$$

Hier ist Ψ^* die zu Ψ konjugiert komplexe Größe, d.h. eine komplexe Zahl mit gleichem

Betrag, jedoch einem Imaginärteil mit umgekehrten Vorzeichen. Für eine allgemeine komplexe Zahl $z = x + jy$ gilt aufgrund von $j^2 = -1$:

$$z^* = (x + jy)^* = x - jy \quad (1.15)$$

und

$$|z|^2 = z \cdot z^* = x^2 + y^2 \quad (1.16)$$

Die Wahrscheinlichkeit dP_Ψ , dass ein durch Ψ beschriebenes Teilchen zur Zeit t am Ort \vec{r} innerhalb eines Volumenelementes dV zu finden, beträgt:

$$dP_\Psi(\vec{r}, t) = |\Psi(\vec{r}, t)|^2 dV = \Psi(\vec{r}, t) \cdot \Psi(\vec{r}, t)^* dV \quad (1.17)$$

Eine Normierung wird dadurch erreicht, dass das Teilchen bei Integration über alle Orte ($V \rightarrow \infty$) mit Sicherheit gefunden wird, und diese völlige Sicherheit der Wahrscheinlichkeit $1 = 100\%$ entspricht:

$$\int_V dP_\Psi(\vec{r}, t) = \int_V \Psi(\vec{r}, t) \cdot \Psi(\vec{r}, t)^* dV = 1 \quad \text{für } V \rightarrow \infty \quad (1.18)$$

1.3 Heisenberg'sche Unschärferelation

In den vorangegangenen Kapiteln haben wir beiden grundlegenden Erscheinungsformen der physikalischen Realität, der Strahlung einerseits und der Materie in Form von Teilchen andererseits, sowohl Teilchen als auch Wellencharakter zuschreiben müssen und dieses als Dualismus bezeichnet. Zur vollständigen Beschreibung mikroskopischer Systeme ist die Kenntnis beider Aspekte notwendig. Die Quantentheorie führt aber auf die Erkenntnis, dass bestimmte Paare von Wellen- bzw. Teilchengrößen nicht gleichzeitig beliebig genau bestimmt werden können. *Heisenberg* fand dies erstmals für die Paare Ort und Impuls bzw. Energie und Zeit. Genauer gilt für die Unbestimmtheiten des Ortes Δx und des Impulses Δp bzw. der Energie ΔW und der Zeit Δt :

$$\begin{aligned} \Delta x \cdot \Delta p &\geq \hbar/2 \\ \Delta W \cdot \Delta t &\geq \hbar/2 \end{aligned} \quad (1.19)$$

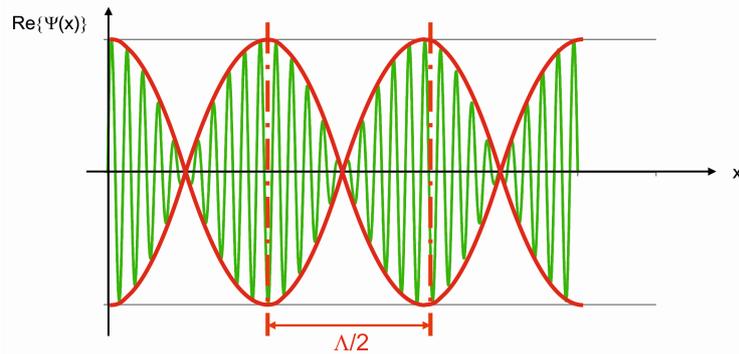


Abbildung 1.5: Schwebung durch Addition zweier Wellenfunktionen

Die Orts-Impuls-Unschärfe in (1.19) lässt sich veranschaulichen, wenn man das betrachtete Teilchen als Überlagerung von zwei Wellen beschreibt, die sich in Frequenz und Wellenlänge bzw. Wellenzahl nur geringfügig unterscheiden.

$$\begin{aligned}\Psi_1 &= C_1 \cdot \exp \{j [kx - \omega t]\} \\ \Psi_2 &= C_2 \cdot \exp \{j [(k + \Delta k) x - (\omega + \Delta \omega) t]\} \\ \Psi &= \Psi_1 + \Psi_2\end{aligned}\tag{1.20}$$

Diese Überlagerung führt zu einer Schwebung, wie sie in Abb. 1.5 dargestellt ist. Die hochfrequente Schwingung wird durch eine Schwebung der Wellenlänge $\Lambda = 4\pi/\Delta k$ moduliert. Ein Wellenpaket hat die Ausdehnung und damit in etwa die Ortsunschärfe

$$\Lambda/2 = 2\pi/\Delta k \leq \Delta x\tag{1.21}$$

Ferner verwenden wir von (1.6) und erhalten für die Impulsunschärfe $\Delta p = \hbar\Delta k$ und somit

$$\Delta p \cdot \Delta x \geq h\tag{1.22}$$

Dieses Ergebnis stimmt nicht exakt mit (1.19) überein, veranschaulicht aber, dass wegen $\Delta p \propto \Delta k$ und $\Delta x \propto 1/\Delta k$ keine genaue Bestimmung der einen Größe die größere Unsicherheit in der Bestimmung der anderen Größe nach sich zieht. Es sei angemerkt, dass die Unschärferelation nicht etwa eine Unzulänglichkeit der Messapparaturen be-

schreibt, die ja durch Verfeinerung der Messtechnik behebbar wären. Sie geben vielmehr die Tatsache wieder, dass ganz fundamental die Beobachtbarkeit entsprechender Größenpaare prinzipiell eingeschränkt ist.

1.4 Schrödinger-Gleichung

Wir wissen nun, wie man Strahlung und Teilchen quantentheoretisch beschreibt. Wir benötigen aber noch eine mathematische Formulierung, um den zeitlichen Ablauf mikroskopischer Prozesse zu beschreiben. Besonders interessiert uns in der Festkörperelektronik das Verhalten von Elektronen. Die gesuchte Formulierung fand *Schrödinger* im Jahr 1925 in Form der folgenden nach ihm benannten Gleichung:

$$\frac{\hbar^2}{2m} \Delta \Psi - V \cdot \Psi = -j\hbar \frac{\partial \Psi}{\partial t} \quad (1.23)$$

Dabei ist Δ der Laplace-Operator

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \quad (1.24)$$

und V die potentielle Energie (Potential). $V = V(\vec{r}, t)$ und $\Psi = \Psi(\vec{r}, t)$ sind im allgemeinen Funktionen des Ortes und der Zeit. Diese Gleichung kann nicht exakt hergeleitet, sondern höchstens durch Analogieschlüsse plausibel gemacht werden. Letztlich aber ist allein die Tatsache, dass sie bisher alle quantenmechanischen Vorgänge richtig beschreibt, der entscheidende Beleg für ihre Gültigkeit.

Bei der folgenden Beschäftigung mit der Schrödinger-Gleichung führen wir folgende Vereinfachungen ein:

1. $V = f(x)$, d.h. eindimensionale Berechnung
2. $V \neq f(t)$, d.h. V ist zeitlich konstant

Die zweite Bedingung trifft recht gut auf den Festkörper zu, in dem sich Elektronen im zeitlich nur relativ langsam verändernden Potential der Atomrümpfe bewegen. Die Masse der Atomrümpfe übersteigt die Masse der Elektronen um Größenordnungen. Ihre charakteristischen Schwingungsfrequenzen (Gitterschwingungen, thermische Anregung) sind dadurch weit von den charakteristischen Frequenzen der Elektronen entfernt. Mit den Abkürzungen $\Psi' = \partial \Psi / \partial x$ bzw. $\dot{\Psi} = \partial \Psi / \partial t$ lautet die eindimensionale

Schrödinger-Gleichung:

$$\frac{\hbar^2}{2m} \Psi'' - V\Psi = -j\hbar\dot{\Psi} \quad (1.25)$$

Zur Prüfung, wie die Beschreibung der Elektronen durch diese Gleichung mit den bisher erreichten Ergebnissen übereinstimmt, werden nun einige mathematische Umformungen zur Lösung dieser Gleichung vorgenommen. Trennung der Variablen:

$$\Psi(x, t) = \Theta(x) \cdot \phi(t) \quad (1.26)$$

Einsetzen in(1.25) ergibt

$$\frac{\hbar^2}{2m} \Theta'' \cdot \phi - V \cdot \Theta \cdot \phi = -j\hbar\Theta\dot{\phi} \quad (1.27)$$

Division durch $\Theta \cdot \phi$ ergibt

$$\frac{\hbar^2}{2m} \frac{\Theta''}{\Theta} - V = -j\hbar \frac{\dot{\phi}}{\phi} \quad (1.28)$$

Die linke Seite dieser Gleichung ist nur noch von x , die rechte nur noch von t abhängig. Die Gleichheit kann nur dann gelten, wenn beide Seiten konstant sind; die Konstante wird $-W$ genannt, sie hat (wie V) die Dimension einer Energie.

rechte Seite

$$-j\hbar \frac{\dot{\phi}}{\phi} = -W \quad (1.29)$$

Hier gibt es die unabhängige Variable t und die von ihr abhängige Funktion ϕ . Die Ableitung nach t kann daher als totales Differential geschrieben werden

$$-j\hbar \frac{1}{\phi} \frac{d\phi}{dt} = -W \quad (1.30)$$

Anwendung der Differenzialrechnung

$$\frac{d\phi}{\phi} = -j \frac{W}{\hbar} dt \quad (1.31)$$

Einfache Integration ergibt

$$\begin{aligned} \ln \phi &= +A \\ \Rightarrow \phi &= \exp\left(-j \frac{W}{\hbar} \cdot t\right) \end{aligned} \quad (1.32)$$

Hier wurde die Integrationskonstante $A = 0$ gesetzt, da sie nur eine beliebige Phase darstellt. Mit der Gesamtenergie $W = h \cdot f$ des Teilchens ergibt sich ϕ als der zeitlich periodische Faktor der De Broglie Welle (1.10)

$$\phi = \exp(-j\omega t) \quad (1.33)$$

linke Seite Nach Multiplikation mit Θ gilt:

$$\frac{\hbar^2}{2m} \Theta'' + (W - V)\Theta = 0 \quad (1.34)$$

Dies ist die zeitunabhängige Schrödinger-Gleichung, die somit stationäre, d. h. zeitlich nicht veränderliche Zustände beschreibt. Sie kann auch angewendet werden, wenn quasistationäre Zustände vorliegen, beispielsweise Elektronen im Halbleiter, die durch ein veränderliches äußeres elektrisches Feld beeinflusst werden, wie z.B. in einem Transistor.

Zur quantitativen Auswertung, beispielsweise bei der Bestimmung der Gesamtenergie W , muss die Normierung (1.18) berücksichtigt werden:

$$\int_{V \rightarrow \infty} \Psi \Psi^* dV = 1 \quad \Rightarrow \quad \int_{-\infty}^{+\infty} \Theta \Theta^* dx = 1 \quad (1.35)$$

Der Beweis für die zweite Gleichung ist leicht durchzuführen. Damit liegt uns jetzt ein Werkzeug vor, mit dem wichtige Vorgänge in Atomen und deren Eigenschaften sowie der daraus aufgebauten Festkörpern erklärt werden können.

1.5 Freie Teilchen

Elektronen sind nicht immer an Atome gebunden, sondern können im Vakuum frei beweglich sein, d. h. sie erfahren keine Kräfte

$$\vec{F} = -\text{grad}V = -\nabla V = 0 \quad (1.36)$$

Hier wird, wie üblich, die Kraft als der Gradient des Potentials erhalten. Aus (1.36) folgt, dass in diesem Fall das Potential konstant ist; der einfachste Fall ist $V = 0$. Bei konstantem $V = 0$ vereinfacht sich die eindimensionale Schrödingergleichung zu

$$\Theta'' = -\frac{2mW}{\hbar^2} \cdot \Theta \quad (1.37)$$

und die Gesamtenergie W besteht nur aus der kinetischen Energie W_{kin} , da die potentielle Energie $V=0$ ist

$$W = W_{kin} = \frac{p^2}{2m} = \frac{\hbar^2 k^2}{2m} \quad (1.38)$$

Mit der Abkürzung $k^2 = 2mW/\hbar^2$ ist die eindimensionale Schrödingergleichung ohne äußeres Potential

$$\Theta'' = -k^2 \Theta \quad (1.39)$$

Diese Wellengleichung wird gelöst durch den Ansatz

$$\Theta(x) = Ae^{jkx} + Be^{-jkx} \quad (1.40)$$

was durch Einsetzen leicht nachgewiesen werden kann. Die zeitabhängige Wellenfunktion (1.25) lautet dann

$$\Psi(x, t) = \Theta(x) \cdot \phi(t) = Ae^{j(kx-\omega t)} + Be^{-j(kx+\omega t)} \quad (1.41)$$

Der erste Term beschreibt eine in positive x -Richtung laufende Welle, der zweite eine in negative x -Richtung. Wenn nur einer der beiden Koeffizienten A oder B ungleich 0 ist, so findet die Bewegung des Teilchens eindeutig in die entsprechende Richtung statt. Man beachte, dass in diesem Fall eine ebene monochromatische Welle vorliegt, die

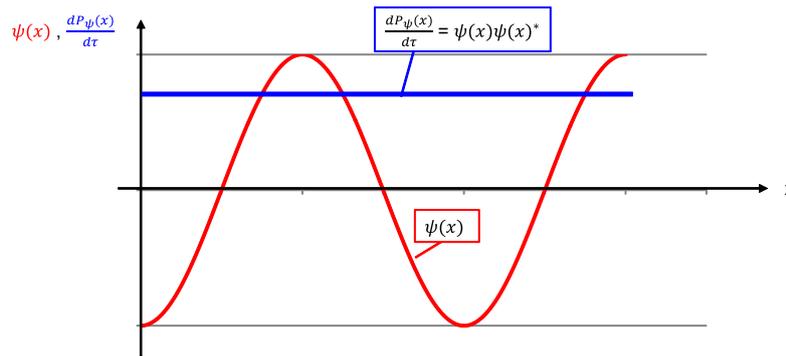


Abbildung 1.6: Aufenthaltswahrscheinlichkeit und Wellenfunktion eines freien Teilchens

eine uniforme Teilchenaufenthaltswahrscheinlichkeit im ganzen Raum (hier: auf der gesamten x -Achse) ergibt.

Selbst für $A \neq 0$ und $B \neq 0$, d.h. bei Ausbildung einer stehenden Welle, ist der Abstand der ortsfesten Maxima oder Minima so gering, dass ein "Wellenbauch" nicht als Wellengruppe im Sinne einer Teilchenaufenthaltswahrscheinlichkeit interpretiert werden kann. Dies würde auf einen Widerspruch zur Heisenberg'schen Unschärferelation (1.19) führen. Diese Wellenfunktion ist auch nicht normierbar (Abb. 2.6)

1.5.1 Teilchen an einer Potentialstufe

Wir wollen nun einfache Modelle betrachten, in denen ein Elektron äußeren Beeinflussungen unterworfen wird. Die Näherung, zunächst nur ein Elektron zu betrachten, stellt sich später auch als zielführend für den Festkörper heraus. Die Wechselwirkung der Elektronen untereinander ist viel geringer, als die Wechselwirkung mit dem periodischen elektrischen Potential aufgrund der Atomrümpfe im Festkörper. Zuerst liegen die Potentialverhältnisse nach Abb. 1.7 vor. Dieses einfache Modell gibt recht gut die Verhältnisse für die folgenden Systeme wieder:

- Elektron an einer Festkörper-Vakuum Grenzfläche (vgl. Photoeffekt),
- Elektron an einem pn-Übergang und
- Elektron an einem Metall-Halbleiterkontakt

Die Elektronen treffen, von links kommend, auf die Potentialstufe und werden – entsprechend ihrer Energie – die Stufe überwinden können oder reflektiert werden. Klassisch

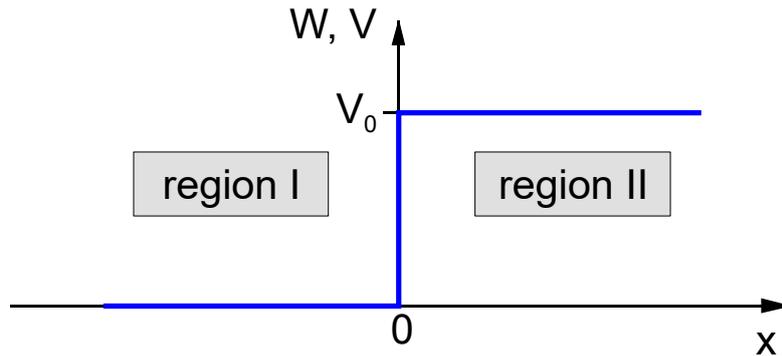


Abbildung 1.7: einseitige Potentialstufe

erwartet man für

$$W_{kin} = \begin{cases} \leq V_0 & \text{vollständige Reflexion} \\ > V_0 & \text{vollständige Transmission} \end{cases} \quad (1.42)$$

Was liefert nun die Schrödinger-Gleichung? Die Wellenfunktion des Elektrons soll das gesamte Problem lösen. Allerdings können in den beiden Bereichen I und II zunächst getrennte Ansätze verfolgt werden, die dann zu einer vollständigen Lösung zusammengesetzt werden. Aus der Abb. 1.7 lesen wir die Werte für das konstante Potential in den beiden Bereichen I und II ab:

$$V = \begin{cases} V_I = 0 & \text{für } x \leq 0 \\ V_{II} = V_0 & \text{für } x > 0 \end{cases} \quad (1.43)$$

Für die Gesamtenergie $W = W_{kin} + V$ gilt in den beiden Bereichen

$$W = \begin{cases} W^I = W_{kin}^I & \text{für } x \leq 0 \\ W^{II} = W_{kin}^{II} + V_0 & \text{für } x > 0 \end{cases} \quad (1.44)$$

Vorgehen:

1. Ψ wird in den Bereichen I und II jeweils getrennt berechnet
2. die Lösungen werden bei $x = 0$ aneinander angepasst, indem die Forderung nach **Stetigkeit** und **Differenzierbarkeit** ausgenutzt wird.

Berechnung:

$$\Psi_I'' = -k_I^2 \Psi_I \quad i = \text{I, II} \quad (1.45)$$

mit

$$k_I = \sqrt{\frac{2m(W - V_I)}{\hbar^2}} \quad i = \text{I, II} \quad (1.46)$$

Die Ansätze entsprechen (1.40), allerdings können die Koeffizienten A_I und B_I in den beiden Bereichen verschiedene Werte annehmen

$$\Psi_I(x) = A_I e^{jk_I x} + B_I e^{-jk_I x} \quad i = \text{I, II} \quad (1.47)$$

Nachdem die k_I bekannt sind, reduziert sich das Problem auf die Bestimmung der vier Konstanten $A_{\text{I,II}}$ und $B_{\text{I,II}}$. Da im Bereich II keine weitere Reflexionsstelle existiert, ist hier nur eine in positive x -Richtung laufende Welle sinnvoll, d.h. $B_{\text{II}} = 0$.

Ferner interessiert im Bereich I nur das Verhältnis von hin- zu rücklaufender Teilchenintensität, d.h. der Reflexions- und Transmissionskoeffizient; damit reduziert sich das Problem weiter auf die Bestimmung von nur zwei Unbekannten. Die dazu notwendigen zwei Gleichungen liefern folgende Forderungen

Stetigkeit:

$$\Psi_I(x=0) = \Psi_{\text{II}}(x=0) \quad (1.48)$$

Differenzierbarkeit:

$$\Psi_I'(x=0) = \Psi_{\text{II}}'(x=0) \quad (1.49)$$

Einsetzen ergibt:

$$\begin{aligned} A_I &= \frac{A_{\text{II}}}{2} \left(1 + \frac{k_{\text{II}}}{k_I} \right) \\ B_I &= \frac{A_{\text{II}}}{2} \left(1 - \frac{k_{\text{II}}}{k_I} \right) \end{aligned} \quad (1.50)$$

und somit können Ψ_I und Ψ_{II} allein in A_{II} , k_I und k_{II} ausgedrückt werden:

$$\begin{aligned}\Psi_I &= \frac{A_{II}}{2} \left[\left(1 + \frac{k_{II}}{k_I}\right) e^{jk_I x} + \left(1 - \frac{k_{II}}{k_I}\right) e^{-jk_I x} \right] \\ \Psi_{II} &= A_{II} \cdot e^{jk_{II} x}\end{aligned}\tag{1.51}$$

Wie bereits erwähnt, interessieren nur Reflexions- und Transmissionskoeffizient R bzw. T , die den relativen Teilchenstrom in beide Richtungen angeben, bezogen auf den Gesamtstrom einfallender Teilchen

$$R = \frac{\text{reflektierter Teilchenstrom}}{\text{einfallender Teilchenstrom}}\tag{1.52}$$

$$T = \frac{\text{transmittierter Teilchenstrom}}{\text{einfallender Teilchenstrom}}\tag{1.53}$$

Die Teilchenstromdichte \vec{S} beschreibt den Fluss einer Anzahl Teilchen durch eine Referenzebene, in Einheiten $\text{cm}^{-2}\text{s}^{-1}$:

$$\vec{S} = n \cdot \vec{v}\tag{1.54}$$

mit der Teilchenkonzentration n mit Einheit $[n] = \text{cm}^{-3}$ und Teilchengeschwindigkeit v mit Einheit $[v] = \text{cm s}^{-1}$. Die Konzentration n kann über die quantentheoretische Aufenthaltswahrscheinlichkeit für ein Teilchen nach (1.17) angegeben werden:

$$N \cdot \Psi\Psi^* = n\tag{1.55}$$

wobei die gesamte Teilchenzahl N verwendet wurde. Da die Gesamtzahl der Teilchen unverändert bleibt, kann man durch N teilen und die Rechnung auf den Strombeitrag eines einzelnen Teilchens beziehen. Dies führt auf die normierte Teilchenstromdichte $\vec{s} = \vec{S}/N$

$$\vec{s} = \Psi\Psi^*\vec{v}\tag{1.56}$$

Unter Verwendung der Ausdrücke für Ψ_I und Ψ_{II} sowie unter Ausnutzung von (1.56)

folgt:

$$R = \frac{\left| \overleftarrow{s}_{\text{I}} \right|}{\left| \overrightarrow{s}_{\text{I}} \right|} = \frac{B_{\text{I}} B_{\text{I}}^* \left| \overleftarrow{v}_{\text{I}} \right|}{A_{\text{I}} A_{\text{I}}^* \left| \overrightarrow{v}_{\text{I}} \right|} \quad (1.57)$$

$$T = \frac{\left| \overrightarrow{s}_{\text{II}} \right|}{\left| \overrightarrow{s}_{\text{I}} \right|} = \frac{A_{\text{II}} A_{\text{II}}^* \left| \overrightarrow{v}_{\text{II}} \right|}{A_{\text{I}} A_{\text{I}}^* \left| \overrightarrow{v}_{\text{I}} \right|} \quad (1.58)$$

Es folgt nun die Unterscheidung für verschiedene Energie des einfallenden Teilchens in die Fälle $W > V_0$ bzw. $W < V_0$, d.h. Teilchenenergie höher oder geringer als die Teilchenstufe. Im klassischen Fall würde man vollständige Transmission bzw. vollständige Reflexion erwarten.

Fall 1: $W < V_0$ Hier ist k_{II} rein imaginär, und man führt als neue Größe k'_{II} ein, für die gilt:

$$jk'_{\text{II}} = k_{\text{II}} \quad (1.59)$$

mit

$$k'_{\text{II}} = \sqrt{\frac{2m(V_0 - W)}{\hbar^2}} \quad (1.60)$$

Setzt man ferner voraus, dass die Reflexion elastisch erfolgt, d.h. es gilt

$$\left| \overrightarrow{v}_{\text{I}} \right| = \left| \overleftarrow{v}_{\text{I}} \right|$$

und mit dem allgemeinen Zusammenhang

$$\overrightarrow{v}_{\text{I}} = \frac{\hbar}{m} \overrightarrow{k}_{\text{I}}$$

erhält man

$$R = \frac{B_{\text{I}} B_{\text{I}}^*}{A_{\text{I}} A_{\text{I}}^*} = 1 \quad (1.61)$$

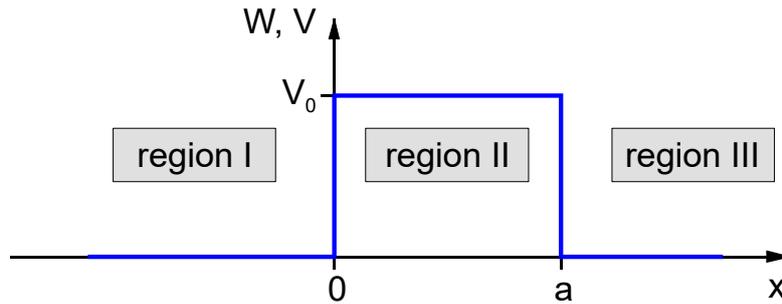


Abbildung 1.8: endliche Potentialbarriere

was durch Verwendung von (1.50) leicht belegbar ist und auch durch das Experiment bestätigt wird, in Übereinstimmung mit der klassischen Physik.

Anzumerken ist, dass der Ausdruck $e^{-jk_{II}x}$ rein reell wird und somit eine mit x exponentiell abklingende Funktion im Bereich der Potentialstufe darstellt. Bei der Reflexion dringt die Welle also ein Stück in die Stufe ein, wird dann aber doch vollständig reflektiert.

Fall 2: $W > V_0$ k_{II} ist reell und (1.51) können in dieser Form verwendet werden. Für die Geschwindigkeiten verwendet man hier:

$$v_i = \hbar/mk_i \quad i = \text{I, II} \quad (1.62)$$

und erhält für den Transmissionskoeffizienten T mit (1.53) und (1.58)

$$T = \frac{4 \frac{k_{II}}{k_I}}{\left(1 + \frac{k_{II}}{k_I}\right)^2} \quad (1.63)$$

(2.49) Nur für $W \gg V_0$, d. h. $k_I \approx k_{II}$ folgt $T = 1$ in Übereinstimmung mit der klassischen Physik. Für $W \gtrsim V_0$ gilt $T < 1$ und damit auch $R > 0$ im Gegensatz zur klassischen Physik, aber in völliger Übereinstimmung mit Experimenten der Atom, Kern- und Festkörperphysik.

1.5.2 Teilchen an einer Potentialbarriere

Jetzt wollen wir eine endlich ausgedehnte Barriere untersuchen, die die Breite a und die Höhe V_0 besitzt. Das Potential ist wieder als Rechteck vorgegeben. Aus der Abb. 1.8

lesen wir die Werte für das Potential in den drei Regionen ab.

$$V = \begin{cases} V_{\text{I}} = 0 & \text{für } x \leq 0 \\ V_{\text{II}} = V_0 & \text{für } 0 < x \leq a \\ V_{\text{III}} = 0 & \text{für } x > a \end{cases} \quad (1.64)$$

Für die Gesamtenergie $W = W_{\text{kin}} + V$ gilt in den drei Bereichen

$$W = \begin{cases} W^{\text{I}} = W_{\text{kin}}^{\text{I}} & \text{für } x \leq 0 \\ W^{\text{II}} = W_{\text{kin}}^{\text{II}} + V_0 & \text{für } 0 < x \leq a \\ W^{\text{III}} = W_{\text{kin}}^{\text{III}} & \text{für } x > a \end{cases} \quad (1.65)$$

Entsprechend den drei Bereichen existieren Wellenfunktionen Ψ_i jeweils mit den Koeffizienten A_i, B_i ($i = \{\text{I}, \text{II}, \text{III}\}$). Analog zur Behandlung der unendlich ausgedehnten Potentialstufe führt man für $W < V_0$ wieder $k_{\text{II}} = jk'_{\text{II}}$ ein und verwendet die Bedingungen Stetigkeit und Differenzierbarkeit der Wellenfunktionen, die jetzt aber sowohl bei $x = 0$ als auch bei $x = a$ zu fordern sind. Ferner ist man wieder nur an den Verhältnissen der Teilchenströme interessiert und nicht an den Strömen selbst. Mit dem gleichen Argument wie im vorangegangenen Kapitel existiert im Bereich III keine rücklaufende Welle, so dass $B_{\text{III}} = 0$.

Da zwischen dem Reflexions- und Transmissionskoeffizienten aufgrund der Teilchenerhaltung immer die Beziehung gilt

$$R + T = 1 \quad (1.66)$$

genügt es, T zu bestimmen. Im vorliegenden Fall gilt:

$$k_{\text{III}} = k_{\text{I}} \quad (1.67)$$

und die Rechnung ergibt:

$$T = \frac{A_{\text{III}} A_{\text{III}}^*}{A_{\text{I}} A_{\text{I}}^*} \cdot \frac{|\vec{v}_{\text{III}}|}{|\vec{v}_{\text{I}}|} = \left[1 + \frac{\sinh^2(k_{\text{II}} a)}{4 \frac{W}{V_0} \left(1 - \frac{W}{V_0}\right)} \right]^{-1} \quad \text{für } W < V_0 \quad (1.68)$$

und

$$T = \frac{A_{\text{III}} A_{\text{III}}^*}{A_{\text{I}} A_{\text{I}}^*} \cdot \frac{|\vec{v}_{\text{III}}|}{|\vec{v}_{\text{I}}|} = \left[1 + \frac{\sin^2(k_{\text{II}} a)}{4 \frac{W}{V_0} \left(\frac{W}{V_0} - 1 \right)} \right]^{-1} \quad \text{für } W > V_0 \quad (1.69)$$

Man sieht, dass im Gegensatz zur klassischen Physik für

1. $W < V_0$ nicht $T = 0$, sondern $T > 0$ gilt und für
2. $W > V_0$ nicht $T = 1$, sondern $T < 1$.

Der 1. Fall beschreibt den quantenmechanischen Tunneleffekt, der kein klassisches Analogon hat, aber in Mikrosystemen beobachtet wird. Abb.1.9 zeigt den Transmissionskoeffizienten in Abhängigkeit vom Verhältnis W/V_0 für zwei Potentialbarrieren, die sich sowohl hinsichtlich der Barrierenhöhe V_0 als auch der Breite a unterscheiden. Im Fall (a) liegen Verhältnisse vergleichbar mit Atompotentialen im Festkörper vor, und im Fall (b) ähnliche Barrieren wie in Metall-Halbleiter- und pn -Übergängen.

Im Wellenbild lässt sich das Tunneln der Anschauung zugänglich machen: schon bei der Behandlung der Potentialstufe war mit rein imaginärem k_{II} die mit x exponentiell abklingende Funktion $e^{-jk'x}$ aufgetreten. Dies wurde als Eindringen der Welle – und damit des Teilchens – in die Stufe gedeutet. Die Tunnelwahrscheinlichkeit entspricht der Amplitude der abklingenden Welle am “Austrittsort”, also bei $x = a$. Als experimentelle Belege für den Tunneleffekt seien zwei Beispiele genannt:

1. Radioaktive Elemente, z.B. α -Strahler: α -Teilchen sind Heliumatomkerne und somit positiv geladen. Bestandteile des Atomkerns müssen für kleine Abstände vom Kernzentrum eine anziehende Kraft erfahren, sonst würde der Kern nicht stabil sein, sondern die einzelnen Bestandteile würden auseinanderfliegen. Andererseits wirkt die für gleich geladene Teilchen (Protonen) eine abstoßende Coulombkraft, die bei größerer Entfernung vom Kernzentrum überwiegt. Daraus ergibt sich der in Abb.1.10 dargestellte Potentialverlauf.

Die Beobachtung liefert α -Teilchen außerhalb des Atomkerns (z. B. U-238) mit einer Energie W_0 , die erheblich kleiner ist als $V_m a x$, die zur Überwindung der Potentialbarriere notwendig wäre. Die einzige Erklärung dieser Beobachtung liefert der eben behandelte Tunneleffekt.

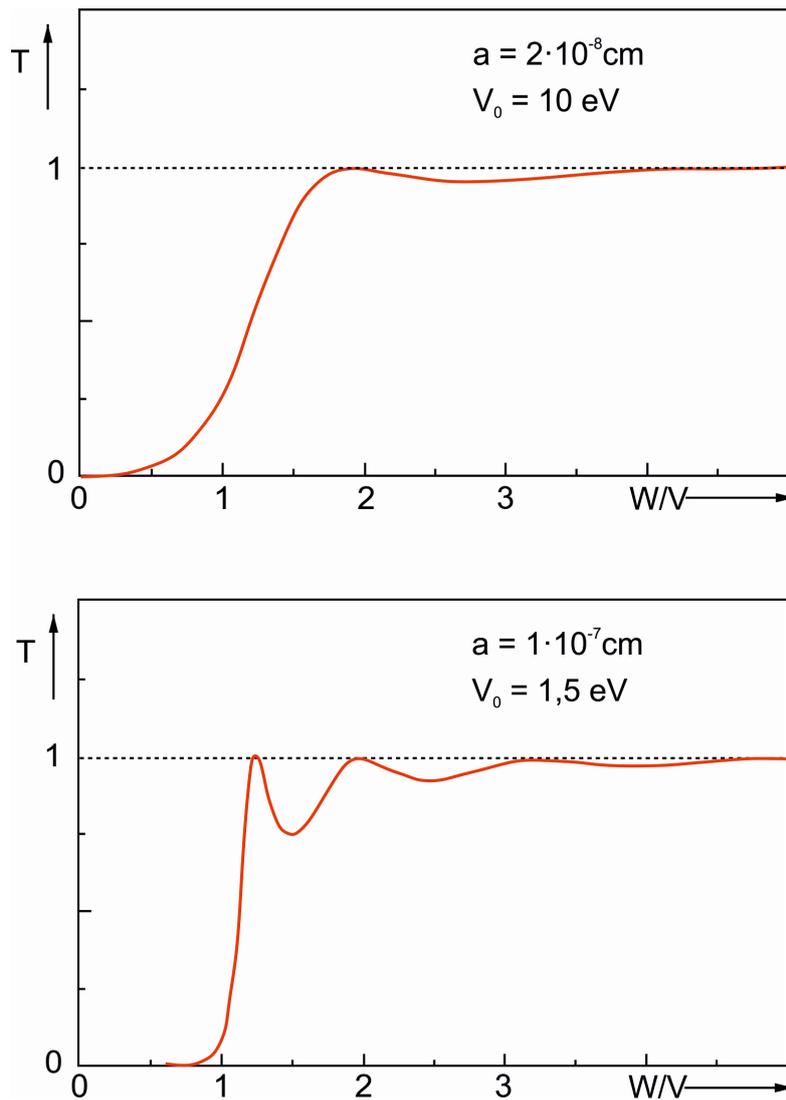


Abbildung 1.9: Transmissionskoeffizient T als Funktion der Teilchenenergie für unterschiedliche Breite und Höhe der Potentialbarriere

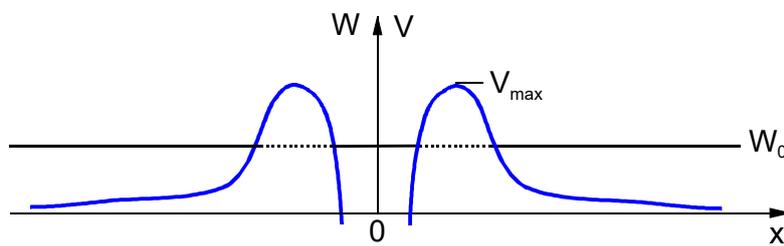


Abbildung 1.10: Potentialverlauf in Atomkernnähe

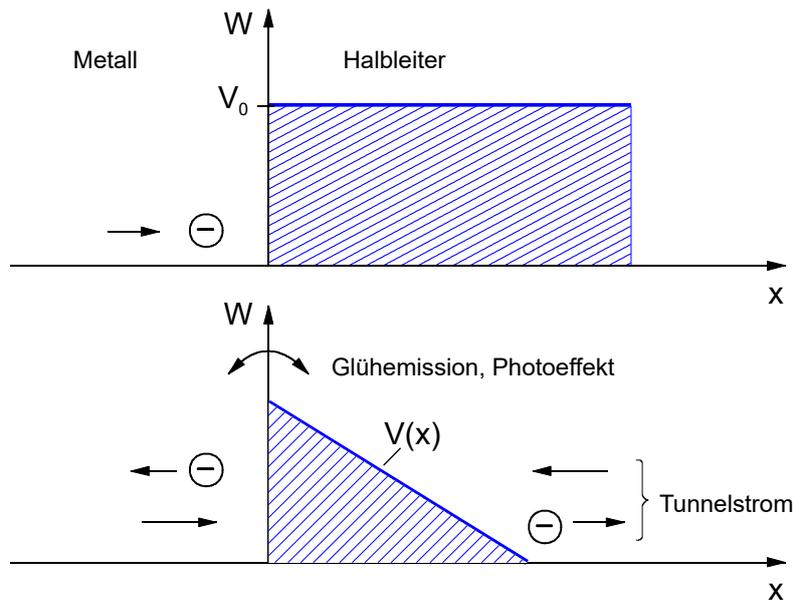


Abbildung 1.11: Tunneleffekt am Metall-Halbleiterkontakt

2. Metall-Halbleiter (Schottky-Kontakt): Wir wissen bereits, dass zum Befreien von Elektronen aus einer Metalloberfläche Energie notwendig ist, die z. B. beim Photoeffekt (vgl. Kap. 2.1) in Form von Photonen bereitgestellt werden kann. Die Elektronen finden an der Metall-Vakuum Grenzfläche eine Potentialstufe vor, die sie durch Energiezufuhr überwinden können. Ähnliche Verhältnisse liegen an der Grenzfläche Halbleiter-Metall vor (Abb. 2.11a).

Wird nun eine Spannung zwischen Metall und Halbleiter angelegt, so verändert sich der Potentialverlauf gemäß Abb. 2.11.b (genauer wird gerade dieser Fall später behandelt, vgl. Kap. 6.2). Natürlich kann man durch Energiezufuhr auch jetzt Elektronen zum Überwinden der jetzt entstandenen dreieckförmigen Potentialbarriere verhelfen (Glühemission \rightarrow thermische Energie, Photoeffekt \rightarrow Lichtenergie). Man beobachtet aber auch bereits ohne Bestrahlung oder Erhitzung einen merklichen elektrischen Strom, der je nach Polung in beide Richtungen fließen kann. Dieser kann ebenfalls nur durch den Tunneleffekt erklärt werden. Sperrfreie Metall-Halbleiterkontakte, wie sie für jedes Halbleiterbauelement benötigt werden, beruhen im allgemeinen auf dem quantenmechanischen Tunneleffekt!

In diesem Kapitel wurden Systeme behandelt, in denen sich Teilchen, insbesondere Elektronen, frei bewegen können, jedenfalls in ausgedehnten Bereichen, die allerdings

durch Potentialstufen bzw. -barrieren begrenzt sind. Ähnliche Bedingungen liegen für Elektronen in Festkörpern wie Metallen und Halbleitern vor: die Potentialbarrieren und -stufen bilden Grenzflächen zu anderen Materialien oder dem Vakuum. Allerdings ist die Bewegung der Elektronen in Festkörpern nicht völlig frei, sie wird vielmehr auch von atomaren Bindungskräften beeinflusst. Es gibt insofern im Festkörper "quasi"-freie Elektronen, die für den Stromfluss verantwortlich sind, aber auch gebundene Elektronen, die andere Festkörpereigenschaften wie dielektrisches und magnetisches Verhalten beeinflussen. Den gebundenen Elektronen wollen wir uns im folgenden widmen.

1.6 Gebundene Teilchen

Gebundene Teilchen sind solche, die einen eng begrenzten Bereich nicht verlassen können. Das typische Beispiel ist ein Elektron, das an den zugehörigen Atomkern gebunden ist. Die Bindung wird durch ein entsprechendes Potential bewerkstelligt. Im System Atomkern-Elektron bewirkt das Coulomb-Potential die Bindung. Im nächsten Unterkapitel werden wir dafür ein erstes Modell kennenlernen.

1.6.1 Elektron im Potentialtopf

Im einfachsten Fall können wir Potentialverhältnisse nach Abb. 1.12 annehmen, mit unendlich hohen Wänden. Welche Eigenschaften können wir für dieses System mit Hilfe der Schrödinger-Gleichung ausfindig machen? Es gilt mit (1.39) für die zeitunabhängige Wellenfunktion Θ

$$\begin{aligned} \Theta''(x) &= -k^2\Theta(x) \quad \text{für} \quad -\frac{a}{2} \leq x \leq \frac{a}{2} \\ \Theta(x) &= 0 \quad \text{für} \quad x < -\frac{a}{2} \quad \text{oder} \quad x > \frac{a}{2} \end{aligned} \quad (1.70)$$

und für das Potential V

$$V = \begin{cases} 0 & \text{für} \quad -\frac{a}{2} \leq x \leq \frac{a}{2} \\ \infty & \text{für} \quad x < -\frac{a}{2} \quad \text{oder} \quad x > \frac{a}{2} \end{cases} \quad (1.71)$$

Das Elektron ist also im Potentialtopf eingesperrt, denn Θ ist außerhalb des Topfs gleich null, und damit verschwindet die Aufenthaltswahrscheinlichkeit $\Theta\Theta^*$ außerhalb des

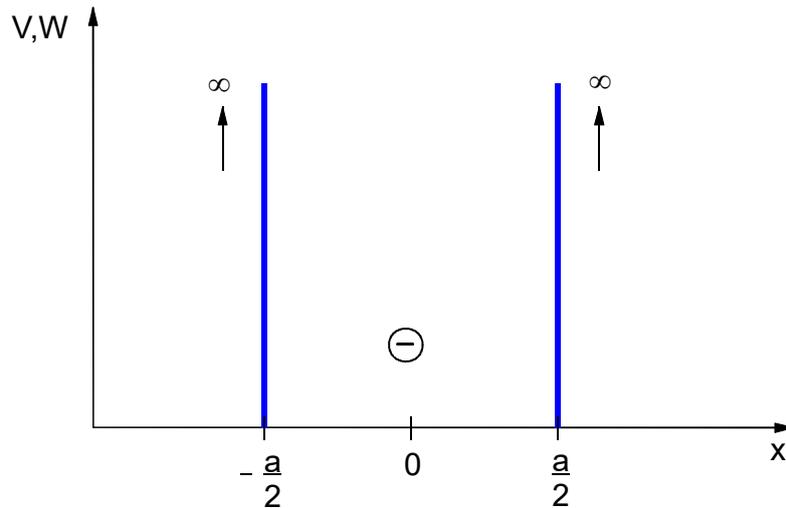


Abbildung 1.12: Elektron im Potentialtopf mit unendlich hohen Wänden

Potentialtopfs. Exakter lautet die Randbedingung:

$$\theta\left(\pm\frac{a}{2}\right) = 0$$

Wie bereits bekannt, wird die eindimensionale Schrödinger-Gleichung durch folgende Wellenfunktion gelöst

$$\Theta(x) = Ae^{j k x} + Be^{(-j k x)} \quad \text{mit} \quad k = \sqrt{\frac{2mW}{\hbar^2}}$$

Wir nutzen auch hier diesen Ansatz. Analog zur Beschreibung der Potentialstufe, bei der die Lösung als Summe von in positiver bzw. negativer x-Richtung laufender Welle zusammengesetzt war, kann man für den jetzt zu behandelnden Potentialtopf die Elektronenbewegung als Hin- und Herbewegung mit elastischer Reflexion an den unendlich hohen Potentialwänden auffassen. Bei den Stößen soll keine Energie verloren gehen (elastisch), daher sind die Amplituden der vor- und rücklaufenden Wellen gleich groß:

$$|A| = |B| \tag{1.72}$$

Somit gibt es zwei Lösungsfunktionen

$$\begin{aligned}\Theta_g &= A \left(e^{jk_g x} + e^{-jk_g x} \right) = 2A \cos(k_g x) \\ \Theta_u &= A \left(e^{jk_u x} - e^{-jk_u x} \right) = 2jA \sin(k_u x)\end{aligned}\tag{1.73}$$

Beide Lösungen müssen die Randbedingung erfüllen. Für die gerade Funktion Θ_g ($\cos(x) = \cos(-x)$ gerade, weil spiegelsymmetrisch in x) gilt am Rand des Potentialtopfs bei $x = \pm a/2$

$$\begin{aligned}\Theta_g(\pm a/2) &= 0 = 2A \cos(k_g x) \\ \Rightarrow k_g \cdot \frac{a}{2} &= g \cdot \frac{\pi}{2} \\ \Rightarrow k_g &= g \cdot \frac{\pi}{a} \\ \text{mit } g &= \{1, 3, 5, \dots\}\end{aligned}\tag{1.74}$$

Für die ungerade Funktion Θ_u ($\sin(x) = -\sin(-x)$ ungerade, weil rotationssymmetrisch um den Ursprung) gilt am Rand des Potentialtopfs bei $x = \pm a/2$

$$\begin{aligned}\Theta_u(\pm a/2) &= 0 = 2A \sin(k_u x) \\ \Rightarrow k_u \cdot \frac{a}{2} &= u \cdot \frac{\pi}{2} \\ \Rightarrow k_u &= u \cdot \frac{\pi}{a} \\ \text{mit } u &= \{2, 4, 6, \dots\}\end{aligned}\tag{1.75}$$

Die gesamte Menge der Lösungsfunktionen setzt sich also zusammen aus geraden (\cos) und ungeraden (\sin) Funktionen, deren Wellenzahlen k_n ausschließlich diskrete Werte annehmen, die das Produkt aus einer positiv ganzen Zahl n und der halben Breite $a/2$ des Potentialtopfs sind.

$$\Theta = \begin{cases} 2A \cos\left(n \cdot \frac{\pi}{a} x\right) & \text{für } n = \{1, 3, 5, \dots\} \\ 2jA \sin\left(n \cdot \frac{\pi}{a} x\right) & \text{für } n = \{2, 4, 6, \dots\} \end{cases}\tag{1.76}$$

Wir bezeichnen den Parameter n als **Quantenzahl**. Die Wellenzahlen k_n sind **quantisiert**. Aus den Wellenzahlen k_n können mit der Beziehung $W = \hbar^2 k^2 / 2m$ die Energiewerte, d.h. die **Quantenenergien** W_n oder auch **Energie-Eigenwerte** der Zustände

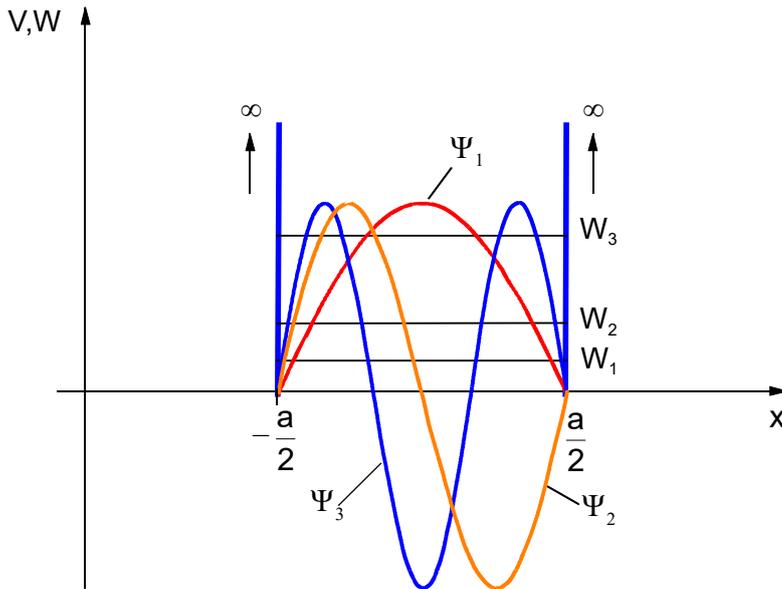


Abbildung 1.13: Energieeigenwerte für ein Elektron im Potentialtopf und zugehörige Wellenfunktionen

berechnet werden

$$W_n = \frac{h^2}{8ma^2} \cdot n^2 \quad \text{mit } n = \{1, 2, 3, 4, \dots\} \quad (1.77)$$

Die Gleichung (1.77) stellt das Eigenwertspektrum der Schrödinger-Gleichung für das System "Elektron im Potentialtopf mit unendlich hohen Wänden" dar. Man erkennt, dass das Elektron, ähnlich der Quantenhypothese von M. Planck (vgl. Abb. (1.1)), nur bestimmte "diskrete" Energiewerte W_n annehmen kann. Im Gegensatz dazu kann das freie Elektron beliebige Energiewerte annehmen.

Abb. 1.13 stellt die ersten erlaubten Energiewerte im Potentialtopf dar, zusammen mit den zugehörigen Wellenfunktionen Ψ_n . Dieses System entspricht völlig dem mechanischen System einer bei $\pm a/2$ eingespannten schwingenden Saite, die auch nur zu diskreten Schwingungen fähig ist, nämlich denen, für die nur Vielfache einer halben Wellenlänge zwischen die Einspannpunkte passen. Mit Hilfe der Normierungsvorschrift (1.18) können die Koeffizienten der Wellenfunktionen quantitativ bestimmt werden zu

$$A = B = \pm \sqrt{\frac{2}{a}} \quad (1.78)$$

In Abb.1.14 ist der Zusammenhang zwischen der Energie W_n und dem Wellenzahlvektor k aufgetragen. Die erlaubten Werte sind als Punkte gekennzeichnet. Es fällt auf, dass sie auf der Parabel liegen, die die kontinuierlichen Energiewerte eines freien Teilchens gleicher Masse beschreibt. Durch die Einschränkung der Bewegung des Elektrons durch den Potentialtopf werden offensichtlich aus dem kontinuierlichen Spektrum des freien Elektrons die Werte des diskreten Spektrums des gebundenen Elektron als nunmehr einzig mögliche ausgewählt. Es sei zusätzlich angemerkt, dass der Wert $W = 0$ durch die Unschärferelation verboten ist und somit nicht zum Eigenwertspektrum gehört.

Die endliche, minimal erlaubte Energie wird auch Nullpunktenergie (Nullpunktschwingung) genannt. Aus (1.76) ist einfach erkennbar, dass für $n = 0$ die Lösungsfunktion konstant und gleich null wäre, dies beschreibt das Nichtvorhandensein eines Teilchens. Das soeben beschriebene, stark vereinfachte System kommt real nicht in der Natur vor. Es beinhaltet bei genauer Betrachtung Probleme in der mathematischen Formulierung, da die Wellenfunktionen bei $\pm a/2$ nicht stetig differenzierbar sind und somit prinzipiell keine Beschreibung der Realität darstellen können!

Die stetige Differenzierbarkeit, also hier die Stetigkeit der ersten Ableitung nach x , bedeutet eine stetige Entwicklung des Teilchenimpulses. Ein physikalisch reales System entsteht, wenn man den Wänden des Potentialtopfs eine endliche Höhe zuschreibt und als Konsequenz an den Topfrändern $\Theta(\pm a/2) \neq 0$ zulässt. Man erhält dann ein erstes einfaches Atommodell. Weiterhin beschreibt das im folgenden Kapitel behandelte System mathematisch exakt die Verhältnisse, die in sogenannten Halbleiter-Heterostrukturen vorherrschen. In Heterostrukturen werden verschiedene Halbleitermaterialien in direkten Kontakt gebracht. Dies hat Diskontinuitäten zur Folge im Potential, das von den sich im Material bewegenden Elektronen gespürt wird. Hierdurch werden Elektronen auf kleine Volumen im Bauelement eingeschränkt – die Grundlage u.a. von Halbleiterlasern und effizienten Leuchtdioden.

1.6.2 Teilchen im endlich tiefen Potentialtopf, Atommodell

Im endlich tiefen Potentialtopf nach Abb.1.15 wird beidseitig eine gleich hohe Potentialstufe mit der Höhe V_0 angenommen. Grundsätzlich erwarten wir ein Eindringen der Wellenfunktionen an den Rändern gemäß Abb.1.16, und ein Verhalten der Wellenfunktion in der Mitte des Potentialtopfs, das dem Fall mit unendlich hohen Wänden entspricht.

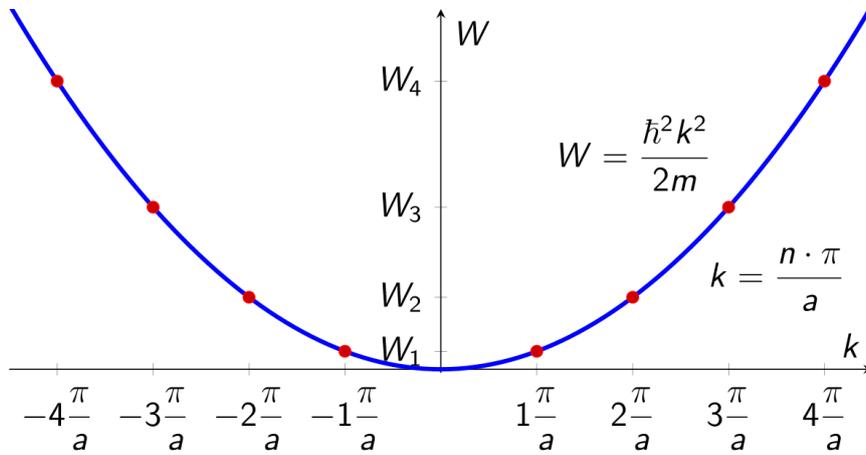


Abbildung 1.14: Erlaubte Energiewerte freier (Kurve) und gebundener (Punkte) Elektronen

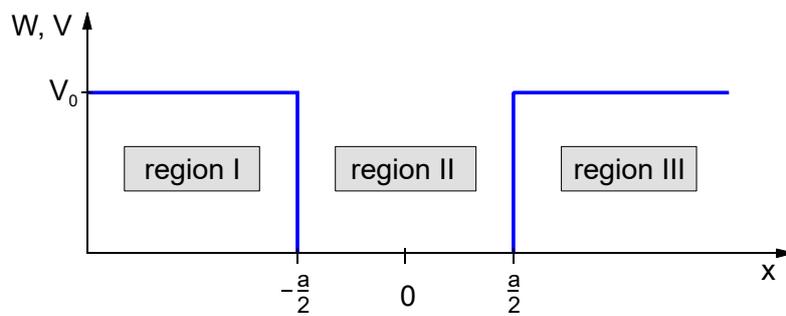


Abbildung 1.15: Potentialtopf mit endlich hohen Wänden

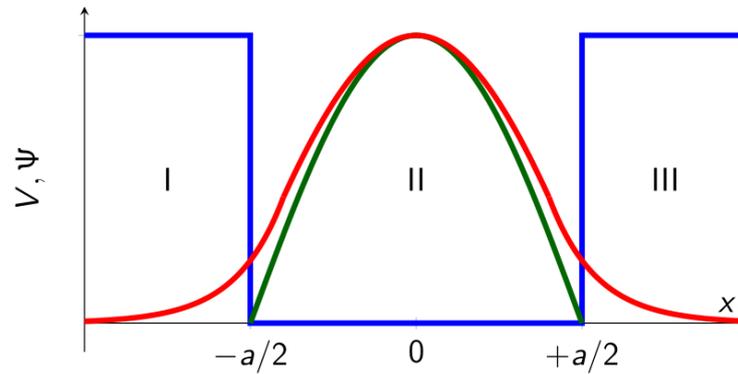


Abbildung 1.16: Wellenfunktion im endlichen Potentialtopf. Zum Vergleich ist in grün die erste Wellenfunktion des unendlich hohen Potentialtopfs eingezeichnet

Es gilt

$$|k|_{\text{I}} = |k|_{\text{III}}$$

$$V_{\text{I}} = V_{\text{III}} = V_0$$

$$V_{\text{II}} = 0$$

Wie bei der Behandlung der Potentialbarriere unterscheiden wir die Fälle

1. $W > V_0$
2. $W < V_0$

Die Beschreibung von gebundenen Teilchen entspricht dem zweiten Fall, der im Folgenden genauer untersucht werden soll. Ähnlich wie bei der Potentialbarriere werden als Randbedingungen die Stetigkeit und Differenzierbarkeit an den Grenzen $\pm a/2$ gefordert. Einsetzen dieser Randbedingungen in die Lösungsfunktionen der Schrödinger-Gleichung führt auf die Gleichungen

$$\begin{aligned} k_{\text{II},u} \cdot \frac{a}{2} \tan \left(k_{\text{II},u} \cdot \frac{a}{2} \right) &= k_{\text{I}} \cdot \frac{a}{2} = \sqrt{\beta_0^2 - \beta^2} \\ k_{\text{II},g} \cdot \frac{a}{2} \cot \left(k_{\text{II},g} \cdot \frac{a}{2} \right) &= k_{\text{I}} \cdot \frac{a}{2} = \sqrt{\beta_0^2 - \beta^2} \end{aligned} \quad (1.79)$$

mit

$$\beta_0^2 \equiv \frac{2mV_0}{\hbar^2} \cdot \left(\frac{a}{2}\right)^2 \quad \text{und} \quad \beta^2 \equiv \frac{2mW}{\hbar^2} \cdot \left(\frac{a}{2}\right)^2 \quad (1.80)$$

Diese Gleichungen sind analytisch nicht allgemein lösbar. Sie können aber numerisch oder, anschaulicher, zeichnerisch gelöst werden, indem man jeweils die rechte und linke Seite der Gleichungen (1.79) über $k_{\text{T}} \cdot (a/2)$ aufträgt. Am jeweils letzten Teil der Gleichungen erkennt man, dass in dieser Darstellung für die rechten Seiten Kreisbögen mit einem Radius von β_0 um den Koordinatenursprung resultieren. Die Schnittpunkte dieser Kreisbögen mit den periodischen \tan - bzw. \cot -Funktionen der linken Seiten stellen dann die Lösungen dar, aus denen man auch die erlaubten Energiewerte berechnen kann. Abb. 1.17 gibt diese Verhältnisse wieder.

Man erkennt, dass nunmehr nur **endlich viele** Energiewerte bzw. Wellenfunktionen erlaubt sind und dass deren Anzahl und Parameter von der Tiefe V_0 und der Breite a des Potentialtopfs abhängen. Wieder können durch die Normierungsvorschrift die Wellenfunktionen quantitativ bestimmt werden. Für den unendlich tiefen Potentialtopf ($V_0 \rightarrow \infty$) gilt $\beta_0 \rightarrow \infty$, sodass dort die Schnittpunkte mit den Asymptoten (in Abb. 1.17 gestrichelte vertikale Linien) die Lösungen darstellen; diese sind mit den Lösungen für den unendlich tiefen Topf identisch, wie es die Grenzwertbetrachtung verlangt.

1.6.3 Wasserstoffatom (H-Atom)

Nach diesen Vorbereitungen können wir uns der Behandlung des einfachsten atomphysikalischen Systems zuwenden, dem Wasserstoffatom. Es besteht aus einem Elektron im Coulombpotential des einfach positiv geladenen Atomkerns. Für die spätere Beschreibung des Festkörpers kann man sich diesen aus Einzelbausteinen ähnlich dem Wasserstoffatom zusammengesetzt denken.

Das Potential des Atomkerns, das das Elektron sieht, ist:

$$V(r) = \frac{Q_1 Q_2}{4\pi\epsilon_0\epsilon_r} \cdot \frac{1}{r} \quad (1.81)$$

wobei Q_1 und Q_2 die Ladungen des Kerns (H^+) bzw. des Elektrons sind, r der Abstand der Ladungen voneinander und ϵ_0 bzw. ϵ_r die dielektrischen Eigenschaften der Umgebung beschreiben. Aus dem Potential kann durch Gradientenbildung die bekannte

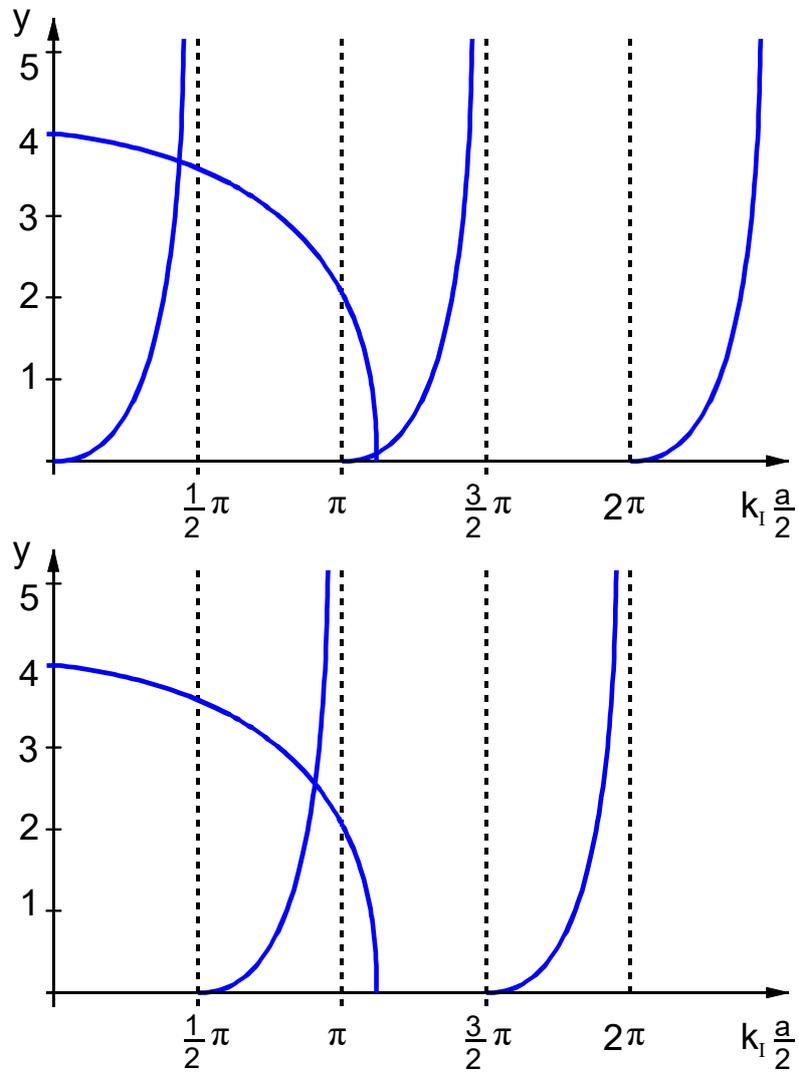


Abbildung 1.17: Graphische Lösung der Gleichungen (1.79) zum endlichen Potentialtopf

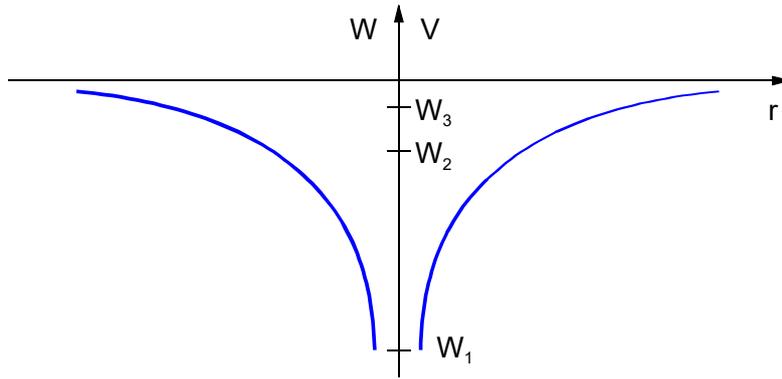


Abbildung 1.18: Coulombpotential des H-Atoms und Eigenwert (schematisch)

Formel für die Kraft zwischen zwei punktförmigen Ladungen hergeleitet werden:

$$F = -\frac{dV}{dr} = 4\pi\epsilon_0\epsilon_r \cdot \frac{1}{2} \quad (1.82)$$

Für das Wasserstoffatom gilt $Q_1 = +q$, $Q_2 = -q$, $\epsilon_r = 1$ und somit

$$V(r) = -\frac{q^2}{4\pi\epsilon_0} \cdot \frac{1}{r} \quad (1.83)$$

Es hat im Grenzfall $r \rightarrow \infty$ den Wert $V = 0$ (siehe Abb. 1.18). Dieses Potential muss jetzt in der Schrödinger-Gleichung verwendet werden. Es ist offensichtlich nur vom relativen Abstand r der Ladungen und nicht von der Richtung abhängig. Für dieses zentralsymmetrische Problem stellen Kugelkoordinaten das natürliche Bezugssystem dar.

Man verwendet den Laplace-Operator Δ (1.24) daher auch in Kugel-Koordinaten durch Ersetzung $(x, y, z) \rightarrow (r, \theta, \phi)$:

$$\Delta = \frac{\partial}{\partial r^2} + \frac{2}{r} \frac{\partial}{\partial r} + \frac{1}{r^2} \left[\cot \theta \frac{\partial}{\partial \theta} + \frac{\partial^2}{\partial \theta^2} + \frac{1}{\sin^2 \theta} \frac{\partial^2}{\partial \phi^2} \right] \quad (1.84)$$

Die Berechnung selbst ist sehr aufwendig und kann im Rahmen dieser Vorlesung nicht durchgeführt werden. Sie liefert unter anderem das Eigenwertspektrum des H-Atoms:

$$W_n = -\frac{q^2}{8\pi\epsilon_0 r_0} \cdot \frac{1}{n^2} \quad \text{mit } n = \{1, 2, 3, \dots\} \quad (1.85)$$

mit dem Bohr'schen Atomradius r_0 , benannt nach dem Physiker Niels Bohr

$$r_0 = \frac{4\pi\epsilon_0\hbar^2}{mq^2} = 5.29 \times 10^{-9} \text{ cm} \quad (1.86)$$

Wie erwartet, sind wieder nur diskrete Energiewerte erlaubt. Im Gegensatz zum unendlich tiefen, rechteckigen Potentialtopf (1.77) rücken die W_n mit wachsenden Quantenzahlen n näher zusammen. Dies ist anschaulich auch verständlich, denn mit wachsender Energie W_n des Elektrons nimmt die Bindung an den Kern aufgrund des Verlaufs von $V(r)$ ab.

Mit $W_n \rightarrow 0$ ist der Fall des freien Elektrons realisiert, das nun jede beliebige Energie aufnehmen kann, also ist auch $\Delta W_n \rightarrow 0$ erlaubt. Das negative Vorzeichen der W_n spiegelt den Bindungscharakter dieser Energie wider. Eine Energiezugabe ist erforderlich, um das Elektron vom Kern zu befreien. Im Gegensatz dazu ist die Energie des freien Elektrons als rein kinetische Energie positiv gezählt.

In diesem Zusammenhang ist der Begriff der Ionisierungsenergie W_{ion} wichtig. Diese muss aufgebracht werden, um das Elektron völlig vom H^+ Kern zu trennen, d.h. um das Elektron zu einem ungebundenen freien Elektron zu machen:

$$W_{\text{ion}} = |W_{n \rightarrow \infty} - W_{n=1}| = \frac{q^2}{8\pi\epsilon_0 r_0} = 13.6 \text{ eV} \quad (1.87)$$

1.6.4 Bohr'sches Atommodell

Diese Eigenschaften lassen sich in diesem einfachsten Fall des H-Atoms auch sehr anschaulich in Form des Bohr'schen Atommodells darstellen, das – ähnlich der Planetenbewegung – von mechanischen Bahnen des Elektrons um den Kern ausgeht, für die nach der Planck'schen Quantenhypothese nur Vielfache eines Elementardrehimpulsquants zulässig sind. Aus dem Gleichgewicht zwischen Coulombkraft und Zentrifugalkraft ergeben sich dann für diese Bahnen die erlaubten Energieniveaus. Ferner liefert dieses Bild einen Zugang zu den sogenannten Nebenquantenzahlen l, m, s , die zusätzlich zur Hauptquantenzahl n auftreten, wenn man Unsymmetrien einbezieht, die teilweise durch äußere Felder auftreten. Zu jedem n existiert nämlich ein Satz von Energiewerten $W_{n,l,m,s}$, die unter sich geringfügig unterschiedlich sind, aber zur gleichen "Elektronenschale" gehören. Eine Schale mit dem Index n besteht aus elliptischen Bahnen unterschiedlicher Exzentrizität, die durch die Drehimpulsquantenzahl l charakterisiert wird. Abb. 1.19 stellt dies für unterschiedliche Hauptquantenzahlen n dar. Die Bahn mit der

1 Grundlagen der Quantentheorie

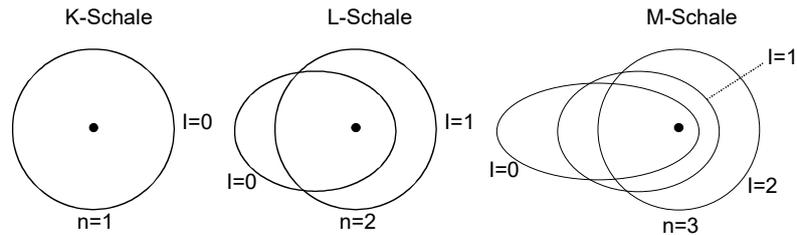


Abbildung 1.19: Elektronenbahnen des Bohr'schen Atommodells

höchsten Exzentrizität, z. B. $l = 0$ bei $n = 3$, entspricht der niedrigsten Energie in dieser Schale. Als Bahnen geladener Teilchen (Elektronen) tragen sie ein magnetisches Moment, das sich in äußeren Feldern ausrichten kann, vgl. Abb. 1.20. Diese Ausrichtung beschreibt die Orientierungs-Nebenquantenzahl (m). Schließlich besitzt jedes Elektron entweder den Spin $s = +1/2$ oder $s = -1/2$, der sich ebenfalls in äußeren Feldern in Feldrichtung ($s = +1/2$) oder entgegengesetzt dem Feld ($s = -1/2$) ausrichten kann. Die exakte Quantentheorie liefert für die Quantenzahlen Auswahlregeln in folgender Form:

$$\begin{aligned}
 n &= 1, 2, 3, \dots \\
 l &\leq n - 1 \\
 -l &\leq m \leq +l \\
 s &= \pm 1/2
 \end{aligned}
 \tag{1.88}$$

Je nach Anregungsstärke, d.h. Energiezufuhr, kann das einzelne Elektron des H-Atoms all diese Energiewerte $W_n l m s$ annehmen. Weiterhin kann mit diesem Modell der Aufbau der Elemente mit steigender Protonenzahl im Kern und entsprechend steigender Elektronenzahl in der Atomhülle beschrieben werden, wenn man zusätzlich das Pauli-Prinzip beachtet. Dies besagt, dass in einem atomaren System ein irgendwie ausgewähltes Teilchenpaar nicht exakt die gleiche Energie haben darf, d.h., jeder Zustand kann genau nur einmal besetzt werden (gilt für Elementarteilchen, die sog. Fermionen sind, darunter fallen die Elektronen). Daher werden die Energiewerte $W_n l m s$, die für jedes Element aufgrund der unterschiedlichen Kernladungszahl unterschiedliche Wertesätze darstellen, jeweils von niedrigstem Energiewert ansteigend aufgefüllt, indem sie mit den Hüllenelektronen besetzt werden. Wegen der Gültigkeit der Auswahlregeln (1.88) entsteht das Periodensystem der Elemente in Abb. 1.21.

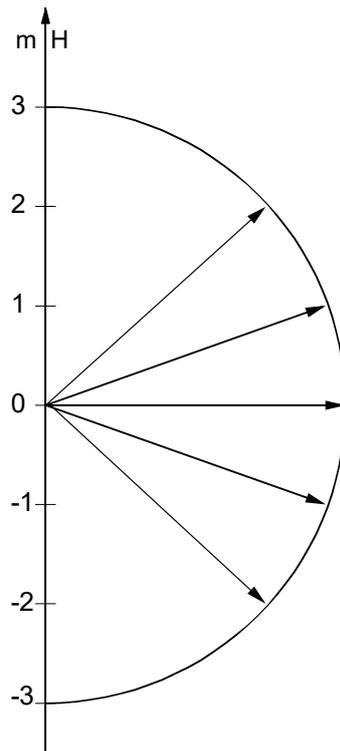


Abbildung 1.20: Mögliche Werte der Orientierungsquantenzahl m als Projektion des Drehimpulses (für $l = 3$)

1 Grundlagen der Quantentheorie

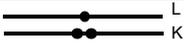
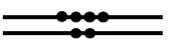
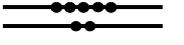
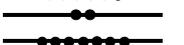
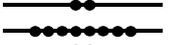
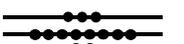
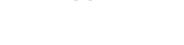
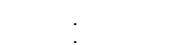
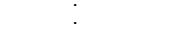
n	l	m	s	Schale	max. Anz. der Elektronen in der Schale	Element (Kernladung)	Besetzung der Schalen	Anz. der verschiedenen Elektronen in der jeweiligen Schale				
1	0	0	$\pm \frac{1}{2}$	K	2 (1s)	¹ H ² He		1x1s 2x1s				
2	0	0	$\pm \frac{1}{2}$	L	2 (2s)	³ Li		1x2s 2x1s				
		1	-1			$\pm \frac{1}{2}$			6 (2p)	⁴ Be		2x2s 2x1s
		0	0			$\pm \frac{1}{2}$			⁵ B	2x2s + 1x2p		
	1	0	$\pm \frac{1}{2}$		⁶ C	6 (2p)	⁷ N		2x2s + 2x2p			
	0	0	$\pm \frac{1}{2}$		⁸ O				...			
	1	+1	$\pm \frac{1}{2}$		⁹ F					...		
	<hr style="width: 50%; margin: auto;"/>						¹⁰ Ne		2x2s + 6x2p			
	<hr style="width: 50%; margin: auto;"/>						$\Sigma 8$					
	3	0	0		$\pm \frac{1}{2}$	M	2 (3s)	¹¹ Na		1x3s + 6x2p 2x2s + 2x1s		
1			-1	$\pm \frac{1}{2}$	6 (3p)			¹² Mg				...
0			0	$\pm \frac{1}{2}$	¹³ Al			...				
1		0	$\pm \frac{1}{2}$	10 (3d)	¹⁴ Si			2x3s + 2x3p 2x2s + 6x2p 2x1s				
1		+1	$\pm \frac{1}{2}$		¹⁵ P			...				
2		-2	$\pm \frac{1}{2}$		¹⁶ S			...				
<hr style="width: 50%; margin: auto;"/>							¹⁷ Cl		...			
<hr style="width: 50%; margin: auto;"/>							¹⁸ Ar		2x3s + 6x3p 2x2s + 6x2p 2x1s			
<hr style="width: 50%; margin: auto;"/>							$\Sigma 18$					
		2	-2	$\pm \frac{1}{2}$						...		
		-1	$\pm \frac{1}{2}$					...				
		0	$\pm \frac{1}{2}$					...				
		+1	$\pm \frac{1}{2}$					...				
		+2	$\pm \frac{1}{2}$					...				

Abbildung 1.21: Periodensystem

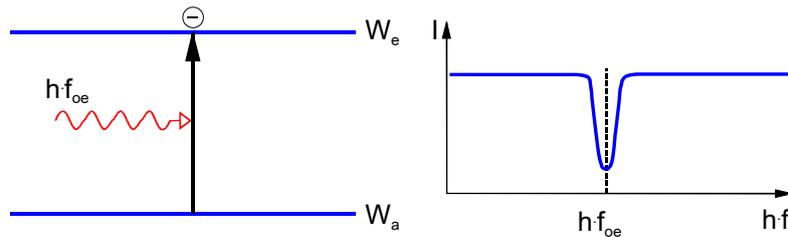


Abbildung 1.22: Anregung eines elektronischen Übergangs und zugehörige Absorptionsspektren

Die $W_n l m s$ bilden für jedes Element charakteristische Wertesätze, die auch messtechnisch mit zwei grundsätzlich unterschiedlichen Methoden erfassbar sind. Diese sollen kurz betrachtet werden, da sie auch für die Festkörperanalyse eine wichtige Rolle spielen.

1.6.5 Absorptionsspektroskopie

Eine Ansammlung von Atomen der gleichen Sorte wird mit Licht bestrahlt und die Schwächung der Strahlung wird in Abhängigkeit von der Frequenz gemessen. Die Schwächung erfolgt durch Absorption von Photonen, indem in den Atomen elektronische Übergänge zwischen den erlaubten Niveaus $W_n l m s$ von einem Ausgangswert in unbesetzten höhere Werte stattfinden. Abb. 1.22 zeigt vereinfacht ein System aus nur zwei Energiewerten und das dazu gehörige "Absorptionsspektrum", in diesem Fall bestehend aus nur einer Absorptionslinie, für die gilt:

$$h \cdot f_{ae} = W_e - W_a \quad (1.89)$$

Im Fall realer Atome mit vielen $W_n l m s$ entsteht entsprechend dem Energieniveauschema ein Absorptionsspektrum mit vielen Linien (Abb. 1.23).

1.6.6 Emissionsspektroskopie

Bei der Emissionsspektroskopie wird nicht die Schwächung des Anregungslichtes spektral vermessen, sondern die Strahlung, die das System als Reaktion auf die Anregungsstrahlung selbst aussendet. Dazu muss es derart angeregt werden, dass im zu untersuchenden Frequenzbereich eine Besetzungsumkehr stattfindet, die umgekehrt ablaufenden Übergänge ermöglicht, d.h. eine Bestrahlung mit höherenergetischem Licht kürze-

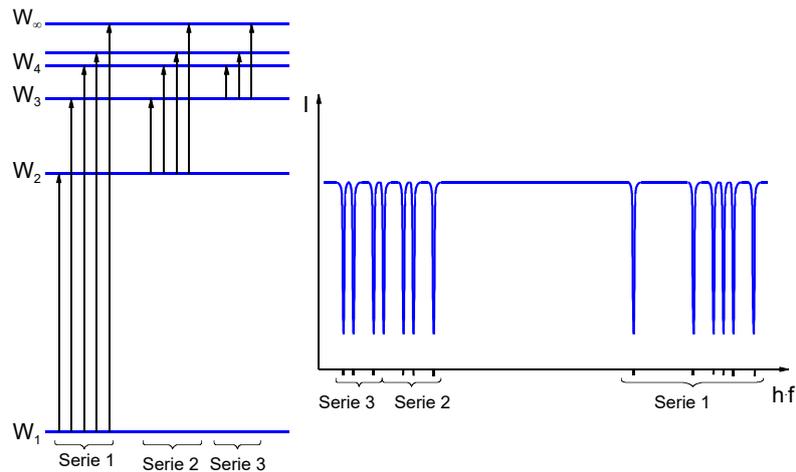


Abbildung 1.23: Absorptionsübergänge und zugehöriges Absorptionslinienspektrum (schematisch)

rer Wellenlänge muss stattfinden. Dann ergibt sich das Emissionsspektrum in Abb. 1.24, das dem Absorptionsspektrum in Abb. 1.23 entspricht.

Diese beiden Verfahren werden bei der Spektralanalyse verwendet und bilden die wichtigsten Methoden bei der Analyse von Proben unbekannter Zusammensetzung. Speziell die Emissionsspektroskopie kann sogar auf weit entfernte Objekte wie die Sonne oder andere "angeregte" Himmelskörper angewandt werden, deren Zusammensetzung wir daher sehr genau kennen (zumindest der äußeren Hülle).

Bis hierher haben wir mit Hilfe der Schrödinger-Gleichung die Eigenschaften von freien und gebundenen Teilchen, speziell Elektronen, beschreiben können. Im Festkörper treten Elektronen sowohl gebunden als auch "quasi"-frei in Erscheinung, d.h. die Elektronen können sich durch den Festkörper bewegen. Wir stellen dies makroskopisch als elektrische Leitfähigkeit fest. Im Folgenden wird ein einfaches, aber trotzdem aussagekräftiges Modell des Festkörpers auf der Grundlage der bisherigen Erkenntnisse entwickelt.

1.7 Kronig-Penny-Modell und Bandstruktur des Festkörpers

In diesem Kapitel werden wir die grundsätzlichen Eigenschaften der Elektronen in kristallinen Festkörpern kennenlernen. Wir werden dabei so vorgehen, dass wir ein einzelnes Elektron betrachten und dessen Eigenschaften auf alle anderen übertragen. Dies ist nur erlaubt, wenn die Wechselwirkung zwischen den einzelnen Elektronen im Ver-

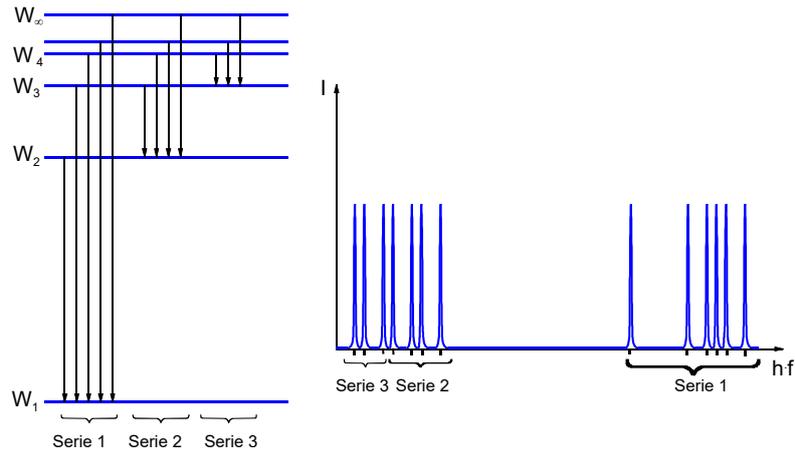


Abbildung 1.24: Emissionsübergänge und zugehöriges Emissionslinienspektrum (schematisch)

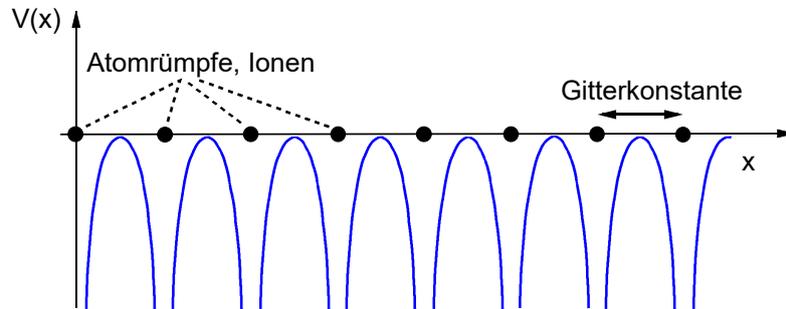


Abbildung 1.25: Potentialverlauf in einem Festkörper (eindimensional)

gleich zur Wechselwirkung mit dem Kristall vernachlässigbar ist, was für nicht zu hohe Elektronendichten und damit speziell für Halbleiter der Fall ist.

Somit haben wir die Schrödinger-Gleichung für ein Elektron zu lösen, das sich in einem Potential bewegt, das die Verhältnisse in einem Festkörper weitgehend realistisch, aber möglichst einfach beschreibt. Das Auffinden dieses Potentials ist daher das Kernproblem. Ist dieses gelöst, können wir nach dem bereits behandelten Verfahren die Lösung der Schrödinger-Gleichung durchführen.

Der ideale Festkörper-Kristall besteht aus einer periodischen Anordnung von Einzelatomrümpfen bzw. Gitterbausteinen, die, jeder für sich, ein Potential ähnlich dem H-Atom besitzen. Zur Vereinfachung der Rechnung beschränken wir uns zunächst auf eine räumliche Dimension. Dann hat das Potential den in Abb. 1.25 schematisch dargestellten Verlauf. Zur weiteren Vereinfachung nähern wir den Verlauf des Potentials durch Rechtecke an, weil wir dann an den Grenzen zwischen den Bereichen I und II

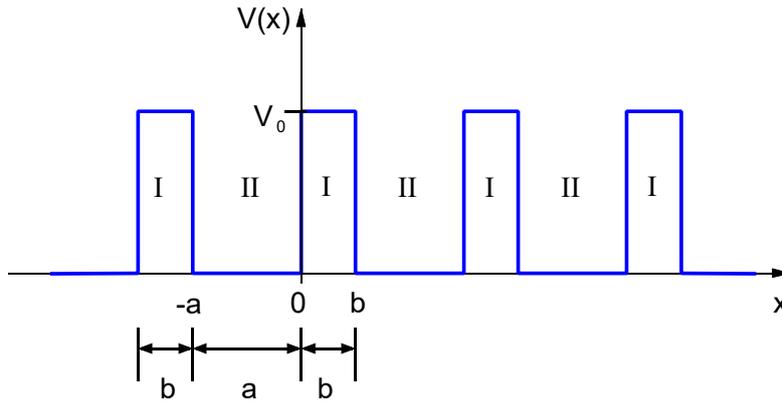


Abbildung 1.26: Idealisierter Potentialverlauf im Kronig-Penney-Modell des eindimensionalen Festkörpers

(vgl. Abb. 1.26) definierte Randbedingungen ansetzen können und somit eine wesentliche Schwierigkeit umgehen, die durch den nichtkonstanten Verlauf des Potentials bei der Behandlung des H-Atoms auftrat.

Das Potential mit der Periode $(a + b)$ und der Barrierenbreite b ist in den Bereichen I und II gegeben als

$$V(x) = \begin{cases} V_0 & \text{für } n(a + b) \leq x \leq n(a + b) + b \\ 0 & \text{sonst} \end{cases} \quad (1.90)$$

Hier ist $n = \{1, 2, 3, \dots\}$ ein ganzzahliger Parameter, der alle Bereiche durchläuft. Die in Abb. 1.26 gezeigte Situation stellt gewissermaßen die Vermischung der Verhältnisse "freies Elektron vor endlicher Potentialbarriere" und "Elektron im endlichen Potentialtopf" dar. Dieses Potential muss in die Schrödinger-Gleichung eingesetzt werden:

$$\Psi'' + \frac{2m}{\hbar^2} [W - V(x)] \Psi = 0 \quad (1.91)$$

Wir interessieren uns besonders für Elektronen, die sich "quasi"-frei im Festkörper bewegen können, weil diese die elektrische Leitfähigkeit tragen. Ein freies Elektron in konstantem Potential ($V(x) = \text{const.}$) kann durch diese Wellenfunktion beschrieben werden

$$\Psi(x) = A \cdot e^{jkx} \quad \text{mit} \quad A = \text{const.} \neq f(x) \quad (1.92)$$

1.7.1 Bloch-Theorem

Da im Festkörper, wie auch in unserem Modell, das Potential periodisch mit dem Kristallgitter schwankt, ist es naheliegend, sich die Wellenfunktion nach F. Bloch als mit gleicher Periode moduliert vorzustellen. Die folgenden beiden Ausdrücke beschreiben das Bloch'sche Theorem, in dem die Lösung eines periodischen Problems selbst einen Teil mit gleicher Periode (hier mit Periode $a + b$) besitzt.

$$\begin{aligned} \Psi(x) &= U(x)e^{jkx} \\ \text{mit } U(x) &= U(x + a + b) \end{aligned} \quad (1.93)$$

Diesen Ansatz verwendet man in der Schrödinger-Gleichung und erhält die folgende Differentialgleichung für $U(x)$:

$$U''(x) + 2jkU'(x) - \left\{ k^2 - \frac{2m}{\hbar^2} [W - V(x)] \right\} U(x) = 0 \quad (1.94)$$

Das weitere Vorgehen ist bereits bekannt:

1. getrennte Lösung von (1.94) in den Bereichen I und II
2. Forderung der Eindeutigkeit und stetigen Differenzierbarkeit an den Grenzen zwischen den Bereichen I und II (Randbedingungen)

Folgende Abkürzungen erleichtern die Schreibearbeit:

$$\begin{aligned} \alpha^2 &= \frac{2mW}{\hbar^2} \\ \beta^2 &= \frac{2m(W - V_0)}{\hbar^2} \end{aligned} \quad (1.95)$$

So entstehen aus (1.94)

$$\begin{aligned} U_{\text{I}}'' + 2jkU_{\text{I}}' - (k^2 - \beta^2) U_{\text{I}} &= 0 \\ U_{\text{II}}'' + 2jkU_{\text{II}}' - (k^2 - \alpha^2) U_{\text{II}} &= 0 \end{aligned} \quad (1.96)$$

Im Gegensatz zur früheren Behandlung wird eine einheitliche Wellenzahl k für beide Bereiche benutzt. Die zu berechnenden Energieeigenwerte gelten für den ganzen Festkörper und schwanken nicht mit dem periodischen Potential. Die Ansätze zur Lösung

von (1.96) haben die Form

$$\begin{aligned} U_{\text{I}}(x) &= A_{\text{I}}e^{j(\beta - k)x} + B_{\text{I}}e^{-j(\beta+k)x} \\ U_{\text{II}}(x) &= A_{\text{II}}e^{j(\alpha - k)x} + B_{\text{II}}e^{-j(\alpha+k)x} \end{aligned} \quad (1.97)$$

mit den Randbedingungen der Stetigkeit an den Grenzen der Regionen I und II

$$\begin{aligned} U_{\text{I}}(0) &= U_{\text{II}}(0) \\ U_{\text{I}}(b) &= U_{\text{II}}(-a) \end{aligned} \quad (1.98)$$

und der stetigen Differenzierbarkeit

$$\begin{aligned} U'_{\text{I}}(0) &= U'_{\text{II}}(0) \\ U'_{\text{I}}(b) &= U'_{\text{II}}(-a) \end{aligned} \quad (1.99)$$

Die zweite Zeile in (1.98) und (1.99) spiegelt besonders deutlich die Periodizität auch der Randbedingungen wider. Mit (1.98) und (1.99) liegen vier Gleichungen vor, mit denen die vier Koeffizienten $A_{\text{I}} \dots B_{\text{II}}$ eindeutig bestimmt werden können. Sie bilden ein homogenes Gleichungssystem

$$(a_{ik}) \cdot \begin{pmatrix} A_{\text{I}} \\ A_{\text{II}} \\ B_{\text{I}} \\ B_{\text{II}} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (1.100)$$

das nur unter der Bedingung

$$\det \{a_{ik}\} = 0 \quad (1.101)$$

nichttriviale Lösungen hat.

Nach einigen Umformungen unter Verwendung der Winkelfunktionen, dargestellt durch Exponentialausdrücke, liefert Gleichung (1.101):

$$-\frac{\alpha^2 + \beta^2}{2\alpha\beta} \sin(\alpha a) \sin(\beta b) + \cos(\alpha a) \cos(\beta b) = \cos[k(a + b)] \quad (1.102)$$

Auch hier ist eine Fallunterscheidung angebracht:

1.7.2 Fall 1: Elektronenenergie kleiner als Potentialbarriere

Für $W < V_0$ ist β mit (1.95) rein imaginär. Man setzt:

$$\begin{aligned}\beta &= j\gamma \\ \gamma^2 &= \frac{2m(V_0 - W)}{\hbar^2}\end{aligned}\tag{1.103}$$

Mit Hilfe der Beziehungen für komplexe Winkelfunktionen

$$\begin{aligned}\sin(j\varphi) &= j \sinh(\varphi) \\ \cos(j\varphi) &= \cosh(\varphi)\end{aligned}\tag{1.104}$$

gilt:

$$\frac{\gamma^2 - \alpha^2}{2\alpha\gamma} \sin(\alpha a) \sinh(\gamma b) + \cos(\alpha a) \cosh(\gamma b) = \cos[k(a + b)]\tag{1.105}$$

Schon hier erkennt man, dass die Lösungsmenge für α und γ bzw. β dadurch eingeschränkt ist, dass die Kosinusfunktion auf der rechten Seite nur Werte zwischen -1 und +1 annehmen kann. Im Gegensatz zum Potentialtopf-Problem, wo nur diskrete Energiewerte erlaubt waren, sind jetzt aber ganze Energiebereiche, die **Energiebänder**, erlaubt. Auf diesem Sachverhalt beruhen die Begriffe **Bandstruktur** oder **Bändertheorie** des Festkörpers.

Zur weiteren Vereinfachung, aber ohne grundlegend die Verhältnisse zu verändern, wird folgender Grenzübergang durchgeführt:

$$\begin{aligned}b &\rightarrow 0 \\ V_0 &\rightarrow \infty \\ \Rightarrow V_0 \cdot b &= \text{const.} \equiv \Gamma\end{aligned}\tag{1.106}$$

Diese Annahme ist deshalb sinnvoll, weil mit der Bedingung $V_0 \cdot b = \Gamma = \text{const.}$ die Tunnelwahrscheinlichkeit durch die Barrierenbereiche I annähernd konstant bleibt, vgl. (1.69), und damit die Bewegung der Elektronen im periodischen Potential fast unverändert bleibt.

In (1.105) stehen Ausdrücke, die γb enthalten. Für große V_0 ist $\gamma \propto \sqrt{V_0}$. Für das

Produkt γb gilt nach (1.106)

$$\lim_{V_0 \rightarrow \infty} (\gamma b) = \lim_{V_0 \rightarrow \infty} (\sqrt{V_0 b}) = \lim_{V_0 \rightarrow \infty} \left(\frac{\sqrt{\Gamma}}{\sqrt{b}} b \right) = 0 \quad (1.107)$$

Damit können die Funktionen in (1.105), die γb enthalten, angenähert werden mit $\cosh(\gamma b) \rightarrow 1$ und $\sinh(\gamma b) \rightarrow \gamma b$. Es ergibt sich aus (1.105) der vereinfachte Ausdruck für $W < V_0$

$$\frac{(\gamma^2 - \alpha^2)ab}{2} \cdot \frac{\sin(\alpha a)}{\alpha a} + \cos(\alpha a) = \cos(ka) \quad (1.108)$$

Als nächstes untersuchen wir den Term $\gamma^2 - \alpha^2$ für $V_0 \gg W$ mithilfe von (1.95) und (1.103).

$$\gamma^2 - \alpha^2 = \frac{2m(V_0 - W)}{\hbar^2} - \frac{2mW}{\hbar^2} = \frac{2m(V_0 - 2W)}{\hbar^2} \rightarrow \frac{2mV_0}{\hbar^2} \quad \text{für } V_0 \gg W \quad (1.109)$$

Damit wird (1.108) zu

$$\frac{amV_0b}{\hbar^2} \cdot \frac{\sin(\alpha a)}{\alpha a} + \cos(\alpha a) = \cos(ka) \quad (1.110)$$

Wenn man Grenzwerte von Größen bildet, muss man das Konvergenzverhalten des gesamten Ausdrucks prüfen. In (1.110) ist der erste Faktor

$$P \equiv \frac{amV_0b}{\hbar^2}$$

beschränkt, da $V_0 b = \Gamma = \text{const.}$ eine Konstante ist, genau wie der Atomabstand a und die Elektronenmasse m .

Die sinc-Funktion (Abb. 1.27) ist definiert als $\text{sinc}(x) = \text{si}(x) = \sin(x)/x$. Sie konvergiert für $x \rightarrow 0$ gegen 1, und verschwindet für $x \rightarrow \infty$. Übrig bleibt nun mit der o.a. Abkürzung P , die vom Potential V_0 abhängt, und dem Einsetzen der sinc-Funktion der folgende Ausdruck:

$$P \cdot \text{sinc}(\alpha a) + \cos(\alpha a) = \cos(ka) \quad (1.111)$$

Die Gleichung (1.111) verknüpft die Teilchenenergie W über $\alpha^2 = 2m/\hbar^2 \cdot W$ mit der Wellenzahl k des Elektrons. Es handelt sich um eine transzendente Gleichung, die nicht geschlossen lösbar ist. Aufschlussreich ist zunächst eine Betrachtung der möglichen Lö-

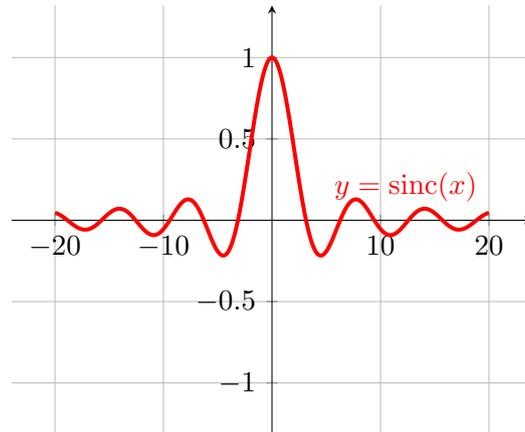


Abbildung 1.27: sinc-Funktion

sungsmenge. Die rechte Seite der Gleichung kann nur Werte im Bereich $-1 \dots 1$ annehmen.

Aus Abb. 1.28 ist ersichtlich, dass dies zu einer Einschränkung der möglichen $W(k)$ -Wertepaare führt. Nur die Abschnitte der roten Kurve, deren (grün unterlegte) Wertebereiche zwischen -1 und $+1$ liegen, sind Lösungen. In den blau unterlegten Bereichen gibt es keine Lösung. Auf der x -Achse ist die Energie als $(\alpha \cdot a)$ aufgetragen. Nur die grün unterlegten Energiewerte sind zulässig.

Berechnung der Energiebänder Wie können wir aus (1.111) die $W(k)$ -Beziehung für Elektronen in unserem eindimensionalen Kronig-Penney-Modell berechnen? Als Parameter lassen wir (αa) entlang der x -Achse von Abb. 1.28 laufen. Wir erhalten für jeden Wert von (αa) die Energie W direkt aus der Definition von α

$$W(\alpha a) = \frac{\hbar^2(\alpha a)^2}{2ma^2}$$

Den folgenden Ausdruck können wir aus (1.111) für k bilden

$$k(\alpha a) = \frac{1}{a} \cdot \arccos [P \cdot \text{sinc}(\alpha a) + \cos(\alpha a)]$$

Somit bekommen wir eine parametrisierte Funktion $(k, W)(\alpha a)$, deren Werte wir aufzeichnen können. Diese ist in Abb. 1.29 dargestellt. Die Grafik zeigt deutlich, dass gewisse Energie-Bereiche nicht zugänglich sind, diese werden als **Energielücke** bezeichnet.

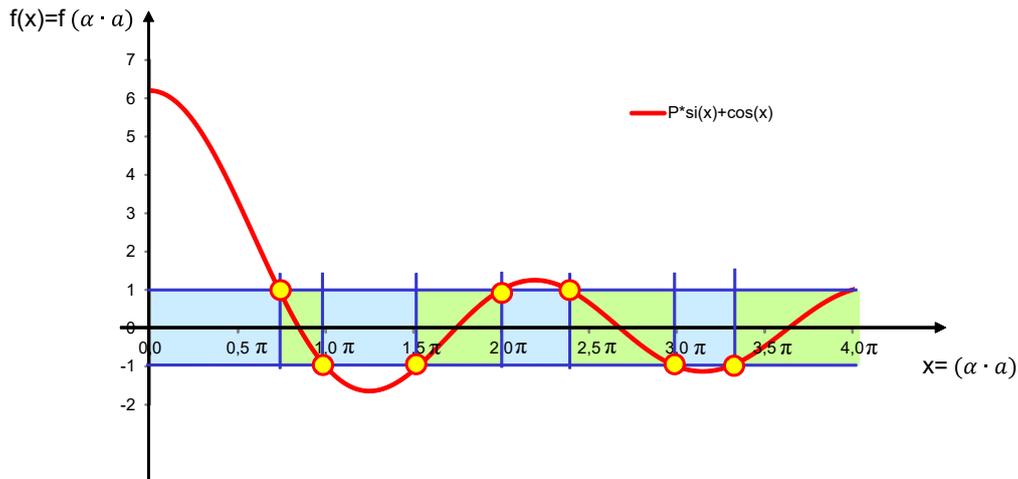


Abbildung 1.28: Grafische Lösung der Bestimmungsgleichung für das Kronig-Penney-Modell

Zwischen den Lücken liegen **Bänder**, die kontinuierliche Energiewerte zulassen.

Aus (1.111) folgt als weiterer wichtiger Punkt, dass die Wellenzahl k nicht eindeutig durch die Energie W bestimmt ist. Die Wellenzahl k kommt in der Bestimmungsgleichung nur auf der rechten Seite als Argument im Kosinus vor. Die Addition von Vielfachen von π/a ändert nichts am Wert der rechten Seite der Gleichung! Die Äste der W - k -Beziehung in Abb. 1.29a können daher in x -Richtung zusammengeschoben werden. Es entsteht ein sogenanntes reduziertes Energieschema (Abb. 1.29b), dessen k -Werte auf die erste Brillouin-Zone beschränkt sind.

Im realen Gitter der Ortskoordinaten wird die Einheitszelle durch gleichförmig wiederholte Atomabstände (in unserem Beispiel der eindimensionale Abstand a) definiert. Im eindimensionalen reziproken Gitter der Wellenzahlen ist die Größe der Einheitszelle durch π/a gegeben. Die Einheitszelle des reziproken Gitters wird als Brillouin-Zone bezeichnet. Genau wie sich die realen Kristallstrukturen gitterperiodisch wiederholen, so wiederholt sich die Bandstruktur $W(k)$ im reziproken Gitter. Die Periodizität des reziproken Gitters beruht auf der Periodizität des realen Gitters.

Die Periodizität des realen Gitters hatten wir als erste Annahme im Kronig-Penney-Modell als periodisches Potential formuliert. Als Ergebnis erhalten wir eine reziprok-periodische Bandstruktur.

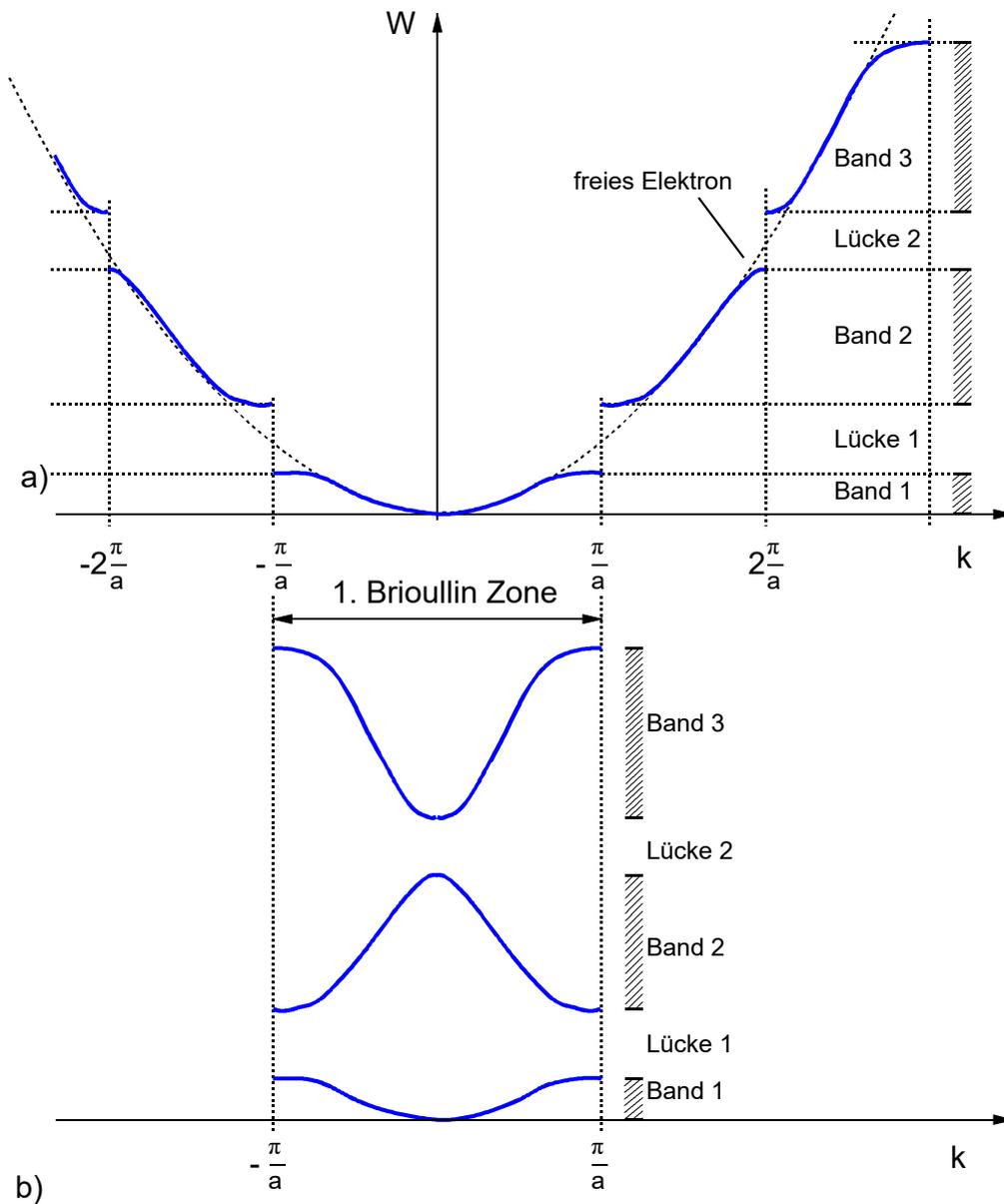


Abbildung 1.29: Energiebänder des Festkörpers im Kronig-Penney-Modell mit a) vollständigem und b) reduziertem Energieschema

Für jedes erlaubte Energieband im vollständigen Schema durchläuft k die Werte

$$n\pi/a < k < (n+1)\pi/a \quad \text{mit } n = \{ \dots, -3, -2, -1, 0, 1, 2, 3, \dots \} \quad (1.112)$$

Dabei gehören jeweils zwei unterschiedliche Energiewerte, der höchste eines Bandes und der niedrigste des nächst höheren Bandes, zum gleichen k -Wert. Negative k -Werte bedeuten Elektronenbewegung in negativer x -Richtung und führen für gleiche Absolutwerte natürlich zu gleichen Energiewerten.

Man erkennt, dass für Energiewerte im Zentrum eines erlaubten Bandes das $W(k)$ -Diagramm fast mit dem eines freien Elektrons identisch ist: ein Elektron tief im erlaubten Band verhält sich wie ein freies Elektron, es kann beispielsweise jede auch beliebig kleine Energiemenge, z. B. durch Beschleunigung aus einem elektrischen Feld, aufnehmen und somit **elektrische Leitfähigkeit** verursachen. Andererseits treten in der Nähe der Werte für $\pm(n \cdot \pi/a)$ große Unterschiede zum freien Elektron auf. Hier kommen die Aspekte des gebundenen Teilchens (im Potentialtopf) zum Ausdruck.

1.7.3 Fall 2: Elektronenenergie größer als Potentialbarriere

Die Situation $W > V_0$ wollen wir nur in Form einer anschaulichen Betrachtung behandeln, die aus dem Fall $W < V_0$ hervorgeht. Man erkennt bereits an Abb. 1.28 und 1.29, dass mit zunehmender Energie W die erlaubten Bänder immer breiter und die verbotenen Energielücken immer schmaler werden. Schließlich sind oberhalb einer bestimmten Energie alle Werte erlaubt – wie wir bereits wissen, verhalten sich Elektronen mit Energiewerten tief in erlaubten Bändern wie freie Elektronen. Dies gilt erst recht für Elektronen mit $W > V_0$.

1.7.4 effektive Elektronenmasse; Elektronen und Löcher

Für **freie Elektronen** ist der Zusammenhang zwischen den Größen Energie W , Impuls p , Wellenzahl k , Geschwindigkeit v und Masse m bekannt:

$$W(k) = \frac{p^2}{2m} = \frac{(mv)^2}{2m} = \frac{\hbar^2 k^2}{2m} \quad (1.113)$$

Wir erhalten die Geschwindigkeit v aus (1.113) als erste Ableitung von W nach k

$$v = \frac{1}{\hbar} \cdot \frac{dW}{dk} \quad (1.114)$$

Eine zweite Differentiation der Energie W nach k ergibt die **reziproke Masse**

$$\frac{d^2W}{dk^2} = \frac{\hbar^2}{m} \quad (1.115)$$

Diese Zusammenhänge übernehmen wir für die Elektronen im Festkörper, nennen aber insbesondere die Masse nun effektive Masse m^* , weil sie im Gegensatz zum freien Elektron von k abhängig ist: $m^* = m^*(k)$. Wir legen jetzt das niedrigste erlaubte Energieband zugrunde und ermitteln durch Differentiation nach k den Verlauf der Geschwindigkeit $v(k)$ und der effektiven Masse m^* , qualitativ in Abb. 1.30 gezeigt.

Für Wellenzahlen k nahe bei Null verhält sich das Elektron wie ein freies Elektron: mit steigendem k , d.h. steigender Impuls, wachsen Geschwindigkeit und Energie kontinuierlich an und die Masse hat einen nahezu konstanten Wert $m^* \approx m^*(k=0)$. Mit wachsendem k erreicht aber die Geschwindigkeit etwa in der Mitte des erlaubten Bandes ein Maximum, um darüber hinaus wieder abzufallen; d.h. obwohl der Impuls $\hbar k$ weiter steigt, fällt die Geschwindigkeit. Das ist formal nur dadurch zu erklären, dass die zugehörige Masse unendlich wird und das Vorzeichen wechselt. An den Bandkanten schließlich ist $v=0$, auch wenn die Wellenzahl $k = \pm(\pi/a) \neq 0$ ist.

An der Bandkante ruht das Elektron, oder anders ausgedrückt: es ist **fest gebunden**. Im Wellenbild entspricht dies stehenden, nicht fortlaufenden Wellen. Die Energie des Elektrons ist an den Bandkanten eine rein potentielle Energie, die wir als Bindungsenergie auffassen können.

Besonders ungewöhnlich ist das Verhalten der effektiven Masse m^* , die für steigenden Betrag von k zunächst unendlich groß und schließlich in der oberen Hälfte eines jeden Bandes sogar **negativ** wird. Eine anschauliche Erklärung erlaubt die Einführung des Begriffs Defektelektron oder Loch. Der Massebegriff hängt wesentlich mit der trägen Masse und somit mit dem Beschleunigungsbegriff zusammen: Für die Beschleunigung einer Masse m^* benötigt man die Kraft F

$$\vec{F} = m^* \cdot \vec{a} = q \cdot \vec{E} \quad (1.116)$$

wenn die Masse eine Ladung q trägt (Elektron) und sich im elektrischen Feld \vec{E} befindet.

Wenn nun die Masse negativ wird, so kann dies dadurch kompensiert werden, dass man gleichzeitig das Vorzeichen der Ladung umkehrt. Statt $-|m^*|$ und $-|q|$ verwendet man nach dem Vorzeichenwechsel der Masse $+|m^*|$, $+|q|$. Das dadurch beschriebene

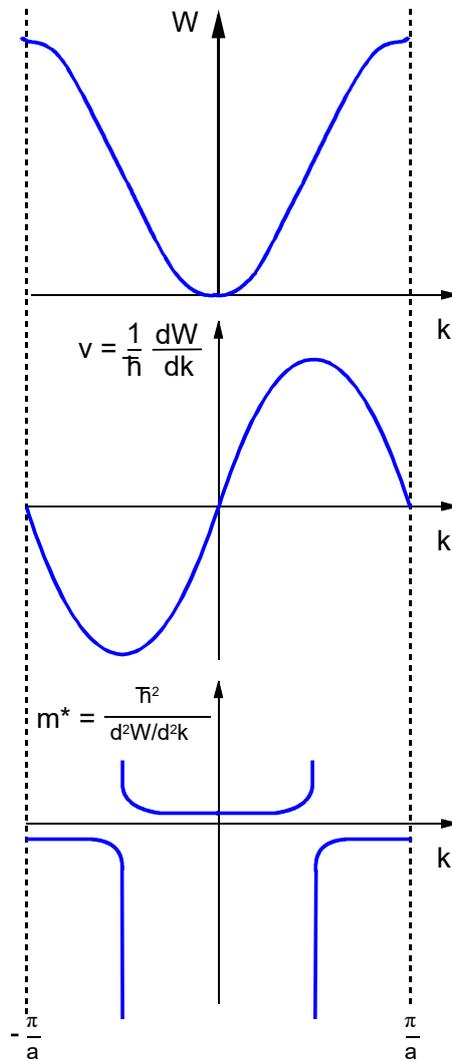


Abbildung 1.30: Zusammenhang zwischen Energie, Geschwindigkeit und effektiver Masse für Elektronen im Festkörper

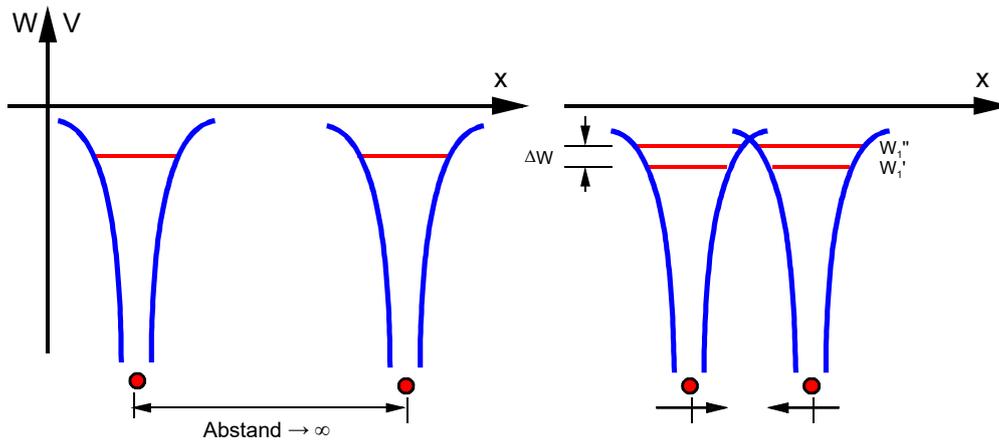


Abbildung 1.31: Aufspaltung zweier identischer Energieniveaus aufgrund des Pauli-Prinzips

Teilchen trägt also eine positive Masse und eine **positive Ladung**. Es lässt sich als fehlendes Elektron, Defektelektron oder Loch nahe der Oberkante eines Bandes interpretieren, da dieses in der Umgebung anderer negativ geladener Elektronen eine fehlende negative Ladung, und somit eine positive Ladung trägt. Diese Begriffsbildung gilt genau genommen nur für das energetisch höchst liegende Band, das mit Elektronen besetzt ist. Dieses ist das Valenzband. Das darüber liegende, somit weitestgehend leere erlaubte Band ist das Leitungsband.

1.7.5 Pauli-Prinzip im Festkörper

Wir haben gelernt, dass Elektronen in einem Potentialtopf (z.B. im H-Atom) als erlaubte Energiewerte eine Serie diskreter, scharfer Energiewerte $W_{n,l,m,s}$ annehmen können. Andererseits haben wir für den Festkörper, dargestellt als periodisch angeordnete Potentialtöpfe mit gegenseitiger Kopplung, erlaubte und verbotene Energiebänder kennengelernt, die aber – zugegebenermaßen wenig anschaulich – aus der quantentheoretischen Rechnung folgten. Mit Hilfe des Pauli-Prinzips kann eine anschauliche Interpretation dieses Ergebnisses angegeben werden.

Dazu stellt man sich zunächst zwei Atome mit jeweils einem Elektron im energetischen Zustand W_1 vor, wie in Abb. 1.31 gezeigt. Nach Pauli kann diese Situation aber nur dann vorliegen, wenn beide Atome sehr weit voneinander entfernt sind. Sobald die Entfernung abnimmt und eine Wechselwirkung auftritt, greift das Pauli-Prinzip: in einem System kann ein diskreter Energiezustand von nur einem Elektron eingenommen wer-

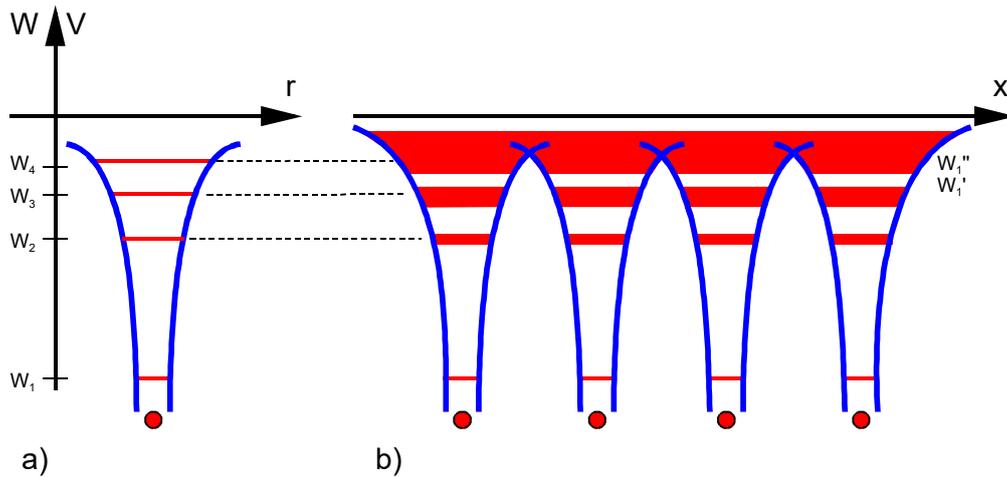


Abbildung 1.32: Zusammenhang zwischen den Energieniveaus des Einzelatoms und den Energiebändern des Festkörpers

den. Wie in Abb. 1.31 dargestellt, spalten in beiden Atomen die Energieniveaus in jeweils zwei leicht unterschiedliche Werte W_1' und W_1'' auf, weil die Elektronen nicht nur, wie bei weit voneinander getrennten Atomen, das Potential ihres Kerns, sondern auch das des anderen Atomkerns "sehen". Je näher die beiden Atome zusammengebracht werden, desto stärker ist die gegenseitige Wechselwirkung und desto größer die Energieaufspaltung ΔW . In einem Festkörper sind nun außerordentlich viele Atome mit jeweils mehr als nur einem Elektron auf engem Raum zusammengebracht ($N \approx 5 \times 10^{22} \text{ cm}^{-3}$). Entsprechend viele Energieniveaus müssen zur Erfüllung des Pauli-Prinzips entstehen, und diese liegen derart dicht beieinander, dass sie wegen der Gültigkeit der Unschärferelationen nicht unterscheidbar sind, wie in Abb. 1.32 angedeutet. Durch die starke Bindung der Elektronen in tieferen Niveaus wechselwirken diese nur schwächer mit den Nachbar-Atomen, und daher ist auch die Energieaufspaltung geringer, die Bänder also schmaler als die höher liegenden. Andererseits können die Elektronen in diesen höher liegenden Bändern nicht mehr nur einem Atom zugeordnet werden, sondern gehören wegen der schwachen Bindung gewissermaßen dem ganzen Festkörperverband an. Diese Elektronen tragen die elektrische Leitfähigkeit. Somit ist mit Hilfe des Pauli-Prinzips eine anschauliche Deutung der quantentheoretischen Berechnung der Verhältnisse im Festkörper gelungen. Wir nutzen diese Grundlagen nun zum weiteren detaillierten Studium der elektronischen Eigenschaften des Festkörpers.

2 Elektronische Eigenschaften von Festkörpern

2.1 Einteilung in Isolatoren, Metalle und Halbleiter

Die gegenseitige chemische Bindung der Festkörperatome bewirken die Elektronen im obersten besetzten Band, weil gerade sie ja die Wechselwirkung mit den Nachbaratomen bewerkstelligen. Man nennt sie Valenzelektronen und das dazugehörige Band **Valenzband**. Die Besetzung des Valenzbandes und des nächst höheren **Leitungsbandes** entscheidet über die elektrische Leitfähigkeit des Festkörpers. Wie wir bereits wissen und in Abb. 2.1 gezeigt, hängt die Energieaufspaltung in Bänder vom Atomabstand a ab.

Je stärker die Einzelatome samt deren Elektronen in Wechselwirkung treten, desto weiter spalten die aus den diskreten Niveaus entstehenden Bänder auf, bis sie sich schließlich überlappen. In Abb. 2.1 sind qualitativ die Bandverhältnisse für die drei hinsichtlich der elektrischen Leitfähigkeit zu unterscheidenden Festkörper Isolator, Halbleiter und Metall eingezeichnet: Isolator und Halbleiter unterscheiden sich nur quantitativ durch die Breite des verbotenen Energiebandes, der Energielücke W_g , die das Valenzband vom Leitungsband trennt. Bei Metallen liegen die Verhältnisse etwas komplizierter, wie die folgende Diskussion zeigen wird.

2.1.1 Isolatoren

In Isolatoren ist das Valenzband vollständig mit Elektronen besetzt und das nächst höhere Band, das Leitungsband, ist völlig leer (Abb. 3.2).

Dies gilt zumindest bei $T = 0\text{ K}$. Für die Bewegung eines Elektrons, d.h. für den Ladungstransport und somit den Stromfluss, muss dieses Elektron kinetische Energie aufnehmen, also einen anderen energetischen Zustand einnehmen. Dies ist aber im Valenzband bei nicht möglich, da alle verfügbaren Plätze bereits mit Elektronen besetzt

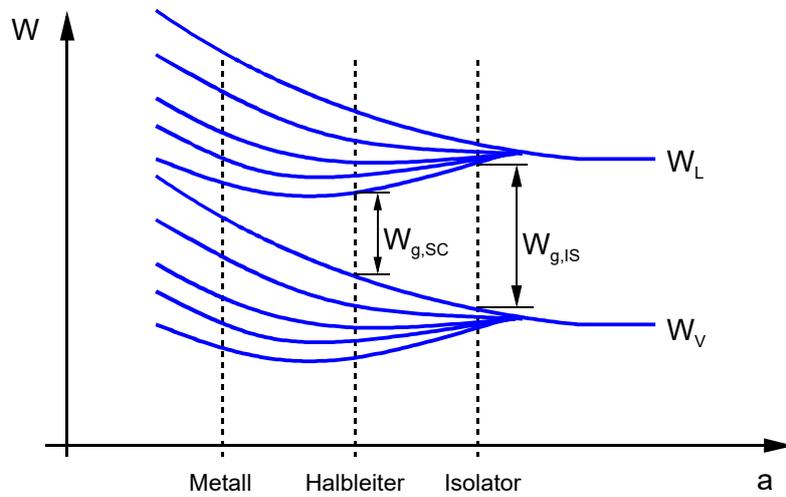


Abbildung 2.1: Aufspaltung der Niveaus als Funktion vom Atomabstand a

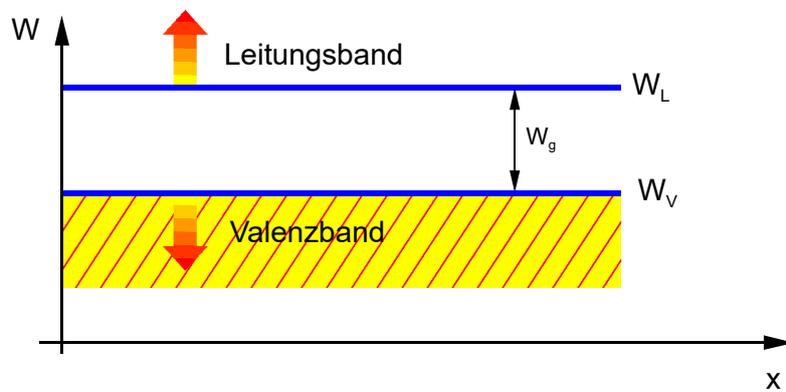


Abbildung 2.2: Bändermodell des Isolators bei $T = 0$ K (gilt auch für intrinsische Halbleiter)

sind. Die nächst höher liegenden erlaubten Energiewerte sind erst im Leitungsband anzutreffen, das aber durch die sehr große Energielücke von mehreren Elektronenvolt (eV) vom Valenzband getrennt ist.

Derart hohe kinetische Energien kann das Elektron üblicherweise, z.B. aus einem von außen an den Festkörper angelegten elektrischen Feld, nicht aufnehmen. Das bedeutet, dass im Isolator kein Strom fließt. Selbst wenn eine endliche Temperatur $T > 0$ K zugelassen wird, besitzen laut Statistik nur ganz vereinzelte Elektronen eine derart hohe thermische Energie, dass sie die Energielücke W_g überspringen können.

2.1.2 Halbleiter

Bei reinen Halbleitern liegen prinzipiell die gleichen Verhältnisse wie beim Isolator in Abb. 2.2 vor, nur ist nun die Energielücke W_g geringer, etwa 1 eV. Hier ist bei endlichen Temperaturen $T > 0$ K die Wahrscheinlichkeit erheblich höher, dass einige Elektronen ausreichend thermische Energie zur Überwindung der Lücke besitzen. Sie befinden sich dann im Leitungsband mit beliebig vielen freien Plätzen, die für die Elektronen nach Aufnahme kinetischer Energie zur Verfügung stehen. Damit ist Elektronenbewegung und folglich Stromtransport möglich. Die Leitfähigkeit ist somit deutlich höher als bei Isolatoren, aber immer noch außerordentlich gering im Vergleich zu Metallen (siehe unten).

Durch gezieltes Einbringen von Fremdatomen in das Gitter des reinen Halbleiters kann man die elektrische Leitfähigkeit ganz erheblich erhöhen. Durch dieses sogenannte Dotieren kann gezielt sowohl Elektronen- als auch Löcherleitfähigkeit erzeugt werden (vgl. Kap. 2.3.4).

Typische Halbleiter sind Silizium (Si) und Germanium (Ge) aus der 4. Hauptgruppe des Periodischen Systems. Es gibt aber neben diesen Elementhalbleitern auch Verbindungshalbleiter, die sich aus Atomen der 3. und 5. Gruppe (III-V Halbleiter, z.B. GaAs, GaN, InP) oder auch aus der 2. und 6. Gruppe (II-VI Halbleiter, z.B. ZnSe) zusammensetzen. Zweikomponentige Halbleiter wie GaN werden als binäre Halbleiter bezeichnet. Es gibt auch ternäre (z.B. GaAlAs) und quaternäre (z.B. InGaAsP), die sich aus drei bzw. vier Komponenten zusammensetzen, und deren Eigenschaften sich in weiten Grenzen über den Mischungsanteil der einzelnen Komponenten gezielt verändern lässt.

Neue Entwicklungen beinhalten zweidimensionale Halbleitermaterialien wie z.B. Graphen. Graphen ist ein besonderes Halbleitermaterial, indem es keine klassische Band-

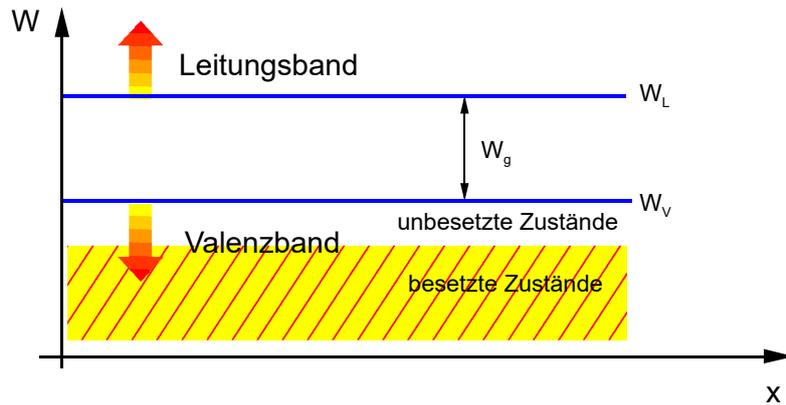


Abbildung 2.3: Bänderdiagramm für ein monovalentes Metall

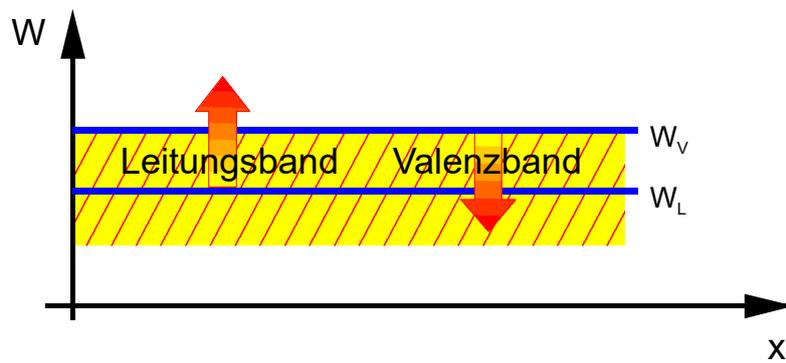


Abbildung 2.4: Bänderdiagramm eines Metalls mit überlappenden Bändern

lücke besitzt.

Neben kristallinen Halbleitermaterialien sind amorphe und polykristalline Halbleiter von technischer Bedeutung. Die weitere Betrachtung in dieser Vorlesung bezieht sich auf kristalline Halbleiter wie Silizium, um die Grundlagen des Elektronentransports und der Statistik der Elektronen herzuleiten.

2.1.3 Metalle

Die Metalle unterscheiden sich grundsätzlich von Halbleitern und Isolatoren dadurch, dass besetzte und freie Zustände nicht durch eine Energielücke voneinander getrennt sind. Im Fall der monovalenten Metalle ist das Valenzband nur etwa zur Hälfte gefüllt (vgl. Abb. 2.3).

Freie Plätze für Elektronen, die eine beliebige (kleine) Bewegungsenergie aufgenommen haben, finden also noch im Valenzband ihren neuen Platz, d.h. Stromtransport

kann stattfinden.

Dies gilt sogar für alle im Valenzband befindlichen Elektronen, sodass im Vergleich zu den außerordentlich wenigen Elektronen im reinen Halbleiter, die die Energielücke überspringen können, eine um Größenordnungen höhere Elektronenkonzentration am Stromtransport teilnehmen kann, und daher auch eine entsprechend höhere elektrische Leitfähigkeit resultiert. Es gibt darüber hinaus einen anderen Typ Metall, bei dem das Leitungs- und das Valenzband überlappen (vgl. Abb. 2.4).

Ein Vergleich der drei Typen von Festkörpern Isolator, Halbleiter und Metall liefert also:

1. Isolatoren und Halbleiter unterscheiden sich hinsichtlich der elektronischen Leitfähigkeit nicht qualitativ, sondern nur quantitativ in Form der Breite der Energielücke. Für beide ist die Zahl der Elektronen im Leitungsband bzw. die Zahl der freien Plätze im Valenzband maßgebend für die elektrische Leitfähigkeit. Im Vergleich zur Gesamtzahl der besetzbaren Zustände im Leitungsband bzw. der besetzten Zustände im Valenzband sind diese verschwindend gering. Daraus resultiert die niedrige Leitfähigkeit.
2. Bei den Metallen tragen nahezu alle Valenzelektronen zur Leitfähigkeit bei, weil im Valenzband etwa ebenso viele freie wie besetzte Plätze und etwa so viele Valenzelektronen wie Atome vorhanden sind. Abb. 2.5 zeigt im logarithmischen Maßstab einige Beispiele des spezifischen Widerstandes (das ist der Kehrwert der Leitfähigkeit) von Metallen bis zu Isolatoren

2.2 Konzentration von Elektronen und Löcher in Festkörpern

Nach der qualitativ ausgerichteten Einführung der Festkörper im vorhergehenden Kap. 2.1 soll jetzt die quantitative Herleitung der Ladungsträgerkonzentration im Festkörper erfolgen. Hierzu ermitteln wir zunächst die Anzahl der möglichen Zustände in einem Energiebereich – die sogenannte **Zustandsdichte**. Im nächsten Schritt bestimmen wir, wieviele dieser Zustände besetzt sind, in Abhängigkeit von der Gesamtenergie des Systems. Dies führt uns auf den Begriff der **Verteilungsfunktion**.

Die Zustandsdichte $g(W)$ und die Fermi-Dirac-Verteilungsfunktion $f(W)$ ergeben die

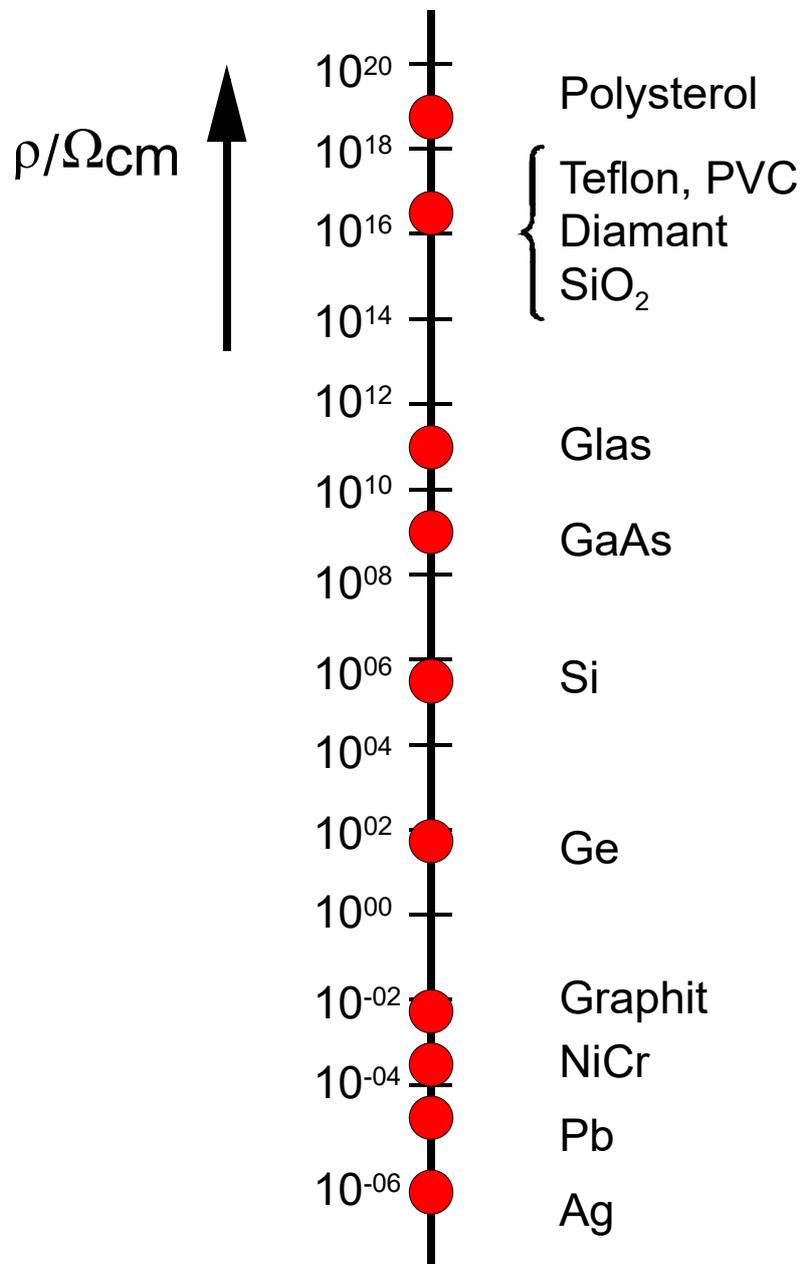


Abbildung 2.5: spezifischer Widerstand einiger Festkörper

Ladungsträgerkonzentration gemäß dem Zusammenhang

$$n(W) = g(W) \cdot f(W) \quad (2.1)$$

Dabei gibt die Zustandsdichte $g(W)$ die Dichte der möglichen besetzbaren Zustände pro Energieintervall und $f(W)$ die Verteilung der tatsächlich besetzten Zustände auf der Energieachse an.

Uns interessiert die Gesamtdichte n der Elektronen, die zur Leitfähigkeit beiträgt. In (2.1) gibt die Elektronenkonzentration $n(W)$ nur die Konzentration der Elektronen, die exakt die Energie W besitzen. Um die Gesamtzahl der Elektronen zu ermitteln, muss über die relevanten Energiebänder integriert werden, mit W als Integrationsvariable. Als zu integrierende Bänder kommen nur diese infrage, die nur zu einem Teil gefüllt sind: die vollständig gefüllten Bänder besitzen keine freien Plätze, die eine Aufnahme von kinetischer Energie einzelner Elektronen ermöglichen würden; und die komplett leeren Bänder besitzen keine Elektronen, die zur Leitfähigkeit beitragen können. Im Fall der Halbleiter müssen daher nur das Valenzband und das Leitungsband betrachtet werden.

Die Gesamtzahl der Elektronen, die in den Bändern zur Leitfähigkeit beitragen, können wir also wie folgt schreiben

$$n = \int_{\text{Band}} n(W) dW = \int_{\text{Band}} g(W) f(W) dW \quad (2.2)$$

Zustandsdichte $g(W)$

Zur Berechnung der Zustandsdichte beschränken wir uns auf die Bereiche in unmittelbarer Nähe der Bandkanten, weil beim Halbleiter gerade die Elektronen an der Unterkante des Leitungsbandes bzw. an der Oberkante des Valenzbandes die Leitfähigkeit tragen (vgl. Abb. 2.2). In diesen Bereichen kann der $W(k)$ -Verlauf durch eine Parabel angenähert (vgl. Abb. 1.26) und die effektive Masse als konstant angesehen werden (vgl. Abb. 1.28). Somit gilt für die kinetische Energie des Elektrons (z. B. im Leitungsband):

$$W_{kin} = W(k) - W_{\text{Kante}} = \frac{1}{2}mv^2 = \frac{\hbar^2 k^2}{2m^*} = \frac{\hbar^2}{2m^*} (k_x^2 + k_y^2 + k_z^2) \quad \text{wobei } m^* \neq f(k) \quad (2.3)$$

Dieser Zusammenhang liefert den Zugang zur Dichte der erlaubten Zustände auf der

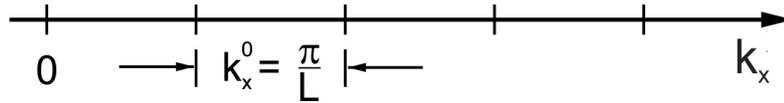


Abbildung 2.6: Erlaubte k -Werte in einem eindimensionalen Potentialtopf der Breite L

Energieachse, da er den Zusammenhang zum Wellenzahlvektor k enthält.

Aus Kap. 1.6 wissen wir, dass gebundene Teilchen in einem Potentialtopf nur diskrete erlaubte k -Werte besitzen (stehende Wellen im Potentialtopf). Nehmen wir jetzt an, die Elektronen im Leitungsband seien "eingesperrt" in einem würfelförmigen Volumen mit Kantenlänge L

$$V = L_x \cdot L_y \cdot L_z \equiv L^3 \quad (2.4)$$

In diesem Bereich können sich die Elektronen quasi frei bewegen, da sie sich im Leitungsband befinden. Die erlaubten k -Werte sind nach Kap. 1.6 mit (1.75) bei eindimensionaler Rechnung:

$$k_{x,n} = n \cdot \frac{\pi}{L} = n \cdot k_{x,0} \quad \text{mit} \quad k_{x,0} = \frac{\pi}{L} \quad \text{und} \quad n = \{1, 2, 3, \dots\} \quad (2.5)$$

pro Länge k_x^0 auf der k -Achse gibt es einen erlaubten k -Wert:

$$N_{\text{abs}}^{(1)} = \frac{1}{k_x^0} = \frac{L}{\pi} \quad (2.6)$$

Diese "Dichteäuf der k -Achse nennen wir absolute Zustandsdichte $N_{abs}^{(1)}$. In zwei Dimensionen muss nicht nur k_x , sondern auch k_y die Bedingung (2.5) erfüllen. Die erlaubten k_x - und k_y -Werte sind in Abb. 2.7 als Punkte wiedergegeben.

Pro Fläche $k_x^0 \cdot k_y^0$ der k -Ebene gibt es einen erlaubten k -Wert:

$$N_{abs}^{(2)} = \frac{1}{k_x^0 \cdot k_y^0} = \left(\frac{L}{\pi}\right)^2 \quad (2.7)$$

Dieses Verfahren lässt sich in drei Dimensionen fortsetzen: Pro Volumen $k_x^0 \cdot k_y^0 \cdot k_z^0$ des

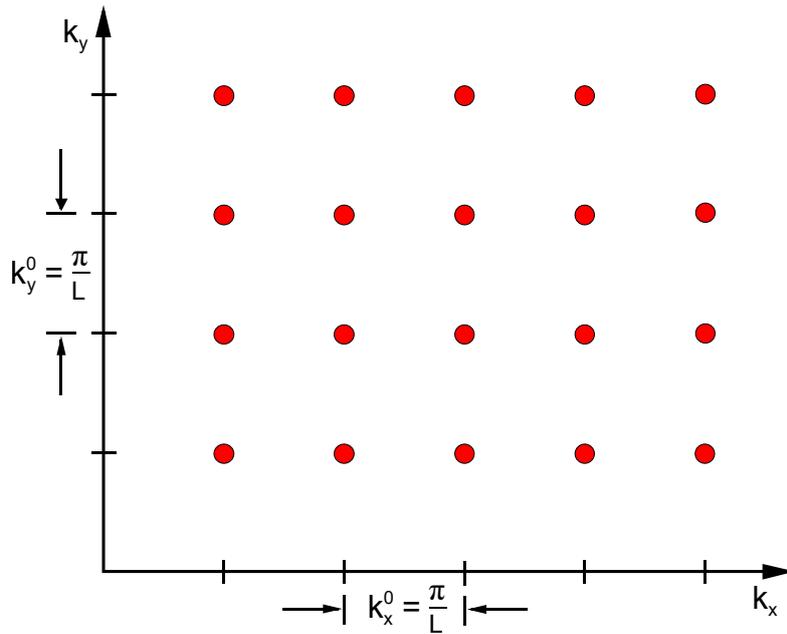


Abbildung 2.7: Erlaubte k -Werte im zweidimensionalen Potentialtopf mit den Abmessungen L^2

k -Raumes gibt es einen erlaubten Wert:

$$N_{abs}^{(3)} = \frac{1}{k_x^0 \cdot k_y^0 \cdot k_z^0} = \left(\frac{L}{\pi}\right)^3 \quad (2.8)$$

Das Volumen L^3 wurde zu Beginn willkürlich gewählt. Es ist üblich, die absoluten Zustandsdichten N_{abs} auf die Länge, Fläche oder das Volumen im Ortsraum zu beziehen. So gelangt man zu den relativen Zustandsdichten.

$$\begin{aligned} N^{(1)} &= \frac{N_{abs}^{(1)}}{L} = \frac{1}{\pi} \\ N^{(2)} &= \frac{N_{abs}^{(2)}}{L^2} = \frac{1}{\pi^2} \\ N^{(2)} &= \frac{N_{abs}^{(3)}}{L^3} = \frac{1}{\pi^3} \end{aligned} \quad (2.9)$$

Dies sind die Zustandsdichten im k -Raum, d.h. im Impulsraum. Uns interessieren allerdings die Anzahl der Zustände als Funktion der Energie. Über (2.3) lassen sich die Zustandsdichten $N^{(d)}$ in eine auf die Energie bezogenen Zustandsdichte umrechnen.

Wir wollen dies gleich für den real auftretenden dreidimensionalen Fall durchführen. Das Ziel ist, die Anzahl der erlaubten Zustände dn pro Energieintervall dW_{kin} zu bestimmen. Diese Größe stellt unsere gesuchte Zustandsdichte $g(W)$ aus (2.1) dar:

$$g(W) = 2 \cdot \frac{dn}{dW_{kin}} = 2 \cdot \frac{\partial n}{\partial k} \cdot \frac{\partial k}{\partial W_{kin}} \quad (2.10)$$

Der Faktor 2 resultiert aus der Tatsache, dass jeder Zustand mit 2 Elektronen besetzbar ist, die sich lediglich im Spin unterscheiden (vgl. Kap. 1.6). Im dreidimensionalen k -Raum stellen wegen des quadratischen Zusammenhangs $W \propto k^2$ (2.3) die Kugelflächen um den Ursprung Flächen konstanter Energie dar, die in Abb. 2.8 gezeigte Energiefläche.

Die Kugelschale der Dicke dk im k -Raum hat das Volumen

$$dV_k^* = 4\pi k^2 dk \quad (2.11)$$

Nach (2.5) können n und somit auch k_x , k_y und k_z nur positive Werte annehmen, so dass nur ein Oktant der Kugel zugelassen ist. Somit ist

$$dV = \frac{1}{8} dV_k^* = \frac{1}{2} \pi k^2 dk \quad (2.12)$$

Mit Kenntnis der Dichte der Zustände im k -Raum kann die Anzahl der Zustände in dieser Kugelschale bestimmt werden:

$$dn = dV_k N^{(3)} \quad (2.13)$$

Die dreidimensionale Zustandsdichte $N^{(3)}$ ist mit (2.9) auch bereits bekannt. Damit folgt:

$$dn = \frac{1}{2} \pi k^2 dk \frac{1}{\pi^3} \quad (2.14)$$

Somit ist unter Verwendung von (2.3)

$$\frac{\partial n}{\partial k} = \frac{1}{\pi^3} \cdot \frac{1}{2} \pi k^2 = \frac{1}{\pi^3} \cdot \frac{1}{2} \pi \cdot \frac{2W_{kin} m^*}{\hbar^2} \quad (2.15)$$

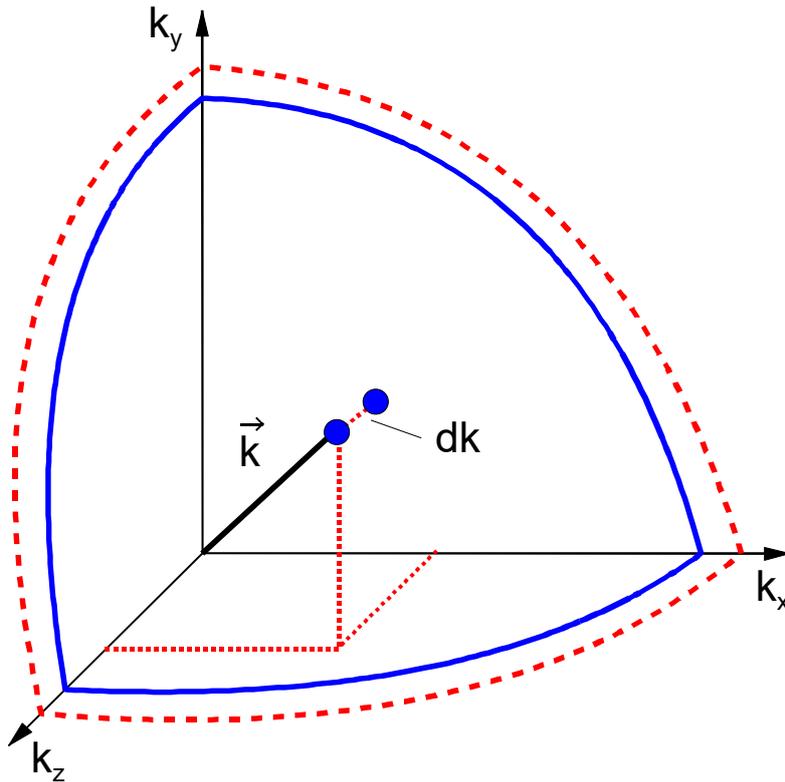


Abbildung 2.8: Flächen konstanter Energie für einen 3-dimensionalen Kristall

Der zweite in (2.10) benötigte Differentialkoeffizient ist

$$\frac{\partial k}{\partial W_{kin}} = \sqrt{\frac{m^*}{2\hbar^2 W_{kin}}} \quad (2.16)$$

erneut unter Verwendung von (2.3), sodass schließlich für die Zustandsdichte $g(W)$ auf der Energieskala geschrieben werden kann

$$g(W) = 2 \cdot \frac{dn}{dW_{kin}} = 2 \cdot \frac{\partial n}{\partial k} \cdot \frac{\partial k}{\partial W_{kin}} = \frac{\sqrt{2} \cdot (m^*)^{3/2}}{\pi^2 \hbar^3} \cdot \sqrt{W_{kin}} \quad (2.17)$$

Die Zustandsdichte im Leitungsband nimmt also von der Leitungsbandunterkante, beginnend mit dem Wert Null, mit der Quadratwurzel der Energie zu. Die durchgeführte Rechnung galt aber exakt nur für Energiebereiche nahe der Bandkante, für die die effektive Masse m^* als konstant angesetzt werden darf.

Abb. 2.9a zeigt den Verlauf der Zustandsdichte für die vereinfachende Annahme, dass mit zwei unterschiedlichen effektiven Massen, die für die Bandunterkante (m_1^*) bzw.

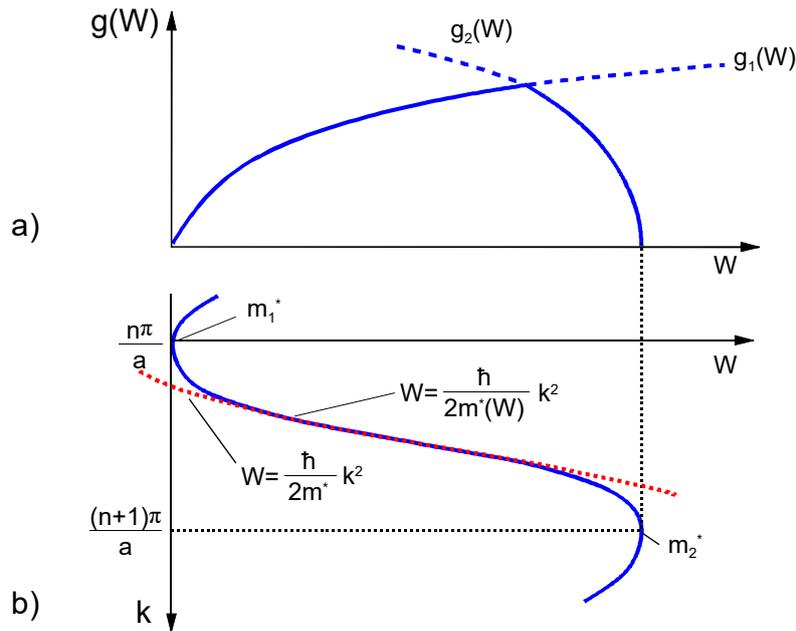


Abbildung 2.9: (a) Zustandsdichten $g_1(W)$ und $g_2(W)$ berechnet für konstante effektive Massen m_1^* bzw. m_2^* , und (b) zugehöriges Banddiagramm

die Bandoberkante (m_2^*) gelten. Abb. 2.9b zeigt das dazugehörige Banddiagramm für die energieabhängige effektive Masse und für die Näherung. Für die Unterkante des Leitungsbandes gilt dann in Anlehnung an (2.17):

$$g_L(W) = \frac{\sqrt{2}}{\pi^2 \hbar^3} (m^*)^{3/2} \cdot \sqrt{W - W_L} \quad \text{für } W > W_L \quad (2.18)$$

und für die Oberkante des Valenzbandes:

$$g_V(W) = \frac{\sqrt{2}}{\pi^2 \hbar^3} (m^*)^{3/2} \cdot \sqrt{W_V - W} \quad \text{für } W < W_V \quad (2.19)$$

mit der Energie der Leitungsbandkante W_L bzw. der Valenzbandkante W_V .

2.2.1 Fermi-Energie W_F für monovalentes Metall

Nehmen wir an, wir haben einen Festkörper mit bekannter Elektronenkonzentration und interessieren uns für den Energiewert, der beim lückenlosen Auffüllen der besetzbaren Zustände mit Elektronen unter Beachtung des Pauli-Prinzips erreicht wird. Dies ist die anschauliche Erklärung der Fermi-Energie W_F . Eine allgemeingültige Definition

folgt im nächsten Kapitel.

Besonders einfach liegen die Verhältnisse für ein monovalentes Metall, dessen Valenzband etwa halb besetzt ist (vgl. Abb. 2.3) und dessen Elektronenkonzentration $n = 5 \times 10^{22} \text{ cm}^{-3}$ bekannt ist. Da das Auffüllen lückenlos erfolgen soll, gilt für die Verteilungsfunktion (genauer erklärt im folgenden Kapitel):

$$f(W) = 1 \quad \text{für} \quad W < W_F \quad (2.20)$$

Dann können wir gemäß (2.2) schreiben:

$$n = \int_0^{W_F} g(W) \cdot 1 dW \quad (2.21)$$

mit $g(W)$ aus (2.17) und $m^* = m_0$ gilt:

$$n = \int_0^{W_F} \frac{\sqrt{2}}{\pi^2 \hbar^3} m_0^{3/2} \sqrt{W} dW = \frac{2}{3} \frac{\sqrt{2}}{\pi^2 \hbar^3} m_0^{3/2} W^{3/2} \Big|_0^{W_F} \quad (2.22)$$

und aufgelöst nach W_F :

$$W_F = (3n\pi^2)^{(2/3)} \cdot \frac{\hbar^2}{2m_0} \quad (2.23)$$

Setzt man die Werte für n und m_0 (Elektronenmasse) ein, so ergibt sich ein Wert von

$$W_F \approx 4.94 \text{ eV}$$

Damit ergibt sich mit der aus der Wärmelehre bekannten Beziehung $W = k \cdot T_F$ eine Fermi-Temperatur

$$T_F \approx 57\,308 \text{ K}$$

Auf diese extrem hohe Temperatur müsste man das Metall aufheizen, um ein Elektron durch Zufuhr von thermischer Energie von der Valenzbandunterkante auf die Fermi-Energie anzuheben. Auf die Besetzung der Zustände im monovalenten Metall hat daher offensichtlich die Temperatur im Bereich bis zum Schmelzpunkt (etwa 1000 K bis 3000 K) kaum einen Einfluss. Das bedeutet, dass die elektronischen Eigenschaften der

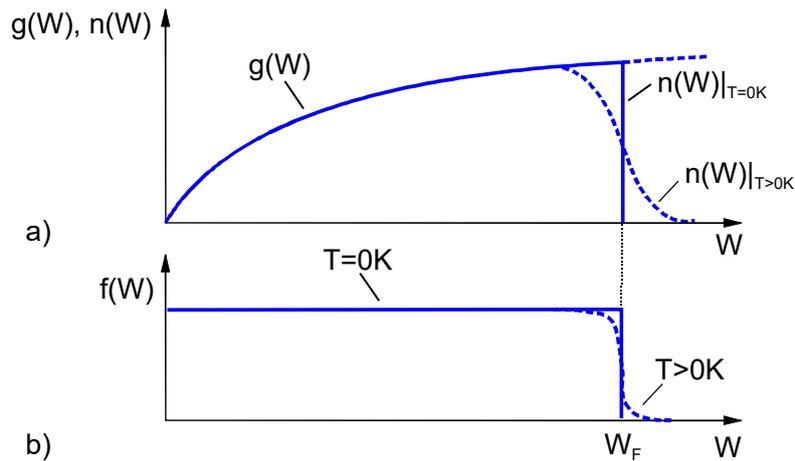


Abbildung 2.10: (a) Zustandsdichte und Elektronenverteilung in einem Metall ($T = 0$ K) und (b) daraus abgeleitete Verteilungsfunktion

Metalle kaum temperaturabhängig sind. Bei Halbleitern ist dies grundsätzlich anders!

2.2.2 Die Fermi-Dirac-Verteilungsfunktion $f(W)$ am Beispiel des monovalenten Metalls)

Die im letzten Kapitel untersuchten Verhältnisse sind in Abb. 2.10 dargestellt, und zwar sowohl für $T = 0$ K als auch für $T > 0$ K.

Für $T = 0$ K bildet die Fermi-Energie eine scharfe Grenze zwischen besetzten und unbesetzten Zuständen:

- alle Zustände unterhalb von W_F sind besetzt, d.h. die Verteilungsfunktion hat den Wert $f(W) = 1$
- alle Zustände oberhalb von W_F sind unbesetzt, d.h. $f(W) = 0$

Wird die Temperatur erhöht, so kann auf einige Elektronen thermische Energie übertragen werden. Als Folge davon nehmen einige Elektronen Energiewerte $W > W_F$ an und hinterlassen bei $W < W_F$ freie, unbesetzte Plätze. Die Wahrscheinlichkeit, besetzte Plätze oberhalb von W_F und unbesetzte Plätze unterhalb von W_F zu finden, wächst also mit der Temperatur. Diese Wahrscheinlichkeit wird durch die Fermi-Dirac-Verteilungsfunktion $f(W)$ beschrieben. Nach der Quantenstatistik gilt (ohne Herlei-

tung):

$$f(W) = \frac{1}{e^{\frac{W-W_F}{kT}} + 1} \quad (2.24)$$

Diese Funktion liefert auch eine allgemeine, d.h. auch für alle Temperaturen gültige Definition der Fermi-Energie:

$$f(W_F) = \frac{1}{2} \quad (2.25)$$

2.2.3 Besetzungsstatistik in Halbleitern, Eigen- und Störstellenhalbleiter

Wir werden uns jetzt auf den Festkörper **Halbleiter** konzentrieren, da aus diesem Material die technisch wichtigsten elektronischen Bauelemente hergestellt werden. Man unterscheidet in dieser Klasse die **Eigenhalbleiter**, das sind solche ohne Störstellen und Verunreinigungen, und die **Störstellenhalbleiter**, die meist gezielt mit Fremdatomen "dotiert" sind.

Eigenhalbleiter (intrinscher Halbleiter)

Wie bereits bekannt, besitzen Halbleiter eine im Vergleich zu Isolatoren kleinere Lücke W_g zwischen Leitungs- und Valenzband, so dass bei Raumtemperatur (300 K) schon einige Elektronen durch Zufuhr thermischer Energie aus dem Valenz- ins Leitungsband angehoben werden. Dann liegen zum einen freie Elektronen im Leitungsband vor, und zum anderen unbesetzte Plätze im Valenzband, die man als Löcher, also als Ladungsträger mit positiver Ladung interpretieren kann. Da Elektronen und Löcher im Eigenhalbleiter immer als Paar erzeugt werden, ist deren Konzentration n (Elektronen) und p (Löcher) gleich:

$$n = p = n_i \quad (2.26)$$

(3.23) Hier ist n_i die intrinsische oder Eigenleitungskonzentration. Die soeben beschriebenen Verhältnisse erfordern, dass die Verteilungsfunktion im Valenzband Werte < 1 und im Leitungsband Werte > 0 annimmt. Somit muss die Fermi-Energie in der Bandlücke zwischen W_L und W_V liegen.

Elektronenkonzentration im Leitungsband

Nach (2.2) gilt:

$$n = \int_{\text{LB}} n(W) dW = \int_{\text{LB}} g_L(W) \cdot f(W) dW \quad (2.27)$$

Für die Verteilungsfunktion kann, da die Fermi-Energie W_{Fi} im intrinsischen Halbleiter um mehrere kT unterhalb der Leitungsbandkante liegt, die folgende Boltzmann-Näherung benutzt werden:

$$f_L(W) = \frac{1}{e^{\frac{W-W_{Fi}}{kT}} + 1} \approx e^{-\frac{W-W_{Fi}}{kT}} \quad \text{für } W_L - W_{Fi} \gg kT \quad (2.28)$$

Unter Verwendung von (2.18) folgt:

$$n = \frac{\sqrt{2}}{\pi^2 \hbar^3} (m_n^*)^{3/2} \int_{\text{LB}} \sqrt{W - W_L} \cdot e^{-\frac{W-W_{Fi}}{kT}} dW \quad (2.29)$$

Die Integration ist von der Unterkante bis zur Oberkante des Leitungsbandes durchzuführen. Man macht aber keinen großen Fehler, wenn man die Integrationsobergrenze ins Unendliche verlegt, weil an der Leitungsbandoberkante die Verteilungsfunktion praktisch den Wert Null hat und somit die Integration für Werte oberhalb des Leitungsbandes keine Beiträge liefert. Mit diesem "Trick" kann aber manchmal ein analytisch geschlossener Ausdruck für das Integral gegeben werden.

Verwendet man:

$$W - W_{Fi} = (W - W_L) + (W_L - W_{Fi})$$

so erhält man:

$$n = \frac{\sqrt{2}}{\pi^2 \hbar^3} (m_n^*)^{3/2} \cdot e^{-\frac{W_L - W_{Fi}}{kT}} \cdot \int_{\text{LB}} \sqrt{W - W_L} \cdot e^{-\frac{W - W_L}{kT}} dW \quad (2.30)$$

Das sogenannte Fermi-Dirac-Integral hat den Wert $\sqrt{\pi}/2$, sodass resultiert:

$$n = N_L \cdot \exp\left(-\frac{W_L - W_{Fi}}{kT}\right) \quad (2.31)$$

mit

$$N_L = \frac{1}{\sqrt{2}\hbar^3} \left(\frac{m_n^* kT}{\pi} \right)^{3/2} \quad (2.32)$$

(Hinweis: Für die Durchführung der Integration wurde die Substitution $x = \frac{W-W_L}{kT}$ und damit $dx = \frac{1}{kT} dW$ benutzt. Daraus folgt der zusätzliche Faktor $(kT)^{3/2}$ in (2.32)).

N_L wird als effektive Zustandsdichte des Leitungsbandes bezeichnet. Sie gibt die Konzentration der Elektronen im Leitungsband an, die vorliegt, wenn die Fermi-Energie mit der Leitungsbandkante zusammenfällt. Diese Interpretation wird sich bei der Behandlung des Störstellenhalbleiters als nützlich erweisen.

Sowohl n als auch N_L sind Konzentrationen und werden in Einheiten cm^{-3} angegeben.

Löcherkonzentration im Valenzband

Analog zu (2.2) gilt:

$$p = \int_{\text{VB}} p(W) dW = \int_{\text{VB}} g_V(W) \cdot [1 - f(W)] dW \quad (2.33)$$

Statt $f(W)$ tritt hier für die Besetzung der Niveaus der Ausdruck $[1 - f(W)]$ auf, da die Löcher fehlende Elektronen an der Oberkante des Valenzbandes darstellen: Sind keine Löcher vorhanden, sind also alle Plätze im Valenzband mit Elektronen besetzt, so ist $f(W) = 1$ und $[1 - f(W)] = 0$.

Unter Verwendung von (2.24) gilt:

$$1 - f(W) = \frac{\exp\left(-\frac{W_{Fi}-W}{kT}\right)}{\exp\left(-\frac{W_{Fi}-W}{kT}\right) + 1} \quad \text{für } W \leq W_V \quad (2.34)$$

Wegen $W_{Fi}-W \gg kT$ (die Fermi-Energie liegt mehrere kT oberhalb des Valenzbandes) ist die Exponentialfunktion im Nenner gegenüber 1 vernachlässigbar, so dass für die Verteilungsfunktion der Löcher im Valenzband gilt:

$$f_V(W) = 1 - f(W) = \exp\left(-\frac{W_{Fi}-W}{kT}\right) \quad (2.35)$$

und damit:

$$p = \frac{\sqrt{2}}{\pi^2 \hbar^3} (m_p^*)^{3/2} \cdot \int_{\text{LB}} \sqrt{W_V - W} \cdot e^{-\frac{W_{Fi} - W}{kT}} dW \quad (2.36)$$

Mit ähnlichen Umformungen und Substitutionen wie für die Elektronenkonzentration und durch die analoge Verschiebung der unteren Integrationsgrenze nach $-\infty$ folgt:

$$p = N_V \cdot \exp\left(-\frac{W_{Fi} - W_V}{kT}\right) \quad (2.37)$$

mit der effektiven Zustandsdichte des Valenzbandes

$$N_V = \frac{1}{\sqrt{2} \hbar^3} \left(\frac{m_p^* kT}{\pi}\right)^{3/2} \quad (2.38)$$

Auch p und N_V sind Konzentrationen mit der Einheit cm^{-3} .

Eigenleitungskonzentration n_i

Nach (2.26) erhält man in Anlehnung an ähnliche Sachverhalte in der Chemie das Massenwirkungsgesetz für Elektronen und Löcher im Halbleiter:

$$\sqrt{n \cdot p} = n_i \quad (2.39)$$

Mit (2.31) und (2.37) folgt:

$$n_i = \sqrt{N_L N_V} \cdot e^{-\frac{W_L - W_V}{2kT}} = \sqrt{N_L N_V} \cdot e^{-\frac{W_g}{2kT}} \quad (2.40)$$

mit

$$\sqrt{N_L N_V} = \frac{1}{\sqrt{2} \hbar^3} \left[\sqrt{m_n^* m_p^*} \cdot \frac{kT}{\pi} \right]^{3/2} \equiv A \cdot T^{3/2}$$

Die effektiven Massen m_n^* und m_p^* sind nur schwach von der Temperatur abhängig. Wir definieren daher eine von der Temperatur T (näherungsweise) unabhängige Abkürzung A als

$$A = \frac{1}{\sqrt{2} \hbar^3} \left[\sqrt{m_n^* m_p^*} \cdot \frac{k}{\pi} \right]^{3/2}$$

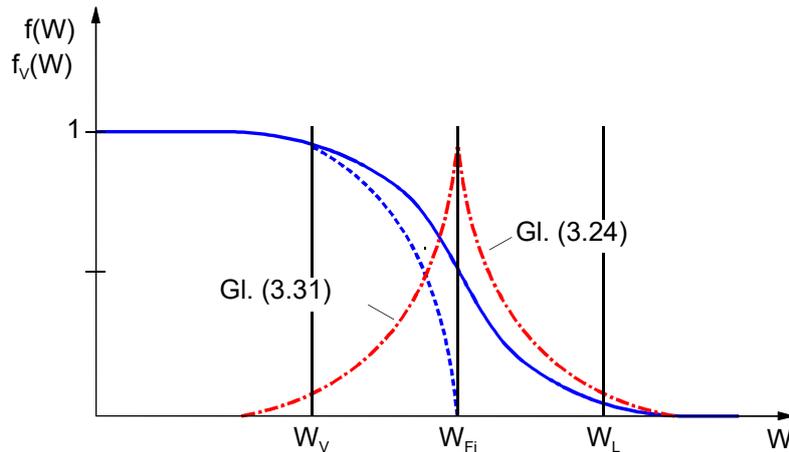


Abbildung 2.11: Verteilungsfunktionen für das Leitungs- und das Valenzband im intrinsischen Halbleiter; exakte Form und Näherungen

und erhalten für die intrinsische Ladungsträgerdichte n_i im undotierten Halbleiter somit einen Ausdruck, der neben einer charakteristischen Abhängigkeit von der Temperatur auch eine exponentielle Abhängigkeit von der Bandlücke W_g besitzt:

$$n_i = A \cdot T^{3/2} \cdot \exp\left(-\frac{W_g}{2kT}\right) \quad (2.41)$$

Abb. 2.11 zeigt die Verteilungsfunktionen $f(W)$ für Elektronen und $f_v(W)$ für Löcher in exakter Form und mit den Näherungen, die durch (2.28) und (2.35) wiedergegeben sind.

Die berechneten Konzentrationen der Elektronen und Löcher stellen gemäß (2.2) und (2.33) die Faltung der Verteilungsfunktionen (Abb. 2.11) mit den Zustandsdichten dar. Damit ergibt sich Abb. 2.12.

Die Ladungsträgerkonzentrationen $n = p = n_i$ hängen gemäß (2.41) von Temperatur und Bandlücke ab. Man erkennt, dass sowohl mit steigender Temperatur als auch mit fallender Energielücke W_g die Ladungsträgerkonzentration exponentiell ansteigt. Beispielsweise für Raumtemperatur unterscheiden sich daher die n_i verschiedener Halbleiter erheblich (Tab. 2.1). Die Temperaturabhängigkeit dieser Halbleiter ist in Abb. 2.13 dargestellt.

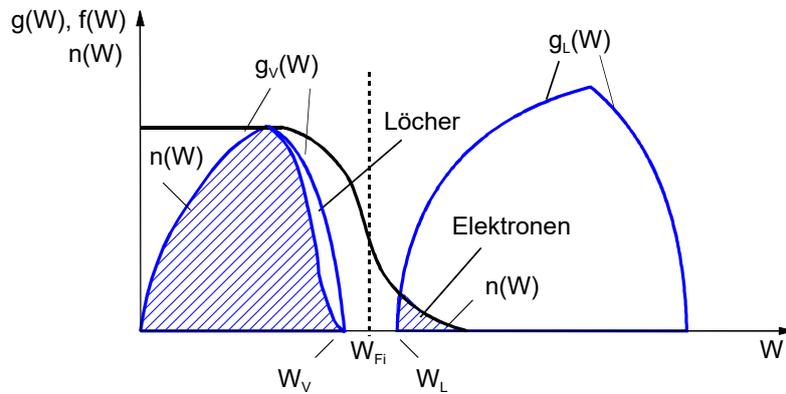


Abbildung 2.12: Zusammenhang zwischen Verteilungsfunktion f , Zustandsdichte g und Ladungsträgerkonzentrationen n und p für einen intrinsischen Halbleiter bei $T > 0$ K

Tabelle 2.1: Daten einiger Festkörper bei $T = 300$ K

	m_n^*/m_0	m_p^*/m_0	W_g/eV	n_i/cm^{-3}	N_L/cm^{-3}	N_V/cm^{-3}	ϵ_r
Ge	0.55	0.37	0.66	2.4×10^{13}	1×10^{19}	6×10^{18}	16.0
Si	0.98	0.49	1.1	1.4×10^{10}	3×10^{19}	1.5×10^{19}	11.9
GaAs	0.067	0.45	1.42	2.5×10^6	4.7×10^{17}	7×10^{18}	13.1
InAs	0.022	0.41	0.35	2.4×10^{15}	2.3×10^{17}	1.8×10^{19}	12.5
InP	0.078	0.56	1.35	6.6×10^6	5.4×10^{17}	2.9×10^{18}	12.4
InGaAs	0.045	0.43	0.76	1.7×10^{12}	6.8×10^{17}	2.0×10^{19}	12.8
4H-SiC							
GaN							
Diamant							

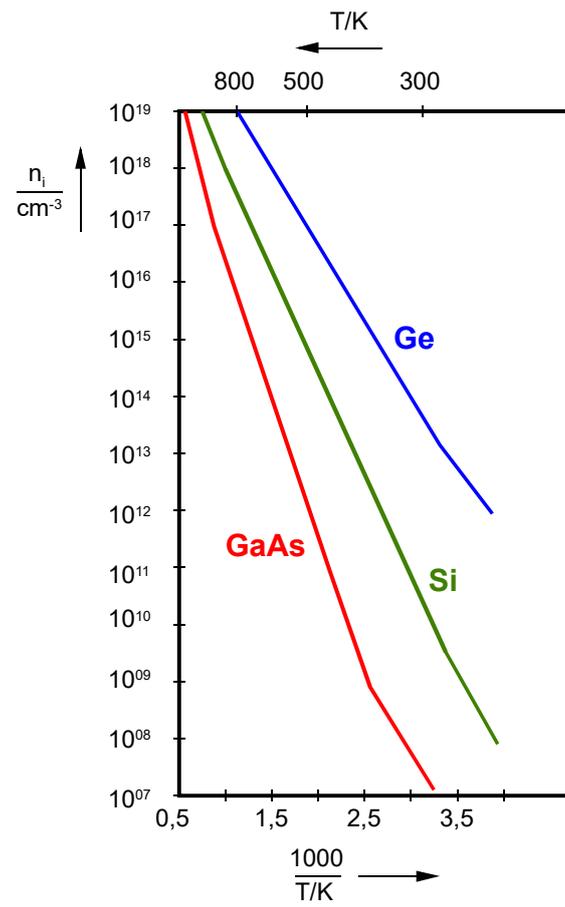


Abbildung 2.13: Temperaturabhängigkeit der Eigenleitungskonzentration

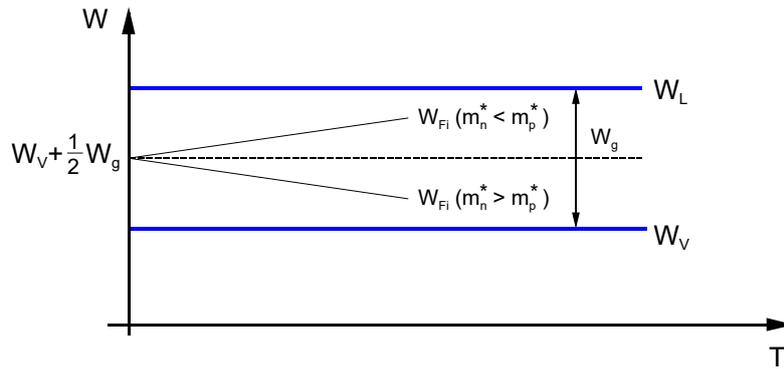


Abbildung 2.14: Temperaturabhängigkeit der Fermi-Energie im intrinsischen Halbleiter

2.2.4 Fermi-Energie W_{Fi} und Temperaturabhängigkeit

Im intrinsischen Halbleiter gilt wegen der paarweisen Erzeugung von Elektronen und Löchern $n = p$ für alle Temperaturen. (2.31) und (2.37) stellen den Zusammenhang mit der Fermi-Energie W_{Fi} her:

$$N_L \cdot \exp\left(-\frac{W_L - W_{Fi}}{kT}\right) = N_V \exp\left(-\frac{W_{Fi} - W_V}{kT}\right)$$

Einsetzen der Größen N_L und N_V aus (2.32) und (2.38) liefert:

$$W_{Fi} = W_V + \frac{1}{2}W_g + \frac{3}{4}kT \ln\left(\frac{m_p^*}{m_n^*}\right) \quad (2.42)$$

Dabei wurde davon Gebrauch gemacht, dass $W_L = W_V + W_g$ gilt. Aus (2.42) entnimmt man, dass die Fermi-Energie immer in der Bandlückenmitte liegt:

$$W_{Fi} = W_V + \frac{1}{2}W_g \quad \text{für } m_n^* = m_p^* \text{ und alle Temperaturen } T$$

Für $m_n^* \neq m_p^*$ ergibt sich über der Temperatur das Verhalten nach Abb. 2.14. Dabei wurde die Temperaturabhängigkeit der Bandlückenenergie W_g , die noch nicht angesprochen wurde, aber real existiert, nicht berücksichtigt. Das Verhalten nach Abb. 2.14 widerspricht keineswegs der Eigenschaft $n = p$, sondern spiegelt den Zusammenhang der effektiven Massen mit den Zustandsdichten wieder: Bei einem typischen $g(W)$ -Verlauf nach Abb. 2.15 ist $m_p^* > m_n^*$, da

$$g(W) \propto (m^*)^{3/2} \sqrt{W} \quad (2.43)$$

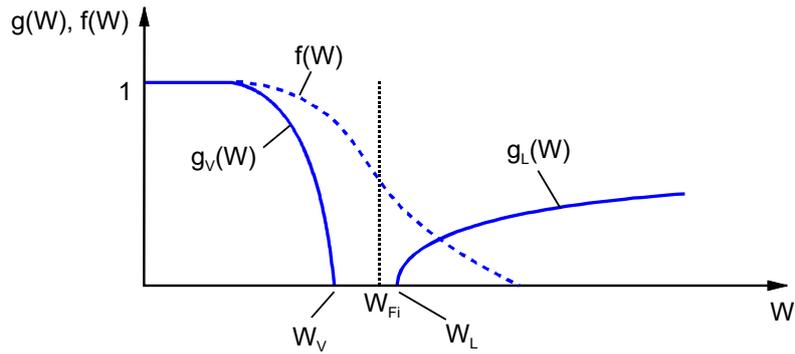


Abbildung 2.15: Temperaturabhängigkeit der Fermi-Energie

und somit mit der Energie für das Valenzband stärker ansteigt als für das Leitungsband. Für eine bestimmte Konzentration von Löchern im Valenzband wird daher nur ein geringerer Energiebereich benötigt als für Elektronen im Leitungsband. Diese Asymmetrie wird dadurch ausgeglichen, dass die Verteilungsfunktion und somit die Fermi-Energie zum Leitungsband hin verschoben wird, sodass die Faltung aus Zustandsdichte und Verteilungsfunktion für Valenz- und Leitungsband wieder gleiche Werte besitzen. Die Verschiebung der Fermi-Energie kommt besonders zum Tragen bei Halbleitern mit großem Unterschied zwischen effektiven Löcher- und Elektronenmassen, bei kleiner Bandlücke sowie bei hohen Temperaturen. Beispielsweise hat der Halbleiter InSb folgende Eigenschaften:

$$\text{InSb} m_p^*/m_n^* = 50, W_g = 0.16 \text{ eV}$$

Mit diesen Werten ist aus (2.42) einfach ersichtlich, dass bei Raumtemperatur die Fermi-Energie mit der Leitungsbandkante zusammenfällt:

$$W_{Fi} \approx W_g$$

Unter anderem hat das die Konsequenz, dass die Boltzmann-Näherung (2.28) nicht mehr zulässig ist für die Beschreibung des Halbleiters im Hinblick auf die Ladungsträgerkonzentrationen, so wie wir sie durchgeführt haben!

Ein Halbleiter, für den die Fermi-Energie $W_{Fi} > W_L$ (bzw. $W_{Fi} < W_V$) ist, heißt **entarteter Halbleiter**. Man beachte, dass diese Eigenschaft temperaturabhängig ist!

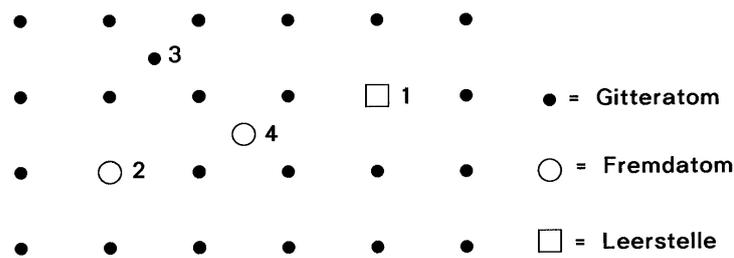


Abbildung 2.16: Punktdefekte im Festkörper

2.3 Störstellenhalbleiter

Der bisher behandelte Eigenhalbleiter stellt eine Idealisierung dar, die in reiner Form in der Natur nicht vorkommt und auch durch eine noch so ausgereifte und aufwendige Technologie nicht realisiert werden kann. Selbst wenn man Verunreinigungen durch Fremdatome völlig unterdrücken könnte (was nicht der Fall ist), ist aus thermodynamischen Gründen die exakte Einhaltung der periodischen Gitterstruktur nicht möglich. In den meisten technisch interessanten Anwendungen ist aber gerade ein gestörter Halbleiter erwünscht, der durch gezielte Zugabe von Fremdatomen (Dotierung) erzeugt werden kann. Beispielsweise können so n -Typ ($n \gg p$) oder p -Typ ($p \gg n$) Halbleiter hergestellt werden.

Störungen im Halbleiterkristall haben i.A. außerordentlich starke Einflüsse auf praktisch alle Festkörpereigenschaften, auch auf die elektrische Leitfähigkeit. Entsprechend der Geometrie und Ausdehnung der Kristallstörungen (Defekte) unterscheidet man nach punktförmigen, linienförmigen und dreidimensionalen Defekten.

2.3.1 Punktförmige Defekte

Denkbar sind die in Abb. 2.16 dargestellten Punktdefekte.

- (1) an einer **Leerstelle** fehlt ein Gitteratom im Kristallverband (engl. *vacancy*)
- (2) in einer substitutionellen Störstelle befindet sich ein Fremdatom anstelle eines Gitteratoms auf dem entsprechenden Gitterplatz, es wird also ein Gitteratom durch das Fremdatom ersetzt (engl. *substitutional defect*)
- (3) eine **Zwischengitterstörstelle** entsteht, wenn ein Kristallatom nicht an einem regulären Platz in den Verbund eingefügt wird, sondern an einem räumlich dazwi-

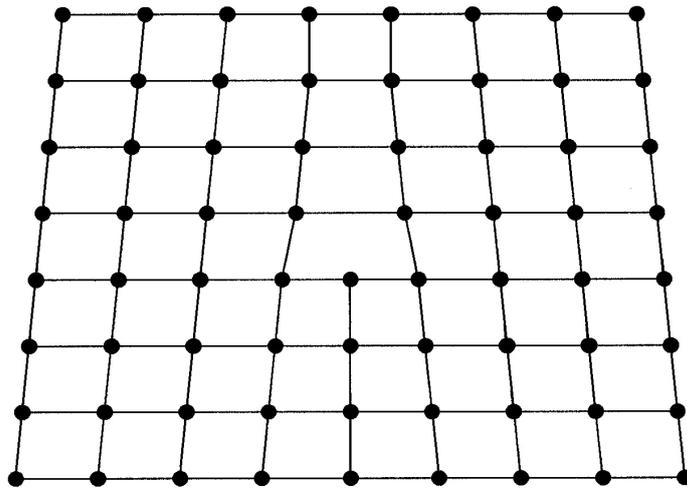


Abbildung 2.17: Versetzung im Kristall

schen gelegenen Platz (engl. *interstitial defect*)

- (4) auch Fremdatome können sich auf Zwischengitterplätzen befinden

Durch kontinuierliche Entwicklung der Kristallzucht-Technologie können heute Halbleiterwafer mit einer sehr geringen Defektdichte hergestellt werden. Auf $1 \times 10^{10} \dots 1 \times 10^{12}$ Atome kommt nur eine einzige punktförmige Störung. Bezogen auf das Volumen sind dies noch $1 \times 10^{10} \dots 1 \times 10^{12}$ Defekte pro cm^3 .

2.3.2 Liniendefekte, Versetzungen

Bei der Kristallzucht kann es vorkommen, dass von einem bestimmten Punkt an eine ganze Gitterebene nicht fortgesetzt wird (Abb. 2.17).

Eine von außen angelegte mechanische Verspannung, hervorgerufen z.B. durch Temperaturunterschiede, kann solche Versetzungen erzeugen. Eine einmal erzeugte Versetzung verursacht auch ihrerseits in ihrer Umgebung eine mechanische Verspannung, weil hier die Atomabstände nicht ihrer Gleichgewichtslage entsprechen. Entlang solcher Versetzungslinien sammeln sich bevorzugt Fremdatome an mit meist unerwünschten Folgen. Insofern versucht man grundsätzlich, die Zahl der Versetzungen möglichst gering zu halten. Typische Versetzungsdichten liegen bei $1 \times 10^3 \text{ cm}^{-2}$ bis $1 \times 10^4 \text{ cm}^{-2}$, bezogen auf die Oberfläche des Halbleiterkristalls.

2.3.3 Zwei- und dreidimensionale Defekte

In diese Klasse gehören alle anderen Defekte. Zu erwähnen sind hier insbesondere räumlich ausgedehnte Ansammlungen von Fremdatomen oder auch, bei Verbindungshalbleitern, von Atomen nur einer Sorte ("Clusterbildung").

Im Folgenden wollen wir uns genauer mit den **Punktdefekten** auseinandersetzen, weil diese, wie bereits erwähnt, besonders in Form von Fremdatomen eine hohe technische Bedeutung haben.

2.3.4 Dotierung von Halbleitern; Donatoren, Akzeptoren, tiefe Störstellen

Die Dotierung von Halbleitern mit Fremdatomen erzeugt in erster Linie punktförmige Defekte der Art (2) in Abb. 2.16, d.h. es werden Gitteratome durch Fremdatome ersetzt. Am Beispiel des Siliziums sollen die zwei wichtigen zu unterscheidenden Fälle verdeutlicht werden.

Silizium ist vierwertig, im idealen ungestörten Kristall ist jedes Siliziumatom durch kovalente Bindungen, bestehend aus jeweils zwei Elektronen, an vier benachbarte Atome gebunden, die räumlich in den vier Ecken eines Tetraeders liegen. Schematisch kann man diesen Sachverhalt in einer ebenen Zeichnung nach Abb. 2.18 bzw. 2.19 darstellen, in der die Atome vierfach koordiniert in einem Rechteckgitter sitzen. Die Anzahl der Bindungen ist in dieser Darstellung korrekt wiedergegeben, nicht aber ihre dreidimensionale Ausrichtung. Da die Elektronen die Bindung zwischen den Siliziumatomen bewirken, liegen sie hier fest und stehen für einen Leitungsvorgang nicht zur Verfügung (Valenzelektronen).

In den Abb. 2.18 bzw. 2.19 sind bereits Dotierstoffatome eingezeichnet, und zwar Arsen (As) in Abb. 2.18 und Bor in Abb. 2.19. Zunächst sei der Fall "Arsen in Silizium" behandelt.

Donatoren

Arsen ist fünfwertig, d.h. es hat fünf Bindungs- oder Valenzelektronen in seiner äußeren Schale und somit eines mehr als zur Absättigung der vier möglichen Bindungen im Siliziumgitter notwendig ist. Das fünfte Elektron des Arsen-Atoms ist daher nicht in eine Bindung eingebaut, sondern es kann sich (ziemlich) frei im Kristall bewegen, sodass es zur elektrischen Leitfähigkeit beitragen kann: Das Arsenatom stellt im Siliziumkristall ein quasi-freies Elektron zur Verfügung, es wirkt als **Donator** (Spender). Im Bänderdia-

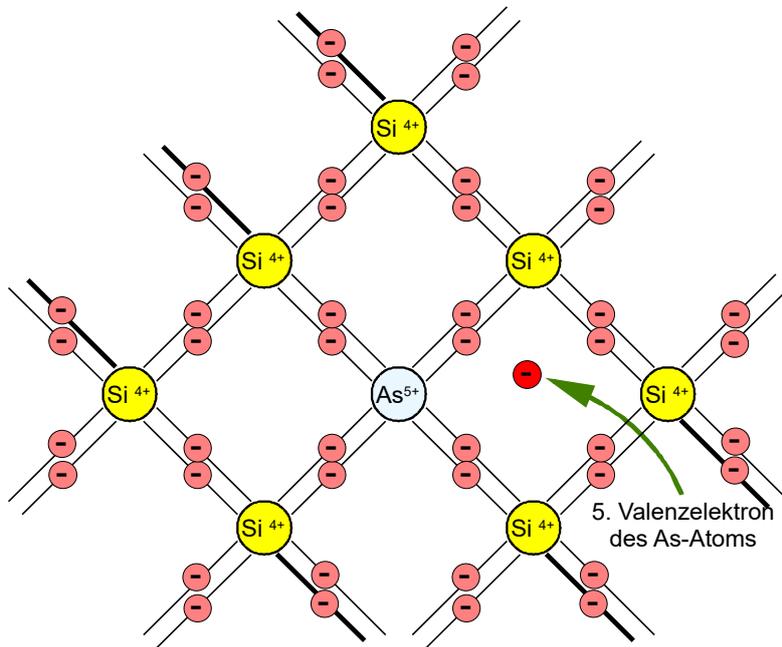


Abbildung 2.18: Donator Arsen in Silizium

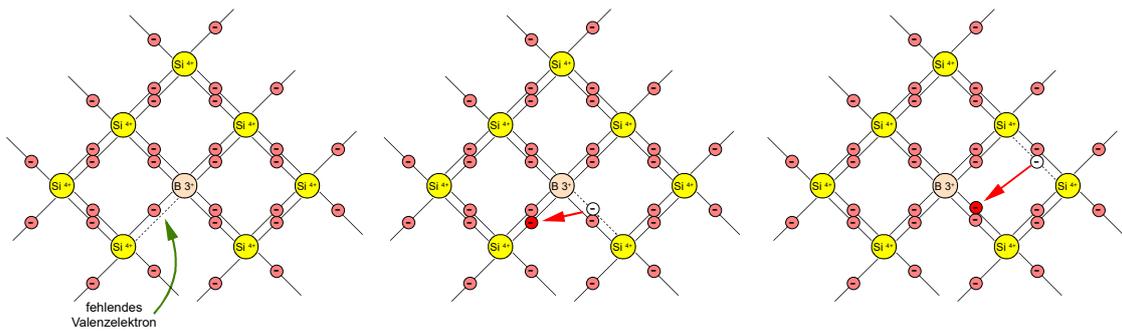


Abbildung 2.19: Akzeptor Bor in Silizium

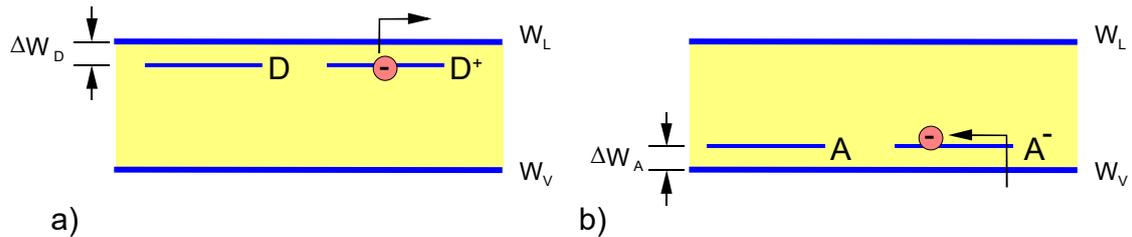


Abbildung 2.20: Dotierniveau im Banddiagramm: (a) Donator für n-Leitfähigkeit und (b) Akzeptor für p-Leitfähigkeit

gramm wird das Donatorverhalten folgendermaßen beschrieben: Zu den Elektronen, die das Valenzband füllen (vier pro Atom), wird das überzählige Elektron des Arsenatoms hinzugefügt. Wenn dieses vollkommen frei wäre, würde es das niedrigste Niveau des Leitungsbandes besetzen. Tatsächlich aber wird es von der einen positiven Netto-Ladung des Arsenatom-Rumpfes schwach angezogen (As_5^+ minus 4 Valenzen gibt eine positive Elementarladung). Die Anziehung zwischen Elektron und Donatorrumpf mit einer positiven Elementarladung können wir als eine schwache Bindungs- oder auch Ionisierungsenergie verstehen. Die Coulomb-Anziehung der beiden Ladungen ist abgeschwächt (s.u.), der resultierende Energiebetrag ist relativ klein. Im Banddiagramm liegen die Donatorzustände etwas unterhalb des Leitungsbandes, in einem Abstand, der der Ionisierungsenergie ΔW_D entspricht (Abb. 2.20).

Akzeptoren

In Abb. 2.19 wird in das Siliziumgitter ein dreiwertiges Atom (Bor) eingebaut. Im Vergleich zum Silizium fehlt ein Valenzelektron, sodass eine der vier Bindungen im Kristall nur mit einem Elektron statt mit den üblichen zwei Elektronen besetzt ist. In dieser Lücke, die einem Loch im Valenzband entspricht, kann leicht ein Elektron eingefangen werden, das aus Bindungen benachbarter Atome stammt. Das dreiwertige Bor-Atom nimmt in seiner Umgebung also ein zusätzliches Elektron auf, es wirkt als Akzeptor. Im Gegenzug fehlt ein Elektron in der Umgebung eines benachbarten Silizium-Atoms, da die Anzahl der Elektronen im Kristall erhalten ist.

Weitere Vorgänge können im Kristall stattfinden, in denen sich jeweils ein Elektron aus einem vollständigen Bindungspaar in diese Lücke umlagert. Das "fehlende" Elektron kann sich also durch den Kristall bewegen. Wir bezeichnen das fehlende Elektron als **Loch**. Der Leitungsmechanismus, getragen von Löchern im Valenzband, ist in

Abb. 2.19 angedeutet.

Im Banddiagramm kann das Akzeptorverhalten nach Abb. 2.20b dargestellt werden. Die geringe Energie, die benötigt wird, um ein Elektron aus einer benachbarten Bindung, d.h. aus dem Valenzband, in die "Bindungslücke" zu heben, ist die Ionisierungsenergie ΔW_A des Akzeptors.

Je nach Dotierung mit Donatoren oder Akzeptoren werden mehr Elektronen bzw. Löcher als im intrinsischen Halbleiter vorliegen, da die Ionisierungsenergie der Störstellenatome sehr viel geringer ist als die Bandlückenenergie. Im Fall der Donatordotierung spricht man von n -Typ Halbleitern, die eine erhöhte Elektronenkonzentration besitzen, und im Falle der Akzeptordotierung von p -Typ Halbleitern mit erhöhter Löcherkonzentration. Nach wie vor gilt jedoch das Massenwirkungsgesetz

$$n \cdot p = n_i^2$$

Damit sinkt im n -Typ Halbleiter die Löcherkonzentration tief unter die des intrinsischen Halbleiters, und entsprechend im p -Typ Halbleiter die Elektronenkonzentration weit unter die des intrinsischen Halbleiters. Die Ionisierungsenergien ΔW_D und ΔW_A können in Anlehnung an die Ionisierungsenergie des H-Atoms abgeschätzt werden (1.87).

$$W_{ion,H} = \frac{q^2}{8\pi\epsilon_0 r_0} = 13.6 \text{ eV} \quad \text{mit} \quad r_0 = \frac{4\pi\epsilon_0 \hbar^2}{m_0 q^2}$$

Im Falle des dotierten Halbleiters ersetzen wir die freie Elektronenmasse m_0 mit der effektiven Masse m^* des Elektrons bzw. des Lochs im Kristall, und wir berücksichtigen die relative Dielektrizitätskonstante ϵ_r des Halbleitermaterials:

$$\Delta W_D = \frac{m_n^* q^4}{2(4\pi\epsilon_0 \epsilon_r \hbar)^2} \quad \text{und} \quad \Delta W_A = \frac{m_p^* q^4}{2(4\pi\epsilon_0 \epsilon_r \hbar)^2} \quad (2.44)$$

Für einige wichtige Halbleiter sind die Ionisierungsenergien in Tab. 2.2 angegeben: Ebenfalls angegeben sind die Bohr'schen Radien für das dem H-Atom ähnliche System "Ladungsträger – Störstellenatom". Die Ausdehnung erstreckt sich über viele hundert Gitterzellen. Für **Verbindungshalbleiter** wie z.B. GaAs sind die Verhältnisse meist komplizierter als beim Elementhalbleiter Silizium: Gallium (Ga) ist dreiwertig (3 Valenzelektronen) und Arsen (As) ist fünfwertig (5 Valenzelektronen), sodass zusammen – wie im Fall des Si – acht Elektronen für die Bindung mit den vier nächsten Nachbaratomen

Tabelle 2.2: Donator- und Akzeptor-Ionisierungsenergien verschiedener Halbleiter

	$\Delta W_D/\text{meV}$	r_e^*/r_0	$\Delta W_A/\text{meV}$	r_h^*/r_0
Si	94	12	47	24
Ge	29	29	20	43
GaAs	5,3	196	36	29
InAs	1,9	568	36	30
InP	8	159	57	22
InGaAs	3,7	284	36	30
SiC				
GaN				

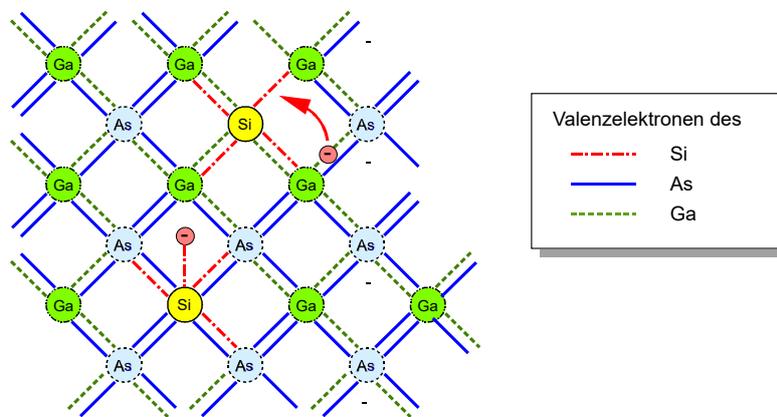


Abbildung 2.21: Amphoterer Dotierverhalten von Si in GaAs

zur Verfügung stehen. Jedes Ga-Atom ist von vier nächsten As-Atomen, jedes As-Atom von vier nächsten Ga-Atomen umgeben.

Bei der Dotierung im Verbindungshalbleiter kommt es darauf an, auf welchem Platz das Dotieratom eingebaut wird. Betrachten wir beispielsweise ein Silizium-Dotieratom (vierwertig) im GaAs-Kristall.

Auf dem Ga-Platz liefert Si ein überzähliges Elektron und wirkt als Donator, auf einem As-Platz fehlt ein Elektron und es wirkt daher als Akzeptor (Abb. 2.21). Dieses Verhalten wird als **amphoter** bezeichnet.

Tiefe Störstellen

Bei komplizierter aufgebauten Störstellen können auch Zustände tiefer in der Bandlücke vorliegen. Diese werden daher auch als tiefe Störstellen bezeichnet. Sie können anhand des Bänderdiagramms nicht eindeutig als Donator bzw. Akzeptor eingeordnet werden. Über derartige Störstellen kann sehr effektiv die Rekombination eines Elektron-Loch-Paares ablaufen, sodass diese auch als Rekombinationszentren bezeichnet werden.

Für den Fall, dass trotz der energetisch tiefen Lage eine tiefe Störstelle eindeutig entweder Donator- oder Akzeptorcharakter besitzt, kann sie zur Kompensation von flachen Störstellen verwendet werden. Beispielsweise ist Chrom (Cr) in GaAs ein tiefer Akzeptor. Üblicherweise existiert in GaAs eine unerwünschte, aber unvermeidbare Verunreinigung mit Si auf Ga-Plätzen, die n -Typ Leitfähigkeit verursacht. Durch eine gezielte Dotierung mit Cr können die von den Si-Atomen stammenden Elektronen unwirksam gemacht werden, da sie von den Cr-Akzeptoren eingefangen werden. Wegen des großen energetischen Abstandes zu den Bandkanten kann das Elektron von dort aus nicht durch thermische Energie befreit werden. Die Folge ist ein hochohmiges, "semi-isolierendes" GaAs, das gute Hochfrequenzeigenschaften besitzt und dadurch als Substratmaterial für HF-Halbleiterchips verwendet werden kann.

Störstellenbänder

Bisher haben wir vorausgesetzt, dass die Störstellenatome (Donatoren bzw. Akzeptoren) nur als vereinzelte Störung des Kristalls auftreten und entsprechend weit voneinander entfernt sind. Andererseits wissen wir aber bereits, dass das System "Störstelle-Elektron" bei der geringen effektiven Masse und hohem ϵ_r -Wert eine große Ausdehnung besitzt. Sobald die Störstellen miteinander in Wechselwirkung treten, greift aber das Pauliprinzip. Daher spalten dann die diskreten Störstellenniveaus in Störstellenbänder auf, wie in Abb. 2.22 gezeigt, ähnlich dem Aufspalten der diskreten Atomenergieniveaus beim Zusammenbau zum Festkörper. Mit steigender Donatorkonzentration nimmt dadurch auch die wirksame Ionisierungsenergie ab, sie verschwindet schließlich sogar. Die Ausbildung der Störstellenbänder setzt gerade bei Halbleitern mit geringer effektiver Masse (s.o.) schon bei relativ geringen Störstellenkonzentration ein.

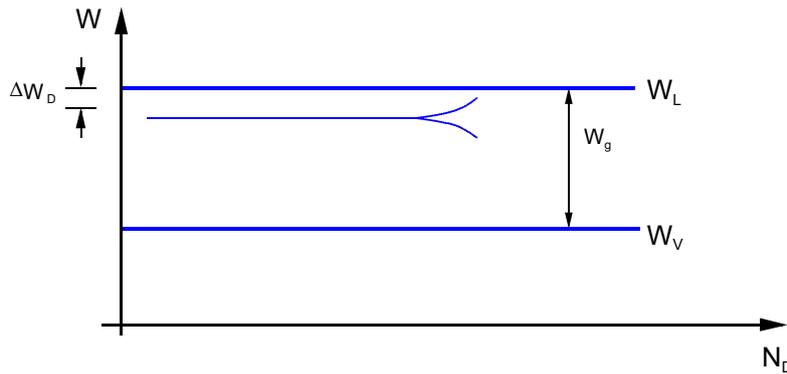


Abbildung 2.22: Aufspaltung des Donatorbiveaus bei steigender Dotierung

2.3.5 Ladungsträgerstatistik im Störstellenhalbleiter

Die Ladungsträgerkonzentrationen (n bzw. p) im intrinsischen Halbleiter werden durch die drei Größen Zustandsdichte $g(W)$, Fermi-Dirac-Verteilungsfunktion $f(W)$ sowie Fermi-Energie W_F bestimmt. Dies gilt auch für den Störstellenhalbleiter (extrinsischer Halbleiter), allerdings sind Unterschiede zum Eigenhalbleiter zu beachten, es existieren nämlich Zustände in der Energielücke, und es gilt hier allgemein $n \neq p$.

Nach wie vor gelten die Gleichungen (2.31) und (2.37) für die Elektronen- bzw. Löcherkonzentrationen:

$$n = N_L \cdot \exp\left(-\frac{W_L - W_F}{kT}\right) \quad \text{und} \quad p = N_V \cdot \exp\left(-\frac{W_F - W_V}{kT}\right) \quad (2.45)$$

Die Fermi-Energie im dotierten Fall unterscheidet sich von der im Eigenhalbleiter: im n -Halbleiter liegt sie näher an der Leitungsbandunterkante, im p -Halbleiter näher an der Valenzbandoberkante. Dies tritt auf, da mehr (n) bzw. weniger (p) der möglichen Zustände mit Elektronen besetzt sind, verglichen mit dem intrinsischen, undotierten Halbleiter. Die hier angegebene Fermienergie E_F entspricht nicht der intrinsischen Fermienergie E_{Fi} . Für n -Dotierung ist also $E_{Fn} > E_{Fi}$ und für p -Dotierung ist $E_{Fp} < E_{Fi}$.

Weiterhin gilt das Massenwirkungsgesetz, das die Konzentrationen in der "Reaktion" von Elektron-Loch-Paaren

$$n + p \leftrightarrow 0$$

beschreibt als

$$\begin{aligned} n \cdot p &= N_L \cdot N_V \cdot \exp\left(-\frac{W_L - W_F}{kT}\right) \cdot \exp\left(-\frac{W_F - W_V}{kT}\right) \\ &= N_L \cdot N_V \cdot \exp\left(-\frac{W_L - W_V}{kT}\right) = N_L \cdot N_V \cdot \exp\left(-\frac{W_g}{kT}\right) = n_i^2 \end{aligned} \quad (2.46)$$

Im Störstellenhalbleiter wird die Besetzung der Dotierniveaus innerhalb der Energielücke durch die Fermi-Energie W_F festgelegt. Die Besetzung der Donatorstörstellen mit Elektronen ist:

$$N_D^0 = g_D \cdot f(W_D) = N_D \cdot \frac{1}{\exp\left(\frac{W_D - W_F}{kT}\right) + 1} \quad (2.47)$$

$$N_D = N_D^0 + N_D^+ \quad (2.48)$$

mit $g_D = N_D$ Zustandsdichte der Donatoren, Konzentration N_D der Donatoratome insgesamt, N_D^0 der mit Elektronen besetzten, N_D^+ der ionisierten Donatoren.

Die Zustandsdichte der Donatoren g_D entspricht gerade ihrer Konzentration. Die Zustandsdichte beschreibt die möglichen Zustände pro Energieintervall und pro Volumen. An jedem Dotieratom besteht genau ein möglicher Zustand bei einer festen Energie, dies kann mathematisch auch ausgedrückt werden durch die Kronecker-Delta-Funktion

$$g_D(W) = \delta(W - W_D)$$

In (2.47) ist auf der linken Seite die Dichte der mit einem Elektron besetzten Dotieratome N_D^0 angegeben. Ein besetztes Dotieratom ist neutral, da die zusätzliche positive Ladung des Dotieratomkerns ausgeglichen ist – in diesem Fall ist das zusätzlich durch Dotierung entstandene Elektron eingefangen, d.h. es trägt nicht zur Leitfähigkeit bei.

Die für die Leitfähigkeit wichtige Größe N_D^+ der ionisierten Donatoren kann nach (2.48) als Differenz zur bekannten Dichte der Dotieratome N_D (technologisch bestimmt) bestimmt werden.

Für Akzeptoren gilt analog:

$$N_A^- = g_A \cdot f(W_A) = N_A \cdot \frac{1}{\exp\left(\frac{W_A - W_F}{kT}\right) + 1} \quad (2.49)$$

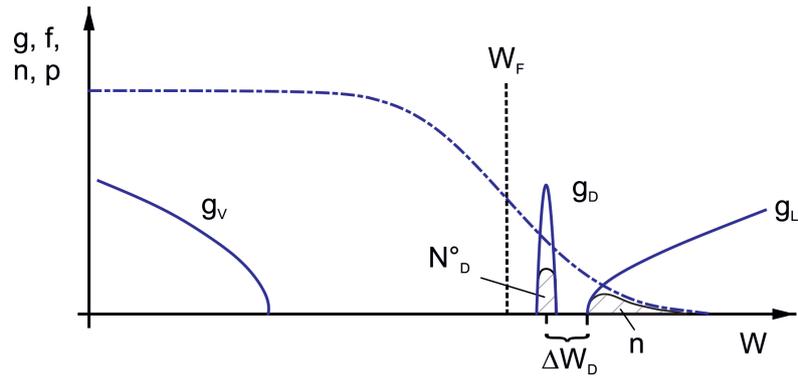


Abbildung 2.23: Besetzungsstatistik im n -Halbleiter

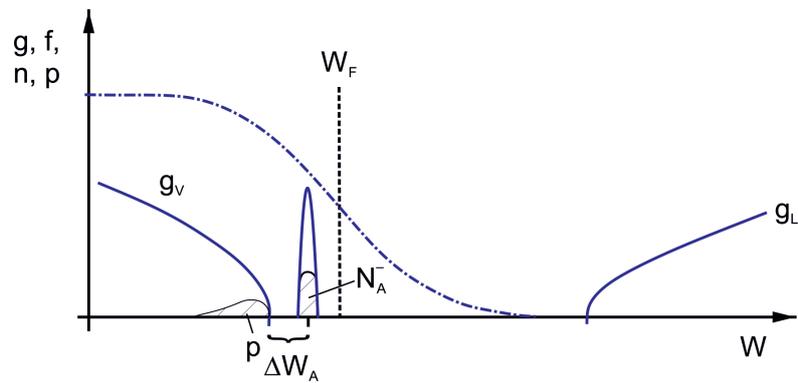


Abbildung 2.24: Besetzungsstatistik im p -Halbleiter

$$N_A = N_A^0 + N_A^- \quad (2.50)$$

In den Abb. 2.23 und 2.24 sind analog zur Abb. 2.12 die für die Ladungsträgerstatistik entscheidenden Größen $g(W)$, $f(W)$ und W_F und die zugehörigen Ladungsträgerkonzentrationen über der Energie dargestellt, und zwar unterschieden nach n - bzw. p -Halbleiter. Da jedes Donator- und Akzeptoratom für sich neutral ist, gilt auch die Ladungsneutralität für dotierte Halbleiter, auch für eine simultane p - und n -Dotierung, solange sich der Halbleiter im thermodynamischen Gleichgewicht befindet:

$$n + N_A^- = p + N_D^+ \quad (2.51)$$

Mit

$$N_D^+ = N_D - N_D^0 \quad (2.52)$$

und (2.45) und (2.49) folgt:

$$\begin{aligned} N_L \cdot \exp\left(-\frac{W_L - W_F}{kT}\right) + N_A \cdot \frac{1}{\exp\left(\frac{W_A - W_F}{kT}\right) + 1} \\ = N_V \cdot \exp\left(-\frac{W_F - W_V}{kT}\right) + N_D \cdot \left[1 - \frac{1}{\exp\left(\frac{W_D - W_F}{kT}\right) + 1}\right] \end{aligned} \quad (2.53)$$

Diese Gleichung beschreibt die gegenseitige Abhängigkeit der Konzentrationen von Elektronen n und Löchern p , sowie der Fermienergie W_F als Funktion der gegebenen Donator- und Akzeptordichten N_D , N_A und den effektiven Zustandsdichten des Halbleiters im Valenz- und Leitungsband, N_V und N_L .

Besonders deutlich wird die zentrale Rolle der Fermi-Energie W_F , die grundsätzlich bei Kenntnis der anderen Parameter aus (2.53) bestimmt werden kann. Auch der starke Einfluss der Temperatur ist ersichtlich. Eine analytische Auswertung lässt sich nur für spezielle Voraussetzungen durchführen. Dies soll hier beispielhaft für einen n -dotierten Halbleiter getan werden. Dabei soll insbesondere Wert auf die Temperaturabhängigkeit gelegt werden.

Dazu wird angenommen, dass

1. nur ein Donatortyp vorliegt, charakterisiert durch die Parameter ΔW_D und N_D
2. zunächst die Eigenleitungskonzentration n_i vernachlässigt wird ($n_i \approx 0$ und $n \approx N_D^+$)

Dann gilt:

$$\begin{aligned} n = N_D^+ &= N_D \cdot [1 - f(W_D)] \\ &= N_D \cdot \left[1 - \frac{1}{\exp\left(\frac{W_D - W_F}{kT}\right) + 1}\right] \\ &= N_D \cdot \left[1 - \frac{1}{e^{-\Delta W_D/kT} \cdot e^{(W_L - W_F)/kT} + 1}\right] \end{aligned} \quad (2.54)$$

Dabei wurde benutzt:

$$W_D - W_F = -W_L + W_D + W_L - W_F = -\Delta W_D + W_L - W_F$$

Ferner gilt mit (2.45):

$$e^{(W_L - W_F)/kT} = N_L/n$$

und somit

$$n = -\frac{N_L}{2} \cdot e^{-\Delta W_D/kT} \pm \sqrt{\left[\frac{N_L}{2} e^{-\Delta W_D/kT}\right]^2 + N_L N_D \cdot e^{-\Delta W_D/kT}} \quad (2.55)$$

Nach einer weiteren Umformung folgt:

$$n = -\frac{N_L}{2} \cdot e^{-\Delta W_D/kT} \pm \left[\sqrt{1 + 4 \frac{N_D}{N_L} \cdot e^{\Delta W_D/kT}} - 1 \right] \quad (2.56)$$

Dieser Ausdruck für die Elektronenkonzentration im n -Halbleiter soll im folgenden näher untersucht werden.

Schwache Dotierung und hohe Temperatur

Genauer soll gelten:

1. $N_D \ll N_L$ und
2. $kT \gg \Delta W_D$

Dann ist

$$4 \frac{N_D}{N_L} e^{\Delta W_D/kT} \ll 1$$

Der Wurzelausdruck kann für kleine x als $\sqrt{1+x} \approx x/2$ entwickelt werden. Dann folgt aus (2.56)

$$n = N_D \quad (2.57)$$

Die Elektronenkonzentration ist identisch mit der Donatorkonzentration, denn alle Donatoren haben ihr Elektron abgegeben und sind somit ionisiert. Man sagt, die Störstellen

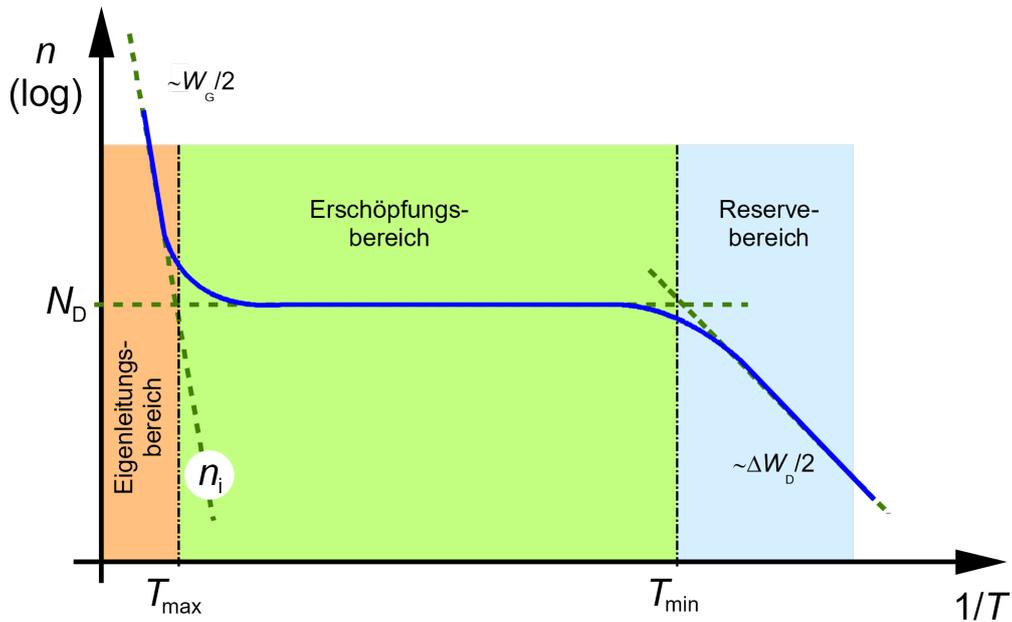


Abbildung 2.25: temperaturabhängige Elektronenkonzentration in n -dotiertem Halbleiter, Erschöpfungsbereich mit konstanter Konzentration

sind erschöpft. Dies ist auch einleuchtend, weil wegen (a) $N_D \ll N_L$ für die Donatorelektronen genügend freie Plätze im Leitungsband vorhanden sind und wegen (b) $kT \gg \Delta W_D$ die thermische Energie sehr viel höher ist, als zur Ionisierung der Donatoren nötig ist.

Tiefe Temperaturen

In diesem Fall ist $kT \ll \Delta W_D$ und somit auch

$$4 \frac{N_D}{N_L} \cdot e^{\Delta W_D/kT} \gg 1$$

In dieser Näherung wird (2.56) zu:

$$n = \sqrt{N_D N_L} e^{-\Delta W_D/2kT} \tag{2.58}$$

Mit steigender Temperatur nimmt die Elektronenkonzentration also stark zu. Es befindet sich eine "Elektronen-Reserve" auf den noch nicht ionisierten Donatoren, daher wird dieser Temperaturbereich Reservebereich genannt. Bei der Temperatur T_{max}

(Abb. 2.24) geht dieser in den Erschöpfungsbereich mit $n = N_D$ über.

Wird die Temperatur weiter gesteigert bis T_{max} in Abb. 2.25, werden Elektron-Loch-Paare durch Band-Band-Übergänge über die Bandlücke hinweg erzeugt, wie bereits für den Fall des Eigenhalbleiters besprochen. In diesem Temperaturbereich überwiegt die Eigenleitung, dementsprechend heißt er **Eigenleitungsbereich**. Die Elektronenkonzentration folgt hier (2.40)

$$n_i = \sqrt{N_L N_V} \cdot e^{-W_g/2kT} \quad (2.59)$$

Die Abb. 2.25 gibt schematisch die Temperaturabhängigkeit eines n -dotierten Halbleiters von niedrigsten bis zu höchsten Temperaturen wieder. Analoge Überlegungen lassen sich für einen p -dotierten Halbleiter durchführen.

Bauelemente mit dotierten Halbleitern werden im Erschöpfungsbereich betrieben, also zwischen T_{min} und T_{max} . Denn Elemente, die einen pn -Übergang enthalten, verlieren oberhalb von T_{max} ihre Funktion, weil ein Eigenhalbleiter per Definition weder n - noch p -leitend, sondern gleich viele Elektronen wie Löcher besitzt. Für Temperaturen unterhalb T_{min} "frieren die Ladungsträger aus", die Ladungsträgerkonzentration hat keinen festgelegten Wert mehr.

Bezieht man sich auf eine n -Dotierung von $N_D = 1 \times 10^{17} \text{ cm}^{-3}$, dann ergibt sich für Silizium ein Wert für T_{max} für GaAs wegen des höheren Bandabstandes ein deutlich höherer Wert von T_{max} . Dabei ist zu beachten, dass diese Temperaturen z.B. in Bipolartransistoren am Ort des pn -Überganges nicht überschritten werden dürfen, so dass die Gehäusetemperatur noch deutlich niedriger liegen muss. Ferner wird klar, dass für elektronische Bauelemente, die in einer Hochtemperaturumgebung (Automobil, Energietechniksteuerungen usw.) arbeiten müssen, vorteilhaft Halbleiter mit größerer Bandlücke verwendet werden.

Wenn man gemäß der oben durchgeführten Überlegungen im n -Halbleiter die Elektronenkonzentration berechnet hat, so ist die Bestimmung der dann vorliegenden Löcherkonzentration sehr einfach durch Anwendung des Massenwirkungsgesetzes durchzuführen. Es gilt im Erschöpfungsbereich:

$$p = \frac{n_i^2}{n} = \frac{n_i^2}{N_D} \quad \text{mit} \quad p \ll n_i \ll n \quad (2.60)$$

Wegen der um viele Größenordnungen unterschiedlichen Konzentrationen nennt man

Elektronen im n -Halbleiter Majoritätsladungsträger und Löcher die Minoritätsladungsträger.

Für den technisch wichtigen Fall der Ladungsträgererschöpfung lässt sich unter Verwendung von (2.45) die Fermi-Energie leicht bestimmen für n - und p -Halbleiter:

$$\begin{aligned} W_{Fn} &= W_L - kT \cdot \ln(N_L/n) \approx W_L - kT \cdot \ln(N_L/N_D) \\ W_{Fp} &= W_V + kT \cdot \ln(N_V/p) \approx W_V + kT \cdot \ln(N_V/N_A) \end{aligned} \quad (2.61)$$

Diese Gleichungen werden später bei der Behandlung des pn-Übergangs eingesetzt.

2.4 Ladungstransport in Festkörpern

Für die elektrische Leitfähigkeit von und die Stromdichte in Festkörpern ist sicher die im vorhergehenden Kapitel behandelte Ladungsträgerkonzentration von großer Bedeutung. Darüber hinaus spielt aber auch die Geschwindigkeit eine Rolle, mit der sich die Ladungsträger unter Einwirkung einer äußeren Kraft bewegen. Sehen wir im Festkörper von einem Ionen-transport ab, so wird die Leitfähigkeit von Elektronen und Löchern getragen.

Allgemein gilt dann für die Stromdichte

$$\vec{J} = -qn\vec{v}_n + qp\vec{v}_p \quad (2.62)$$

mit den Ladungsträgerkonzentrationen n und p , der Elementarladung $\pm q$ mit neg. und pos. Vorzeichen für Elektronen und Löcher, und der Elektronen- und Löchergeschwindigkeit \vec{v}_n und \vec{v}_p . Wird die äußere Kraft durch ein elektrisches Feld hervorgerufen, so ist wegen des Vorzeichenunterschiedes der Ladung \vec{v}_n antiparallel und \vec{v}_p parallel zum elektrischen Feld gerichtet.

Unter \vec{v}_n und \vec{v}_p verstehen wir die mittlere Driftgeschwindigkeit der Elektronen und Löcher im Festkörper. Elektronen und Löcher bewegen sich bei Raumtemperatur auch ohne eine von außen wirkende Kraft mit erheblichen Geschwindigkeiten, da sie thermische Energie besitzen. Die mittlere thermische Energie jedes Ladungsträgers beträgt:

$$W_{th} = \frac{3}{2}kT = \frac{1}{2}m^*v_{th}^2 = \frac{1}{2}m^* \left(\frac{\hbar k}{m^*} \right)^2$$

oder, umgestellt nach der thermischen Geschwindigkeit:

$$v_{th} = \sqrt{\frac{3kT}{m^*}} \quad (2.63)$$

Für gängige Halbleiter beträgt die thermische Geschwindigkeit v_{th} bei $T = 300$ K etwa $v_{th} \approx 1 \times 10^7$ cm/s.

Die Geschwindigkeitsrichtungen sind allerdings völlig unkorreliert und unterschiedlich, und nicht jedes Elektron oder Loch besitzt v_{th} . Die thermische Geschwindigkeit kann man vielmehr als Maß für die mittlere Energie der Elektronen verstehen gemäß (2.63).

Die Geschwindigkeit jedes einzelnen Elektrons mit der individuellen Energie W erhalten wir aus der $W - k$ -Beziehung, die im vorhergehenden Kapitel beschrieben wurde.

Die Verteilung der Elektronen auf die erlaubten Energieniveaus erfolgt im Gleichgewicht unter Beachtung der Zustandsdichte $g(W)$, der Fermi-Dirac-Verteilungsfunktion $f(W)$ und der Fermi-Energie W_F (2.1)

$$n(W) = g(W) \cdot f(W)$$

Die Geschwindigkeit ist nur mit dem kinetischen Anteil der Gesamtenergie W verknüpft, für Elektronen:

$$W = W_L + W_{kin} = W_L + \frac{\hbar^2 \vec{k}^2}{2m^*} \quad (2.64)$$

mit (2.64) und unter Einsatz der Boltzmann-Näherung kann $f(W)$ aufgeteilt werden in:

$$f(W) = e^{-(W-W_F)/kT} = \exp\left(-\frac{W_L - W_F}{kT}\right) \cdot \exp\left(-\frac{\hbar^2 k^2}{2m^* kT}\right) \quad (2.65)$$

Das heißt, dass die Besetzungswahrscheinlichkeit der Zustände mit Elektronen mit steigender kinetischer Energie, d.h. mit steigendem Wellenzahlvektor k und damit auch mit steigender Geschwindigkeit $(\hbar k)/m^*$ rasch abnimmt, wie in Abb. 2.26 dargestellt. Die mit der Zustandsdichte $g(W)$ zusammenhängende Verteilung der erlaubten k -Werte (2.1) ist für zwei Dimensionen noch einmal in Abb. 2.27 wiedergegeben. Wird nun eine äußere Kraft, beispielsweise in Form eines elektrischen Feldes, angelegt, so erhält jedes

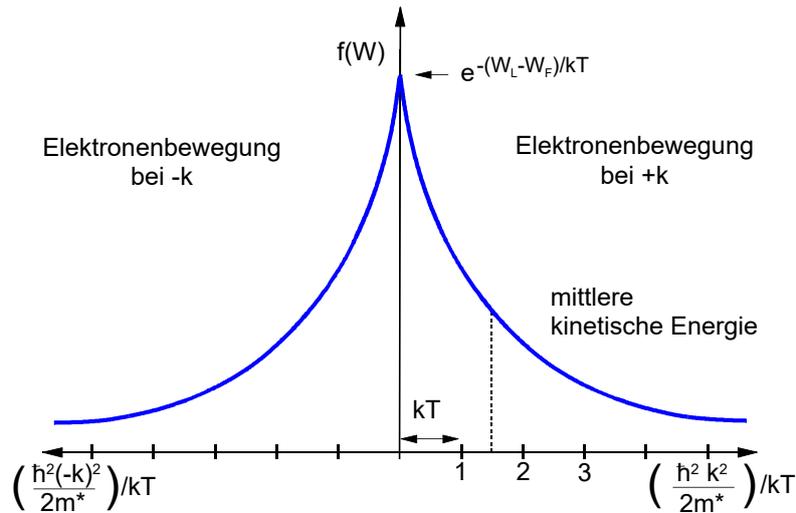


Abbildung 2.26: Besetzungswahrscheinlichkeit $f(W)$ in Abhängigkeit von der kinetischen Energie ohne äußere Kraft

einzelne Elektron einen zusätzlichen Anteil zum Wellenzahlvektor $-k_D$ (negatives Vorzeichen aufgrund negativer Elektronenladung). Dies ist in Abb. 2.28 zusammen mit der daraus resultierenden zusätzlichen Geschwindigkeit

$$\vec{v}_D = \frac{\hbar}{m^*} \vec{k}_D \quad (2.66)$$

dargestellt. Es stellt sich nun die Frage, wie groß das elektrische Feld für eine bestimmte zusätzliche Geschwindigkeitskomponente v_D , der mittleren Driftgeschwindigkeit, sein muss. Denn nur diese Komponente trägt zur Stromdichte J aus (2.61) bei, da sie jedem einzelnen Elektron (bzw. Loch) aufgeprägt wird, während die rein thermische Bewegung zu einer Gesamtstromdichte $J = 0$ führt.

2.4.1 Ladungsträgerbeweglichkeit und elektrische Leitfähigkeit

Zur Klärung dieser Frage betrachten wir genauer den mikroskopischen Prozess der Geschwindigkeitsaufnahme eines Elektrons im elektrischen Feld. Die Kraft, die ein geladenes Teilchen im elektrischen Feld \vec{E} erfährt, ist

$$\vec{F}_n = -q \cdot \vec{E} \quad (2.67)$$

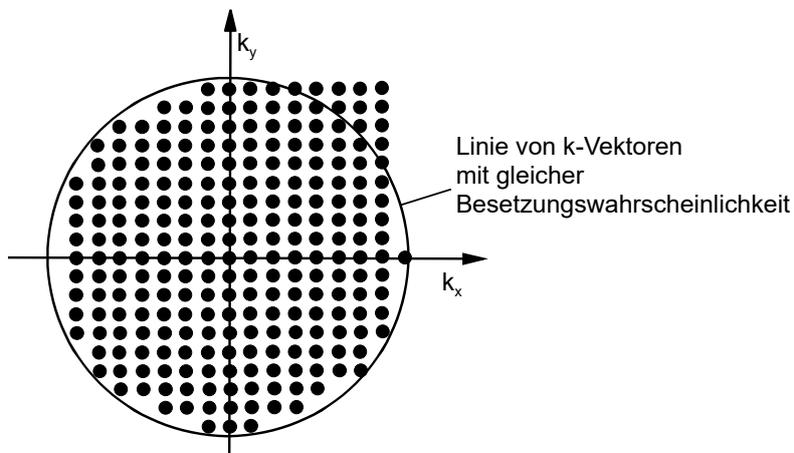


Abbildung 2.27: Quantentheoretisch erlaubte k -Werte in zwei Dimensionen ohne äussere Kraft. Für drei Dimensionen ist der Kreis gleicher Besetzungswahrscheinlichkeit durch eine entsprechende Kugel, die Fermi-Kugel", zu ersetzen

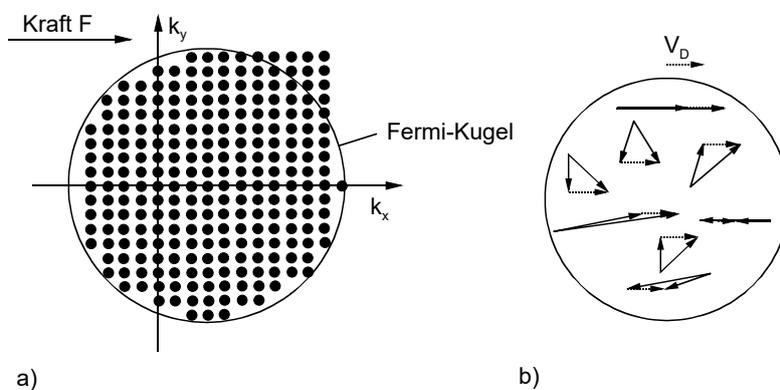


Abbildung 2.28: (a) Verschiebung der Fermi-Kugel (Fermi-Kreises) und (b) Veränderung der Geschwindigkeiten der einzelnen Elektronen durch Wirkung einer äusseren Kraft

Dabei wird ein Elektron mit der Ladung $-q$ betrachtet. Die Folge ist eine Beschleunigung \vec{a} des Elektrons nach dem mechanischen Kraftgesetz:

$$\vec{F} = m_n^* \cdot \vec{a} \quad (2.68)$$

Die Kombination beider Gleichungen führt auf:

$$\vec{a}_n = -\frac{q}{m_n^*} \cdot \vec{E} \quad (2.69)$$

Beschleunigung bedeutet eine ständige Veränderung der Geschwindigkeit. Warum wurde im vorangegangenen Kapitel die Wirkung der Feldkraft als Geschwindigkeitskomponente mit festem Wert v_D beschrieben? Die Begründung ist folgende: Die Beschleunigung, d.h. die Geschwindigkeitszunahme, erfolgt nicht zeitlich unbegrenzt, sondern findet durch einen Zusammenstoß des Elektrons mit einem anderen Teilchen (Gitteratom, Fremdatom, Elektron) ein Ende. Beim Zusammenstoß wird das Elektron nicht auf $v = 0$ abgebremst, denn es besitzt ja nach wie vor thermische Energie. Hier soll nur die Komponente betrachtet werden, die der ganzen Elektronengesamtheit durch das E-Feld aufgeprägt wird, da nur dieser Anteil zur Stromdichte beiträgt. Für diese Komponente kann man annehmen, dass sie unmittelbar nach dem Zusammenstoß gleich Null ist. Anschließend erfolgt erneut die Beschleunigung des Elektrons bis zum nächsten Zusammenstoß. Die Zeit zwischen zwei Zusammenstößen ist nicht immer gleich, sondern besitzt eine statistische Verteilung, der man durch Einführen einer mittleren Zeit τ_n^* Rechnung tragen kann. In dieser Zeit erreicht das Elektron die Geschwindigkeit

$$\vec{v}_{Dn}^* = \int_0^{\tau_n^*} \vec{a}_n dt = \vec{a}_n \cdot \tau_n^* \quad (2.70)$$

Im Mittel bewegen sich die Elektronen mit der bereits in (2.66) angegebenen Driftgeschwindigkeit

$$\vec{v}_{Dn} = \frac{1}{2} \vec{a}_n \cdot \tau_n^* = \frac{\hbar}{m^*} \vec{k} \equiv \vec{a}_n \cdot \tau_n \quad (2.71)$$

da sich die Geschwindigkeit zwischen $v = 0$ (unmittelbar nach einem Zusammenstoß) und v_{Dn}^* (unmittelbar vor einem Zusammenstoß) linear mit der Zeit verändert. Diese Mittelwertbildung der Geschwindigkeit wurde im letzten Teil der Gleichung durch

Einführen einer neuen Zeitkonstanten τ_n berücksichtigt. Somit ist die Wertung der Feldkraft in Form einer Geschwindigkeitskomponente v_{Dn} vernünftig. Mit Hilfe der Gleichungen (3.63) und (3.65) folgt

$$\vec{v}_{Dn} = -\frac{q}{m_n^*} \cdot \tau_n \cdot \vec{E} \quad (2.72)$$

Den Proportionalitätsfaktor (ohne Vorzeichen) zwischen der resultierenden Geschwindigkeit \vec{v}_D und der diese verursachenden Feldstärke \vec{E} nennt man – sehr anschaulich – Beweglichkeit der Elektronen μ_n im Festkörper: je beweglicher das Elektron ist, desto höher wird bei vorgegebener Feldstärke \vec{E} die Driftgeschwindigkeit \vec{v}_D sein.

$$\mu_n = \frac{q\tau_n}{m_n^*} \quad (2.73)$$

Analoge Überlegungen für Löcher führen zu den Gleichungen:

$$\begin{aligned} \vec{a}_p &= \frac{+q}{m_p^*} \cdot \vec{E} \\ \vec{v}_{Dp} &= \vec{a}_p \cdot \tau_p \\ \mu_p &= \frac{q\tau_p}{m_p^*} \end{aligned} \quad (2.74)$$

Einsetzen in (2.61) liefert:

$$J = q(n \cdot \mu_n + p \cdot \mu_p) \quad (2.75)$$

Aus (2.72) und (2.73) bzw. (2.74) folgt:

$$\vec{v}_{Dn} = -\mu_n \cdot \vec{E} \quad \text{bzw.} \quad \vec{v}_{Dp} = \mu_p \cdot \vec{E} \quad (2.76)$$

sowie eine allgemeingültige Definition der Ladungsträgerbeweglichkeit:

$$\mu = \frac{d|\vec{v}_D|}{d|\vec{E}|} \quad (2.77)$$

Dieser Ausdruck wird genauer Driftbeweglichkeit genannt. Er ist nicht identisch mit der später zu behandelnden "Hall"-Beweglichkeit, jedoch liegen beide Werte meist dicht beieinander.

Der Ausdruck (2.75) liefert ferner die elektrische Leitfähigkeit als Proportionalitätsfaktor zwischen Stromdichte und elektrischer Feldstärke

$$\vec{J} = \kappa \cdot \vec{E} = (\kappa_n + \kappa_p) \cdot \vec{E} \quad (2.78)$$

Diese Gleichung gibt erneut die Aufteilung in Elektronen- und Löcheranteil wieder. Der Kehrwert von κ wird als spezifischer Widerstand ρ bezeichnet:

$$\rho = 1/\kappa \quad (2.79)$$

Im Ausdruck $\mu = (q \cdot \tau)/m^*$ treten neben der absoluten konstanten Elementarladung q nur noch die effektive Masse m^* und die mittlere Zeit zwischen den Stößen τ als von inneren oder äußeren Gegebenheiten abhängige Größen auf. Hier sollen kurz die wichtigsten Abhängigkeiten des Transportverhaltens der Ladungsträger von der Temperatur, der Dotierung und der elektrischen Feldstärke diskutiert werden.

Temperaturabhängigkeit

Die effektive Masse m^* besitzt eine nur geringe Temperaturabhängigkeit, die in der Bandstruktur steckt: (1.105) beinhaltet die Atomabstände und die Form des Potentialtopfes, beides ändert sich nur schwach mit der Temperatur.

Eine weitaus stärkere Temperaturabhängigkeit besitzt τ , die letztlich die Anzahl der Zusammenstöße der Elektronen pro Sekunde wiedergibt. Für $T = 0$ kann in einem perfekten Festkörperkristall die Beweglichkeit sehr hoch sein, weil die Wechselwirkung mit dem Gitterpotential bereits in der effektiven Masse berücksichtigt ist (vgl. Kap. 1.7). Mit steigender Temperatur schwingen die Gitteratome immer stärker um ihre Ruhelage und stören somit die strenge Periodizität der Gitteranordnung. Diese Gitterschwingungen werden **Phononen** genannt, die man ähnlich den Photonen und Elektronen als Teilchen behandeln kann. Die Störung der Elektronen- (Löcher-)bewegung (Phononenstreuung) kann man in diesem Bild als Zusammenstoß mit Phononen interpretieren, der umso häufiger auftritt, je höher die Temperatur ist.

Die Beweglichkeit und die Driftgeschwindigkeit werden also mit steigender Temperatur abnehmen. Man findet:

$$\mu_{phon}(T) \propto T^{-3/2} \quad (2.80)$$

mit Schwankungen des Exponenten um den Wert $-(3/2)$. Abb. 2.29 zeigt diesen Verlauf für sehr schwache Dotierungen.

Dotierungsabhängigkeit

Eine Dotierung beeinflusst die effektive Masse kaum, da wegen der im Vergleich zur Konzentration der Gitteratome geringen Dotierstoffkonzentration die Bandstruktur nicht wesentlich verändert wird. Oben wurde bereits dargelegt, dass die Störung der Periodizität des Gitters zu häufigeren Zusammenstößen und somit zur Beeinträchtigung der Beweglichkeit führt. Eingebaute Störungen, insbesondere in Form von geladenen Donatoren oder Akzeptoren, erniedrigen dann erst recht die Beweglichkeit. Durch die Ladung ist ihr Wirkungsbereich besonders groß, der durch r_0^* (Tab. 2.2) gegeben ist. Mit steigender Störstellenkonzentration wird daher die Beweglichkeit abnehmen. Die Streuung der Elektronen an ionisierten (geladenen) Störstellen (Coulombstreuung) hat ebenfalls eine charakteristische Temperaturabhängigkeit:

$$\mu_{ion} \propto T^{3/2} \quad (2.81)$$

Die Beweglichkeit steigt also mit zunehmender Temperatur an (Abb. 2.29, für höhere Störstellenkonzentrationen). Dies ist leicht einzusehen, da mit steigender Temperatur die thermische Ladungsträgergeschwindigkeit ansteigt und damit die Wechselwirkung mit den Störstellen abnimmt.

Je nach Dotierung durchläuft daher die Beweglichkeit mit der Temperatur ein Maximum. Entscheidend ist hier die Summe von Donator- und Akzeptorkonzentrationen, da beide Anteile die Beweglichkeit reduzieren. Für die Ladungsträgerkonzentration ist dagegen die Differenz von Donator- und Akzeptorkonzentration entscheidend, z.B. $n = N_D^+ - N_A^-$, wenn $N_D^+ > N_A^-$.

Abhängigkeit von der elektrischen Feldstärke

Hier spielen sowohl m^* als auch τ eine Rolle. Zunächst, d.h. bei kleinen Feldstärken, sind beide Größen konstant (Abb. 2.30) und gemäß (2.76) steigt die Driftgeschwindigkeit linear mit der Feldstärke.

Übertragen auf Stromdichte und Spannung ist dies der Bereich, in dem das Ohm'sche

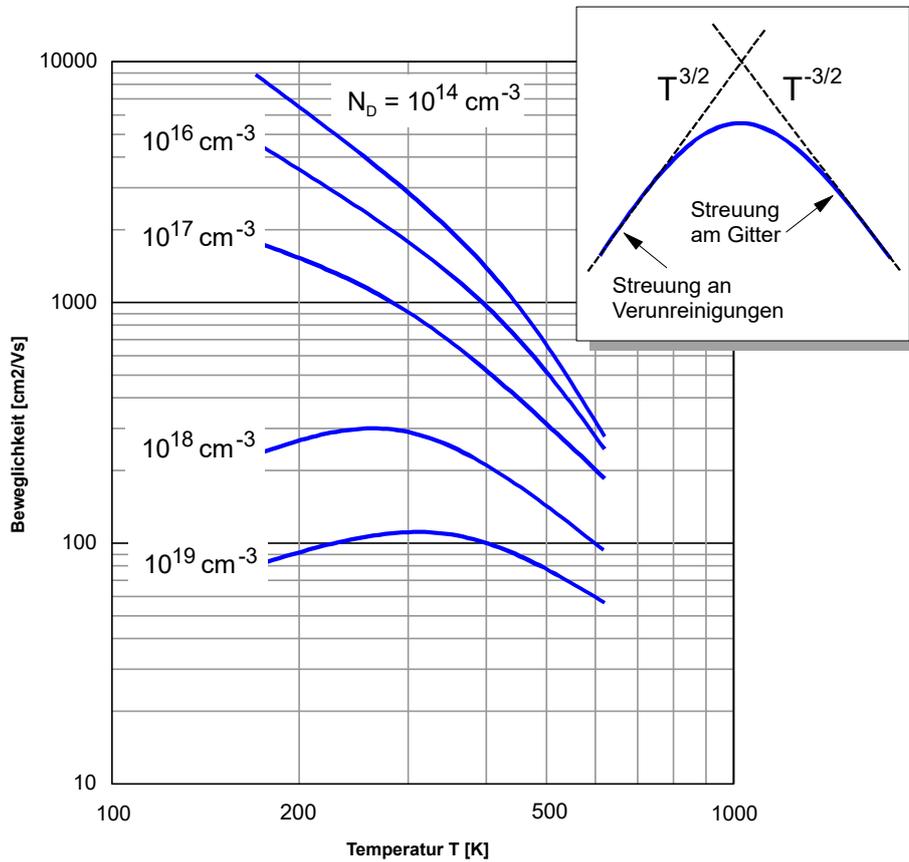


Abbildung 2.29: Abhängigkeit der Elektronendriftbeweglichkeit von der Temperatur und der Dotierung in n-dotiertem Silizium, nach H. Schaumburg, "Halbleiter", B. G. Teubner-Verlag, 1991, S. 145

Gesetz gültig ist:

$$J = \kappa \cdot E \quad \text{bzw.} \quad I = U/R \quad (2.82)$$

mit dem elektrischen Feld $E = U/L$, das die Spannung U über den Kontaktabstand L hervorruft, dem Widerstand

$$R = 1/\kappa \cdot L/A = \rho \cdot L/A$$

und dem Leitungsquerschnitt A . Dieser bildet die Proportionalität von Stromdichte J und Strom I

$$I = J \cdot A$$

Mit steigender Feldstärke und somit steigender Geschwindigkeit nimmt die kinetische Energie der Elektronen zu. Nach Kap. 1.7 gilt dann nicht länger die Parabelnäherung der konstanten effektiven Masse, sondern m^* steigt an. Somit fällt gemäß (2.73) bzw. (2.74) die Ladungsträgerbeweglichkeit.

Ferner werden die Zeitabstände zwischen den Zusammenstößen bei steigender Geschwindigkeit (aufgrund steigender Feldstärke) immer kürzer, sodass τ ebenfalls für eine fallende Beweglichkeit sorgt. Dies führt schließlich, z.B. bei Silizium, zu einer Sättigung der Geschwindigkeit mit der Feldstärke (vgl. Abb. 2.30). Eine Geschwindigkeits-sättigung ist gleichbedeutend mit einer verschwindenden Beweglichkeit (2.77).

Verbindungshalbleiter wie z.B. GaAs weisen die Besonderheit auf, dass bei Erhöhung des angelegten elektrischen Felds die Elektronendriftgeschwindigkeit nach Durchlaufen eines Maximums sogar wieder fällt, obwohl die Feldstärke ansteigt. Dieser Sachverhalt ist gleichbedeutend mit einer negativen Beweglichkeit und damit auch mit einem **negativen Widerstand**. Bauelemente aus solchen Materialien (z.B. Gunn-Elemente) können zur Verstärkung und Schwingungserzeugung benutzt werden.

Diesem Effekt liegt die in Abb. 3.32 gezeigte Bandstruktur zugrunde. Neben dem Hauptminimum des Leitungsbandes mit der kleinen effektiven Masse m_1^* existiert noch ein zweites, energetisch etwas höherliegendes Minimum mit der erheblich höheren Masse m_2^* .

Steigt nun die Feldstärke, so bewegen sich die Elektronen zunächst, entsprechend ihrer geringen Masse m_1^* , mit hoher Driftgeschwindigkeit. Wenn ihre kinetische Energie

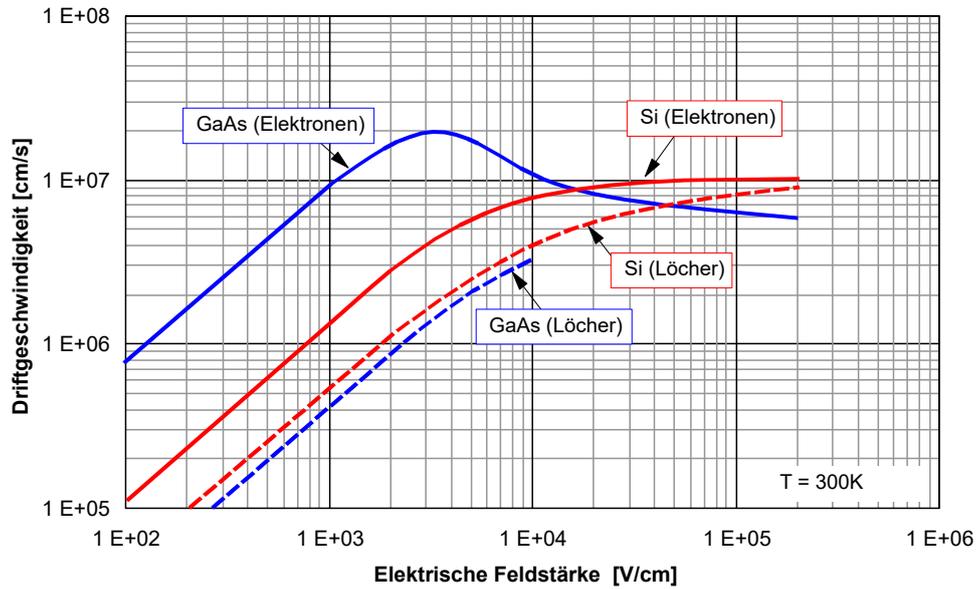


Abbildung 2.30: Feldstärkeabhängigkeit von Ladungsträgern in unterschiedlichen Halbleitern, nach H. Schaumburg, "Halbleiter", B. G. Teubner-Verlag, 1991, S. 147

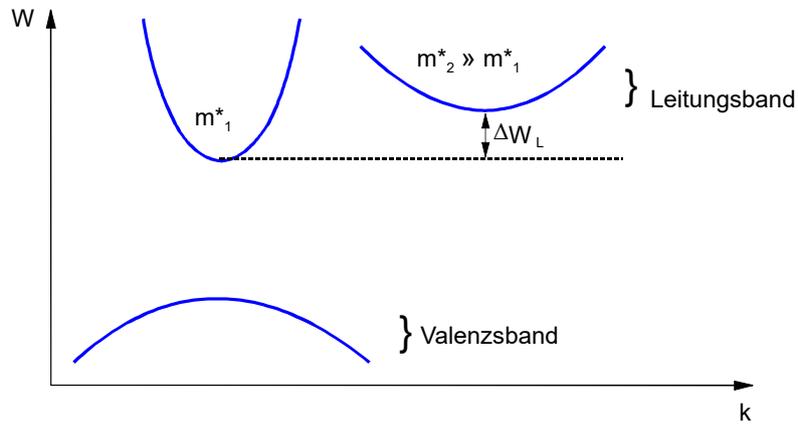


Abbildung 2.31: $W(k)$ -Verlauf von GaAs (InP)

größer wird als der Abstand beider Minima, können viele Elektronen ins Minimum mit der großen Masse m_2^* "hinübergestreut" werden. Hier besitzen sie dann wegen der hohen Masse eine geringe Geschwindigkeit, sodass der Strom trotz steigender Feldstärke sinkt. Hier sei noch angemerkt, dass es sich um einen negativen differentiellen Widerstand handelt, d.h., mit diesem Bauelement kann keine Energie erzeugt werden.

Unterschiedliche Streumechanismen (Phononenstreuung, Störstellenstreuung) führen zu unterschiedlichen Beweglichkeitsanteilen. Wirken diese simultan, so bestimmt sich die Gesamtbeweglichkeit nach der Matthiesen-Regel, die besagt, dass die Beweglichkeitsanteile reziprok addiert werden:

$$1/\mu = \sum_i \frac{1}{\mu_i} \quad (2.83)$$

Die Beweglichkeiten μ_i sind antiproportional zu den entsprechenden Streuzeiten τ_i , d.h., hier werden die Streuzeiten verschiedener physikalischer Prozesse aufaddiert.

2.4.2 Ströme infolge eines elektrischen Feldes: Feldströme

Grundsätzlich wurde bereits im vorangegangenen Kapitel der Zusammenhang zwischen Stromdichte und elektrischer Feldstärke (2.78) angegeben:

$$\vec{J}_{\text{Feld}} = \kappa \cdot \vec{E} = (\kappa_n + \kappa_p) \cdot \vec{E}$$

Dabei wurde angenommen, dass die Leitfähigkeit κ , und damit auch die Ladungsträgerkonzentrationen und die Beweglichkeiten, ortsunabhängig waren. Bei räumlich konstanter Feldstärke ist dann auch die Stromdichte ortsunabhängig.

Im folgenden soll ein Zusammenhang der für den Feldstrom relevanten Größen mit dem Bänderdiagramm hergestellt werden, das sich bei der Beschreibung der elektronischen Eigenschaften von Festkörpern bereits ausgezeichnet bewährt hat.

Für den Ladungsträgertransport ist die Geschwindigkeit der Elektronen und Löcher, und damit deren kinetische Energie entscheidend (2.61), die, wie bereits mehrfach begründet (2.64), für Elektronen ihren Bezugspunkt an der Leitungsbandunterkante hat. Andererseits bestehen zwischen der Energie W einer Ladung in einem elektrischen Feld \vec{E} , diesem Feld und dem dazugehörigen elektrostatischen Potential φ , die Beziehungen

$$\begin{aligned} W &= -q\varphi + \text{const.} \\ \nabla\varphi &= -\vec{E} \end{aligned} \quad (2.84)$$

Gemäß der Wahl des Nullpunktes der kinetischen Energie wird aus diesen Gleichungen

$$\begin{aligned} W_L &= -q\varphi \\ \nabla W_L &= -q\nabla\varphi = q\vec{E} \end{aligned} \quad (2.85)$$

Wegen $W_g = \text{const.}$ für einen bestimmten Halbleiter gilt auch

$$\nabla W_L = \nabla(W_V + W_g) = \nabla W_V \quad (2.86)$$

Das bedeutet, dass bei Anlegen eines E-Feldes die Bandkanten W_L und W_V nicht mehr waagrecht, sondern geneigt, und somit W_L und W_V **ortsabhängig** sind. Nach unserer Voraussetzung sollen jedoch die Ladungsträgerkonzentrationen ortsunabhängig sein, d.h. nach (2.45) muss $(W_L - W_F)$ bzw. $(W_F - W_V)$ konstant sein. Somit ergibt sich die Abb. 2.32. Der Potentialunterschied zwischen linkem und rechtem Ende des betrachteten Halbleiterkörpers ist identisch mit der angelegten Spannung, die das E-Feld verursacht.

Es gilt also (eindimensional):

$$-U = \varphi(x=L) - \varphi(x=0) = \int_{\varphi(x=0)}^{\varphi(x=L)} d\varphi = -\frac{1}{q} \int_{W_L(x=0)}^{W_L(x=L)} dW_L \quad (2.87)$$

Die Neigung der Bandkanten ist offensichtlich gleichbedeutend mit einem Potentialunterschied und dem Vorliegen eines elektrischen Feldes. Dieses Ergebnis ist wichtig für die Verhältnisse beim *pn*-Übergang, aber auch beim Metall-Halbleiterkontakt und an der Halbleiteroberfläche.

2.4.3 Ströme infolge eines Konzentrationsgradienten: Diffusionsströme

Von Gasen weiß man, dass sie die Tendenz haben, den ganzen verfügbaren Raum gleichmäßig auszufüllen. Anders ausgedrückt besagt dies, dass Konzentrationsunterschiede der frei beweglichen Gasteilchen ausgeglichen werden. Dieses Verhalten gilt ganz all-

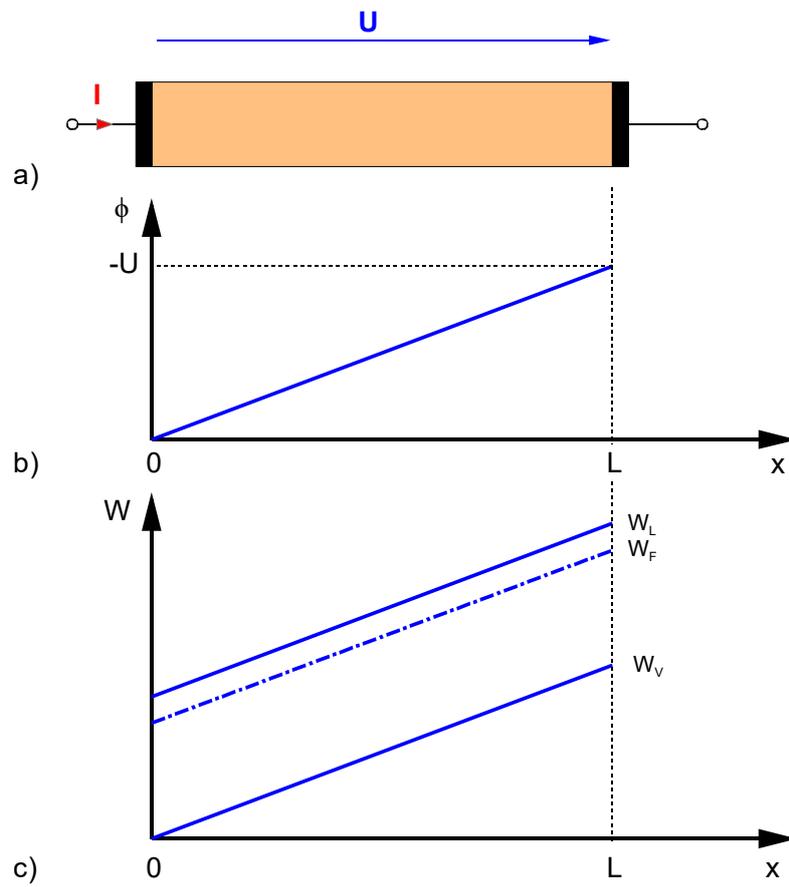


Abbildung 2.32: Verlauf der Bandkanten und des Potentials für einen Halbleiter im E-Feld

gemein für bewegliche Teilchen, auch für Elektronen und Löcher im Festkörper. Der Mechanismus des Ausgleichs von Konzentrationsunterschieden durch die Bewegung von Teilchen heißt Diffusion.

Für die Teilchenstromdichte gilt:

$$\vec{s} = -D\nabla n(x) \quad (2.88)$$

Sie ist umso größer, je steiler das Konzentrationsprofil ist, und ist in Richtung kleinerer Konzentrationen gerichtet. D wird Diffusionskoeffizient genannt. Er ist von der Temperatur abhängig, da die Energiequelle für die Diffusionsbewegung die thermische Energie ist, die für die thermische Geschwindigkeit verantwortlich ist (vgl. (2.62) und (2.63)). Nur für kleine Konzentrationen ist er näherungsweise konzentrationsunabhängig.

Sind die Teilchen elektrisch geladen, so ist mit dem Teilchendiffusionsstrom ein elektrischer Strom verbunden. Für Elektronen bzw. Löcher findet man

$$\begin{aligned} J_{n,\text{diff}} &= -q \cdot \vec{s}_n = qD_n \nabla n(x) \\ J_{p,\text{diff}} &= q \cdot \vec{s}_p = -qD_p \nabla p(x) \end{aligned} \quad (2.89)$$

In dem später zu behandelnden pn -Übergang werden diese Diffusionsströme erhebliche Beiträge liefern, da hier die Elektronen- und Löcherkonzentrationen sich um Größenordnungen ändern, weil sie beim Überschreiten des pn -Überganges, dessen Ausdehnung bei nur etwa $1 \mu\text{m}$ liegt, von Majoritäten zu Minoritäten werden (2.60). Im pn -Übergang müssen Diffusions- und Feldströme gemeinsam betrachtet werden.

Ohne eine detaillierte Behandlung kann bereits an dieser Stelle die wichtige **Einstein-Beziehung** zwischen Beweglichkeit und Diffusionskonstante hergeleitet werden, da sie auf ganz allgemein gültigen Prinzipien beruht: Zunächst gilt nach dem thermodynamischen "Prinzip des detaillierten Gleichgewichts", dass im Gleichgewicht zu jedem einzelnen Mikroprozess ein Gegenprozess existiert. Im Fall des pn -Überganges ist der zum Diffusionsprozess der Elektronen existierende Gegenprozess der Elektronenfeldstrom. Das treibende Feld entsteht aus dem Diffusionsladungsträgertransport, der Ladungen verschiebt und somit einen Potentialunterschied aufbaut. Wenn von außen keine Spannung anliegt, herrscht Gleichgewicht, und außen kann auch kein Strom fließen; sonst

hätte man ein Perpetuum Mobile gebaut. Es gilt also:

$$\vec{J}_n = 0 = \vec{J}_{n,\text{diff}} + J_{n,\text{Feld}} = q \left[\mu n \vec{E}(x) + D_n \nabla n(x) \right] \quad (2.90)$$

Mit

$$n(x) = N_L \cdot e^{-(W_L(x) - W_F)/kT}$$

gewinnt man aus (3.89)

$$0 = qnE(x) [\mu_n - D_n q/kT]$$

Ohne Herleitung wurde dabei benutzt, dass die Fermi-Energie W_F im Gleichgewicht konstant im ganzen betrachteten System, also ortsunabhängig ist. Feldstärke und Elektronenkonzentration sind aber ortsabhängig anzusetzen. Diese Gleichung hat als allgemeine Lösung die Einsteinsche Beziehung für die Elektronen:

$$\mu_n = D_n q/kT \quad (2.91)$$

bzw. für die Löcher

$$\mu_p = D_p q/kT \quad (2.92)$$

2.4.4 Ladungsträgertransport bei Einwirkung eines magnetischen Feldes

Auf in einem Magnetfeld \vec{B} mit der Geschwindigkeit \vec{v} bewegte Ladung Q wirkt die Lorentz-Kraft \vec{F}_L :

$$\vec{F}_L = Q \cdot (\vec{v} \times \vec{B}) \quad (2.93)$$

Die betrachtete Ladung sei ein Elektron ($Q = -q$). Besonders einfach wird (2.93), wenn \vec{v} und \vec{B} senkrecht aufeinander stehen. (vgl. Abb. 2.33). Das Kreuzprodukt \times kann dann einfach ausgewertet werden und führt zu einer skalaren Beziehung.

$$\vec{F}_L = -q \cdot v \cdot B \cdot \vec{e}_y \quad (2.94)$$

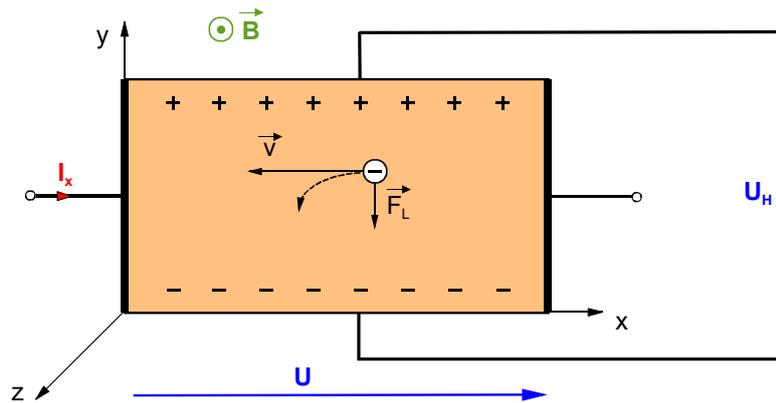


Abbildung 2.33: Ablenkung von bewegten Elektronen in einem Magnetfeld

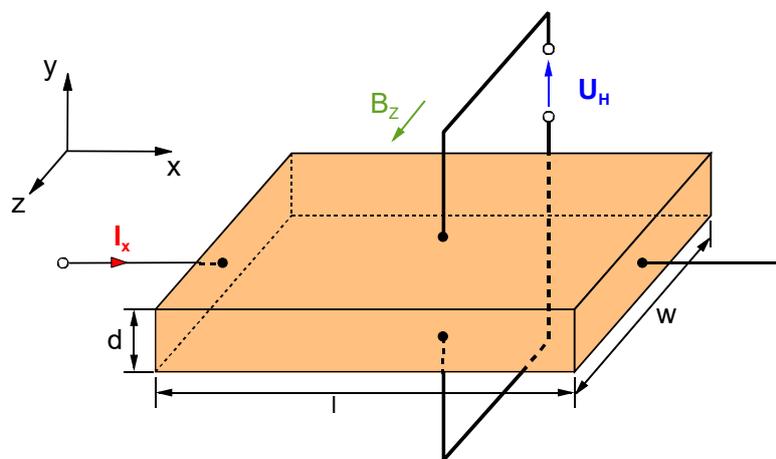


Abbildung 2.34: Richtung der Felder und Ströme im Hall-Effekt

Die Elektronen werden unter den Randbedingungen der Abb. 2.33 in negativer y -Richtung abgelenkt und führen zu einer negativen Aufladung der Probenunterseite, während durch Elektronenverarmung an der Oberseite dort eine positive Ladung aufgebaut wird. Diese Ladungstrennung erzeugt einen Potentialunterschied zwischen Ober- und Unterseite der Probe, der Hall-Spannung U_H . Diese Erscheinung ist als Hall-Effekt bekannt. Wie im Folgenden dargelegt, kann er vorteilhaft zur Messung der Ladungsträgerbeweglichkeit in Halbleitern benutzt werden.

2.4.5 Hall-Effekt

Es liegen die Verhältnisse nach Abb. 2.34 vor. Ferner nehmen wir an, dass zum Stromfluss nur Elektronen beitragen (n -Halbleiter), und diese sich mit der Geschwindigkeit

v bewegen, bevor das \vec{B} -Feld eingeschaltet wird. Die Bewegung mit v wurde durch eine Spannung U verursacht, die den Strom I_x einprägt. Die Lorentz-Kraft weist wieder in $(-y)$ -Richtung (2.94), und erzeugt aufgrund der Ladungstrennung ein elektrisches Feld in $(-y)$ -Richtung. Mit der Hall-Spannung U_H besteht der Zusammenhang:

$$U_H = -E_y \cdot d \quad (2.95)$$

Der Aufbau des Ladungsunterschiedes zwischen Ober- und Unterseite der Probe setzt sich aber nicht unbegrenzt fort, sondern wird durch das entstehende Feld E_y selbst begrenzt. Die Gleichgewichtsbedingung lautet

$$|\vec{F}_L| = |\vec{F}_{E_y}| \quad \text{bzw.} \quad qE_y = qvB_z \quad (2.96)$$

Mit (2.95) folgt:

$$U_H = -vdB_z$$

Für den Strom I_x gilt mit der Elektronenkonzentration n

$$I_x = qnvdW$$

so dass schließlich

$$U_H = -\frac{1}{nq} \cdot \frac{I_x B_z}{W} \quad (2.97)$$

Der erste Term wird als Hall-Faktor R_H bezeichnet

$$R_H = -\frac{1}{nq} \quad (2.98)$$

Es wurde bereits darauf hingewiesen, dass der Hall-Effekt zur messtechnischen Bestimmung der Beweglichkeit μ benutzt werden kann. Die verbindende Größe ist die Ladungsträgergeschwindigkeit, die sowohl für die Beweglichkeit (2.76) als auch für die Lorentz-Kraft eine entscheidende Rolle spielt. Die Kopplung beider Größen geschieht durch eine zusätzlich durchgeführte Messung der Leitfähigkeit κ **ohne** angelegtes Magnetfeld. Die Leitfähigkeit κ beinhaltet das Produkt von Ladungsträgerkonzentration n

und Beweglichkeit μ_n

$$\kappa_n = q n \mu_n$$

Während R_H nur die Ladungsträgerkonzentration n beinhaltet

$$R_H = -\frac{1}{qn}$$

Die Kombination beider Gleichungen ergibt die Messung der Beweglichkeit μ_n

$$\mu_n = \frac{1}{qn} \cdot \kappa_n = -R_H \kappa_n \quad (2.99)$$

Die Leitfähigkeit kann durch Messung des Ohmschen Widerstands in x -Richtung ermittelt werden

$$\kappa_n = I_x/U \cdot \frac{l}{dw} \quad (2.100)$$

Hier sind die Länge l , die Breite w und die Dicke d der Probe zu berücksichtigen.

Für einen p -Halbleiter, bei dem der Strom idealisiert nur von Löchern getragen wird, erhält man analog

$$R_H = +\frac{1}{qp} \quad (2.101)$$

Man beachte das invertierte Vorzeichen des Hall-Faktors und damit auch der Hallspannung im Vergleich zum n -Halbleiter. Ferner ist der Hall-Faktor eindeutig durch die Ladungsträgerkonzentration bestimmt.

Misst man also bei vorgegebener Probengeometrie, magnetischer Induktion B_z und eingepprägtem Strom I_x die Spannung U (bei abgeschaltetem Magnetfeld), anschließend die Hallspannung U_H , so kann aus dem Vorzeichen der Hall-Spannung sofort der Halbleitertyp (p - oder n -Typ) bestimmt werden.

Die Ergebnisse unserer Betrachtungen waren an die Probengeometrie gemäß Abb. 2.34 geknüpft. Van der Pauw hat gezeigt, dass die Ergebnisse auf beliebige, insbesondere auch planare Kontaktanordnungen übertragen werden können.

Die o.a. Ergebnisse für den Hall-Effekt gelten nur für

$$n \gg p \quad \text{oder} \quad p \gg n$$

In kompensationsdotierten Proben mit signifikanter Konzentration des jeweils anderen Dotiertyps wird mit dem Hall-Effekt eine mit den Ladungsträgerkonzentrationen gewichtete Mischung der Beweglichkeiten erhalten, aus einem komplizierteren Ansatz.

Die Geschwindigkeit der Ladungsträger wurde als uniform konstant angenommen. Es treten kleine Modifikationen am Wert des Hall-Faktors auf durch eine statistische Verteilung der Ladungsträgergeschwindigkeiten.

2.4.6 Kontinuitätsgleichungen

Bei den bisher behandelten Transporterscheinungen ist vorausgesetzt worden, dass die Ladungsträgerkonzentration konstant ist. Bei der Behandlung der Diffusionsströme wurde die örtliche Konstanz aufgegeben, nach wie vor galt aber die zeitliche Konstanz. Diese wollen wir jetzt ebenfalls aufgeben. Zwar kann man oft mit örtlichen und zeitlichen Mittelwerten der Konzentrationen rechnen, grundsätzlich führen Stromflussänderungen, Generation und Rekombination zu Konzentrationsänderungen gegenüber dem thermodynamischem Gleichgewicht. Diese Änderungen sowie die stationären Gleichgewichtswerte sollen im folgenden bestimmt werden.

Die drei Anteile Stromdichteänderungen, Generation sowie Rekombination sollen in dieser Reihenfolge behandelt werden.

Stromdichteänderungen

Betrachtet werden soll die Situation in Abb. 2.35. Die Teilchenzahl N ist gegeben durch

$$N = n \cdot \Delta x \cdot A = n \cdot (v \cdot \Delta t) \cdot A \quad (2.102)$$

wenn die Teilchen mit der Geschwindigkeit v gerade in der Zeit Δt die Strecke Δx durchlaufen. Die Teilchenstromdichte beträgt dann, z.B. am Ort x

$$s(x) = n \cdot v = \frac{N(x)}{\Delta t \cdot A} \quad (2.103)$$

Fließt nun nicht der gleiche Strom durch die Flächen A_1 bzw. A_2 , so ist die Teilchenzahldifferenz der in das Volumenelement $A \cdot \Delta x$ hinein- bzw. herausfließenden Teilchen

$$\Delta N = N(x) - N(x + \Delta x) = [s(x) - s(x + \Delta x)] \cdot A \Delta t \quad (2.104)$$

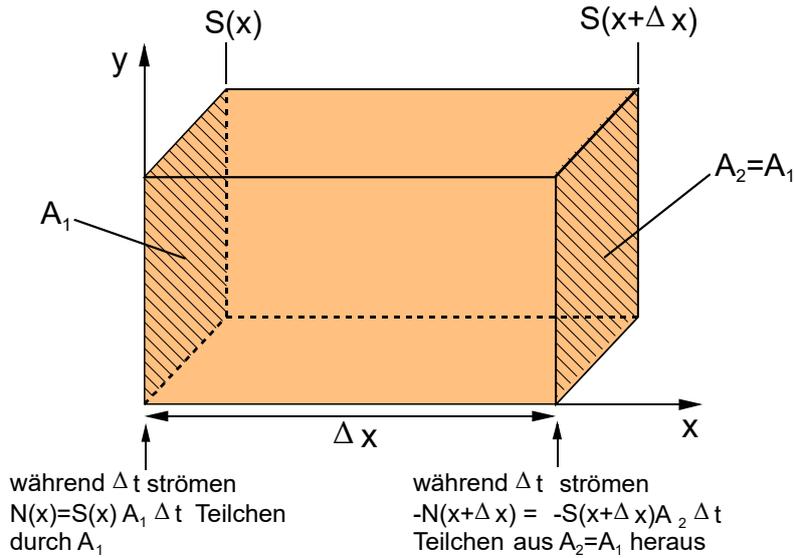


Abbildung 2.35: Kontinuität in beliebigem quaderförmigen Volumen

Eine Taylorentwicklung liefert

$$\Delta N = \left\{ s(x) - \left[s(x) + \Delta x \frac{\partial s(x)}{\partial x} \right] \right\} \cdot A \cdot t = -\frac{\partial s(x)}{\partial x} \cdot \Delta x \cdot A \cdot t \quad (2.105)$$

Die Teilchenzahländerung zieht eine Teilchendichteänderung nach sich:

$$\Delta n = \frac{\Delta N}{A \Delta x} \quad (2.106)$$

Somit folgt

$$\frac{\Delta N}{A \Delta x \Delta t} = \frac{\Delta n}{\Delta t} = -\frac{\partial s}{\partial x} \quad (2.107)$$

bzw. mit dem Grenzübergang $\Delta t \rightarrow 0$ und dreidimensionaler Rechnung

$$\frac{\partial n}{\partial t} = \dot{n} = -\nabla \cdot \vec{s} \equiv K_1 \quad (2.108)$$

als ersten Anteil K_1 zur zeitlichen Teilchendichteänderung.

Generation

Während durch die Stromdichteänderungen eine Zu- oder Abnahme der Teilchenkonzentration auftreten kann, so führt die Generation prinzipiell zu einer Zunahme. Als Generationsrate wird definiert

$$g = \frac{\text{Generationsprozesse}}{\text{Zeit} \cdot \text{Volumen}}$$

Daraus folgt ein zweiter Anteil K_2 zur zeitlichen Teilchedichteänderung

$$\frac{\partial n}{\partial t} = g \equiv K_2 \quad (2.109)$$

Rekombination

Dieser Prozess führt grundsätzlich zu einer Konzentrationsreduktion. Man definiert analog zu (2.109) als Rekombinationsrate

$$r = \frac{\text{Rekombinationsprozesse}}{\text{Zeit} \cdot \text{Volumen}}$$

und somit gilt für den dritten Anteil K_3 zur zeitlichen Teilchedichteänderung

$$\partial n \partial t = -r \equiv K_3 \quad (2.110)$$

Die vollständige Änderung der Teilchenkonzentration ist dann:

$$\frac{\partial n}{\partial t} = \sum_{i=1}^3 K_i = -\nabla \cdot \vec{s} + g - r \quad (2.111)$$

In Halbleitern werden die stromtragenden Teilchen entweder durch Elektronen oder Löcher repräsentiert. Teilchen- und elektrischer Strom von Elektronen und Löchern sind verknüpft über

$$\begin{aligned} \vec{J}_n &= -q\vec{s}_n \\ \vec{J}_p &= +q\vec{s}_p \end{aligned} \quad (2.112)$$

Somit folgt die endgültige Form der **Kontinuitätsgleichungen** für Elektronen

$$\frac{\partial n}{\partial t} = \dot{n} = \frac{1}{q} \nabla \cdot \vec{j}_n + g_n - r_n \quad (2.113)$$

und für Löcher

$$\frac{\partial p}{\partial t} = \dot{p} = -\frac{1}{q} \nabla \cdot \vec{j}_p + g_p - r_p \quad (2.114)$$

2.4.7 Poisson-Gleichung

Im Kap. 2.4.2 "Feldströme" wurde die elektrische Feldstärke als vorgegeben betrachtet, beispielsweise hervorgerufen durch eine außen angelegte Spannung. Im Kapitel 2.4.5 "Hall-Effekt" waren aber bereits Ladungen, hervorgerufen durch innere Effekte, aufgetreten, die zur Hall-Feldstärke bzw. -Spannung geführt haben. In diesem Kapitel soll der allgemein gültige Zusammenhang zwischen elektrischer Feldstärke und zugeordneter Raumladungsdichte behandelt werden.

Jede Raumladung ist Quelle bzw. Senke von Verschiebungsflussdichtelinien und damit auch Feldstärkelinien. Dies drückt eine der Maxwellgleichungen aus:

$$\nabla \cdot \vec{D} = \rho \quad (2.115)$$

mit der elektrischen Verschiebungsdichte $\vec{D} = \varepsilon_0 \varepsilon_r \vec{E}$. Mit (2.84)

$$\vec{E} = -\nabla \varphi$$

folgt

$$\nabla \cdot \vec{D} = -\varepsilon_0 \varepsilon_r \nabla \cdot \nabla \varphi = -\varepsilon_0 \varepsilon_r \Delta \varphi \quad (2.116)$$

Dabei wurde der Laplace'sche Operator

$$\Delta = \nabla \cdot \nabla = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

verwendet. Damit folgt die **Poisson-Gleichung**

$$\Delta \varphi = -\rho \varepsilon_0 \varepsilon_r \quad (2.117)$$

In Halbleitern setzt sich die Raumladung aus geladenen Störstellen sowie Elektronen und Löchern zusammen:

$$\rho = q(p - n + N_D^+ - N_A^-)$$

In den meisten bisher behandelten Fällen wurde davon ausgegangen, dass sich die beweglichen Ladungen (Elektronen, Löcher) durch die Ladung der ionisierten Störstellen etwa kompensieren. Dies ist der Fall in homogenen n - oder p -Halbleitern. In der Nähe von metallischen Kontakten und pn -Übergängen überwiegen oft die ortsfesten Ladungen und führen zu merklichen Raumladungen, die gemäß (2.115) ein elektrisches Feld aufbauen.

2.4.8 Zusammenstellung der Halbleitergrundgleichungen für den Ladungstransport

Hier soll der vollständige Satz der für den Ladungstransport relevanten Halbleitergrundgleichungen aufgelistet werden:

$$\begin{aligned} \vec{J}_n &= \vec{J}_{n,\text{Feld}} + \vec{J}_{n,\text{Diff}} = q\mu_n \left(n\vec{E} + \frac{kT}{q}\nabla n \right) \\ \vec{J}_p &= \vec{J}_{p,\text{Feld}} + \vec{J}_{p,\text{Diff}} = q\mu_p \left(p\vec{E} - \frac{kT}{q}\nabla p \right) \\ \frac{\partial n}{\partial t} &= \frac{1}{q}\nabla \cdot \vec{J}_n + g_n - r_n \\ \frac{\partial p}{\partial t} &= -\frac{1}{q}\nabla \cdot \vec{J}_p + g_p - r_p \\ \Delta\varphi &= -\nabla \cdot \vec{E} = -\frac{\rho}{\varepsilon_0\varepsilon_r} \end{aligned} \tag{2.118}$$

Daraus können grundsätzlich die Ströme, die Ladungsträgerkonzentrationen sowie die elektrische Feldstärke berechnet werden, wenn die Temperatur, die Beweglichkeiten, die Generations- und Rekombinationsraten und zusätzlich die außen angelegte Spannung vorgegeben sind.

3 Wechselwirkung zwischen elektromagnetischer Strahlung und Halbleitern

Aus Kap. 1.1 ist bereits bekannt, dass die Photonen die Elementarteilchen der elektromagnetischen Strahlung und damit auch des Lichts sind. Ihre Energie ist der Schwingfrequenz f und ihr Impuls dem Wellenzahlvektor \vec{k} proportional.

$$\begin{aligned} W &= h \cdot f = h \cdot c/\lambda \\ |\vec{p}| &= \hbar \cdot |\vec{k}| = h/\lambda = hf/c \end{aligned} \tag{3.1}$$

Im folgenden wird die Wechselwirkung dieser Strahlung bzw. dieser Teilchen mit Halbleitern behandelt. Grundsätzlich können die Photonen vom Halbleiter entweder absorbiert oder emittiert werden. Diese beiden Prozesse, Absorption und Emission, bilden die Grundlage für eine Fülle von Bauelementen der Optoelektronik.

3.1 Absorption von Licht in Halbleitern

Im Vakuum breitet sich Licht ungeschwächt aus, da eine Wechselwirkung mangels Partnern nicht vorhanden ist. Beim Durchgang durch Materie tritt jedoch eine Wechselwirkung in Form der Lichtabsorption in Erscheinung: pro Längeneinheit dx nimmt die Intensität der Strahlung um dI ab:

$$dI = -\alpha \cdot I dx \tag{3.2}$$

Bei nicht zu großen Intensitäten ist diese Abnahme proportional zur durchlaufenen Wegstrecke dx und der Intensität I . Der Faktor α heißt Absorptionskoeffizient und ist abhängig von der durchstrahlten Materie und von der Wellenlänge bzw. Photonenenergie.

3 Wechselwirkung zwischen elektromagnetischer Strahlung und Halbleitern

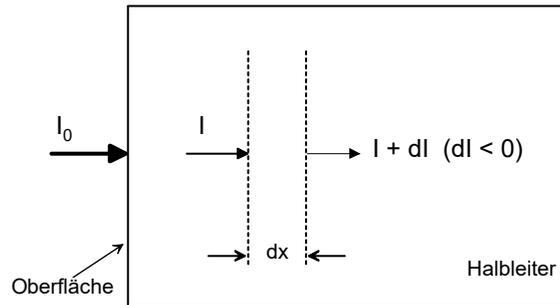


Abbildung 3.1: Lichtabsorption

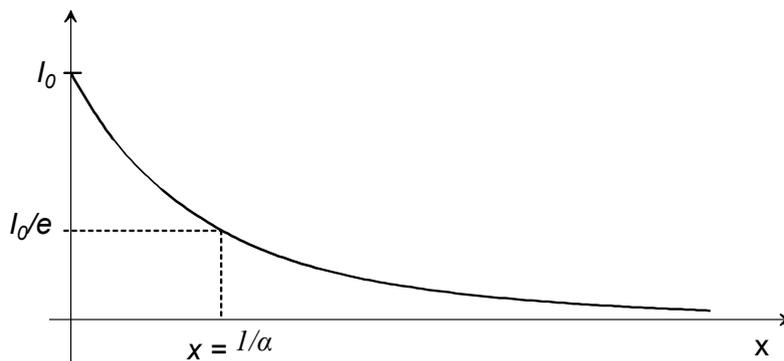


Abbildung 3.2: zum Absorptionsgesetz

gie des Lichtes (vgl. Abb. 3.3). Aus (3.2) kann man leicht das makroskopische Absorptionsgesetz herleiten:

$$I(x) = I_0 \cdot e^{-\alpha x} \quad (3.3)$$

Die Intensität klingt also exponentiell mit der Wegstrecke durch die Materie ab. Nach der Länge $x = 1/\alpha$ ist eine Abschwächung auf den Wert I_0/e messbar (Abb. 3.2).

In Abb. 3.3 ist die bereits angesprochene Abhängigkeit des Absorptionskoeffizienten α von der Materie und der Wellenlänge λ des Lichtes für die Halbleiter Silizium, Germanium und Galliumarsenid wiedergegeben. Das Verhalten des GaAs unterscheidet sich im Kurvenverlauf grundsätzlich von dem des Siliziums (Si) und des Germaniums (Ge) dadurch, dass der Anstieg von α zu kurzen Weglängen stärker ist und nur eine einzige (GaAs) statt zwei (Si, Ge) Schwellenwellenlängen erkennbar ist. Dieses Verhalten soll nun auf mikroskopischer Basis erläutert werden.

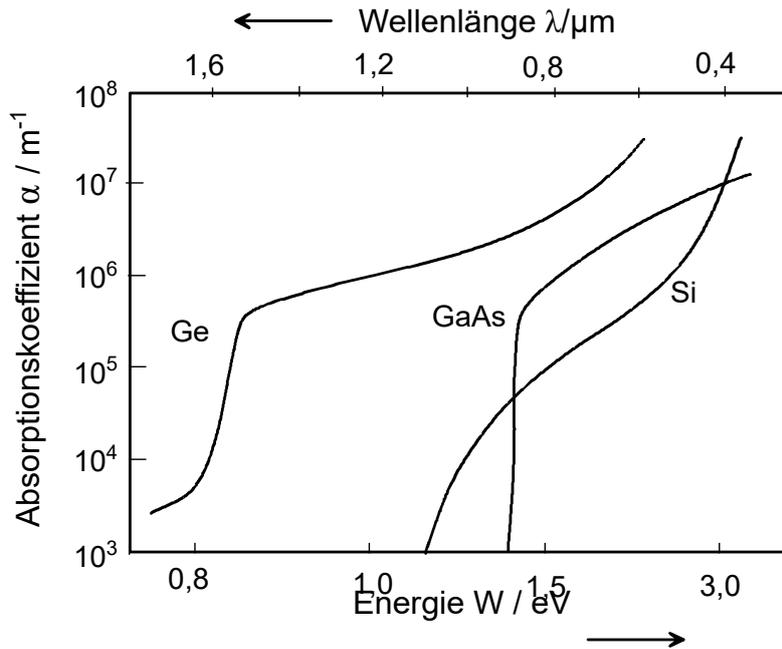


Abbildung 3.3: Absorptionskoeffizient in Abhängigkeit von der Wellenlänge für verschiedene Halbleiter

3.2 Einfluss der Bandstruktur auf die Absorption

Beim mikroskopischen Absorptionsprozess wird die Energie eines Photons auf ein Elektron des Festkörpers übertragen. Im intrinsischen Halbleiter befinden sich unter Normalbedingungen fast alle Elektronen im Valenzband oder noch tiefer liegenden Bändern. Die Elektronen besetzen auch nahezu alle verfügbaren Plätze in diesen Bändern. Sie können die Energie der Photonen nur derart aufnehmen, dass sie die Energielücke zwischen Valenz- und Leitungsband überspringen und Leitungsbandniveaus einnehmen, d.h. für eine merkliche Absorption muss gelten:

$$h \cdot f \geq W_g \quad (3.4)$$

In Abb. 3.4 ist der Bandverlauf $W(k)$ für einen direkten Halbleiter (z.B. GaAs) und der zugehörige Verlauf von α schematisch dargestellt. Bei nicht zu großen Intensitäten ist diese Abnahme proportional zur durchlaufenen Wegstrecke dx und der Intensität I . Der Faktor α heißt Absorptionskoeffizient und ist abhängig von der durchstrahlten Materie und von der Wellenlänge bzw. Photonenenergie des Lichts (vgl. Abb. 3.3). Aus (3.2) kann

3 Wechselwirkung zwischen elektromagnetischer Strahlung und Halbleitern

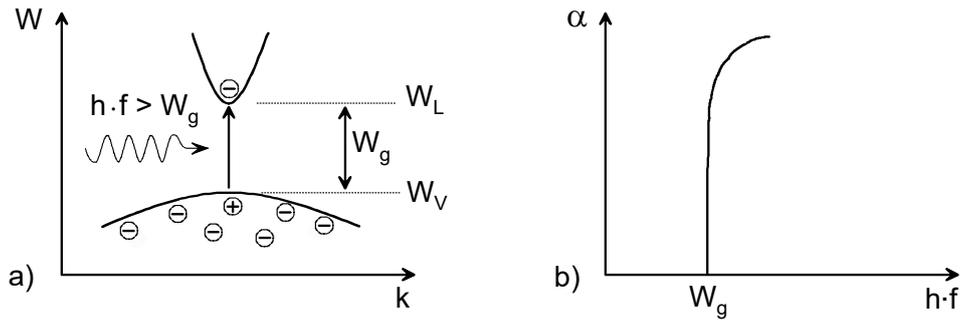


Abbildung 3.4: a) Bandverlauf für einen direkten Halbleiter und b) zugehöriger Verlauf von $\alpha(h\nu)$

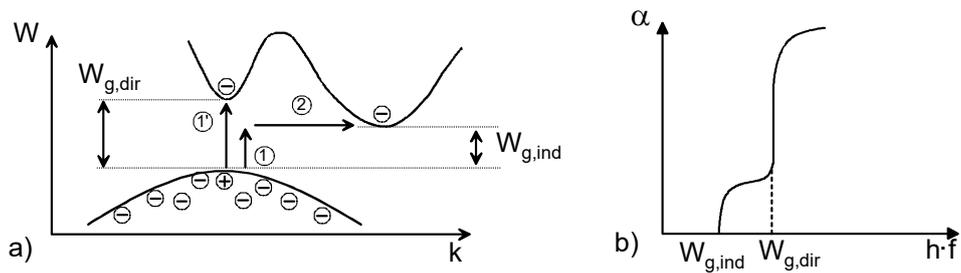


Abbildung 3.5: a) Bandverlauf für einen indirekten Halbleiter und b) zugehöriger Verlauf von $\alpha(h\nu)$

man leicht das makroskopische Absorptionsgesetz herleiten:

$$I(x) = I_0 \cdot e^{-\alpha x} \quad (3.5)$$

Die Intensität klingt also exponentiell mit der Wegstrecke durch die Materie ab. Nach der Länge $x = 1/\alpha$ ist eine Abschwächung auf den Wert I_0/e messbar (Abb. 3.2). Abb. 3.5 zeigt in Analogie zu Abb. 3.4 die Verhältnisse für einen indirekten Halbleiter (z.B. Si). Der direkte Übergang vom Valenz- ins Leitungsband, wie in Abb. 3.4 dargestellt, kann erst bei der höheren Energie $W_{g,dir}$ erfolgen. Zwar ist die Bandlücke $W_{g,ind}$ kleiner als $W_{g,dir}$, d.h. schon für $hf \geq W_{g,ind}$ liegen bereits verfügbare Zustände im Leitungsband vor. Sie können aber nur durch Mitwirkung eines Phonons (Gitterschwingung, vgl. Kap. 2.4.1) erreicht werden (Teil 2 des Prozesses 1,2), weil der Impulserhaltungssatz erfüllt werden muss (zur Erinnerung; Photonen besitzen einen vergleichsweise sehr kleinen Impuls).

Damit sind die Übergänge im $W(k)$ -Diagramm senkrecht. Phononen besitzen als Bewegungsform von massebehafteten Teilchen einen merklichen Impuls $\hbar \cdot k$, der im $W(k)$ -

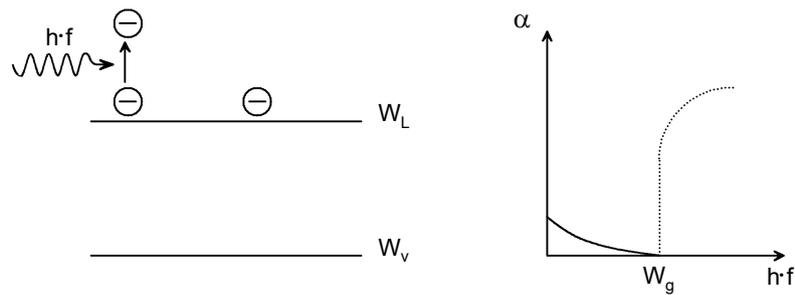


Abbildung 3.6: Absorption durch freie Ladungsträger

Diagramm als waagerechter Prozess (2) gezählt wird. Die gleichzeitige Wechselwirkung von drei Teilchen (Elektron, Photon und Phonon) ist aber sehr viel unwahrscheinlicher als die von zwei Teilchen (nur Elektron und Photon). Deshalb ist der Absorptionskoeffizient im Bereich $W_{g,ind} \leq hf \leq W_{g,dir}$ erheblich kleiner als für $hf \geq W_{g,dir}$, wie in Abb. 3.5b skizziert und in Abb. 3.3 für Si und Ge als Messkurve dargestellt ist.

3.2.1 Absorption durch freie Ladungsträger

Befinden sich bereits freie Ladungsträger im Leitungsband (oder Valenzband), so können diese innerhalb ihres Bandes Energie von Photonen aufnehmen. Dies gilt besonders für intrinsische Halbleiter bei hohen Temperaturen (oder für Halbleiter mit kleinem W_g) sowie für dotierte Halbleiter.

Der mikroskopische Prozess ist in Abb. 3.6 gezeigt zusammen mit dem Verlauf von $\alpha(hf)$. Die Abnahme von α mit zunehmendem hf hängt erneut mit der Impulserhaltung zusammen.

3.2.2 Störstellenabsorption

Zusätzliche Absorptionsprozesse laufen im extrinsischen Halbleiter unter Beteiligung von Störstellenniveaus ab. Abb. 3.7 zeigt schematisch die elektronischen Übergänge und den entsprechenden Verlauf von $\alpha(hf)$. Prinzipiell sind auch die in Abb. 3.7 gestrichelt eingezeichneten absorptiven Übergänge möglich und werden auch beobachtet.

3.3 Photoleitung

Die Photoleitung wird durch die Absorption von Photonen und die dadurch generierten Ladungsträger hervorgerufen. Außerdem ist sie ein einfaches Beispiel für die Anwen-

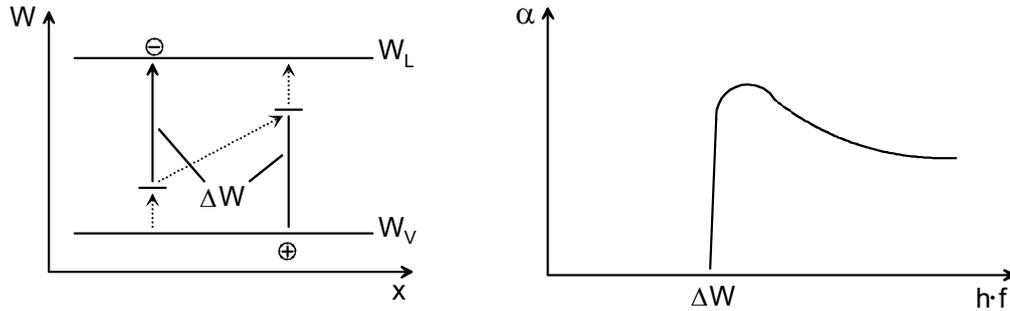


Abbildung 3.7: Störstellenabsorption im extrinsischen Halbleiter

derung der Kontinuitätsgleichungen.

Die Photoleitfähigkeit κ_{ph} sei definiert als die Differenz der Leitfähigkeiten mit (κ_1) und ohne (κ_0) Bestrahlung

$$\begin{aligned} \kappa_{ph} &= \kappa_1 - \kappa_0 \\ &= q \cdot [(qn_1\mu_n + p_1\mu_p) - (n_0\mu_n + p_0\mu_p)] \\ &= q \cdot [\Delta n \cdot \mu_n + \Delta p \cdot \mu_p] \quad (3.6) \end{aligned}$$

mit

$$\begin{aligned} \Delta n &= n_1 - n_0 \\ \Delta p &= p_1 - p_0 \end{aligned} \quad (3.7)$$

Dabei wurde vorausgesetzt, dass die Beweglichkeiten μ_n, μ_p sich durch die Beleuchtung nicht verändern. Ferner gelte in beiden Fällen (mit bzw. ohne Beleuchtung) das Ohm'sche Gesetz.

Die Größen Δn und Δp sind im allgemeinen Fall zeitabhängig und stellen die durch optische Anregung hervorgerufenen Überschusskonzentrationen dar.

Im folgenden wird die Berechnung der Photoleitfähigkeit exemplarisch für die Größe Δn durchgeführt, die Berechnung für Δp verläuft analog. Ausgangspunkt ist die Kontinuitätsgleichung für Elektronen (2.118)

$$\frac{\partial n}{\partial t} = \frac{1}{q} \nabla \cdot \vec{J}_n + g_n - r_n \quad (3.8)$$

Bei gleichmäßiger Beleuchtung und flächigem Aufbau ist der Strom im wesentlichen

ortsunabhängig, d.h.

$$\nabla \cdot \vec{J}_n = 0 \quad (3.9)$$

Auf den Photoleiter strahlen Photonen ein. Von diesen werden die Anzahl L pro Volumen und Sekunde absorbiert und erzeugen im Photoleiter jeweils ein Elektron-Loch-Paar, daher ist die Generationsrate

$$g_n = g_p = L \quad (3.10)$$

Die Einheiten der Generationsrate sind

$$[L] = [g] = \text{cm}^{-3}\text{s}^{-1} \quad (3.11)$$

Für die Rekombination unterscheidet man zwei Grenzfälle:

3.3.1 Reine Elektron-Loch-Rekombination

Dabei ist ein undotierter, im Idealfall intrinsischer Halbleiter, zugrunde gelegt. Wegen der paarweisen Erzeugung gilt dann:

$$\Delta n \approx \Delta p \quad (3.12)$$

Die Rekombinationsrate r_n ist dann sowohl zu Δn als auch zu Δp proportional, weil für die Rekombination das Zusammentreffen eines Elektrons und eines Lochs notwendig ist:

$$r_n = B \cdot \Delta n(t) \cdot \Delta p(t) \quad (3.13)$$

mit einer Proportionalitätskonstanten B .

Somit lautet mit (3.8) und (3.12) die Kontinuitätsgleichung:

$$\frac{\partial \Delta n(t)}{\partial t} = L - B \cdot \Delta n(t) \cdot \Delta p(t) \quad (3.14)$$

Stationärer Zustand mit Beleuchtung

In diesem Fall gilt:

$$\frac{\partial \Delta n(t)}{\partial t} = 0 \quad (3.15)$$

das heißt

$$\Delta n_1 = n_1 - n_0 = \text{const.} \quad (3.16)$$

und weiter folgt aus (3.14) mit (3.12)

$$\Delta n_1 = \sqrt{\frac{L}{B}} \quad (3.17)$$

Für $n_1 \gg n_0$ gilt schließlich

$$n_1 = \sqrt{\frac{L}{B}} \quad (3.18)$$

Einschaltvorgang

bei $t = 0$:

Ohne Beleuchtung sind die Ladungsträger im Gleichgewicht, $n = n_0$. Wegen

$$\Delta n(t \gtrsim 0) \approx 0$$

unmittelbar nach dem Einschalten der Beleuchtung gilt nach (3.13)

$$r_n(t \approx 0) \approx 0$$

Dann folgt aus (3.14):

$$\frac{\partial \Delta n(t)}{\partial t} = L \quad \text{für } t \gtrsim 0 \quad (3.19)$$

bzw.

$$\Delta n(t) = L \cdot t \quad \text{für } t \gtrsim 0 \quad (3.20)$$

und

$$n(t) = n_0 + L \cdot t \quad (3.21)$$

Mit zunehmender Zeit t wird $r_n \neq 0$ und n nähert sich schwächer als linear dem stationären Wert n_1 .

Ausschaltvorgang

bei $t = t_1$:

Der Ausschaltvorgang startet mit der erhöhten Konzentration $n(t_1) = n_1$ aufgrund der stationären Beleuchtung vor dem Ausschalten bei $t = t_1$. Nach dem Ausschalten ($L = 0$) gilt auch $g_n = 0$ (und $g_p = 0$) für $t > t_1$. Die Kontinuitätsgleichung lautet dann

$$\frac{\partial \Delta n(t)}{\partial t} = -B \cdot (\Delta n)^2 \quad (3.22)$$

Als Anfangsbedingung bei $t = t_1$ gilt $n(t = t_1) = n_1$ und $\Delta n(t = t_1) = n_1 - n_0 \equiv \Delta n_1$. Trennung der Variablen und Integration von (3.22) liefert

$$\int_{\Delta n_1}^{\Delta n(t)} \frac{d(\Delta n)}{(\Delta n)^2} = -B \int_{t_1}^t dt \quad (3.23)$$

mit der Lösung:

$$\Delta n(t) = \frac{\Delta n_1}{1 + \Delta n_1} \cdot B \cdot (t - t_1) \quad (3.24)$$

Die Überschusskonzentration $\Delta n(t)$ klingt also für $t \gg t_1$ mit $1/t$ ab. Definiert man als Zeitkonstante τ die Zeitspanne, nach der $\Delta n(t)$ auf $\Delta n_1/2$ abgenommen hat, so gilt :

$$\tau = \frac{1}{B \cdot \Delta n_1} = (L \cdot B)^{-1/2} \quad (3.25)$$

Die Zeitkonstante hängt daher von der Beleuchtungsstärke L ab: je stärker am Anfang beleuchtet wird, umso mehr weicht die Ladungsträgerkonzentration vom Gleichgewicht ab, und umso schneller klingt der Überschuss ab.

3.3.2 Rekombination über Rekombinationszentren

In diesem Fall sind sehr viele Störstellen (Donatoren, Akzeptoren, tiefe Störstellen) im Halbleiter vorhanden, über die die Elektronen bevorzugt einen Übergang ins Valenzband erreichen – im Gegensatz zur gerade behandelten direkten Elektron-Loch Rekombination.

Wenn ein räumlich festes Rekombinationszentrum den Weg für das Elektron ins Valenzband ermöglicht, dann ist die Löcherkonzentration nicht maßgeblich für den Rekombinationsprozess, sondern die Rekombinationsrate hängt nur noch proportional mit einer Konstanten C von der Elektronenüberschusskonzentration $\Delta n(t)$ ab

$$r_n = C \cdot \Delta n \quad (3.26)$$

Daraus kann direkt eine Zeitkonstante abgeleitet werden:

$$\tau_n = \frac{1}{C} \quad (3.27)$$

und

$$r_n = \frac{\Delta n}{\tau_n} \quad (3.28)$$

Somit schreibt sich die Kontinuitätsgleichung

$$\frac{\partial(\Delta n)}{\partial t} = L - \frac{\Delta n}{\tau_n} \quad (3.29)$$

Stationärer Zustand

In analoger Weise zur reinen Elektron-Loch-Rekombination in 3.3.1 ergibt sich

$$\Delta n_1 = L \cdot \tau_n \quad (3.30)$$

bzw. für $n_1 \gg n_0$

$$n_1 = L \cdot \tau_n \quad (3.31)$$

Einschaltvorgang

bei $t = 0$:

Hier erhält man das gleiche Ergebnis wie im Fall der reinen Elektron-Loch-Rekombination,

$$\Delta n = L \cdot t$$

und

$$n(t) = L \cdot t + n_0$$

Ausschaltvorgang

bei $t = t_1$: Es ergibt sich mit

$$\Delta n = \Delta n_1 \cdot e^{-(t-t_1)/\tau_n} \quad (3.32)$$

ein typisches exponentielles Abklingen der Überschusskonzentration.

In Abb. 3.8 ist das zeitliche Verhalten von Δn für beide Fälle wiedergegeben. Mit analogen Berechnungen für die Überschusskonzentration Δp der Löcher kann die gesamte Photoleitfähigkeit κ_{ph} auch zeitabhängig berechnet werden:

$$\kappa_{ph} = q (\Delta n \cdot \mu_n + \Delta p \cdot \mu_p)$$

.

3.3.3 Verstärkung in Photoleitern

Der Verstärkungsfaktor V ist definiert als (Abb. 3.9)

$$V = \frac{\text{Zahl der Ladungsträger, die den Photoleiter pro Sekunde durchlaufen}}{\text{Zahl der Photonen, die pro Sekunde absorbiert werden}} = \frac{\Delta n \cdot v_n \cdot A}{L \cdot A \cdot l} = \frac{\Delta n \cdot v_n}{L \cdot l} \quad (3.33)$$

Die Driftzeit durch den Photoleiter $T_d = l/v_n$ und die Lebensdauer der Ladungsträger führen mit dieser Definition auf

$$V = \tau/T_d \quad \text{bzw.} \quad V = \tau_n/T_d \quad (3.34)$$

3 Wechselwirkung zwischen elektromagnetischer Strahlung und Halbleitern

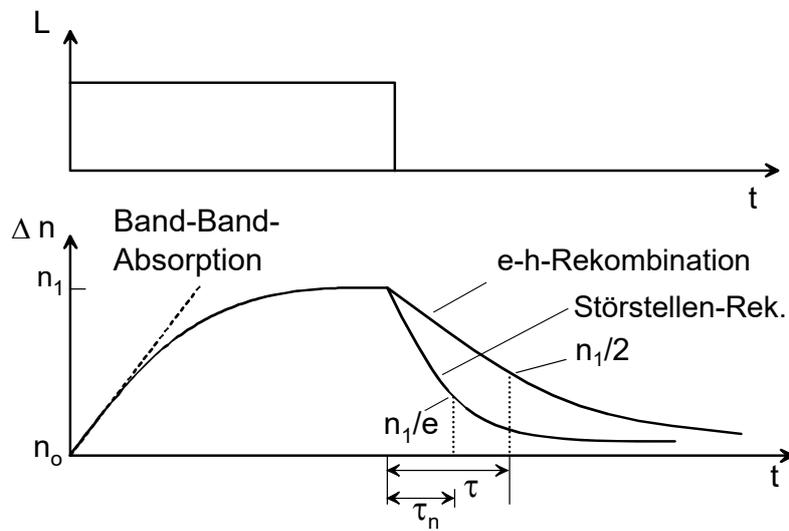


Abbildung 3.8: Zeitliches Verhalten der Überschusskonzentration

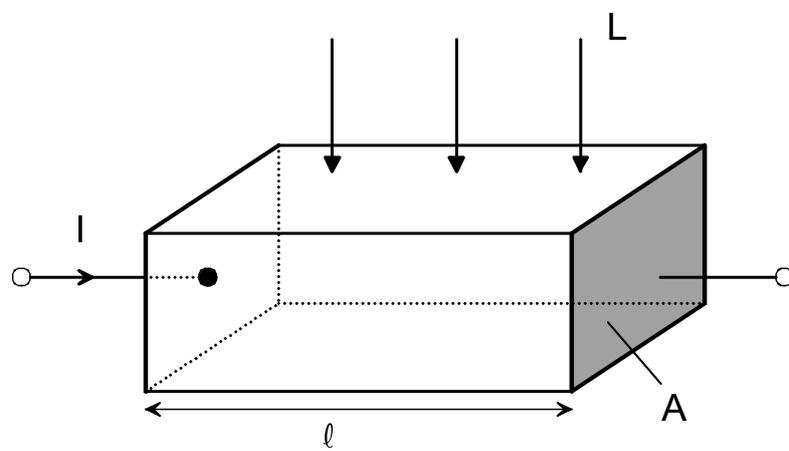


Abbildung 3.9: Aufbau eines Photoleiters

Hohe Verstärkungen erreicht man daher mit Photoleitern, bei denen die Trägerlebensdauer groß im Vergleich zur Driftzeit ist. Hier spielen sowohl Materialeigenschaften (v_n bzw. μ_n, τ_n) als auch die Bauelementgeometrie (l , fingerartige Kontaktstrukturen) eine wichtige Rolle.

3.4 Emission von Licht aus Halbleitern

Die Emission von Photonen aus Halbleitern und, allgemeiner, aus Festkörpern ist der zur Absorption inverse Prozess: Es befinden sich bereits Elektronen in höher liegenden Energieniveaus, meist dem Leitungsband, und diese fallen in tiefer liegende Niveaus unterhalb des Fermi-niveaus (Valenzband, Störstellenniveaus) unter Aussendung von Photonen der Energie, die der Niveaudifferenz entspricht.

Grundsätzlich wurde dieser Sachverhalt bereits im Kap. 1.6.6 bei der Behandlung der Spektrallinien besprochen. In diesem Fall der Photonenabstrahlung spricht man von strahlenden Übergängen, die von den nichtstrahlenden zu unterscheiden sind, bei denen die Niveaudifferenzenergie in Gitterschwingungen, also Phononen umgesetzt wird. Für den Aufbau effektiver Lichtemitter ist der letztgenannte Prozess so weit wie möglich zu unterdrücken.

Prinzipiell sind zwei Typen der Lichtemission zu unterscheiden, die spontane Emission, die Grundlage aller Lumineszenzerscheinungen ist, und die stimulierte Emission, die das Funktionsprinzip des Lasers ausmacht. Beide sollen im folgenden behandelt werden.

3.4.1 Spontane Emission

Wie der Name bereits aussagt, findet der elektronische Übergang spontan statt. Der genaue Zeitpunkt des Prozesses ist also nicht vorhersagbar und auch nicht zu beeinflussen, somit sind die Einzelprozesse völlig unkorreliert. Die mittlere Zeitspanne zwischen den Rekombinationsprozessen wird spontane Trägerlebensdauer τ_{sp} genannt.

Für die Emission von Photonen müssen Elektronen in höheren Niveaus bereitstehen, der Halbleiter befindet sich in einem angeregten Zustand (Nichtgleichgewicht).

Je nach Anregungsart hat die in der Folge auftretende Lumineszenzerscheinung unterschiedliche Namen:

Tabelle 3.1: verschiedene Anregungsarten der Lumineszenz

Anregung	Bezeichnung
Lichteinstrahlung	Photo-Lumineszenz (PL)
Elektroneneinstrahlung	Kathodo-Lumineszenz (KL)
Ladungsträgerinjektion	Elektro-Lumineszenz (EL)
Erwärmung	Thermo-Lumineszenz
Chemische Reaktion	Chemo-Lumineszenz
Mechanische Beanspruchung	Tribo-Lumineszenz

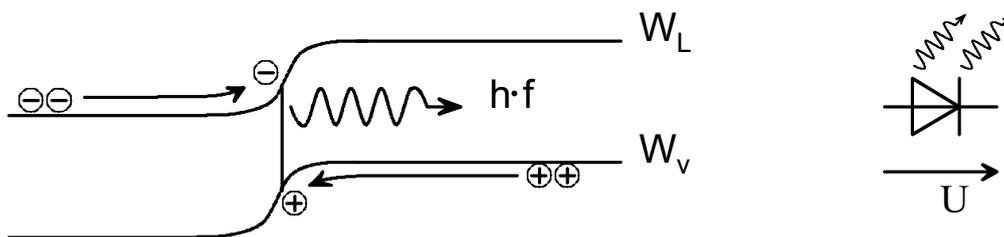


Abbildung 3.10: Wirkungsweise (a) und Symbol der Leuchtdiode (b)

Die beiden ersten Verfahren werden intensiv zur Analyse von Halbleitermaterial eingesetzt. Das dritte bildet die Grundlage für Leuchtdioden und Halbleiterlaser.

Die z.B. in der Beleuchtungstechnik eingesetzten Leuchtdioden (LED = light emitting diode) haben als Kernbereich einen pn -Übergang, in dem bei Polung in Durchlassrichtung Elektronen aus dem n -Gebiet und Löcher aus dem p -Gebiet aufeinander zugetrieben, d.h. injiziert werden, und im Bereich des Überganges zum großen Teil strahlend rekombinieren (Abb. 3.10).

Bei Leuchtdioden, auch Lumineszenzdioden genannt, läuft die Rekombination spontan ab, die Leuchterscheinung ist daher eine Lumineszenz. Bei Halbleiterlasern tritt durch zusätzliche Maßnahmen die stimulierte Emission auf, die im folgenden Abschnitt behandelt wird.

Die fünfte Art der Anregung (Chemo-Lumineszenz) machen sich beispielsweise die Glühwürmchen zunutze, indem durch eine chemische Reaktion Energie freigesetzt wird, die in Form von Licht abgestrahlt wird.

Die Lumineszenzerscheinungen werden in die Typen Fluoreszenz und Phosphoreszenz eingeteilt, wobei grundsätzlich für den ersten Typ die Lichtemission nach Abschalt-

ten der Anregung unmittelbar abbricht, und für den zweiten Typ noch eine Weile andauert (Nachleuchten). Der Übergang ist in der Realität allerdings fließend und hängt von der Lebensdauer der Elektronen im angeregten Zustand ab.

3.4.2 Stimulierte Emission

Das Prinzip der stimulierten Emission wurde schon lange vor der experimentellen Bestätigung und der technischen Nutzung von A. Einstein erkannt und formuliert.

Der Gedankengang sei kurz skizziert: Bei der Wechselwirkung eines Strahlungsfeldes (Photonen) mit einem System von Atomen (z.B. Festkörper, aber auch Gase) müssen sich im thermodynamischen Gleichgewicht die Absorptions- und Emissionsvorgänge einander die Waage halten. Kann man das System von Atomen durch zwei Energieniveaus W_2 und W_1 mit $W_2 > W_1$ beschreiben (für Halbleiter ist $W_2 = W_L$ und $W_1 = W_V$), so treten Emission und Absorption resonanzartig und somit näherungsweise ausschließlich bei $hf = W_2 - W_1$ auf.

Nach Einstein sind nun drei Arten von Wechselwirkungsprozessen zwischen den Niveaus zu unterscheiden:

- (a) **Spontane Emission:** bereits oben beschrieben. Wesentlich ist, dass er unbeeinflusst vom Strahlungsfeld abläuft (vgl. Abschnitt 3.4.1). Man kann ihn mit Hilfe der Wahrscheinlichkeit P_{sp} für den elektronischen Übergang $W_2 \rightarrow W_1$ beschreiben:

$$P_{sp} = P_{21}^A = \frac{1}{\tau_{sp}} \quad (3.35)$$

Sie ist von der Struktur des atomaren Systems (Bandstruktur des Festkörpers) und der Temperatur abhängig, aber nicht von der Anregungsstärke.

- (b) **Stimulierte Emission:** der elektronische Übergang ist derselbe wie bei der spontanen Emission, die damit verbundene Emission des Photons wird jedoch von dem vorliegenden Strahlungsfeld, also von bereits vorhandenen Photonen, verursacht, "induziert" bzw. "stimuliert". Es ist naheliegend, die Übergangswahrscheinlichkeit proportional zur Energiedichte des Strahlungsfeldes $u(h \cdot f)$ anzusetzen:

$$P_{st} = u(hf) \cdot P_{21}^B = \frac{1}{\tau_{st}} \quad (3.36)$$

3 Wechselwirkung zwischen elektromagnetischer Strahlung und Halbleitern

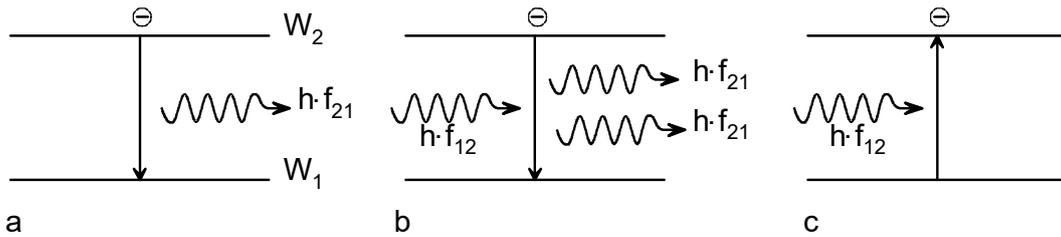


Abbildung 3.11: Die drei Wechselwirkungsprozesse zwischen Strahlungsfeld und atomarem System: a) spontane Emission, b) stimulierte Emission, c) (stimulierte) Absorption

(c) **Absorption:** der elektronische Übergang vom tiefer liegenden Niveau W_1 zum höher liegenden Niveau W_2 "verbraucht" ein Photon. Dieser Prozess kann nur bei Anwesenheit von Photonen stattfinden, und stellt daher einen stimulierten Prozess dar. Als Umkehrprozess zum stimulierten Emissionsprozess kann die Absorption analog beschrieben werden:

$$P_{abs} = u(hf) \cdot P_{12}^B \quad (3.37)$$

Das Analogon zur spontanen Emission fehlt. In Abb. 3.11 sind diese Prozesse schematisch dargestellt.

Im thermodynamischen Gleichgewicht müssen sich Absorptions- und Emissionsprozesse die Waage halten. Die Besetzung der beiden Niveaus folgt aus der Boltzmann Verteilung (vgl. Kap. 2.2) proportional zu $\exp(-W/kT)$, sodass die folgende Gleichgewichtsbedingung zwischen Emissionsrate (linke Seite) und Absorptionsrate (rechte Seite) formuliert werden kann:

$$e^{-W_2/kT} \cdot \left(\underbrace{P_{21}^A}_{\text{spontan}} + \underbrace{u(hf) \cdot P_{21}^B}_{\text{stimuliert}} \right) = e^{-W_1/kT} \cdot u(hf) \cdot P_{12}^B \quad (3.38)$$

Im Planck'schen Strahlungsgesetz finden wir neben A_{21} auch B_{12} und B_{21}

$$u(hf) = \frac{P_{21}^A}{P_{12}^B \cdot e^{hf/kT} - P_{21}^B} \quad (3.39)$$

Wenn die Temperatur T gegen unendlich strebt, sollte auch die Energiedichte u unend-

lich werden:

$$\lim_{T \rightarrow \infty} u(hf) = \infty$$

Da P_{21}^A konstant ist, wird der Bruch unendlich, wenn der Nenner gegen Null strebt:

$$\lim_{T \rightarrow \infty} P_{12}^B \cdot e^{hf/kT} - P_{21}^B = P_{12}^B \cdot 1 - P_{21}^B = 0$$

Daraus folgt

$$P_{21}^B = P_{12}^B \quad (3.40)$$

Dies demonstriert deutlich die Analogie von stimulierter Emission und Absorption: die Übergangswahrscheinlichkeiten der beiden B -Prozesse sind identisch. Ferner ist aus (3.38) ersichtlich, dass mit stärker werdendem Strahlungsfeld die stimulierte Emission gegenüber der spontanen zunimmt. Stellt man also in geeigneter Weise zusätzlich Photonen der Energie $h \cdot f_{21}$ zur Verfügung, so werden dadurch weitere Photonen erzeugt. Wie Abb. 3.11b zeigt, tritt dann eine Lichtverstärkung auf, von welcher der LASER seinen Namen ableitet (Light Amplification by Stimulated Emission of Radiation).

3.4.3 Halbleiter-Laser

Wie bei der Lumineszenzdiode ist der für die Bauelementfunktion wichtige Bereich des Lasers als pn -Übergang ausgebildet, der in Durchlassrichtung betrieben wird. Die Elektronen und Löcher werden in den Übergangsbereich an der Grenzfläche zwischen p - und n -Gebiet injiziert und rekombinieren teilweise strahlend.

Die Injektion von Ladungsträgern durch einen von außen aufgeprägten Strom (äußere Spannungsquelle) führt dem System elektrische Energie zu, und sorgt für eine Abweichung der Ladungsträgerkonzentrationen vom Gleichgewicht, die durch Rekombinationsprozesse abgebaut wird.

Der strahlende Prozess ist für Halbleiter mit **direkter Bandstruktur** (vgl. Abb. 3.4 und 3.5) besonders effektiv, weil sich hier Elektronen und Löcher bei identischem Wellenzahlvektor \vec{k} "gegenüberstehen" und im Rekombinationsprozess keinen weiteren Partner (Phonon) für die Erfüllung der Impulserhaltung benötigen. Im **indirekten Halbleiter** befinden sich nahezu alle Elektronen im Nebenminimum, für die kein Loch mit

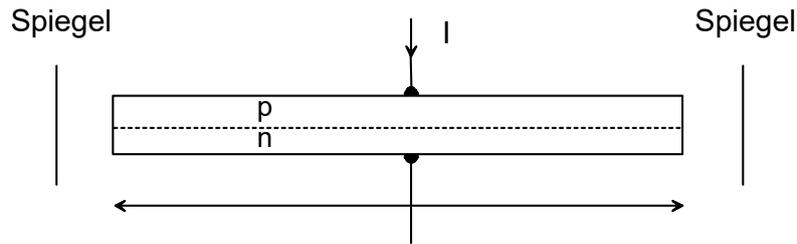


Abbildung 3.12: prinzipieller Aufbau eines Halbleiterlasers

passendem \vec{k} existiert.

Wird der Durchlassstrom des pn -Überganges stetig erhöht, so tritt eine wachsende spontane Emission von Photonen auf. Für eine hohe Photonendichte werden große Ladungsträgerdichten benötigt: der Halbleiter muss p - und n -seitig sehr hoch dotiert, d.h. **entartet** sein, so dass gelten kann:

$$W_{Fn} > W_L \quad \text{und} \quad W_{Fp} < W_V$$

oder allgemein:

$$W_{Fn} - W_{Fp} > W_L - W_V = W_g \quad \text{Laserbedingung 1} \quad (3.41)$$

Nur so kann eine ausreichende "Besetzungsinversion" aufgebaut werden, d.h. im höheren Niveau (Leitungsbandkante) befinden sich mehr Elektronen als im tiefer liegenden (Valenzbandkante).

Ohne zusätzliche Maßnahmen kommt es allerdings nicht zum Auftreten stimulierter Emission, weil das Strahlungsfeld aufgrund der stetig mit Lichtgeschwindigkeit abfließenden Photonen zu schwach ist. Zur weiteren Steigerung der Photonendichte werden die ausgesandten Photonen durch etwas durchlässige Spiegel zum größten Teil wieder in den Bereich des pn -Übergangs zurückgekoppelt (Abb. 3.12). Das Volumen zwischen den Spiegeln bildet einen optischen Resonator, die Lichtwellen laufen zwischen den Spiegeln hin und her. Dies ist analog zu einem elektrischen Schwingkreis, in dem Ladung zwischen Kapazität und Induktivität oszilliert. Anhand der Anordnung in Abb. 3.12 kann eine zweite Bedingung für eine Nettoverstärkung und damit die Laseraktion hergeleitet werden.

Man nimmt an, dass am linken Rand der Anordnung eine Elektron-Loch Rekombination spontan stattfindet und ein Photon in Richtung des rechten Randes startet. Die

entsprechende Lichtwelle kann auf ihrem Weg sowohl absorbiert werden, als auch weitere Rekombinationen und Photonen stimulieren. Ferner wird ein Teil der Energie der Lichtwelle durch die teilweise durchlässigen Spiegel der Resonatoranordnung verlorengehen. Es gibt also Verluste (Absorption, Spiegel) und Gewinne (Emission), und eine Netto-Verstärkung ist nur gegeben, wenn die Gewinne die Verluste bei einem vollständigen, geschlossenen Umlauf der Lichtwelle im Resonator zumindest minimal übersteigen.

Für die Berechnung eines Umlaufs verwenden wir gleiche Reflexionsfaktoren $R < 1$ für beide Spiegel und das Absorptionsgesetz nach (3.3) mit einer Weglänge $2l$. Für den Gewinn g wird ein analoger Zusammenhang mit positivem Vorzeichen im Argument der Exponentialfunktion angesetzt.

$$R^2 \cdot \exp(-2\alpha l + 2gl) \geq 1$$

bzw. nach Ziehen der Wurzel

$$R \cdot \exp(-\alpha l + gl) \geq 1 \quad (3.42)$$

Für den Gewinn g gilt danach:

$$g \geq \frac{1}{l} \cdot \ln\left(\frac{1}{R}\right) + \alpha \quad \text{Laserbedingung 2} \quad (3.43)$$

Die Spiegelverluste nehmen im Vergleich zu den Absorptionsverlusten mit kürzer werdenden Resonatoren zu. In Abb. 3.13 ist schematisch ein Halbleiterlaser aus dem Materialsystem AlGaAs/GaAs dargestellt. Die Verwendung unterschiedlicher Materialien (GaAs/AlGaAs-Heterostruktur) dient sowohl einer effektiven Konzentration von Elektronen und Löchern im Bereich des pn -Überganges (Abb. 4.14) als auch der Konzentration der Lichtwelle im "aktiven" Bereich. Die optische Welle wird durch unterschiedliche Brechungsindizes geführt: $n(\text{GaAs}) > n(\text{AlGaAs})$, die optische Welle wird zum dichteren Medium (höheres n) hin gebrochen und wird daher im GaAs-Material "geleitet", man spricht in diesem Zusammenhang auch von einem Wellenleiter. Die Resonatorlänge beträgt typisch $50 \mu\text{m}$ bis $500 \mu\text{m}$.

Die Emissionswellenlänge ist von der Bandlücke abhängig. Für verschiedene photonische Anwendungen werden daher unterschiedliche Materialsysteme gewählt. Einige

3 Wechselwirkung zwischen elektromagnetischer Strahlung und Halbleitern

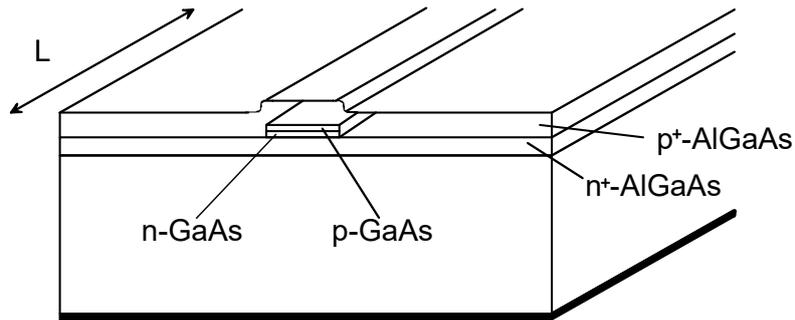


Abbildung 3.13: AlGaAs/GaAs Streifenlaser mit vergrabener Doppelheterostruktur (Buried Heterostructure (BH)-Laser)

Tabelle 3.2: Halbleitersysteme und optische Anwendungen

Material	Wellenlänge/nm	Anwendungsbeispiele
GaAs / $\text{Ga}_x\text{Al}_{1-x}\text{As}$	9XX	kW-Pumplaser, Materialbearbeitung
$\text{Ga}_x\text{In}_{1-x}\text{As}_y\text{P}_{1-y}$	1300 und 1550	Datenübertragung mit Glasfasern
$\text{In}_x\text{Ga}_{1-x}\text{N}$	450 – 520	Beleuchtungstechnik (LED), Farblaser (blau, grün), Bildprojektoren
$\text{Al}_x\text{Ga}_{1-x}\text{N}$	100 – 380	UV-LED (Desinfektion, Lithografie)

für *pn*-Halbleiterlaser verwendete Materialsysteme sind in Tab. 3.2 zusammengestellt.

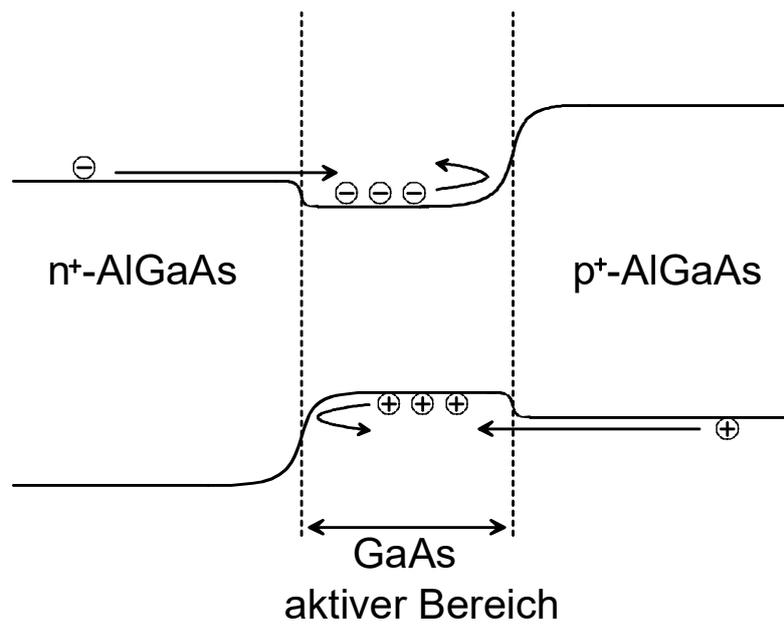


Abbildung 3.14: Doppel-Heterostruktur zur Erhöhung der lokalen Ladungsträgerkonzentration im aktiven Laserbereich

4 Halbleiteroberfläche

Bisher wurden – bis auf einige vorweggenommene Anwendungsbeispiele – ausschließlich homogene Halbleiter behandelt. Insbesondere wurden **Randeffekte** und **Grenzflächen** nicht diskutiert. In allen Halbleiterbauelementen spielen diese aber gerade die funktionsbestimmende Rolle!

In diesem Kapitel soll, gewissermaßen als Einführung in die Problematik, die Halbleiteroberfläche, d.h. die Halbleiter-Vakuum-Grenzfläche, besprochen werden. Zunächst ist klar, dass an der Halbleiteroberfläche aufgrund des Fehlens der Nachbaratome offene Bindungen vorliegen (Abb. 5.1).

Diese offenen Bindungen verursachen, ähnlich wie die nicht abgesättigten Bindungen im Inneren von p-Halbleitern (vgl. Abschnitt 2.3), zusätzliche Zustände in der Bandstruktur: die **Oberflächenzustände**. Die Situation in Abb. 4.1 ist allerdings stark idealisiert, sie kann beispielsweise durch Spalten einer Halbleiterprobe im Ultrahochvakuum höchstens einige Sekundenbruchteile aufrechterhalten werden. Die folgenden Prozesse überführen die idealisierte, aber energetisch ungünstige Situation in eine reale Oberfläche.

Oberflächenkonstruktion Die freien Bindungen orientieren sich zurück ins Halbleiterinnere, so dass wieder Paarbindungen entstehen. Da mit chemischen Bindungen auch immer Kraftwirkungen verbunden sind, kommt es bei dieser Umorientierung zu Verschiebungen der Oberflächenatome. Die Anzahl der Oberflächenzustände wird dabei

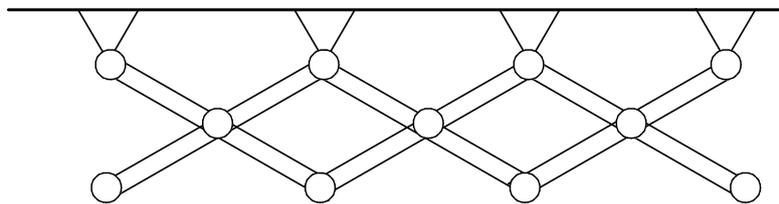


Abbildung 4.1: Bindungsverhältnisse an einer idealisierten Halbleiteroberfläche

gegenüber der idealisierten Oberfläche reduziert.

Chemische Reaktionen Die offenen Bindungen werden durch Fremdatome aus der Umgebung abgesättigt. Je nach der Chemie der Fremdatome können donator- oder akzeptorähnliche Oberflächenzustände erzeugt oder abgesättigt werden. Die Oxidation der Siliziumoberfläche, bildet die Grundlage für alle MOS (Metal-Oxide-Semiconductor)-Bauelemente.

4.1 Bändermodell der Halbleiteroberfläche

Es soll beispielhaft die Oberfläche eines p -Halbleiters behandelt werden, die donatorartige Oberflächenzustände (bei der Energie W_s) ausbildet. Im Augenblick des Entstehens (z. B. durch Spalten im Vakuum) sind diese Zustände noch neutral, und bis auf die erzeugten Zustände selbst entspricht das Bänderdiagramm dem Inneren des homogenen Halbleiters (Abb. 4.2a). Da aber W_s deutlich oberhalb der Fermi-Energie liegt, werden die meisten donatorartigen Oberflächenzustände ihr Elektron abgeben und für eine positive Ladung Q_{ss} der Oberfläche sorgen, die proportional zur Dichte N_s dieser Zustände ist:

$$\begin{aligned} Q_{ss} &= qN_s \cdot [1 - f(W_s)] \\ &= qN_s \left(1 - \frac{1}{e^{(W-W_s)/kT} + 1} \right) \approx qN_s(1 - 0) \approx qN_s \quad (4.1) \end{aligned}$$

Die freigesetzten Elektronen werden in Oberflächennähe mit den Löchern des p -Halbleiters rekombinieren und hinterlassen nicht kompensierte, negativ geladene Akzeptorrümpfe. Aus der Forderung der Ladungsneutralität folgt für diese Raumladung Q_R

$$Q_R = -qN_A^- \cdot l = -Q_{ss} \quad (4.2)$$

mit l als die Tiefe, bis zu der die Oberfläche an Löchern verarmt und die Akzeptoren nicht kompensiert sind: Es tritt also eine Verarmung der Oberfläche an freien Ladungsträgern auf, in diesem Falle von Löchern. Aus (4.2) folgt für die Konzentrationen

$$N_s^+ = N_A^- \cdot l \quad (4.3)$$

4 Halbleiteroberfläche

Ladungen verursachen ein elektrisches Feld, das nach Abschnitt 2.4.2 grundsätzlich einhergeht mit geneigten Bandkanten und einem entsprechenden ortsabhängigen Potentialverlauf. Bandkanten- und Potentialverlauf können mit Hilfe der Poisson-Gleichung im Bereich der Oberflächenverarmung (mit Tiefe l) berechnet werden:

$$\Delta\varphi = -\frac{\rho}{\varepsilon_0\varepsilon_r}$$

mit

$$\rho = \begin{cases} q(p(x) - N_A^-) & \text{für } 0 \leq x \leq l \\ 0 & \text{sonst} \end{cases} \quad (4.4)$$

wobei angenommen wird, dass $p(x)$ bei $x = l$ abrupt von p_0 auf Null abfällt unter der Voraussetzung, dass für das Oberflächenpotential φ_s gilt, dass $\varphi_s \gg kT$ (vgl. Abb. 4.2). Für den hier zu betrachtenden eindimensionalen Fall ist also die folgende Differentialgleichung zu lösen

$$\frac{d^2\varphi}{dx^2} = \frac{qN_A^-}{\varepsilon_0\varepsilon_r} \quad (4.5)$$

mit den Randbedingungen

$$\varphi = 0 \quad \text{und} \quad \frac{d\varphi}{dx} = 0 \quad \text{für } x = l$$

Die Lösungen lauten:

$$\begin{aligned} \frac{d\varphi}{dx} &= -\frac{qN_A^-}{\varepsilon_0\varepsilon_r} (l - x) = -E(x) \\ \varphi(x) &= +\frac{qN_A^-}{2\varepsilon_0\varepsilon_r} (l - x)^2 \end{aligned} \quad (4.6)$$

Insbesondere ist bei $x = 0$ das Oberflächenpotential φ_s gegeben als

$$\varphi(x = 0) = \varphi_s = \frac{qN_A^-}{2\varepsilon_0\varepsilon_r} l^2 \quad (4.7)$$

4 Halbleiteroberfläche

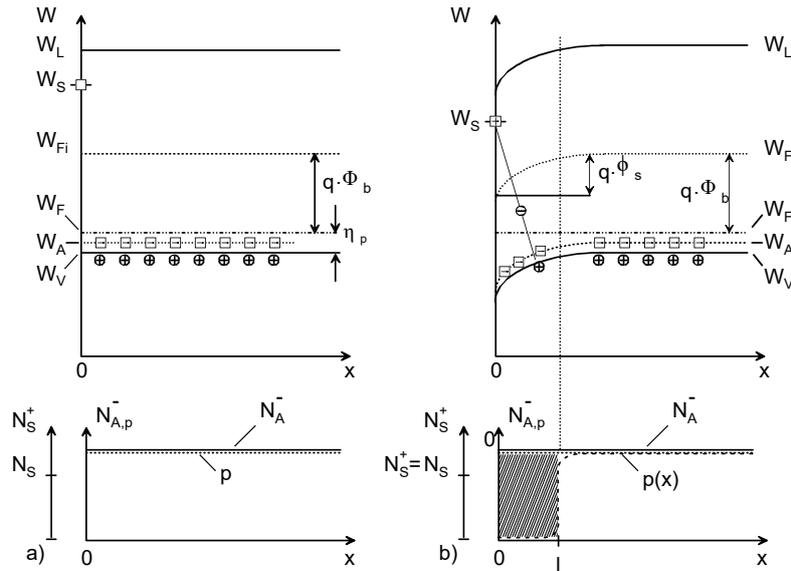


Abbildung 4.2: Bändermodell und Ladungsverteilung an einer Halbleiteroberfläche: a) unmittelbar nach Erzeugung", b) im Gleichgewicht

Die Bandkanten folgen gemäß (2.85), (2.86) dem Potential, sodass an der Oberfläche gegenüber dem Inneren des Halbleiters gilt

$$\Delta W = \Delta W_L = \Delta W_V = \Delta W_A = \Delta W_{Fi} = -q\varphi_s = \frac{q^2 N_A^- l^2}{2\epsilon_0 \epsilon_r} \quad (4.8)$$

Hier ist zu beachten, dass die Verbiegung von Leitungs- und Valenzband gleichförmig ist, da an jeder Position im Raum der Energieabstand $W_L - W_V = W_g$ der Bandlücke entsprechen muss – die Bandlücke W_g ist unverändert durch äußere Felde, sie hängt nur von den inneren Gegebenheiten des Kristalls ab (vgl. Kronig-Penney-Modell in 1.7).

Mit Hilfe von (4.3) werden (4.6) und (4.8) eindeutig lösbar, wenn die Konzentrationen bekannt sind:

$$\varphi_s = \frac{q(N_s^+)^2}{2\epsilon_0 \epsilon_r N_A^-} \quad (4.9)$$

Beispiel Man kann leicht nachrechnen, dass bei typischen Werten $\epsilon_r = 12$, $N_A = 5 \times 10^{16} \text{ cm}^{-3}$, $N_s = 5 \times 10^{11} \text{ cm}^{-2}$ folgt, dass die Verarmungstiefe $l = 0.1 \mu\text{m}$ und das Oberflächenpotential $\varphi_s = 0.4 \text{ V}$ ist. Mit Hilfe von (4.6) kann dann auch $p(x)$ angegeben

4 Halbleiteroberfläche

werden. Allgemein gilt (2.37) und (2.46) für die Löcherkonzentration

$$p = N_V \cdot e^{-(W_F - W_V)/kT} = N_V \cdot e^{-\eta_p/kT} \quad \text{mit } \eta_p = W_F - W_V$$

Im Fall der Halbleiteroberfläche sind W_V und somit auch p ortsabhängig

$$p(x) = N_V \cdot e^{-(W_F - W_V(x))/kT} = N_V \cdot e^{-\eta_p(x)/kT}$$

Verwendet man den Abstand von Valenzband und Fermienergie η_p im Volumen des Halbleiters, weit weg von der Verarmungszone, also für $x \rightarrow \infty$

$$\eta_p(x) = \eta_p(x = \infty) + q\varphi(x)$$

so folgt

$$p(x) = p(x = \infty) \cdot e^{-q\varphi(x)/kT} \tag{4.10}$$

mit $\varphi(x)$ aus (4.6). An der Oberfläche ist die Konzentration der Löcher um den Faktor $e^{-q\varphi(x)/kT}$ abgeschwächt im Vergleich zum Volumen des Halbleiters – die sogenannte **Oberflächenverarmung**.

4.2 Debye-Länge

Aus (4.7) folgt für Verarmungstiefe l

$$l = \sqrt{\frac{2\varepsilon_0\varepsilon_r\varphi_s}{qN_A^-}} \tag{4.11}$$

Nimmt man eine Normierung von φ_s in Bezug auf die Temperaturspannung $U_{th} = kT/q$ vor, so lässt sich l berechnen zu

$$l = \sqrt{\frac{\varepsilon_0\varepsilon_r}{qN_A^-} \cdot U_{th}} \cdot \sqrt{2\frac{\varphi_s}{U_{th}}} \equiv L_{Dp} \cdot \sqrt{2\frac{\varphi_s}{U_{th}}} \tag{4.12}$$

mit der Debye-Länge L_{Dp}

$$L_{Dp} = \sqrt{\frac{\varepsilon_0\varepsilon_r}{qN_A^-} \cdot \frac{kT}{q}} = \sqrt{\frac{\varepsilon_0\varepsilon_r}{qN_A^-} \cdot U_{th}} \tag{4.13}$$

4 Halbleiteroberfläche

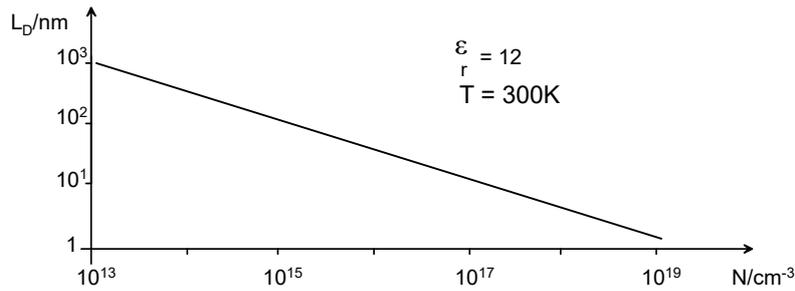


Abbildung 4.3: Debye-Länge in Silizium für verschiedene Dotierstoffkonzentrationen

Aus (4.12) gewinnt man:

$$\frac{L_{Dp}}{l} = \frac{\sqrt{U_{th}/2}}{\sqrt{\varphi_s}}$$

Mit dem Wissen, dass die Spannung φ_s/q über der Länge l der Verarmungszone abfällt, ist L_{Dp} die Länge, über der die Spannung $U_{th}/2$ abfällt. Die Debye-Länge ist somit ein Maß für den Abstand, über den in einem gleichförmig dotierten Halbleiter die Inhomogenität freier Ladungsträger (hier Löcher) abklingt. Die Inhomogenität der freien Ladungsträger wird durch das Oberflächenpotential oder andere äußere Potentiale hervorgerufen. Sie stellt einen wichtigen Parameter zur Charakterisierung von Halbleitermaterial dar. In Abb. 4.3 ist sie für Silizium bei 300 K über der Dotierungskonzentration aufgetragen.

4.3 Oberflächenleitfähigkeit

Im Kapitel 4.1 wurde exemplarisch ein p -Halbleiter mit donatorartigen Oberflächenzuständen behandelt, die aufgrund ihrer positiven Polarität eine **Oberflächenverarmung** zur Folge haben. Setzt man für den gleichen Halbleitertyp aber akzeptorähnliche Zustände an (wie sie real ebenfalls auftreten), so folgt eine **Anreicherung der Oberfläche** mit Löchern.

Oberflächenzustände führen zu einer **Bandverbiegung**. Es gelte allgemein für das Oberflächenpotential:

$$q \cdot \varphi_s = W_{Fi}(x \rightarrow \infty) - W_{Fi}(x = 0)$$

Das entsprechende Bänderdiagramm stellt Abb. 4.4 dar. Es gibt eine weitere Form der

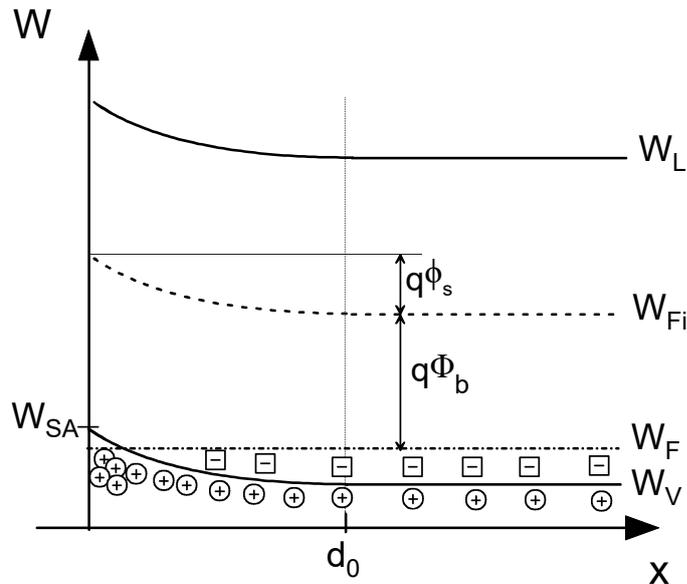


Abbildung 4.4: Anreicherung einer p -Halbleiteroberfläche

Oberflächenladung, die auf einer Anreicherung von Minoritäten statt Majoritäten beruht: die **Inversionsladung**. Sie tritt für p -Halbleiter bei hohen positiven Werten von φ_s auf, weil dann die Bandkanten so weit nach unten gebogen werden (s. Abb. 4.2), dass die Leitungsbandkante dem Fermi-Niveau näher ist als die Valenzbandkante, und somit die Besetzung des Leitungsbandes mit Elektronen an der Oberfläche wahrscheinlicher wird als die Besetzung des Valenzbandes mit Löchern. Das ursprünglich p -dotierte Halbleitermaterial direkt unter der Oberfläche nimmt daher n -leitenden Charakter an!

Dieser Zusammenhang folgt direkt aus dem Massenwirkungsgesetz $n = n_i^2/p$. Wenn in unserem Fall die Löcherkonzentration p durch die Oberflächenladung immer kleiner wird, dann überwiegt ab einem bestimmten Punkt die lokale Elektronenkonzentration n , da die intrinsische Konzentration n_i im dotierten Halbleiter nur ein (temperaturabhängiger) Materialparameter ist, der nicht von den lokalen Ladungsträgerkonzentrationen n und p abhängt.

Der umgekehrte Ladungsträgertyp herrscht vor, wenn ein Oberflächenpotential φ_b angelegt wird, das dem Abstand von W_F und intrinsischem Fermi-niveau W_{F_i} entspricht (**schwache Inversion**)

$$q \cdot \varphi_b = W_{F_i}(x \rightarrow \infty) - W_F \quad (4.14)$$

Tabelle 4.1: Oberflächenleitfähigkeit

	p -Halbleiter	n -Halbleiter
Anreicherung	$\varphi_s < 0$	$\varphi_s > 0$
Verarmung	$0 < \varphi_s < \varphi_b$	$\varphi_b < \varphi_s < 0$
Inversion	$0 < \varphi_b < \varphi_s$	$\varphi_s < \varphi_b < 0$

Sobald die entgegengesetzte Bandkante das Fermi-niveau W_F schneidet, steigt die Inversionsladung sprunghaft an. In unserem Beispiel wäre die **starke Inversion** gegeben für

$$\begin{aligned}
 q \cdot \varphi_i &= W_L(x \rightarrow \infty) - W_F W_g - [W_F - W_V(x \rightarrow \infty)] \\
 &= W_g - W_F + W_V(x \rightarrow \infty) \quad (4.15)
 \end{aligned}$$

Tabelle 4.1 listet alle Kombinationen auf. Es ist unmittelbar ersichtlich, dass die Anreicherung und die Inversion zu einer erhöhten Oberflächenleitfähigkeit führen, und die Verarmung zu einer verringerten Leitfähigkeit der Oberfläche. Ähnliche Betrachtungen führen uns im folgenden Abschnitt auf die Funktionsweise der technisch außerordentlich wichtigen MOS-Struktur.

5 Halbleiter-Festkörper-Grenzflächen

In Kapitel 4 wurden erstmals die Einflüsse einer geometrischen Begrenzung von Halbleiterkörpern, im speziellen die Grenzfläche zwischen Halbleiter und Vakuum, behandelt. Es ist klar, dass für jedes reale Halbleitergebilde Grenzflächen eine entscheidende Rolle spielen. Beispielsweise werden die Eigenschaften von speziellen Halbleitern erst durch metallische Kontakte, d.h. über Metall-Halbleiter-Grenzflächen verfügbar. Die CMOS-Technologie, die heute alles von CPUs über Solid State Disks ("Festkörper-Platten"), Chips für Grafik, KI und Hochfrequenzanwendungen sowie Leistungsschalter ermöglicht, basiert auf Eigenschaften der Halbleiter-Isolator-Grenzflächen.

Bipolare Bauelemente wie *npn*-Transistoren für höchste Frequenzen, Leistungsdioden und photonische Quellen (LEDs und Laser) sowie Detektoren beziehen ihre funktionsbestimmenden Merkmale aus Halbleiter-Halbleiter-Grenzflächen, an denen zunehmend nicht allein der Leitfähigkeitstyp (*p* bzw. *n*), sondern auch die Halbleitersubstanz eine Veränderung erfährt, um Grenzübergänge von Materialien mit verschiedenen Bandlücken auszunutzen (z.B. Heterostruktur-Bipolartransistoren, Heterostruktur-Laser und -Detektoren).

In diesem Abschnitt wird zunächst das Metal-Insulator-Semiconductor (MIS)-System bearbeitet. Daran anschliessend wird der Metall-Halbleiterkontakt sowohl als sperrender (Schottky-) als auch als sperrfreier, ohmscher Kontakt behandelt. Der pn-Übergang als Kontakt zwischen gleichen Halbleitersubstanzen unterschiedlichen Leitfähigkeitstyps vermittelt das grundlegende Verständnis für alle bipolaren Bauelemente. Schließlich werden kurz die wichtigsten Eigenschaften von Heterostruktur-Grenzflächen besprochen. Damit sind die Grundlagen gelegt für die Behandlung sämtlicher Halbleiterbauelemente, die heute von technischer Bedeutung sind, und deren Funktion in weiterführenden Vorlesungen im Detail behandelt werden.

5.1 Das Metall-Isolator-Halbleiter-System, MIS- und MOS-Kondensatoren

Dieses System ist das in der Mikroelektronik mit Abstand am weitesten verbreitete System. In Form von CPU- und Speicherchips mit höchster Integrationsdichte (Stand 2024: knapp 100 Mrd. Transistoren auf einem Chip) hat sie die moderne Informationsgesellschaft mit leistungsfähigen Smartphones, Workstations und Cloud-Infrastruktur erst möglich gemacht, nicht zu vergessen die verbindende Netzwerktechnik mit Routern und Switches, die höchste Bandbreiten mit hoher Effizienz verarbeiten können.

Das Metall-Isolator-Halbleiter (MIS)-System enthält bereits zwei Festkörper-Grenzflächen, von denen die Isolator-Halbleitergrenzfläche die wichtigste ist. Das Metall dient lediglich der isolatorseitigen Kontaktierung der Anordnung. Für den Zweck des chemischen und mechanischen Schutzes der Halbleiteroberfläche (Passivierung) zwischen kontaktierenden Flächen wird nur das Oxid eingesetzt.

Die technisch wichtigste Kombination von Halbleiter und Isolator ist die des Silizium/SiO₂ mit einer SiO₂-Schichtdicke im Bereich von nur wenigen Nanometern. Es wird hier das natürliche Oxid des Halbleitermaterials (Metall-Oxid-Semiconductor, MOS) als Isolator verwendet, das mit sehr niedriger Defektdichte hergestellt werden kann. In den neuesten CMOS Technologien werden auch andere Isolatoren eingesetzt, deren Materialeigenschaften sich günstig auf das Bauelement-Verhalten auswirken (z.B. high-*k*-Materialien mit höherem ϵ_r als SiO₂, um höhere Drainströme zu erreichen, dazu später mehr).

5.1.1 Bändermodell und Ladungsverteilung

Grundsätzlich liegen an der Grenzfläche zwischen Isolator und Halbleiter ähnliche Verhältnisse vor wie an der unbedeckten Halbleiteroberfläche aus Abschnitt 4: aus Oberflächenzuständen werden hier **Grenzflächenzustände**. Allerdings können jetzt noch zusätzlich Zustände und Ladungen im Isolator selbst vorliegen. Ihre Wirkung überlagert sich additiv mit den Grenzflächenzuständen bzw. -ladungen, so dass sie keine zusätzliche Problematik aufwerfen.

5.1.2 Ideale MIS-Struktur

Zuerst wird zur Vereinfachung die ideale MIS-Struktur ohne Grenzflächen- und Isolatorladungen behandelt. Sie ist folgendermaßen definiert (vgl. Abb. 5.1:

1. Die Austrittsarbeiten für Metall und Halbleiter sind identisch,

$$q \cdot \Phi_M = q \cdot \Phi_{HL} \quad \text{bzw.} \quad q \cdot \Phi_{MHL} = q \cdot (\Phi_M - \Phi_{HL}) = 0 \quad (5.1)$$

Die Austrittsarbeiten entsprechen der Höhe V_0 der Potentialstufe aus Kap. 1.5.1, also der Energie, die man dem Elektron mitgeben muss, damit es den Festkörper (hier Metall bzw. Halbleiter) durch die Oberfläche verlassen kann.

2. Grenzflächen- sowie Isolatorladungen liegen nicht vor
3. Der Isolator ist ideal und lässt somit keinen Stromfluss zu. Legt man keine Spannung zwischen Metall und Halbleiter an, so liegt thermodynamisches Gleichgewicht vor und die Fermi-Energie W_F verläuft in der gesamten Anordnung waagrecht: sie dient als Referenzlinie für die weiteren Betrachtungen, die man anhand der Abb. 5.1 durchführen kann. Dabei ist im Teil b) ein n -Typ und im Teil c) ein p -Typ Halbleiter zugrunde gelegt.

Die im Zusammenhang verwendeten Größen sind

$q\Phi_M$	Austrittsarbeit für das Metall (vereinfachend werden unterschiedliche Metalle ($q \cdot \Phi_{M,1}$ und $q \cdot \Phi_{M,2}$) angenommen, sodass W_0 im Halbleiter und Metall gleich ist.)
$q\Phi_B$	Austrittsarbeit für Elektronen, die aus dem Metall ins Leitungsband des Isolators gelangen
W_0	Vakuumniveau - Energie der Elektronen, die den Festkörper verlassen haben
$W_{L,IS}$	Leitungsbandkante des Isolators
$W_{g,IS}$	Bandlücke des Isolators
$q\chi_{IS}$	Elektronenaffinität des Isolators

$q\chi$	Elektronenaffinität des Halbleiters
W_L, W_V	Leitungs- und Valenzbandkante des Halbleiters
W_F, W_{Fi}	Fermi-Niveau des dotierten bzw. intrinsischen Halbleiters
$q\psi_b$	Volumenpotential des Halbleiters
η_n, η_p	Abstand von W_F zu den jeweiligen Bandkanten

Für diese ideale MIS-Struktur verlaufen die Bandkanten des Halbleiters und des Isolators horizontal, d.h. W_L, W_V und $W_{L,IS} = \text{const.}$ Da die Austrittsarbeiten aus thermodynamischen Gründen auf W_F bezogen werden, sind sie für n - und p -Typ Halbleiter unterschiedlich. Daher müssen in beiden Fällen unterschiedliche Metalle zur Realisierung der idealen Struktur verwendet werden. Aus Abb. 5.1 gewinnt man mit $W_g/2 = W_L - W_{Fi}$

$$q\Phi_M = q\Phi_{HL} \begin{cases} = q\chi + W_g/2 - q\Phi_b = q\chi + \eta_n & n\text{-Typ} \\ = q\chi + W_g/2 + q\Phi_b = q\chi + W_g - \eta_p & p\text{-Typ} \end{cases} \quad (5.2)$$

Jetzt werden Spannungen an die Struktur angelegt. Das dadurch erzeugte Feld muss seinen Ursprung in Ladungen haben (Poisson-Gleichung), die im Falle der idealen MIS-Struktur (MIS-Diode) wegen der Voraussetzung (2) nur auf der Metallplatte und als Raumladung im Halbleiter vorliegen können. Unter dem Einfluss eines elektrischen Felds müssen die Bandkanten des Halbleiters geneigt sein. In den Abbildungen 5.2 und 5.3 sind Bänderdiagramm und Ladungsverteilung für einen n -Halbleiter mit $U > 0$ bzw. $U < 0$ dargestellt. In Abb. 5.1 führt die angelegte Spannung $U > 0$ an der Metallelektrode zu einer positiven Aufladung und einer Absenkung von W_F gegenüber dem Halbleiter. Ladungsneutralität und Poisson-Gleichung fordern eine negative Ladung im n -Typ Halbleiter, die durch eine Anreicherung von Elektronen an der Grenzfläche über den durch die Dotierung vorgegebenen Wert hinaus bereitgestellt wird, sowie die damit einhergehende nach unten gerichtete Bandverbiegung im Halbleiter über die Länge l_a . Wegen der extrem hohen Ladungsträgerdichte (Elektronen) im Metall ist die Ausdehnung (Dicke) der Raumladungszone hier verschwindend gering und die Raumladung entartet zu einer Flächenladung.

Für $U < 0$ (Abb. 5.3a) wird die Metallplatte negativ aufgeladen, W_F gegenüber dem Halbleiter angehoben und somit die Bänder im Halbleiter nach oben verbogen. Die Fol-

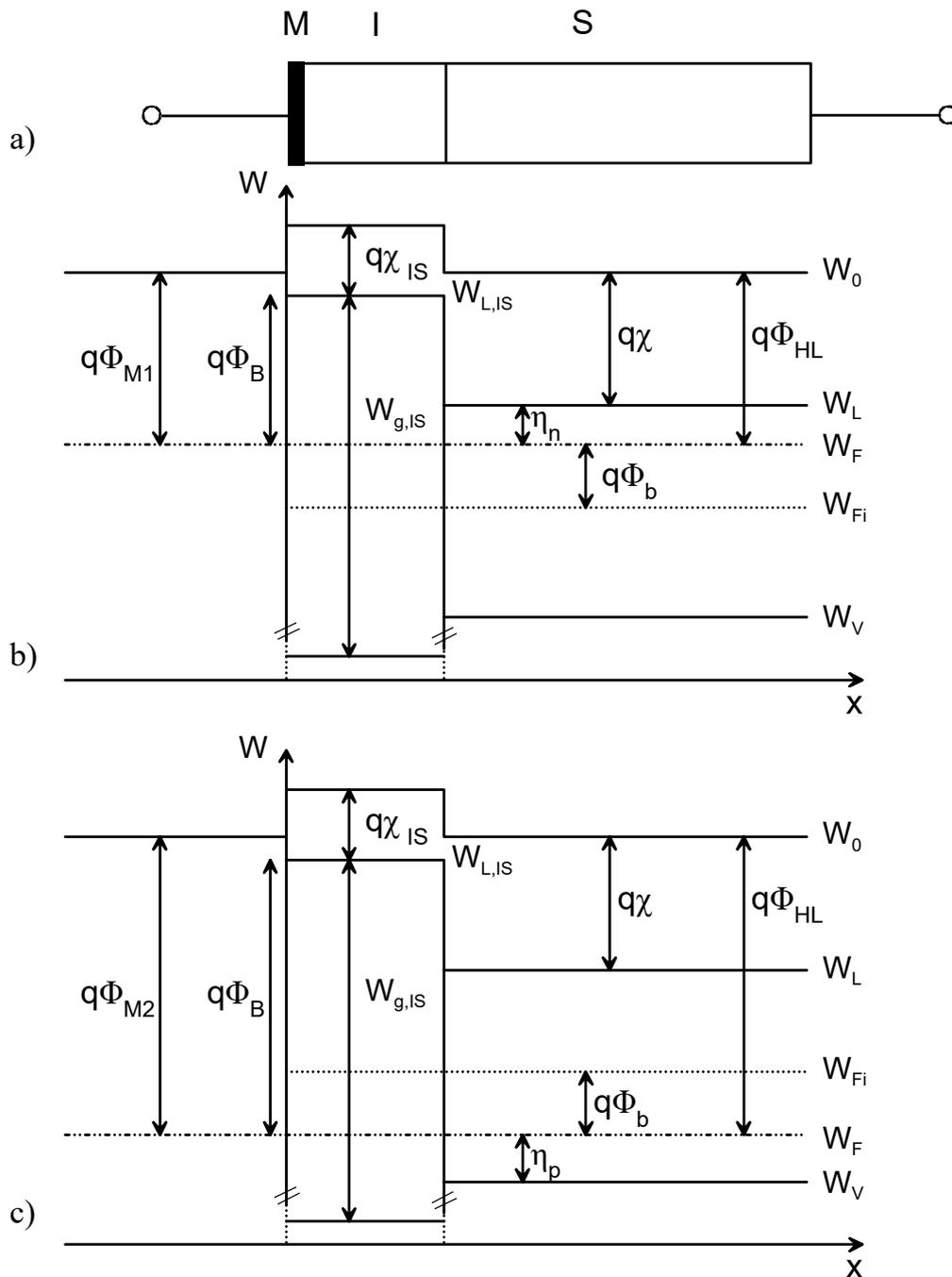


Abbildung 5.1: Aufbau a) und Bändermodell der idealen MIS-Struktur bei äußerer Spannung $U = 0$ für b) n -Typ und c) p -Typ Halbleiter

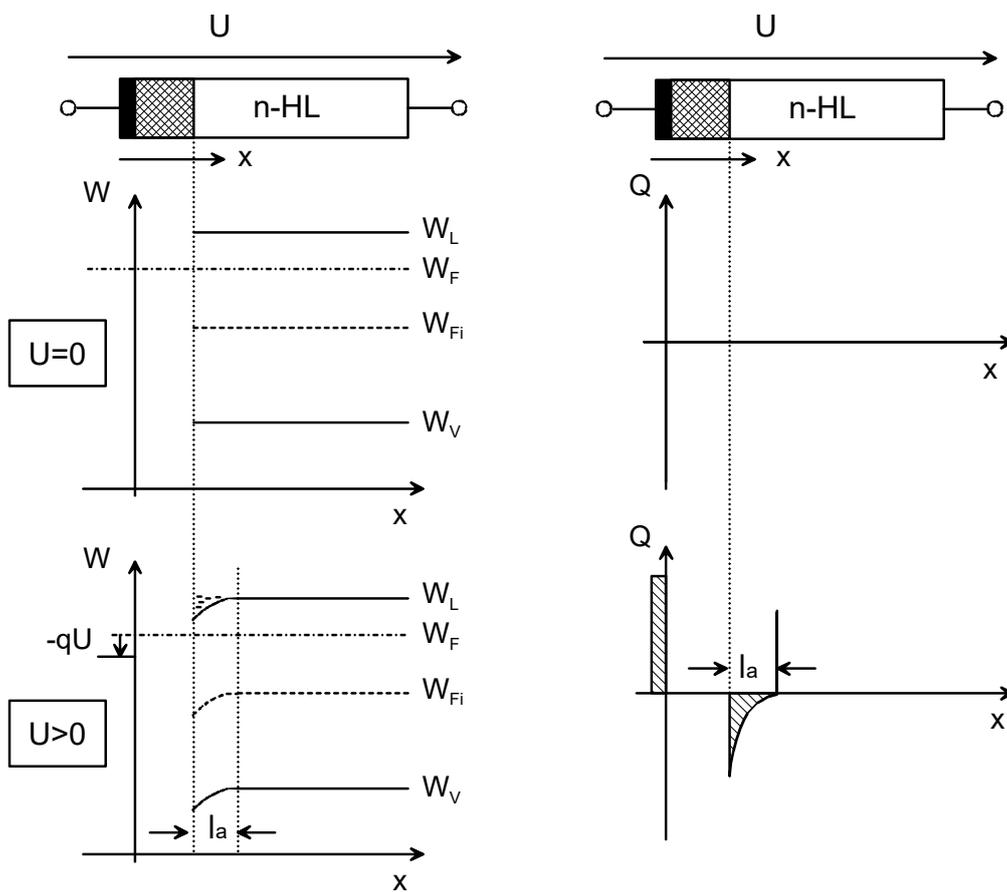


Abbildung 5.2: Aufbau und Bändermodell einer idealen MIS-Struktur a) $U = 0$ (Flachband), b) $U > 0$ (Anreicherung)

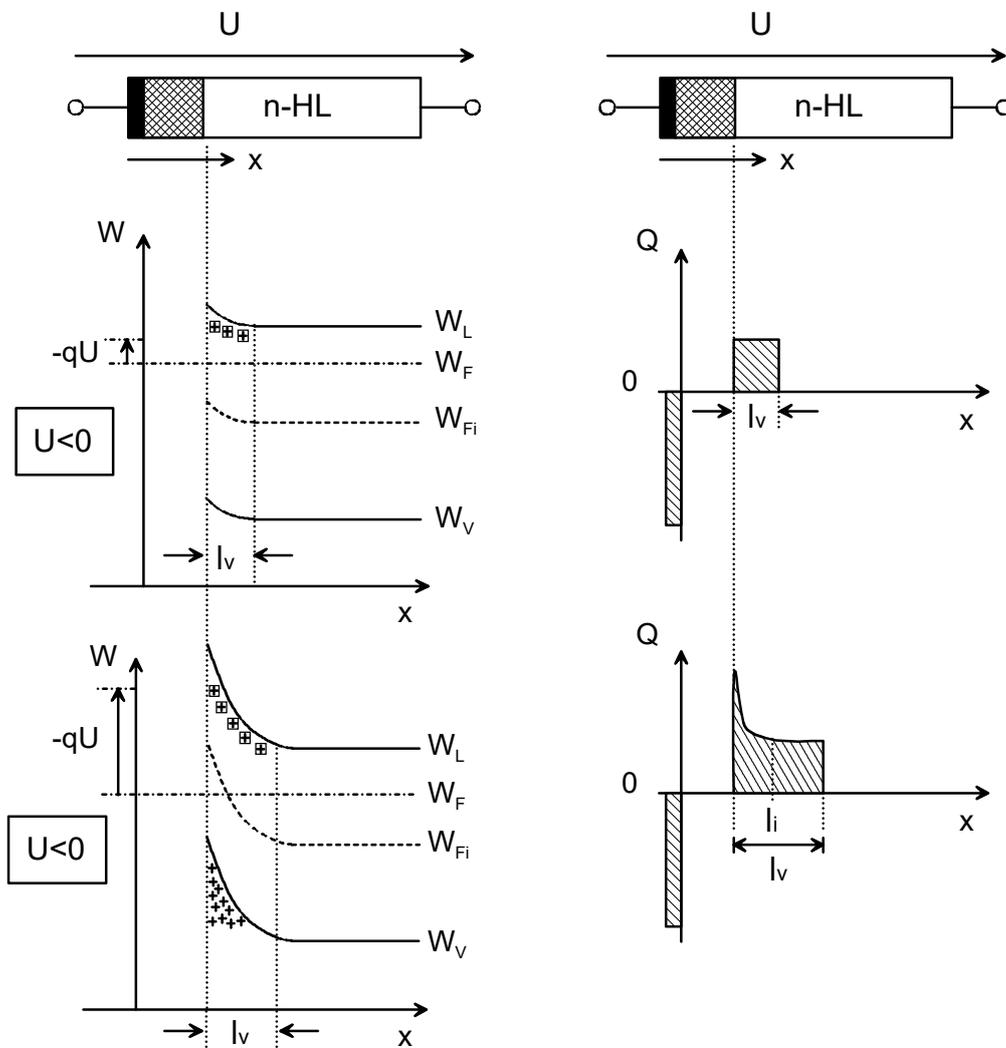


Abbildung 5.3: Aufbau und Bändermodell einer idealen MIS-Struktur a) $U < 0$ (Verarmung), b) $U < 0$ (Inversion)

ge ist eine Verarmung an Elektronen an der Grenzfläche Isolator-Halbleiter, die die notwendige positive Gegenladung aus ionisierten positiv geladenen Donatoren bereitstellt. Die Ausdehnung der verarmten Halbleiter-Raumladungszone ist l_v .

In Abb. 5.3b wurde die negative Spannung weiter gesteigert, bis an der Grenzfläche $W_{Fi} > W_F$ eintritt. Das Resultat ist die bereits in Kapitel 4 erwähnte Inversion: die weiter erhöhte negative Ladung auf der Metallelektrode wird nicht mehr ausschließlich durch positiv geladene ionisierte Donatoren bereitgestellt, sondern auch durch positiv geladene Löcher unmittelbar an der Grenzfläche. Dies wird möglich, weil die Ladungsträgerkonzentrationen vom Abstand der jeweiligen Bandkante zum Fermi-Niveau bestimmt werden.

Für den Fall $W_{Fi} > W_F$ in **Inversion** folgt trotz n -Dotierung

$$p(x=0) = N_V \cdot e^{-(W_F - W_V)/kT} > N_L \cdot e^{-(W_L - W_F)/kT} = n(x=0) \quad (5.3)$$

Die durch diese Inversion erzeugte Grenzflächenleitfähigkeit, die offensichtlich durch die angelegte Spannung gesteuert werden kann, wird in Feldeffekttransistoren ausgenutzt: Durch Anlegen eines Feldes über den Gate-Kontakt können Ströme nahezu leistungslos, eben durch einen reinen Feldeffekt, gesteuert werden.

Eine analoge Behandlung kann für die MIS-Struktur mit einem p -Halbleiter erfolgen, soll hier aber nicht durchgeführt werden.

In allen Fällen gilt für die ideale MIS-Struktur:

$$W_{F,HL} - W_{F,M} = -q \cdot U \quad (5.4)$$

Das heißt aber nicht, dass die Bandverbiegung im Halbleiter identisch ist mit $-qU$. Vielmehr fällt ein Teil der Spannung und damit der möglichen Bandverbiegung über dem Isolator ab, im Falle der Anreicherung und der Inversion sogar der deutlich größere. Dies ist zu bedenken, wenn (5.1) für $U \neq 0$ umgeschrieben wird, und hat ebenfalls Einfluss auf die jetzt zu behandelnden Abweichungen von der idealen MIS-Struktur.

5.1.3 Nicht-ideale MIS-Struktur

Die nicht-ideale MIS-Struktur unterscheidet sich von der idealen durch

- unterschiedliche Austrittsarbeiten von Metall und Halbleiter

$\Phi_M < \Phi_{HL}$	Anreicherung
$\Phi_M > \Phi_{HL}$ und $\Phi_M - \Phi_{HL} < \psi_b$	Verarmung
$\Phi_M > \Phi_{HL}$ und $\Phi_M - \Phi_{HL} > \psi_b$	Inversion

Tabelle 5.2: Vorgang an der MIS-Grenzfläche für unterschiedliche Differenzen der Austrittsarbeiten im n -Halbleiter

- Grenzflächenladungen
- Ladungen im Isolator

Je nach technologischer Realisierung spielen diese zusätzlichen Beiträge eine unterschiedlich große Rolle. Die Abbildung 5.4 stellt die Konsequenzen für das Bänderdiagramm und die Ladungsverteilung dar, und zwar für $U = 0$.

In allen Fällen tritt im Gegensatz zur idealen MIS-Diode bereits bei $U = 0$ eine Bandverbiegung im Halbleiter auf, die in der Folge Influenzladungen im Halbleiter oder auch auf der Metallelektrode verursacht. Für den Fall in Abb. 5.4a gilt für die Austrittsarbeitendifferenz:

$$q\Phi_{MHL} = q(\Phi_M - \Phi_{HL}) = q\Phi_M - (q\chi + \eta_n) = q\Phi_M - (q\chi + W_g/2 - q\Psi_b) \quad (5.5)$$

In der Abbildung wurde angenommen, dass $\Phi_M > \Phi_{HL}$ ist und eine Verarmung der Halbleiteroberfläche resultiert. Tabelle 5.2 fasst alle Möglichkeiten für einen n -Halbleiter zusammen. Die Fälle Anreicherung, Verarmung und Inversion können gleichfalls durch Grenzflächen- und Oxidladungen je nach Vorzeichen und Konzentration realisiert werden.

Die charakteristische Größe, die in Summe die Einflüsse der unterschiedlichen Austrittsarbeiten und Grenz- sowie Isolatorladungen beschreibt, ist die **Flachbandspannung** U_{FB} . Dies ist die Spannung, die man an die reale MIS-Struktur anlegen muss, um das Bänderdiagramm der idealen MIS-Struktur für $U = 0$, eben den flachen Bandverlauf, zu erzeugen.

Abb. 5.5 zeigt für die unterschiedlichen Ursachen der Bandverbiegung, die in Abb. 6.4) die Ladungsverteilung in der MIS-Struktur, sowohl für $U = 0$ als auch für $U = U_{FB}$. Da für $U = U_{FB}$ im Halbleiter keine Bandverbiegung, somit kein Potentialabfall und gemäss der Poisson-Gleichung keine Ladung auftritt, beschränkt sich die Ladungsver-

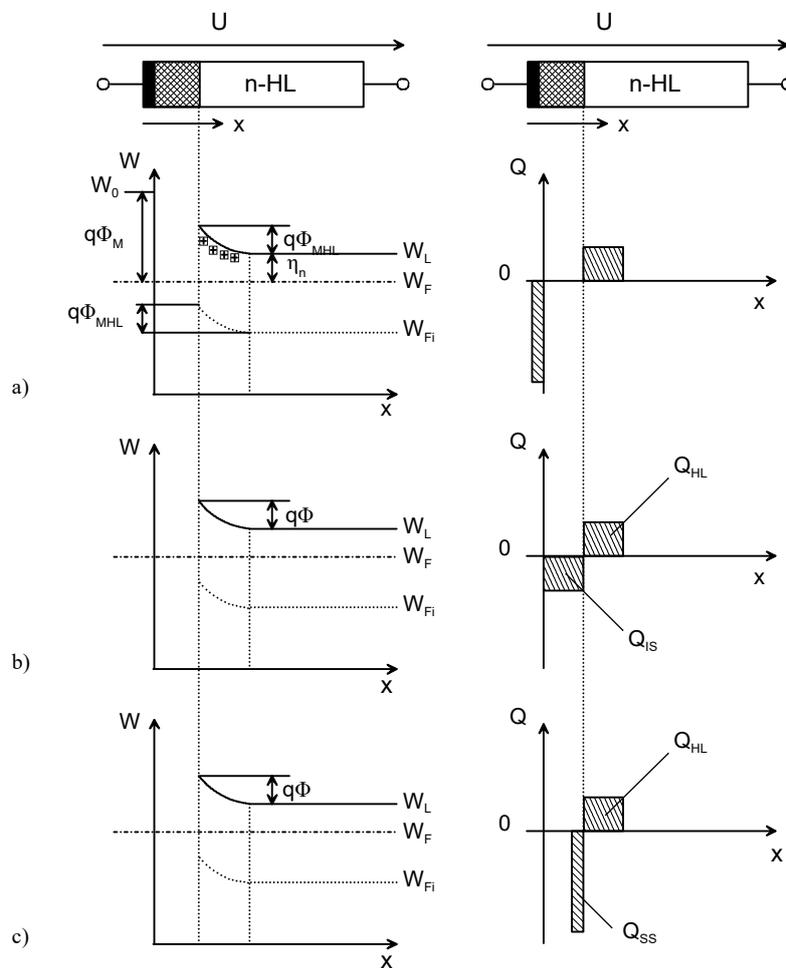


Abbildung 5.4: MIS-Struktur, Bändermodell und Ladungsverteilung für einen n -Halbleiter mit a) unterschiedliche Austrittsarbeiten Φ_M und Φ_{HL} , mit b) negative Oxidladungen oder mit c) negativen Grenzflächenladungen

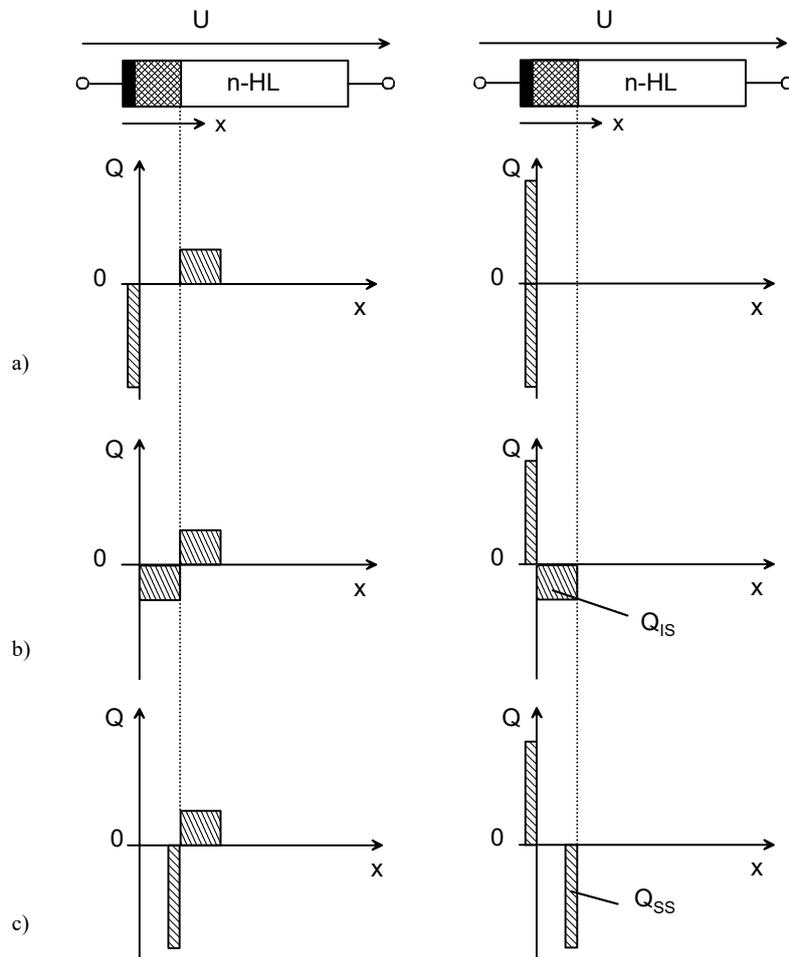


Abbildung 5.5: Ladungsverteilung bei den äußeren Spannungen $U = 0$ (links) und $U = U_{FB}$ für die Fälle a) unterschiedliche Austrittsarbeiten Φ_M und Φ_{HL} , b) örtlich feste negative Oxidladungen oder c) örtlich feste negative Grenzflächenladungen

teilung auf die Metallplatte und den Isolator bzw. dessen Grenzfläche. Für die unterschiedlichen Fälle ergibt sich für die Flachbandspannung

$$U_{\text{FB}} = \Phi_{\text{MHL}} = \Phi_{\text{M}} - \Phi_{\text{HL}} \quad (5.6)$$

wenn der Unterschied zur idealen MIS-Struktur allein aufgrund der Differenzen der Austrittsarbeiten zwischen Metall und Halbleiter besteht, und

$$U_{\text{FB}} = -Q_{\text{IS}}/c_{\text{IS}} \quad (5.7)$$

mit der Isolatorkapazität pro Fläche $c_{\text{IS}} = (\varepsilon_0 \varepsilon_{r,\text{IS}})/d$ und der festen Oxidladungsdichte pro Fläche Q_{IS} , die z.B. durch den Einschluss von geladenen Ionen im Isolator entsteht, bzw.

$$U_{\text{FB}} = -Q_{\text{ss}}/c_{\text{IS}} \quad (5.8)$$

mit der (festen) Grenzflächen-Ladungsdichte pro Flächeneinheit Q_{ss} , die z.B. aufgrund besetzter akzeptorartiger Grenzflächenzustände besteht.

Treten alle Fälle gleichzeitig auf, so gilt

$$U_{\text{FB}} = \Phi_{\text{M}} - \Phi_{\text{HL}} - Q_{\text{IS}}/c_{\text{IS}} - Q_{\text{ss}}/c_{\text{IS}} \quad (5.9)$$

Reale MIS-Strukturen verhalten sich bei angelegter Spannung wie ideale MIS-Strukturen, wenn man eine Verschiebung der Bezugsspannung von $U = 0$ zu $U = U_{\text{FB}}$ vornimmt und die zusätzlichen Ladungen gemäß Abb. 5.5 berücksichtigt. Bei der folgenden Behandlung der Kapazitäts-Spannungskennlinien von MIS-Strukturen wird dies deutlich werden.

5.1.4 Kapazitäts-Spannungs-Kennlinie der MIS-Diode

Die MIS-Struktur mit n -Halbleiter stellt eine typische Kondensatoranordnung dar, bei der zwei elektrisch leitfähige Bereiche, Metallplatte und Halbleiter, durch ein Dielektrikum, Isolator und Raumladungszone, getrennt sind. Abb. 5.6 zeigt den Querschnitt und die beiden in Serie geschalteten Kapazitätsanteile c_{IS} (Isolator-Kapazität) und c_{R} (Raumladungskapazität im Halbleiter). Die Isolatorkapazität c_{IS} entspricht der Kapazität eines Plattenkondensators und ist in (5.9) gegeben. Die Raumladungskapazität c_{R}

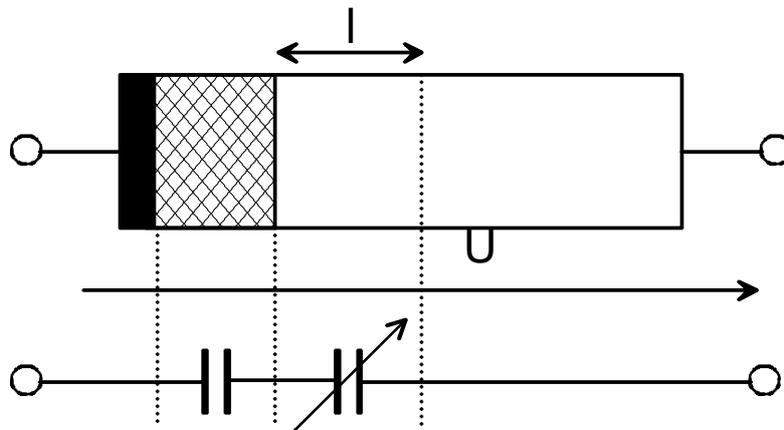


Abbildung 5.6: Anteile der Isolator- (c_{IS}) und der Raumladungskapazität (c_R) an der gesamten MIS-Kapazität

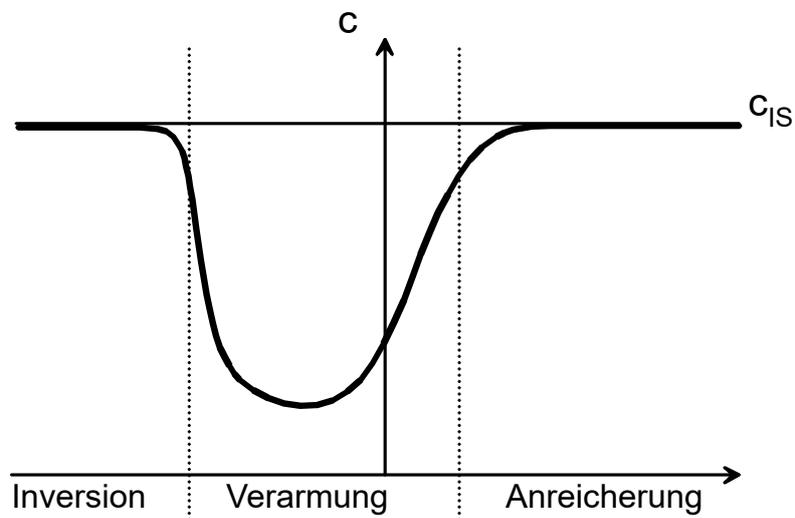


Abbildung 5.7: Spannungsabhängigkeit der MIS-Kapazität

ist im Gegensatz zu c_{IS} spannungsabhängig, da sich die Dicke l der Raumladungszone, die für c_R als Dielektrikum wirkt, mit der angelegten Spannung ändert. Setzt man zunächst den Fall der Verarmung an, so gilt für die Raumladung Q_R pro Fläche in der Raumladungszone

$$Q_R = q \cdot N_D^+ \cdot l \quad (5.10)$$

analog zu (4.2), wenn eine konstante Dotierung des n -Halbleiters vorausgesetzt wird. Allgemein gilt für die Kapazität (pro Flächeneinheit)

$$c_R = \frac{dQ_R}{dU_R} \quad (5.11)$$

und mit (5.10)

$$c_R = qN_D^+ \frac{dl}{dU_R} \quad (5.12)$$

Der Spannungsabfall U_R über der RLZ entspricht der Potentialdifferenz zwischen der Halbleiteroberfläche $\Phi(x=0) = \Phi_S$ und dem Ende der RLZ (bei $x=l$)

$$U_R = \Phi(x=0) - \Phi(x=l) = \Phi_S$$

In Kapitel 4 wurde die Länge l der RLZ in Abhängigkeit des Oberflächenpotentials berechnet, wobei $\Phi(x=0) = 0$ gesetzt wurde:

$$l(\Phi_S) = \sqrt{\frac{2\varepsilon_0\varepsilon_{r,HL}}{qN_D^+} \Phi_S} \quad (5.13)$$

Damit lässt sich berechnen:

$$c_R = q \cdot N_D^+ \cdot \frac{dl}{d\Phi_S} = q \cdot N_D^+ \cdot \sqrt{\frac{2\varepsilon_0\varepsilon_{r,HL}}{qN_D^+}} \cdot \frac{d\sqrt{\Phi_S}}{d\Phi_S} \quad (5.14)$$

$$c_R = \sqrt{\frac{\varepsilon_0\varepsilon_{r,HL} \cdot qN_D^+}{2 \cdot \Phi_S}} \quad (5.15)$$

Das Oberflächenpotential Φ_S entspricht der Flachbandspannung U_{FB} , die über die am Metall angelegte Spannung U abgebaut wird:

$$\Phi_S = U_{FB} - U \quad \text{mit } U < U_{FB} \quad (5.16)$$

Damit gilt:

$$c_R = \sqrt{\frac{\varepsilon_0 \varepsilon_{r,HL} q N_D^+}{2(U_{FB} - U)}} \quad \text{für } U < U_{FB} \quad (5.17)$$

Unter Verwendung der Debye-Länge L_{Dn} (4.13) kann man (5.17) auch schreiben

$$c_R = (L_{Dn} q N_D^+) / \sqrt{2U_{th}(U_{FB} - U)} \quad \text{für } U < U_{FB} \quad (5.18)$$

Die Gesamtkapazität der MIS-Struktur ist die Serienschaltung von c_R und c_{IS} wie in Abb. 5.6 schematisch gezeigt

$$c_{MIS}(U) = \frac{c_{IS} \cdot c_R(U)}{c_{IS} + c_R(U)} \quad \text{für } U < U_{FB} \quad (5.19)$$

Der Verlauf von c_{MIS} und ist in Abb. 6.7 dargestellt. Man erkennt, dass c_{MIS} nur im Bereich der Verarmung der Halbleitergrenzfläche spannungsabhängig ist: schrumpft die Verarmungszone zusammen, so ist c_R im Vergleich zu c_{IS} sehr groß, der Halbleiter ist bis zur Grenzfläche zum Isolator leitfähig, und als Dielektrikum bleibt nur der Isolator übrig, dessen Dicke spannungsunabhängig ist. Bei weiterer Erhöhung der angelegten Spannung werden Elektronen an der Grenzfläche angereichert. Sie erhöhen zwar die Leitfähigkeit an der Grenzfläche, ändern aber nichts an der Dicke des Dielektrikums. Erniedrigt man, ausgehend von der Verarmung, die angelegte Spannung, so werden mit Auftreten der Inversion wieder Ladungsträger (jetzt Minoritäten, hier Löcher) an der Grenzfläche erzeugt. Die halbleiterseitige "Kondensatorplatte" verschiebt sich somit aus dem Halbleiterinneren wieder an die Grenzfläche, so dass erneut nur die Isolatorkapazität wirksam ist. Die Kapazität c_{MIS} ist nicht nur von der angelegten Spannung, sondern auch noch von der Betriebsfrequenz abhängig. Zunächst können die Ladungen Q_{ss} spannungsabhängig sein, so dass bei Wechselspannung die entsprechenden Zustände umgeladen werden können. Für die Umladung ist im Mittel eine Umladezeit τ_s nötig. Für Frequenzen $f > 1/\tau_s$ werden daher nur ein Teil der Zustände umgela-

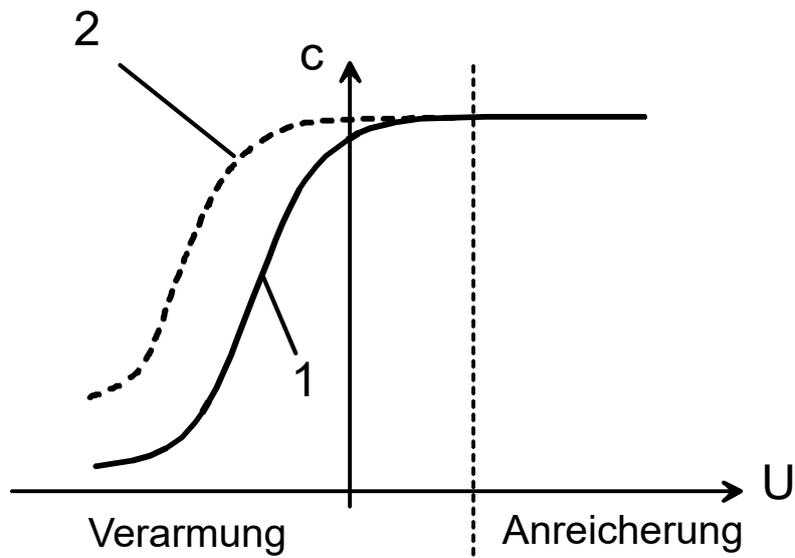


Abbildung 5.8: Einfluss der Messfrequenz auf die Kapazitäts-Spannungsabhängigkeit (schematisch) für die Bereiche Verarmung und Anreicherung (1) ohne Umladungen (hohe Frequenzen) und (2) mit Umladungen (tiefe Frequenzen)

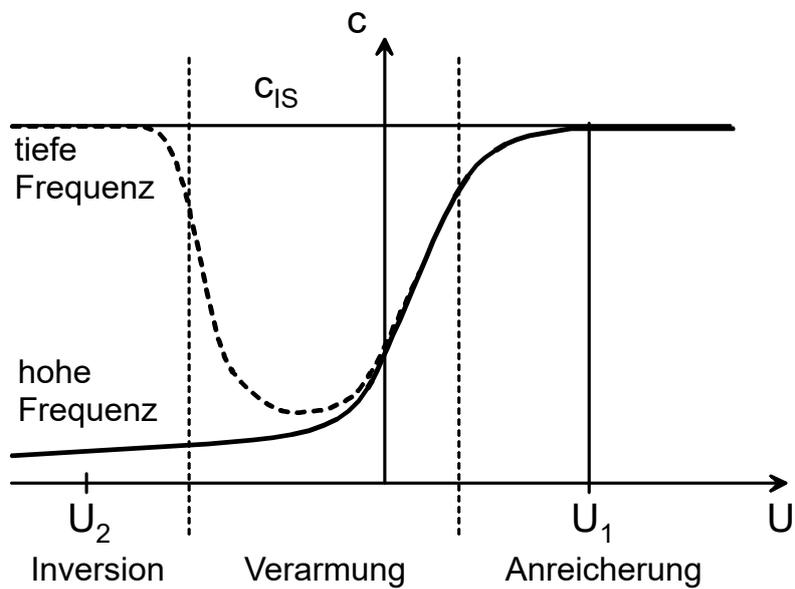


Abbildung 5.9: Spannungsabhängigkeit der Gesamtkapazität einer MIS-Struktur und Einfluss der Messfrequenz im Inversionsbereich

den und gemäss $c = dQ/dU$ ist c_{MIS} für hohe Frequenzen kleiner als für niedrige (vgl. Abb. 5.8).

Besonders ausgeprägt ist die Frequenzabhängigkeit im Bereich der Inversion (vgl. Abb. 5.9): Für hohe Frequenzen bleibt die Gesamtkapazität niedrig. Um das zu verstehen, muss man überlegen, wie die Inversion erzeugt wird: Es handelt sich um die Bereitstellung von Minoritäten an der Grenzfläche, also in einem n -Halbleiter um Löcher. Diese können nur mittels thermischer Generation von Elektron-Loch-Paaren durch Aufbringen der Bandlückenenergie W_g erzeugt werden, und dies ist für 300 K ein extrem unwahrscheinlicher, d.h. langsamer Prozess. Je nach Bandabstand und Reinheit des Materials liegen die Zeitkonstanten für diesen Prozess bei $\tau_{\text{inv}} = 1 \times 10^{-6}$ s bis 1 s. Für $f \gg 1/\tau_{\text{inv}}$ gilt gemäß $dQ_{\text{inv}}/dU \approx 0$ also wieder $c \approx c_{\text{min}}$.

Im Gegensatz dazu handelt es sich bei der Verarmung am Rande der Raumladungszone lediglich um eine Hin- und Herbewegung von Majoritäten, die reichlich vorhanden sind und nicht erst erzeugt werden müssen.

5.2 Metall-Halbleiter-Grenzflächen; Schottky-Dioden und Ohm-Kontakte

Erst über metallische Kontakte, d.h. durch Metall-Halbleiter-Grenzflächen, werden die Vorgänge im Innern von Halbleitern nach außen verfügbar. Bauelemente und erst recht Schaltungen sind also ohne Metall-Halbleiter-Kontakte nicht denkbar, so dass der Grenzfläche zwischen diesen beiden Festkörpern eine für die Technik zentrale Bedeutung zukommt. Grundsätzlich unterscheidet man zwischen sperrfreien (Ohm-) und sperrenden (Schottky-) Kontakten. Der erste hat eine lineare Strom-Spannungskennlinie und dient der reinen, niederohmigen Kontaktierung, der zweite besitzt eine stark nichtlineare Diodencharakteristik und kann bereits Träger einer Bauelementfunktion (Diode, Transistor) sein.

5.2.1 Bändermodell des Metall-Halbleiter-Kontaktes

Das Bändermodell des Metall-Halbleiter-Übergangs liefert den Zugang zum elektrischen Verhalten der Anordnung. Es bestimmt die Ladungsträgerkonzentration, den Feld- und den Potentialverlauf im Halbleiter an der Grenzfläche zum Metall. In Abb. 5.12 sind die Bandverläufe idealer Metall- und n -Halbleiteroberflächen getrennt gezeigt, wobei

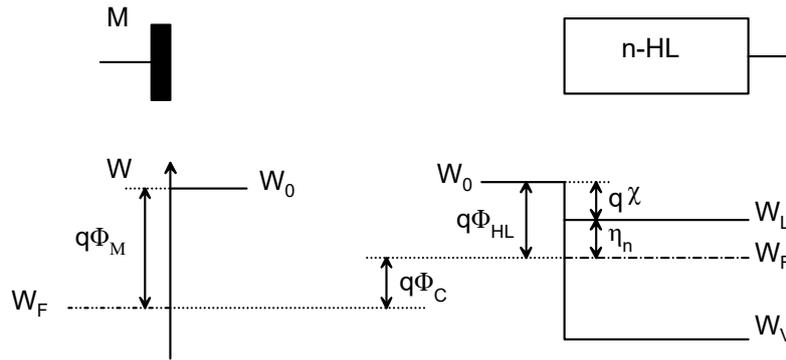


Abbildung 5.10: Zum Entstehen eines Metall-Halbleiter-Kontaktes: getrennte Systeme

als Bezugspunkt die Energie W_0 gewählt wird, die ein Elektron nach Aufbringen der Austrittsarbeit mit $W_{kin} = 0$ besitzt. Da beide Festkörper noch getrennte Systeme darstellen, liegen die Fermi-Energien auf unterschiedlichen Niveaus, weil sie i.A. unterschiedliche Austrittsarbeiten aufweisen.

Werden beide idealen Festkörper zusammengefügt (Abb. 5.11) unter der Voraussetzung, dass keine Grenzflächenladungen auftreten (ideales Verhalten), so bilden sie ein gemeinsames System, das sich ohne von außen angelegte Spannung (oder andere äußere Einflüsse) im thermodynamischen Gleichgewicht befindet. Daher hat die Fermi-Energie W_F im ganzen System den gleichen Wert! Im gewählten Fall $\Phi_M > \Phi_{HL}$ ist es für Elektronen einfacher, aus dem Halbleiter ins Metall zu wechseln als umgekehrt aus dem Metall in den Halbleiter. Es wird sich daher ein Gleichgewicht derart einstellen, dass an der Grenzfläche im Halbleiter eine an Elektronen verarmte und durch die zurückbleibenden Donatorrümpfe positiv geladene Zone ausbildet, deren Ladung durch eine entsprechende Elektronenladung metallseitig an der Grenzfläche kompensiert wird. Die Ausdehnung dieser zusätzlichen Elektronenanreicherung ist im Metall wegen der um Größenordnungen höheren Elektronenkonzentration extrem klein. Elektronen, die vom Metall in den Halbleiter übergehen, müssen die Metall-Halbleiter-Barriere $q \cdot \Phi_{Bn}$ überwinden, unter Verwendung des Volumenniveaus als Referenzpotential.

$$q\Phi_{Bn} = q \cdot (\Phi_M - \chi) \tag{5.20}$$

und für den Übergang Halbleiter \rightarrow Metall

$$qU_D = q\Phi_{Bn} - \eta_n \tag{5.21}$$

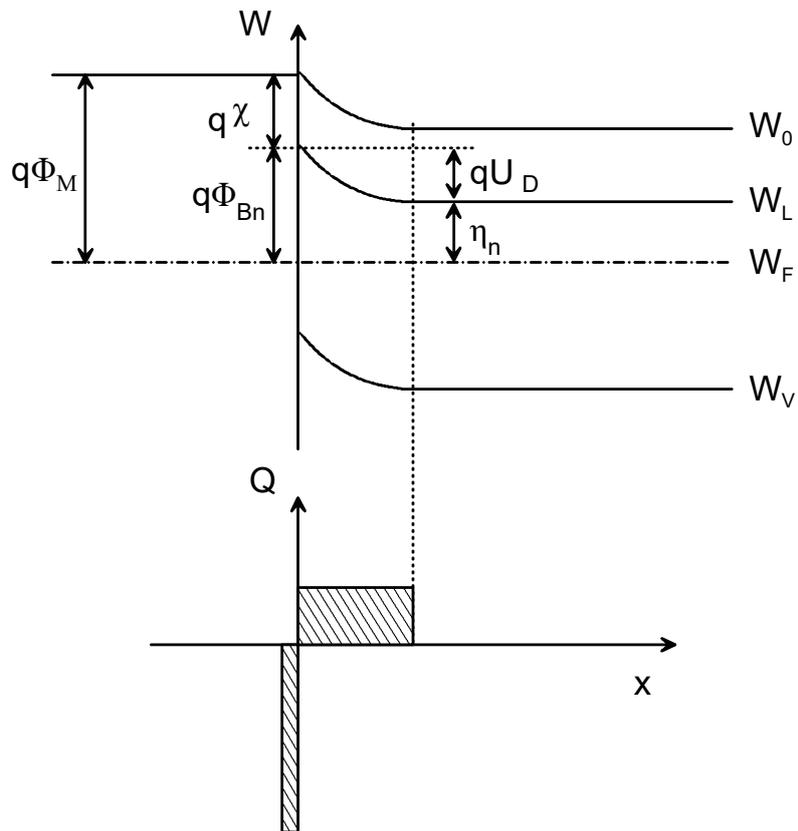


Abbildung 5.11: Zum Entstehen des Metall-Halbleiter-Kontaktes: vereinte Systeme ($U = 0$)

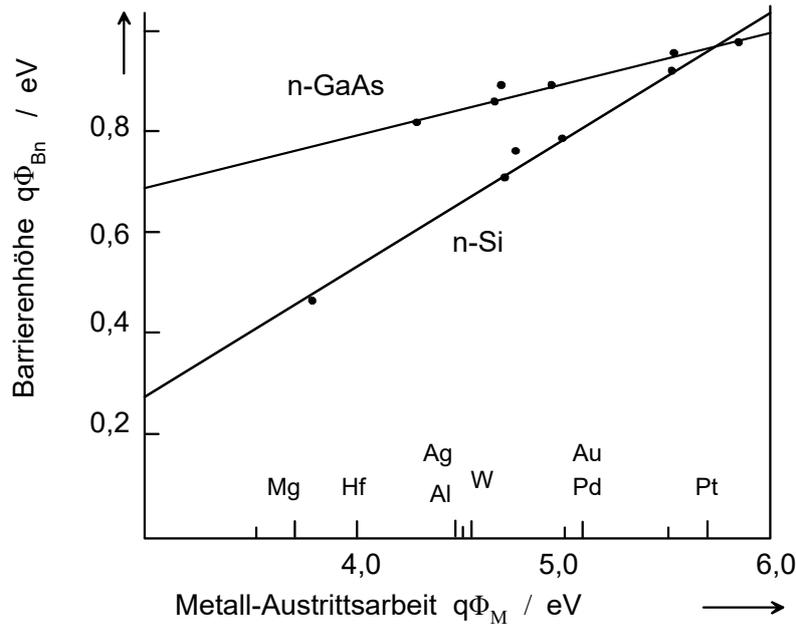


Abbildung 5.12: Gemessene Barrierrhöhe auf n-Si bzw. n-GaAs für unterschiedliche Metalle

übrig (siehe Abb. 5.11). Φ_{Bn} wird Schottky-Barriere bzw. Barrierrhöhe genannt, U_D heißt Diffusionsspannung (vgl. Kap. 5.3). Die Barrierrhöhe wird maßgeblich vom Metall durch Φ_M bestimmt, ferner vom intrinsischen Halbleiterparameter χ (Elektronenaffinität). U_D ist über $\eta_n = W_L - W_F$ noch von der Dotierung, also einem extrinsischen Halbleiterparameter, abhängig.

Abb. 5.12 zeigt die Abhängigkeit der Barrierrhöhe von der Metallaustrittsarbeit Φ_M für zwei verschiedene Halbleiter, *n*-Silizium und *n*-Galliumarsenid. In beiden Fällen, insbesondere aber für *n*-GaAs, ist experimentell eine deutlich schwächere Abhängigkeit $\Phi_{Bn}(\Phi_M)$ festzustellen als aufgrund von (5.20) zu erwarten wäre. Die Ursache für diese Diskrepanz liegt in der Vernachlässigung der Grenzflächenzustände. In Abb. 5.13 sind sie im Bänderdiagramm und bei der Ladungsverteilung berücksichtigt. Der größte Teil der die Bandverbiegung verursachenden Differenz der Austrittsarbeiten $q(\Phi_M - \chi)$ wird nicht durch die Raumladung im Halbleiter, sondern durch Umladung der mit hoher Konzentration vorhandenen Grenzflächenzustände verursacht. Wegen dieser in der GaAs-Oberfläche hohen Konzentration werden schon bei sehr kleinen Änderungen $\Delta W = W_L - W_F$ sehr große Änderungen der Grenzflächenladung ΔQ_{ss} hervorgerufen. Man sagt, das Fermi-Niveau ist durch die Grenzflächenzustände "festgenagelt", engl.

“pinned”. Durch das “pinning” ändert sich gemäß Abb. 5.13 die Barrierenhöhe kaum, selbst wenn durch die Wahl eines anderen Metalls die Größe $q\Phi_M$ sich um mehrere eV ändert. Für n -Si findet man experimentell statt (5.20)

$$q\Phi_{Bn} = 0.27 \cdot q\Phi_M - 0.55 \text{ eV} \quad (5.22)$$

Bisher wurde nur der Metall-Halbleiterkontakt auf n -Halbleitern behandelt. Für p -Halbleiter muss die Barrierenhöhe in die negative Energierichtung gezählt werden, da die Löcher ihre energetisch günstigste Lage bei möglichst hohen Elektronen-Energiewerten haben. Aus Abb. 5.18 folgt dann für den Fall, dass man dasselbe Kontaktmetall einmal auf dem n -, und das andere Mal auf dem p -Material der gleichen Halbleitersubstanz (z.B. n - bzw. p -Silizium) verwendet. Hier ist “ W_F -pinning” vorausgesetzt! Dabei gilt

$$q\Phi_{Bn} + q\Phi_{Bp} = W_g \quad (5.23)$$

5.2.2 Kapazitäts-Spannungskennlinie

Der Metall-Halbleiterkontakt stellt, ähnlich wie die MIS-Struktur, eine Kondensatoranordnung dar, bei der die an freien Ladungsträgern verarmte Raumladungszone das Dielektrikum darstellt. Wie in der MIS-Struktur heißt die entsprechende Kapazität c_R , die Kapazität c_{IS} fehlt jedoch. Völlig analog gelten wieder für den Kapazitätsbelag, d.h. die Kapazität c_R pro Fläche des Halbleitermaterials

$$c_R = \frac{dQ_R}{dU} \quad (5.24)$$

und für die Ladung pro Fläche in der Raumladungszone

$$Q_R = qN_D^+ l \quad (5.25)$$

Für $U = 0$ stellt sich eine Raumladungszone mit der Dicke l_0 ein

$$l_0 = \sqrt{\frac{2\varepsilon_0\varepsilon_{r,HL}}{qN_D^+} U_D} \quad \text{für } U_D > 0 \quad (5.26)$$

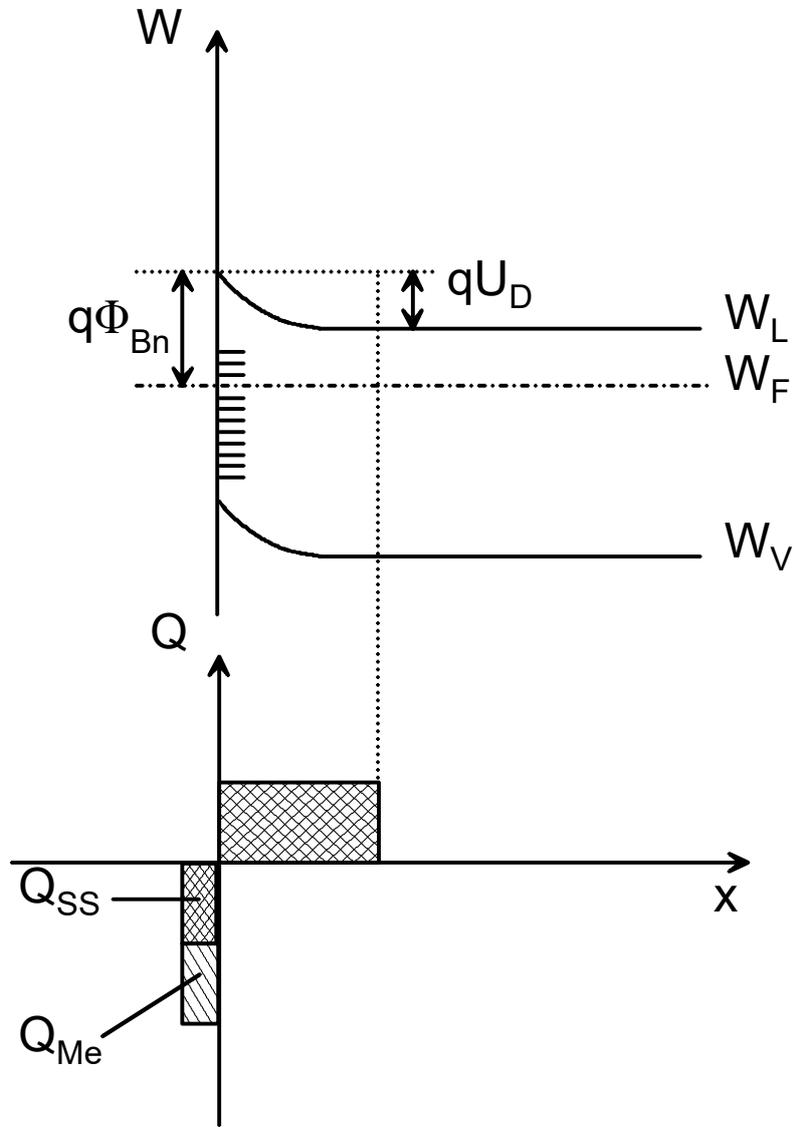


Abbildung 5.13: Bändermodell und Ladungsträgerverteilung unter Berücksichtigung von Grenzflächenzuständen

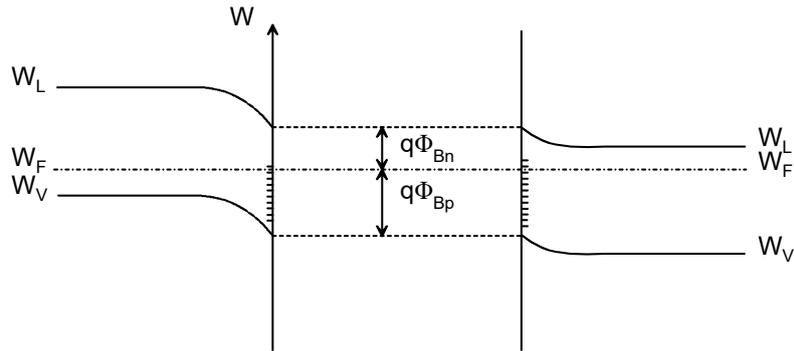


Abbildung 5.14: Komplementäres Verhalten von Barrierenhöhen auf n - bzw. p -Halbleitern

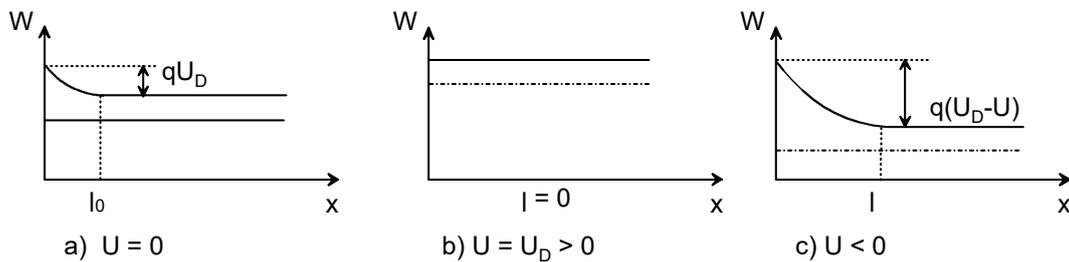


Abbildung 5.15: Verlauf der Leitungsbandkante für unterschiedliche Spannungen

Dabei entspricht U_D der Flachbandspannung U_{FB} der realen MIS-Struktur. Durch eine außen angelegte Spannung kann die Dicke der Raumladungszone verändert werden:

$$l = \sqrt{\frac{2\varepsilon_0\varepsilon_{r,HL}}{qN_D^+}(U_D - U)} \quad (5.27)$$

Für $U = U_D$ ist die Raumladungszone völlig abgebaut, d.h. $l = 0$. Mit (5.11) und (5.12) folgt für den Kapazitätsbelag

$$c_R = \sqrt{\frac{\varepsilon_0\varepsilon_{r,HL}qN_D^+}{2(U_D - U)}} \quad \text{für } U \leq U_D \quad (5.28)$$

Der Fall $U > U_D$ ist für die Kapazitätsberechnung uninteressant, da kein Dielektrikum mehr vorhanden ist. Für die Strom-Spannungs-Kennlinie wird dieser Fall wieder aufgegriffen. In Abb. 5.15 sind die Bandverläufe für unterschiedliche Spannungen gezeigt. Die Gln. (5.27) und (5.28) zeigen, dass mit steigender negativer Spannung die Ausdehnung der Raumladungszone mit $\sqrt{U_D - U}$ zunimmt, und die Kapazität mit dem

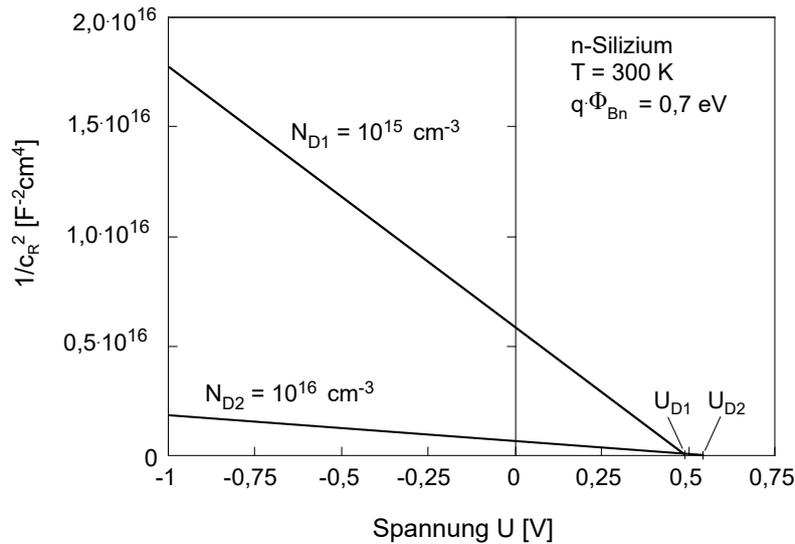


Abbildung 5.16: Spannungsabhängigkeit der Raumladungskapazität für zwei verschiedene Dotierungen

Kehrwert $1/\sqrt{U_D - U}$ abnimmt: das Bild des Plattenkondensators mit dem Plattenabstand l ist auch für die Schottky-Diode ein funktionsfähiges Modell. Ferner ist c_R von der Dotierung des Halbleiters abhängig. Auf der Spannungsabhängigkeit der Schottky-Diode beruht sogar das (neben dem Hall-Effekt) wichtigste Verfahren zur Bestimmung der Dotierstoffkonzentration. Aus (5.28) folgt

$$\frac{1}{c_R^2} = \frac{2}{q\varepsilon_0\varepsilon_r\text{HL}N_D^+}(U_D - U) = \frac{K}{N_D^+} \cdot (U_D - U) \quad \text{für } U \leq U_D \quad (5.29)$$

Für eine bestimmte Halbleitersubstanz ist K eine Konstante. Aus einer grafischen Darstellung $1/c_R^2 = f(U)$ folgt zunächst aus der Steigung die unbekannte Dotierstoffkonzentration N_D , vorausgesetzt, dass alle Donatoren ionisiert sind ($N_D = N_D^+$). Ferner ist aus dem Achsenabschnitt die Diffusionsspannung bestimmbar. Abb. 5.16 zeigt das Diagramm für zwei unterschiedliche Konzentrationen $N_{D,1}$, $N_{D,2}$. Man beachte, dass sich auch zwei verschiedene U_D ergeben, da die Dotierung auch die Größe $\eta = W_L - W_F$ verändert. Neben dieser messtechnischen Anwendung wird die Schottky-Diode als Varaktor (variable reactor) meist in der Hochfrequenztechnik als einstellbare Kapazität eingesetzt. Jeder Radioempfänger enthält diese Abstimmelemente in Schwingkreisen.

5.2.3 Strom-Spannungs-Kennlinie des Metall-Halbleiter-Kontaktes

Der Stromfluss durch die Metall-Halbleitergrenzfläche muss von Ladungsträgern getragen werden, die die Barriere an der Grenzfläche überwinden können. Bei Beschränkung auf n -Halbleiter sind dies Elektronen, die die Barriere $q \cdot \Phi_{Bn}$ bzw. $q \cdot U_D$ überwinden. Nach klassischer Physik können das nur Elektronen, deren kinetische Energie größer ist als der Höhe der Potentialbarriere entspricht. Für Bauelemente, die eine Diodencharakteristik (Sperr- und Durchlassbereich) aufweisen, bildet dieser Elektronenstrom auch den Hauptanteil des Diodenstromes. Im folgenden Abschnitt 5.2.4 wird auf klassischer Grundlage die Strom-Spannungs-Kennlinie der Schottky-Diode berechnet. Die Funktion des sperrfreien Metall-Halbleiterkontaktes (Kap. 5.2.5) lässt sich aber erst durch Einbeziehen des quantenmechanischen Tunneleffektes (vgl. Kap. 1.4) erklären.

5.2.4 Schottky-Diode

Man geht davon aus, dass die kinetische Energie der zum Stromtransport beitragenden Elektronen mindestens so hoch ist wie $q\Phi_{Bn}$ bzw. qU_D . In Kapitel 2.4 wurde festgestellt, dass bei Raumtemperatur und nicht zu großen Feldern der weitaus größte Teil der kinetischen Energie thermischen Ursprungs ist

$$v_{th} \gg v_D$$

mit der thermischen Geschwindigkeit v_{th} und der Driftgeschwindigkeit v_D .

Man spricht daher auch von **thermionischer Emission** von Elektronen über die Barriere. Je nach Breite der Barriere und "mittlerer freier Weglänge" (MFW) zwischen zwei Zusammenstößen mit dem Gitter des Kristalls gibt es zwei Extremfälle für die Durchquerung der Strecke l , über die sich die Potentialbarriere erstreckt (Abb. 5.17).

- (a) Die Elektronen durchqueren von $x = l$ aus den Bereich der Barriere (Länge l) ohne Zusammenstöße, d.h. $MFW \gg l$. Die darauf aufbauende Beschreibung heißt Diodentheorie (nach H. Bethe, 1922).
- (b) Beim Durchqueren des Barriembereiches finden sehr viele Zusammenstöße mit dem Gitter statt, aber ohne dass die Elektronen dabei ihre kinetische Energie verlieren (elastische Stöße). Diese Art der Bewegung ist charakteristisch für die Diffusion. Es gilt dann $MFW \ll l$, diese Beschreibung heißt Diffusionstheorie (nach W. Schottky, 1938).

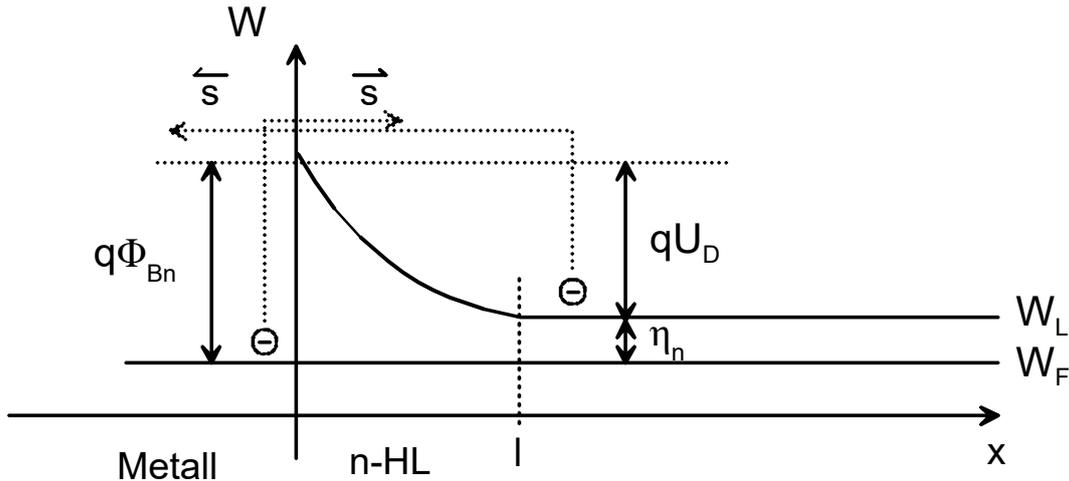


Abbildung 5.17: Thermionische Elektronenemission am Schottky-Kontakt ($U = 0$)

Keine der beiden Beschreibungen gilt in der Realität in reiner Form, vielmehr treten beide Fälle gemischt auf. Es wird jedoch gezeigt, dass die Ergebnisse beider Berechnungen zu sehr ähnlichen Ergebnissen führen. Zunächst sei die Diodentheorie behandelt:

a) Diodentheorie ($MFW \gg l$)

Die gesamte anliegende Spannung fällt über der Raumladungszone ab. Die Teilchenstromdichte über dem Kontakt ist die Summe bestehend aus dem Anteil "aus dem Metall in den Halbleiter" $s_{M \rightarrow HL}$ und "aus dem Halbleiter ins Metall" $s_{HL \rightarrow M}$.

Für $s_{M \rightarrow HL}$ gilt am Kontakt ($x = 0$):

$$s_{M \rightarrow HL}(0) = B \cdot v_{th}(0) \cdot n(0) = v_R \cdot n(0) \quad (5.30)$$

wobei v_R die mittlere Geschwindigkeit "nach rechts" bedeutet, mit der thermischen Geschwindigkeit v_{th} , der Elektronenkonzentration n , und einer Proportionalitätskonstanten B können wir unter der Annahme schreiben, dass v_R in erster Näherung linear mit v_{th} zusammenhängt, vgl. Kap. 2.2.3, d.h.

$$v_R = B \cdot v(0)$$

mit

$$n(0) = N_L \cdot e^{-q\Phi_{Bn}/kT} \quad (5.31)$$

sodass gilt:

$$s_{M \rightarrow HL}(0) = v_R \cdot N_L \cdot e^{-q\Phi_{Bn}/kT} \quad (5.32)$$

Für $s_{HL \rightarrow M}$ gilt bei $x = l$

$$s_{HL \rightarrow M}(l) = -Cv_{th} \cdot n(l) = -v_R \cdot n(l) \quad (5.33)$$

Dabei ist zu berücksichtigen, dass sich das Minuszeichen aus der Elektronenbewegung in $(-x)$ -Richtung ergibt. B und C sind Konstanten, die sich aus einer Mittelwertbildung über die Geschwindigkeitsverteilung ergeben und auch berücksichtigen, dass nur Elektronen, die sich auf den Kontakt zu bewegen, zum Strom beitragen.

Von der Konzentration $n(l)$ erreicht nur der Anteil den Kontakt bei $x = 0$, dessen kinetische Energie bei $x = 0$ größer oder gerade gleich der halbleiterseitigen Barriere ist. Der Anteil kann über die Boltzmannverteilung mit einem Faktor $e^{-q(U_D - U)/kT}$ ermittelt werden

$$\begin{aligned} s_{HL \rightarrow M}(0) &= -v_R \cdot n(0) = -v_R \cdot n(l) \cdot e^{-q(U_D - U)/kT} \\ &= -v_R \cdot N_L \cdot e^{-\eta_n/kT} \cdot e^{-qU_D/kT} \cdot e^{qU/kT} \\ &= -v_R \cdot N_L \cdot e^{-\Phi_{Bn}/kT} \cdot e^{qU/kT} \end{aligned} \quad (5.34)$$

da gilt

$$qU_D + \eta_n = q\Phi_{Bn} \quad (5.35)$$

Für die elektrische Gesamtstromdichte findet man

$$\begin{aligned} J &= -q \cdot [s_{M \rightarrow HL}(0) + s_{HL \rightarrow M}(0)] \\ &= qv_R \cdot N_L \cdot e^{-q\Phi_{Bn}/kT} \cdot (e^{qU/kT} - 1) = J_0 \cdot (e^{qU/kT} - 1) \end{aligned} \quad (5.36)$$

Allgemein ist die Temperaturabhängigkeit der berechneten Größen von hoher Bedeutung. Um dies weiter zu verfolgen, setzen wir die Definition der thermischen Geschwindigkeit ein, die ja eine mittlere kinetische Energie $mv^2/2$ beschreibt

$$\frac{1}{2}mv_{th}^2 = \frac{3}{2}kT$$

die Geschwindigkeit "nach rechts" findet in einer Dimension statt, d.h. es wird $kT/2$ für die Proportionalität angesetzt

$$v_R \propto T^{1/2}$$

weiterhin wissen wir bereits, dass die effektive Zustandsdichte N_L auch von der Temperatur abhängt (2.32)

$$N_L \propto T^{3/2} \quad (5.37)$$

Damit kann aus dem Faktor J_0 die Temperaturabhängigkeit herausgezogen werden:

$$J_0 = A^* \cdot T^2 \cdot e^{-q\Phi_{Bn}/kT} = q \cdot v_R \cdot n(0) \quad (5.38)$$

A^* wird effektive Richardson-Konstante genannt. Der Strom $I_{Du} = J_{Du} \cdot A$ wird Durchlassstrom genannt. Er steigt mit wachsender positiver Spannung exponentiell an. Wird die Spannung umgepolt, so liegt die Sperrichtung vor:

$$J_{Sp} = -q \cdot [s_{M \rightarrow HL}(0) + s_{HL \rightarrow M}(0)] = -J_0 \cdot \left(1 - e^{qU/kT}\right) \quad \text{für } U < 0 \quad (5.39)$$

Für große negative Werte von U gilt:

$$J_{Sp} = -J_0 \quad (5.40)$$

unabhängig von U . Der Sperrstrom J_{Sp} ist sehr viel kleiner als der Durchlassstrom J_{Du} aus (5.36), sodass eine typische Diodencharakteristik mit Gleichtrichterwirkung resultiert (Abb. 5.18 bzw. Abb. 5.19).

b) Diffusionstheorie ($MFW \ll l$)

Im vorhergehenden Fall der Diodentheorie wurde bereits benutzt, dass aufgrund der Bandverbiegung die Elektronenkonzentration am Kontakt ($x = 0$) um den Faktor

$$\exp(-q(U_D - U)/kT)$$

gegenüber der Konzentration am Raumladungszonenrand verringert ist. Mit diesem Konzentrationsgefälle ist laut Kap. 2.4 ein Teilchenstrom verknüpft. Nach (2.91) gilt

$$n(x) = N_L \cdot e^{-(W_L(x)-W_F)/kT} = N_L \cdot e^{-\eta_n(x)/kT}$$

und mit (2.88)

$$s_{\text{HL} \rightarrow \text{M}} = -D_n \cdot \nabla n(x) = -\mu_n kT/q \cdot \nabla n(x) \quad (5.41)$$

Ferner verursacht die Bandverbiegung einen ortsabhängigen Potentialverlauf, wie bereits in Abschnitt 2.4.3 mit (2.85) beschrieben:

$$\nabla W_L(x) = -q \cdot \nabla \Phi(x) = q \cdot E \quad (5.42)$$

Das so berechnete E-Feld treibt seinerseits einen Feldstrom. Dieser ist aber als Teilchenstrom vom Metall in den Halbleiter gerichtet:

$$s_{\text{M} \rightarrow \text{HL}}(x) = -\mu_n \cdot n(x) \cdot \nabla \Phi(x) \quad \text{mit } \nabla \Phi(x) < 0 \quad (5.43)$$

Je nach Polarität der angelegten Spannung gilt wieder:

$$\begin{aligned} J_{Du} &= -q[s_{\text{HL} \rightarrow \text{M}}(x) + s_{\text{M} \rightarrow \text{HL}}(x)] \\ J_{Sp} &= +q[s_{\text{HL} \rightarrow \text{M}}(x) + s_{\text{M} \rightarrow \text{HL}}(x)] \end{aligned} \quad (5.44)$$

Die Ströme müssen am gleichen Ort x summiert werden. Der Gesamtstrom ist aber wegen der vorausgesetzten Kontinuität ($\nabla \cdot \vec{J} = 0$) ortsunabhängig, da es sich um einen geschlossenen Stromkreis handelt.

Mit Verwendung von (5.41) und (5.43) folgt aus (5.44):

$$J_{Du} = \mu_n \cdot kT \cdot \nabla n(x) + q \cdot \mu_n \cdot n(x) \cdot \nabla \Phi(x) \quad (5.45)$$

$$J_{Sp} = -\mu_n \cdot kT \cdot \nabla n(x) - q \cdot \mu_n \cdot n(x) \cdot \nabla \Phi(x) \quad (5.46)$$

Eine geschlossene Lösung dieser Differentialgleichungen erhält man durch Ersetzen des ortsabhängigen Potentialgradienten durch eine konstante Feldstärke E_R am Kontakt.

Man findet dann:

$$J_{Du} = q \cdot n(0) \cdot \mu_n \cdot E_R \cdot \left(e^{qU/kT} - 1 \right) = J_0 \cdot \left(e^{qU/kT} - 1 \right) \quad \text{für } U \geq 0 \quad (5.47)$$

$$J_{Sp} = -q \cdot n(0) \cdot \mu_n \cdot E_R \cdot \left(1 - e^{qU/kT} \right) = -J_0 \cdot \left(1 - e^{qU/kT} \right) \quad \text{für } U < 0 \quad (5.48)$$

Die Größe $\mu_n \cdot E_R = v_D$ kann dabei als Driftgeschwindigkeit der Elektronen am Kontakt interpretiert werden, so dass

$$J_0 = q \cdot v_D \cdot n_R \quad (5.49)$$

c) Diskussion von Dioden- und Diffusionstheorie

Beide Beschreibungen liefern sehr ähnliche Ergebnisse: Es ergibt sich jeweils eine Diodencharakteristik mit Durchlass- und Sperrbereich. Sogar die Ansätze für die Sperrstromdichte J_0 unterscheiden sich lediglich durch die Größen v_R (5.38) bzw. v_D (5.49). Eine Abschätzung liefert für beide ähnliche Werte von $v_D \approx v_R \approx 5 \times 10^6$ cm/s. Allerdings ist v_R wegen des direkten Zusammenhanges mit der thermischen Geschwindigkeit spannungsunabhängig, v_D dagegen hängt über die Randfeldstärke $E_R = (U_D - U)/l$ von der angelegten Spannung ab. Der Sperrstrom I_{Sp} ist daher für die Diodentheorie konstant, für die Diffusionstheorie steigt er schwach mit der Sperrspannung an (Abb. 5.18). In beiden Fällen steigt der Durchlassstrom I_{Du} für $U \gtrsim 3kT/q$ exponentiell an, so dass man näherungsweise schreiben darf:

$$I_{Du} = I_0 \cdot e^{qU/kT} \quad \text{für } U \gtrsim 3kT/q \quad (5.50)$$

Die hier entwickelten Modelle gelten nur bis zum Flachbandfall, d.h. bis $U \approx 0.6$ V für Silizium. Darüber hinaus fällt die zusätzlich angelegte Spannung im wesentlichen in den Bahngebieten in größerer Entfernung vom Kontakt selbst ab. Dieser Einfluss wird später noch diskutiert. Für sehr kleine Spannungen kann man die e -Funktion entwickeln:

$$I/I_0 = e^{qU/kT} - 1 \approx (1 + U/U_T) - 1 = U/U_T \quad \text{mit } kT/q = U_T \quad (5.51)$$

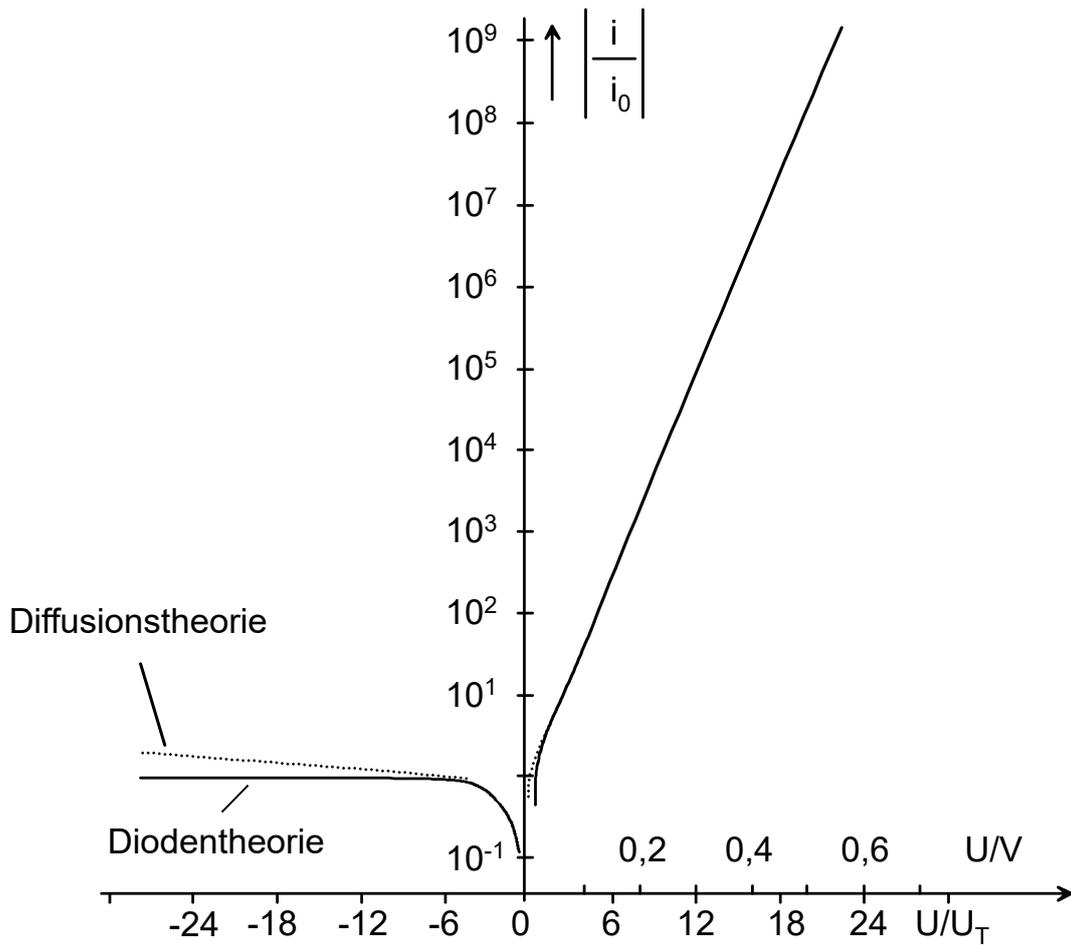


Abbildung 5.18: Normierte Strom-Spannungskennlinien einer Schottky-Diode bei $T = 300 \text{ K}$

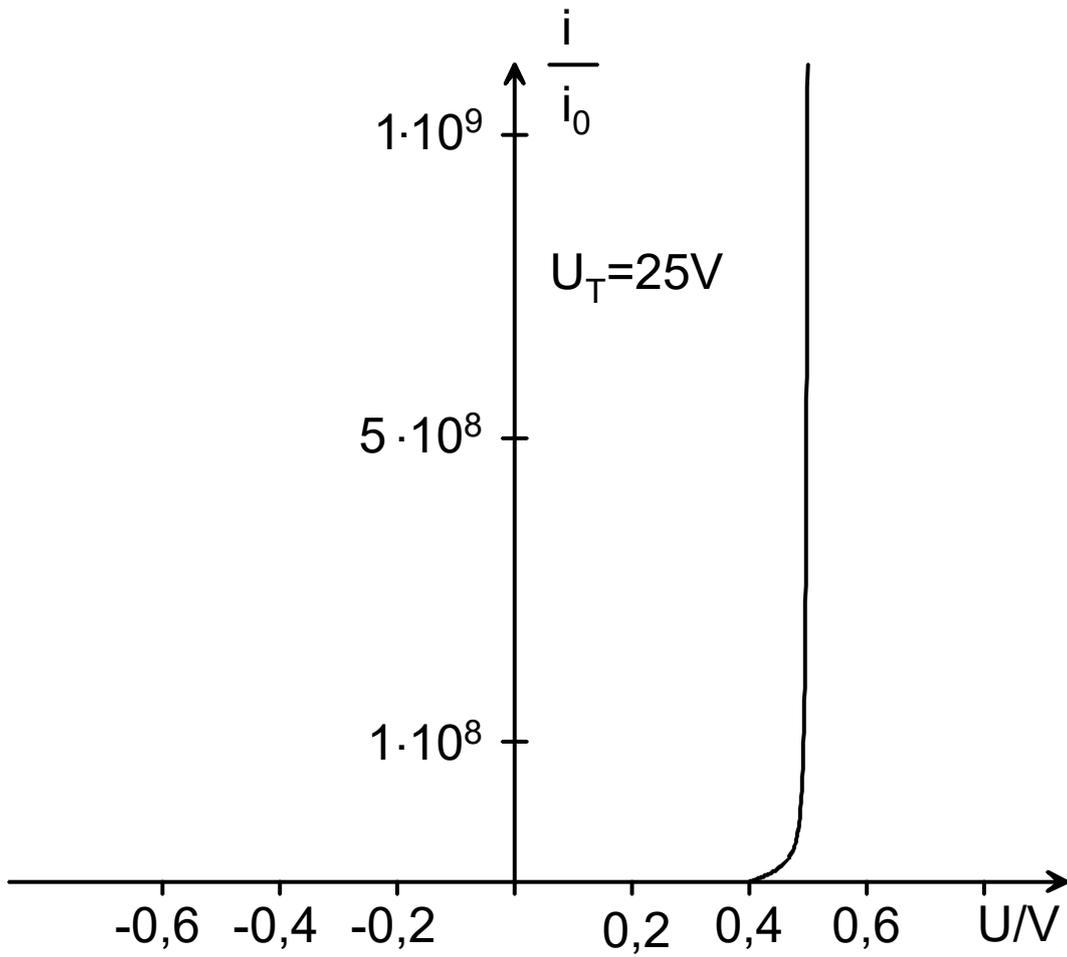


Abbildung 5.19: wie Abb.5.18, jedoch linearer Massstab

Das Verhalten ist linear, sodass sich die Schottky-Diode bei kleinen Spannungen wie ein Ohm'scher Widerstand verhält.

Einfluss des Bahnwiderstandes

Mit steigendem Strom fällt ein steigender Anteil der angelegten Spannung über den Bahngebieten ausserhalb der Raumladungszone ab. Die Bahngebiete können als Ohmscher Serienwiderstand R_b dargestellt werden. Um diesen Betrag vermindert sich die Spannung über der eigentlichen (intrinsischen) Schottky-Diode:

$$U_{Diode} = U - I \cdot R_b \quad (5.52)$$

Die resultierende Kennliniengleichung lautet mit der Diodenfläche A , die den absoluten Strom $I = J \cdot A$ bzw. $I_0 = J_0 \cdot A$ mit den entsprechenden Stromdichten J verknüpft.

$$I = I_0 \cdot \left[e^{(U - I \cdot R_b)/U_T} - 1 \right] \quad (5.53)$$

Diese Gleichung ist explizit nicht nach I auflösbar. Die Konstruktion der Kennlinie gelingt aber leicht auf graphischem Weg bei Kenntnis der Kennlinie der idealen Diode und des Bahnwiderstandes (Abb. 5.20), indem der Spannungsabfall an R_b bei jedem Stromwert zur Diodenspannung U_{Diode} addiert wird.

5.2.5 Ohmsche Kontakte

Ohm-Kontakte sind solche mit einer Ohmschen, d.h. linearen Strom-Spannungskennlinie über den ganzen Betriebsbereich. Nach der bisherigen Behandlung existieren derartige Kontakte nicht, da der Metall-Halbleiter-Kontakt immer auch einen Sperrbereich der Kennlinie aufweist. Für einen sperrfreien Kontakt muss zusätzlich zur bisherigen Behandlung der quantenmechanische Tunneleffekt berücksichtigt werden. Dieser tritt aber nach Kap. 1.5 nur bei sehr dünnen Barrieren in Erscheinung, wie sie bei einfachem Aufbringen von Metallen auf Halbleiteroberflächen nicht realisierbar sind.

Die Breite der Potentialbarriere hängt nach (5.27) von der Dotierung ab: wenn diese im Bereich des Kontaktes stark erhöht wird, so wird die Barriere sehr schmal und der Tunnelstrom überwiegt den im vorangegangenen Kapitel berechneten Strom bei weitem. Dabei ist das Vorzeichen der angelegten Spannung unerheblich, weil die Elektronen im Metall die gleiche Barrierendicke wie die im Halbleiter sehen. Ferner ist auch die

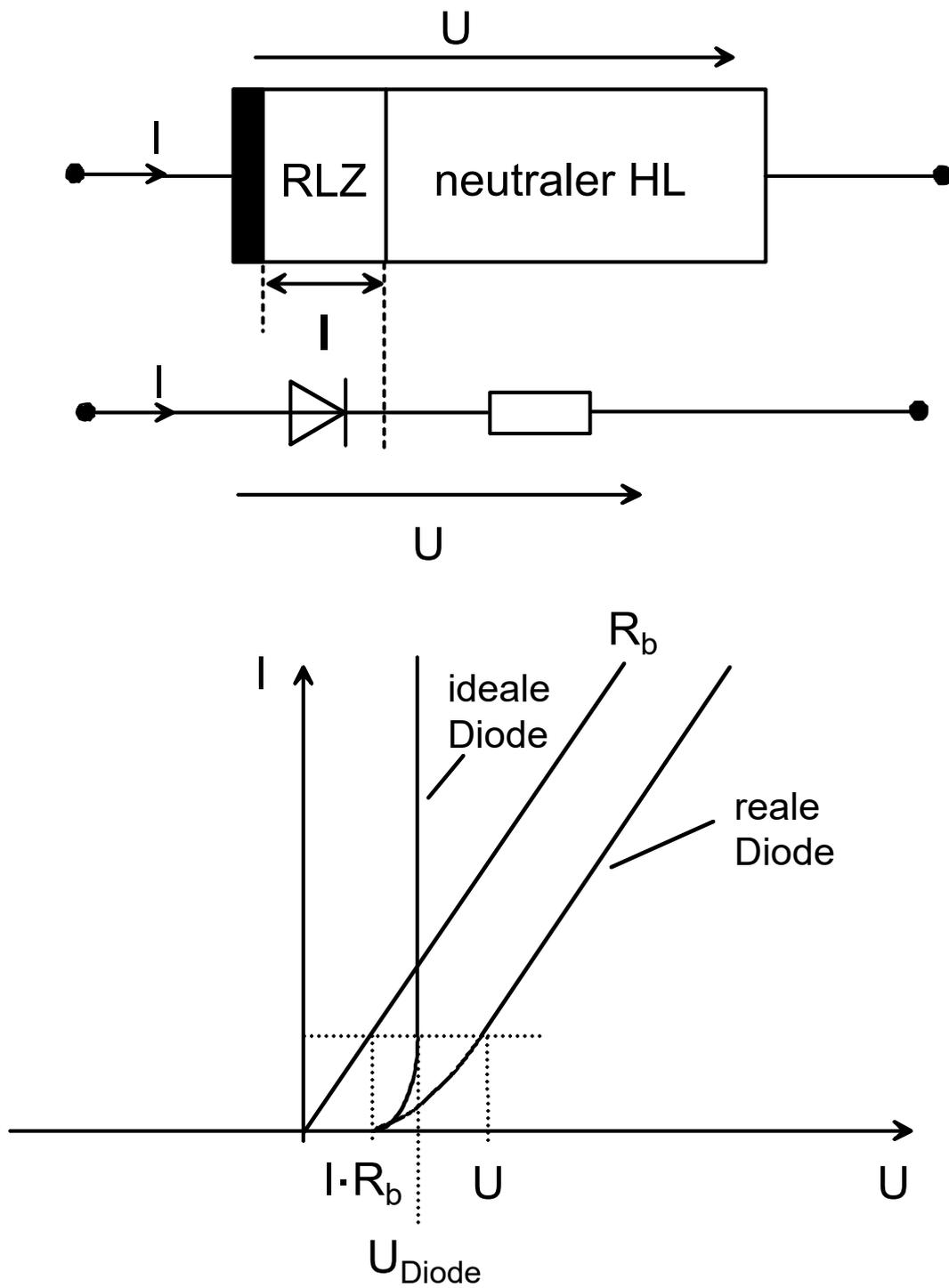


Abbildung 5.20: Einfluss des Bahnwiderstandes R_b auf die Kennlinie der Schottky-Diode

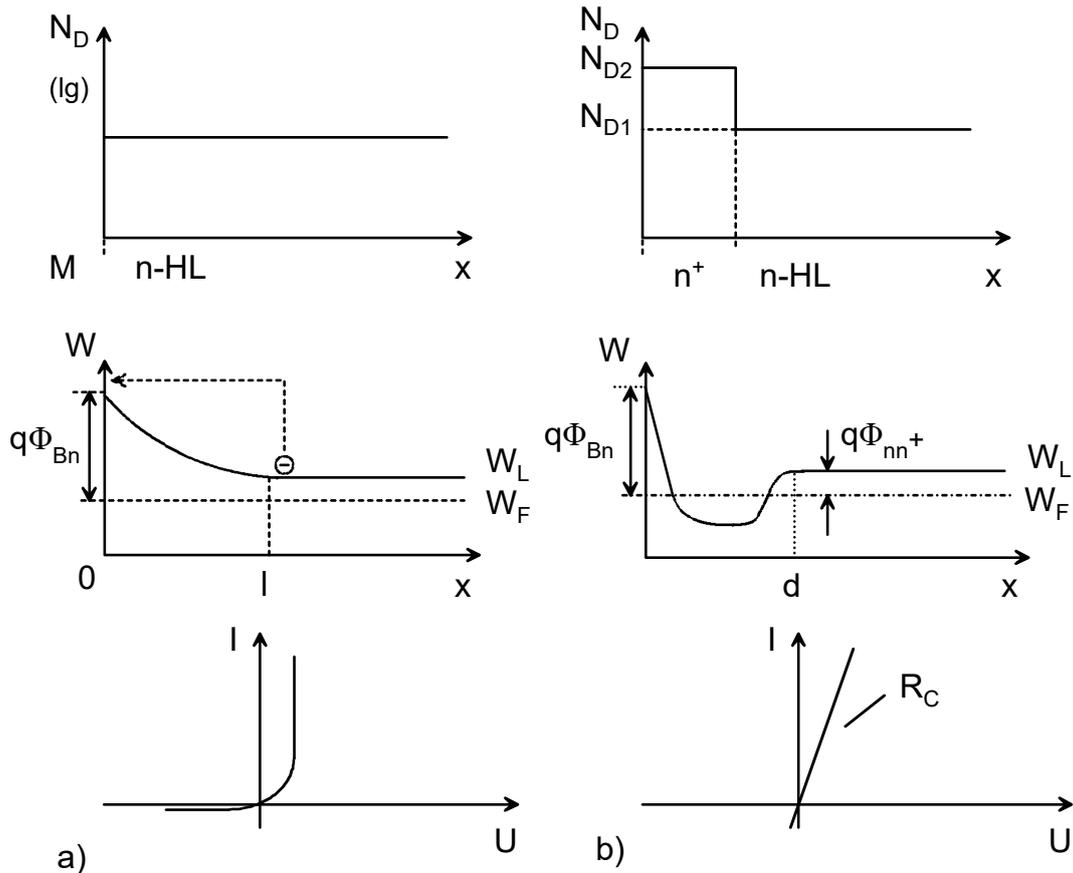


Abbildung 5.21: Bändermodell und Kennlinie eines Schottky- (a) und Ohm-Kontaktes (b)

Barrierenhöhe für beide etwa gleich, wie aus Abb. 5.21 hervorgeht. Hier ist angenommen, dass in einer Dicke d der Halbleiteroberfläche die Dotierung von $N_{D,1}$ auf $N_{D,2}$ so stark erhöht wurde, dass der Halbleiter entartet ist und folglich das Fermi-Niveau oberhalb der Leitungsbandkante liegt. In Abb. 5.21 ist neben der eigentlichen Schottky-Barriere eine weitere mit $q\Phi_{nn^+}$ gekennzeichnete Barriere erkennbar. Bei geringer Dotierung $N_{D,1}$ ist sogar diese und nicht die eigentliche Schottky-Barriere bestimmend für den Kontaktwiderstand R_C . Die Herstellung von Ohm-Kontakten ist in Abb. 5.22 skizziert. Zunächst wird im Hochvakuum Dotierstoff aufgebracht und das Kontaktmetall aufgedampft oder gesputtert. Alternativ wird zuerst die Dotierung mittels Ionenimplantation eingebracht. Anschließend wird die ganze Probe kurzzeitig stark erhitzt, sodass der n -Dotierstoff aus der Aufdampfschicht kontrolliert in die Halbleiteroberfläche eindiffundiert und hier eine dünne n^+ -Schicht erzeugt. Nach dem Abkühlen auf

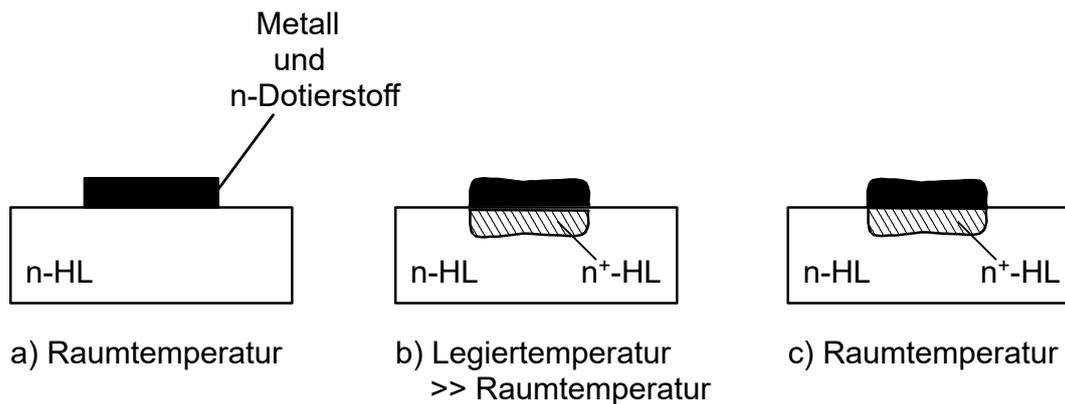


Abbildung 5.22: Herstellung eines Ohm-Kontaktes durch Legieren

Raumtemperatur wird dieser Zustand eingefroren, und damit liegen die Verhältnisse nach Abb. 5.21 vor.

Die Abb. 5.23 zeigt die Abhängigkeit des spezifischen Kontaktwiderstandes $\rho_C = R_C/A$ von der Dotierung N_D für n -GaAs. Der Kontaktwiderstand ändert sich um mehr als 3 Größenordnungen mit logarithmisch steigender Dotierung. Er sättigt bei hohen Konzentrationen, weil hier die Barriere $q\Phi_{nn}^+$ abgebaut ist. Einem geringen Kontaktwiderstand kommt hohe technologische Bedeutung zu, weil die Verlustleistung elektronischer Bauelemente so gering wie möglich sein muss. Dies gilt insbesondere bei hoch integrierten Schaltungen. Zweitens führen Serienwiderstände zu Signalverzögerungen: das Produkt $\tau = RC$ bildet eine Zeitkonstante, die die Bandbreite der elektronischen Schaltung bestimmt. Über die Serienwiderstände werden intrinsische (im Bauelement) und extrinsische (in den Zuleitungen) Kapazitäten geladen und entladen.

5.3 Halbleiter-Halbleiter-Grenzflächen

Man unterscheidet zwischen Homoübergängen mit gleichem Halbleitermaterial, aber unterschiedlicher Dotierung, und Heteroübergängen, bei denen an der Grenzfläche das Halbleitermaterial wechselt, also beispielsweise auch der Bandabstand W_g oder die Valenz- und Leitungsbandkanten Diskontinuitäten aufweisen.

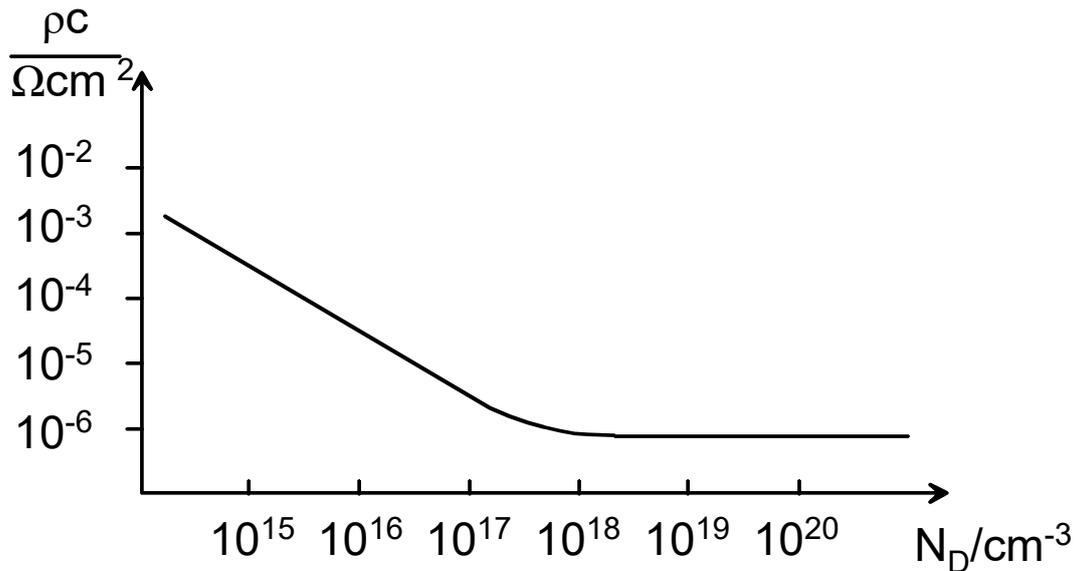


Abbildung 5.23: Typischer spezifischer Kontaktwiderstand in Abhängigkeit von der Dotierung

5.3.1 pn -Übergang zwischen Halbleitern gleichen Materials (Homoübergang)

Der pn -Übergang in ein und demselben Halbleitermaterial ist das Standardbeispiel für Halbleiter-Homoübergänge. Er bildet das Kernstück vieler Bauelemente wie pn -Diode, Bipolartransistor, Thyristor, Photodiode, Solarzelle und auch LED und Laserdiode. Selbst im CMOS-Transistor sind pn -Übergänge zu finden.

Zunächst wird der strom- und spannungslose pn -Übergang behandelt. Es zeigt sich, dass der stromdurchflossene pn -Übergang anschließend als System behandelt werden kann, bei dem die Unterschiede zum stromlosen Fall als nur kleine Störung betrachtet werden können.

5.3.2 Stromloser pn -Übergang; Bänderdiagramm und Ladungsverteilung

Dieses Kapitel kann als Vorbereitung für die Beschreibung des stromdurchflossenen pn -Überganges aufgefasst werden. Ströme werden durch zwei Mechanismen angetrieben: durch ein elektrisches Feld und bzw. oder durch Ladungsträgerdiffusion. Elektrische Felder korrespondieren mit Bandverbiegungen, Ladungsträgerdiffusion mit Konzentrationsgradienten. Daher sollen bereits in diesem Kapitel die Ortsabhängigkeit des Bänderdiagramms sowie der Konzentrationsverlauf der Ladungsträger und der Dotierstoffato-

me behandelt werden. Ohne Verfälschung der grundsätzlichen Verhältnisse werden zur Vereinfachung folgende Voraussetzungen gemacht:

1. Abruptes Dotierstoffprofil: die Dotierstoffkonzentrationen N_D bzw. N_A sind im n - bzw. p -Gebiet konstant und fallen an der Grenzfläche abrupt auf Null (vgl. Abb. 5.24a), d.h.

$$N_D = \begin{cases} \text{const.} & \text{für } x > l_g \\ 0 & \text{sonst} \end{cases} \quad (5.54)$$

$$N_A = \begin{cases} \text{const.} & \text{für } x < l_g \\ 0 & \text{sonst} \end{cases} \quad (5.55)$$

2. Alle Dotierstoffatome sind ionisiert:

$$\begin{aligned} N_A^- &= N_A \\ N_D^+ &= N_D \end{aligned} \quad (5.56)$$

3. Andere Störstellen, insbesondere Grenzflächenzustände, existieren nicht. Weit entfernt von der Grenzfläche gelten ferner (vgl. Abb. 5.24):

$$\begin{aligned} p_{p0} &= N_A^- = N_A \\ n_{n0} &= N_D^+ = N_D \end{aligned} \quad (5.57)$$

Die Indizes n bzw. p in (5.56) geben das n - bzw. p -Gebiet an, der Index 0 kennzeichnet den Gleichgewichtszustand ($U = 0, I = 0$). Für diesen Gleichgewichtszustand (thermodynamisches Gleichgewicht) gilt das Massenwirkungsgesetz, das die Minoritätenkonzentrationen festlegt, und zwar die Minoritäten-Elektronen im p -Gebiet:

$$n_{p0} = \frac{n_i^2}{p_{p0}} = \frac{n_i^2}{N_A} \quad (5.58)$$

und die Löcher-Minoritäten im n -Gebiet:

$$p_{n0} = \frac{n_i^2}{n_{n0}} = \frac{n_i^2}{N_D} \quad (5.59)$$

In Abb. 5.24 wurde zur Konstruktion wieder $W_F = \text{const.}$ für den stromlosen Fall (ther-

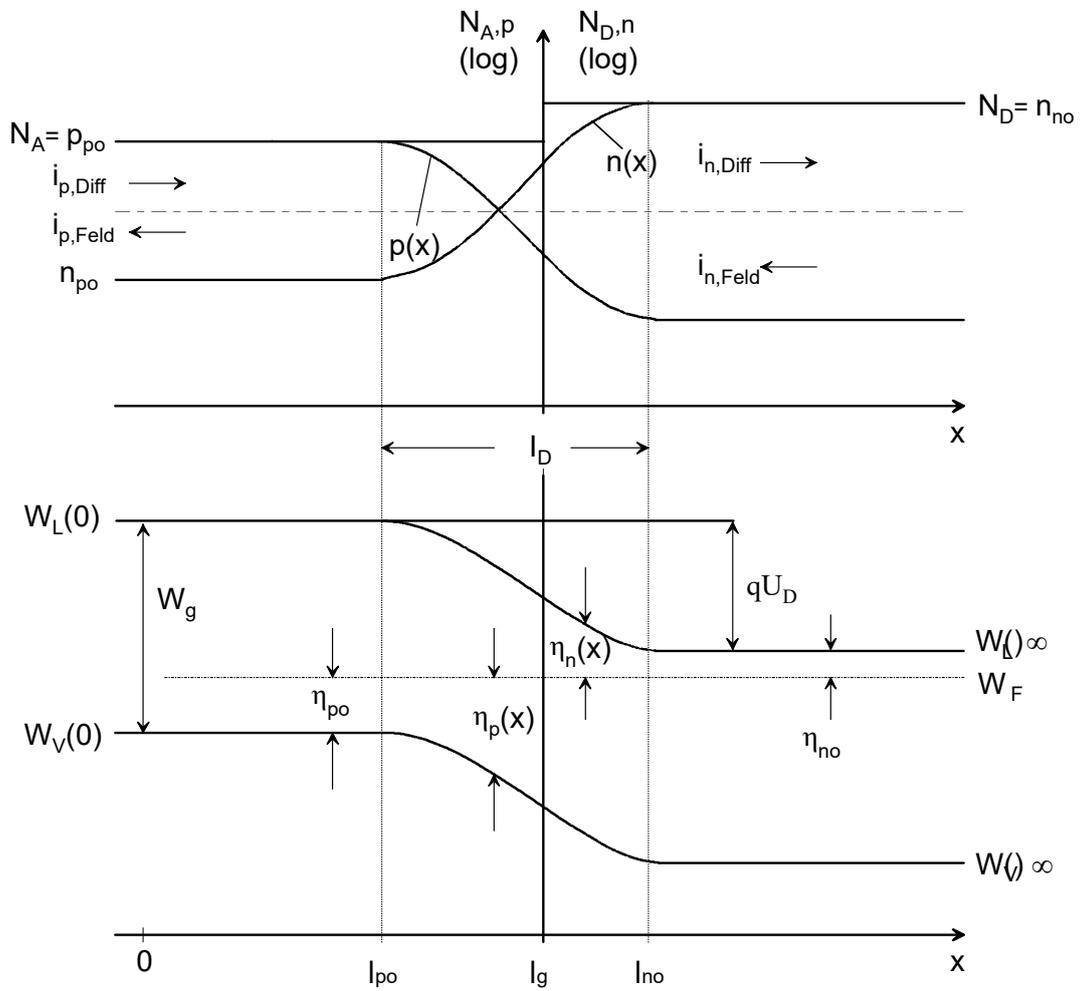


Abbildung 5.24: (a) Dotierstoff- und Ladungsträgerkonzentration sowie (b) Bändermodell eines abrupten pn -Überganges

modynamisches Gleichgewicht) benutzt. (5.57) stellt in den von der Grenzfläche weit entfernten Bereichen die lokale Ladungsneutralität sicher, da hier die Dotierstoffladung (N_D^+ bzw. N_A^-) jeweils von den freien Ladungsträgern (n bzw. p) kompensiert wird. Im Bereich der Grenzfläche werden die freien Ladungsträger aufgrund ihrer Beweglichkeit nicht dem abrupten Profil der Dotierstoffatome folgen, weil die Diffusion die Elektronen vom n - ins p -Gebiet und die Löcher vom p - ins n -Gebiet treibt. Hier findet daher keine Kompensation von freien (n, p) und ortsfesten (N_D^+, N_A^-) Ladungen statt, und somit ist die lokale Ladungsneutralität nicht mehr gegeben: Im n -Gebiet sorgen die Donatorrümpfe (N_D^+) für eine positive Raumladung, im p -Gebiet die Akzeptorrümpfe (N_A^-) für eine negative.

Gemäß der Poisson-Gleichung entsteht eine Bandverbiegung und damit ein elektrisches Feld (Abb. 5.24b), das der Diffusionsbewegung der Elektronen und Löcher entgegenwirkt und zu einem Gleichgewichtszustand führt. Der Potentialunterschied zwischen n - und p -Seite ist aus der Abbildung direkt ablesbar:

$$qU_D = W_L(x = 0) - W_L(x = \infty) \quad (5.60)$$

und ferner

$$W_g - \eta_{p0} = qU_D + \eta_{n0} \quad (5.61)$$

bzw.

$$qU_D = W_g - \eta_{p0} - \eta_{n0}$$

Mit (2.37) und (2.46) gilt:

$$\begin{aligned} n_{n0} &= N_L \cdot e^{-\eta_{n0}/kT} \\ p_{p0} &= N_V \cdot e^{-\eta_{p0}/kT} \\ n_i^2 &= N_L \cdot N_V \cdot e^{-W_g/kT} \end{aligned} \quad (5.62)$$

Ferner gilt mit (5.57)

$$\begin{aligned} n_{n0} &= N_D \\ n_{p0} &= N_A \end{aligned} \quad (5.63)$$

Nach geeigneter Umformung findet man:

$$qU_D = kT \left[\ln \frac{N_L \cdot N_V}{n_i^2} - \ln \frac{N_V}{N_A} - \ln \frac{N_L}{N_D} \right] = kT \ln \frac{N_A N_D}{n_i^2}$$

bzw. mit $U_{th} = kT/q$

$$U_D = U_{th} \ln \frac{N_A \cdot N_D}{n_i^2} \quad (5.64)$$

Aus Abb. 5.24b erkennt man, dass qU_D gerade die Barriere ist, die sowohl Elektronen als auch Löcher für den Strom über den pn -Übergang überwinden müssen. Anders ausgedrückt: zur Erzeugung des Flachbandfalles muss außen an den pn -Übergang die Spannung $U_D > 0$ angelegt werden. Dies ist beim realen pn -Übergang nicht möglich, da dies durch den Spannungsabfall in den Bereichen außerhalb des eigentlichen Überganges (Bahngebiete) verhindert wird.

Aus (5.64) ist die Abhängigkeit der Diffusionsspannung von der Dotierung (N_D, N_A) und den intrinsischen Halbleiterdaten (n_i und damit N_L, N_V, W_g) ersichtlich. Da diese Größen als Argumente der (ln)-Funktion auftreten, ist deren Einfluss auf U_D nur schwach. Sowohl (5.64), aber auch direkt aus Abb. 5.24b, ist abzulesen, dass für eine ansteigende Dotierung N_D und N_A die Fermi-Energie W_F im n -Bereich gegen W_L , und im p -Bereich gegen W_V strebt, sodass sich U_D dem Wert W_g/q nähert. Für die Dotierungen $N_D = N_A = 1 \times 10^{17} \text{ cm}^{-3}$ wird für Silizium $U_D \approx 0.8 \text{ V}$ und für GaAs $U_D \approx 1.3 \text{ V}$. Wegen der geringen Zustandsdichte im Leitungs- und Valenzband des GaAs liegt speziell dieser U_D -Wert nahe dem Wert, der der Bandlückenenergie entspricht: $W_g/q = 1.42 \text{ V}$.

Es sollen jetzt die Konzentrationen n und p sowie der Bandverlauf, die gemäß Abb. 5.24 bisher nur qualitativ beschrieben wurden, mit Hilfe der Poisson-Gleichung berechnet werden. Wie bereits erwähnt, können aus dem ortsabhängigen Verlauf von n und p die Diffusionsströme, und aus dem Bandverlauf (und damit dem Potentialverlauf) die Feldströme berechnet werden.

Es gilt die Poisson-Gleichung (2.118):

$$\Delta\Phi = -\nabla \cdot \vec{E} = -\frac{\rho}{\varepsilon_0 \varepsilon_r} \quad \text{mit } \rho = q \cdot [N_D^+ + p(x) - N_A^- - n(x)] \quad (5.65)$$

Da die Konzentration der freien Ladungsträger n, p innerhalb der Raumladungszone am pn -Übergang stark reduziert ist, können sie in einem ersten Ansatz gegenüber den

ortsfesten Ladungen (N_D^+ , N_A^-) vernachlässigt werden.

Hierzu zwei Anmerkungen:

1. Selbst wenn beispielsweise nahe dem Rand der Raumladungszone n nur $n \approx 0.1 \cdot N_D^+ \approx 1 \times 10^{16} \text{ cm}^{-3}$ ist, und analog $p \approx 0,1 \cdot N_A^-$, so kann ∇n (bzw. ∇p) noch sehr hoch sein, weil die Minoritätenkonzentrationen auf Werte von nur etwa $1 \times 10^5 \text{ cm}^{-3}$ fallen. Somit steht diese Vernachlässigung nicht im Widerspruch zu hohen Diffusionsströmen im pn -Übergang.
2. Wenn mit diesem vereinfachenden Ansatz ($n = p = 0$) innerhalb der Raumladungszone der Bandverlauf $W(x)$ in zulässiger Näherung ermittelt worden ist, können nachträglich die Konzentrationen gemäß (2.45) berechnet werden:

$$\begin{aligned} n(x) &= N_L \cdot e^{-(W_L(x)-W_F)/kT} = N_L \cdot e^{-\eta_n(x)/kT} \\ p(x) &= N_V \cdot e^{-(W_F-W_V(x))/kT} = N_V \cdot e^{-\eta_p(x)/kT} \end{aligned}$$

Somit gilt vereinfachend (vgl. Abb. 5.25) und mit den Voraussetzungen (1)-(2):

$$\rho = \begin{cases} +qN_D^+ & \text{für } l_g \leq x \leq l_{n0} \quad \text{n-Seite} \\ -qN_A^- & \text{für } l_{p0} \leq x \leq l_g \quad \text{p-Seite} \end{cases} \quad (5.66)$$

Die Poisson-Gleichung kann nun für die n - bzw. p -Seite getrennt integriert werden, und an der Grenzfläche ($x = l_g$) müssen beide Lösungen einander angepasst werden. Die Lösungen auf beiden Seiten müssen an der Grenze zwischen p und n zusammenpassen. Dazu muss bei $x = l_g$ Stetigkeit

$$W_L^p(l_g) = W_L^n(l_g) \quad (5.75)$$

und stetige Differenzierbarkeit gegeben sein

$$\frac{dW_L^p}{dx}(l_g) = \frac{dW_L^n}{dx}(l_g) \quad (5.76)$$

Aus (5.75) und (5.76) folgt sofort die Ladungsneutralität der Gesamtanordnung

$$qN_A(l_g - l_{p0}) = qN_D(l_{n0} - l_g) \quad (5.77)$$

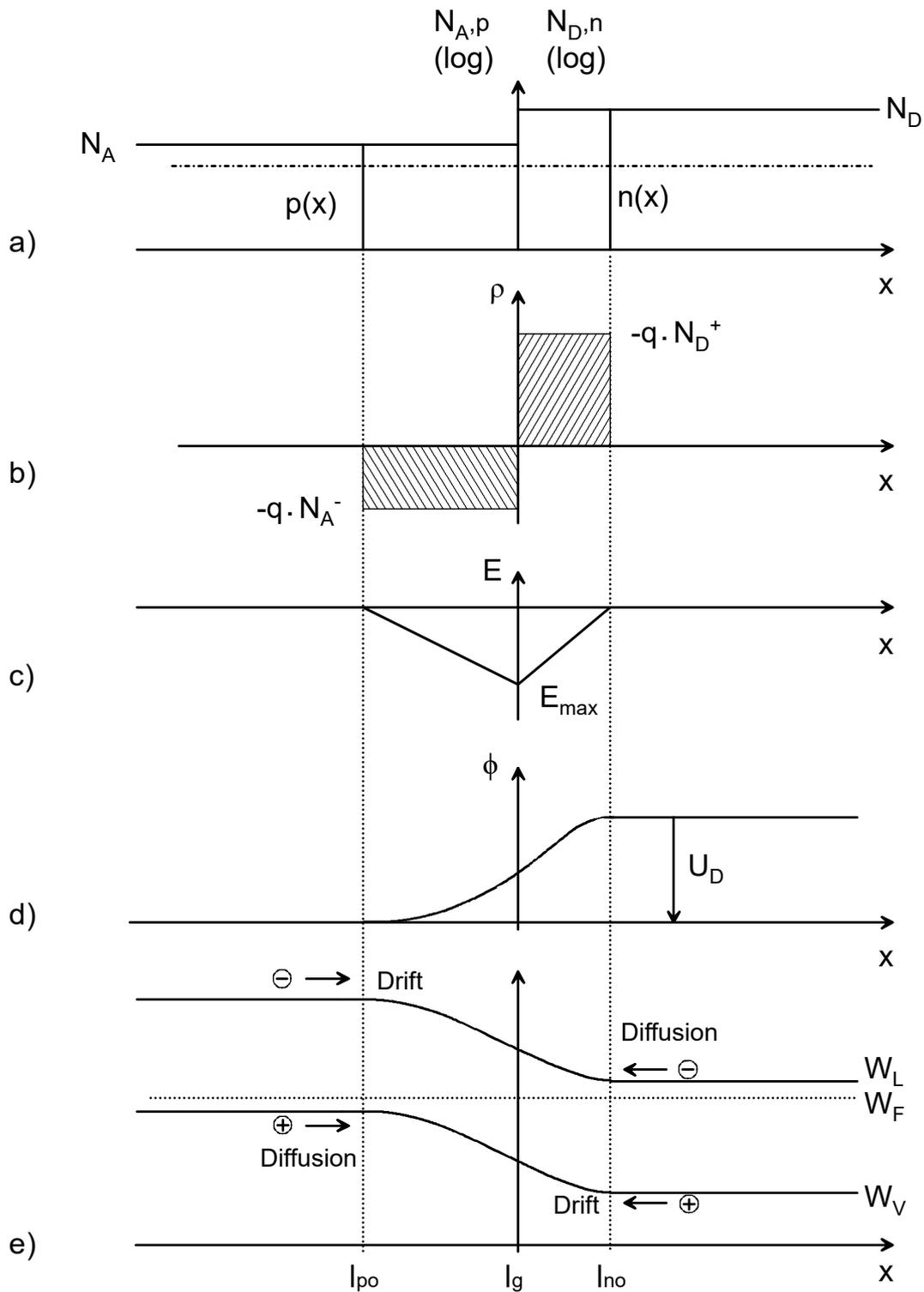


Abbildung 5.25: Abrupter pn -Übergang im stromlosen Zustand, (a) Konzentration der Ladungsträger und Dotierstoffe, (b) Raumladung, (c) Feldverlauf, (d) Potentialverlauf, (e) Bandverlauf

p-Seite	n-Seite
$\frac{d^2\Phi}{dx^2} = -\frac{\rho}{\epsilon_0\epsilon_r} = \frac{qN_A^-}{\epsilon_0\epsilon_r}$ $= -\frac{1}{q} \frac{d^2W_L}{dx^2} \quad (5.67)$	$\frac{d^2\Phi}{dx^2} = -\frac{\rho}{\epsilon_0\epsilon_r} = -\frac{qN_D^+}{\epsilon_0\epsilon_r}$ $= -\frac{1}{q} \frac{d^2W_L}{dx^2} \quad (5.68)$
1. Integration	
$\frac{dW_L^p}{dx} = -\frac{q^2N_A}{\epsilon_0\epsilon_r} \cdot x + \text{const.}$	$\frac{dW_L^n}{dx} = \frac{q^2N_D}{\epsilon_0\epsilon_r} \cdot x + \text{const.}$
mit Randbedingung fürs Feld an äußeren Grenzen der RLZ	
$\frac{dW_L^p}{dx}(l_{p0}) = 0 \quad (5.69)$	$\frac{dW_L^n}{dx}(l_{n0}) = 0 \quad (5.70)$
ergibt für das Feld $qE = -dW/dx$	
$\frac{dW_L^p}{dx} = -\frac{q^2N_A}{\epsilon_0\epsilon_r} \cdot (x - l_{p0})$ (5.71)	$\frac{dW_L^n}{dx} = -\frac{q^2N_D}{\epsilon_0\epsilon_r} \cdot (l_{n0} - x)$ (5.72)
2. Integration	
$W_L^p(x) = -\frac{q^2N_A}{\epsilon_0\epsilon_r} \cdot \int (x - l_{p0}) dx$	$W_L^n(x) = -\frac{q^2N_D}{\epsilon_0\epsilon_r} \cdot \int (l_{n0} - x) dx$
mit Randbedingung für die Bänder an äußeren Grenzen der RLZ	
$W_L^p(0) = W_L^p(l_{p0})$	$W_L^n(l_{n0}) = W_L^n(\infty)$
ergibt für den Bandverlauf $W(x)$	
$W_L^p(x)$ $= W_L^p(0) - \frac{q^2N_A}{\epsilon_0\epsilon_r} \cdot (x - l_{p0})^2$ (5.73)	$W_L^n(x)$ $= W_L^n(\infty) + \frac{q^2N_D}{\epsilon_0\epsilon_r} \cdot (l_{n0} - x)^2$ (5.74)

Tabelle 5.3: Berechnung des Bandverlaufs im pn-Übergang

Aus (5.73), (5.75) und mit (5.60) folgt das eingebaute Potential U_D des pn -Übergangs

$$W_L(0) - W_L(\infty) = qU_D = \frac{q^2}{2\varepsilon_0\varepsilon_r} \left[N_A (l_g - l_{p0})^2 + N_D (l_{n0} - l_g)^2 \right] \quad (5.78)$$

Aus (5.77) und (5.78) können getrennt die n - bzw. p -seitigen Ausdehnungen der Raumladungszone bestimmt werden:

$$l_g - l_{p0} = \sqrt{\frac{2\varepsilon_0\varepsilon_r N_D}{qN_A(N_A + N_D)}} \cdot \sqrt{U_D} \quad (5.79)$$

$$l_{n0} - l_g = \sqrt{\frac{2\varepsilon_0\varepsilon_r N_A}{qN_D(N_A + N_D)}} \cdot \sqrt{U_D} \quad (5.80)$$

und somit ist die Gesamtausdehnung der Raumladungszone

$$l_0 = l_{n0} - l_{p0} = (l_{n0} - l_g) + (l_g - l_{p0}) = \sqrt{\frac{2\varepsilon_0\varepsilon_r}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right)} \cdot \sqrt{U_D} \quad (5.81)$$

Die Ausdehnung der Raumladungszone ist für $U = 0$ (stromloser Fall) schwach abhängig von intrinsischen Halbleitergrößen (ε_r, n_i über U_D), wesentlich stärker aber von den Dotierungskonzentrationen N_A und N_D . Die Gleichung (5.77) zeigt insbesondere, dass für $N_D \neq N_A$ die Ausdehnung in den schwächer dotierten Bereich größer ist als in den höher dotierten. Durch stark unterschiedliche Dotierung kann man erreichen, dass die Raumladungszone fast völlig auf der n - oder p -Seite liegt: Für einen p^+n -Übergang mit $N_A \gg N_D$ ist

$$l_{n0} - l_g \gg l_g - l_{p0}$$

und für einen pn^+ -Übergang mit $N_D \gg N_A$ ist

$$l_g - l_{p0} \gg l_{n0} - l_g$$

Für $N_D = N_A = 1 \times 10^{18} \text{ cm}^{-3}$ beträgt die Gesamtausdehnung für Silizium nur $l_0 = 0.04 \mu\text{m}$. Wenn an den pn -Übergang von außen eine Spannung angelegt wird, verändern sich die Abmessungen der Raumladungszone; dies wird später beim stromdurchflos-

senen pn -Übergang noch behandelt.

Abb. 5.25c gibt den Verlauf des elektrischen Felds im pn -Übergang wieder, das bei $x = l_g$ sein Maximum erreicht. Mit (5.69) lässt sich dieses berechnen zu

$$E_{\max} = \frac{1}{q} \frac{dW_L}{dx}(l_g) = -\frac{qN_A}{\varepsilon_0\varepsilon_r}(l_g - l_{p0}) = -\frac{qN_D}{\varepsilon_0\varepsilon_r}(l_{n0} - l_g) = \frac{U_D}{l_0/2} \quad (5.82)$$

E_{\max} ist daher stark abhängig von der Dotierstoffkonzentration und beträgt für das oben angegebene Beispiel (Si, $N_D = N_A = 1 \times 10^{18} \text{ cm}^{-3}$)

$$E_{\max} = \frac{0.9 \text{ V}}{0.02 \text{ } \mu\text{m}} = 450 \text{ kV/cm}$$

Mit dem Ergebnis von 5.3 ist der vollständige ortsabhängige Verlauf der Leitungsbandkante, und wegen $W_L - W_V = W_g = \text{const.}$, auch der Verlauf der Valenzbandkante bekannt. Wie bereits angekündigt, können jetzt die ortsabhängigen Konzentrationsverläufe der freien Ladungsträger $n(x)$ und $p(x)$ berechnet werden (s. Anmerkung 2). Für die Majoritäten gilt:

$$\begin{aligned} n(x) &= N_L \cdot e^{-(W_L(x) - W_F)/kT} = N_L \cdot e^{-\eta_n(x)/kT} \\ &= N_L \cdot \exp\left(-\frac{W_L(\infty) - W_F}{kT}\right) \cdot \exp\left[-\frac{q^2 N_D (l_{n0} - x)^2}{2\varepsilon_0\varepsilon_r}\right] \end{aligned} \quad (5.83)$$

Unter Verwendung der Debye-Länge (4.13)

$$L_{Dp} = \sqrt{\frac{\varepsilon_0\varepsilon_r}{qN_A^-} \cdot \frac{kT}{q}} = \sqrt{\frac{\varepsilon_0\varepsilon_r}{qN_A^-} \cdot U_{th}} \quad (5.84)$$

und (2.45)

$$n_{n0} = N_D = N_L \cdot e^{-(W_L(\infty) - W_F)/kT}$$

folgt:

$$n_n(x) = N_D \cdot \exp\left[-\frac{1}{2} \left(\frac{l_{n0} - x}{L_{Dp}}\right)^2\right] \quad \text{für } l_g \leq x \leq l_{n0} \quad (5.85)$$

Analog findet man

$$p_p(x) = N_A \cdot \exp \left[-\frac{1}{2} \left(\frac{x - l_{p0}}{L_{Dn}} \right)^2 \right] \quad \text{für } l_{p0} \leq x \leq l_g \quad (5.86)$$

Die Minoritätsträgerkonzentration gewinnt man aus der Bedingung

$$n_n \cdot p_p = n_i^2$$

Somit ist die Elektronenkonzentration (Minoritäten) auf der p -Seite

$$n_p(x) = \frac{n_i^2}{p_p(x)} = \frac{n_i^2}{N_A} \cdot \exp \left[\frac{1}{2} \left(\frac{x - l_{p0}}{L_{Dp}} \right)^2 \right] \quad \text{für } l_{p0} \leq x \leq l_g \quad (5.87)$$

und die Löcherkonzentration auf der n -Seite

$$p_n(x) = \frac{n_i^2}{n_n(x)} = \frac{n_i^2}{N_D} \cdot \exp \left[\frac{1}{2} \left(\frac{l_{n0} - x}{L_{Dn}} \right)^2 \right] \quad \text{für } l_g \leq x \leq l_{n0} \quad (5.88)$$

Unter Verwendung der Diffusionsspannung

$$U_D = U_{th} \ln \left(\frac{N_A N_D}{n_i^2} \right)$$

findet man weiter für die Minoritäten-Elektronen

$$n_p(x) = N_D \cdot e^{-U_D/U_{th}} \cdot \exp \left[\frac{1}{2} \left(\frac{x - l_{p0}}{L_{Dn}} \right)^2 \right] \quad \text{für } l_{p0} \leq x \leq l_g \quad (5.89)$$

und für die Minoritäten-Löcher

$$p_n(x) = N_A \cdot e^{-U_D/U_{th}} \cdot \exp \left[\frac{1}{2} \left(\frac{l_{n0} - x}{L_{Dp}} \right)^2 \right] \quad \text{für } l_g \leq x \leq l_{n0} \quad (5.90)$$

Diskussion der Gleichungen (5.85), (5.86), (5.89), (5.90): Die Konzentrationen werden üblicherweise im logarithmischen Maßstab aufgetragen, in dem die Konzentrationen quadratisch vom Ort abhängen. An den äußeren Grenzen der RLZ ergeben sich die

Minoritäten-Konzentrationen auf der p -Seite bei $x = l_{p0}$

$$n = n_p = N_D \cdot e^{-U_D/U_{th}} = \frac{n_i^2}{N_A} = \frac{n_i^2}{p_{p0}} \quad (5.91)$$

und auf der n -Seite bei $x = l_{n0}$

$$p = p_n = N_A \cdot e^{-U_D/U_{th}} = \frac{n_i^2}{N_D} = \frac{n_i^2}{n_{n0}} \quad (5.92)$$

Für die Majoritäten-Konzentration der Löcher auf p -Seite findet man an der Grenze der RLZ bei $x = l_g$:

$$p = p_p = N_A$$

und für die Majoritäten-Elektronen auf n -Seite bei $x = l_{n0}$

$$n = n_n = N_D$$

Dieses Verhalten ist in Abb. 5.24a bereits vorweggenommen und entsprechend eingezeichnet. Man erkennt ferner, dass die Minoritäten an den Raumladungszonen-Grenzen gegenüber der zugehörigen Dotierstoffkonzentration um den Faktor $e^{-U_D/U_{th}}$ herabgesetzt sind und dass dies in Übereinstimmung mit dem Massenwirkungsgesetz $n \cdot p = n_i^2$ ist.

In diesem Kapitel sind die ortsabhängigen Verläufe von Leitungs- und Valenzband in (5.7) und (5.8) sowie die Konzentrationsverläufe von Elektronen und Löchern (5.85)-(5.90) berechnet worden. Aus den Konzentrations- und Bandverläufen können nun die Diffusions- und Feldströme berechnet werden

Das lokale Feld $E(x) = -\nabla\Phi(x) = 1/q\nabla W(x)$ liefert mit der lokalen Elektronenkonzentration $n(x)$ den Feldstrom am Ort x

$$J_{n,\text{Feld}} = -qn(x)\mu E(x)$$

und den Diffusionsstrom

$$J_{n,\text{Diff}} = qD_n\nabla n$$

Analoge Ausdrücke werden für die Löcher erhalten. Im spannungslosen Fall (thermo-

dynamisches Gleichgewicht) muss der Gesamtstrom gleich Null sein

$$J_{\text{ges}} = 0 = J_{\text{Feld}} + J_{\text{Diff}} = J_{n,\text{Feld}} + J_{n,\text{Diff}} + J_{p,\text{Feld}} + J_{p,\text{Diff}}$$

Genauer gilt aufgrund des Prinzips des detaillierten Gleichgewichts – d.h. der Erhaltung der Elektronen und der Löcher – sogar bei $U = 0$

$$\begin{aligned} J_n &= J_{n,\text{Feld}} + J_{n,\text{Diff}} = 0 \\ J_p &= J_{p,\text{Feld}} + J_{p,\text{Diff}} = 0 \end{aligned} \tag{5.93}$$

weil jeder Einzelprozess (Elektronen und Löcher getrennt) für sich im Gleichgewicht ist. Diese Gleichgewichtsforderung muss für den nun folgenden Fall des stromdurchflossenen p - n Überganges aufgegeben werden.

5.3.3 Strom-Spannungscharakteristik des pn -Übergangs

Durch eine außen angelegte Spannung wird das durch (5.93) beschriebene Gleichgewicht gestört. Es zeigt sich aber, dass diese Störung in den meisten Fällen nur sehr gering ist, obwohl sie den Stromfluss durch den pn -Übergang trägt. Das liegt daran, dass die Feld- bzw. Diffusionsströme für sich genommen aufgrund der hohen Konzentrationen, deren hoher Gradienten sowie des hohen elektrischen Felds entsprechend groß sind. Im Vergleich zu den Einzelkomponenten Feld- oder Diffusionsstrom ist aber der Gesamtstrom als Differenz beider Größen sehr klein. Grundsätzlich liegen daher die gleichen Verhältnisse wie im stromlosen Fall vor, und es gilt für nicht zu große, außen anliegende Spannungen $U \lesssim U_D$

$$\begin{aligned} J_n &= J_{n,\text{Feld}} + J_{n,\text{Diff}} \approx 0 \\ J_p &= J_{p,\text{Feld}} + J_{p,\text{Diff}} \approx 0 \end{aligned} \tag{5.94}$$

Zur Herleitung der Strom-Spannungscharakteristik wird angenommen, dass die gesamte außen angelegte Spannung über der Raumladungszone abfällt. Nach der obigen Darstellung ist dann in den Gleichungen für den stromlosen Fall lediglich die Größe U_D durch $U_D - U$ zu ersetzen.

Insbesondere gilt für die Ausdehnung der Raumladungszone

$$\begin{aligned}
 l_g - l_p &= \sqrt{\frac{2\varepsilon_0\varepsilon_r N_D}{qN_A(N_A + N_D)}} \cdot \sqrt{U_D - U} \\
 l_n - l_g &= \sqrt{\frac{2\varepsilon_0\varepsilon_r N_A}{qN_D(N_A + N_D)}} \cdot \sqrt{U_D - U} \\
 l &= l_n - l_p = (l_n - l_g) + (l_g - l_p) = \sqrt{\frac{2\varepsilon_0\varepsilon_r}{q} \left(\frac{1}{N_A} + \frac{1}{N_D} \right)} \cdot \sqrt{U_D - U}
 \end{aligned} \tag{5.95}$$

Man erkennt, dass je nach Vorzeichen von U die Ausdehnung der Raumladungszone vergrößert oder verringert wird: für positives U (d.h. positive Spannung am p -Gebiet) verringert und für negative U vergrößert sich die an freien Ladungsträgern verarmte Zone. Diese stellt für den elektrischen Strom einen großen Widerstand dar, der mit der Ausdehnung wächst, so dass für negative U die Sperrrichtung und für positive U die Durchlassrichtung des pn -Überganges vorliegt.

Mit der Verschiebung der Raumladungszonengrenzen verschieben sich auch die Majoritätenkonzentrationen, die aber nach wie vor an den (jetzt neuen) Grenzen nahezu ihre Gleichgewichtskonzentration n_{n0} bzw. p_{p0} annehmen (vgl. Abb. 5.26a und 5.27a). Dagegen erreichen die Minoritäten ihre Gleichgewichtskonzentrationen nicht an den Raumladungszonengrenzen, sondern erst tiefer in den Bahngebieten, da nach (5.89) und (5.90) gilt

$$\begin{aligned}
 n_p(x) &= N_D \cdot e^{(-U_D+U)/U_{th}} \cdot \exp \left[\frac{1}{2} \left(\frac{x - l_p}{L_{Dp}} \right)^2 \right] \quad \text{für } l_p \leq x \leq l_g \\
 p_n(x) &= N_A \cdot e^{(-U_D+U)/U_{th}} \cdot \exp \left[\frac{1}{2} \left(\frac{l_n - x}{L_{Dn}} \right)^2 \right] \quad \text{für } l_g \leq x \leq l_n
 \end{aligned} \tag{5.96}$$

An den Grenzen der RLZ $x = l_p$ bzw. $x = l_n$ sind die Minoritätenkonzentrationen daher

$$\begin{aligned}
 n_p(l_p) &= N_D \cdot e^{-U_D/U_{th}} \cdot e^{U/U_{th}} = n_{p0} \cdot e^{U/U_{th}} \\
 p_n(l_n) &= N_A \cdot e^{-U_D/U_{th}} \cdot e^{U/U_{th}} = p_{n0} \cdot e^{U/U_{th}}
 \end{aligned} \tag{5.97}$$

Die Minoritätenkonzentrationen sind also an den Raumladungszonengrenzen gegenüber dem stromlosen Fall je nach Vorzeichen von U um den Faktor $\exp U/U_{th}$ angehoben

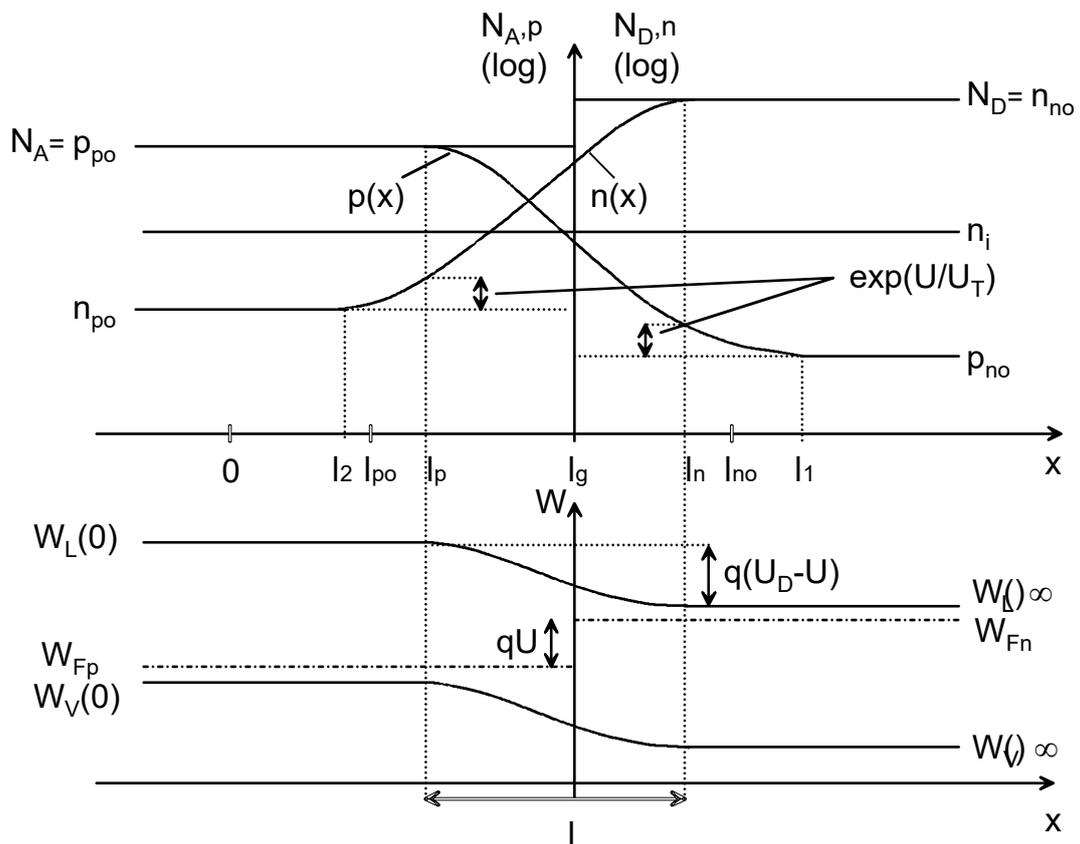


Abbildung 5.26: Stromdurchflossener pn -Übergang ($U > 0$); Ladungsträgerverteilung und Bändermodell

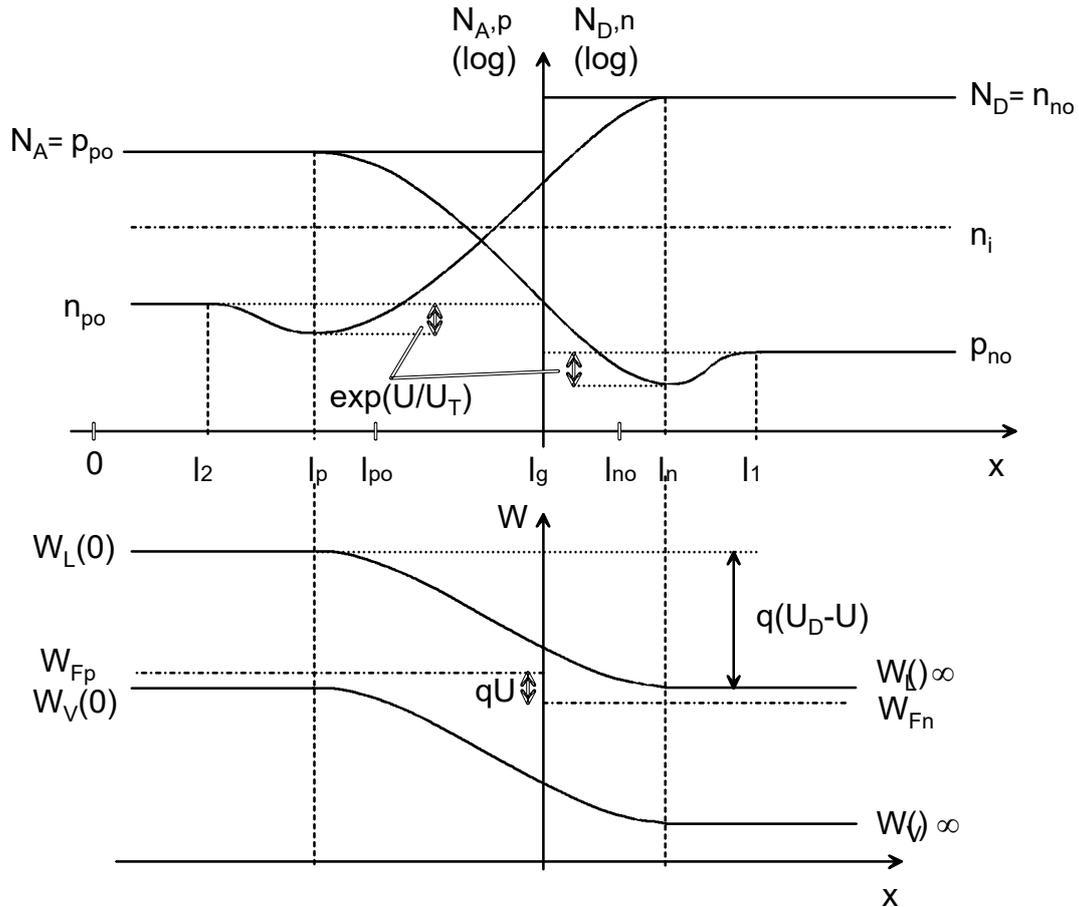


Abbildung 5.27: Stromdurchflossener pn -Übergang ($U < 0$); Ladungsträgerverteilung und Bändermodell

oder abgesenkt (vgl. Abb. 5.26 und 5.27). Wegen $n_{n0} \gg p_n(l_n)$ und $p_{p0} \gg n_p(l_p)$ bleiben jedoch, wie bereits erwähnt, die Majoritätenkonzentrationen nahezu unverändert. Die außen angelegte Spannung U tritt (mit der Elementarladung q multipliziert) als Unterschied der Fermi-Energie im n - bzw. p -Gebiet auf (vgl. Schottky-Kontakt).

Man erkennt, wie durch U die Bandverbiegung und damit auch das elektrische Feld erhöht (Sperrichtung) oder erniedrigt (Durchlassrichtung) wird.

Berechnung der Diffusionsanteile der Stromdichte

Diese sei exemplarisch für den Löcherstromanteil $J_{pn}(x)$ auf der n -Seite durchgeführt.

Es gilt:

$$J_{pn}(x) = -q\mu_p U_{th} \frac{dp_n(x)}{dx} \quad \text{für } x \geq l_n \quad (5.98)$$

Ferner gilt die Kontinuitätsgleichung:

$$\frac{dp_n(x)}{dt} = -\frac{1}{q} \frac{dJ_{pn}(x)}{dx} - r_p + g_p \quad (5.99)$$

Für den stationären Fall ist:

$$\frac{dp_n(x)}{dt} = 0 \quad (5.100)$$

Außerdem ist $g_p = 0$, da eine Lochgeneration z.B. durch Anregung mit Licht außer acht gelassen werden soll. Für die Rekombinationsrate wird der Ansatz:

$$r_p = \frac{p_n(x) - p_{n0}}{\tau_p} \quad (5.101)$$

wie er bereits bei der Behandlung der Photoleitfähigkeit in Abschnitt 3.3 verwendet wurde. Damit wird (5.99):

$$\frac{dJ_{pn}(x)}{dx} = -q \frac{p_n(x) - p_{n0}}{\tau_p} \quad (5.102)$$

Die Ableitung von (5.98) wird in (5.102) eingesetzt. Somit folgt die Differentialgleichung für die Minoritätenkonzentration:

$$\frac{d^2 p_n(x)}{dx^2} = \frac{1}{L_p^2} (p_n(x) - p_{n0}) \quad (5.103)$$

mit

$$L_p^2 = \tau_p \mu_p U_{th} \quad (5.104)$$

als Diffusionslänge der Minoritäten (Löcher) im n -Gebiet. Analog gilt für die Elektronenkonzentration im p -Gebiet:

$$d^2 n_p x dx^2 = \frac{1}{L_n^2} (n_p(x) - n_{p0}) \quad (5.105)$$

Die allgemeine Lösung der Differentialgleichungen, hier von (5.103), hat die Form:

$$p_n(x) - p_{n0} = A \cdot e^{x/L_p} + B \cdot e^{-x/L_p} \quad (5.106)$$

Als Randbedingungen gelten:

$$p_n(l_n) = p_{n0} \cdot e^{U/U_{th}} \quad (5.107)$$

und

$$p_n(l_1) = p_{n0} \quad (5.108)$$

(vgl. Abb. 5.26 und 5.27). Diese legen die Konstanten A und B fest, so dass folgt:

$$p_n(x) - p_{n0} = p_{n0} \cdot \left(e^{U/U_{th}} - 1 \right) \frac{\sinh\left(\frac{l_1-x}{L_p}\right)}{\sinh\left(\frac{l_1-l_n}{L_p}\right)} \quad (5.109)$$

Nach Differentiation kann $dp_n(x)/dx$ in (5.98) eingesetzt werden mit dem Ergebnis:

$$J_{pn}(x) = \frac{q\mu_p U_{th}}{L_p} \cdot p_{n0} \cdot \left(e^{U/U_{th}} - 1 \right) \frac{\cosh\left(\frac{l_1-x}{L_p}\right)}{\sinh\left(\frac{l_1-l_n}{L_p}\right)} \quad (5.110)$$

Analog findet man für

$$J_{np}(x) = \frac{q\mu_n U_{th}}{L_p} \cdot n_{p0} \cdot \left(e^{U/U_{th}} - 1 \right) \frac{\cosh\left(\frac{x-l_2}{L_n}\right)}{\sinh\left(\frac{l_1-l_2}{L_n}\right)} \quad (5.111)$$

Die Minoritätenausläufer in die Bahngebiete hinein werden Diffusionsschwänze genannt. Durch folgende Näherungen und darauf aufbauenden Überlegungen kann man zeigen, dass der Strom durch den pn -Übergang von den gerade berechneten Diffusionsstromanteilen der Minoritäten im Bereich der Diffusionsschwänze getragen wird:

1. Keine Rekombination in der Raumladungszone (Shockley-Näherung). Dies ist gleichbedeutend mit der Bedingung

$$L(n, p) \gg l_n - l_p \quad (5.112)$$

2. Kein Spannungsabfall in den Bahngebieten, das heißt, dass an den Raumladungszonengrenzen und in den Bahngebieten keine Feldströme fließen. Real müssen aber Felder in den Bahngebieten existieren, da hier Diffusionsströme mangels Diffusionsgradienten nicht auftreten und somit die Kontinuitätsgleichungen verletzt wären.

Mit diesen Annahmen sind die Minoritätenkonzentrationsgradienten innerhalb der Raumladungszone in etwa konstant:

$$\frac{dp_n(x)}{dx} \approx \text{const.} \quad \text{und} \quad \frac{dn_p(x)}{dx} \approx \text{const.} \quad \text{für } l_p \leq x \leq l_n$$

Diese Konstanz überträgt sich auf die Diffusionsströme:

$$J_{pn}(l_n) \approx J_{pn}(l_g) \quad \text{und} \quad J_{np}(l_n) \approx J_{np}(l_g) \quad (5.113)$$

Da wegen der Annahme 2 an den Grenzen der Raumladungszone die Feldströme verschwinden, die Diffusionsströme also die einzig verbleibenden Anteile darstellen, kann mit Anwendung der Kontinuität des Stromes die Gesamtstromdichte folgendermaßen bestimmt werden:

Man berechnet den Diffusionslöcherstrom an der Stelle $x = l_n$ und den Diffusionselektronenstrom an der Stelle $x = l_p$; anschließend addiert man die Ströme bei $x = l_g$, was nach (5.93) unter Verwendung der bei l_n bzw. l_p errechneten Ergebnisse leicht möglich ist.

Es folgt die Stromdichte-Spannungskennlinie des pn -Überganges:

$$J = J_0 \cdot \left(e^{U/U_{th}} - 1 \right) \quad (5.114)$$

mit

$$J_0 = qU_{th} \cdot \left[\frac{\mu_p}{L_p} \cdot p_{n0} \cdot \coth \left(\frac{l_1 - l_n}{L_p} \right) + \frac{\mu_n}{L_n} \cdot n_{p0} \cdot \coth \left(\frac{l_p - l_2}{L_n} \right) \right] \quad (5.115)$$

Mit Hilfe von

$$p_{n0} = \frac{n_i^2}{N_D} \quad \text{und} \quad n_{p0} = \frac{n_i^2}{N_A}$$

gilt auch

$$J_0 = qU_{th}n_i^2 \cdot \left[\frac{\mu_p}{L_p N_D} \coth\left(\frac{l_1 - l_n}{L_p}\right) + \frac{\mu_n}{L_n N_A} \coth\left(\frac{l_p - l_2}{L_n}\right) \right] \quad (5.116)$$

J_0 enthält dann die technologisch einstellbaren Dotierstoffkonzentrationen N_A und N_D . Die Form von (5.114) ist wie bereits bei der Schottky-Diode die einer Diode mit Durchlass- und Sperrichtung:

$$J = J_0 \cdot \left(e^{U/U_{th}} - 1 \right) \quad \text{für } U > 0 \text{ (Durchlassrichtung)} \quad (5.117)$$

bzw.

$$J = J_0 \cdot e^{U/U_{th}} \quad \text{für } U \gg U_{th} \quad (5.118)$$

und

$$J = -J_0 \quad \text{für } U < 0 \text{ (Sperrichtung)} \quad (5.119)$$

Die Kennlinien haben die gleiche Form wie die der Schottky-Diode (vgl. Abb. 5.18 und 5.19). Für den Einsatz in Transistoren und Dioden ist eine Fallunterscheidung wichtig, die eine Vereinfachung und Spezialisierung des Ausdrucks für J_0 liefert.

1. Langbasisdiode: In diesem Fall können auch die Minoritäten in der pn -Struktur ihren Gleichgewichtswert annehmen. Dann wird vereinfachend angenommen:

$$l_1 - l_n \gg L_p \quad \text{und} \quad l_p - l_2 \gg L_n$$

Wegen $\coth x \approx 1$ für $x \gg 1$ folgt für J_0 der vereinfachte Ausdruck aus (5.116):

$$J_{0,l} = qU_{th}n_i^2 \cdot \left[\frac{\mu_p}{L_p N_D} + \frac{\mu_n}{L_n N_A} \right] \quad (5.120)$$

2. Kurzbasisdiode, in diesem Fall gilt umgekehrt:

$$l_1 - l_n \ll L_p \quad \text{und} \quad l_p - l_2 \ll L_n$$

Mit $\coth(x) \approx 1/x$ für $x \ll 1$ gilt nun

$$J_{0,k} = qU_{th}n_i^2 \left[\frac{\mu_p}{(l_1 - l_n)N_D} + \frac{\mu_n}{(l_p - l_2)N_A} \right] \quad (5.121)$$

Die jeweils kürzere der beiden Längen $\{L_n, L_p\}$ bzw. $\{(l_p - l_2), (l_1 - l_n)\}$ bestimmt also in beiden Fällen den Sättigungsstrom J_0 (vgl. (5.120) und (5.121)). Eine weitere Vereinfachung folgt für stark unsymmetrische pn -Übergänge mit entweder $N_A \gg N_D$ (p^+n -Übergang) bzw. $N_A \ll N_D$ (n^+p -Übergang)

$$J_{0,l} = qU_{th}n_i^2 \frac{\mu_p}{L_p N_D} \quad \text{für } p^+n \quad (5.122)$$

$$J_{0,l} = qU_{th}n_i^2 \frac{\mu_n}{L_n \cdot N_A} \quad \text{für } n^+p \quad (5.123)$$

$$J_{0,k} = qU_{th}n_i^2 \frac{\mu_p}{(l_1 - l_n)N_D} \quad \text{für } p^+n \quad (5.124)$$

$$J_{0,k} = qU_{th}n_i^2 \frac{\mu_n}{(l_p - l_2)N_A} \quad \text{für } n^+p \quad (5.125)$$

Der Fall der Kurzbasisdiode gewinnt für den Bipolartransistor tragende Bedeutung. Hier liegen zwei pn -Übergänge so dicht benachbart, dass die Minoritätenkonzentrationen des einen pn -Überganges wesentlich von den Diffusionsausläufern des anderen beeinflusst werden.

Die Sperrströme J_0 sind über die Größe n_i^2 stark temperaturabhängig. In den Durchlassströmen macht sich zusätzlich noch die Temperaturabhängigkeit über U_{th} mit

$$J = J_0 \cdot \left(e^{U/U_{th}} - 1 \right)$$

bemerkbar.

Sperrschichtkapazität des pn -Übergangs

Ähnlich wie der Schottky-Kontakt besitzt der pn -Übergang eine Raumladungs- oder Sperrschichtkapazität, bei der die Raumladungszone von freien Ladungsträgern verarmt ist und das Dielektrikum bildet. Unter Verwendung von (5.95) gilt:

$$c = \frac{\varepsilon_0 \varepsilon_r}{l} = \sqrt{\frac{q \varepsilon_0 \varepsilon_r N_A N_D}{2(N_A + N_D)}} \cdot \frac{1}{\sqrt{U_D - U}} \quad (5.126)$$

Mit steigender Sperrspannung wächst die Ausdehnung der Raumladungszone, somit fällt die Kapazität. In Flussrichtung ist zusätzlich zur Sperrschichtkapazität noch die Diffusionskapazität zu berücksichtigen. Letztere hat folgende Ursachen: Spannungsänderungen verändern nicht nur die Ausdehnung der Raumladungszone, sondern auch die Minoritätenkonzentrationen an deren Grenzen. Der damit verbundene Transportprozess ist die Diffusion, die aber nicht momentan, sondern zeitversetzt der angelegten Spannung folgt. Dieses Verhalten lässt sich als Kapazität beschreiben.

5.3.4 Halbleiter-Halbleitergrenzflächen zwischen unterschiedlichen Halbleiterwerkstoffen; Heteroübergänge

Übergänge zwischen zwei unterschiedlichen Halbleiterwerkstoffen werden Heteroübergänge genannt. Üblicherweise haben beide Werkstoffe unterschiedliche Bandlücken. Diese Strukturen erweitern die Realisierungsmöglichkeiten und die Konzeptionierung von Bauelementen außerordentlich. Beispielsweise können mit ihrer Hilfe die Kräfte auf Elektronen und Löcher weitgehend unabhängig voneinander eingestellt werden, während sie in Homo- pn -Übergängen grundsätzlich entgegengesetzt gleich groß sind (vgl. Abb. 5.28). In Abb. 5.28b ändert sich die Bandlücke kontinuierlich mit dem Ort x , dies kann durch eine entsprechende kontinuierliche Veränderung der Materialzusammensetzung erreicht werden. Technologisch gestalten sich solche Strukturen schwierig, weil mit wechselnder Zusammensetzung i.A. auch der Gitterabstand der Atome verändert wird. Für diese Veränderung der Gitterkonstante besteht aber nur ein sehr enger Spielraum, da damit starke mechanische Verspannungen einhergehen, die schließlich den regelmäßigen Kristallaufbau und damit auch die elektrischen und optischen Eigenschaften des Halbleitermaterials zerstören können. In der Realität muss man daher statt

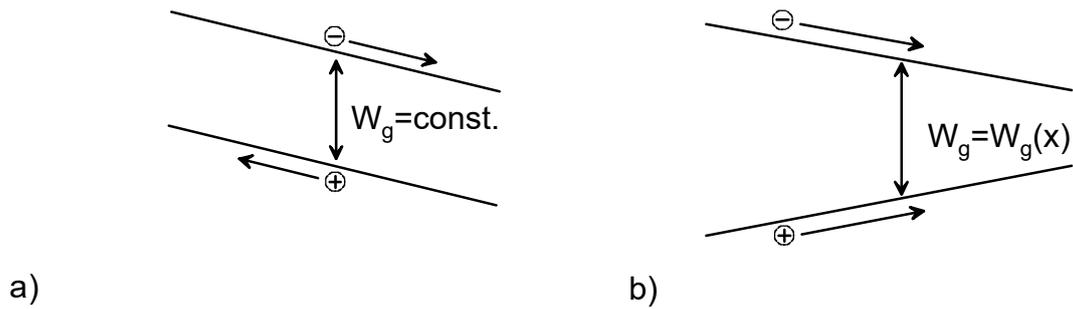


Abbildung 5.28: Ladungstransport in (a) Homo- und (b) Heteroübergängen mit variierender Bandlücke W_g

binärer chemischer Verbindungen solche aus vier verschiedenen Komponenten, „quaternäre“ Materialien verwenden. Mit den zwei zusätzlichen Komponenten gelingt es in ausgewählten Fällen, die zwei Materialeigenschaften Bandlücke und Gitterkonstante für bestimmte Wertebereiche unabhängig voneinander einzustellen.

Bändermodell von Heteroübergängen

Im folgenden wird davon ausgegangen, dass die Gitterkonstanten beider Halbleiterwerkstoffe gleich sind und der Übergang abrupt ist. Ferner treten keine Grenzflächenladungen auf und die Dotierungen im jeweiligen Halbleiter sind konstant. Gegenüber den Homoübergängen gibt es bei den Heteroübergängen eine Vielzahl zusätzlicher Kombinationsmöglichkeiten.

Zunächst muss man unterscheiden, ob jeweils die Leitungs- und Valenzbandkanten in entgegengesetzter (Typ I) oder gleicher Richtung (Typ II) verschoben sind (Abb. 5.29). Dies hängt ab von den Elektronenaffinitäten χ_1 und χ_2 sowie von den Bandabständen beider Halbleiter (vgl. Abb. 5.30a).

Weiterhin kann der Übergang aus Materialien vom gleichen Leitungstyp (nn) oder (pp) gebildet werden, wobei sich jedoch die Materialien beiderseits des Übergangs durch die Bandlücke unterscheiden. Möglich ist auch (pn). Manchmal wird das Material mit der größeren Bandlücke mit einem Großbuchstaben verdeutlicht, z.B. (Npn). In Abb. 5.30 wurde der Typ I Heteroübergang zugrunde gelegt, weil er eine größere technische Be-

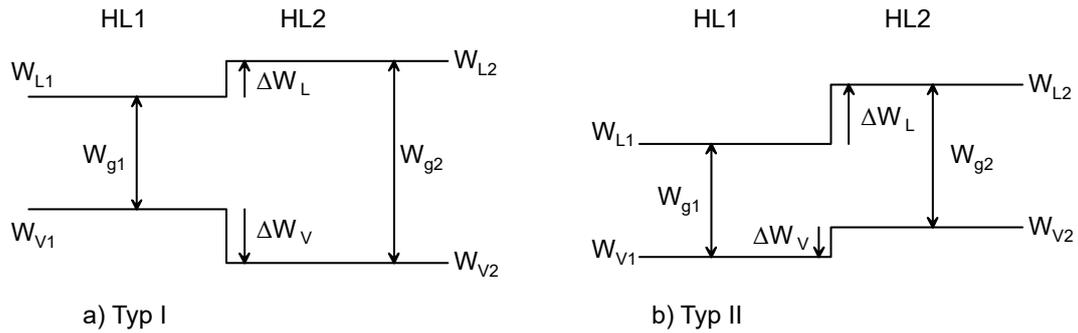


Abbildung 5.29: Abrupte Typ I (a) und Typ II (b) Heteroübergänge (schematisch)

deutung besitzt. Aus Abb. 5.30 liest man ab

$$\Delta W_L = q(\chi_1 - \chi_2) = q\Delta\chi \quad (5.127)$$

$$\Delta W_V = W_{g1} - W_{g2} - q(\chi_1 - \chi_2) = \Delta W_g - q\Delta\chi \quad (5.128)$$

$$\Delta W_g = \Delta W_L - \Delta W_V \quad (5.129)$$

mit

$$\Delta W_g = W_{g1} - W_{g2}$$

$$\Delta W_L = W_{L1} - W_{L2}$$

$$\Delta W_V = W_{V1} - W_{V2}$$

Man erkennt, dass schon beim nn- sowie beim pp-Übergang Raumladungszonen auftreten, weil für den Ausgleich der Banddiskontinuitäten ΔW_L , ΔW_V Bandverbiegungen erforderlich sind. Dies gilt erst recht bei Hetero-*pn*-Übergängen. In allen Fällen treten daher bei Anlegen einer äußeren Spannung Diodenkennlinien auf, da je nach Polung die Ausdehnung der Raumladungszone entweder vergrößert oder reduziert wird. Der relative Anteil der Elektronen bzw. Löcher am Gesamtstrom hängt jetzt aber nicht nur von der Dotierung (und der Beweglichkeit) im jeweiligen Halbleiter, sondern ganz entscheidend von den Banddiskontinuitäten ΔW_L und ΔW_V ab, da diese die wesentlichen Barrieren für die jeweiligen Ladungsträger darstellen. Auf die exakte Berechnung der Bandkantenenergien $W_L(x)$, $W_V(x)$ sowie der Konzentrationen $n(x)$, $p(x)$ und den dar-

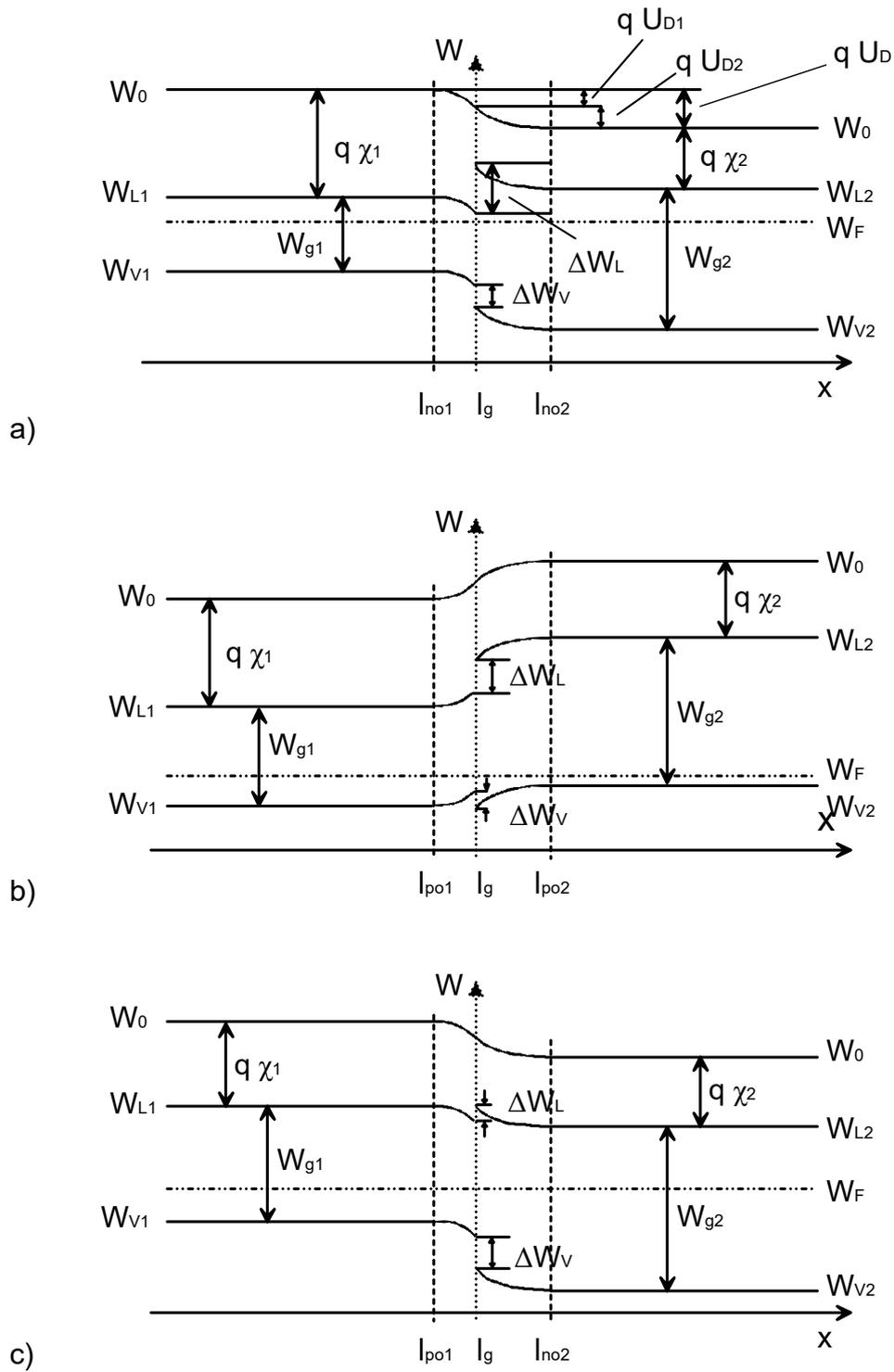


Abbildung 5.30: (a) *nn*-, (b) *pp*- und (c) *pn*-Heteroübergang; (zugrunde gelegt wurde der Typ I Heteroübergang) Eingezeichnet wurde zusätzlich W_0 , um die unterschiedlichen Elektronenaffinitäten beider Halbleiter und die beiden Diffusionsspannungen zu verdeutlichen

aus berechenbaren Strömen wird hier verzichtet. Vielmehr sollen schematisch einige Anwendungsbeispiele beschrieben werden.

5.3.5 Anwendungsbeispiele

Heterostruktur-Feldeffekttransistor

Beim Heterostruktur-Feldeffekttransistor (HFET), auch mit High-Electron-Mobility Transistor (HEMT) bezeichnet, kommt üblicherweise ein Typ $I-nn$ - oder np -Heteroübergang zum Einsatz (Abb. 5.31a). Bekannte Materialkombinationen für diese Anwendung sind AlGa_N/Ga_N und AlGaAs/GaAs, wobei die Al-haltige Schicht jeweils einen höheren Bandabstand besitzt.

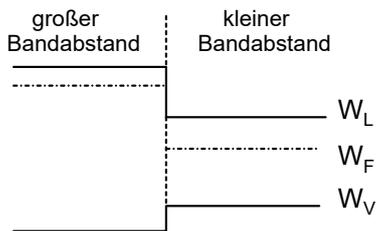
Das AlGaAs ist bis auf eine sehr dünne Schicht unmittelbar an der Heterogrenzfläche (Spacer-Schicht) hoch n -dotiert, während die GaAs-Schicht undotiert ist (Abb. 5.31b). Je nach Schichtwachstumsprozess ist diese Schicht, durch unvermeidbare Verunreinigungen, entweder schwach n - oder p -dotiert. Durch die eingebauten Felder ($W_F = \text{const.}$) entsteht an der Heterogrenzfläche durch den Verlauf der Leitungsbandkante ein nahezu dreieckförmiger Potentialtopf für Elektronen, in dem gemäß Kap. 1.6 nur diskrete Energieniveaus erlaubt sind. In der Praxis werden nur ein oder höchstens zwei Niveaus mit Elektronen besetzt. Die Elektronen sammeln sich hier, weil sie bestrebt sind, energetisch tiefliegende Zustände zu besetzen. Aus Gründen der Ladungsneutralität muss diese negative Elektronenladung durch eine positive Ladung der Donatorrümpfe im AlGaAs kompensiert werden.

Vereinfachend kann man sich vorstellen, dass die Elektronen direkt von den Donatoren im AlGaAs in den Potentialtopf an der Heterogrenzfläche übergehen ("real space transfer"). Der Vorteil gegenüber Standard-Feldeffekttransistoren besteht in der räumlichen Trennung von Donatoren (im AlGaAs) und Elektronen (im GaAs), denn die elektronischen Transporteigenschaften (Beweglichkeiten, Geschwindigkeit) sind in undotierten Materialien weit besser als in dotierten (vgl. Kap. 2.4). In Volumenhalbleitern ist andererseits bei undotierten Materialien die Ladungsträgerkonzentration nur sehr gering. Erst die Heterostruktur gestattet es, hohe Ladungsträgerkonzentrationen und exzellente Transporteigenschaften gleichzeitig bereitzustellen.

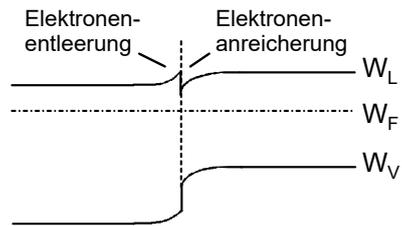
Da der Potentialtopf extrem schmal ist, spricht man auch von Ladungsträgertransport in einem zweidimensionalen Elektronengas (2-DEG). Transistoren, die den Stromtransport im 2-DEG ausnutzen, heißen HFET (s.o.). Die Stromsteuerung erfolgt durch einen

Halbleiter...

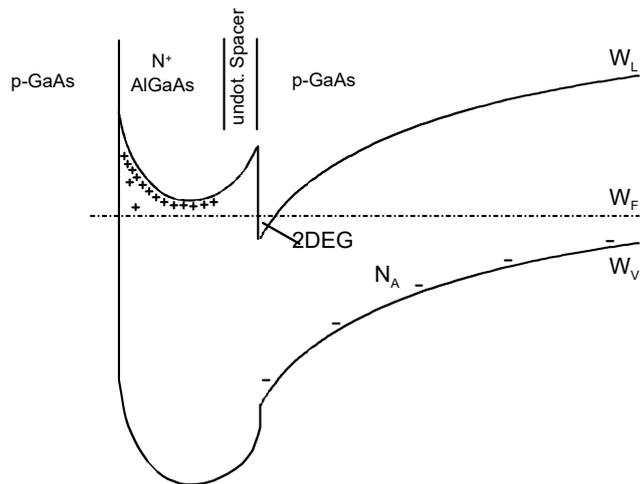
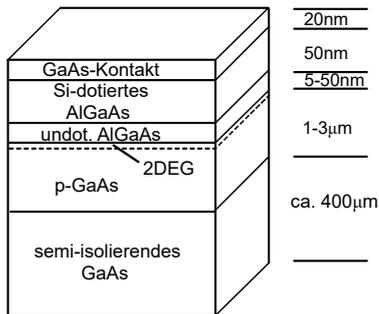
1.) vor Einstellung des therm. Gleichgewichts



2.) nach Einstellung des therm. Gleichgewichts



a)



b)

Abbildung 5.31: (a) Entstehung des Potentialtopfes und (b) Schichtaufbau sowie Bandstruktur eines HFET

Schottky-Kontakt, mit dessen Hilfe die Ladungsträgerdichte im 2-DEG verändert werden kann. Derartige Transistoren, aufgebaut aus einer InAlAs/InGaAs Heterostruktur, besitzen die zur Zeit besten verfügbaren Hochfrequenzeigenschaften mit Grenzfrequenzen von $f_{\max} > 1000$ GHz!

Heterostruktur-Bipolartransistor (HBT)

Der "Wide-Gap-Emitter" findet Verwendung in Heterostruktur-Bipolartransistoren. Er besteht wieder aus einer Typ I-Heterostruktur, in der jetzt beide Halbleiter eine hohe, aber vom Typ her unterschiedliche Dotierung besitzen. Abb. 5.32 zeigt die Bänderstruktur eines HBT, in der der linksseitige np^+ -Übergang eine Wide-Gap-Emitter-Struktur darstellt.

Dieser Hetero- np^+ -Übergang besitzt gegenüber einem Homoübergang folgende Eigenschaften: trotz höherer p -Dotierung des Materials mit kleinerem Bandabstand (z.B. GaAs) im Vergleich zur n -Dotierung des Materials großer Bandlücke (z.B. AlGaAs) fließen bei Anlegen einer äußeren Spannung mehr Elektronen als Löcher über den Hetero- pn -Übergang.

Dies liegt an der gegenüber einem Homo- pn -Übergang höheren Barriere ΔW_V für die Löcher im Valenzband, wie in Abb. 5.32 dargestellt. Es werden also sehr effizient Elektronen aus dem "Wide Gap"-Material emittiert, ohne dass ein nennenswerter Löcherstrom in entgegengesetzter Richtung fließt. Ein weiterer pn -Homo- oder auch Heteroübergang, angefügt auf der rechten Seite, vervollständigt die Npn -Bipolar-Transistorstruktur.

Der Vorteil gegenüber dem konventionellen Bipolartransistor ohne Heterostruktur liegt in dem hohen Verhältnis von Elektronen- zu Löcherstrom, das unmittelbar die erreichbare Verstärkung beeinflusst. Die dünne Basisschicht kann sehr hoch p -dotiert sein, so dass ein geringer Basisbahnwiderstand erreicht werden kann, der aufgrund reduzierte extrinsischer RC -Zeitkonstante zu verbessertem Hochfrequenzverhalten führt. Genauer wird der Bipolar- und Heterostruktur-Bipolartransistor in der Vorlesung "Höchstfrequenz- und THz-Halbleitertechnologien" besprochen.

Doppelheterostruktur-Laser

Abb. 5.33 zeigt die Schichtstrukturen (a), das Bänderdiagramm (b) und das Brechzahlprofil (c) eines Doppelheterostruktur-Laser (DH-Laser). Im Beispiel wurde das InP/InGaAsP-Materialsystem verwendet, das die Lichterzeugung im für die optische Nachrichtentechnik wichtigen spektralen Bereich von $\lambda = 1 \mu\text{m}$ bis $1.55 \mu\text{m}$ gestattet.

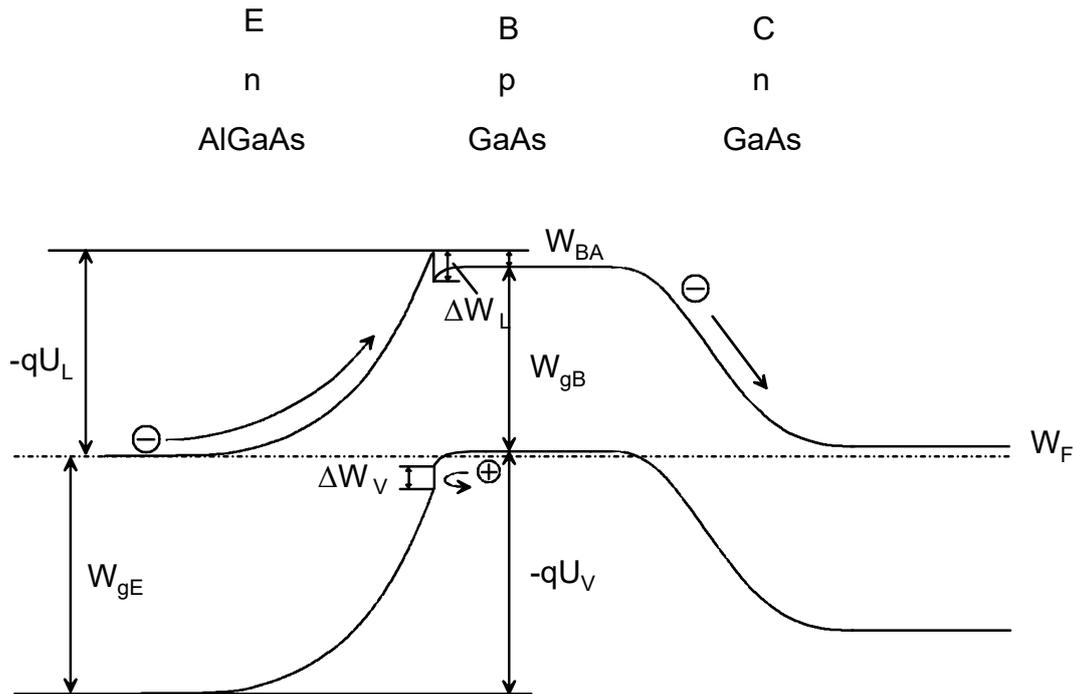


Abbildung 5.32: Bänderdiagramm eines Heterostruktur-Bipolartransistors (HBT)

Die aktive Zone, in der die Lichterzeugung durch strahlende Rekombination von Elektronen und Löchern stattfindet, besteht aus dem Material geringeren Bandabstandes (InGaAsP), das beidseitig von InP mit großem Bandabstand umgeben ist, aber auf der einen Seite p - und auf der anderen n -dotiert ist.

Bei Anlegen einer Flussspannung werden sowohl Elektronen als auch Löcher in das InGaAsP getrieben und wegen der Barrieren an den Heteroübergängen (vgl. Abb. 5.33b) am Abfließen in den jeweils anderen InP-Bereich gehindert. Dies führt im Vergleich zu Homo- pn -Übergängen zu sehr hoher Elektronen- und Löcherkonzentration im InGaAsP-Bereich und entsprechend intensiver Rekombinations-Strahlung. Die so entstehende Lichtwelle wird darüberhinaus durch die unterschiedlichen optischen Brechzahlen n_a und n_s im aktiven Lasermedium (InGaAsP) geführt und sorgt so für eine effiziente stimulierte Emission (vgl. Kap. 3.4), sodass ausgezeichnete Lasereigenschaften resultieren.

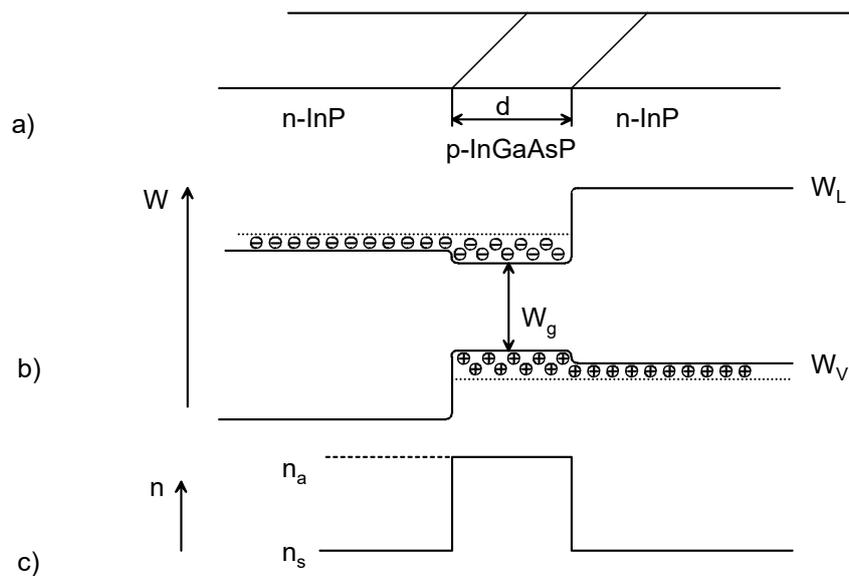


Abbildung 5.33: Doppel-Hetero-Laserstruktur: (a) Schicht- und (b) Bandstruktur sowie (c) Brechzahlprofil

Übergitter-Heterostrukturen

Durch periodische Anordnung von Heteroschichtstrukturen entstehen sogenannte Übergitterstrukturen, die schichtweise (zweidimensionale) Potentialverhältnisse erzeugen, wie sie bereits bei der Behandlung des Festkörpers mit Hilfe des Kronig-Penney-Modells behandelt wurden. Man erzeugt quasi einen künstlichen Kristall im Festkörper, und entsprechend bildet sich eine "Minibandstruktur" in der Bandstruktur, wie in Abb. 5.34 dargestellt.

Durch Auswahl der Materialkombination und verwendeter Schichtdicken können die Miniband-Struktur und somit die Eigenschaften dieses künstlichen Festkörpers gezielt eingestellt und auf bestimmte Anwendungen maßgeschneidert werden. Dies gilt sowohl für elektronische als auch optische Eigenschaften.

Diese Strukturen finden Anwendung in Quantenkaskadenlasern (QCL), die zur Erzeugung von Infrarotstrahlung definierter Wellenlänge eingesetzt werden. Die Miniband-Struktur erlaubt langwellige (niederenergetische) Übergänge bei gleichzeitig hoher Besetzungsinversion, dies kann mit einfachen Band-Band-Übergängen nur bei sehr tiefen Temperaturen erreicht werden.

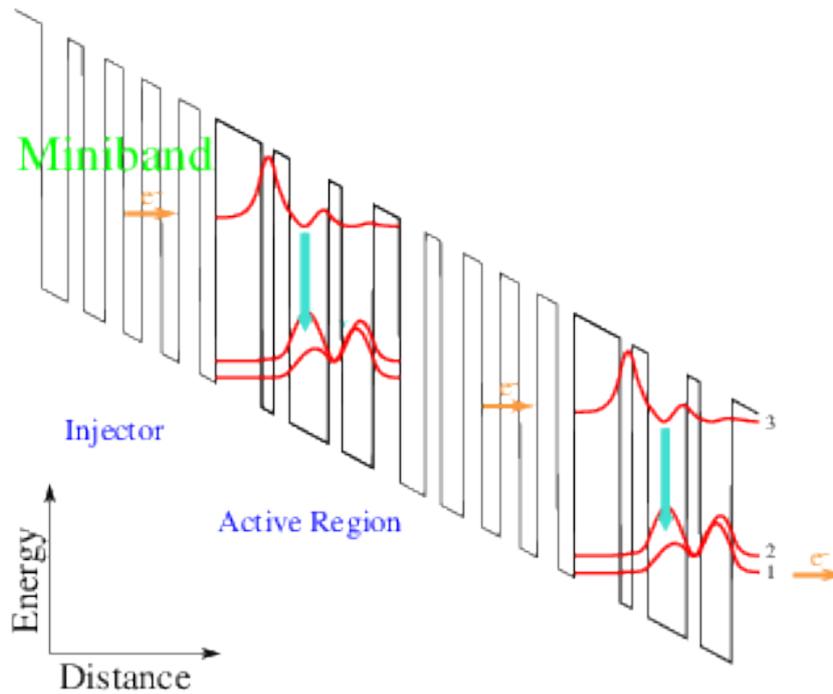


Abbildung 5.34: Bänderdiagramm einer QCL-Struktur mit Minibändern. Der vertikale (energetische) Abstand zwischen den Minibändern ist geringer als der Abstand zwischen den Bandkanten der generierenden Übergitter-Struktur