Halbleiterbauelemente

Vorlesung von Prof. Dr. J. Leuthold [1]
Institut für Photonik und Quantenelektronik (IPQ)
Fakultät für Elektrotechnik und Informationstechnik
Karlsruhe Institute of Technology (KIT)
D-76131 Karlsruhe, Germany

WS 2011/2012

Inhaltsverzeichnis

Einleitung

Die Vorlesung „Halbleiterbauelemente“ soll einerseits einen Einblick in die Physik und Wirkungsweise der Halbleiterbaulemente vermitteln und zum andern eine Einführung in die Vielzahl der heute verwendeten Halbleiterbauelemente sein. Der Schwerpunkt dieser Vorlesung liegt auf dem Erarbeiten von Bauelementmodellen unterschiedlicher Komplexität. Ausgefeilte Computermodelle werden hier nicht angestrebt, da diese, so wichtig sie für den tatsächlichen Schaltungsentwurf sind, in der Regel keine tiefere Einsicht in die Bauelementfunktion und die prinzipiellen Grenzen der Bauelementanwendung erlauben.

Die Vorlesung ist wie folgt organisiert: In den ersten Kapiteln werden die für die Halbleiterbauelemente wichtigen festkörperphysikalischen Grundlagen zusammengestellt. Teile davon wurden auch schon in der Vorlesung Festkörperelektronik behandelt. Die weiteren Kapitel befassen sich mit den Halbleiter-Grundgleichungen, d. h. dem elektronischen Transport und den pn-Übergängen. Die eigentliche Halbleiterbaulemente Diskussion beginnt mit dem Kapitel über pn-Dioden und den Bipolar-Transistoren. Anschließend diskutieren wir Halbleiter-Grenzflächen und Metallübergänge sowie die große Familie der Feldeffekt-Transistoren.

Die Vorlesung lehnt sich an die von den Vorgängern, Prof. Grau und Prof. Kärtner, an der Universität Karlsruhe gehaltenen Vorlesungen über Halbleiterbauelemente aus den Jahren 1997/98 und 2000/2001 an. Teile der Vorlesung wurden aus den in der Literaturliste angegebenen Lehrbüchern entnommen.

Begleitend zur Vorlesung bieten sich insbesondere das Lehrbuch von Thuselt oder die Lehrbücher von B.G. Streetman, R.F. Pierret an. Das Buch von Sze, ‚Kwok und Ng beschreibt eine große Anzahl von Halbleiterbauelementen sehr ausführlich. Die Bücher von Ashcroft und Kittel vermitteln einen tieferen Einblick in die Physik der Halbleiter - geben aber keine Überblick über die Halbleiterbauelemente.

Literaturempfehlungen

Einführende Literatur:

- F. Thuselt; Physik der Halbleiterbauelemente"; Springer-Verlag, 2. Auflage, 2011
- B.G. Streetman, S.K. Banerjee; SSolid State Electronic Devices", 6th ed., Pearson Prentice Hall, 2006
- Pierret, R.F.: Semiconductor Device Fundamentals. Addison Wesley 1996
- Sze, S. M.: Semiconductor Devices, Physics and Technology. New York, John Wiley 1985.
- Reisch, M.: Halbleiter-Bauelemente. Berlin-Heidelberg, Springer-Verlag 2005

Weiterführende Literatur:

- S.M. Sze, Kwok K. Ng: Physics of Semiconductor Devices. New York: Wiley-Interscience, 3rd Ed., 2007
- Ashcroft, N. W.; Mermin, N. D.: Solid State Physics, Saunders College (1976)
- Kittel, Ch.; Einführung in die Festkörperphysik, 7. Auflage, Oldenburg Verlag, 1988

Kapitel 1

Grundlegende Eigenschaften von Halbleitern

In diesem Kapitel sollen die grundlegende Eigenschaften von Halbleitern diskutiert werden. Ausgehend von der physikalisch messbaren Leitfähigkeit werden wir die Halbleiterkristallstruktur und die Halbleiterbandstruktur diskutieren.

1.1 Halbleitermaterialien

Festkörper kann man in Abhängigkeit von der elektrischen Leitfähigkeit in drei große Klassen unterteilen: Isolatoren, Halbleiter und Leiter (Metalle).

Als ***Halbleiter (semiconductor)*** werden die Elemente bzw. Verbindungen bezeichnet, deren Leitfähigkeit bei Zimmertemperatur und bei höchster Reinheit zwischen jener der Metalle und jener von Isolatoren liegt. In Fig. 1.1 sind die elektrische Leitfähigkeit σ und der spezifische Widerstand ($\rho = \frac{1}{\sigma}$) für verschiedene Stoffe abgebildet. Je nach Reinheit des Halbleiters ändert sich der spezifische Widerstand um mehrere Dekaden.

Die Leitfähgikeit nimmt bei Metallen mit steigender Temperatur ab. Im Gegensatz dazu verbessert sich diese in Halbleitern mit steigender Temperatur. In den nachfolgenden Kapiteln werden wir der Ursache auf den Grund gehen.

Man unterscheidet zwischen ***Element-Halbleitern*** wie dem Si oder Ge, die alle in der IV-ten Hauptgruppe des Periodensystems der Elemente, siehe Fig. 1.2, auftreten und zwischen ***Verbindungshalbleitern (compound semiconductors).***Verbindungshalbleiter gehen aus Elementen der IV-ten mit der IV-ten Gruppe, Elementen der III-ten mit der V-ten Gruppe oder Elementen der II-ten mit der VI Gruppe hervor. Die wichtigsten Halbleiter sind in Tabelle 1.1 zusammengestellt, in die auch zusätzlich IV-IV-Verbindungshalbleiter eingetragen sind. Besteht der Verbindungshalbleiter aus Elementen der Gruppen III und V, so werden sie $A^{III}B^{V}-$Verbindungen genannt usw.

Aus 2 Elementen bestehende Verbindungshalbleiter nennt man ***binäre Halbleiter.*** Entsprechend den binären Halbleitern lassen sich die aus drei Elementen bestehenden ***ternäre Halbleiter*** herstellen. Dazu zählen z.B. $Al_{1-x}Ga_xAs$ oder $Al_{1-x}In_xAs$,... . Ferner kennen wir die aus vier Ele-

Element-HL	Verbindungs-HL		
	IV-IV-Verbindungen	III-V-Verbindungen	II-VI-Verbindungen
C,	SiC	AlP, AlAs, AlSb,	ZnS, ZnO, ZnSe, ZnTe
Si,	SiGe	GaN, GaP, GaAs, GaSb	CdS, CdSe, CdTe
Ge		InP, InAs, InSb	HgS, HgSe, HgTe

Tabelle 1.1: Die wichtigsten Element- und Verbindungs-Halbleiter

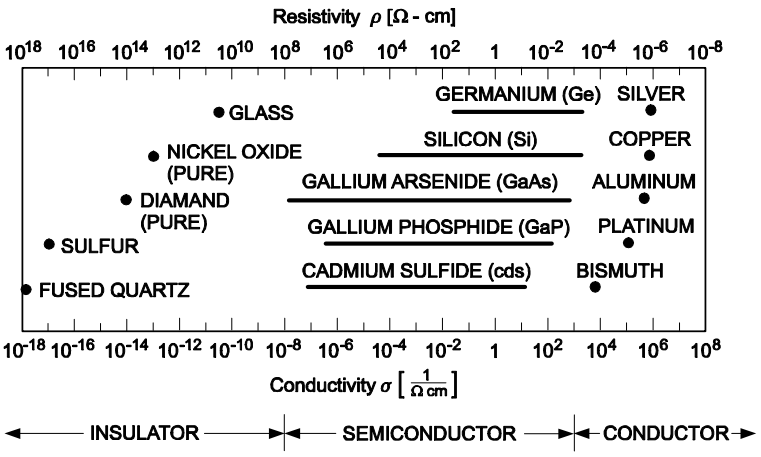


Abbildung 1.1: Spezifischer Widerstand verschiedener Stoffe bei Zimmtertemperatur

menten zusammengesetzten ***quaternären Halbleiter***, z.B. $In_{1-x}Ga_xAs_{1-y}P_y$.

1.2 Chemische Bindung und Kristallstruktur

1.2.1 Chemische Bindungen

Der Übersicht halber seien hier die wichtigsten chemischen Bindungen aufgeführt:

1. **Ionische Bindung**: Starke Polarisation der beteiligten Atome, so dass sich letztlich stabile Edelgas-Elektronenkonfigurationen heranbilden. Da die Materialien keine freien Elektronen haben, sind diese schlechte Wärme- und Elektriztäs-Leiter. Die Bindungen sind stabil und schlecht verformbar. Bsp.: Salze, MgO,...

2. **Kovalente Bindung**: Benachbarte Atome teilen sich ein Elektronenpaar und bilden so eine abgeschlossene Elektronenschale. Dabei muss man sich vorstellen, dass sich die Atomorbitale benachbarter Atome gegenseitig durchdringen. Die Bindungen sind stabil, die Stoffe sind hart, schwer verformbar. Bsp: Isolatoren, Halbleiter

3. **Metallische Bindung**: Die Atome geben die Elektronen der äussern Schalen ab. Es bildet sich ein quasi-freies Elektronengas welches die positiven Atomrümpfe bindet. Die metallischen Bindungen sind gut verformbar und bilden gute thermische und elektrische Leiter.

4. **Van-der Waals'sche Bindung**: Es kommt zu keinem Elektronenaustausch zwischen den Atomen. Stattdessen influenzieren Makromoleküle Dipole und binden sich via Dipolkräfte an andern Molekülen. Es handelt sich dabei um eine schwache Bindungsart.

5. **Mischbindungen**: Obige Bindungen treten in der Natur häufig kombinert auf.

1.2.2 Festkörperklassifikation

Drei große Klassen von Festkörpern lassen sich unterscheiden. Diese sind in Bild 1.3 schematisch aufgeführt.

1.2.3 Halbleiterkristallbindungen - kovalente Bindungen

Die Eigenschaften der Halbleiter ergeben sich im wesentlichen aus der Anordnung der Einzel-Elektronen im Halbleiteratom. Beispielhaft sind die Elektronenkonfigurationen von Si und Ge in

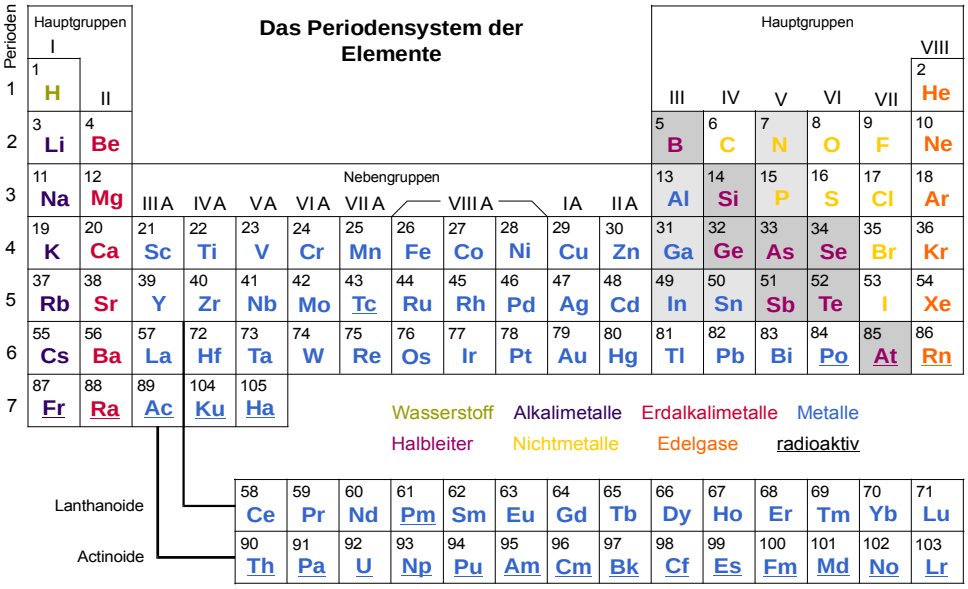


Abbildung 1.2: Periodensystem der Elemente: Die Element-Halbleiter sind zwischen den Metallen und Nichtmetallen zu finden. Verbindungshalbleiter können auch aus der Verbindung von einem Metall mit einem Nichtmetall hervorgehen.

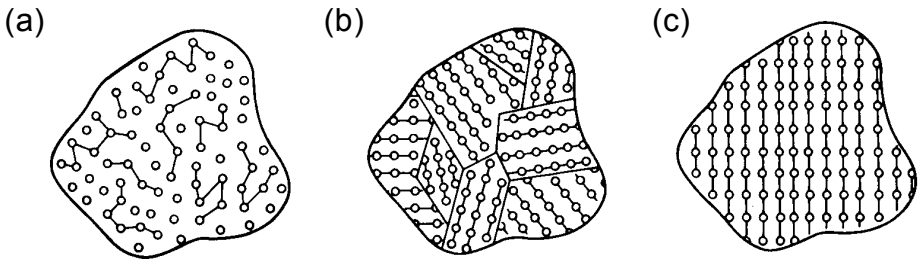


Abbildung 1.3: (a) Amorphe, (b) polykristalline und (c) kristalline Festkörper.

Fig. 1.4 aufgezeichnet. Die abgeschlossenen Elektronenschalen bilden zusammen mit dem Atomkern den ***Atomrumpf***. Die Elektronen in den nicht abgeschlossenen äußeren Schalen können mit anderen Atomen sogenannte kovalente Bindungen eingehen. Durch das Eingehen von kovalenten Bindungen und das gemeinsame Nutzen von Elektronen können die äußeren Schalen vervollständigt werden. Da abgeschlossene Schalen einen energetisch günstigen Zustand darstellen führt das auch zur Bildung von stabilen Kristallen. Wie man der Anordnung der Elektronenkonfigurationen in Fig.1.4 entnehmen kann, haben sowohl Ge als auch Si in den äußeren s- und p- Orbitalen 4 Elektronen. Durch das Eingehen von kovalenten Bindungen mit Elektronen von benachbarten Atomen kann sich eine stabile Verbindung mit abgeschlossenen s- und p- Orbitalen bilden. Ganz ähnlich wie bei Si und Ge aus der VI-ten Gruppe lassen sich durch Auffüllen der Schalen stabile Verbindungen mit abgeschlossen Schalen aus Elementen der III/V oder II/VI Gruppen bilden.

Die Wechselwirkung zwischen den Atomen ist zunächst gekennzeichnet durch die Coulomb-Wechselwirkung zwischen den Ladungen Die Coulomb-Kraft bewirkt, dass die positiv geladenen Kerne von den negativen Elektronen angezogen werden und führt zu einer Verdichtung. Sobald sich die Atome aber zu nahe kommen und sich die Elektronenhülle verdichten, überwiegt die abstossende Coulomb Kraft. Als Folge davon, bildet sich ein energetisch günstiger Gitterabstand aus. Die exakte Anordnung und die exakten Energieniveaus, welche sich aus dem Aneinanderfügen der Atome ergibt, kann nur durch eine quantenmechanische Rechnung unter Einbezug der

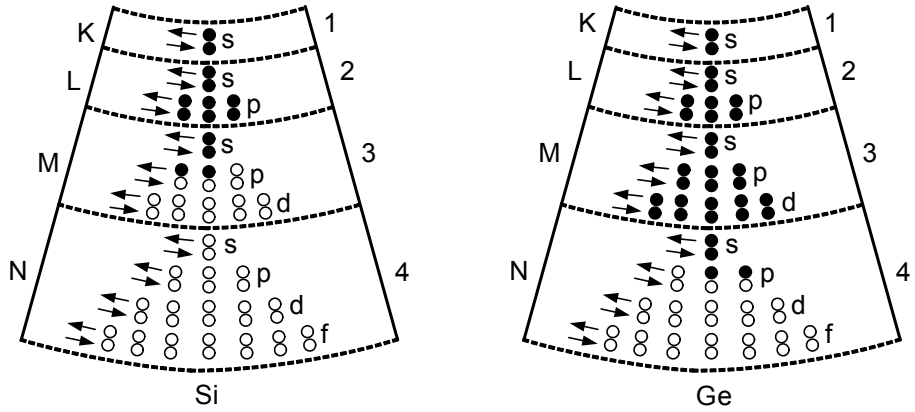


Abbildung 1.4: Elektronenkonfiguration von Si und Ge. Die Elektronenschalen sind durch die Buchstaben K, L, M und N und die dazugehörigen Hauptquantenzahlen gekennzeichnet. Die Pfeile geben die Spinorierntierung der Zustände an.

Coulomb-Kräfe, der Bohrschen Postulate und des Pauli-Vebotes berechnet werden. Wir werden dieses Verfahren weiter unten qualitativ diskutieren.

1.2.4 Einheitszelle

Die ***Einheitszelle*** oder ***Elementarzelle (unit cell)***, ist eine möglichst kleine Volumeneinheit des Gitters, welche dazu benutzt werden kann, um das ganze Gitter zu beschreiben. Diese enthält sämtliche Symmetrieeigenschaften des ganzen Gitters. Die ***primitive Einheitszelle*** ist dabei die kleinst mögliche Einheitszelle.

Einheitszellen sind deshalb so wichtig, weil sich die Eigenschaften entlang des Gitters in repetierender Weise ändern. Man sucht sich dann die kleinst mögliche Gitterstruktur, welche alle Eigenschaften des Gesamtgitters enthält und diskutiert nur diese Zelle um auf die physikalischen Eigenschaften des gesamten Kristalls zu schliessen.

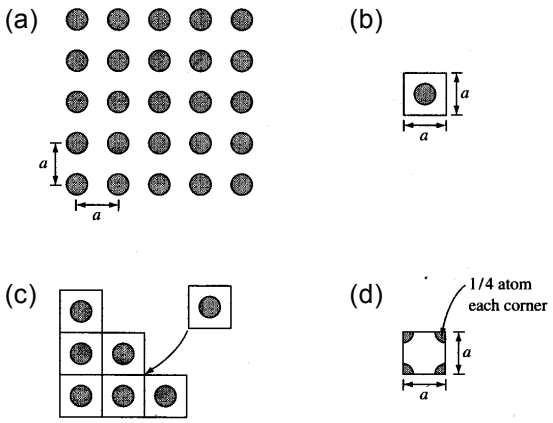


Abbildung 1.5: Einführung zur Einheitszelle: (a) Zwei-dimensionaler Kristall, (b) Einheitszelle, (c) der Kristall kann aus der Einheitszelle wiederhergestellt werden, (d) eine andere mögliche Einheitszelle.

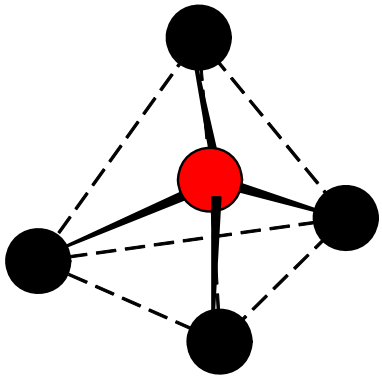


Abbildung 1.6: Siliziumatom mit den vier nächsten Nachbarn [6]

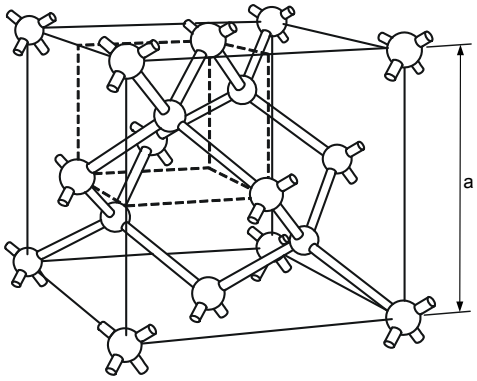


Abbildung 1.7: Einheitszelle des Diamantgitters. Fortgesetzte Wiederholung dieser Zelle bildet das erweiterte Diamantgitter, wobei bei Silizium der Gitterabstand ungefähr 5,43Å beträgt und es ungefähr 5×10^{22} Si-Atome pro Kubikzentimeter gibt [6].

1.2.5 Die Halbleiterkristallstrukturen

Die Kristallstruktur des Halbleiters bildet sich beim Erstarren des Halbleiters aus der Schmelze. Die einzelnen Atome lagern sich in der energetisch günstigsten Art und unter Befolgung der Bohr-Postulate und des Pauli-Verbotes in eine Kristallstruktur ein. Die energetisch günstigste Art ist erreicht, wenn die Elektronen der äußeren Schale wegen der abstoßenden Columb-Wechselwirkung maximal voneinander entfernt sind. Diese maximale Abstoßung ist dann gegeben, wenn die vier Elektronen der äußeren Schalen sich tetraedrisch um den Atomrumpf anordnen. Bild 1.6 zeigt die tetraedrische Struktur - wie man sie etwa bei Si oder Ge vorfindet.

Der daraus resultierende ideale Kristall für C, Si oder Ge ist wird in Bild 1.7 gezeigt. Bei Fig. 1.7 handelt es sich um die primitive Einheitszelle von C, Si oder auch Ge. Man kann sich diese aus zwei, sich gegenseitig durchdringenden, kubisch-flächenzentrierten Gittern zusammengesetzt vorstellen, wobei ein Gitter gegenüber dem anderen um ein Viertel der Würfel-Diagonale entlang dieser verschoben ist. Dieses Gitter wird auch als ***Diamant-Gitter*** bezeichnet, da das Kristallgitter von Kohlenstoff in Form des Diamanten dieselbe Gitterstruktur besitzt. Dieser Sachverhalt ist nochmals in den Bildern 1.8(a) und 1.8(b) dargestellt.

Bild 1.9 zeigt die Elektronenkonfiguration des dreiwertigen Ga und des fünfwertigen As, die den Verbindungshalbleiter GaAs ergeben. Das Kristallgitter von GaAs hat die ***Zinkblendestruktur,*** was der Diamantstruktur nach Bild 1.9 entspricht, nur dass abwechselnd Si durch Ga oder As ersetzt wird. Bild 1.10(b) zeigt gerade die Einheitszelle von GaAs bzw der Zinkblendestruktur.

Die II-VI Verbindungen haben in der Regel Zinkblende- oder Wurzit- Struktur, siehe Bild 1.10,

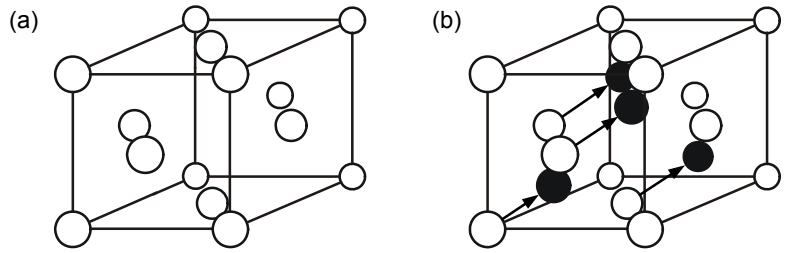


Abbildung 1.8: (a) Einheitszelle eines kubisch flächenzentrierten (fcc) Kristalles. (b) Einheitszelle des Diamantgitters, zusammengesetzt aus zwei gegeneinder verschobenen fcc Gittern. Beim Diamantgitter sind beide Untergitter aus denselben Atomen zusammengesetzt. Bestehen die Untergitter aus verschiedenen Materialien, so wird die Gesamtstruktur als Zinkblende-Gitter bezeichnet. In (b) sind nur die Atome des ersten Gitters (schwarz) eingezeichnet, welche in die Elementarzelle des ersten Gitters fallen. [6]

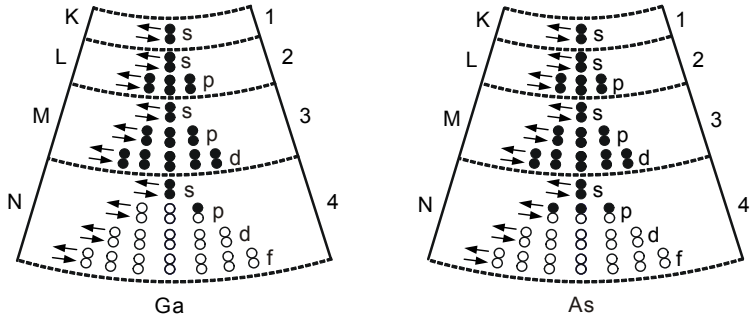


Abbildung 1.9: Elektronenanordnung der Elemente Ga und As. [3]

z. B. ZnSe (Zb) oder ZnO (W), CdS (W). Die ***Wurzit-Struktur*** besteht aus zwei Hexagonalgittern, bei dem das zweite Gitter gegen das erste Gitter um 1/3 in Richtung der Längsachse verschoben ist. Bei CdS besteht das erste Hexagonalgitter aus Cadmium und das zweite aus Schwefel.

Da sich die Eigenschaften eines Kristall je nach Kristallrichtung ändern, ist es wichtig, dass man schon bei der Herstellung von Halbleiterbauelementen auf die Ausrichtung des Kristalls und der Bauteile gegenüber den Kristallachsen achtet. Per Konvention werden Kristallausrichtungen mit den sogenannten ***Miller-Indizes*** angegeben. Dabei definiert man:

[hkl]: Angabe der Kristallrichtung. Z.B. bezeichnet [100] die x-Achse.

(hkl): Angabe der Ebene, welche normal zur Achse [hkl] steht.

Beispielhaft zeigt Fig. 1.11 verschiedene Ebenen und deren Notation.

1.3 Energiebänder

Im Folgenden wollen wir uns Gedanken über die Bindungsstärke der kovalenten Bindungen im Kristall machen und anhand der erlaubten und verbotenen Elektronen-Zustände zu einer neuen Definition dessen, was ein Isolator, ein Halbleitern oder ein Metall ist, gelangen.

1.3.1 Atom-Energiezustände

Im weiteren wollen wir uns ein Bild über die im Kristall vorhandenen Energieniveaus machen. Dazu betrachten wir vorerst die Energieniveaus im isolierten Atom. Grundsätzlich bewegen sich die Elektronen mit großer Geschwindikeit um die positiven Atomkerne herum. Dabei werden sie durch die Coulombkräfte am Entweichen gehindert. Die Coulomb-Kräfte allein können aber die

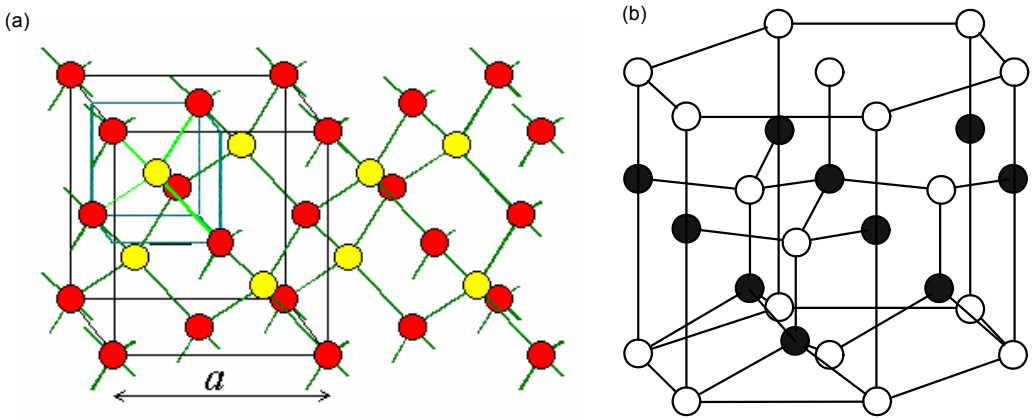


Abbildung 1.10: (a) Zinkblende-Gitter, (b) Wurzit-Gitter [5]

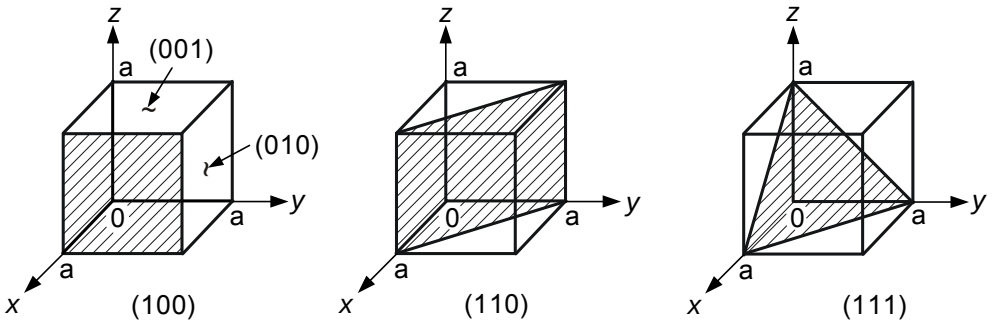


Abbildung 1.11: Die schraffierten Ebenen sind (100), (110) und (111) Ebenen.

Existenz von Atomen noch nicht erklären. Denn eigentlich müssten Elektronen auf Atombahnen sofort verstrahlen. Diese wechseln nämlich Ihre Bewegungsrichtung konstant und müssten - wie man von einem Hertzschen Dipol weiss, sofort ihre Energie durch Strahlung abgeben und dann in den Kern fallen. Bohr hat deshalb postuliert, dass Elektronen wie Wellen um den Kern oszillieren. Falls die Wellenbahn um das Atom zu einer konstruktiven Interfernz der Wellebewegung führt, ist die Bahn stabil und zerfällt auch nicht. Natürlich hat er das nicht so anschaulich formuliert. Seine etwas nüchterne Ausformulierung lautet:

1. **BOHRsches Postulat:** Es sind nur diskrete Bahnen erlaubt. Festlegung der diskreten Bahnen erfolgt durch Quantelung des Bahndrehimpulses.

2. **BOHRsches Postulat:** Bewegung auf den erlaubten Bahnen erfolgt strahlungslos. Emission und Absorption von Energie (Licht) erfolgt durch Übergänge zwischen den Bahnen.

Das erklärt aber immer noch nicht die verschiedenen Anordnungen der Elektronen und der Atomdurchmesser. Denn nun könnte es immer noch passieren, dass alle Elektronen mit der gleichen Wellenbewegung im Gleichrhytmus um das Atom herum kreisen. Das Pauli-Ausschliessungsprinzip ***(Pauli-Verbot)*** verbietet genau dies. Es besagt, dass es zwei Arten von Elementarteilchen gibt. Zum einen Fermionen und zum andern Bosonen. Elektronen (und auch Protonen) sind Fermionen. Alle Teilchen in der Natur, welche Fermioneigenschaft besitzen unterliegen dem sogenannten Pauli-Verbot, welches besagt, dass es in einem System keine zwei Fermionen mit identischen Zuständen geben darf. Die Fermionen unterscheiden sich damit von den Photonen, welche zu den Bosonen zählen und welche, wie man weiss, beliebig oft identische Zustände annehmen können (siehe. Laserlicht). Bis dato gibt es keine vernünftige Erklärung dafür, warum die beobachtbaren Teilchen entweder Fermionen oder Bosonennatur haben. Die Natur scheint so gesetzt zu sein.

Im Fall des Wasserstoffatoms, ergeben sich aufgrund der Coulombkräfte und unter Berücksichtigung der Bohr-Postulate und des Pauli-Verbotes exakt berechenbare Atomorbitale, in welchen sich Elektronen mit diskreten Energiezuständen bewegen. Die für die Elektronen erlaubten Energieniveaus sind nach dem Bohrschen Model:

$$E_H = -\frac{m_0 q^4}{8\epsilon_0^2 h^2 n^2} = \frac{-13.6}{n^2}[eV],$$

mit der Masse des Elektron m_0, der Elektronladung q, der Dielektrizitätskonstanten ϵ_0 , dem Planckschen Wirkungsquantum h und der Bahnhauptquantenzahl n . Diese Energieniveaus sind voneinander getrennt und die Energieabstände können mit der Bohrschen Formel berechnet werden.

Genau das gleiche Prinzip kann auf alle andern Atome angewendet werden und es zeigt sich, dass die Elektronen diskrete und von der Kernzahl abhängige Energiewerte annehmen.

1.3.2 Kristall-Bandstruktur

Wenn wir nun zunächst zwei isolierte Atome betrachten, dann haben diese exakt identische Energieniveaus. Bringen wir diese räumlich nahe zusammen, dann bilden sich einerseits neue, energetisch günstigere Niveaus mit tieferer Gesamtenergie und anderseits fächern sich die Niveaus energetisch um kleine Energiebeträge auf, Bild 1.12. Wenn wir N Atome zusammenbringen, dann spalten sich diese in N Niveaus auf. Um die isolierten Energieniveaus des Einzelatoms bilden sich mit zunehmender Teilchenzahl immer feinere Aufspaltungen bis man letztlich nur noch von Energiebändern spricht. Der tiefere Grund, weshalb es zu dieser Auffächerung kommt liegt einmal mehr an der Ferminatur der Elektronen. Das Pauli-Verbot gilt nicht nur innerhalb der Atome, das Pauli-Verbot muss auf alle sich überlappenden Teilchen angewendet werden und damit auf all Elektronen innerhalb des gleichen Festkörpers.

In Figur 1.13 ist die Aufspaltung der Atomorbitale im Si aufgezeichnet. Da im Si nur vier der acht möglichen Niveaus der s- und p-Orbitale besetzt sind, sind auch nur die Hälfte der möglichen

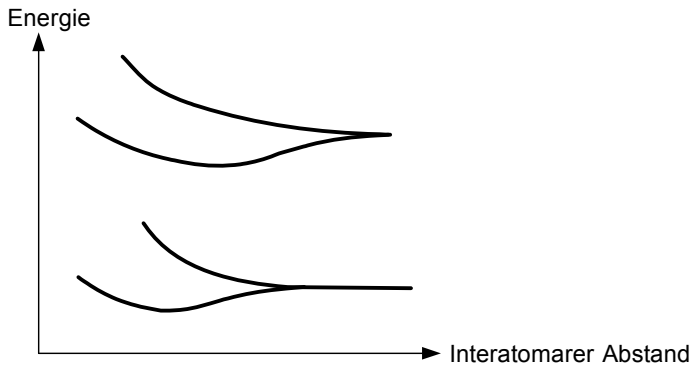


Abbildung 1.12: Aufspalten von Energiezuständen zweier Atome als Funktion des Abstandes. Die Entartung wird aufgehoben [3].

Bandniveaus besetzt. Zufälligerweise bildet sich beim Atomabstand von $r_0 = 0.234\,\text{nm}$ zwischen den energetisch tiefen und den energetisch höhernem Bändern gerade eine Energielücke aus, so dass man zwischen dem mit Elektronen besetzten Valenzband und dem energetisch über der Bandlücke liegenden Leitungsband unterscheiden kann.

1.3.3 Die Bandlücke

Zur Veranschaulichung der Energiebänder im Halbleiter, wechselt man üblicherweise auf das sog. eindimensionale Bandschema, Bild 1.14. Im Halbleiter wird bei der Temperatur $T = 0$ K das höchste, vollbesetzte Band als ***Valenzband VB (valence band)*** bezeichnet. Die höchste Energie für ein Elektron in diesem Band ist E_V (Valenzbandkante). Der darauf folgende Bereich, die ***Bandlücke (bandgap)*** bietet keine möglichen Plätze für Elektronen bis zur Energie E_L (der Leitungsbandkante). Die Größe

$$E_G = E_L - E_V > 0 \tag{1.1}$$

wird als Bandabstand bezeichnet. Bei der Temperatur $T = 0$ ist das ***Leitungsband LB (conduction band)*** völlig leer. Höher liegende Bänder, die das Leitungsband LB überkreuzen und zu höheren Energien $E > E_L$ immer Plätze für Elektronen ohne weitere Energielücken anbiete, sind nicht eingezeichnet. E_G ist beim Halbleiter in der Größenordnung von 1 eV (im Gegensatz zum Isolator mit E_G in der Größenordnung 10 eV, so dass bei Raumtemperatur keine merkliche Leitfähigkeit vorhanden ist). Bei Metallen, Bild 1.14, hingegen ist das energetisch am höchsten liegende nicht leere Band nur teilweise besetzt, so daß Ladungsträger zum Stromtransport beitragen können.

Es sei an dieser Stelle angemerkt, dass die Bandlücke E_G eine Funktion der Temperatur ist, was in Bild 1.15 dargestellt ist. Bei Raumtemperatur T_0 ist eine lineare Approximation der Bandlücke oftmals ausreichend

$$E_g(T) = E_g(T_0) + \frac{dE_g}{dT}\Delta T. \tag{1.2}$$

In Bild 1.15 sind verschiedene Werte für $E_g(T_0)$ und dE_g/dT aufgeführt. Bei großen Temperaturschwankungen empfiehlt es sich ein parabolisches Modell für die Bandlücke zu verwenden

$$E_g(T) = E_g(T = 0) - \frac{\alpha T^2}{\beta + T}. \tag{1.3}$$

Typische Werte für Si sind $\alpha = 4.73\ 10^{-4}$eV/K und $\beta = 636$K und für GaAs $\alpha = 5.4\ 10^{-4}$eV/K und $\beta = 204$K.

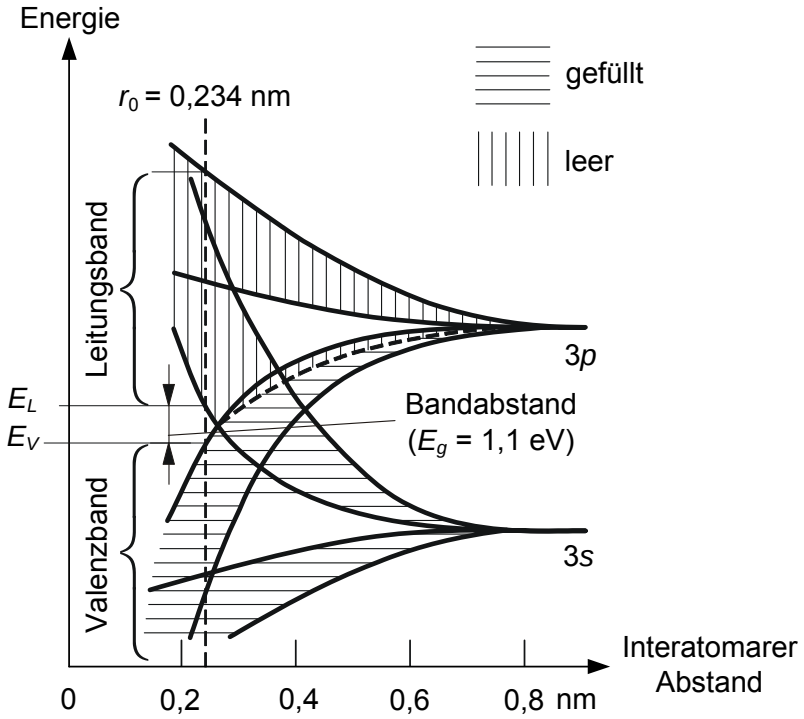


Abbildung 1.13: Energiebänder in Silizium als Funktion des interatomaren Abstands. [3]

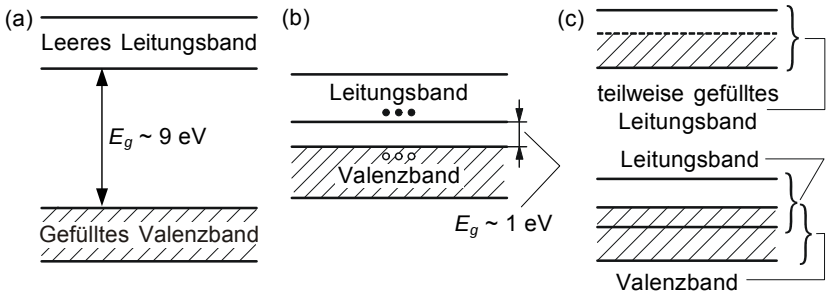
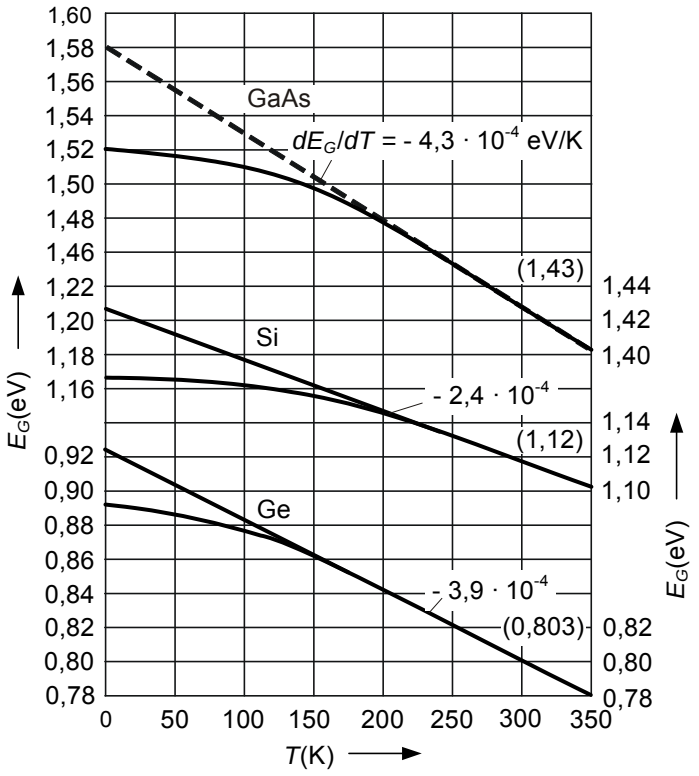


Abbildung 1.14: Schematische Darstellung der Bandstruktur (genauer Extrema der Banstruktur) eines (a) Isolators, (b) Halbleiters und (c) Metalls.



von GeSiGaAs.wmf

Abbildung 1.15: Bandabstand von Ge, Si und GaAs als Funktion der Temperatur. Für Ge ist der direkte Bandabstand angegeben [3].

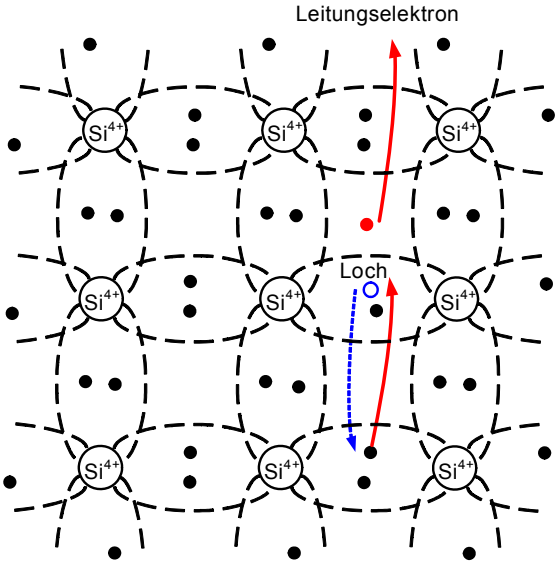


Abbildung 1.16: Zweidimensionale Abbildung eines Si-Kristalls. Der vierfach positive Atomrumpf ist von Elektronenwolken umgeben. Eine kovalente Bindung ist aufgebrochen. Es bildet sich ein freies Elektron, welches ein Loch in der kovalenten Bindung hinterlässt. Jeder Punkt repräsentiert ein Elektron in einer Bindung zwischen zwei Nachbaratomen.

1.3.4 Leitungselektronen, Löcher und Leitfähigkeit

Der Einfachheit halber wollen wir das dreidimensionale Gitter auf ein zweidimensionales Gitter projizieren, wie dies in Bild 1.16 dargestellt ist.

Die kovalenten Bindungen, welche dem Kristall die Struktur geben gelten als stabil. Sie bleiben im entsprechenden Atomorbital gebunden. Da es keine freien Ladungsträger gibt, gibt es auch keine Stromleitung. Dies bleibt bei genügend tiefen Temperaturen auch so. Bei höheren Temperaturen können thermische Schwingungen dazu führen, dass kovalente Bindungen aufbrechen. Damit bilden sich nun aber zwei leitende Teilchen:

1.) Es bildet sich ein freies Elektron, welches energetisch nicht mehr an die Orbitale gebunden ist und sich somit als freies ***Leitungselektron*** an einem Stromfluss beteiligen kann.

2.) Das freie Elektron hinterlässt aber auch ein ***Loch (hole)*** im Bindungsorbital des Kristalls. Dieses Loch kann sich innerhalb der Atomorbitale gleicher Energie verschieben. Genauger gesagt verschiebt sich nicht das Loch sondern ein anderes Elektron von einem Orbital gleicher oder ähnlicher Energie kann die Stelle des Lochs einnehmen und an einer andern Stelle ein Loch hinterlassen. Da das Loch oder die Elektronen, welche die Löcher besetzten nicht frei sind, sondern sich nur innerhalb der vorgegebenen Orbitale bewegen können, ist es nur natürlich, dass sie nicht die gleiche Beweglichkeit haben und somit in nur geringerem Masse zum Stromfluss beitragen können. Und in der Tat bewegen sich die Löcher (oder die an die Löcher gebundenen Elektronen) auch physikalisch messbar viel langsamer und träger als freie Elektronen. Phänomenologisch kann man ihnen einen größere Masse zuordnen um dem Umstand des trägeren Verhaltens gerecht zu werden.

Damit ein Kristall den Strom leiten kann, darf die zur Erzeugung freier Elektronen notwendige Energie also nicht zu hoch sein.

Kapitel 2

Bandstruktur der Festkörper

Im Folgenden wollen wir ein tieferes Verständnis der Energieniveaus, respektive der Bandstrukturen im Festkörper, gewinnen. Die Bandstrukturen beeinflussen nämlich die Eigenschaften der Kristalle nachhaltig. Wenn wir die Halbleitereigenschaften im vorhergehenden Kapitel qualitativ diskutiert haben, so möchten wir in diesem Kapitel physikalisch korrekte Aussagen zu diesen machen können. Des Weiteren vermittelt dieses Kapitel grundlegende physikalische Einsichten der Physik von Festkörpern.

2.1 Quantenmechanische Grundlagen

2.1.1 Welle-Teilchen Dualismus und Schrödingergleichung

Einstein postulierte 1905 die Welle-Teilchen Dualität von Licht. Ganz analog postulierte L. de Broglie die Welle-Teilchen Dualität von Materie. Demzufolge kann man jedem Teilchen aufgrund seiner Energie eine Frequenz ω zuordnen und dem Impuls p eine Wellenzahl k.

$$E = \hbar\omega \tag{2.1}$$

$$p = \hbar k \tag{2.2}$$

Im Vergleich zu den identischen Gleichungen mit dem Photon gibt es allerdings einen gewichtigen Unterschied. Beim Licht ist die Energie eine absolute Größe und man kann den Photonen eindeutig eine Wellenlänge zuordnen. So z.B. über

$$c = v_{ph} = \omega/k = E/(2\pi/\lambda) \tag{2.3}$$

wobei $k = 2\pi/\lambda$ uns auf die Wellenlänge führt. Bei einem Materieteilchen, welches sehr stark lokalisiert ist, ist die einzig relevante physikalische Geschwindigkeit jedoch die Gruppengeschwindigkeit $v_{gr} = d\omega/dk$. Da ein zusätzlicher konstanter Energie-Term diese nicht verändert, muss die Energie E in der Gleichung für Materiewellen als relative Energie bezüglich einer frei wählbaren Eichung interpretiert werden.

Akzeptiert man einmal, dass Materieteilchen ebenso Wellencharakter haben wie man das von elektromagnetischer Strahlung oder Wasserwellen gewohnt ist, so ist es nur logisch, dass man Materieteilchen zu einer bestimmten Zeit keinen exakten Ort mehr zuordnen kann sondern diese durch eine komplexe Wahrscheinlichkeitsdichteamplitude $\Psi(x,t)$ beschreiben muss. Die Wahrscheinlichkeit das Teilchen zur Zeit t am Ort x zu finden ist dann

$$w = |\Psi(x,t)|^2.$$

1926 hat Erwin Schrödinger die ***Wellengleichung für quantenmechanische Teilchen*** formuliert. Diese lautet mit der Hamiltonfunktion $H(x,t)$

$$-\mathrm{j}\,\hbar\frac{\partial}{\partial t}\Psi(x,t) = H(x,t)\Psi(x,t). \tag{2.4}$$

Für Teilchen in einem Potential $V(x)$ lautet lautet die Gleichung

$$-\mathrm{j}\,\hbar\frac{\partial}{\partial t}\Psi(x,t) = \left[-\frac{\hbar^2}{2m}\Delta + V(x)\right]\Psi(x,t). \tag{2.5}$$

In dieser Vorlesung werden wir natürlich nicht in aller Breite auf alle Implikationen dieser Gleichung eingehen. Es ist aber dennoch wichtig einzusehen, dass diese Gleichung nicht aus der Luft gegriffen ist. In diesem Sinne wollen wir sie anhand des Beispiel des freien Elektrons im Potential $V(x)$ veranschaulichen.

Wie eingangs diskutiert, hat de Broglie den Teilchen Wellencharakter attestiert. In diesem Sinne ist es auch gerechtfertigt, diese wie elektromagnetische Wellen zu beschreiben. Konkret könnte man schreiben:

$$\Psi(x,t) = \psi(x)e^{\mathrm{j}\,\omega t} = Ae^{-\mathrm{j}(kx-\omega t)}, \tag{2.6}$$

wobei der Ausdruck $\Psi(x,t)$ nicht mehr ein elektromagnetisches Feld darstellt, sondern als Wahrscheinlichkeitsamplitude interpretiert werden muss. Mit den von de Broglie postulierten Gleichungen wird (2.6) zu:

$$\Psi(x,t) = \psi(x)e^{\mathrm{j}\,\frac{E}{\hbar}t} = Ae^{-j\left(\frac{p}{\hbar}x-\frac{E}{\hbar}t\right)}. \tag{2.7}$$

Setzt man diese Wahrscheinlichkeitsamplitude in die Schrödingergleichung ein, und nimmt man zusätzlich an, dass sich E mit der Zeit nicht ändert und leitet zunächst lediglich nach der Zeit ab, so kommt man auf die ***zeitunabhängige Schrödingergleichung***,

$$E\psi(x) = \left[-\frac{\hbar^2}{2m}\Delta + V(x)\right]\psi(x) \tag{2.8}$$

Nach Einsetzen des ortsabhängigen Ansatzes, findet man schließlich:

$$E\psi(x) = \frac{p^2}{2m}\Psi(x) + V(x)\psi(x)\ . \tag{2.9}$$

Multipliziert man $\Psi(x,t)$ mit der konjugiert komplexen Wahrscheinlichkeitsamplitude $\Psi^*(x,t)$, dann erhält man die Aufenthaltswahrscheinlichkeit am Ort x und Gl. 2.9 stellt die Energieerhaltungsgleichung der klassischen Mechanik dar.

Damit ergibt sich für quantenmechanische Probleme folgendes Lösungsschema. In einem ersten Schritt geht es darum, die zum Problem gehörende Hamiltonfunktion zu finden. Falls wir an der Berechnung von Halbleitern interessiert sind muss man im Wesentlichen den korrekten Ausdruck für das Potential $V(x)$ hinschreiben. In einem zweiten Schritt geht es darum, einen Ansatz für die Wahrscheinlichkeitsamplitude zu finden, welche die Schrödinger-Differentialgleichung löst. Durch Einsetzen von letzterer gelangt man zu einer Energie-Impulsbeziehung. Im folgenden Abschnitt wollen wir darlegen, weshalb diese Energie-Impulsbeziehung von großem Interesse ist.

2.1.2 Die Halbklassischen Bewegungsgleichungen

Zunächst wollen wir jedoch den Begriff Dispersion (dispersion) definieren:

Chromatische Dispersion bezeichnet ganz allgemein die Abhängigkeit einer für uns interessanten Größe von der Ausbreitungskonstanten k bzw. der Wellenlänge (Chroma) λ.

Wenn wir also in unserem Fall die Energie-Impulsbeziehung $E(p)$ studieren, dann ist dies wegen den de Broglie-Beziehungen gleichbedeutend mit dem Studium des Zusammenhangs zwischen

der Kreisfrequenz und der Ausbreitungskonstanten, d.h von $\omega(k)$. $\omega(k)$ wird im Folgenden der Einfachheit halber die Dispersionsrelation genannt.

Die Dispersionsrelation ist deshalb so wichtig, weil sich aus dieser Beziehung sowohl die Teilchengeschwindigkeit v_g als auch die effektive Masse m^* für Materiewellen berechnen lassen. So gilt:

$$\frac{1}{\hbar}\frac{\partial E}{\partial k} = \frac{\hbar}{\hbar}\frac{\partial \omega}{\partial k} = \frac{\partial \omega}{\partial k} = v_g \tag{2.10}$$

$$\frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k^2} = \frac{1}{m^*}\,. \tag{2.11}$$

Ferner gilt für die Impulsänderungen aufgrund einer externen Kraft die Beziehung

$$\frac{\mathrm{d}}{\mathrm{d}\,t}(\hbar k) = F_{\mathrm{ext}}\,. \tag{2.12}$$

Wir wollen diese Relationen hier nicht weiter beweisen. Es sei aber an dieser Stelle gesagt, dass diese Formeln glücklicherweise mit den bekannten Formeln der klassischen Mechanik identisch sind. So erhält man beispielsweise die Gruppengeschwindigkeit eines klassischen Teilchens aus:

$$\frac{\partial E}{\partial p} = \frac{\partial\left(\frac{p^2}{2m}\right)}{\partial p} = \frac{p}{m} = v \tag{2.13}$$

oder auch

$$\frac{\partial^2 E}{\partial p^2} = \frac{\partial^2\left(\frac{p^2}{2m}\right)}{\partial p^2} = \frac{1}{m} \tag{2.14}$$

und es ist in der klassischen Mechanik nach Newton

$$\frac{\mathrm{d}}{\mathrm{d}t}p = F_{\mathrm{tot}}\,. \tag{2.15}$$

Die Gleichungen 2.10 bis 2.12 werden als ***halbklassische Bewegungsgleichungen*** bezeichnet. Dies deshalb, weil wir in einem korrekt quantenmechanischen Modell selbst die von außen einwirkenden Kräfte F_{ext} quantisieren müssten. Auch sind die Gleichungen insofern gefährlich, als sie in Anlehnung an die uns bekannten klassischen Gleichungen aufgeschrieben wurden. In Wirklichkeit muss man diese aber etwas anders interpretieren. Der wichtigste Unterschied zu den klassischen Gleichungen ist, dass im halbklassischen Modell der Einfluss des periodischen Kristall-Potentials in die Energie-Impulsbeziehung, sprich die Dispersionsrelation, eingeht.

Wie geht denn nun das periodische Kristall-Potential in die die Energie-Impulsbeziehung ein? Insofern, als das periodische Potential dafür verantwortlich ist, dass die Elektronen-Wellen am periodischen Potential in Ihrer freien Bahn gestört werden und sich viel langsamer bewegen (nur noch mit der Gruppengeschwindigkeit), als wie man das von einem freien Elektron erwarten würde (Phasengeschwindigkeit). So kann es beispielsweise passieren, dass ein Elektron zwar eine beachtliche kinetische Energie hat, sich aber wegen Reflexionen im periodischen Potential nicht vom Ort bewegt (stehende Welle) und es trotz der hohen kinetischen Energie eine verschwindende Gruppengeschwindigkeit hat. Im klassischen Modell wäre das Elektron ein streng lokalisiertes Teilchen, dass zwar auch reflektiert würde, aber wie ein Tennisball hin- und herspringen würde. Dem Tennisball könnten wir dann zu jeder Zeit eine endliche Geschwindigkeit zuordnen, welche natürlich mit steigender kinetischer Energie zunimmt.

Desgleichen steht in unseren halbklassischen Bewegungsgleichungen nicht mehr die Ruhemasse der Teilchen sondern eine sogenannte effektive Masse, welche den Einfluss des Potentials berücksichtigt. Dies ist wiederum so zu verstehen, dass unsere Elektronen sich in einem periodischen Potential befinden und für gewisse kinetische Energien und bei bestimmten Wellenlängen stark

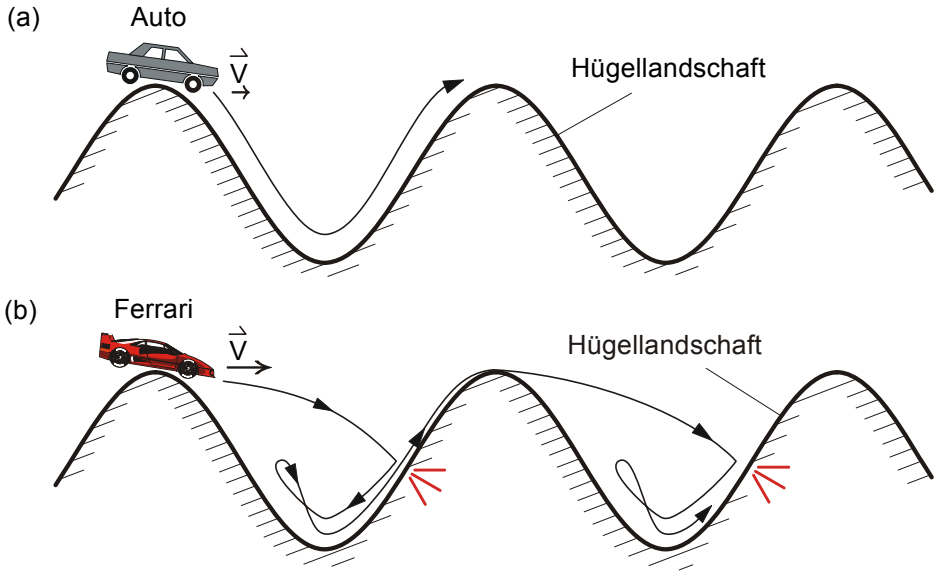


Abbildung 2.1: Trotz hoher Geschwindigkeit (Phasengeschwindigkeit) bewegt sich der Ferrari-Fahrer nur langsam vorwärts (Gruppengeschwindigkeit). In dieser Zeichnung verträgt sich seine Phasengeschwindigkeit mit den Hügeln nur schlecht und er prallt an denselben zurück.

reflektiert werden. Die Reflexionen lassen sie als sich langsam bewegende Teilchen mit einer entsprechend kleineren Gruppengeschwindigkeit erscheinen. Wenn in Tat und Wahrheit die Reflexionen beim Erhöhen des Energiezustandes zunehmen und sich die Gruppengeschwindigkeit nicht wie erwartet erhöht, so drückt sich das in den halbklassischen Bewegungsgleichungen durch eine Erhöhung der Masse aus. Es ist dann so als ob man in einem Auto fährt und aufs Gaspedal drückt und sich trotzdem nicht in dem Maße schneller bewegt wie man es erwartet, weil man in einen Geschwindigkeitsbereich gekommen ist, in dem sich die Eigenmasse des Autos erhöht. Dieses Ausbremsen durch das periodische Gitter wird in den halbklassischen Bewegungsgleichungen durch eine Erhöhung der Masse ausgedrückt. Wenn unser Autofahrer nun weiter aufs Gaspedal drückt und sich die Geschwindigkeit sogar verlangsamt, so müsste er in diesem Modell negative Masse geladen haben, Abb. 2.1. Da in den halbklassischen Bewegungsgleichungen der Kräfteeinfluss des Potentials aufs Teilchen bereits in der effektiven Masse berücksichtigt ist, braucht man in Gl. (2.12) nicht mehr die Impulsänderung aller Kräfte zu berücksichtigen, sondern nur noch die Impulsänderung der externen Kräfte auf das Teilchen.

Jetzt ist auch klar, weshalb wir weiter oben erwähnt haben, dass der ***Kristallimpuls*** $\hbar k$ vom rein mechanischen Impuls des Elektrons verschieden ist.

Beim freien Elektron verändert eine äußere Kraft den mechanischen Impuls gemäß $\hbar k = m_0 v$. Im Kristall ist das Elektron jedoch mit der Wechselwirkung des Gitters "bekleidet". Bei Einwirkung einer äußeren Kraft F, z. B. infolge eines angelegten elektrischen Feldes, wird nicht nur der mechanische Impuls des Elektrons geändert, sondern zufolge seiner Wechselwirkung mit dem Gitter auch Impuls auf das Gitter übertragen.

Der Kristallimpuls hat aber dennoch eine unmittelbare physikalische Bedeutung. Bei der Wechselwirkungen eines Kristallelektrons mit Photonen (Quanten der elektromagnetischen Strahlung) oder Phononen (Quanten der Gitterschwingungen) gilt Energie- und Impulserhaltung. Dabei ist für Elektronen nicht ihr mechanischer, sondern ihr Kristall-Impuls maßgeblich.

2.1.3 Quantisierung

Last but not least, fehlt uns noch ein quantenmechanisches Konzept. Wir müssen verstehen, warum es in einem endlich ausgedehnten Raum nur diskrete, sprich quantisierte Zustände gibt.

Falls der Raum sich nur über die Länge L erstreckt, so könnten wir in einer ersten Arbeitshypothese fordern, dass die Wellenamplitude an den Rändern verschwindet, d.h. die Randbedingungen seien $\Psi(x = 0) = \Psi(x = L) = 0$. Diese Randbedingung wäre aber etwas zu stark, sie würde uns nämlich nur symmetrische oder antisymmetrische stehende Wellen erlauben. Eine etwas schwächere Randbedinung ist

$$\Psi(x = 0) = \Psi(x = L) \ . \tag{2.16}$$

Diese Randbedingung erlaubt uns nun innerhalb des Raums der Länge L alle möglichen Lösungen von stehenden symmetrischen und antisymmetrischen Wellen anzusetzen. Wenn wir gar keine Randbedingungen fordern würden, bzw. $\Psi(x = 0) \neq \Psi(x = L)$ ansetzten würden, so erhielte man als Lösungen keine stehenden, sondern laufende Wellen. Damit könnten die Elektronen dann den Kristall verlassen - und zwar einseitig, so dass der Kristall sich der Elektronen entleert, ohne dass neue Elektronen nachkommen könnten.

In Zukunft werden wir allerdings auch laufende Wellen betrachten. Schließlich wollen wir ja auch Phänomene wie den Ladungstransport verstehen. Auch dann ist es allerdings wichtig, dass für jedes Elektron, welches auf der einen Seite den Raum verlässt ein gleichwertiges auf der anderen Seite des Raumes nachkommt [8, ashcroft88]. Auch für diesen Fall, ist dann die Randbedingung (2.16) gültig.

Damit ergibt sich mit (2.6) für k:

$$Ae^{-\mathrm{j}(k\cdot 0-\omega t)} = Ae^{-\mathrm{j}(kL-\omega t)} \tag{2.17}$$

$$kL = k_i L = 2\pi i, \quad i = 0, \pm 1, \pm 2, \ldots \tag{2.18}$$

Es sind also nicht mehr alle möglichen Wellenzahlen erlaubt, sondern nur diskrete Wellenzahlen. Wird das Quantisierungsintervall bzw. Volumen nur groß genug gewählt, so spielen die Details der Randbedingungen (periodische Randbedingungen oder harte Randbedingungen, d. h. das Elektron kann das Volumen nicht verlassen) keine Rolle mehr und die Elektronen können wie im klassischen Fall alle möglichen Zustände annehmen.

2.2 Das Elektron im Festkörper

Nachdem wir die Grundlagen der Quantenmechanik aufgefrischt haben, wollen wir diese auf das Elektron im Festkörper anwenden. Wir beginnen mit dem Studium des freien Elektrons. Danach werden wir das Elektron im idealen 1-dimensionalen Kristall betrachten und letztlich die Situation im realen 3-dimensionalen Festkörper diskutieren.

2.2.1 Freies Elektron

Im Falle eines freien Elektrons in einem konstanten äußeren Potential $V(x) = V_0$ lautet die dazugehörige Schrödinger-Gleichung:

$$-\mathrm{j}\,\hbar\frac{\partial}{\partial t}\Psi(x,t) = \left[-\frac{\hbar^2}{2m}\Delta + V_0\right]\Psi(x,t)\,. \tag{2.19}$$

Da das Potential nicht zeitabhängig ist, können wir als Lösungsansatz stationäre Zustände des Elektrons mit der exakten Energie E ansetzen

$$\Psi(x,t) = \psi_{\ w}(x)\,\mathrm{e}^{\mathrm{j}\,\omega t} = A\,\mathrm{e}^{-\mathrm{j}\,kx}\,\mathrm{e}^{\mathrm{j}\,\omega t}, \quad \omega = \frac{E}{\hbar}\ . \tag{2.20}$$

Durch Einsetzen des Ansatzes in der Schrödingergleichung findet man die Dispersionsrelation:

$$E = \hbar\omega = V_0 + \frac{(\hbar k)^2}{2m_0}. \tag{2.21}$$

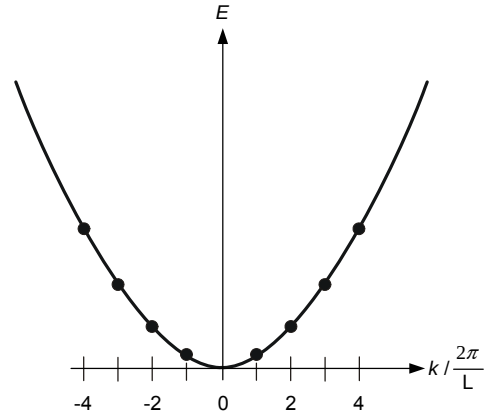


Abbildung 2.2: Dispersionsrelation des freien Elektrons (—). Diskretisierung der Energie, bzw. Impulswerte bei periodischen Randbedingungen mit Periode L.

Die zugehörige Teilchengeschwindigkeit v ergibt sich aus

$$v = \frac{\partial \omega}{\partial k} = \frac{1}{\hbar}\frac{\partial E}{\partial k} = \frac{\hbar k}{m_0} = \frac{p}{m_0}. \tag{2.22}$$

Man beachte, daß $p = \hbar k = m_0 v$ die Bedeutung des mechanischen Impulses des Elektrons hat.

Die effektive Masse ist mit der Masse des Elektrons identisch:

$$\frac{1}{m_0} \quad = \quad \frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k^2}. \tag{2.23}$$

Wegen der periodischen Randbedingungen $\psi(x) = \psi(x \pm L)$ welche für alle ω gelten muss, gilt $1 = \mathrm{e}^{-\mathrm{j}\,kx}$. Damit kann k nur die Werte

$$kL = k_i L = 2\pi i, \quad i = 0, \pm 1, \pm 2, \ldots \tag{2.24}$$

annehmen, siehe Bild 2.2.

Ferner wissen wir, dass die Aufenthaltswahrscheinlichkeit des Elektrons im Raum gerade 1 ist. Dies erlaubt es uns, die Normierungskonstante A zu bestimmen:

$$\int_0^L w(x)\,\mathrm{d}x = 1 \text{ mit } w(x) = |\psi\ (x)|^2 \tag{2.25}$$

Damit erhalten wir für die Wellenfunktion dann:

$$\psi(x,t) = \frac{1}{\sqrt{L}}\,\mathrm{e}^{-\mathrm{j}\,kx}\,\mathrm{e}^{\mathrm{j}\,\omega t}\,. \tag{2.26}$$

2.2.2 Elektronen im 1-dim idealen Kristall (1-dim period. Potential)

In diesem Kapitel soll untersucht werden, wie sich die Dispersionsrelation der äußeren Elektronenwellen in einem Festkörper ändert, wenn Atome zu einem Kristall zusammengefügt werden. Bild 2.3 zeigt das periodische Potential (eindimensional), welches die periodische Anordnung der Atomrümpfe im Abstand a zueinander erzeugt

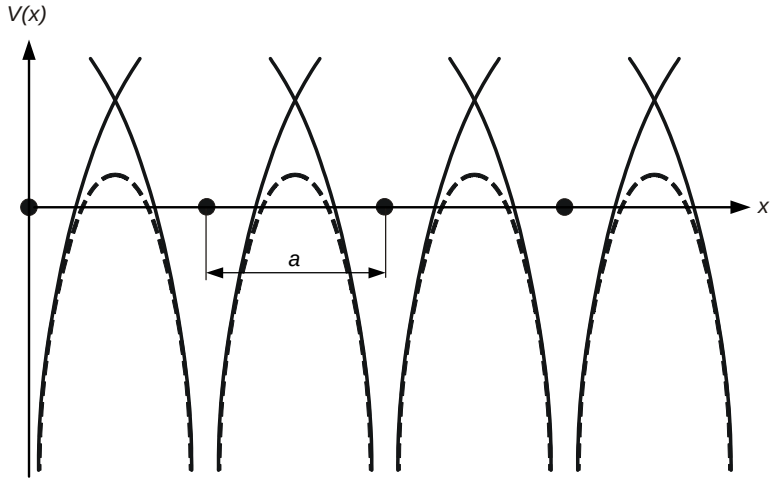


Abbildung 2.3: Periodisches Potential $V(x) = V(x+a)$, gestrichelt gezeichnet, erzeugt durch die periodische Anordnung von Atomrümpfen, deren jeweiliges Potenial durchgezogen gezeichnet ist.

Die gesuchten Zustände der äußeren Elektronen im periodischen Potential $V(x) = V(x+a)$ der Atomrümpfe ergeben sich aus der Lösung der zugehörigen stationären Schrödinger-Gleichung gemäß Gl.(2.5).

$$-\mathrm{j}\,\hbar\frac{\partial}{\partial t}\Psi(x,t) = \left[-\frac{\hbar^2}{2m}\Delta + V(x)\right]\Psi(x,t) \text{ mit } V(x) = V(x+a) \tag{2.27}$$

Es geht nun darum, den geeigneten Lösungsansatz für das periodiodische Potential zu finden. Zunächst beschränken wir uns auf zeitunabhängige Lösungen, wenn keine äußern Kräfte die Gesamtenergie verändern. Damit können wir folgenden ersten Ansatz schreiben:

$$\psi(x,t) = \psi_w(x)\,\mathrm{e}^{\mathrm{j}\,\omega t}\,.$$

Da sich unser physikalisches Problem nicht verändert, wenn wir die ganze Anordnung der Atomrümpfe um eine Gitterkonstante a verschieben, dürfen sich die stationären Lösungen der Schrödinger-Gleichung bei Verschiebung um eine Gitterkonstante höchstens um einen Phasenfaktor verändern. Gemäß Bloch (Bloch-Theorem) ist folgender Ansatz zu wählen:

$$\begin{aligned}&\psi_w(x+a) = e^{-\mathrm{j}\,ka}\psi_w(x), \quad \text{bzw.} \\ &\psi_w(x) = e^{-\mathrm{j}\,kx}u_{k,w}(x), \text{ wobei } u_{k,w}(x+a) = u_{k,w}(x)\end{aligned} \tag{2.28}$$

Ein konstanter Phasenfaktor vor einer Wellenfunktion ändert an der Aufenthaltswahrscheinlichkeit des Elektrons nichts, da die entsprechende Wahrscheinlichkeitsdichteverteilung unverändert bleibt. Würde sich die Aufenthaltswahrscheinlichkeit von Gitterposition zu Gitterposition ändern, dann hätte man von EZ zu EZ anders geladene Bereiche - damit wäre es keine EZ mehr.

Um zu verstehen, wie die Bloch-Bedingungen die Dispersionsrelationen verändern, betrachten wir die Bandstruktur eines Elektrons in einem periodischen Gitter mit Periode a und zunächst verschwindendem Potential der Atomrümpfe, $V(x) = 0$. Dies entspricht einem freien Elektron und die entsprechende Dispersionsrelation ist in Bild 2.4(a) dargestellt. Es gilt die Wellenfunktion für das freie Elektron, welches der Blochbedingung gehorcht, zu finden. Diese Funktion ist

$$\psi_w(x) = \frac{1}{\sqrt{L}}e^{-\mathrm{j}\,kx}e^{\pm jnKx} \equiv \psi_{k,n}(x) \text{ mit } K = \frac{2\pi}{a} \tag{2.29}$$

$$\text{damit ist } u_{k,n}(x) = \frac{1}{\sqrt{L}}e^{\pm\mathrm{j}\,nKx} \tag{2.30}$$

mit $n = 0, 1, 2, ...$ Die Dispersionsrelation ergibt sich dann zu:

$$E_n(k) = \hbar\omega_n(k) = \frac{\hbar^2 \left(k \pm nK\right)^2}{2m_0}$$

Die Dispersionsrelation für das Elektron im Blochpotential ist in Fig. 2.4(b) eingezeichnet. Das Einführen eines periodischen Potentials führt also dazu, dass man einer bestimmten Wellenzahl mehrere mögliche Energiezustände zuordnen kann. Um die verschiedenen Zustände eindeutig identifizieren zu können, bekommen die so entstehenden Bänder einen Index $n = 0, 1, 2, ...$#

Wegen der Periodizität macht es Sinn, auf die reduzierte Darstellung zu wechseln, 2.4(c). Dabei wird die Energie $E(k)$ jetzt nur noch im Intervall $-\frac{\pi}{a} < k < \frac{\pi}{a}$, was als erste ***Brillouin-Zone*** bezeichnet wird, aufgetragen. Zur Konstruktion der ersten Brillouin-Zone, kann man sich vorstellen, dass man dabei die Dispersionsrelation des freien Elektrons, dargestellt als Funktion der Ausbreitungskonstanten von $-\infty < k < \infty$, auf die erste Brillouin-Zone faltet. Die Ausbreitungskonstante k, welche zum mechanischen Impuls proportional ist, wird dann durch $k \rightarrow k \pm nK$ ersetzt. Die neue Ausbreitungskonstante k ist jetzt nur noch bis auf ein Vielfaches der Gitterwellenvektors $\pm nK$ proportional zum mechanischen Impuls des Elektrons. Dieses Vorgehen ist sinnvoll, da der Phasenfaktor vor der Wellenfunktion in k periodisch ist mit dem Gitterwellenvektor $K = \frac{2\pi}{a}$. Zu jedem k-Wert findet man aber jetzt mehrere Energiebänder $W_n(k)$ mit $n = 0, 1, 2, ...$

Was passiert, wenn wir jetzt das periodische Potential der Atomrümpfe einschalten? Es entsteht Bragg-Reflexion am periodischen Gitter, wie wir es aus den Versuchen von Davisson und Germer (1927) kennen. Diese Streuungen entstehen z.B. an den Stellen $1/2K$, K, ... Man sieht, dass dies gerade bei den entartet gezeichneten Zuständen am Rand oder im Zentrum der Brillouinzone der Fall ist. An diesen Stellen kommt es zur Reflektion sowohl der nach rechts als auch der nach links laufenden Wellen, welche beide identische stehende Wellen ausbilden. Nach Fermis-Ausschschulssprinzip dürfen aber keine zwei Zustände die gleiche Energie annehmen. Die Entartung wird nun aufgehoben indem sich die Energien anheben bzw. senken. Es entstehen Energie-Lücken, Bild 2.4(d), (e) und (f). Elektronen mit einer Energie, die einem Wert in der Lücke entspricht, können in diesem Kristall nicht propagieren, sie würden durch Bragg-Reflektion ständig in die entgegengesetzte Richtung zurückreflektiert werden. (Vergleiche Bragg-Spiegel in der Optik bzw. Photonic-Bandgap-Strukturen). Die Zustände, die unmittelbar an die Energielücke anschließen entsprechen stehenden Wellen. Deshalb strebt auch die Gruppengeschwindigkeit der Teilchen bei erreichen der Grenze der Brillouinzone gegen Null, d. h.

$$v = \frac{1}{\hbar}\frac{\partial E}{\partial k}\bigg|_{\text{Grenze BZ}} = 0.$$

2.2.3 Elektronen im idealen dreidimensionalen Kristall

Dieselben Überlegungen, die wir gerade in einer Dimension aufgestellt haben, lassen sich natürlich auch für reale dreidimensionale Kristalle durchführen.

Die Bandstruktur eines freien Elektrons in einem kubisch-flächenzentrierten (fcc = face centered cubic) Kristall mit Gitterabstand a ist in Bild 2.5 dargestellt. Im Dreidimensionalen gibt es natürlich einen dreidimensionalen $\vec{k}$-Raum (Raum der Wellenvektoren) und die Energieeigenwerte $E(\vec{k})$ dieser Wellen entsprechen ineinander geschachtelten, dreidimensionalen Flächen. Um diese Schwierigkeit der Darstellung zu umgehen, werden nur bestimmte Schnitte durch diese Flächen wiederum in der Brillouin-Zone gezeichnet. Einige Punkte im Raum der Wellenvektoren sind besonders wichtig, z.B. der $\vec{\Gamma}$-Punkt der durch den Ursprungsvektor $\vec{\Gamma} = (0, 0, 0)$ definiert ist und die Punkte $\vec{L} = \frac{K}{2}(1, 1, 1)$, $\vec{X} = K(1, 0, 0)$.

Die Bilder 2.6, 2.7 und **??** zeigen die berechneten Bandstrukturen von Si, Ge und GaAs, wobei nicht nur das periodische Potential der Atomrümpfe mitberücksichtigt wurde, sondern auch die elektrostatische Energie der anderen Elektronen (Hartree-Fock-Ansatz) und die Spin-Bahn-Kopplung der Elektronen. Der Vergleich mit Bild 2.4(f) zeigt, dass die Bandstruktur realer Kristalle in vielen Zügen bereits aus der des freien Elektrons im entsprechenden Gittertyp hervorgeht.

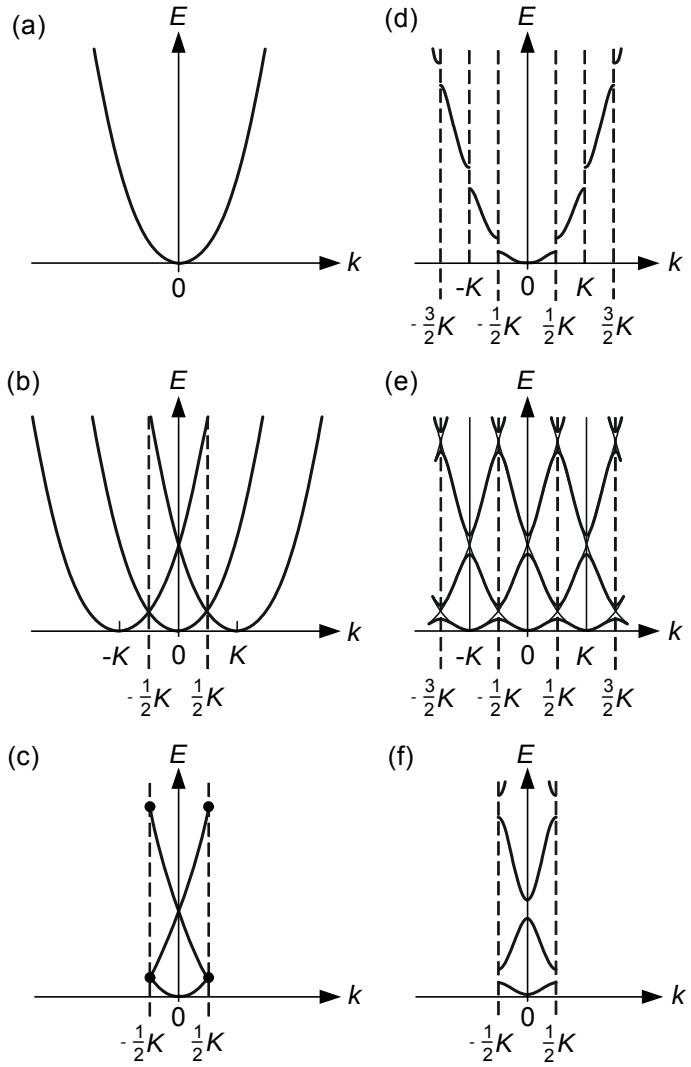


Abbildung 2.4: a) Dispersionsrelation des freien Elektrons. b) Schritt 1 in der Konstruktion der Dispersionsrelation des freien Elektrons, gestört von einem schwachen periodischen Gitter, durch Verschieben der Dispersionsrelation des freien Elektrons um ein Vielfaches des Gitterwellenvektors K. c) Reduktion auf die Darstellung in der ersten Brillouin-Zone. Die so entstandenen scheinbar entarteten Zustände, die sich eigentlich durch ein Vielfaches des Gitterwellenvektors unterscheiden, werden durch Aufnahme bzw. Abgabe einer Gitterwellenvektors bzw. Vielfachen der Gitterwellenvektors miteinander gekoppelt. Die Entartung wird aufgehoben. d) Verzerrung der ursprünglichen Dispersionsrelation. e) Darstellung im erweiterten Zonenschema. f) Reduzierte Darstellung in der ersten Brillouin-Zone. [7]

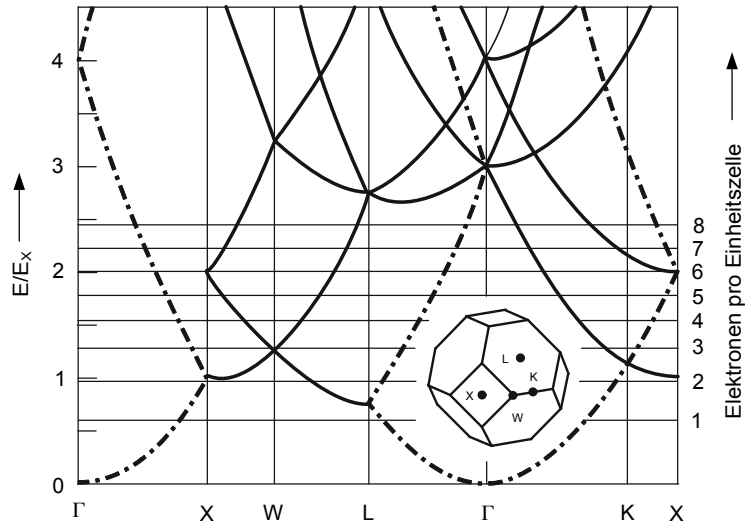


Abbildung 2.5: Energie-Niveaus eines kubisch flächenzentrierten Gitters. Die Energie ist entlang von Linien in der ersten Brillouin-Zone, welche die Punkte $\vec{\Gamma}(\vec{k}=0), \vec{K}, \vec{L}, \vec{W}$ und $\vec{X}$ miteinander verbinden, gezeichnet. W_X ist die Energie im Punkt $\vec{X}$. Die horizontalen Linien geben das Energieniveau an, bis zu welchem alle Zustände besetzt sind, falls eine gegebene Anzahl von Elektronen pro Einheitszelle untergebracht wird (Fermi-Energie). Die Anzahl der Punkte auf einer Kurve charakterisiert die Entartung der entsprechenden Niveaus [7].

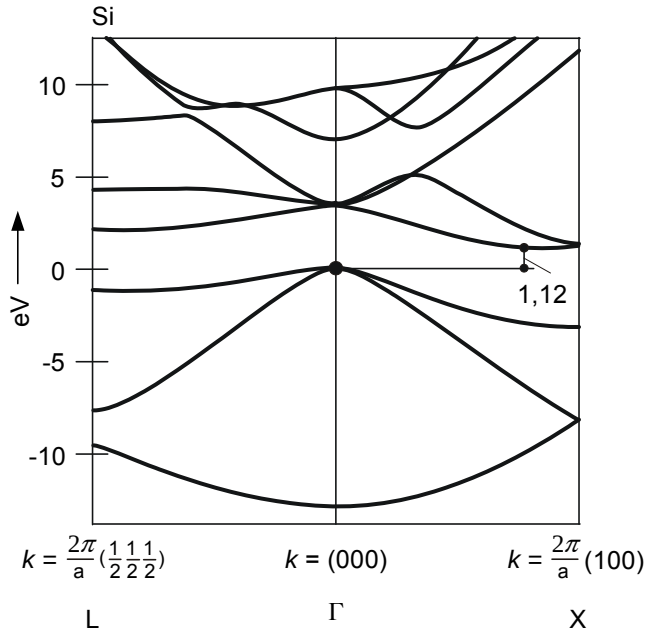


Abbildung 2.6: Berechnete Bandstruktur von Silizium.

2.2.4 Bloch-Oszillationen

Als Modellsystem für die Bewegung eines Elektrons im Leitungsband eines eindimensionalen Kristalls kann folgende cosinus-förmige Abhängigkeit der Leitungsbandenergie als Funktion der Wellenzahl betrachtet werden:

$$E(k) = \frac{\Delta E}{2}(1 - \cos(ka)). \tag{2.31}$$

Mit den Formeln (2.10) und (2.11) erhalten wir für die Gruppengeschwindigkeit und effektive Masse der Elektronen als Funktion seiner Wellenzahl k

$$\begin{aligned} \upsilon &= \frac{1}{\hbar}\frac{\partial E}{\partial k} = \frac{\Delta E a}{2\hbar}\sin(ka), \\ \frac{1}{m_{\text{eff}}} &= \frac{1}{\hbar^2}\frac{\partial^2 E}{\partial k^2} = \frac{\Delta E a^2}{2\hbar^2}\cos(ka), \end{aligned} \tag{2.32}$$

die in Bild 2.9 dargestellten funktionalen Zusammenhänge für die Gruppengeschwindigkeit und die effektive Masse für ein Elektron im Gitter.

An den Grenzen der Brillouinzone, $k = \pm\pi/a$, wird die Geschwindigkeit $\upsilon = 0\,(\partial W/\partial k = 0)$. Dort ist der Gitterabstand gerade halb so groß wie die Elektronenwellenlänge, $a = \lambda/2$; alle von Gitteratomen reflektierten Teilwellen interferieren konstruktiv (Bragg-Bedingung für Reflexionsgitter) und führen zu einer stehenden Welle. An der Oberkante des Valenzbandes ist für Elektronen $m_{\text{eff}} < 0$; dies bedeutet, dass sie in die entgegengesetzte Richtung einer äußeren auf sie wirkenden Kraft beschleunigen.

Legen wir nun eine äußere Kraft an das Elektron an, so ändert sich dessen Impuls, aber gleichzeitig wird auch Impuls ans Gitter weitergegeben (Das Elektron ist mit der Gitterwechselwirkung bekleidet). Ist die Kraft $F = konst.$, so durchläuft das Elektron die erste Brillouinzone gemäß

$$k(t) = k(0) + Ft/\hbar, \tag{2.33}$$

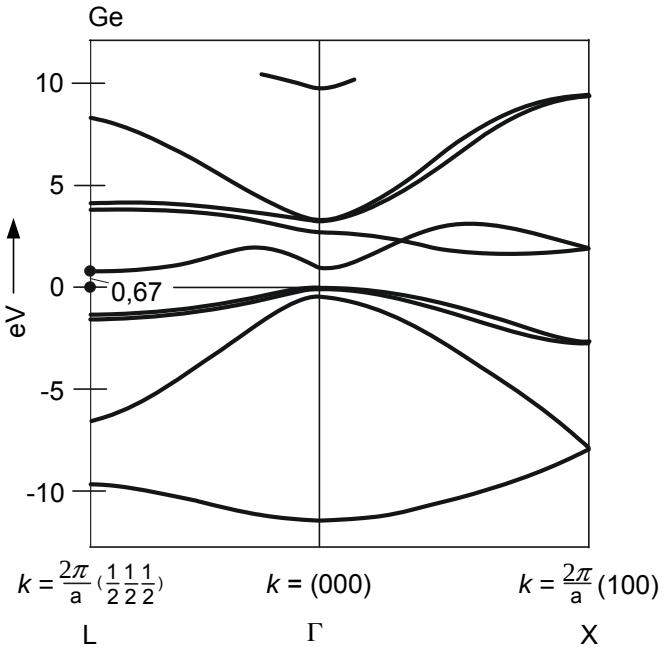


Abbildung 2.7: Berechnete Bandstruktur von Germanium.

"verschwindet" (falls $F > 0$ gilt) am rechten Rand bei $k = \pi/a$, und taucht am Punkt $k = -\pi/a$ am linken Rand der Brillouinzone wieder auf (die Werte $k = \pm\pi/a$ unterscheiden sich nur durch einen Gitterimpuls und führen daher zur selben Wahrscheinlichkeitsdichteamplitude).

Durch Einsetzen des Kristallimpulses in die Gleichung (2.32) findet man für die Gruppengeschwindigkeit und den Ort des Elektrons für ein konstantes elektrisches Feld, $E = 100\,\mathrm{kVcm}^{-1}$ respektive Anlegen einer Kraft $F = -eE$, die in Fig. 2.10 gezeigten Oszillationen.

Während das Elektron die Brillouinzone durchläuft, ändert sich seine Teilchengeschwindigkeit ($\hbar v = \partial E/\partial k$) periodisch zwischen positiven und negativen Extremwerten, es führt eine Schwingung aus (sogenannte Bloch-Oszillation). Die Zeit zum Durchlaufen des Bereichs $-\pi/a \leq k \leq \pi/a$ unter dem Einfluss einer elektrischen Feldstärke, (T = Periodendauer der Bloch-Oszillation), ist gerade:

$$T = \frac{h}{aeE}. \tag{2.34}$$

Für $a = 0.5$ nm (typische Gitterkonstante von Halbleitern), $E = 100\,\mathrm{kVcm}^{-1}$ erhält man $T = 0.8$ ps.Die sich damit ergebende Schwingung in Geschwindigkeit und Ort ist in Bild 2.10 dargestellt. So lange bleibt aber die Wellenfunktion eines Elektrons nicht unbeeinflußt von Störungen, wie z.B. Wechselwirkung mit thermisch angeregten Phononen, Gitterversetzungen, Elektron-Elektron-Streuung u.s.w. Es wird innerhalb von Bruchteilen einer Pikosekunde durch unelastische Stöße mit Gitterschwingungen in andere Zustände im gleichen Band transferiert, wobei sich Energie, Materiewellenlänge und Phase der Elektronenwelle diskontinuierlich ändern. Die Bloch-Oszillation ist in Kristallen daher nicht direkt beobachtbar. In künstlichen Kristallen, basierend auf periodischen Folgen dünner Schichten (Übergitter [16, esaki70]) kann man eine effektive Gitterkonstante a von der Größe der Periodenlänge des Übergitters vortäuschen (vielleicht $50\,\mathring{A}$ oder $500\,\mathring{A}$). Diese Bloch-Oszillationen wurden kürzlich nachgewiesen und entsprechende Bauelemente könnten in naher Zukunft eine effiziente Quelle für THz- Strahlung werden [17, leo91] [18, feldmann92].

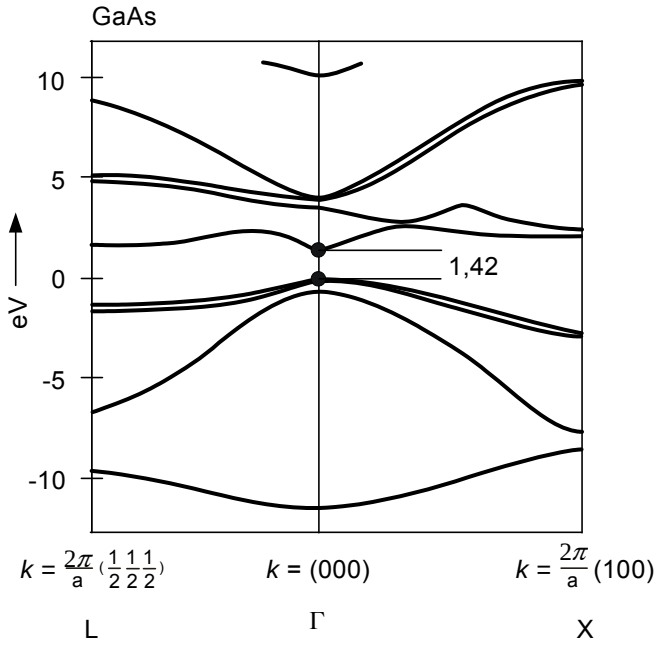


Abbildung 2.8: Berechnete Bandstruktur von GaAs.

2.2.5 Parabolische Näherung der Bandstruktur

Die Bandstruktur ist im Zentrum und an den Grenzen der ersten Brillouinzone eine symmetrische Funktion von k. In der Nähe von $k = 0$ und $k = \pm\pi/a$ läßt sich die Energie im Valenz- und Leitungsband daher durch eine Parabel approximieren.

$$E = E_0 + \frac{\hbar^2(k-k_i)^2}{2m_{\text{eff}}}, \quad \frac{1}{m_{\text{eff}}} = \frac{1}{\hbar^2}\left.\frac{\partial^2 E}{\partial k^2}\right|_{k=k_i}, \quad k_i = 0, \pm\pi/a \tag{2.35}$$

Durch Vergleich mit der Dispersionsrelation vom freien Elektron sieht man, dass sich die Kristallelektronen in der Nähe dieser ausgezeichneten Stellen wie freie Elektronen bzw. Löcher benehmen, wenn man an die Stelle der Ruhemasse m_0 die entsprechenden effektiven Massen einsetzt.

Da im Leitungsbandminimum $m_{\text{eff}} > 0$ gilt, können die dortigen Elektronen als quasifrei bezeichnet werden. Im Valenzbandmaximum dagegen ist $m_{\text{eff}} < 0$. Es ist zweckmäßig, das Konzept eines "Loches" in einem fast zur Gänze mit Elektronen besetzten Band einzuführen. Siehe Bild 2.11.

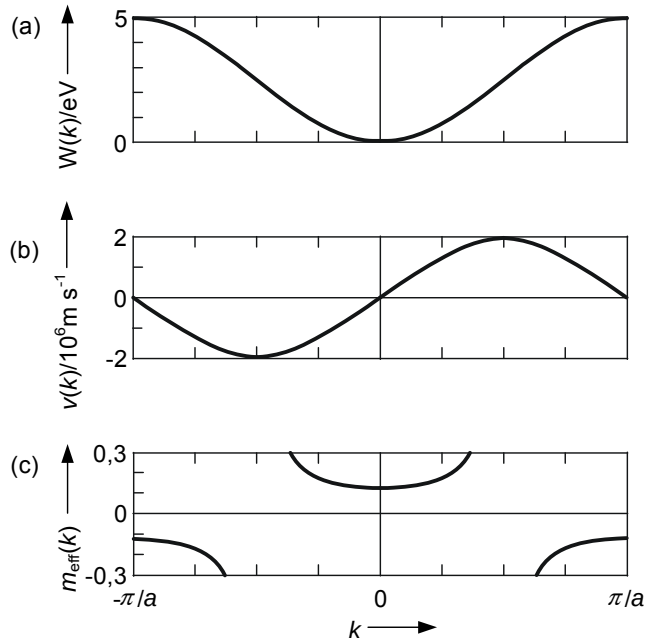


Abbildung 2.9: (a) Energie, (b) Gruppengeschwindigkeit und (c) effektive Masse als Funktion der Wellenzahl für einen cosinus-förmigigen Bandverlauf mit der Breite $\Delta W = 5\text{eV}$, in einem Kristall mit Gitterkonstante $a = 0.5\text{nm}$.

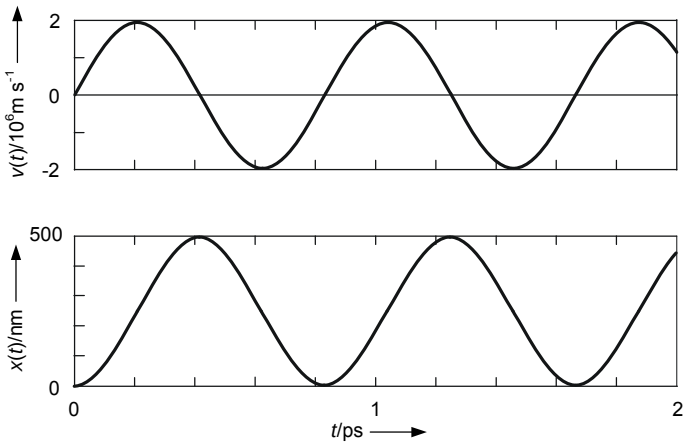


Abbildung 2.10: Blochoszillation eines Elektrons als Funktion der Zeit infolge eines äußeren angelegten Feldes der Stärke $E = 10^7\text{Vm}^{-1}$ in dem oben definierten cosinus-förmigen Leitungsband.

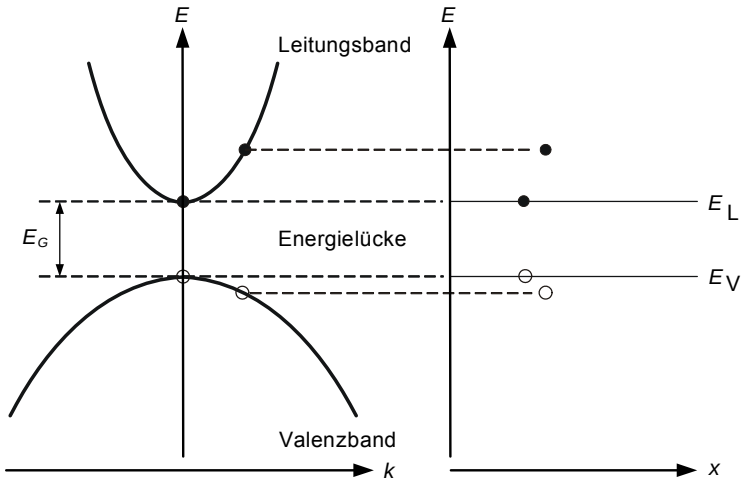


Abbildung 2.11: Parabelnäherung der Bandstruktur eines Halbleiters.

Kapitel 3

Eigenhalbleiter und dotierte Halbleiter

Bisher wurden alle Betrachtungen bei der Temperatur $T = 0$ K durchgeführt. Bei sehr tiefen Temperaturen sind in reinen Halbleitern alle Zustände im Valenzband besetzt und alle Zustände im Leitungsband unbesetzt, so dass kein Strom fließen kann. Was passiert bei endlicher Temperatur? Aufgrund der thermischen Energie kann es passieren, dass ein Elektron eine Bindung verlässt und vom Valenzband ins Leitungsband wechselt. Umgekehrt werden thermisch angeregte Elektronen mit den vorhandenen Löchern rekombinieren. Im thermischen Gleichgewicht werden pro Zeiteinheit genauso viele Bindungen durch thermische Aktivierung aufbrechen wie Rekombinationen zwischen Elektronen und Löcher aufgrund verschiedenster Prozesse stattfinden. Dies führt zu einer statistischen Besetzung der vorhandenen Elektronenzustände. Die dann nicht mehr voll besetzten Bänder erlauben bei angelegter Spannung einen Stromtransport. Damit wir zu quantitativen Aussagen kommen können, werden wir im Modell der parabolischen Bandstuktur die Anzahl der freien Ladungsträger bei gegebener Temperatur bestimmen. Wir werden dabei dotierte und undotierte Halbleiter diskutieren.

3.1 Eigenhalbleiter

Ein ***Eigenhalbleiter (intrinsic semiconductor)*** ist ein Halbleiter, der nur verschwindend kleine Spuren von Verunreinigungen enthält. Verschwindend heißt in diesem Zusammenhang etwa eine Verunreinigung auf 10^9 oder noch mehr Atome.

Im Folgenden wollen wir berechnen, wieviele freie Elektronen n_{th} und Löcher p_{th} bei gegebener Temperatur in einem Band eines Eigenhalbleiters vorhanden sind. Konkret müssen wir dabei folgende Gleichungen berechnen:

$$n_{th} = \int_{W_L}^{\infty} f(E)\, \rho_n(E)\, \mathrm{d}E \tag{3.1}$$

$$p_{th} = \int_{-\infty}^{W_V} (1 - f(E))\ \rho_p(E)\, \mathrm{d}E \tag{3.2}$$

Die Anzahl der freien Elektronen erhält man als Produkt der Wahrscheinlichkeit f, dass ein Zustand mit der Energie E besetzt ist (oder $(1-f)$, dafür dass er unbesetzt ist) und der Anzahl der Zustände ρ, welche es zu dieser Energie gibt. Desgleichen berechnet sich die Löcherkonzentration. Bevor wir diese Rechnung durchführen, werden wir die beiden Terme gesondert betrachten.

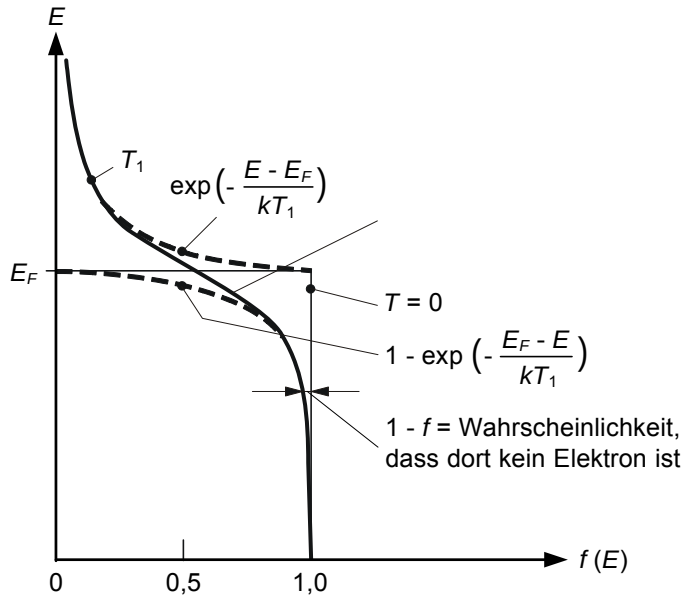


Abbildung 3.1: Fermi-Verteilung: Wahrscheinlichkeit für die Besetzung von Zuständen im Festkörper durch Elektronen als Funktion der Energie. (- - -) Boltzman-Näherungen.

3.1.1 Die Fermi-Dirac-Verteilung

Gesucht ist die Wahrscheinlichkeit $f(E)$, daß ein elektronischer Einteilchenzustand mit der Energie E besetzt ist. Konkret beobachtet man für Fermiteilchen (Elektronen und andere Teilchen welche dem Pauli-Ausschluss Prinzip genügen) eine Besetzungswahrscheinlichkeit, die der Femi-Dirac-Verteilung folgt:

$$f(E) = \frac{1}{1 + e^{(E-E_F)/kT}} \ , \tag{3.3}$$

dabei ist die Konstante E_F dadurch bestimmt, dass bei $E = E_F$ die Besetzungswahrscheinlichkeit gerade $1/2$ ist, $f(E = E_F) = 1/2$, siehe Bild 3.1.

Wichtige Spezialfälle

Für $T \to 0$ strebt diese Verteilung gegen eine Sprungfunktion, bei der nur die Zustände unterhalb des Fermi-Niveaus besetzt sind.

Weit weg von der Fermi-Energie kann die Fermi-Dirac-Verteilung durch eine Boltzmann-Verteilung (***Boltzmann-Näherung***) angenähert werden.

$$\begin{array}{lll} E \gg E_F: & f_c(E) & \cong \ e^{-(E-E_F)/kT} \\ E \ll E_F: & f_v(E) \equiv 1 - f_c(E) & \cong \ e^{(E-E_F)/kT} = e^{-(E_F-E)/kT} \end{array} \tag{3.4}$$

wobei $1 - f(E)$ die Wahrscheinlichkeit für ein Loch darstellt. Für $|(E - E_F)/kT| > 3$ sind die Näherungen besser als 5%. Physikalisch ist E_F dadurch bestimmt, dass diese Statistik für ein gegebenes System im Mittel die Zahl der im System vorhandenen Elektronen richtig wiedergeben muss. Dazu betrachten wir ein Gas freier Elektronen.

Ergänzender Kommentar zur Fermi-Dirac Verteilung Für all jene, die es etwas genauer wissen wollen seien hier ergänzende Erläuterungen angebracht. Aus der statistischen Mechanik wissen wir, dass die Wahrscheinlichkeit,für die Besetzung eines Zustandes der Energie E_n gegeben

ist durch

$$P(E_n) = \frac{g_n}{Z} e^{-E_n/kT}. \tag{3.5}$$

Dabei ist g_n der Entartungsfaktor des Energieeigenwerts E_n, d. h. wieviele verschiedene Zustände es zur selben Energie E_n gibt. Z heißt Zustandssumme und ist der Normierungsfaktor, damit die Wahrscheinlichkeit, dass das System in irgend einem Zustand ist, auf 1 normiert ist

$$\sum_n P(E_n) = 1 \quad \text{d. h.} \tag{3.6}$$

$$Z = \sum_n g(E_n) e^{-E_n/kT}, \tag{3.7}$$

wobei die Summe über alle Energieeigenwerte des Gesamtsystems zu bilden ist. Gl. (3.5) ist bereits aus der klassischen statistischen Mechanik als Arrhenius-Gesetz, Boltzmann-Verteilung oder kanonische Verteilung bekannt. Sie besagt letzlich, dass die Wahrscheinlichkeit für die Besetzung eines Zustandes exponentiell mit der Erhöhung der Energie abnimmt..Auch besagt die Gleichung, dass der exponentielle Abfall von der Temperatur abhängt. Je höher die Temperatur desto flacher die Kurve. Die Quanteneigenschaften der Elektronen als nichtunterscheidbare Teilchen und Fermionen werden dabei durch das Ausschließungs-Prinzip von Pauli berücksichtigt. Wir wollen die Statistik $P(E_i)$ in ihrer Gesamtheit nicht untersuchen. Wechselwirken die Teilchen nicht miteinander, was wir hier voraussetzen wollen, so können die Einteilchenzustände getrennt betrachtet werden. Ein spezieller Einteilchen-Energieeigenzustand mit Energie E kann im Fall von Fermionen besetzt, $n = 1$, oder unbesetzt, $n = 0$, sein und die entsprechenden Wahrscheinlichkeiten sind

$$n = 0 : P(0) = \frac{1}{Z}. \tag{3.8}$$

$$n = 1 : P(E) = \frac{1}{Z} e^{-E/kT}. \tag{3.9}$$

Die Summe dieser beiden Wahrscheinlichkeiten muß eins sein, was $Z = 1 + e^{-E/kT}$ liefert. Die gesuchte Wahrscheinlichkeit, dass der Einteilchen-Zustand mit Energie E besetzt ist, ist dann

$$f(E) = P(E) = \frac{e^{-E/kT}}{1 + e^{-E/kT}} = \frac{1}{1 + e^{E/kT}} \tag{3.10}$$

Da die Energie nur bis auf eine Konstante bestimmt ist, können wir auch E durch $E - E_F$ ersetzen und wir erhalten die gesuchte Fermi-Dirac-Statistik

$$f(E) = \frac{1}{1 + e^{(E-E_F)/kT}}. \tag{3.11}$$

3.1.2 Zustandsdichten

Es sei die Zustandsdichte $\rho(E) = (1/V)\, \mathrm{d}N/\mathrm{d}E$, d. h. die Anzahl der quantenmechanisch erlaubten und besetzbaren Zustände pro Volumen $V = L^3$ in einem Energieinterval dE zu berechnen. Die Zustandsdichte ist wichtig, weil es in einem quantenmechanischen System (begrenzten Raum) nur endlich viele erlaubte Zustände gibt (Quantisierung) und diese, mit den tiefen Energien beginnend, höchstens einfach (Fermistatistik) besetzt werden dürfen.

Da sich die Zustandsdichte im $\vec{k}$-Raum einfacher berechnen lässt, schreiben wir stattdessen

$$\rho(E) = \frac{1}{V} \frac{\mathrm{d}N}{\mathrm{d}E} = \frac{1}{V} \frac{\mathrm{d}N}{\mathrm{d}k} \frac{\mathrm{d}k}{\mathrm{d}E}. \tag{3.12}$$

Wir beginnen mit dem Berechnen der Zustandsdichte $\rho(k) = \mathrm{d}N/\mathrm{d}k$ im $\vec{k}$-Raum. Dazu benötigen wir zuerst einen Audruck für $N(k)$ der Anzahl aller erlaubten Zustände zu einem gegebenen k. Die Dichte der k-Raum Zustände ergibt sich aus den periodischen Bandschema-Randbedingungen

$$\vec{k}L = 2\pi\ (n, m, l)\,, \quad n, m, l = 0, \pm 1, \pm 2, \ldots \tag{3.13}$$

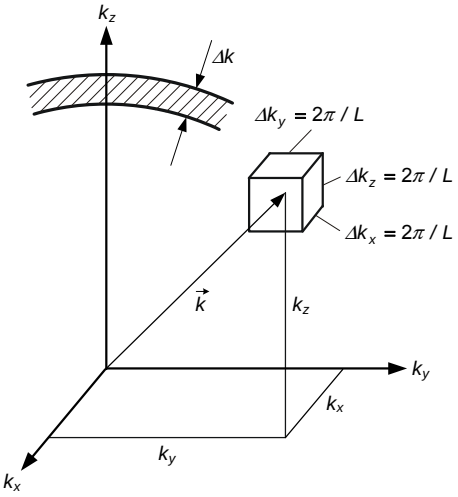


Abbildung 3.2: Zustandsdichte im k-Raum.

oder anders formuliert

$$k_i = i \cdot \frac{2\pi}{L}, \quad i = n, m, l \ . \tag{3.14}$$

Das bedeutet, dass es pro $\vec{k}$-Raum Volumeneinheit $V_k = (2\pi/L)^3$ einen erlaubten k-Zustand gibt. Damit ergibt sich für die Anzahl der Zustände $N(k)$, aus dem Volumen der $\vec{k}$-Raum Kugel, dividiert durch das Volumen $(2\pi/L)^3$, welches für einen einzigen k-Raum Zustand beansprucht wird.

$$N(k) = 2 \times \frac{V_{k-Kugel}}{V_k} = 2 \times \frac{\frac{4\pi}{3}k^3}{\left(\frac{2\pi}{L}\right)^3} \ , \tag{3.15}$$

wobei der Faktor 2 in Gl. (3.15) hinzugefügt wurde, um den beiden möglichen Spineinstellungen der Fermionen Rechnung zu tragen. $\rho(k) = \mathrm{d}N/\mathrm{d}k$ ergibt sich dand durch Ableiten

$$\frac{\mathrm{d}N}{\mathrm{d}k} = 2 \times \frac{3\frac{4\pi}{3}k^2}{\left(\frac{2\pi}{L}\right)^3} \ , \tag{3.16}$$

Die Geometrie der Anordnung ist in 3.2 noch einmal verdeutlicht.

Wir brauchen noch $\mathrm{d}k/\mathrm{d}E$. Diese erhalten wir aus der Dispersionsrelation des Elektrons in der parabolischen Näherung mit $\left|\vec{k}\right| = k$

$$E = \frac{\hbar^2 k^2}{2m_0} \tag{3.17}$$

Durch ableiten findet man

$$\frac{\mathrm{d}k}{\mathrm{d}E} = \frac{1}{2}\sqrt{\frac{2m_{\mathrm{eff}}}{\hbar^2}}\frac{1}{\sqrt{E}} \ . \tag{3.18}$$

Damit folgt für die Zustandsdichte des freien Elektronengases, d. h. die Anzahl der Zustände pro Volumeneinheit und Energieintervall $\rho\,(E)$

$$\rho\,(E) = \frac{1}{V}\frac{\mathrm{d}N}{\mathrm{d}k}\frac{\mathrm{d}k}{\mathrm{d}E} = \frac{1}{L^3}\frac{8\pi k^2}{\left(\frac{2\pi}{L}\right)^3}\frac{1}{2}\sqrt{\frac{2m_{\mathrm{eff}}}{\hbar^2}}\frac{1}{\sqrt{E}} = \frac{4\pi\left(2m_{\mathrm{eff}}\right)^{3/2}}{h^3}\sqrt{E}. \tag{3.19}$$

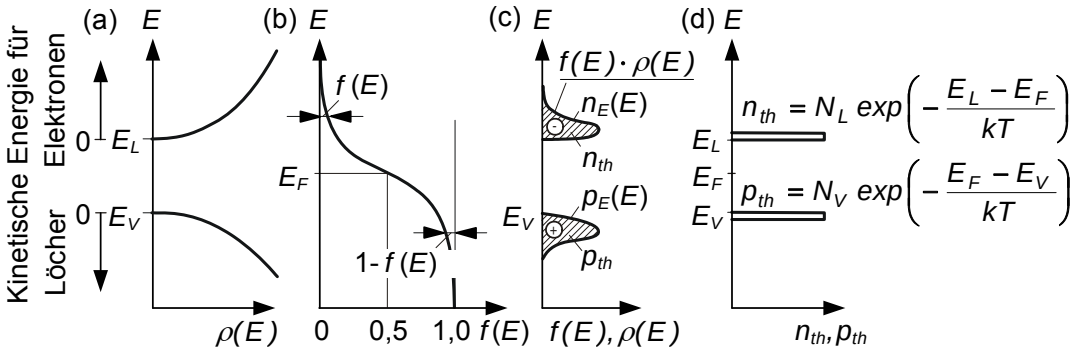


Abbildung 3.3: Bestimmung der Eigenleitungsträgerdichten [3]. (a) Zustandsdichten, (b) Fermi-Dirac Verteilung, (c) Produkt aus Zustandsdichte mit Fermi-Dirac-Verteilung und (d) Ladungsträger in den Bändern.

Wie wir vorher gesehen haben, bewegen sich Elektronen und Löcher im Halbleiter wie freie Elektronen, aber mit einer effektiven Masse. Die entsprechenden Zustandsdichten im Valenz- und Leitungsband für Elektronen und Löcher sind daher

$$\rho_n = \frac{4\pi\,(2m_n)^{3/2}}{h^3}\sqrt{E - E_L}, \tag{3.20}$$

$$\rho_p = \frac{4\pi\,(2m_p)^{3/2}}{h^3}\sqrt{E_V - E}. \tag{3.21}$$

Mit Hilfe der sogenannten ***äquivalenten Zustandsdichten (effective density of states)*** für Leitungs- und Valenzband N_L und N_V

$$N_L = 2\left(\frac{2\pi m_n kT}{h^2}\right)^{3/2},\ \text{bzw. } N_V = 2\left(\frac{2\pi m_p kT}{h^2}\right)^{3/2} \tag{3.22}$$

können die Zustandsdichten als Funktion der Energie kompakter geschrieben werden als

$$\rho_n = \frac{2}{\sqrt{\pi}}\frac{N_L}{kT}\sqrt{\frac{E - E_L}{kT}},\ \text{bzw. } \rho_p = \frac{2}{\sqrt{\pi}}\frac{N_V}{kT}\sqrt{\frac{E_V - E}{kT}}. \tag{3.23}$$

Dies entspricht einer Normierung der Energieskala auf die thermische Energie kT, d. h. $\int_{E_L}^{E_L+kT}\rho_n dE = 0.75N_L$.

3.1.3 Ladungträgerkonzentrationen im Eigenleiter

In Bild 3.3 sind in (a) die Zustandsdichten für Elektronen und Löcher in einem reinen Halbleiter gemäß den Gln.(3.20) und (3.21) skizziert. Die Wahrscheinlichkeiten, dass diese Zustände besetzt sind und die daraus resultierenden Elektronen- und Löcherverteilungen $\rho_n(E)f(E)$ und $\rho_p(E)\,[1 - f(E)]$ sind in (b) bzw. (c) dargestellt.

Für die Löcher- und Elektronen-Gesamtdichten n_{th} und p_{th} gilt:

$$n_{th} = \int_{E_L}^{\infty}\rho_n(E)f_c(E)dE \qquad p_{th} = \int_{-\infty}^{E_V}\rho_p(E)\,[1 - f_c(E)]\,dE \tag{3.24}$$

Die bei Eigenleitung resultierenden Elektronen- und Löcherdichten pro Energieintervall sind aufgrund der Fermi-Verteilung an den Bandkanten lokalisiert und nur innerhalb eines Energieintervalls, welches der thermischen Energie kT entspricht, von Null verschieden. Die Energie $E_{th} = kT = 25\text{meV}$ für $T = 293\text{K}$ ist bei allen Transportprozessen in Festkörpern von größter

Bedeutung und bei den nützlichen Halbleitern klein gegen die Bandlücke $E_G \approx 1eV = 40W_{th}$. Die thermische Energie bestimmt offensichtlich, wieviele bewegliche Ladungsträger beim eigenleitenden Material vorhanden sind.

Falls die Fermienergie weit weg von der Bandkante ist, gilt für die Löcher- und Elektronen-Gesamtdichten n_{th} und p_{th} unter Ausnutzung der Boltzmann-Näherung für die Fermi-Verteilung nach den Gln.(3.4) und (3.23), wobei der Index th für thermisches Gleichgewicht steht:

$$n_{th} = N_L \exp\left[-\tfrac{E_L - E_F}{kT}\right] \qquad p_{th} = N_V \exp\left[-\tfrac{E_F - E_V}{kT}\right] \tag{3.25}$$

Dabei wurde das Integral $\int_0^\infty \sqrt{x}\exp(-x)\,dx = \sqrt{\pi}/2$ verwendet. Die äquivalenten Zustandsdichten sind also Zustandsdichten, die man sich unmittelbar an den Bandkanten lokalisiert vorstellen kann, siehe Bild 3.3(d).

Massenwirkungsgesetz

Das Produkt aus den thermisch angeregten Elektronen und Löchern ist durch

$$n_{th}p_{th} = n_i^2(T) = N_L N_V \exp\left(-\frac{E_G}{kT}\right) \tag{3.26}$$

gegeben. $n_i(T)$ heißt ***Eigenleitungsträgerdichte (intrinsic carrier density)*** und hängt nach (3.26) gemäß einem Ahreniusgesetz von der Bandlückenenergie ab. Die Beziehung in (3.26) $n_{th}p_{th} = n_i^2$ heißt ***Massenwirkungsgesetz (mass action law)*** und kann wie folgt interpretiert werden: $n_{th}p_{th}$ ist proportional zur Rekombinationsrate von Elektronen im LB mit Löchern im VB, n_i^2 ist proportional zur Generationsrate von Trägerpaaren zufolge des Aufreißens von Valenzbindungen. Gleichgewicht herrscht dann, wenn Rekombinationsrate und Generationsrate gleich groß sind. Das Produkt $n_{th}p_{th}$ ist nur dann unabhängig von der Fermi-Energie E_F (allein abhängig vom Bandabstand E_G), wenn die Konzentrationen klein sind gegen die Bandgewichte, d.h. wenn das Fermi-Niveau hinreichend weit von den Bandkanten entfernt ist, so daß die Boltzmann-Näherung für die Fermi-Verteilung gerechtfertigt ist. Diesen Fall bezeichnet man als "Nichtentartung" (Gegensatz: entarteter Halbleiter).

Da im Eigenhalbleiter Elektronen im LB und Löcher im VB paarweise entstehen, ist

$$n_{th} = p_{th} = n_i. \tag{3.27}$$

Beispiel: $T = T_0 = 293\,\mathrm{K}$

Ge	$n_i = 2{,}4 \cdot 10^{13}\,\mathrm{cm}^{-3}$
Si	$n_i = 1{,}5 \cdot 10^{10}\,\mathrm{cm}^{-3}$
InP	$n_i = 1{,}2 \cdot 10^{8}\,\mathrm{cm}^{-3}$
GaAs	$n_i = 1{,}8 \cdot 10^{6}\,\mathrm{cm}^{-3}$

Die Eigenleitungsträgerdichte für die drei wichtigsten Halbleiter sind in Bild 3.4 als Funktion der Temperatur dargestellt. Wie nachfolgend gezeigt wird, ist in Halbleitern mit hohem Bandabstand der Fall der Eigenleitung nicht erreichbar, da die Anzahl der Ladungsträger liefernden Störstellen auch bei höchster Reinheit groß gegen n_i ist.

Lage des Fermi-Niveaus beim intrinsischen HL

Aus (3.27) erhält man durch Einsetzen von n_{th} und p_{th} aus (3.26) unter Beachtung der Beziehung (3.25) für die Bandgewichte eine Beziehung für die Lage des Fermi-Niveaus für einen eigenleitenden Halbleiter

$$E_F = \frac{1}{2}(E_V + E_L) + kT\ln\sqrt{\frac{N_V}{N_L}} = \frac{1}{2}(E_V + E_L) + \frac{3}{4}kT\ln\left(\frac{m_p}{m_n}\right) = E_i. \tag{3.28}$$

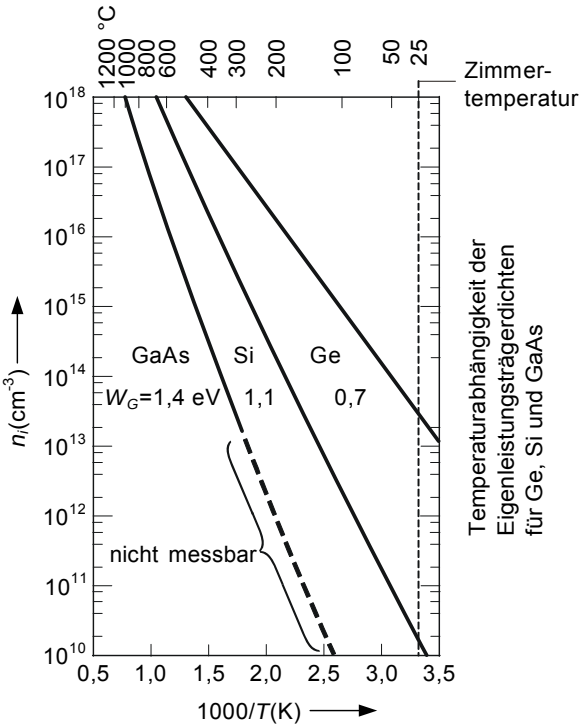


Abbildung 3.4: Temperaturabhängigkeit der Eigenleitungsträgerdichten für Ge, Si und GaAs. [3].

Für $T \to 0$ liegt das Fermi-Niveau im Eigenhalbleiter in der Mitte des verbotenen Bandes. Für hohe Temperaturen nähert sich das Ferm-Niveau dem Band mit dem kleineren Bandgewicht (dem Band der kleineren effektiven Masse), weil bei $n = p$ das Band mit der kleineren Zustandsdichte schneller gefüllt wird und sich dadurch die Lage des Zustands, der mit der Wahrscheinlichkeit 1/2 mit einem Träger besetzt ist, näher an dieses Band rückt.

Die Energie E_i bezeichnet die Lage des Fermi-Niveaus bei einem Eigenhalbleiter ('i' für intrinsisch).

3.2 Dotierte Halbleiter

Die moderne Halbleitertechnologie wurde durch Prozesse zur Herstellung reinster Halbleitermaterialien ermöglicht. Dies ist von größter Wichtigkeit, da die Kontamination mit Fremdatomen zusätzlich Elektronen freisetzen kann. Beim Zusetzen von Fremdatomen geht der Eigenhalbleiter in einen Fremdhalbleiter (extrinsischen) über.

- Um z. B. in Silizium bei Raumtemperatur Eigenleitung zu erreichen, muss die Eigenleitungsträgerdichte $n_i = 1.5 \cdot 10^{10}\text{cm}^{-3}$ größer sein als die Dichte der Fremdatome unter der Voraussetzung, dass jedes Fremdatom nur ein Elektron beiträgt. Für Silizium und Raumtemperatur folgt damit, daß auf $\left(8/\left(5 \cdot 10^{-8}\right)^3\right) / \left(1.5 \cdot 10^{10}\right) = 4 \cdot 10^{14}$ Si-Atome nur ein Fremdatom kommen darf. Es werden heute Reinheitsgrade von $10^{10} - 10^{12}$ Fremdatomen pro cm^3 erreicht.

- Umgekehrt kann bei so hochreinem Silizium die Trägerkonzentration dann reproduzierbar durch kontrollierte Zugabe von entsprechend wenigen Fremdatomen während des Kristallisationsvorganges gezielt verändert werden.

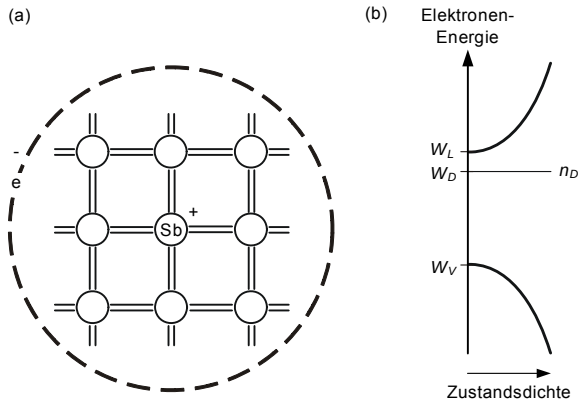


Abbildung 3.5: (a) Antimon-Donatoratom (V. Hauptgruppe) substituiert ein Silizium Atom. (b) Lage des Donatorniveaus in der Bandlücke [6]

Den Vorgang der gezielten Veränderung der Ladungsträgerdichten durch Hinzugabe von Fremdatomen nennt man ***Dotieren***.

3.2.1 Donatoren und Akzeptoren

Bei Halbleitern werden die Valenzbindungen zu den nächsten vier Nachbaratomen im Mittel durch vier Elektronen pro Atom in den äußersten Schalen gebildet. Dem entsprechen in den äußersten Schalen folgende Elektronenkonfigurationen, siehe Bilden 1.4 und 1.9.

Elementhalbleiter	Ge, Si (s^2p^2)
III-V-Halbleiter	Ga (s^2p^1), As (s^2p^3)
II-VI-Halbleiter	Zn (s^2), Se (s^2p^4)

Donator Dotierungen

Bei ***Donatoren*** wird ein Kristall-Atom durch ein anderes substituiert, welches ein zusätzliches Elektron in der äußersten Schale enthält (Substitutionsstörstelle). Dieses kann dann mit geringer Energiezufuhr (je nach Atom im Bereich von $0{,}2\,kT_0$ bis $2\,kT_0$, $kT_0 = 25\,\mathrm{meV}$) an das LB abgegeben werden. Damit wird $n > p$ und man erhält einen ***n-Halbleiter (Überschusshalbleiter)***, siehe Bild 3.5.

Akzeptor Dotierungen

Wird dagegen ein Atom durch ein anderes substituiert, welches ein Elektron weniger in der äußersten Schale enthält, so nimmt dieses mit geringer Energiezufuhr ein Elektron aus einer Valenzbindung auf, wodurch ein Loch im VB entsteht. Damit wird $p > n$ und man erhält einen ***p-Halbleiter (Defekthalbleiter)***. Diese Störstelle heißt ***Akzeptor***, siehe Bild 3.6.

Amphotere Dotierungen

In einem III-V-Halbleiter kann ein Element der IV-ten Gruppe (z. B. Si) sowohl als Donator (auf einem Ga-Platz), als auch als Akzeptor (auf einem As-Platz) eingebaut werden. Man bezeichnet solche Störstellen als amphoter (Altgriech. “zwitterhaft”).

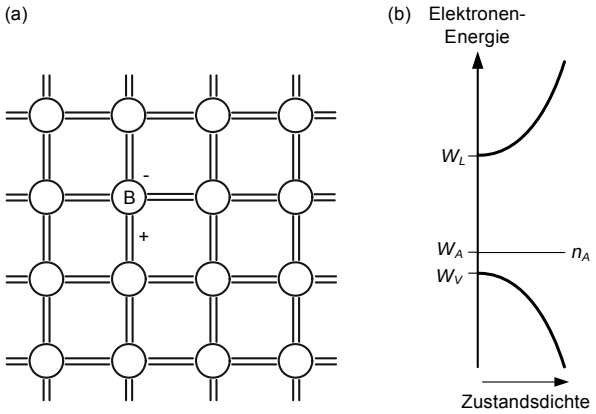


Abbildung 3.6: (a) Bor-Akzeptoratom (III. Hauptgruppe) substituiert ein Silizium Atom. (b) Lage des Akzeptorniveaus in der Bandlücke [6]

3.2.2 Lage der Energieniveaus bei verschiedenen Dotanden

Leitfähigkeitsdotierung

Mit Dotanden, deren Abstand von von der Bandkante in der Größenordnung von $E_L - E_D \approx kT$, $E_A - E_V \approx kT$ sind, läßt sich die Leitfähigkeit durch Dotierung beeinflussen (Leitfähigkeitsdotierung).

Die Lage der Donator- und Akzeptorniveaus ergibt sich aus der notwendigen Ionisierungsenergie der Störstellen und kann daher aus der Ionisierungsenergie eines äquivalenten Wasserstoffatoms

$$E_H = \frac{m_0 e^4}{8\varepsilon_0^2 h^2} = 13,6 \text{ eV} \tag{3.29}$$

abgeschätzt werden. Für einen Donator oder Akzeptor tritt an die Stelle der Masse des freien Elektrons die effektive Masse von Elektron oder Loch im Kristall, und die Dielektrizitätskonstante des Vakuums ist durch die Dielektrizitätskonstante des Kristalls zu ersetzen. In Si ergibt sich z.B. ($m_n/m_o \sim 1,\ m_p/m_o \sim 0.5,\ \varepsilon = \varepsilon_0\varepsilon_{\text{Si}};\ \varepsilon_{\text{Si}} = 12$)

$$\begin{array}{ll} \text{Donator} & \text{Akzeptor} \\ \Delta E_D = \frac{m_n}{m_0}\frac{1}{\varepsilon_{\text{Si}}^2} E_H = 100 \text{ meV}, & \Delta E_A = \frac{m_p}{m_0}\frac{1}{\varepsilon_{\text{Si}}^2} W_H = 50 \text{ meV}. \end{array} \tag{3.30}$$

Diese Abschätzung gibt nur die Größenordnungen richtig wieder, die genaue Lage des Energieniveaus hängt von der Größe des Fremdatoms, dessen chemischen Eigenschaften (wie ist die Ladung des Atomrumpfes des Donators bzw. Akzeptors abgeschirmt), sowie den Eigenschaften des Wirtskristalls ab. Bild 3.7 zeigt die Lage von wichtigen Donator- und Akzeptorniveaus in den drei wichtigsten Halbleitern.

Semiisolierende Dotierung, Lebensdauerdotierung

Ist das Donator- bzw. Akzeptor-Niveau mehr in der Bandmitte (von den Bandkanten viele kT_0 entfernt), so wird auch das Fermi-Niveau dorthin gezogen, und die Leitfähigkeit nimmt ab, bzw. ändert sich gegenüber dem intrinsischen Halbleiter nicht wesentlich. Auf diese Weise kann z.B. durch Dotierung von n-halbleitendem GaAs mit Cr, siehe Bild 3.7, ***semiisolierendes*** GaAs hergestellt werden.

Solche Störstellen in der Mitte des verbotenen Bandes sind zwar nicht leitend, aber sehr wirksam für die Rekombination von Elektronen aus dem LB mit Löchern aus dem VB (Rekombinationsstörstellen; sie helfen, schneller ein Gleichgewicht zwischen Elektronen im LB und Löchern

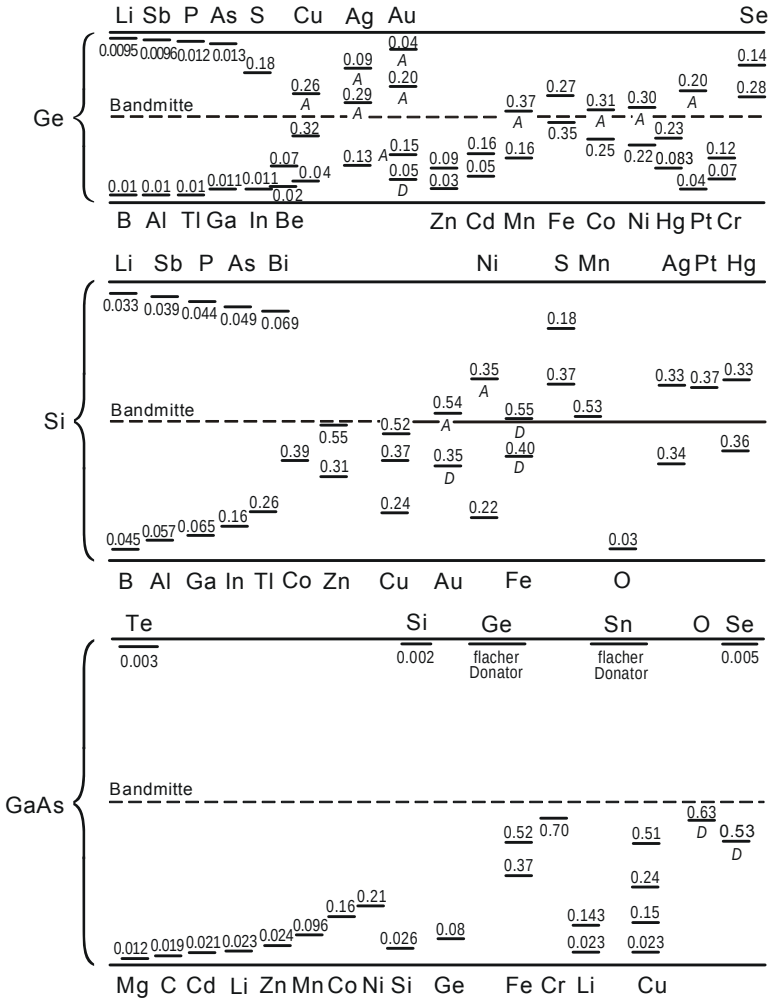


Abbildung 3.7: Gemessene Ionisierungsenergien einiger Dotierungselemente in Ge, Si und GaAs, die Niveaus. Die Niveaus unter der Bandmitte sind von der oberen Bandkante des Valenzbandes gemessen und stellen Akzeptorniveaus dar, wenn nicht durch D (Donator) gekennzeichnet; die Niveaus über der Bandmitte sind von der unteren Bandkante des Leitungsbandes gemessen und stellen Donatorniveaus dar, wenn nicht durch A (Akzeptor) gekennzeichnet [3].

im VB herzustellen, d. h. sie verringern die Trägerlebensdauer). Bei einer gezielten Dotierung mit solchen Materialien spricht man von Lebensdauerdotierung (z. B. Au in Si) .

3.2.3 Donator Konzentrationen

Kleine und mittlere Dotierungen

Donatoren haben bei nicht zu hohen Dotierungen eine relativ lokalisierte Lage im Energiniveau. In dem Sinne braucht man sich nicht um Zustandsdichten bei verschiedenen Energien zu kümmern. Wir wollen folgende Notation verwenden:

- Die Konzentrationen der Donatoren (Akzeptoren) sei n_D, (n_A),
- Die Konzentrationen der ionisierten Störstellen (der Donatoren, die ein Elektron abgegeben haben; der Akzeptoren, die ein Elektron angenommen haben) sei n_D^+, (n_A^-),
- und die Konzentrationen der neutralen Störstellen (Donatoren, die kein Elektron abgegeben haben; Akzeptoren, die kein Elektron angenommen haben) seien $n_D^\times$, $(n_A^\times)$,

Damit lauten die Störstellenbilanzen

$$\begin{aligned} n_D &= n_D^\times + n_D^+, \\ n_A &= n_A^\times + n_A^-. \end{aligned} \tag{3.31}$$

Entartete Halbleiter

Bei hohen Dotierungen entsteht durch die Wechselwirkung zwischen Dotierungsatomen eine Aufspaltung der Donator- bzw. Akzeptorterme, wie in Bild 3.8 dargestellt. Wie wir weiter unten sehen werden, verschiebt sich das Fermi-Niveau in das Leitungsband (n-Dotierung) oder in das Valenzband (p-Dotierung) und der Halbleiter zeigt metallähnliches Verhalten. Man spricht von einem ***entarteten Halbleiter***. Die Leitfähigkeit bleibt auch bei tiefen Temperaturen erhalten. Entartete Halbleiter finden Anwendung in der Tunneldiode und als Kontaktzone im Transistor, wie später noch gezeigt wird.

In extrinsischen bzw. entarteten Halbleitern gibt es also folgende Situationen:

	n-Halbleiter	p-Halbleiter
	$n_n \gg p_n$	$p_p \gg n_p$
Majoritätsträger	Elektronen: n	Löcher: p
Minoritätsträger	Löcher: p	Elektronen: n

3.2.4 Fermiverteilung der Störstellen

Zur Berechnung der in das System zusätzlich eingebrachten freien Ladungsträger infolge der Dotierung brauchen wir neben den Dichten der Donatoren oder Akzeptoren auch noch die Wahrscheinlichkeiten für deren Aktivierung, d.h. dass die Donatoren ionisiert sind und die Akzeptoren ein Elektron aufgenommen haben. Diese ergeben sich gemäß Abschnitt 3.1.1, wobei jetzt verschiedene Entartungsfaktoren für den elektronenreicheren bzw. elektronenärmeren Zustand, je nachdem, ob es sich um einen Akzeptor oder Donator handelt, in Betracht gezogen werden müssen, siehe Bild 3.9.

Im Falle des unbesetzten Donators D^+ (n=0) sind alle Elektronen eingebunden und deren Spin nicht frei wählbar. Im Gegensatz dazu ist beim besetzten Donator D^x(n=1) ein Elektron vorhanden, dessen Spin frei wählbar ist. Es folgt daher für die beiden Zustände, n=0, 1 eine Besetzung g=1, 2. Die Wahrscheinlichkeit für Besetzung bei den Donatoren ergibt sich dann als Quotient von

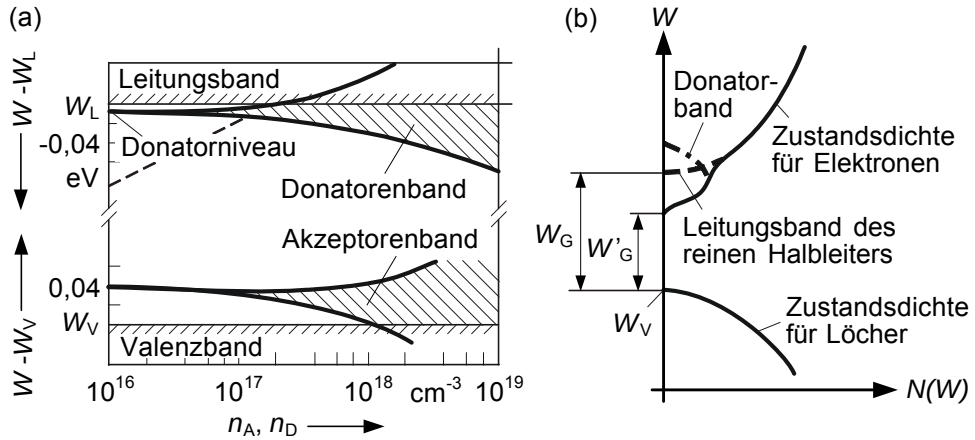


Abbildung 3.8: (a) Aufspaltung von Störstellenniveaus zu Störbändern und Überlappung mit dem Valenz- und Leitungsband in GaAs. (b) Zustandsdichte für einen n-degenerierten Halbleiter [3].

Besetzungswahrscheinlichkeit für n=1: $2 \cdot \exp\left(-(E_D - E_F)/kT\right)$ und der totalen Wahrscheinlchkeit für unbesetzt und besetzt: $2 \cdot 1 + 2 \cdot \exp\left((E_F - E_D)/kT\right)$. Beim Akzeptor errechnet sich die W'keit für Besetzung ähnlich. Die Situation ist in der Tabelle noch einmal aufgeführt:

Donator
$n = 0, \quad g = 1 \Rightarrow$ Besetzungsw'keit 1·1
$n = 1, \quad g = 2 \Rightarrow$ Besetzungsw'keit $2 \cdot \exp\left(-(E_D - E_F)/kT\right)$
W'keit für $n = 1$:

$$f_{D^x}(E_D) = \frac{2\exp\left(-\dfrac{E_D - E_F}{kT}\right)}{1 + 2\exp\left(-\dfrac{E_D - E_F}{kT}\right)}$$

Akzeptor
$n = 0, \quad g = 2 \Rightarrow$ Besetzungsw'keit 2·1
$n = 1, \quad g = 1 \Rightarrow$ Besetzungsw'keit $1 \cdot \exp\left(-(E_A$
W'keit für $n = 1$:

$$f_{A^-}(E_A) = \frac{\exp\left(-\dfrac{E_A - E_F}{kT}\right)}{2 + \exp\left(-\dfrac{E_A - E_F}{kT}\right)} \tag{3.32}$$

Im Falle des Akzeptors gilt gerade das Umgekehrte. Für eine detailliertere Diskussion der Verteilungsfunktionen für Akzeptor und Donator siehe [10] Daraus folgt, dass ein bei der Energie E vorhandener Platz für ein Elektron im thermodynamischen Gleichgewicht mit der Wahrscheinlichkeit

$$f_{B,D^x,A^-}(E) = \frac{1}{1 + 1/g \exp\left(\dfrac{E - E_F}{kT}\right)} \qquad g = \begin{cases} 1 & \textbf{B}\text{andzustände} \\ 2 & \textbf{D}\text{onatoren} \\ 1/2 & \textbf{A}\text{kzeptoren} \end{cases} \tag{3.33}$$

besetzt ist. Der Parameter E_F in der Fermi-Funktion für Donatoren und Akzeptoren hat die Bedeutung, dass die Wahrscheinlichkeit für die Besetzung eines Donators oder Akzeptors mit der Energie E_F gleich 2/3 bzw 1/3 ist, entsprechend des Verhältnises von Anzahl der möglichen Zustände bei Besetzung zur Gesamtzahl der möglichen Zustände.

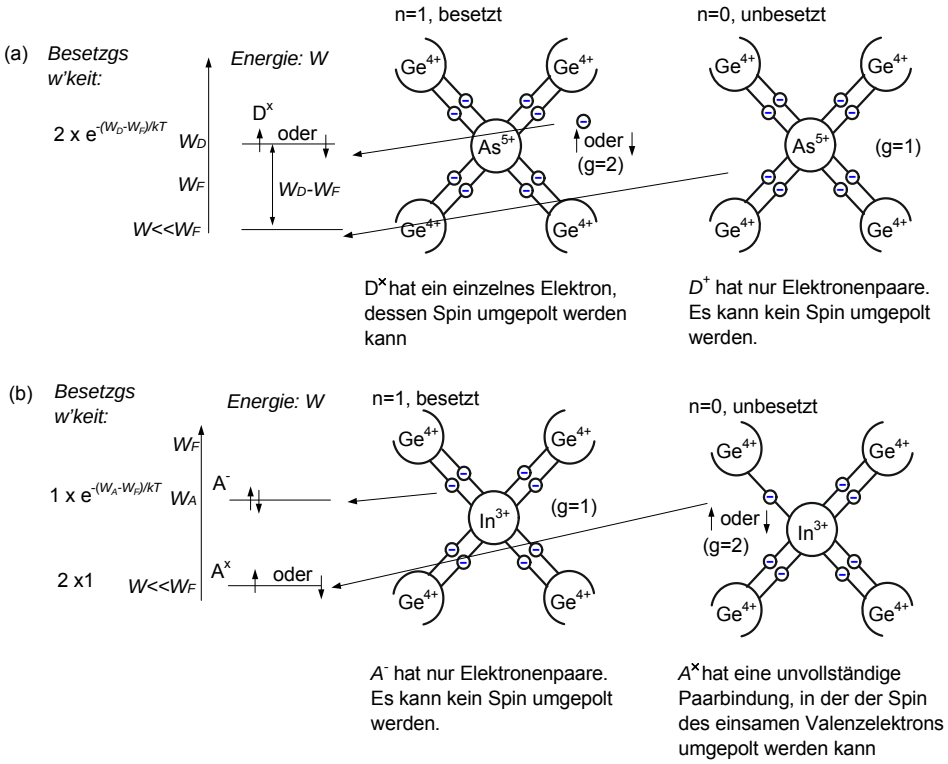


Abbildung 3.9: As und In als Donator bzw. Akzeptor in Germanium. Fim Fall des Donators (a) entspricht der besetzte Zustand dem energetisch unangeregte Fall D^x. Das Elektron kann 2 mögliche Spins annehmen ($g = 2$).Im Fall des Akzeptors (b) entspricht der besetzte Fall dem energetisch angeregten Zustand mit keiner Wahlfreitheit für den Spin ($g = 1$).

3.2.5 Konzentration der Ladungsträger bei dotierten HL

Die Anzahl der aktivierten Donatoren bzw. Akzeptoren ergibt sich ganz einfach aus Multiplikation der Dotierungkonzentration mit der Fermibesetzungswahrscheinlichkeit:

$$\begin{aligned} n_D^+ &= n_D[1 - f_D(E_D)] &= \frac{n_D}{1 + 2\exp\left(\frac{E_F - E_D}{kT}\right)}, \\ n_A^- &= n_A f_A(E_A) &= \frac{n_A}{1 + 2\exp\left(\frac{E_A - E_F}{kT}\right)}. \end{aligned} \tag{3.34}$$

Diese Gleichungen zeigen, dass die Störstellen aktiv bleiben, solange das Fermi-Niveau vom entsprechenden Störstellen-Niveau mehrere kT entfernt liegt.

Die Zahl der Ladungsträger im Leitungs- und Valenzband ergibt sich (bei bekanntem Ferminiveau) ganz anlaog wie in (3.25) und unter der Voraussetzung der Boltzmann-Näherung aus:

$$n = N_L \exp\left[-\tfrac{E_L - E_F}{kT}\right] \qquad p = N_V \exp\left[-\tfrac{E_F - E_V}{kT}\right] \tag{3.35}$$

Das Massenwirkungsgesetz gilt genauso für den dotierten wie den undotierten HL. D.h. es ist

$$n_{th} p_{th} = n_i^2(T) = N_L N_V \exp\left(-\frac{E_G}{kT}\right). \tag{3.36}$$

Um das zu beweisen, genügt es 3.35 miteinander zu multiplizieren.

Man könnte jetzt denken, dass sich die Zahl der Ladungstäger z.B. im Leitungsband aus $n = n_{th} + n_D^+$ berechnen lässt. n_{th} könnte man dann z.B. aus Gl. (3.25) mit $E_F = E_i$ berechnen. Das ist leider nicht so. Die Lage des Ferminiveaus ändert sich nämlich mit der Dotierung. Damit kennt man aber auch die Zahl der aktiven Dotanden n_D^+ nicht mehr. Was wir benötigen ist eine Bestimmungsleichung für das Ferminiveau um damit die Konzentration der Ladungsträger im Dotierten HL berechnen zu können. Diese Gleichung ist die Ladungsneutralitätsgleichung. Im homogenen Halbleiter gilt ***Ladungsneutralität (charge neutrality)*** an jeder Stelle (im inhomogenen Halbleiter hingegen gilt die Ladungsneutralität nur bei Integration der lokalen Raumladungsdichte über das Gesamtvolumen).

$$n + n_A^- = p + n_D^+. \tag{3.37}$$

Lage des Fermi-Niveaus im dotierten Halbleiter

Wir benötigen also eine Bestimmungsgleichung für das Ferminiveau. Setzt man n, p aus (3.35) und n_D^+, n_A^- aus (3.34) in die Neutralitätsbedingung (3.37) ein, so erhält man eine Bestimmungsgleichung für die Fermi-Energie W_F.

Bild 3.10 zeigt die Abhängigkeit des Fermi-Niveaus als Funktion der Temperatur mit der Dotierung als Parameter. Dabei ist auch angemerkt, dass die Bandlücke selbst eine Funktion der Temperatur ist, was in Bild 1.15 bereits zu sehen war.

Solange die Dotierungsdichte klein gegen die äquivalente Zustandsdichte ist, liegt das Fermi-Niveau innerhalb der Bandlücke und alle Donatoren sind ionisiert. Erst bei höherer Dotierung entsteht eine teilweise Ionisierung. Sind sowohl Akzeptoren als auch Donatoren vorhanden, so ist die Differenz $n_D - n_A$ dafür verantwortlich, ob der Halbleiter p oder n leitend ist. Der gesamte Einbau von Dotierstoffen $n_D + n_A$ ist jedoch verantwortlich für die Reduktion der Beweglichkeit, da die Störstellen Streuzentren für die Ladungsträger bilden (Beitrag zur Impulsrelaxationszeit).

Gl. (3.41) zeigt, siehe auch Bild 3.11, dass nicht etwa jeder Donator ein Elektron im LB und jeder Akzeptor ein Loch im VB erzeugt, sondern dass n_A Donatoren ihr Elektron an einen Akzeptor abgeben (energetisch günstigerer Fall), so dass dann noch $n_D - n_A$ Donatoren ihr Elektron an das LB abliefern können.

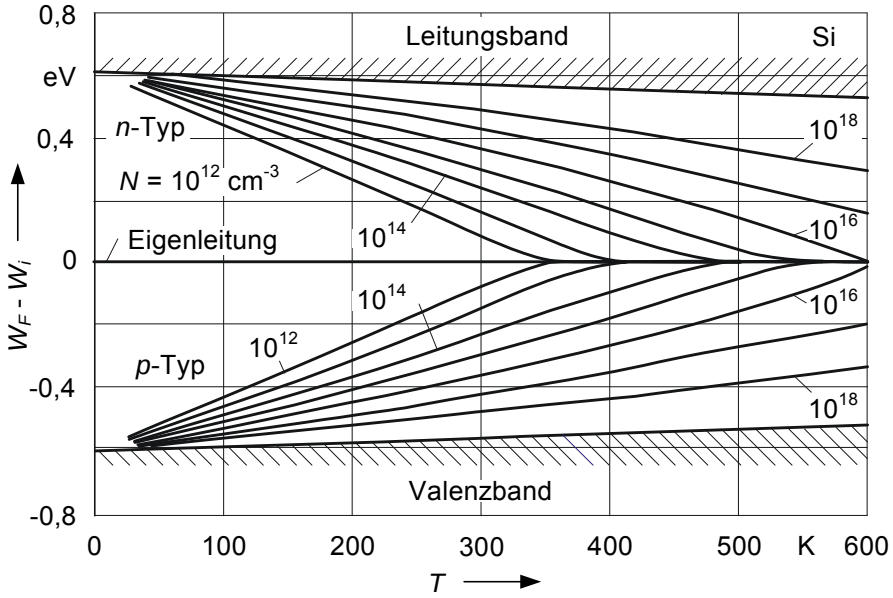


Abbildung 3.10: Fermi-Niveau in Si als Funktion der Temperatur mit der Dotierungskonzentration als Parameter [3].

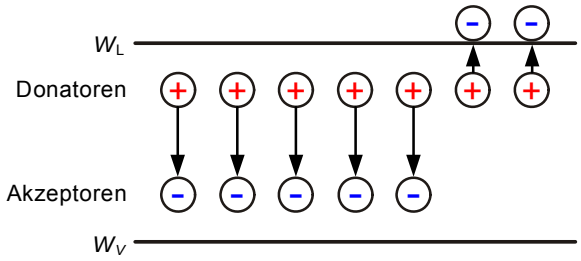


Abbildung 3.11: Elektronentransfer von Donatoren zu Akzeptoren und von Donatoren ins Leitungsband bei einem n-Halbleiter, $n_D > n_A$.

Beispiel für die Berechnung des Ferminiveaus für einen n-HL

Für einen n-Halbleiter, welcher nur mit Donatoren dotiert ist, fordert die Ladungsneutralitätsbedingung

$$n = p + n_D^+ . n_D^+ = n_D \frac{1}{1 + 2\exp\left(\frac{E_F - E_D}{kT}\right)} \tag{3.38}$$

bzw.

$$N_L \exp\left(-\frac{E_L - E_F}{kT}\right) = N_V \exp\left(-\frac{E_F - E_V}{kT}\right) + n_D \frac{1}{1 + 2\exp\left(\frac{E_F - E_D}{kT}\right)} .$$

Dies ist eine Bestimmungsgleichung für E_F. Sie kann graphisch gelöst werden, indem die linke und rechte Seite als Funktion der Fermi-Energie aufgetragen wird. Aus dem Schnittpunkt ergibt sich die gesuchte Fermi-Energie, siehe Bild 3.12. Dort sind zunächst die Dichten n_{th} und p_{th} im logarithmischen Maßstab eingetragen. Ist $n_D = 0$, so ergibt der Schnittpunkt dieser Geraden das

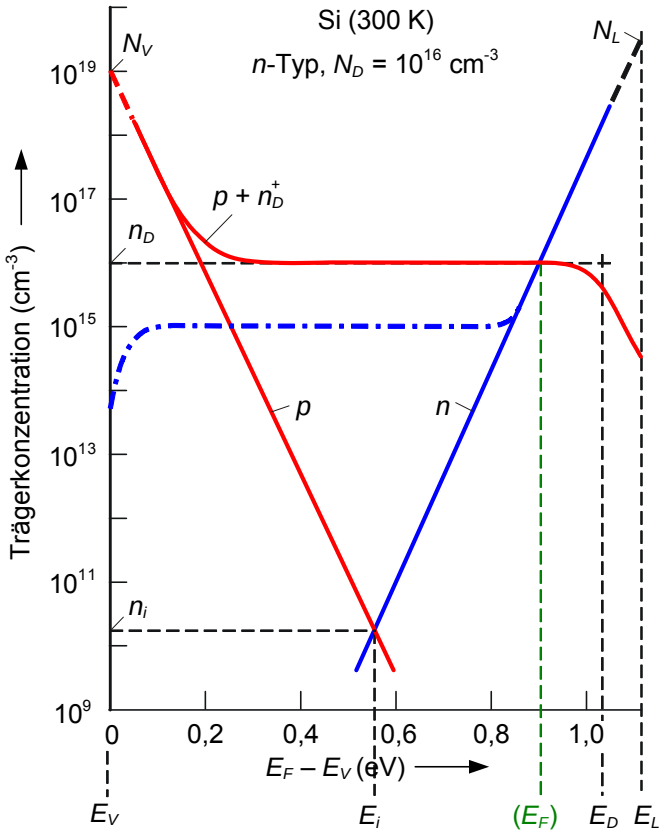


Abbildung 3.12: Ermittelung der Lage des Fermi-Niveaus. [3]

Ferrmi-Niveau im undotierten Zustand gemäß Gl.(3.28); die zugehörige Trägerdichte ist gleich der Eigenleitungsdichte n_i. Bei nicht verschwindender Dotierung muss noch die Anzahl der ionisierten Donatoren eingetragen werden. Durch Addition kann man die Kurve $p + n_D^+$ erhalten. Ist die Dotierdichte allerdings mehr als zehnmal so groß wie die Eigenleitungsdichte so kann man die Löcherdichte p vernachlässigen. Der Schnittpunkt zwischen n_D^+ und n ergibt dann das gesuchte Fermi-Niveau. Wie oben erwähnt, liegt die Fermienergie im n-dotierten HL stark auf der Seite der Leitungsbandkante.

Eine Möglichkeit um sich ein Bild über die vorhande Elektronendichte zu machen ist in Fig. 3.13 gezeigt.

Wenn wir die Elektronendichte im soeben diskutierten n-dotierten HL als Funktion der Temperatur auftragen, gelangen wir zum prinzipiellen Elektronendichteverlauf in Abb. 3.14.

- Bei niedrigen Temperaturen werden sukzessive Elektronen aus den Störstellen ionisiert, n steigt stark an (die durch Aufbrechen von Valenzbindungen erzeugten Trägerpaare sind noch vernachlässigbar). Dies ist der ***Bereich der Störstellenreserve.***

- Ist $kT > E_L - E_D$, aber $kT \ll E_G$, so sind praktisch alle Störstellen ionisiert $n_D^+ = n_D$, aber die Zahl der thermisch generierten Trägerpaare noch vernachlässigbar klein. In diesem ***Bereich der Störstellenerschöpfung*** ist $n \approx n_D$ (aber wegen $np = n_i^2$ ist die Minoritätendichte $p \approx n_i^2/n_D$ stark temperaturabhängig). In diesem Bereich verhält sich der HL ***extrinsisch (extrinsic).***

- Schließlich überwiegt die Zahl der thermisch generierten Trägerpaare die Zahl der aus Störstellen befreiten Träger und nach (3.26) gilt annähernd $n = p = n_i \sim \exp(-E_G/2kT)$. In diesem Bereich verhält sich der HL wieder wie ein ***intrinsischer HL.***

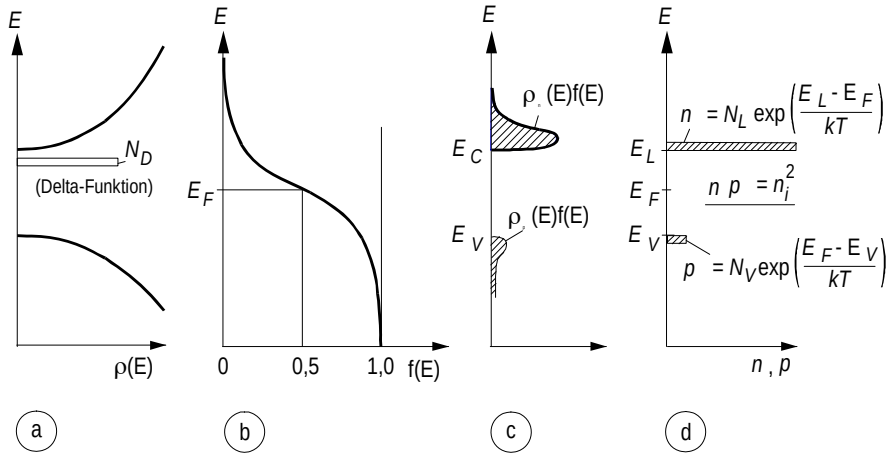


Abbildung 3.13: Verteilung der Elektronen und Löcher über der Energie für n-Halbleiter [3].

Bem: intrinsisch="von Innen herkommend". Eigenschaft, welche zum Gegenstand selbst gehört
extrinsisch="von außen kommend". Eigenschaft, welche von äußerem Einfluss (z.B. Dotierung) herrührt.

Störstellenerschöpfung (extrinsisch)

Für den wichtigen Fall der Störstellenerschöpfung (extrinsischer HL, alle Störstellen ionisiert $n_D^+ = n_{D,}, n_A^- = n_A$) erhält man aus der Neutralitätsbedingung und dem Massenwirkungsgesetz für die Gleichgewichtskonzentrationen n_n, p_n der Elektronen und Löcher für den n-Halbleiter:

$$\left.\begin{array}{rcl} n_n + n_A &=& p_n + n_D \\ n_n p_n &=& n_i^2 \end{array}\right\} \qquad \begin{array}{rcl} n_n &=& \sqrt{\left(\frac{n_D - n_A}{2}\right)^2 + n_i^2} + \frac{n_D - n_A}{2}, \\ p_n &=& \sqrt{\left(\frac{n_D - n_A}{2}\right)^2 + n_i^2} - \frac{n_D - n_A}{2} \end{array} \tag{3.39}$$

und den p-Halbleiter

$$\left.\begin{array}{rcl} n_p + n_A &=& p_p + n_D \\ n_p p_p &=& n_i^2 \end{array}\right\} \qquad \begin{array}{rcl} n_p &=& \sqrt{\left(\frac{n_D - n_A}{2}\right)^2 + n_i^2} + \frac{n_D - n_A}{2}, \\ p_p &=& \sqrt{\left(\frac{n_D - n_A}{2}\right)^2 + n_i^2} - \frac{n_D - n_A}{2} \end{array} \tag{3.40}$$

Vereinfachungen ergeben sich falls der extrinsische HL genügend stark dotiert ist $|n_D - n_A| \gg n_i$. Dann kann man für den n-HL schreiben

$$\left.\begin{array}{rcl} n_n + n_A &=& n_D \\ n_n p_n &=& n_i^2 \end{array}\right\} \qquad \begin{array}{rcl} n_n &=& n_D - n_A \\ p_n &=& \frac{n_i^2}{n_D - n_A} \end{array} \tag{3.41}$$

und für den p-dotierten HL:

$$\left.\begin{array}{rcl} n_A &=& n_D + p_p \\ n_p p_p &=& n_i^2 \end{array}\right\} \qquad \begin{array}{rcl} p_p &=& n_A - n_D \\ n_p &=& \frac{n_i^2}{n_A - n_D} \end{array} \tag{3.42}$$

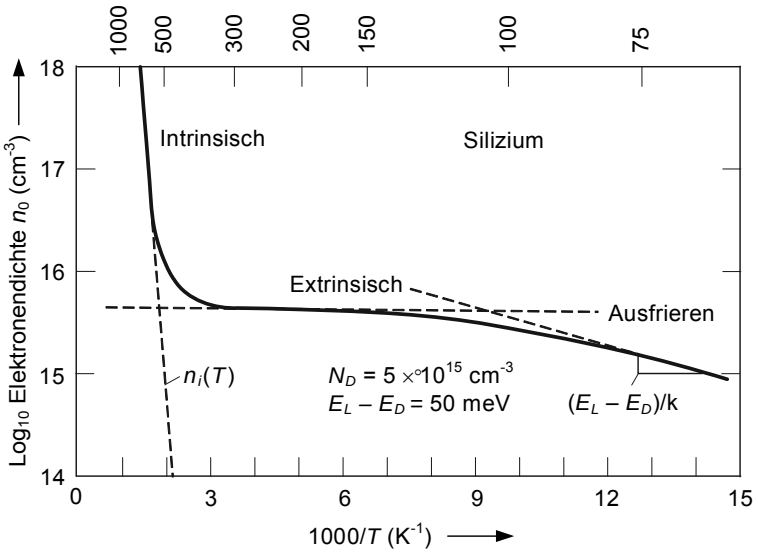


Abbildung 3.14: Prinzipieller Verlauf der Majoritätsträgerdichte als Funktion der Temperatur in einem n-Halbleiter. Die vertikale Achse ist logarithmisch und die horizontale invers zur Temperatur, so daß die asympthotischen Bereiche, Eigenleitung (intrinsisch), Fremdleitung (extrinsisch) und Störstellenreserve (Ausgefrieren der Ladungsträger) als Geraden sichtbar werden. [6]

Für einen n-Halbleiter folgt daraus, dass bei Temperaturen, bei denen Eigenleitung noch nicht eingetreten ist, aber alle Donatoren bereits ionisiert sind, die Majoritätsträgerdichte n_n nicht mehr von der Temperatur abhängig ist. Die Minoritätsträgerdichte p_n dagegen stark temperaturabhängig ist. Analoges gilt für den p-Halbleiter.

Die Ferminiveaus für den n-HL und den p-HL sind dann gerade:

$$E_F = E_L - kT \ln\left(\frac{N_L}{n_D}\right), \tag{3.43}$$

$$E_F = E_V + kT \ln\left(\frac{N_V}{n_A}\right). \tag{3.44}$$

3.3 Direkte und indirekte Halbleiter

1. Wir sprechen von einem ***direkten Halbeiter,*** wenn sowohl das Minimum der Leitungsbandkante als auch das Maximum der Valenzbandkante den gleichen Kristallimpuls aufweisen. Entsprechend haben diese beim ***indirekten Halbleiter*** verschiedene Kristallimpulse

Direkte und indirkete Halbleiter verhalten sich unter Lichteinwirkung ganz unterschiedlich. Damit es nämlich zu einer optischen Anregung eines Elektrons durch ein Lichtquant kommen kann, muss sowohl die Energieerhaltung als auch die Impulserhaltung erfüllt werden können.

Lichtquanten (Index L) haben im optischen Spektralbereich Energien der Größenordnung des Bandabstandes, aber Impulse p_L die viel kleiner sind als $h/(2a)$, da die Lichtwellenlänge (λ=1 μm) mehr als 1000 mal größer ist als der Gitterabstand a. In dem Sinne ändert sich der Impuls bei einer Anregung eines Elektrons durch ein Lichtquant nicht. Da die Maxima und Minima der entsprechenden Bandkanten bei einem direkten HL untereinander liegen, können bei einem direkten Halbleiter Elektronen durch Lichtanregung vom Valenzband ins Leitungsband übergehen, wenn ihnen nur mindestens die Energie E_g zugeführt wird. Bei einem indirekten Halbleiter, muss dem Elektron noch der entsprechende Impuls (z.B. $\hbar k = \pm h/(2a)$ im Bsp. von Fig. 2.11) zugeführt werden, um eine Anregung bei der Bandlückenenergie E_g zu ermöglichen.

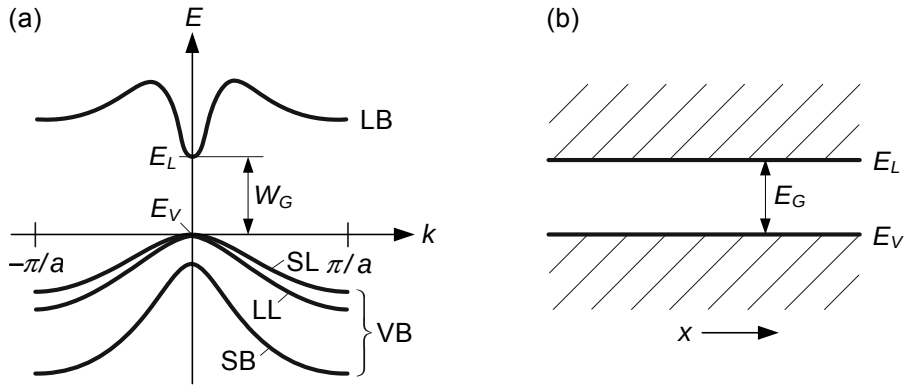


Abbildung 3.15: Bandstruktur eines direkten Halbleiters

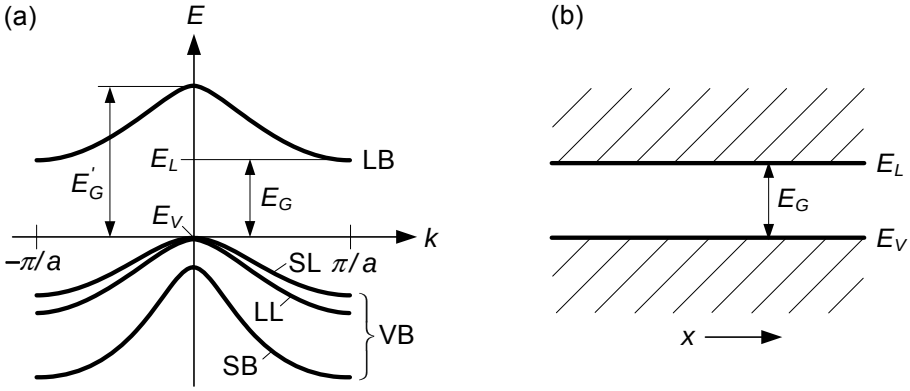


Abbildung 3.16: Bandstruktur eine indirekten Halbleiters

Gitterschwingungsquanten (Schallquanten, Index S) können Impulse dieser Größe haben, ihre Energie ist aber viel kleiner als der Bandabstand (Maximale Energie eines Schallquants ca. 40 meV). Damit kann man auch bei einem indirekten HL einen Übergang bei der minimalen Energie E_g sehen, falls das Elektron unter gleichzeitiger Beteiligung eines Photons und Phonons angeregt wird. Ein derartiger Dreierstoß (Elektron, Photon, Phonon) ist allerdings unwahrscheinlich. Licht wird nur schwach absorbiert und emittiert. Die Lebensdauer für einen angeregten Zustand liegen im Bereich von Millisekunden (und darüber). Indirekte Halbleiter (dazu gehören Ge, Si) sind als Lichtquellen ungeeignet (Dies gilt nicht für amorphe Halbleiter, dort ist keine Kristallstruktur vorhanden und die Prozesse, welche zur Absorption und Emission führen, gehorchen anderen Regeln als den hier diskutierten).

Bei hinreichend kurzwelligem Licht (mit der Photonenenergie E_g', siehe Bild 2.11 benimmt sich ein indirekter Halbleiter für die Lichtabsorption wie ein direkter Halbleiter. Bei k=0 an die höchste Stelle des LB kommende Elektronen relaxieren durch unelastische Stöße (Phononenemission) sehr schnell (˜100 Femtosekunden) ins Minimum des LB bei $k = \pm\pi/a$.

Wegen der direkten Anregung eines Elektrons durch ein Photon ohne Beteiligung von Phononnen wird in direkten Halbleitern Licht hingegen stark absorbiert (und im angeregten Zustand schnell emittiert). Die zugehörigen Lebensdauer für strahlende Übergänge vom LB ins VB liegt im Bereich von Nanosekunden: Lichtquellen (LED, Laser) verwenden direkte Halbleiter, z.B. GaAs, InGaAsP und andere.

$$\text{Photon-Impuls: } p_L = \hbar k_L = \hbar\frac{\omega}{c} = \hbar\frac{2\pi}{\lambda_L} \ll \hbar\frac{2\pi}{2a}, \qquad \text{Photon-Energie: } E_L = h\frac{c}{\lambda_L} \approx E_g \quad (3.45)$$

$$\text{Phonon-Impuls: } p_S = \hbar k_S = \hbar \frac{2\pi}{\lambda_S} \approx \hbar \frac{2\pi}{2a}, \qquad \text{Photon-Energie: } E_S = h \frac{v_S}{\lambda_S} \ll E_g \tag{3.46}$$

Kapitel 4

Ladungsträgertransport im Halbleiter

In diesem Kapitel werden wir die verschiedenen Effekte betrachten, die zum Ladungsträgertransport im Halbleiter beitragen. Die Summe aller Effekte wird uns letztlich zur Kontinuitätsgleichung führen.

4.1 Driftstrom

Unter dem ***Drift- oder Konvektionsstrom (drift current)*** versteht man den Stromfluss aufgrund der elektrischen Leitfähigkeit des Materials bzw. der im Material vorhandenen freien Ladungsträger, welche unter dem Einfluss eines äußeren elektrischen Feldes $\vec{E}$ eine Driftbewegung durchführen. Sowohl Elektronen als auch Löcher tragen zum Driftstrom bei. Die Größe der beiden Anteile hängt im Wesentlichen von der Ladungsträgerkonzentration und den mittleren Driftgeschwindigkeiten v_n, v_p ab. Konkret kann man schreiben:

$$\begin{aligned}\vec{J}_F &= \vec{J}_{n,F} + \vec{J}_{p,F} && (4.1)\\ &= -en\vec{v}_n + ep\vec{v}_p && (4.2)\\ &= \left[en\mu_n + ep\mu_p\right]\vec{E} && (4.3)\\ &= \sigma\vec{E}\ . && (4.4)\end{aligned}$$

Damit ergibt sich für die Leitfähigkeit aufgrund der freien Ladungsträger der Ausdruck

$$\sigma = en\mu_n + ep\mu_p\ . \tag{4.5}$$

Die Leitfähigkeit hängt also von den Konzentrationen der Ladungsträger - aber auch von materialabhängigen Größen, den sogenannten ***Beweglichkeiten (mobilities)*** μ_n und μ_p ab. Bild 4.1 zeigt die Leitfähigkeit von mit Arsen dotiertem Silizium bei verschiedenen Dotierungen als Funktion der Temperatur. Wir erwarten, dass die Leitfähigkeit mit der Dotierung und Temperatur ansteigt. Das ist in einem gewissen Bereich auch der Fall. Bei hohen Temperaturen verringert allerdings die Streuung der Ladungsträger an thermisch angeregten Phononen die Beweglichkeit und damit die Leitfähigkeit. Die Leitfähigkeit eines Halbleiters ist auch nicht streng proportional zur Dotierungsstärke, da die Störstellen nicht nur freie Ladungsträger erzeugen, sondern auch als Streuzentren für die Ladungsträger wirken.

Offensichtlich hängt die Beweglichkeit der Ladungsträger in komplexer Weise von der Temperatur und der Konzentration der Ladungsträger ab. Wir wollen diese im nächsten Abschnitt gesondert betrachten.

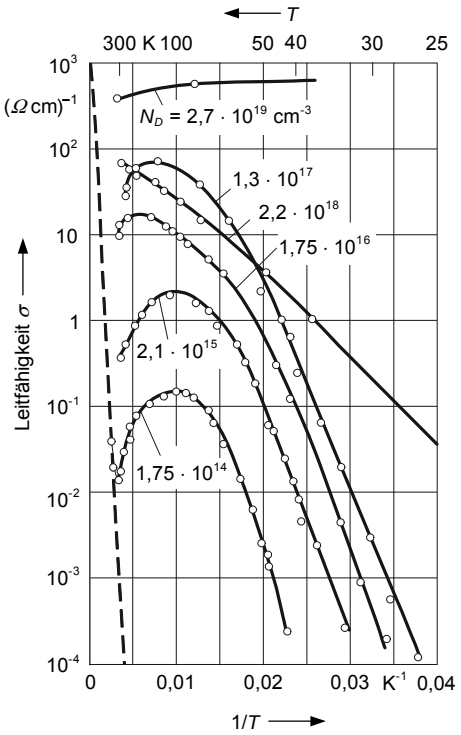


Abbildung 4.1: Temperaturabhängigkeit der Leitfähigkeit für verschieden starke As Dotierungen von Si [3].

4.1.1 Beweglichkeit der Ladungsträger

1. Im Leitungsband

Wir betrachten ein teilweise gefülltes Leitungsband, so dass Ströme auftreten können.

Ohne äußere Kraft kompensieren sich die Strombeiträge einzelner Elektronen paarweise und es tritt kein Strom auf, Bild 4.2(a).

Diese Situation ändert sich in der Gegenwart einer äußeren Kraft. Eine von außen angelegte Kraft ändert die Impulsverteilung der Elektronen im Kristall. Ein Strom fließt, Bild 4.2(b). Idealerweise würde der Strom bei einer angelegten Kraft nun exponentiell anwachsen, weil $F = \hbar\, dk/dt$. Dem ist aber nicht so. In einem realen Kristall, bei endlicher Temperatur und einer angreifenden äußerer Kraft ändert sich der Kristallimpuls nicht über beliebig lange Zeiträume. Im Gegenteil, es stellt sich ein dynamisches Gleichgewicht zwischen der von außen angelegten Kraft und den durch Stöße mit Gitterschwingungen (Wechselwirkung mit Phononen) an den Kristall verlorenen Energie (Erwärmung des Kristalls) ein. Im Fall des dynamischen Gleichgewichtsfall erfahren die Elektronen keine weitere mittlere Beschleunigung, $d\vec{v}_n/dt = 0$. Es fliesst ein konstanter Gesamtstrom (s. Bild 4.2). Physikalisch bedeutet dies, dass die Summe aller angelgeten Kräfte sich gerade aufheben (d.h. die Kraft aus dem anglegeten elektrischen Feld $\vec{E}$, $\vec{F}_{ext} = -e\vec{E}$ und die phänomenologische mittlere Ausbremsungskraft der Gitterstöße):

$$0 = m_n \frac{d\vec{v}_n}{dt} = \sum \vec{F}_i = \vec{F}_{ext} - \vec{F}_{Stoss} = -e\vec{E} - m_n \frac{\vec{v}_n}{\tau_{\mathrm{LB}}}. \tag{4.6}$$

wobei die Gesamtheit der Elektronen im Leitungsband im Fall des bestehenden dynamischen Gleichgewichts durch eine mittlere Driftgeschwindigkeit $\vec{v}_n$ charakterisiert sind. Dabei hat τ_{LB} (***Intrabandimpulsrelaxationszeit des Leitungsbandes***) offenbar die Bedeutung der Zeitkonstante, mit der sich eine endliche Driftgeschwindigkeit $\vec{v}_n \neq 0$ nach Abschalten der äußeren Kraft

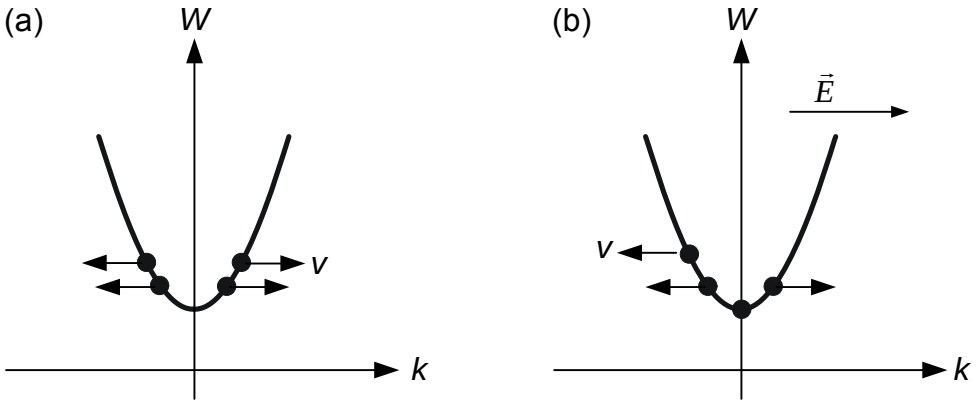


Abbildung 4.2: (a) Elektronen im Leitungsband ohne äußere Kräfte. (b) Einstellung eines stationären Nichtgleichgewichtes bei Vorhandensein einer äußeren Kraft und inelastischen Stößen. v ist die Elektronengeschwindigkeit.

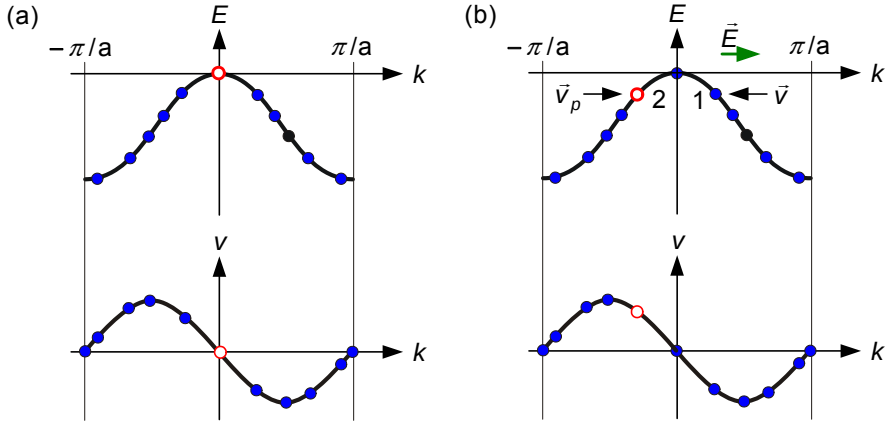


Abbildung 4.3: Verschiebung der Elektronenzustände in der Bandstruktur zufolge eines elektrischen Feldes. Der offene Kreis bezeichnet einen Zustand, der nicht mit einem Elektron besetzt ist.

dem Gleichgewichtsfall ($\vec{v}_n = 0$ nähert, $\vec{v}_n(t) = \vec{v}_n(0)\exp(-t/\tau_{\mathrm{LB}})$). Durch Auflösen findet man für die Elektronendriftgeschwindigkeit

$$\vec{v}_n = -\frac{e\tau_{\mathrm{LB}}}{m_n}\vec{E} = -\mu_n\vec{E}, \quad \mu_n = \frac{e\tau_{\mathrm{LB}}}{m_n}. \tag{4.7}$$

Aus einer Messung der ***Elektronen Beweglichkeit (electron mobility)*** μ_n lässt sich bei bekanntem m_n die Relaxationszeit τ_{LB} ermitteln (z. B. erhält man für InP, $\mu_n = 4600\,\mathrm{cm}^2/\mathrm{Vs}$, $m_n/m_0 = 0{,}077$, $m_0 = 0{,}911 \cdot 10^{-30}\,\mathrm{kg}$, $e = 1{,}602 \cdot 10^{-19}\,\mathrm{As}$ den Wert $\tau_{\mathrm{LB}} = 0{,}2\,\mathrm{ps}$).

2. Im Valenzband

Ein Band, in dem alle Plätze besetzt sind, trägt nicht zum Gesamtstrom bei, da sich wegen der Symmetrie $W(k) = W(-k)$ die Strombeiträge der einzelnen Elektronen paarweise kompensieren.

Betrachtet wird nun ein Band, in dem der höchste Zustand leer ist (etwa das Valenzband in Fig. 4.3(a), aus dem das Elektron mit der höchsten Energie entfernt wurde). Eine Feldstärke $\vec{E}$ ändert den Kristallimpuls der Elektronen gemäß (2.12)

$$\frac{d(\hbar\vec{k})}{dt} = -e\vec{E} \tag{4.8}$$

um einen zu $\vec{E}$ antiparallelen Betrag. In der Bandstruktur von Bild 4.3(b) wandern die Elektronen in jeweils links benachbarte Zustände. In Bild 4.3(b) sind alle Elektronen (auch die Leerstelle) um einen Platz nach links gewandert, das bei $k = -\pi/a$ über die Grenze der Brillouinzone gewanderte Elektron wurde an der äquivalenten Stelle bei $k = +\pi/a$ wieder eingefügt. Der Strom entsteht allein durch das ungepaarte Elektron 1. Seine Teilchengeschwindigkeit $\vec{v}$ ist nach links gerichtet ($\partial W/\partial k < 0$, s. (2.10)). Derselbe Strom würde durch eine positive Ladung $+e$ an der symmetrisch zu $\vec{k} = 0$ liegenden Stelle 2 erzeugt, der man die Geschwindigkeit $\vec{v}_p$ zuschreibt, welche ein an dieser Stelle sitzendes Elektron hätte. Es gilt $\vec{v}_p = -\vec{v}$. Für die Geschwindigkeitsänderung des Elektrons an der Stelle 1 gilt nach

$$\frac{d\vec{v}}{dt} = \frac{-e\vec{E}}{m_n} = \frac{e\vec{E}}{(-m_n)} = \frac{e\vec{E}}{m_p} = \frac{d\vec{v}_p}{dt}, \qquad m_p = -m_n. \tag{4.9}$$

Nach (4.9) erfährt also ein fiktives Teilchen der Ladung $+e$ mit einer effektiven Masse $m_p = -m_n > 0$ dieselbe Geschwindigkeitsänderung. Bem: An der Stelle 2 gilt für die effektive Masse, die man einem dort sitzenden Elektron zuzuschreiben hätte, $m_n < 0$ (da $\partial^2 W/\partial k^2 < 0$).

Ergebnis: Der Strombeitrag eines fast vollen Bandes mit Leerstellen an der Oberkante (in dem Bereich, in dem man den Elektronen negative effektive Massen zuzuschreiben hat) inklusive der Stromänderungen kann dadurch erfasst werden, dass man sich alle Leerstellen mit "Löchern" besetzt denkt, welche die Ladung $+e$ besitzen, dieselbe Geschwindigkeit haben wie ein an den jeweiligen Stellen sitzendes Elektron, und deren effektive Masse das Negative der effektiven Masse ist, die ein dort sitzendes Elektron hätte. Alle mit Elektronen besetzten Plätze bleiben unbeachtet.

Bei Berücksichtigung von Stößen erhält man für die Löcher anstelle von (4.6)

$$0 = m_p \frac{d\vec{v}_p}{dt} \quad = \quad \sum \vec{F}_i = \vec{F}_{ext} - m_p \frac{\vec{v}_p}{\tau_{\mathrm{VB}}} = e\vec{E} - m_p \frac{\vec{v}_p}{\tau_{\mathrm{VB}}}, \tag{4.10}$$

und anstelle von (4.7) für die Löchergeschwindigkeit

$$\vec{v}_p \quad = \quad \mu_p \vec{E},\ \mu_p = \frac{e\tau_{\mathrm{VB}}}{m_p}. \tag{4.11}$$

Dabei ist μ_p die ***Beweglichkeit der Löcher (hole mobility).*** τ_{VB} ist die ***Intrabandimpulsrelaxationszeit des Valenzbandes*** (die Zeitkonstante, mit der sich nach Abschalten der äußeren Kraft die Driftgeschwindigkeit $\vec{v}_p = 0$ einstellt).

- Die getrennte Behandlung von Elektronen im Leitungsband (4.6) und von Löchern im Valenzband (4.11) stellt sich als gerechtfertigt heraus, da die Intrabandimpulsrelaxationszeiten sehr klein sind (fs und ps) im Vergleich zu jenen Zeiten, in denen sich durch Übergänge zwischen dem Valenzband und dem Leitungsband ein kollektives Gleichgewicht zwischen Elektronen im Leitungsband und Löchern im Valenzband einstellt (ns bis ms). τ_{LB} kann daher als die Zeit betrachtet werden, in der nach einer Störung des Gleichgewichts die Elektronen im Leitungsband infolge von Stößen Impuls abgeben und in die tiefsten ihnen zugänglichen Zustände im LB relaxieren, also gemeinsam ins Gleichgewicht kommen. Ähnlich ist τ_{VB} für Löcher im Valenzband die Zeit, in der sie unter Abgabe von Impuls und Energie in die höchsten ihnen zugänglichen Zustände im Valenzband relaxieren (die Löcher gelangen gemeinsam ins Gleichgewicht). Erst durch Wechselwirkung zwischen Elektronen und Löchern (also Band-Band-Übergänge) stellt sich allmählich nach Abschalten der äußeren Kräfte ein thermodynamisches Gleichgewicht zwischen den beiden Trägersorten und dem Gitter ein.

- Eine differenziertere Betrachtung müsste unterscheiden zwischen der eigentlich in (4.6) und (4.11) definierten *Impulsrelaxationszeit* (es sind nämlich die Zeitkonstanten, mit denen die Impulse $m_n\vec{v}_n$ und $m_p\vec{v}_p$ im Elektronen- und Löcherkollektiv gegen Null gehen) und *Energierelaxationszeiten*, mit denen sich die Energie der Elektronen im Leitungsband und die der Löcher im Valenzband der mittleren Energie des Kristallgitters angleicht. Die Energierelaxationszeiten können bis 100 mal so groß sein wie die Impulsrelaxationszeiten. Die Stoßpartner

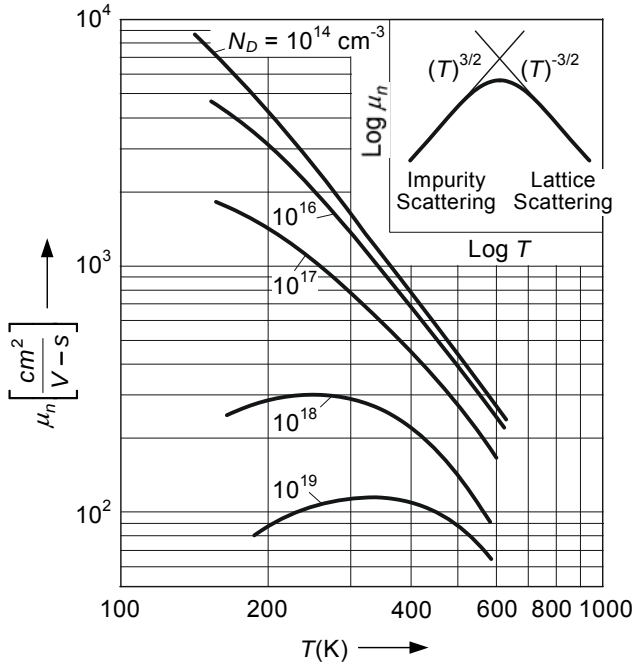


Abbildung 4.4: Die Beweglichkeit hängt stark von der Kollisionswahrscheinlichkeit ab. Bei steigender Dotierung und/oder mit steigender Temperatur sinkt die Bweglchkeit.

der Träger (z. B. ionisierte Störstellen, Wechselwirkung mit akustischen Gitterschwingungen) haben eine fiktive Masse, die sehr groß ist im Vergleich zur effektiven Masse der Träger und tauschen bei einem Stoß daher nur einen kleinen Bruchteil der Trägerenergie mit den Trägern aus. Die Stöße sind im wesentlichen elastisch und daher dauert es lange, bis sich die Energien der Träger an die der Gitterschwingungen angeglichen haben (beachte aber die nach (4.7) erwähnten unelastischen Stöße bei hohen Feldstärken, bei denen longitudinal optische Phononen angeregt werden, deren Quantenenergie vergleichbar ist mit der mittleren thermischen Energie $\frac{3}{2}kT_0$ der Träger; bei $T_0 = 293\,K$ ist $kT_0 = 25\,\text{meV}$).

3. Einfluss von Dotierungen und Temperatur auf die Beweglichkeit

Sowohl die Dotierkonzentration als auch die Temperatur beeinflussen den absoluten Wert der Beweglichkeiten. Hohe Dotierungen reduzieren die Beweglichkeit wegen einer größeren Kollisionswahrscheinlichkeit mit einem Dotanden (Impurity Scatttering). Genauso vermindert eine hohe Temperatur die Beweglichkeit der Ladungsträger (Phonon Scattering).

Allgemein berechnet sich die Relaxationszeit aus der Wahrscheinlichkeit einer Kollision mit einem Phonon und der Wahrscheinlichkeit einer Kollision mit einem Dotanden bzw. einer Verunreinigung.

$$\frac{1}{\tau_{\text{RelaxTot}}} = \frac{1}{\tau_{\text{LB/VB}}} + \frac{1}{\tau_{\text{Verunreinigung}}} \tag{4.12}$$

4. Einfluss von hohen Feldern auf die Beweglichkeit

Für hohe Feldstärken ist (4.7) ungültig. Die Elektronen regen LO-Phononen (das sind longitudinale optische Gitterschwingungen; bei ihnen schwingen benachbarte Atome longitudinal in Gegenphase mit einer Frequenz um 10^{13} Hz, das entspricht einer Phononenenergie von ca. 40 meV) so effizient an, dass sich die Driftgeschwindigkeit einem Sättigungswert nähert. Je nach Material und

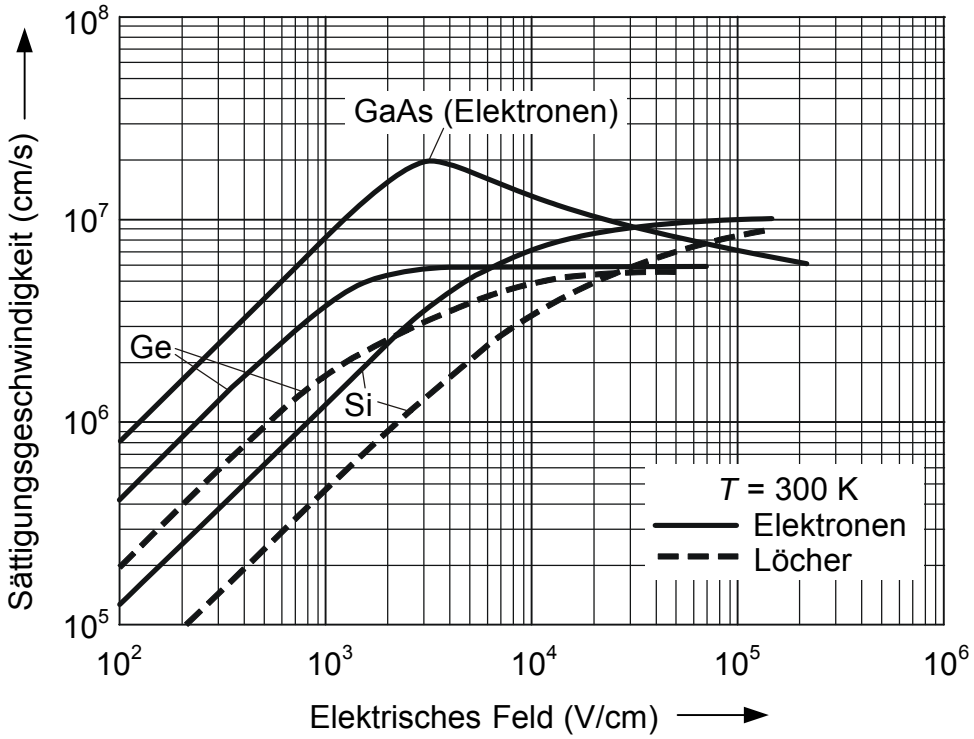


Abbildung 4.5: Ladungsträgergeschwindigkeit als Funktion des elektrischen Feldes für Ge, Si und GaAs. Für stark dotiertes Material ist die Anfangssteigung der Kurven kleiner. Die Sättigungsgeschwindigkeiten sind aber unabhängig von der Dotierung [4].

Trägersorte liegen die ***Sättigungsgeschwindigkeiten (saturation velocity)*** im Bereich von $10 \ldots 100\ \mathrm{nm\,ps^{-1}} \equiv 10^6 \ldots 10^7\ \mathrm{cm\,s^{-1}}$

Bild 4.5 zeigt die Driftgeschwindigkeit als Funktion der angelegten Feldstärke für die drei wichtigsten Halbleiter. Im Falle von GaAs ergibt sich in einem bestimmten Bereich sogar eine negative differentielle Beweglichkeit. Diese rührt davon her, daß die Elektronen im Leitungsbandminimum auf Grund der kleinen effektiven Masse eine sehr hohe Beweglichkeit besitzen. Bei großen Feldstärken nehmen die Elektronen so viel Energie auf, daß diese in das um etwa 0.3 eV höher gelegene Seitental streuen können, siehe Bild **??**. Die effektive Masse dort ist aber um etwa einen Faktor 20 höher und die Beweglichkeit entsprechend geringer, was zu dem in Bild 4.5 gezeigten Einbruch in der effektiven Geschwindigkeits-Feldstärke-Beziehung für GaAs führt. Die experimentell gefundenen Driftgeschwindigkeiten können durch folgenden emprischen Audruck wiedergegeben werden

$$v_{n,p} = \frac{v_s}{\left[1 + (E_0/E)^{\gamma}\right]^{1/\gamma}}\,, \tag{4.13}$$

wobei v_s die Sättigungsgeschwinigkeit ist und E_0 ein konstantes Feld darstellt. Für hochreines Silizium gelten folgende Werte: $v_s = 1{\cdot}10^7\ cm/s$, $E_0 = 7{\cdot}10^3\ V/cm$ für Elektronen und $2{\cdot}10^4\ V/cm$ für Löcher mit $\gamma = 2$ für Elektronen und $\gamma = 1$ für Löcher.

Wie sich später zeigen wird beruhen auf diesem Effekt wichtige Bauelemente zur Mikrowellenerzeugung (Gunn-Element).

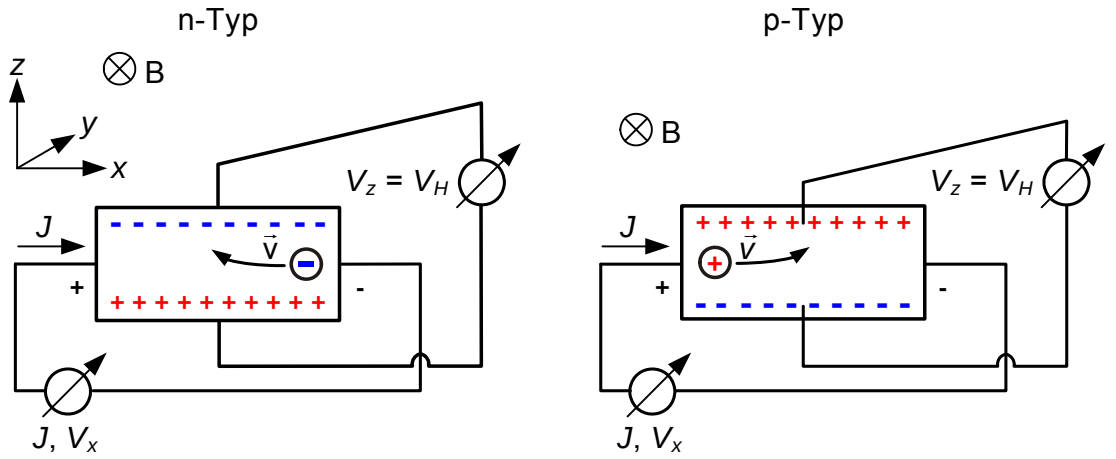


Abbildung 4.6: Anordnung zur Messung des Halleffektes in Halbleitern. Das Magnetfeld hat einen Vektor, welcher entlang der y-Achse ins Blatt hinein ausgerichtet ist.

4.1.2 Messung der Beweglichkeiten und Ladungsträgerkonzentrationen: Hall-Effekt

Die bisherigen Überlegungen haben nur ein äußeres elektrisches Feld mitberücksichtigt. Was passiert aber bei einem konstanten äußeren Magnetfeld (siehe Bild 4.6)? Wie hängen die Spannungen V, welche man in der x- und z-Richtung messen kann vom angelegten Strom ab? Wir wollen die Antwort systematisch erarbeiten.

Ges: Der zusammenhang zwischen $\vec{J}_p$ oder $\vec{J}_n$ mit den messbaren Spannungen V_z und V_x.

Lsg:

i) Je nach Dotierungen fließt ein Strom $\vec{J}_p$ oder $\vec{J}_n$.

$$\begin{array}{ll} \text{p-Typ} & \text{n-Typ} \\ \vec{J}_p = ep\vec{v}_p & \vec{J}_n = -en\vec{v}_n \end{array} \tag{4.14}$$

ii) Ohne Magnetfeld stellt sich infolge der elektrischen Kraft gemäss Gl. (4.7) und (4.11) die Dauergeschwindigkeit $\vec{v}$ ein

$$\begin{array}{ll} \text{p-Typ} & \text{n-Typ} \\ \vec{v} = \mu_p \vec{E} & \vec{v} = -\mu_n \vec{E} \end{array} \tag{4.15}$$

Auf die Teilchen mit Ladung $\pm e$ ($(+)$ für p-und $(-)$ für n-Halbleiter) wirkt nun aber nicht nur das el. Feld, sondern viel allgemeiner die Lorentz-Kraft

$$\vec{F} = \pm e \left(\vec{E} + \vec{v} \times \vec{B} \right). \tag{4.16}$$

Die zusätzliche Kraft infolge des Magnetfeldes führt zu einer weiteren Änderung in der sich stationär einstellenden Geschwindigkeit

$$\begin{array}{ll} \text{p-Typ} & \text{n-Typ} \\ \vec{v} = \mu_p(\vec{E} + \vec{v} \times \vec{B}) & \vec{v} = -\mu_n(\vec{E} + \vec{v} \times \vec{B}). \end{array} \tag{4.17}$$

iii) Wir setzten nun den Ausdruck für $\vec{v}_p$ und $\vec{v}_n$ durch den entsprechenden mit $\mu_p \left(\vec{E} + \vec{v} \times \vec{B} \right)$ und $\mu_n \left(\vec{E} + \vec{v} \times \vec{B} \right)$ im Ausdruck für den Strom

$$\begin{array}{ll} \text{p-Typ} & \text{n-Typ} \\ \vec{J}_p = ep\mu_p(\vec{E} + \vec{v} \times \vec{B}) & \vec{J}_n = -en\mu_n(\vec{E} + \vec{v} \times \vec{B}). \end{array} \tag{4.18}$$

iv) Auflösen nach dem elektrischen Feld ergibt dann:

$$\begin{array}{cc} \text{p-Typ} & \text{n-Typ} \\ \vec{E} = \frac{\vec{J}_p}{ep\mu_p} - \frac{1}{ep}\left(\vec{J}_p \times \vec{B}\right) & \vec{E} = \frac{\vec{J}_n}{en\mu_n} - \frac{-1}{en}\left(\vec{J}_n \times \vec{B}\right) \\ = \frac{\vec{J}_p}{\sigma_p} - R_p\left(\vec{J}_p \times \vec{B}\right) & = \frac{\vec{J}_n}{\sigma_n} - R_n\left(\vec{J}_n \times \vec{B}\right) \\ \text{mit} & \text{mit} \\ \sigma_p = \mu ep, \quad R_p = \frac{1}{ep} & \sigma_n = en\mu_n, \quad R_n = \frac{-1}{en} \end{array} \tag{4.19}$$

wobei $R_{p,n}$ die Hall-Konstante genannt wird. Gleichung (4.19) zeigt, dass der parallel zur Probe fliessende Strom in Folge des statischen Magnetfeldes ein elektrisches Feld erzeugt, welches senkrecht zur Stromrichtung wirkt. Dieses Feld, welches je nach Halbleiter-Typ eine verschiedene Richtung hat (da die Ladungsträger in die gleiche Richtung abgelenkt werden), erzwingt eine Anhäufung der Ladung auf einer Seite der Probe, was das Feld im Inneren der Probe im stationären Fall wieder abschirmt. Man kann an externen Klemmen über dem Halbleiter eine Hall-Spannung V_H abgreifen, welche es erlaubt, mit Hilfe der Probenabmessungen (w ist die Weite der Probe in Richtung der Hallspannung und ℓ die Länge der Probe in Richtung der Längsspannung V_x) die Beweglichkeit der Majoritätsträger in der Halbleiterprobe zu bestimmen:

$$\begin{array}{cc} \text{p-Typ} & \text{n-Typ} \\ V_z = V_H = \frac{J_P B}{ep} w & V_z = V_H = -\frac{J_n B}{en} w, \\ V_x = V_{\shortparallel} = \frac{J_P}{\sigma_p}\ell & V_x = V_{\shortparallel} = \frac{J_n}{\sigma_n}\ell, \\ \mu_p = \frac{V_H \ell}{V_{\shortparallel} B w} & \mu_n = -\frac{V_H \ell}{V_{\shortparallel} B w}, \end{array} \tag{4.20}$$

Hall-Messungen sind neben Leitfähigkeitsmessungen die am meisten verwendete Charakterisierungstechnik von Halbleiterproben. Das Vorzeichen der Hall-Spannung gibt den Halbleiter-Typ an. Der Hall-Effekt stellt auch die Basis für viele Halbleitersensoren dar, z.B. zur Messung des Magnetfeldes. Wir werden diesen Effekt in dieser Vorlesung nicht mehr verfolgen. Im Folgenden soll $B = 0$ gelten.

4.2 Diffusionsstrom

Im vorangehenden Kapitel haben wir Ladungstransport infolge eines angelegten äußeren Feldes betrachtet. Ein weiterer Beitrag zum Ladungstransport stammt von örtlichen Konzentrationsunterschieden der Ladungsträger, z.B. $n = n(x) \neq konst.$ Die Ladungsträger diffundieren nämlich aus Gebieten mit hohen Ladungsträgerkonzentrationen in Gebiete mit niedriger Konzentration. Dieses Diffundieren rührt von der statistischen, aber gleichmäßigen Bewegung der Teilchen in alle Richtungen her. Den resultierenden Strom nennt man ***Diffusionsstrom (diffusion current).*** Der Diffusionsstrom ist gegeben als

$$\vec{J}_D = \vec{J}_{n,D} + \vec{J}_{p,D} \tag{4.21}$$

$$\vec{J}_{n,D} = +eD_n \operatorname{grad} n, \tag{4.22}$$

$$\vec{J}_{p,D} = -eD_p \operatorname{grad} p. \tag{4.23}$$

D.h. also, dass er entgegengesetzt zum Konzentrationsgradienten gerichtet ist und proportional zu den Diffusionskonstanten D_n und D_p der Elektronen und Löcher ist. Diese Diffusionskonstanten sind noch zu bestimmen.

4.2.1 Herleitung der Beziehungen für den Diffusionsstrom

Zunächst wollen wir den Diffusionsprozess selber betrachten. In Fig. 4.7 ist eine entlang der x-Achse variierende Elektronendichte aufgezeichnet. Wir wollen einen Ausdruck für die Anzahl der Elektronen, welche die Fläche bei x=0 passieren, herleiten.

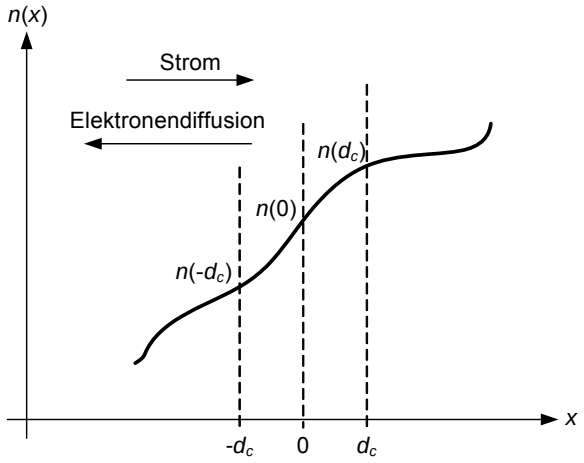


Abbildung 4.7: Elektronenkonzentrationsgradient entlang der x-Achse.

Die Elektronen bewegen sich mit der mittleren Geschwindigkeit $v_{th} = d_c/\tau_{\mathrm{LB}}$, wobei d_c und τ_{LB} die mittlere freie Weglänge und die Leitungsbandrelaxationszeit sind. In unserer Herleitung sollen sich die Elektronen auch nur nach rechts oder links bewegen können. Damit erhält man für den Elektronenfluss pro Einheitsfläche in die rechte Hälfte

$$F_r = \frac{1}{2} n(-d_c)\, \frac{d_c}{\tau_{\mathrm{LB}}} = \frac{1}{2} n(-d_c)\, v_{th} \tag{4.24}$$

und in die linke Hälfte

$$F_l = \frac{1}{2} n(d_c)\, v_{th} \tag{4.25}$$

und damit für den Nettofluss

$$\Delta F = F_r - F_l = \frac{1}{2} v_{th} \left(n(-d_c) - n(d_c) \right). \tag{4.26}$$

Eine Taylorentwicklung der Ladungsträgerkonzentration $n(-d_c)$ und $n(d_c)$ um $x = 0$ ergibt

$$\Delta F = \frac{1}{2} v_{th} \left(\left[n(0) - d_c \frac{dn}{dx} \right] - \left[n(0) + d_c \frac{dn}{dx} \right] \right) = -v_{th} d_c \frac{dn}{dx} \equiv -D_n \frac{dn}{dx}. \tag{4.27}$$

Diese kurze Herleitung zeigt, dass der Teilchenfluss in die zum Konzentrationsgefälle entgegengesetzte Richtung zeigt. Die Diffusionskonstante ist gegeben durch $D_n \equiv v_{th} d_c = v_{th}^2\, \tau_{\mathrm{LB}}$. Allerdings ist sie in dieser Form nicht sehr praktisch. Wir müssen Sie mit messbaren Größen in Verbindung bringen.

4.2.2 Die Diffusionskonstante: die Einstein-Beziehungen

Wir wollen $D_n = v_{th}^2\, \tau_{\mathrm{LB}}$ umschreiben.

Dazu verwenden wir zunächst das Äquipartitionsgesetz

$$\frac{1}{2} v_{th}^2 m_n = \frac{1}{2} kT \tag{4.28}$$

und die Definition der Beweglichkeit (4.7)

$$\mu_n = \frac{e\tau_{\mathrm{LB}}}{m_n} \tag{4.29}$$

um die zwei Größen v_{th} und τ_{LB} im Ausdruck D_n zu eliminieren.

Man erhält dann für die Diffusionskonstanten der Elektronen und Löcher

$$D_n = \frac{kT}{e}\mu_n \quad \text{und} \quad D_p = \frac{kT}{e}\mu_p. \tag{4.30}$$

Die Beziehungen (4.30) nennt man ***Einsteinbeziehungen.***

4.2.3 Thermisches Gleichgewicht im HL mit Konzentrationsgradient (Gleichgewicht zwischen Drift- und Diffusionsstrom)

Im thermischen Gleichgewicht fließen aufgrund eines Ladungsträgerkonzentrationsgradienten solange Ladungen, bis sich aufgrund einer Ladungsumverteilung ein elektrisches Feld aufbaut, welches den Diffusionsströmen entgegenwirkt.

Als Beispiel untersuchen wir einen inhomogen dotierten Halbleiter im thermischen Gleichgewicht.

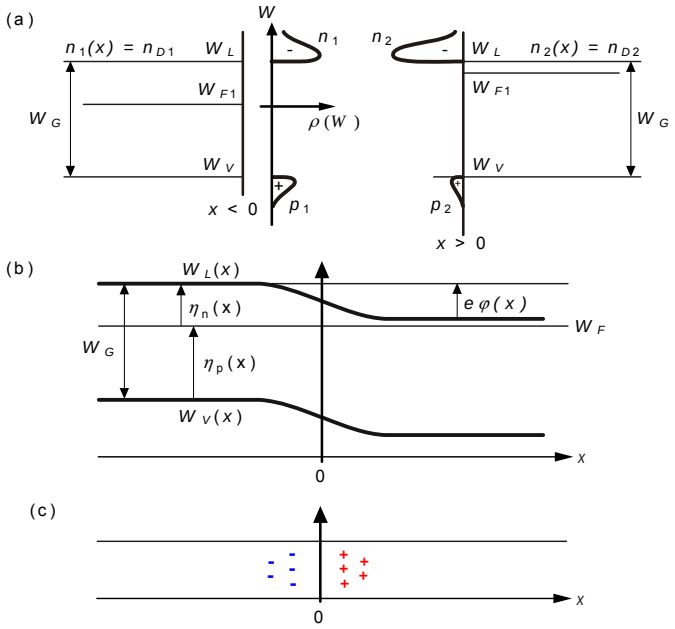


Abbildung 4.8: Diffusion von Elektronen und Löchern aus verschieden stark dotierten n-Halbleitern and der Grenzfläche. (a) Bandverläufe der ungestörten Halbleiter; (b) Bandverlauf nach dem Zusammenbringen der beiden Halbleiter, (c) Ladungsträgerverteilung nach Einstellen des Gleichgewichtes.

Dazu betrachten wir zunächst zwei räumlich getrennte n-Halbleiter mit verschieden starker Dotierung, n_{D1} und n_{D2}, siehe Bild 4.8(a). Somit herrschen in beiden Halbleitern verschiedene Gleichgewichts-Ladungsträgerdichten n_1, und n_2. Jeder frei bewegliche Ladungsträger im Leitungsband besitzt in beiden Proben, welche sich auf gleicher Temperatur befinden sollen, die selbe thermische Energie und bewegt sich mit gleicher Wahrscheinlichkeit in positive oder negative x-Richtung. Bringen wir die beiden Halbleiterproben zusammen, Fig. 4.8(b), so bewegen sich in der Probe 2 aufgrund der höheren Ladungsträgerdichte in einem Volumenelement unmittelbar am Rand zunächst mehr Ladungsträger in negative x-Richtung, als sich in einem Randvolumenelement in Probe 1 Ladungsträger in positive x-Richtung bewegen. Daher fließt ein Elektronen-Diffusionsstrom in negativer x-Richtung. Infolge des Diffusionsstromes kommt es zur Umverteilung der Ladungsträger und es entsteht eine Raumladungszone, da in der linken Halbleiterprobe mehr Elektronen und in der rechten Probe weniger Elektronen als im Gleichgewicht vorhanden sind.

Weit innerhalb der Proben, nach Abklingen der Störung, sollte wieder der gewohnte Gleichgewichtszustand herrschen. Aufgrund der ortsabhängigen Dotierung wäre die thermisch angeregte Elektronen- und Löcherkonzentration im Fall der Störstellenerschöpfung und isolierter Teilbereiche proportional zur Dotierung. Werden die Halbleiterbereiche zusammengeführt, so fließen aufgrund des dadurch entstehenden Konzentrationsgradienten so lange Diffusionsströme, bis sich aufgrund der neuen Ladungsverteilung ein elektrisches Feld aufbaut, bei welchem die Driftströme gerade die Diffusionsströme kompensieren. Dies muss im Gleichgewicht für Elektronen und Löcher gleichermaßen gelten. Damit folgt z.B. für die Elektronen

$$\vec{J}_{n,D} + \vec{J}_{n,F} \quad = \quad eD_n \operatorname{grad} n + en\mu_n \vec{E} \quad = \quad 0 \ ,$$

bzw. mit dem elektrostatischen Potential infolge des resultierenden elektrostatischen Feldes

$$\vec{E} = -\operatorname{grad} \varphi \, , \tag{4.31}$$

folgt

$$\frac{1}{n} \operatorname{grad} n = \frac{\mu_n}{D_n} \operatorname{grad} \varphi. \tag{4.32}$$

Die linke Seite lässt sich umformen. Es gilt nämlich $(\ln f)' = 1/f \ \cdot f'$. Damit erhält man

$$(\ln n)' = \frac{\mu_n}{D_n} \operatorname{grad} \varphi$$

Integrieren wir die Gleichung entlang eines Weges vom Ort x_0 mit Teilchendichte n_0 zum Ort x mit Teilchendichte $n(x)$ und Potentialdifferenz $\varphi(x)$ so erhalten wir

$$\ln n(x) - \ln n_0 = \frac{\mu_n}{D_n} \varphi(x) - 0. \tag{4.33}$$

bzw.

$$n(x) = n_0 \exp\left[\frac{\varphi(x)}{(D_n/\mu_n)}\right] = n_0 \exp\left[\frac{e\varphi(x)}{kT}\right] \quad . \tag{4.34}$$

Die Ladungsträgerkonzentration im Gleichgewicht kann also direkt aus der Raumladungsverteilung, d.h. über Glg. (4.31) und mit (4.34) errechnet werden.

Anmerkung: Diese Beziehung erlaubt es uns nun das Diffusionspotential über dem Halbleiterinterface zu bestimmen. Gemäß Gleichung .(4.34) gilt nämlich

$$n(x = \infty) = n_{D1} \exp\left[\frac{e\varphi(x)}{kT}\right]$$

aber auch

$$n(x = \infty) = n_{D2}$$

Im Übrigen gilt im Fall von Störstellenerschöpfung und unter der Annahme, dass die Boltzmann-Näherung erfüllt ist (siehe Abschnitt 3.1.1):

$$n_{D1} = N_L \exp\left[-\frac{W_{L1} - W_F}{kT}\right] \quad \text{und}$$
$$n_{D2} = N_L \exp\left[-\frac{W_{L2} - W_F}{kT}\right] \ ,$$

wobei W_{L1} und W_{L2} die Bandkanten der beiden Hableiterhälften sind. Auflösen der vier Gleichungen nach φ führt uns auf das Potential $\varphi = \varphi_D$ über dem Interface

$$\varphi_D = W_{L1} - W_{L2} \ .$$

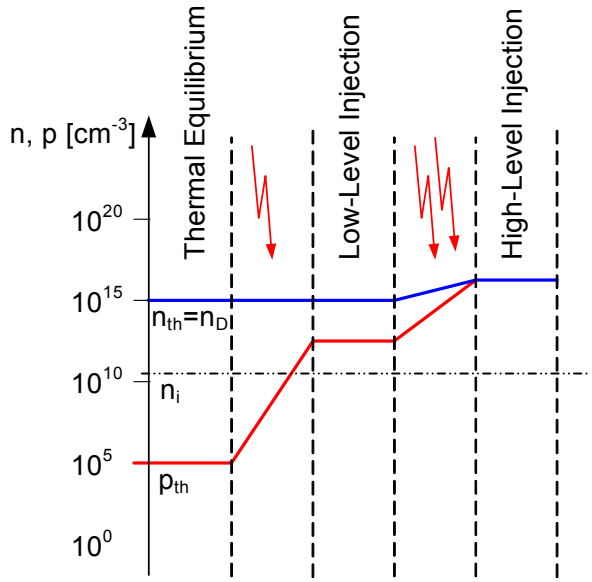


Abbildung 4.9: Ladungsträgerkonzentrationen in n-dotiertem Si ($n_D = 10^{15}\text{cm}^{-1}$) für (a) thermisches Gleichgewicht, (b) Niederinjektion und (c) Hochinjektion.

4.3 Einschuss von Ladungsträgern

Im thermischen Gleichgewicht gilt das Massenwirkungsgesetz

$$pn = n_i^2 \tag{4.35}$$

und die Ladungsträgertransportmechanismen werden sich entsprechend einstellen.

Falls Ladungsträger eingeschossen werden ***(carrier injection)*** stellt sich ein Nichtgleichgewicht ein und es gilt

$$pn > n_i^2 \ . \tag{4.36}$$

Die zusätzlichen Ladungsträger werden ***Überschussladungen (excess carriers)*** genannt. Der Einschuss von Ladungsträgern kann über zahlreiche Mechanismen erfolgen. Zu nennen wären etwa das optische Einschießen mit einer Rate g_{ext}, oder Ladungsträgereinschuss durch Anlegen einer vorwärtsgerichteten Spannung. Im Fall der optischen Anregung werden je ein Elektron im Leitungsband und ein Loch im Valenzband erzeugt, welche die Elektronen und Löcherkonzentrationen der entsprechenden Bänder anheben und das Gleichgewicht weiter stören.

Die Zahl der eingeschossenen Ladungen relativ zu der Menge der Majoritätsträger entscheidet, ob man von ***schwacher Injektion (low-level injection, LLI)*** oder ***starker Injektion*** oder ***Hochinjektion (high-level injection, HLI)*** spricht. Um die Begriffe zu verdeutlichen machen wir ein Beispiel.

Bsp: Es liege ein n-dotierter Si Kristall vor, welcher mit $n_D = 10^{15}\text{cm}^{-1}$ dotiert sei. Dieser weist die folgenden Ladungsträgerkonzentrationen auf: $n_{th} \cong n_D = 10^{15}\text{cm}^{-1}$, $p_{th} = 1,45 \cdot 10^5\text{cm}^{-1}$ mit $n_i = 1,2 \cdot 10^{10}\text{cm}^{-1}$. Zunächst werden $\Delta n = \Delta p = 1 \cdot 10^{12}\text{cm}^{-1}$ Überschussladungen eingeschossen. Die neuen Ladungsträgerkonzentrationen sind nun: $n_{th} = (1 \cdot 10^{15}\text{cm}^{-1} + 1 \cdot 10^{12}\text{cm}^{-1}) \cong 1 \cdot 10^{15}\text{cm}^{-1} = n_D$ und $p_{th} = 1 \cdot 10^{12}\text{cm}^{-1}$. Die Minoritätsträgerkonzentration ändert sich also um 7 Dekaden. Die Majoritätsträgerkonzentration ändert sich aber nur unwesentlich. Dieser Fall wird LLI genannt. Bei einem erneuten Einschuss von Ladungsträgern sollen $\Delta n = \Delta p = 1 \cdot 10^{16}\text{cm}^{-1}$ Überschussladungen eingeschossen werden. Die neuen Ladungsträgerkonzentrationen sind nun: $n_{th} = 1,1 \cdot 10^{16}\text{cm}^{-1}$,

$p_{th} = 1,0 \cdot 10^{16}\text{cm}^{-1}$. Es verändern sich nun sowohl die Majoritätsträgerkonzentration als auch die Minoritätsträgerkonzentration in großem Stil und wir sprechen von HLI. Siehe auch Fig. 4.9.

Mathematisch lässt sich die Situation für den Fall schwacher Injektion im n-dotierten und p-dotierten HL so beschreiben:

$$\left.\begin{array}{ll} \text{n-Typ HL:} & \begin{array}{l} n = n_{th} + \Delta n \cong |n_D - n_A| \\ p = p_{th} + \Delta p \ll |n_D - n_A| \end{array} \\ \text{p-Typ HL:} & \begin{array}{l} n = n_{th} + \Delta n \ll |n_D - n_A| \\ p = p_{th} + \Delta p \cong |n_D - n_A| \end{array} \end{array}\right\} \implies n_i^2 < np \ll (n_D - n_A)^2$$

wohingegen im Fall starker Injektion gilt

$$n,\, p \gg |n_D - n_A| \ .$$

4.4 Generation und Rekombination

Bis anhin haben wir Ladungsträgerdrift und -Diffusion als Transportmechanismen kennengelernt. Insbesondere haben wir gesehen, wie Ladungsträgerkonzentrationsgradienten durch Drift und Diffusion wieder ins Gleichgewicht kommen können. Ladungsträgergeneration und -rekombination stellen weitere Prozesse dar, um Nichtgleichgewichtszustände ($pn \neq n_i^2$) wieder ins Gleichgewicht (d.h. $pn = n_i^2$) relaxieren zu lassen.

4.4.1 Das Prinzip des detaillierten Gleichgewichts

Wird eine dünne Halbleiterprobe mit elektromagnetischer Strahlung der Frequenz f bestrahlt und ist die Quantenenergie größer als die der Energielücke, d.h. $hf > W_G$, so werden durch die gleichförmige Bestrahlung im Halbleiter zusätzliche Elektronen- und Lochpaare erzeugt, siehe Bild 4.10 (a),

$$n = n_{th} + \Delta n \tag{4.37}$$

$$p = p_{th} + \Delta p \tag{4.38}$$

wobei die Δ-Größen die Überschussladungen bezeichnen. Bei n- oder p–Halbleitern ist (n, p, n_{th}, p_{th}) durch (n_n, p_n, n_{nth}, p_{nth}) bzw. (n_p, p_p, n_{pth}, p_{pth}) zu ersetzen.

Die Relaxation der Überschussladungen ins Gleichgewicht wird durch die Differenz zwischen Generationsrate g und Rekombinationsrate r

$$\frac{dn}{dt} = \frac{dp}{dt} = g - r \tag{4.39}$$

bestimmt.

Die totale Generationsrate setzt sich im Allgemeinen aus verschiedenen Beiträgen zusammen

$$g = g_{\text{ext}} + g_{\text{phonon}} + g_{\text{opt}} + g_{\text{t}} + g_{\text{Auger}} \cdots \tag{4.40}$$

dabei beschreibt g_{ext} extern induzierte Generation, g_{phonon} ist die thermische Generationsrate infolge der Absorption von Gitter-Phononen in einem perfekten Halbleiter, g_{opt} die Generationsrate infolge der Absorption von Photonen aus der thermischen Hintergrundstrahlung oder zusätzlicher Bestrahlung wie in Bild 4.10 gezeichnet, g_{t} die Generationsrate infolge von Rekombinationszentren (Traps), usw., siehe Bild 4.11.

Zu jedem physikalisch unterschiedlichen Generationsprozess muss es auch einen entsprechenden Rekombinationsprozess geben

$$r = r_{\text{phonon}} + r_{\text{opt}} + r_{\text{t}} + r_{\text{Auger}} \cdots \ . \tag{4.41}$$

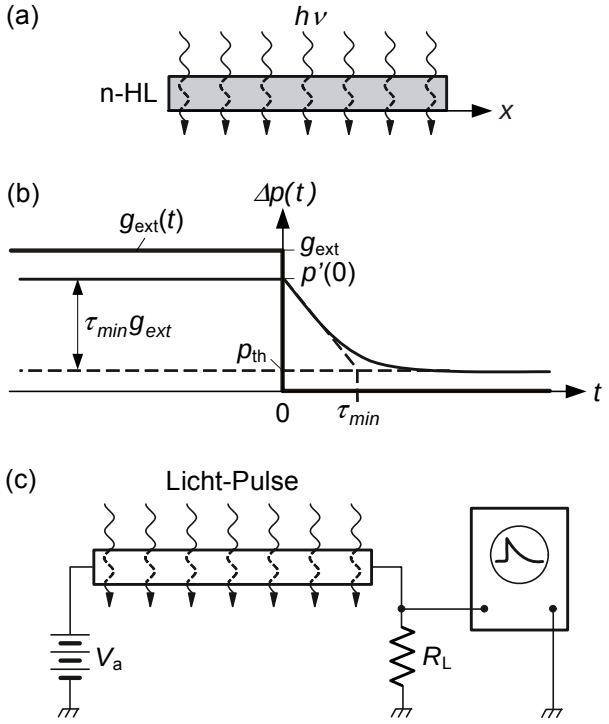


Abbildung 4.10: Bestrahlung und Rekombination optisch angeregeter Ladungsträger. (a) n-Halbleiter bei konstanter Beleuchtung. (b) Rekombination der Minoritätsträgerdichte. (c) Experiment zur Messung der Ladungsträgerlebensdauer [5].

Im **stationären Zustand** gilt das ***Prinzip des detaillierten Gleichgewichts (principle of detailed balance).*** Das heißt, dass die Anzahl der ins Leitungsband oder Valenzband eintretenden bzw. austretenden Ladungsträger sich gerade ausgleicht und damit gilt:

$$\frac{dn}{dt} = g_n - r_n \tag{4.42}$$

$$= g_{n,\text{ext}} + g_{n,\text{phonon}} + g_{n,\text{opt}} + g_{n,\text{t}} + g_{n,\text{Auger}} \ldots - (r_{n,\text{phonon}} + r_{n,\text{opt}} + r_{n,\text{t}} + r_{n,\text{Auger}} \ldots) = 0 \tag{4.43}$$

$$\frac{dp}{dt} = g_p - r_p \tag{4.44}$$

$$= g_{p,\text{ext}} + g_{p,\text{phonon}} + g_{p,\text{opt}} + g_{p,\text{t}} + g_{p,\text{Auger}} \ldots - (r_{p,\text{phonon}} + r_{p,\text{opt}} + r_{p,\text{t}} + r_{p,\text{Auger}} \ldots) = 0, \tag{4.45}$$

wobei die Indizes n und p andeuten, dass nur Prozesse, die auch einen Beitrag zum entsprechenden Band liegen, berücksichtigt werden sollten.

Als besonders nützlich wird sich diese Betrachtung erweisen, wenn wir das **Prinzip des detaillierten Gleichgewichts im thermischen Gleichgewicht** betrachten, wo $g_{\text{ext}} = 0$ gilt und wir mit bekannten Ladungsträgerkonzentrationen $n = n_{th}$ und $p = p_{th}$ rechnen können. Im thermischen Gleichgewicht der Probe mit ihrer Umgebung müssen dann nicht nur die gesamten Generationsrate und g_n und g_p gleich den Rekombinationsrate r_n und r_p sein, sondern für jeden Einzelprozess muss *die entsprechende Generationsrate gleich der Rekombinationsrate sein.* Wäre dies nicht der Fall, so ergäben sich Widersprüche. (Beweis: Es gebe nur Phononen und Photonen Generation und Rekombination. Diese 2 Prozesse seien im stationären und thermischen Gleichgewicht so dass gilt:$g_{\text{phonon}} + g_{\text{opt}} - r_{\text{phonon}} - r_{\text{opt}} = 0$. Falls nun $g_{\text{opt}} - r_{\text{opt}} < 0$ (d.h. der Körper strahlt Licht ab), dann muss $g_{\text{phonon}} - r_{\text{phonon}} > 0$ (d.h. der Körper kühlt sich ab, weil er seine Gitterschwinungen in Form von Elektronen-Loch Paaren abgibt). So ein Halbleiter würde strahlen

und sich dabei abkühlen - was explizit nicht einem thermischen Gleichgewicht entspricht.)

4.4.2 Die Generations- und Rekombinationsmechanismen

1. Strahlende Übergänge: Spontane Emission

Wir unterscheiden zwischen spontaner und stimulierter Emission bzw. stimulierter Absorption. Man spricht von spontaner Emission wenn, es zu einem spontanen Übergang von einem Elektron in einem hochenergetischen Niveau (z.B. Leitungsband) auf ein niederenergetisches Niveau (z.B. Valenzband) kommt und dabei ein Lichtquant emittiert wird. Von stimulierter Emission und Absorption spricht man, wenn die entsprechenden Übergänge von Licht induziert werden. Stimulierte Emission tritt in Lasern auf.

Hier betrachten wir nur spontane Übergänge. Diese dominieren die optischen Effekte in direkten Halbleitern (siehe Abschnitt 3.3) - solange wir diese nicht als Laser betreiben. Die Rekombinationsrate, siehe Bild 4.11(a) muss proportional zur Anzahl der vorhandenen angeregten Elektronen n sein und sie muss proportional zur Anzahl der freien Lochzustände p sein, in welche das Elektron relaxieren kann. Kurz, die spontane Rekombinationsrate kann geschrieben werden als

$$r_{\mathrm{opt}} = r_{\mathrm{sp}} = Bnp. \tag{4.46}$$

Die Rekombinationskoeffizienten B haben Werte $B = 1 \cdot 10^{-10} \ldots 7 \cdot 10^{-10}\,\mathrm{cm^3\,s^{-1}}$ bei GaAlAs und $B = 8{,}6 \cdot 10^{-11}\,\mathrm{cm^3\,s^{-1}}$ bei InGaAsP.

Da im thermischen Gleichgewicht (d.h. $np = n_i^2$) die Generationsrate jedes einzelnen Prozesses gleich der Rekombinationsrate sein muss (Prinzip des detaillierten Gleichgewichts), müssen wir für die Elektron-Loch Paare, welche durch Absorption der thermischen Hintergrundstrahlung erzeugt werden fordern, dass

$$g_{\mathrm{s}} = Bn_i^2\ . \tag{4.47}$$

Denn nur dann wird im thermischen Gleichgewicht $g_{\mathrm{s}} - r_{\mathrm{opt}} = 0$.Die spontane Generationsrate ändert sich mit Dotieren nicht wesentlich, da bei der Generation ja auf das unerschöpfliche Meer von Elektronen im Valenzband und die Unzahl von freien Zuständen im Leitungsband zurückgegriffen wird.

2. Nichtstrahlende Übergänge: Shockley-Read-Hall

Hier wird die nichtstrahlende indirekte Rekombination über eine Rekombinationsstörstelle der Energie W_T (T für Trap/Fangstelle, z.B. tiefer Donator oder Akzeptor) im verbotenen Band untersucht, s. Bild 3.7. Die Rekombinationsraten (Anzahl pro Volumen und Zeit), siehe Bild 4.11(b), von Elektronen und Löchern werden mit r_{nt} und r_{pt}, die Generationsraten mit g_{nt} und g_{pt} bezeichnet. Zunächst wird nur der Transfer über eine Störstelle betrachtet.

Die Erzeugungsraten für ein Elektron bzw. Loch sind proportional zu den Wahrscheinlichkeiten w und $1 - w$, mit der die Störstelle bereits mit einem Elektron besetzt ist, bzw. nicht besetzt ist (s. Bild 4.11 (b)) und proportional zu der Zahl der Störstelen n_{t}. Die Proportionalitätskonstanten sind A_{gn} und A_{gp}, bzw. A_{n} und A_{p} wenn wir die Störstellenzahl n_{t} in die Konstanten einbeziehen. Damit sind die Generationsraten

$$g_{\mathrm{nt}} = A_{\mathrm{gn}}\, n_{\mathrm{t}} w \qquad = A_{\mathrm{n}} w, \tag{4.48}$$

$$g_{\mathrm{pt}} = A_{\mathrm{gp}}\, n_{\mathrm{t}} (1 - w) = A_{\mathrm{p}} (1 - w). \tag{4.49}$$

Die Rekombinationsraten sind proportional zur Anzahl der vorhandenen Träger, n oder p, der Störstellenzahl und deren Besetzungswahrscheinlichkeiten (s. Bild 4.11 (b)). Die Proportionalitätskonstanten sind A_{rn} und A_{rp}. Es ist klar, dass die Störstellen die Lebensdauer der Elektronen und Löcher infolge von Rekombination verringern werden. Die Lebensdauer wird mit zunehmender Störstellendichte abnehmen. In Anbetracht dessen führen wir bereits jetzt die *Elektronen und*

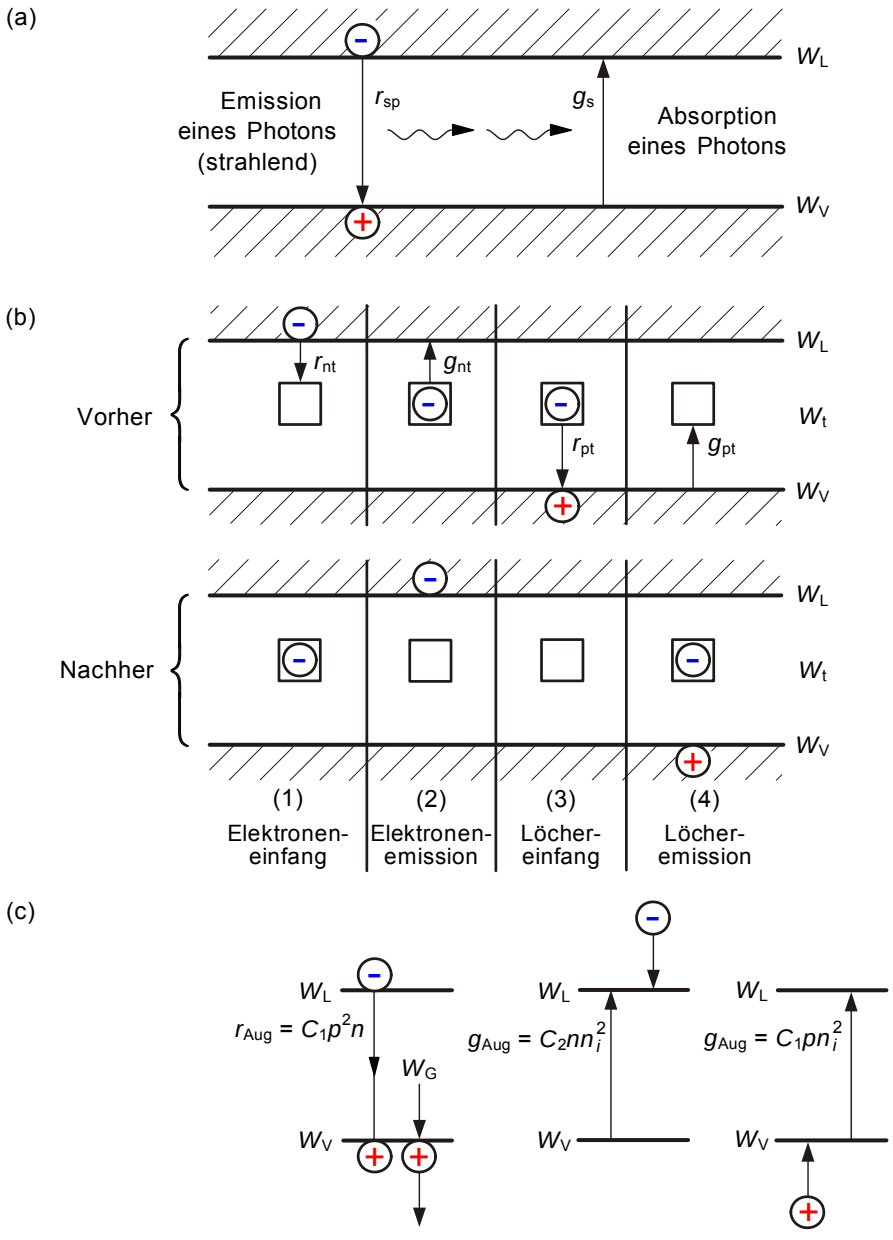


Abbildung 4.11: Einige wichtige Generations- und Rekombinationsprozesse: (a) Strahlende Übergänge. (b) Übergänge vermittelt durch Rekombinationszentren. (c) Auger-Prozesse.

Löcherlebensdauerkonstanten, τ_n und τ_p, ein. Wir schreiben dann für die Rekombinationsraten:

$$r_{\rm nt} = A_{\rm rn}\, n\, n_t(1-w) \equiv \frac{n\,(1-w)}{\tau_n}, \tag{4.50}$$

$$r_{\rm pt} = A_{\rm rp}\, p\, n_t w \qquad\quad \equiv \frac{p\,w}{\tau_p}. \tag{4.51}$$

Damit haben wir neun Unbekannte aber nur vier Gleichungen. Da τ_n und τ_p messbare Größen sind benötigen wir statt der 5 nunmehr noch 3 weitere Gleichungen. Dazu lösen wir das System vorerst im thermischen Gleichgewicht (d.h.$g_{\rm ext} = 0$). Dort gilt nämlich das Prinzip des detaillierten Gleichgewicht. Nach dem Prinzip der detaillierten Gleichgewicht müssen im thermischen Gleichgewicht (d.h $n = n_{th}, p = p_{th}$) auch für diesen Einzelprozess die Generationsraten für Löcher und Elektronen den entsprechenden Rekombinationsraten entsprechen. Das liefert uns zwei Gleichungen

$$g_{\rm nt} = r_{\rm nt} \quad \text{und} \quad g_{\rm pt} = r_{\rm pt}\ . \tag{4.52}$$

Ferner gilt für die Störstellenbesetzung im thermischen Gleichgewicht

$$w_{th} = \frac{1}{1 + e^{(W_T - W_F)/kT}}.$$

Damit erhält man für $A_{\rm n}$ und $A_{\rm p}$ (alles gerechnet für $n = n_{th}, p = p_{th}$:

$$A_{\rm n} = \frac{n_{th}}{\tau_n}\frac{1-w_{th}}{w_{th}} \quad \text{und} \quad A_{\rm p} = \frac{p_{th}}{\tau_p}\frac{w_{th}}{1-w_{th}},$$

bzw. mit

$$\frac{1-w_{th}}{w_{th}} = e^{(W_T - W_F)/kT}$$

folgt

$$A_{\rm n} = \frac{n_{th}\exp\left(-\dfrac{W_F - W_T}{kT}\right)}{\tau_n} \text{ und } \quad A_{\rm p} = \frac{p_{th}\exp\left(-\dfrac{W_T - W_F}{kT}\right)}{\tau_p}. \tag{4.53}$$

Für allgemeine Konzentrationen n und p in den Bändern, welche durch einen Nichtgleichgewichts-Vorgang aufrecht erhalten werden, ergibt sich durch Einsetzen von Gl. (4.53) in die Generationsraten und Rekombinationsraten Gl. (4.48 - 4.51) dann die Dynamik

$$\begin{aligned} g_{\rm nt} &= \frac{n_{th}\exp\left(\dfrac{W_T - W_F}{kT}\right)w}{\tau_n}, & r_{\rm nt} &= \frac{n(1-w)}{\tau_n}, \\ g_{\rm pt} &= \frac{p_{th}\exp\left(\dfrac{W_F - W_T}{kT}\right)(1-w)}{\tau_p}, & r_{\rm pt} &= \frac{pw}{\tau_p}. \end{aligned} \tag{4.54}$$

Man beachte, dass die Generation von Elektronen ins Leitungsband exponentiell wächst, falls die Rekombinationszentren in der Nähe des Leitungsbandes liegen, weil dann $W_T - W_F$ groß wird. Umgekehrt wächst die Generationsrate für Löcher, falls die Rekombinationszentren in der Nähe des Valenzbandes liegen. Es macht Sinn folgende Abkürzungen einzuführen

$$n'_{th} = n_{th}\exp\left(\frac{W_T - W_F}{kT}\right), \tag{4.55}$$

$$p'_{th} = p_{th}\exp\left(\frac{W_F - W_T}{kT}\right). \tag{4.56}$$

Im stationären Zustand unter nicht thermischem Gleichgewicht ($w \neq w_{th}, n \neq n_{th}$ und $n \neq n_{th}$) gilt zudem

$$\frac{dn}{dt} = g_n - r_n = g_{\mathrm{ext}} + g_{\mathrm{nt}} - r_{\mathrm{nt}} = 0 \tag{4.57}$$

$$\frac{dp}{dt} = g_p - r_p = g_{\mathrm{ext}} + g_{\mathrm{pt}} - r_{\mathrm{pt}} = 0\ . \tag{4.58}$$

Nach Elimination von $g_{\mathrm{ext}} \neq 0$ gilt:

$$g_t - r_t = g_{nt} - r_{nt} = g_{pt} - r_{pt}\ . \tag{4.59}$$

Gleichung (4.59) ist eine Bestimmungsgleichung für w. Das Ergebnis lautet:

$$w = \frac{n\tau_p + p'_{th}\tau_n}{(n + n'_{th})\tau_p + (p + p'_{th})\tau_n}, \tag{4.60}$$

$$1 - w = \frac{n'_{th}\tau_p + p\tau_n}{(n + n'_{th})\tau_p + (p + p'_{th})\tau_n}. \tag{4.61}$$

Setzt man Gl. (4.60) in (4.54) ein, so folgt unter Beachtung von $n'_{th}p'_{th} = n_i^2$ das Ergebnis

$$g_t\ -\ r_t\ =\ g_{\mathrm{pt}}\ -\ r_{\mathrm{pt}}\ =\ g_{\mathrm{nt}}\ -\ r_{\mathrm{nt}}\ =\ \frac{n_i^2 - np}{(n + n'_{th})\tau_p + (p + p'_{th})\tau_n} \tag{4.62}$$

Die wirksamste Rekombination erfolgt dann, wenn die Energie W_T der Störstelle so liegt, dass $n'_{th}\tau_p + p'_{th}\tau_n$ minimal wird. Für $\tau_n = \tau_p = \tau_0$ heißt das, dass $n'_{th} + p'_{th}$ (unter der Nebenbedingung $n'_{th}p'_{th} = n_i^2$) minimal werden muß: Das ist für $n'_{th} = p'_{th} = n_i$ der Fall (d. h. die Rekombinationsstörstelle liegt genau in der Mitte des verbotenen Bandes). Man erhält das Ergebnis

$$g_t - r_t = \frac{n_i^2 - np}{(n + p + 2n_i)\tau_0}. \tag{4.63}$$

Bei starker Trägerinjektion schreibt man oft in Analogie zu Gl. (**??**)

$$\boxed{g_t\text{-}r_t = A\sqrt{|n_i^2 - np|}}. \tag{4.64}$$

3. Auger-Prozesse

Ein weiterer Generations-Rekombinationsmechanismus ist der sogenannte Auger-Prozess. Fig. 4.11(c) zeigt das Prinzip im einfachen Bändermodell. Ein Elektron aus dem LB vernichtet ein Loch im VB und die freiwerdende Energie W_G wird dazu verwendet, um ein weiteres Elektron im LB in einen um W_G höherliegenden Leitungsbandzustand zu transferieren. Die Energie W_G kann auch dazu verwendet werden, um ein weiteres Loch, welches sich im VB befindet, in einen um W_G tieferliegenden Valenzbandzustand zu drücken. Die zugehörigen Rekombinationsraten lauten

$$r_{\mathrm{Aug}} = C_1 np^2 + C_2 n^2 p. \tag{4.65}$$

Bei GaAlAs spielen Auger-Prozesse keine merkliche Rolle. Bei InGaAsP ist $C_2 \ll C_1 = 4 \cdot 10^{-29}\,\mathrm{cm}^6\,\mathrm{s}^{-1}$. Bei Si ist $C_1 \ll C_2 = 2 \cdot 10^{-31}\,\mathrm{cm}^6\,\mathrm{s}^{-1}$. Trotz der Kleinheit des Auger-Koeffizienten C_2 kann die Auger-Rekombination in Hochinjektionszonen (wo gilt $n \approx p$) dominierend sein: Nach Gl. (4.64) und Gl. (4.65) steigt in diesem Fall die indirekte Rekombination wie n, die Auger-Rekombination aber wie n^3.

Berücksichtigt man wie oben die Generationseffekte, so erhält man

$$g_{\mathrm{Aug}} - r_{\mathrm{Aug}} = C_1 p(n_i^2 - np) + C_2 n(n_i^2 - np). \tag{4.66}$$

4.4.3 Trägerlebensdauer

In diesem Kapitel wollen wir den Begriff der Minoritätsträgerlebensdauer und deren eventuelle Dichteabhängigkeit untersuchen.

1. Minoritätsträgerlebensdauer der spontanen Emission

Dazu betrachten wir einen Halbleiter mit direkten Übergängen (GaAs), so dass die Generations- und Rekombinationsprozesse durch die optischen Übergänge dominiert werden, und vernachlässigen alle anderen Prozesse. Bei externer Bestrahlung mit Licht, wie in Bild 4.10 dargestellt, muss die totale Generationsrate g durch

$$g = g_s + g_{ext} \tag{4.67}$$

ersetzt werden. Dabei bezeichnet g_s die von der thermischen Hintergrundstrahlung, siehe Abschnitt 4.4.2, und g_{ext} die vom externen Licht pro Zeiteinheit erzeugten Trägerpaare. Einsetzen von (4.46) und (4.47) in (4.39) ergibt

$$\frac{dn}{dt} = \frac{dp}{dt} = g_{ext} + B\left(n_i^2 - np\right) . \tag{4.68}$$

Es interessiert die Abweichung vom Gleichgewicht gemäß (4.37). Wir erhalten mit (4.68).

$$\begin{aligned}\frac{dn}{dt} = \frac{dp}{dt} &= g_{ext} + B\left[n_i^2 - n_{th}p_{th}\right] - B\left[\Delta n\, p_{th} + \Delta p\, n_{th} + \Delta n\, \Delta p\right] \\ &= g_{ext} - B\left[\Delta n\, p_{th} + \Delta p\, n_{th} + \Delta n\, \Delta p\right] ,\end{aligned} \tag{4.69}$$

wobei die erste eckige Klammer nur Gleichgewichtsgrößen beinhaltet und deshalb verschwindet. Da die Überschussträgerdichten aufgrund der Paarerzeugung und Rekombination gleich sind, gilt $\Delta n = \Delta p$ und mit $dn/dt = d\Delta n/dt$ bzw. $dp/dt = d\Delta p/dt$ vereinfachen sich diese Gleichungen zu

$$\frac{d\Delta n}{dt} = g_{ext} - B\left[p_{th} + n_{th} + \Delta n\right]\Delta n. \tag{4.70}$$

$$\frac{d\Delta p}{dt} = g_{ext} - B\left[p_{th} + n_{th} + \Delta p\right]\Delta p \tag{4.71}$$

Fall A: Kleine Überschussträgerdichten: Für kleine Überschußträgerdichten (Low-Level Injection) $\Delta n \ll p_{th} + n_{th}$, und $\Delta p \ll p_{th} + n_{th}$, d.h. die Überschussträgerdichten sind sehr viel kleiner als die der Majoritätsträger, können wir Δn und Δp in der Klammer vernachlässigen und erhalten einfache gewöhnliche Dgl. erster Ordnung

$$\frac{d\Delta n}{dt} = -\frac{1}{\tau_{\min}}\Delta n + g_{ext} , \tag{4.72}$$

$$\frac{d\Delta p}{dt} = -\frac{1}{\tau_{\min}}\Delta p + g_{ext} \tag{4.73}$$

wobei wir die Minoritätsträgerlebensdauer $\tau_{\min}$

$$\tau_{\min}^{-1} = B\left[p_{th} + n_{th}\right] . \tag{4.74}$$

eingeführt haben.

Für einen stufenförmigen Zeitverlauf des eingestrahlten Lichtes ($g_{ext} \neq 0$ für $t \leqq 0$) und damit der Überschussgenerationsrate g_L nach Bild 4.10(b) reduzieren sich die Differentialgleichungen auf den stationären Fall und es gilt

$$0 = -\frac{1}{\tau_{\min}}\Delta n + g_{ext}\,, \tag{4.75}$$

$$0 = -\frac{1}{\tau_{\min}}\Delta p + g_{ext} \tag{4.76}$$

Damit wird die Überschussladungsträgerdichte für $t \leqq 0 : \Delta n\,(0) = \Delta p\,(0) = g_{ext}\tau_{\min}$.

Für $g_{ext} = 0$ für $t > 0$ und damit der Überschussgenerationsrate g_{ext} nach Bild 4.10(b) erhalten wir die Lösung

$$\Delta n\,(t) = g_{ext}\tau_{\min} e^{-t/\tau_{\min}} \tag{4.77}$$

$$\Delta p\,(t) = g_{ext}\tau_{\min} e^{-t/\tau_{\min}}\,. \tag{4.78}$$

Es gilt also: Im Falle schwacher Injektion kann die Majoritätsträgerdichte als konstant angenommen werden und nur die Minoritätsträgerdichte unterliegt einer Dynamik.

	Schwache Injektion	
n-Typ	i-Typ	p-Typ
$n_n \ggg p_n$	$p = n$	$p_p \ggg n_p$
$n_n \approx n_{nth}$	$n \approx n_{th} + \Delta n$	$n_p = n_{pth} + \Delta n$
$p_n = p_{nth} + \Delta p$	$p \approx p_{th} + \Delta p$	$p_p \approx p_{pth}$
$\tau_{\min}^{-1} \sim B n_n$	$\tau_{\min}^{-1} = B\,[p_{th} + n_{th}]$	$\tau_{\min}^{-1} \sim B p_p$
	$\Delta n = \Delta n_0\, e^{-t/\tau_{\min}}$,	$\Delta n = \Delta n_0\, e^{-t/\tau_{\min}}$
$\Delta p = \Delta p_0\, e^{-t/\tau_{\min}}$	$\Delta p = \Delta p_0\, e^{-t/\tau_{\min}}$,	

(4.79)

Gl. (4.74) zeigt, dass die Minoritätsträgerlebensdauer in einem Halbleiter umgekehrt proportional zur Gesamtladungsträgerdichte ist. Diese einfache Berechnung ist allerdings mit Vorsicht zu geniessen, denn in einem realen Halbleiter liefern die unterschiedlichsten Prozesse, strahlende und nichtstrahlende, einen Beitrag. Im Allgemeinen ist die Trägerlebensdauer einer Halbleiterprobe ein experimentell zu bestimmender Parameter, siehe Bild 4.10(c), der ein Maß für die Qualität der Probe darstellt und der für eine reale Probe nur schwer zu berechnen ist. Typische Bereiche für die Minoritätsträgerlebensdauer sind

Ge:	10^{-6} s	bis	10^{-3} s
Si:	10^{-10} s	bis	10^{-3} s
GaAs:	10^{-10} s	bis	10^{-8} s.

Infolge der direkten Übergänge zeigt GaAs auch bei hoher Reinheit nur eine Ladungsträgerlebensdauer im Bereich von einigen Nanosekunden.

Fall B: Starke Trägerinjektion: Wenn die Voraussetzung für schwache Injektion nicht eingehalten wird, kann Gl. (4.70) nicht mehr genähert werden. Wir betrachten also den High-Level-Injection (HLI) Fall. Unter Benutzung der Minoritätsträgerlebensdauer $\tau_{\min}$ kann Gl. (4.70) geschrieben werden als

$$\frac{d\Delta n}{dt} = g_{ext}(t) - \frac{1}{\tau_{\min}}\left[1 + \frac{\Delta n}{p_{th} + n_{th}}\right]\Delta n. \tag{4.80}$$

Wie schon vorher erwähnt, ist die exakte Lösung dieser Gleichung schwierig. Wir können aber sehr einfach die stationäre Trägerdichte finden, wenn $g_{ext}(t) = g_{ext}$ konstant ist oder sich innerhalb der Trägerlebensdauer nur unwesentlich ändert.

Fall B1: Stationäre Trägerdichte Aus $\frac{d\Delta n}{dt} = 0$ folgt durch Lösung der aus Gl. (4.80) folgenden quadratischen Gleichung:

$$\Delta n_0 = \frac{(p_{th} + n_{th})}{2} \left[\sqrt{1 + \frac{4 g_{ext} \tau_{\min}}{p_{th} + n_{th}}} - 1 \right]. \tag{4.81}$$

Durch Entwicklung der Wurzel ist leicht nachzuweisen, dass sich für den Fall schwacher Injektion, $g_{ext}\tau_{\min} \ll p_{th} + n_{th}$, der stationäre Wert für schwache Injektion ergibt. Die allgemeine Lösung (4.81) zeigt aber, dass immer $\Delta n_0 < g_{ext}\tau_{\min}$ gilt. Im Grenzfall sehr hoher Injektion folgt

$$\Delta n_0 = \sqrt{g_{ext}\tau_{\min}\,(p_{th} + n_{th})}, \tag{4.82}$$

d.h. die stationäre Trägerdichte ist nicht mehr proportional zur Generationsrate wie im Falle schwacher Injektion, sondern nur noch zur Wurzel der Generationsrate.

Fall B2: Transienter Vorgang bei starker Trägerinjektion Wird eine starke Überschussträgerdichte aufgebaut, so dass in Gl.(4.80) der quadratische Term überwiegt, so gehorcht die Überschussträgerdichte nach Abschalten der Trägererzeugung der Gleichung

$$\frac{d\Delta p}{dt} = -\frac{1}{\tau_{\min}} \frac{\Delta p^2}{p_{th} + n_{th}}, \tag{4.83}$$

welche durch Separation der Variablen wiederum sehr einfach zu lösen ist. Es folgt mit der Anfangspopulation $\Delta p_0 = \Delta p(0)$

$$\Delta p(t) = \frac{\Delta p_0}{1 + \left[\frac{\Delta p_0}{p_{th}+n_{th}} \frac{t}{\tau_{\min}} \right]}. \tag{4.84}$$

Solange die Überschussträgerdichte groß gegen die thermischen Dichten sind, zerfallen die Überschussträgerdichten wesentlich schneller als nur exponentiell, vgl. (4.80), nämlich umgekehrt proportional zur Zeit:

$$\Delta p(t) \sim \frac{(p_{th} + n_{th})\,\tau_{\min}}{t}. \tag{4.85}$$

Dies gilt für den Bereich $\frac{p_{th}+n_{th}}{\Delta n_0}\tau_{\min} < t < \tau_{\min}$. Vorausgesetzt, die Halbleiterprobe befindet sich zu Beginn im Bereich der starken Injektion so zerfällt die Überschussträgerdichte unabhängig von ihrem Anfangswert innerhalb einer Minoritätsträgerlebensdauer auf die Majoritätsträgerdichte, siehe Bild 4.12.Danach befindet sich die Probe im Bereich schwacher Injektion und die verbleibende Überschussträgerdichte zerfällt nur noch exponentiell mit der Minoritätsträgerlebensdauer.

2. Trägerlebensdauer bei dominanter Rekombination von Störzentren

Wir betrachten einen Halbleiter, der wieder von außen angeregt wird und dessen dominanter Rekombinationsprozess durch Störzentren gegeben sei. In diesem Fall gilt

$$\frac{dn}{dt} = \frac{dp}{dt} = g_{ext} + \frac{n_i^2 - np}{(n + n'_{th})\tau_p + (p + p'_{th})\tau_n}. \tag{4.86}$$

Beschränken wir unsere Betrachtung auf einen n-Halbleiter, so wird eine merkliche Störung der Minoritätsträgerdichte p_n vorgenommen. Konkret werden also $\Delta n = \Delta p$ Ladungsträger eingeschossen (dadurch wird bei einer Neutralstörung die Majoritätsträgerdichte nicht wesentlich geändert) und wir verwenden wieder analog zu oben:

$$\left.\begin{array}{l} n_n = n_{nth} + \Delta n \approx n_{nth} \\ p_n = p_{nth} + \Delta p \approx \Delta p \end{array}\right\} \quad n_{nth} p_{nth} = n_i^2, \quad n_n \gg n'_{th}, p'_{th}, \quad n_n \gg \Delta n. \tag{4.87}$$

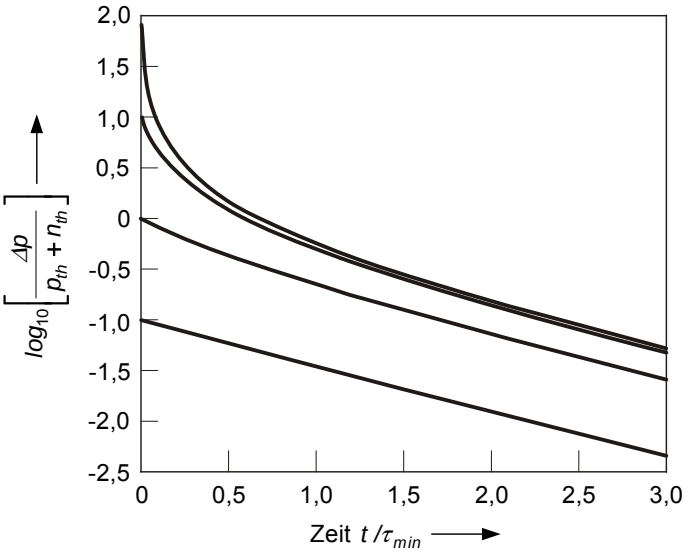


Abbildung 4.12: Relaxation der Überschußladungsträgerdichte für verschiedene Anfangswerte.

Damit erhalten wir nach Einsetzen der Vereinfachungen

$$\frac{d\Delta p}{dt} = g_{ext} - \frac{\Delta p}{\tau_p} \tag{4.88}$$

und als Lösung, falls $g_{ext} = 0$ für $t > 0$:

$$\Delta p\,(t) = g_{ext}\tau_{\min} e^{-t/\tau_{\min}} \tag{4.89}$$

mit

$$\tau_{\min} = \tau_p \text{ für den n-Halbleiter,} \tag{4.90}$$

$$\tau_{\min} = \tau_n \text{ für den p-Halbleiter.} \tag{4.91}$$

Die physikalische Bedeutung der Zeitkonstanten τ_n und τ_p wird damit klar. τ_p ist die Lebensdauer der Löcher (Minoritäten) im n-Halbleiter und wird daher auch als Minoritätsträgerlebensdauer bezeichnet. Analog ist τ_n die Lebensdauer der Elektronen (Minoritäten) im p-Halbleiter. Die Minoritätsträgerlebensdauer $\tau_{n,p}$ ist umgekehrt proportional zur Konzentration der Fangstellen n_T, worauf bereits bei Gl.(4.50) hingewiesen wurde. Die Minoritätsträgerlebensdauer in unreinen Halbleitern wird daher von nichtstrahlenden Prozessen bestimmt.

3. Trägerlebensdauer bei dominanter Augerrekombinationen

Augerrekombination dominiert bei großen Trägerkonzentrationen, weil dieser Term mit der dritten Potenz der Ladungsträger ansteigt. Da Löcher und Elektronen gleichermaßen am Prozess beteiligt sind, sind Trägerinjektionen wichtiger als Dotierungen. Mit $n \approx p$ und $n_i \ll n, p$ ergibt sich

$$\frac{dn}{dt} = \frac{dp}{dt} = g_{ext} + g_{\mathrm{Aug}} - r_{\mathrm{Aug}} = g_{ext} + C_1 p\left(n_i^2 - np\right) + C_2 n\left(n_i^2 - np\right). \tag{4.92}$$

$$\cong g_{ext} - (C_1 + C_2)\, n^3 \tag{4.93}$$

womit sich die Differentialgleichung mit $g_{ext} = 0$ zu

$$\frac{d\Delta n}{dt} = g_{ext} - \left.\frac{\partial(g_{Aug} - r_{Aug})}{\partial n}\right|_n \Delta n = g_{ext} - \frac{1}{\tau_{\mathrm{eff}}}\Delta n \tag{4.94}$$

mit der Lösung, falls $g_{ext} = 0$ für $t > 0$

$$n\,(t) = g_{ext}\tau_{\text{eff}} e^{-t/\tau_{\text{eff}}} \tag{4.95}$$

vereinfacht. Diese Gleichung hat die Lösung

$$\frac{1}{\tau_{\text{eff}}} = 3(C_1 + C_2)\, n^2 \tag{4.96}$$

Für Trägerkonzentrationen größer als etwa $n = p = 10^{17}\,\text{cm}^{-3}$ wird die effektive Trägerlebensdauer τ_{eff} aus (4.62) und (4.65) merklich konzentrationsabhängig. Sie nimmt somit mit wachsender Trägerkonzentration wie $1/n^2$ ab.

4. Diskussion der Hochinjektion von Ladungsträgern

Im Falle von Hochinjektion unter den Annahmen

$$n \approx p, \quad \text{und } n_i \ll n, p \tag{4.97}$$

und unter Berücksichtigung von Rekombinationszentren, spontaner Emission und Auger-Rekombination wird der Generations- bzw. Rekombinationsterm zu

$$g - r = -A\sqrt{np} - Bnp - C_1 np^2 - C_2 n^2 p \tag{4.98}$$

$$\cong -An - Bn^2 - (C_1 + C_2)\, n^3. \tag{4.99}$$

Damit lässt sich die Differentialgleichung für die Ladungsträgerkonzentration schreiben als

$$\frac{dn}{dt} = g_{ext} - [g - r]\,. \tag{4.100}$$

Diese Gleichung lässt sich anders schreiben. Mit $dn/dt = d\Delta n/dt$ und einer Taylorentwicklung um n für den Term $[g - r]$ erhält man:

$$\frac{d\Delta n}{dt} = g_{ext} - \left.\frac{\partial(g - r)}{\partial n}\right|_n \Delta n \tag{4.101}$$

welche als Lösung eine effektive Ladungsträgerlebensdauer von

$$\frac{\partial(g - r)}{\partial n} = A + 2Bn + 3\,(C_1 + C_2)\, n^2 = \frac{1}{\tau_{\text{eff}}} \tag{4.102}$$

liefert. Wenn wir für A die eben Definierten Störstellenlebensdauer einsetzen erhalten wir für die effektive Lebensdauer also

$$\frac{1}{\tau_{\text{eff}}} = \frac{1}{\tau_p + \tau_n} + 2Bn + 3\,(C_1 + C_2)\, n^2\,. \tag{4.103}$$

4.4.4 Photoleitung und Photoleiter

Legen wir ein homogenes $\vec{E}-$Feld an eine Halbleiterprobe an, so hängt der durch die Probe fließende Strom $J = \sigma E$ direkt von der Ladungsträgerdichte

$$\sigma = e\left[\mu_n n + \mu_p p\right]. \tag{4.104}$$

und damit von der Stärke des eingestrahlten Lichtes ab. Dieses Phänomen nennt man ***Photoleitung*** und auf ihm basiert das wohl einfachste und älteste Halbleiterbauelement. CdS-Photodetektoren sind seit langem in der Photographie als Belichtungsmesser im Einsatz. Trotz der langen Geschichte ist die Entwicklung von effizienten und vor allem schnellen Photoleitern noch immer Gegenstand

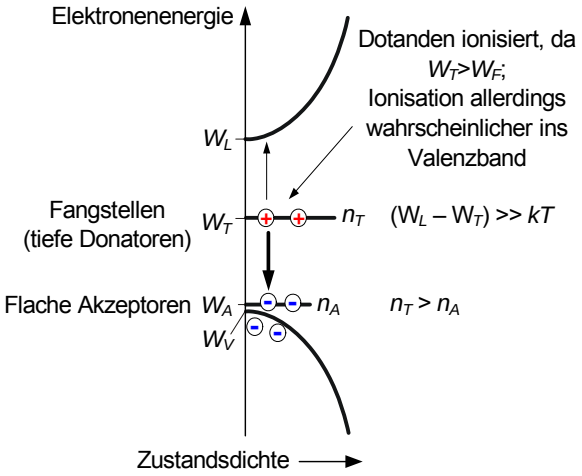


Abbildung 4.13: Relative Lagen von tiefen Donatoren und flachen Akzeptoren in einem Photoleiter [6].

der Forschung. Einige der schnellsten Photodetektoren beruhen auf bei niedrigen Temperaturen gewachsenem oder speziell dotiertem GaAs.

Wir schreiben die Leitfähigkeit in Form der Gleichgewichts- und Überschussladungsträgerdichten

$$\sigma = e\left[\mu_n\left(n_{th} + \Delta n\right) + \mu_p\left(p_{th} + \Delta p\right)\right] \tag{4.105}$$

$$= \sigma_0 + \Delta\sigma \tag{4.106}$$

mit

$$\sigma_0 = q\left[\mu_n n_{th} + \mu_p p_{th}\right], \tag{4.107}$$

$$\Delta\sigma = q\left[\mu_n + \mu_p\right]\Delta n, \tag{4.108}$$

wobei σ_0 die Leitfähigkeit bei thermischen Gleichgewicht und $\Delta\sigma$ die Überschussleitfähigkeit unter Beachtung von $\Delta n = \Delta p$ ist. Der Zusammenhang zwischen der Überschußladungsträgerdichte und der Beleuchtung der Probe ist aus dem vorhergehenden Abschnitt bereits bekannt.

Idealerweise sollte die Leitfähigkeit bei keinem Lichteinfall verschinden und bei wenig Lichteinfall stark ansteigen. Eine erste mögliche zur Lichtdetektion unabhängig von Schaltgeschwindigkeiten wäre wie folgt: Die Überschussleitfähigkeit ist im Allgemeinen klein, wenn die Beweglichkeit von Elektronen und Löchern gleich ist. Als Lösung könnte man dann also schwach dotiertes Material verwenden. Im LLI-Betrieb ist dann der Widerstand groß. Im HLI-Betrieb hingegen verschwindet der hohe Widerstand. Im HLI-Bereich ist der Zusammenhang zwischen Trägerdichte und Beleuchtungsstärke allerdings nicht mehr linear.

Alternative könnte man einen stark p-dotierten HL verwenden. Verwendet man einen p-HL, mit kleinen Löcherbeweglichkeiten kann man in Abwensenheit von Licht tatsächlich nur einen kleinen Strom detektieren. Bei Lichteinfall werden dann gleich viele Löcher wie Elektronen eingeschossen und dank der hohen Elektronenleitfähigkeit kann man einen dem Lichteinfall fast proportionalen Strom messen.

Besser wäre ein Photoleiter, dessen Majoritätsträger effektiv unbeweglich sind und die Minoritätsträger eine hohe Beweglichkeit haben. In solchen Photoleitern, (z.B. ein p-HL mit $\mu_p = 0$) verschwindet die Gleichgewichtsleitfähigkeit und es gilt $\sigma = \Delta\sigma = e\mu_n\,\Delta n$, welche nur bei Beleuchtung existiert.

Diese Situation wird mit der in Bild 4.13 angedeuteten Dotierung erreicht. Den p-Halbleiter erreichen wir durch Dotierung mit sehr flachen Akzeptoren der Konzentration n_A. Zusätzlich

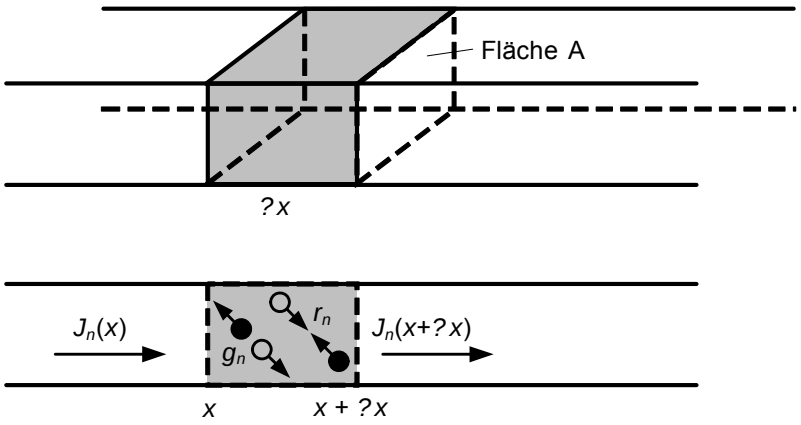


Abbildung 4.14: Stromfluss, Erzeugung und Rekombination in einer Scheibe der Dicke Δx.

wird eine große Anzahl von sehr tiefen Donatoren der Konzentration n_T eingebaut, d.h. für die Donatorenergie W_T gilt $W_L - W_T \gg kT$. Diese wirken, wie in Abschnitt 4.4.2 besprochen, als Rekombinationsstörstellen und verringern die Minoritätsträger-Lebendauer. Der Trick ist dieser: Obwohl die Donatoren im thermischen Gleichgewicht nicht ans weit entfernte Leitungsband ionisiert werden um dort ein Elektron abzugeben, so werden doch n_A von ihnen ionisiert und ionisieren die Akzeptoren. Idealerweise sorgt man dafür, dass gilt

$$n_{T0}^{+} = n_A.$$

Die n_A ionisierten Fangstellen können jetzt keine weiteren Löcher generieren und sind daher absolut unbeweglich, es gilt $\mu_p = 0$. Wird jetzt infolge eines Absorptionsvorganges ein Elektron-Lochpaar erzeugt, so wird das Loch schnell durch weitere Ionisation eines unbeweglichen Donatoratoms aufgefüllt und es bleibt nur ein beweglicher Minoritätsträger übrig, welcher zur Überschussleitfähigkeit beiträgt.

Die äquivalente Bedingung zur schwachen Injektion ist nun, dass die Überschussträgerdichte kleiner als n_{T0}^{+} bleibt, d.h. $n' \ll n_A$. Für den Betrieb des Bauelementes gilt (**??**), wobei jetzt die neue Trägerlebensdauer, welche sich durch den Elektroneneinfang in den ionisierten Fangstellen ergibt, verwendet werden muss.

4.5 Die Kontinuitätsgleichung

In den vorangehenden Unterkapiteln haben wir die verschiedenen Ladungsträgertransporteffekte und Ladungsträgererzeugungsmechanismen systematisch aufgelistet. Wir suchen nun eine Gleichung, welche die Gesamtbilanz aller Beiträge aus Ladungsträgerdrift, -diffusion sowie Rekombination und Generation in einem Volumenelement beschreibt. Diese Gleichung ist die ***Kontinuitätsgleichung.***

Zur Herleitung der Kontinuitätsgleichung betrachten wir zunächst die Veränderung der Elektronendichte in einem Würfel entlang der x-Achse, Fig. 4.14.

Die Rate der Elektronenladungsdichteänderungen ergibt sich als Differenz der Nettoströme durch die Flächen bei x und $x + dx$, sowie der Nettoerzeugungs- und Rekombinationsterme

$$\frac{\partial n}{dt} A \,\mathrm{d}x = \left[\frac{J_n(x)A}{-e} - \frac{J_n(x+\mathrm{d}x)A}{-e} \right] + (g_n - r_n)\, A \,\mathrm{d}x. \tag{4.109}$$

Den Strom-Term $J_n(x+\mathrm{d}x)$ kann als Taylorserie um $J_n(x)$ schreiben

$$J_n(x+\mathrm{d}x) = J_n(x) + \frac{\partial J_n(x)}{\partial x}\mathrm{d}x + ... \tag{4.110}$$

Einsetzen in Gl. (4.109) liefert die Kontinuitätsgleichung für die Elektronen

$$-e\frac{\partial n}{dt} = -\frac{\partial J_n(x)}{\partial x} - e\,(g_n - r_n)\ . \tag{4.111}$$

Analog erhält man die Kontinuitätsgleichung für die Löcher

$$e\frac{\partial p}{dt} = -\frac{\partial J_p(x)}{\partial x} + e\,(g_p - r_p)\ . \tag{4.112}$$

In drei Dimensionen lauten die Kontinuitätsgleichungen also

$$\begin{array}{rcl} \dfrac{\partial(ep)}{\partial t} + \operatorname{div}\vec{J}_p &=& e(g_p - r_p), \\ \dfrac{\partial(-en)}{\partial t} + \operatorname{div}\vec{J}_n &=& -e(g_n - r_n), \end{array} \tag{4.113}$$

wobei die Ströme als Summen des Drift- und des Diffusionstromes aufzufassen sind, Glgn (4.3) und (4.21-4.23)

$$\vec{J}_n = \vec{J}_{nF} + \vec{J}_{nD}, \qquad \vec{J}_n = en\mu_n\vec{E} + eD_n \operatorname{grad} n, \tag{4.114}$$

$$\vec{J}_p = \vec{J}_{pF} + \vec{J}_{pD}, \qquad \vec{J}_p = ep\mu_p\vec{E} - eD_p \operatorname{grad} p. \tag{4.115}$$

Bemerkung:
Bildet man die Summe der Gleichungen (4.113), ersetzt $\operatorname{div}(\vec{J}_p + \vec{J}_n) = \operatorname{div}\vec{J}$ und setzt für die Raumladungsdichte ρ

$$\rho = e(p + n_D^+ - n - n_A^-), \tag{4.116}$$

so erhält man

$$\frac{\partial\rho}{\partial t} + \operatorname{div}\vec{J} = e\,(g_n - r_n) - e\,(g_p - r_p). \tag{4.117}$$

Falls Gleichgewicht herrscht ($g_n = r_n$ und $g_p = r_p$) oder wenn Träger nur in Paaren erzeugt oder vernichtet werden ($g_n - r_n = g_p - r_p$), gilt die vereinfachte Form der Kontinuitätsgleichung. Diese lautet:

$$\frac{\partial\rho}{\partial t} + \operatorname{div}\vec{J} = 0. \tag{4.118}$$

Umgekehrt beinhaltet Gl. (5.2), eine Beziehung für die Änderung der nicht beweglichen Raumladungsdichte zufolge ionisierter Störstellen

$$\frac{\partial}{\partial t}e\,(n_D^+ - n_A^-) = e\,(g_n - r_n) - e\,(g_p - r_p). \tag{4.119}$$

Kapitel 5

Die Grund-Gleichungen und -Konstanten des Halbleiters

In diesem Kapitel werden wir zuerst die grundlegenden Gleichungen der Halbleitertechnologie aufschreiben und dann mit deren Hilfe wichtige Fallbeispiele diskutieren. Ferner werden wir das Konzept der Quasifermienergien einführen.

5.1 Die drei Halbleitergleichungen

5.1.1 Poisson-Gleichung

Im Folgenden machen wir einige Einschränkungen:

- Die betrachteten Materialien seien nicht magnetisch oder magnetisierbar $\mu = 1$.
- Es seien ferner keine äußeren Magnetfelder vorhanden $\overrightarrow{B} = 0$ und die Eigenmagnetfelder der Ströme können vernachlässigt werden (dies trifft immer dann zu, wenn sich Ladungen mit nichtrelativistischen Geschwindigkeiten bewegen). Unter dieser Vorraussetzungen ist das elektrische Feld wirbelfrei ($\operatorname{rot} \vec{E} = 0$) und kann für ein einfach zusammenhängendes Gebiet von einem Skalarpotential φ gemäß $\vec{E} = -\operatorname{grad} \varphi$ abgeleitet werden.

Die Quelle der dielektrischen Verschiebung ist die Raumladungsdichte ρ

$$\operatorname{div} \vec{D} = \rho \,. \tag{5.1}$$

Unter Verwendung von

$$\vec{D} = \varepsilon \vec{E} = \varepsilon_0 \varepsilon_r \vec{E} \quad \text{und} \quad \vec{E} = -\operatorname{grad} \varphi \tag{5.2}$$

(wobei ε_0 die Permitivität des Vakuums, ε_r die Permitivitätszahl und ε die Permitivität (vormals Dielektrizitätskonstante) ist) folgt

$$\operatorname{div} \vec{D} = -\operatorname{div}(\varepsilon \operatorname{grad} \varphi) = \rho = e(p + n_D^+ - n - n_A^-). \tag{5.3}$$

In einem Material mit konstanter Permitivität ε ergibt sich daraus die Poisson-Gleichung

$$\Delta \varphi = -\frac{\rho}{\varepsilon} = -\frac{e}{\varepsilon}(p + n_D^+ - n - n_A^-). \tag{5.4}$$

5.1.2 Die drei Halbleiter-Gleichungen in der Übersicht

Darunter versteht man die Kontinuitätsgleichungen

$$\begin{aligned} \frac{\partial(ep)}{\partial t} + \operatorname{div} \vec{J}_p &= e(g_p - r_p), \\ \frac{\partial(-en)}{\partial t} + \operatorname{div} \vec{J}_n &= -e(g_n - r_n)\,, \end{aligned} \tag{5.5}$$

wobei der Ladungsträgertransport infolge von Drift und Diffusion gegeben ist durch

$$\vec{J}_n = \vec{J}_{nF} + \vec{J}_{nD}, \qquad \vec{J}_n = en\mu_n\vec{E} + eD_n \operatorname{grad} n, \tag{5.6}$$

$$\vec{J}_p = \vec{J}_{pF} + \vec{J}_{pD}, \qquad \vec{J}_p = ep\mu_p\vec{E} - eD_p \operatorname{grad} p, \tag{5.7}$$

sowie die Poisson-Gleichung:

$$\Delta\varphi = -\frac{\rho}{\varepsilon} = -\frac{e}{\varepsilon}(p + n_D^+ - n - n_A^-). \tag{5.8}$$

Ab und zu ist es nützlich, die beiden Kontinuitätsgleichungen aufzusummieren. Falls Gleichgewicht herrscht ($g_n = r_n$ und $g_p = r_p$) oder wenn Träger nur in Paaren erzeugt oder vernichtet werden ($g_n - r_n = g_p - r_p$) gilt auch

$$\frac{\partial\rho}{\partial t} + \operatorname{div} \vec{J} = 0. \tag{5.9}$$

5.1.3 Halbleiter-Gleichungen bei schwacher Injektion

Im Falle von Störstellenhalbleitern mit schwacher Injektion lassen sich die drei Halbleiter-Grundgleichungen wesentlich vereinfachen. Schwache Injektion bedeutet, dass sich die Majoritätsträgerdichte kaum ändert und aufgrund der Quasineutralität die Minoritätsträgerdichte wesentlich kleiner als die Majoritätsträgerdichte ist. Es kann daher der Driftstrom der Minoritätsträger gegenüber dem der Majoritätsträger vernachlässigt werden, und die Majoritätsträgerdichte kann gleich der ungestörten Majoritätsträgerdichte gesetzt werden. Ferner können die Generations-Rekombinationsterme entsprechend Gl. (4.72-4.73) vereinfacht werden. Damit folgt

n-Typ	p-Halbleiter
$n_n \approx n_{nth} \gg p_n$	$p_p \approx p_{pth} \gg n_p$
$\vec{J}_n = en_n\mu_n\vec{E} + eD_n \operatorname{grad} n_n,$	$\vec{J}_n = eD_n \operatorname{grad} n_p,$
$\vec{J}_p = -eD_p \operatorname{grad} p_n.$	$\vec{J}_p = ep_p\mu_p\vec{E} - eD_p \operatorname{grad} p_p,$
$\frac{\partial(p_n)}{\partial t} + \frac{1}{e}\operatorname{div}\vec{J}_p = -\frac{\Delta p_n}{\tau_p} + g_{ext}$	$\frac{\partial(p_p)}{\partial t} + \frac{1}{e}\operatorname{div}\vec{J}_p = -\frac{\Delta n_p}{\tau_n} + g_{ext}$
$\frac{\partial(n_n)}{\partial t} - \frac{1}{e}\operatorname{div}\vec{J}_n = -\frac{\Delta p_n}{\tau_p} + g_{ext}$	$\frac{\partial(n_p)}{\partial t} - \frac{1}{e}\operatorname{div}\vec{J}_n = -\frac{\Delta n_p}{\tau_n} + g_{ext}$

(5.10)

$$\Delta\varphi = -\frac{\rho}{\varepsilon} = -\frac{e}{\varepsilon}(p + n_D^+ - n - n_A^-) \tag{5.11}$$

Wie aus (5.10) zu ersehen ist, entkoppeln im Falle schwacher Injektion die Minoritätsträgerdichten und können daher separat betrachtet werden. g_{ext} ist hier eine Nichtgleichgewichts-Generationsrate von Elektronen und Löchern, z.B. infolge von externer Bestrahlung.

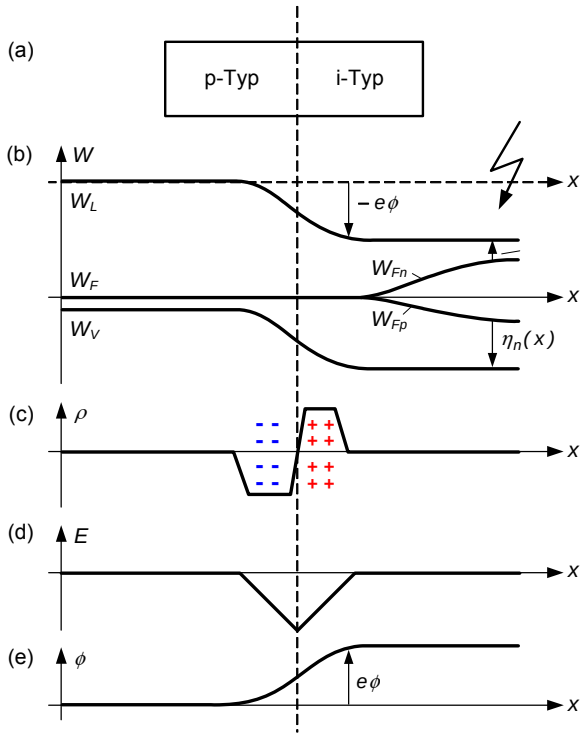


Abbildung 5.1: (a) Situation nach dem Zusammenfügen eines p-dotierten und eines intrisischen Halbleiters. (b) Energiebänder, (c) Ladungsträgerverteilung, (d) elektrische Feldverteilung und (e) Potentialverlauf.

5.2 Quasi-Fermi-Niveaus - ein Konzept zur Beschreibung des Nicht-Gleichgewichts

Zu den drei Halbleitergleichungen gehören drei Unbekannte, die es zu bestimmen gibt. Diese drei Unbekannten sind im Allgemeinen die Ladungsträgerkonzentrationen n und p sowie das Potential φ. Bei bekannter Geometrie und Dotierung sowie bei bekannten Materialparametern lösen diese drei Variabeln das Problem. In der Praxis ist das Auflösen nach n und p nicht immer ratsam. Wie wir gesehen haben, schwanken diese Werte je nach Dotierung und Ladungsträgerinjektion um Dekaden. Im Vergleich dazu ändert sich das Potential φ nur schwach und es wäre deshalb schön, wenn man zur numerischen Simulation ebenfalls mit Ladungsträgerpotentialen statt mit n und p rechnen könnnte. Dieses Potentialkonzept soll hier mit den Quasi-Fermi-Potentialen W_{Fn} und W_{Fp} eingeführt werden. Die äquivalenten Variabeln zur vollständigen Lösung des Problems sind dann W_{Fn}, W_{Fp} und φ.

Im Kapitel 3 haben wir gesehen, dass es zwischen der Ladungsträgerkonzentration und dem Abstand η von der Bandkante zum Ferminieveau eine direkte Beziehung gibt. Im Rahmen der Boltzmann-Näherung lautete diese Beziehung für die Leitungsbandelektronen von Eigenhalbleitern oder dotierten Halbleitern beispielsweise

$$n(x) = N_L \exp\left[-\frac{W_L - W_F}{kT}\right] . \tag{5.12}$$

Später haben wir im Kapitel über Drift und Diffusion, Kap. 4.2.3, festgestellt, dass beim Zusammenfügen von Halbleitern mit verschiedenen Dotierungseigenschaften Diffusions- und Driftströme fließen, bis sich wieder ein Geichgewicht über Ladungsträgeraustausch eingestellt hat. Es bilden sich Raumladungszonen aus, welche zur Bildung eines Potential (***built in potential***) $\varphi(x)$ beitragen. Im Beispiel von Fig. 5.1(a) wurde ein p-dotierter HL an einen intrinsisch dotierten HL gefügt.

Über der Kontaktstelle bildet sich wegen Diffusion eine Raumladungszone, Fig. 5.1(c) und damit ändert sich der Potentialverlauf, Fig. 5.1(e). Für die Elektronen im Leitungsband bedeutet dies, dass sich ihre potentielle Energie bei einer Verschiebung von der p-dotierten Seite zur intrisischen Seite verringert und sich ihre relative Lage zum Ferminiveau W_F deshalb verändert. Mit der Änderung des Abstandes Bandkante - Ferminiveau, ändert sich dann die Ladungsträgerkonzentration gemäß den Glgn. im Kap. 4.2.3. Das Ferminiveau W_F selber ändert sich nicht. Sie ist im thermischen Gleichgewicht eine ortsunabhängige Konstante. W_F ist die absolute Referenz bezüglich welcher sich das Potential oder die Bandkante der Elektronen und Löcher ändert und damit die Anzahl der Ladungsträger. Mit Hilfe des Potentials, können wir die Leitungs- und Valenzbandkanten als ortsabhängige Konstante betrachten und wie in Bild 5.1(b) gezeichnet, schreiben als

$$W_L(x) = W_L - e\varphi(x) \text{ und } W_V(x) = W_V - e\varphi(x) \ , \tag{5.13}$$

so dass damit gilt

$$n(x) = N_L \exp\left[-\frac{W_L(x) - W_F}{kT}\right] = N_L \exp\left[-\frac{W_L - e\varphi(x) - W_F}{kT}\right]. \tag{5.14}$$

(Diese Gleichung wurde eigentlich bereits im Kap. 4.2.3 hergeleitet).

Nun bestrahlen wir die Probe, wie das auf der rechten Seite in Fig. 5.1(b) angedeutet ist, so dass sich Überschussladungsträger bilden. Obwohl sich nun die Ladungsträgerkonzentration ändert, bleiben sowohl das Potential φ als auch die Leitungsbandkanten unverändert. Es bietet sich nun an, diese Ladungsträgerkonzentrationsänderung über eine Nachführung des Fermipotentials zu korrigieren, so dass die Boltzmannnäherung auch in dieser Situation ihre Richtigkeit behält. Da die Ladungsträgerkonzentration letztlich nur vom relativen Abstand η_n und η_p zwischen Ferminieveau und Bandkante abhängt, führt man für die Elektronen und Löcher jeweils eigene Ferminiveaus, sogenannte ***Quasiferminiveaus*** W_{Fn} und W_{Fp} ein. Damit gilt dann für die Ladungsträgerkonzentrationen

$$n(x) = N_L \exp\left[-\frac{W_L - e\varphi(x) - W_{Fn}}{kT}\right] \tag{5.15}$$

$$p(x) = N_V \exp\left[-\frac{W_{Fp} - (W_V - e\varphi(x))}{kT}\right]. \tag{5.16}$$

Bem.: Damit haben wir explizit angenommen, dass die Ladungsträgerverteilung sowohl im Nichtgleichgewicht als auch im Gleichgewicht identisch sind, d.h. der Fermi-Dirac Verteilungswahrscheinlichkeit folgen. Diese Annahme ist wegen der sehr schnellen Intrabandrelaxationszeiten der Ladungsträger in den meisten Fällen gerechtfertigt. Sie gilt allerdings während großen Ladungsträgerinjektionen nicht mehr - ist aber bereits ca. ~1 ps nach dem Ereignis wieder gültig.

Zur Umrechnung von Ladungsträgerkonzentrationen auf die Quasiferminiveaus können die Beziehungen

$$W_{F_n}(x) = W_L - e\varphi(x) + kT \ln\left(\frac{n(x)}{N_L}\right) \tag{5.17}$$

$$W_{F_p}(x) = W_V - e\varphi(x) - kT \ln\left(\frac{p(x)}{N_V}\right) \tag{5.18}$$

verwendet werden.

Durch Ableiten von (5.17) und (5.18) und multiplizieren mit $n\mu_n$ respektive $p\mu_p$ kann man zeigen, dass Quasi-Ferminiveaus und Stromfluss verknüpft sind.

$$\vec{J_n} = n\mu_n \operatorname{grad} W_{F_n} = -en\mu_n \operatorname{grad} \varphi + eD_n \operatorname{grad} n, \tag{5.19}$$

$$\vec{J_p} = p\mu_p \operatorname{grad} W_{F_p} = -ep\mu_p \operatorname{grad} \varphi - eD_p \operatorname{grad} p. \tag{5.20}$$

Elektronen- und Löcherströme sind also das Anzeichen für eine Nichtgleichgewichtssituation und erzeugen einen Gradienten in den Quasi-Fermi-Niveaus.

Das Massenwirkungsgesetz ergibt sich aus (5.15) und (5.16) zu

$$np = n_i^2 \exp\left[\frac{W_{F_n} - W_{F_p}}{kT}\right] . \tag{5.21}$$

Wird die Differenz in den Quasi-Fermi-Niveaus durch eine ***äquivalente Spannung (quasifermilevel separation voltage)*** $U = \left(W_{F_n} - W_{F_p}\right)/e$ ausgedrückt, so ergibt sich mit der Temperaturspannung $U_T = kT/e$ und wegen der Interpretation des Massenwirkungsgesetzes $np = n_i^2$ im Gleichgewicht (Rekombinationsrate = Generationsrate)

$$np = n_i^2 \exp\left(\frac{U}{U_T}\right) \qquad \begin{array}{lll} U > 0: & np > n_i^2 & \text{überwiegende Rekombination} \\ U < 0: & np < n_i^2 & \text{überwiegende Generation} \end{array} \tag{5.22}$$

Wird die das Nichtgleichgewicht hervorrufende Störung abgeschaltet, so stellt sich durch Trägertransport und wegen (5.22) allmählich wieder thermisches Gleichgewicht mit $W_{Fn} = W_{Fp} = W_F$ ein. Diese Zeit, in der die Elektronen im LB mit den Löchern im VB ins Gleichgewicht kommen, ist sehr lang (ns bis ms) im Vergleich zu den Intrabandrelaxationszeiten (einige 100 fs).

5.2.1 Ladungsträgerkonzentrationen und äquivalente Spannung

Wenn die Störung ladungsneutral erfolgt, d.h.

$$n = n_{th} + \Delta \tag{5.23}$$
$$p = p_{th} + \Delta, \tag{5.24}$$

so erhält man im Fall der Störstellenerschöpfung mit der Ladungsneutralität und dem angepassten Massenwirkungsgesetz

$$\left.\begin{array}{l} n + n_A = p + n_D \\ np = n_i^2 \exp\left(\dfrac{U}{U_T}\right) \end{array}\right\} \quad \begin{array}{l} n = \sqrt{\left(\dfrac{n_D - n_A}{2}\right)^2 + n_i^2\, \mathrm{e}^{U/U_T}} + \dfrac{n_D - n_A}{2}, \\ p = \sqrt{\left(\dfrac{n_D - n_A}{2}\right)^2 + n_i^2\, \mathrm{e}^{U/U_T}} - \dfrac{n_D - n_A}{2}, \end{array} \tag{5.25}$$

was eine zu (3.39) analoge Beziehung ist.

Schwache Injektion

Die schwache Injektion (Kap. 4.3, Kap. 4.79) ist definiert durch entweder $n \cong |n_D - n_A|$ mit $p \ll |n_D - n_A|$ oder $p \cong |n_D - n_A|$ mit $n \ll |n_D - n_A|$ was äquivalent ist mit der Aussage $np \ll (n_D - n_A)^2$ oder

$$np = n_i^2\, \mathrm{e}^{U/U_T} \ll (n_D - n_A)^2 \qquad \text{(schwache Injektion)} \tag{5.26}$$

Für einen n-Halbleiter, d.h. $n_D - n_A > 0$, erhält man durch Reihenentwicklung der Wurzeln ($\sqrt{1+x} \simeq 1 + (x/2)$) in (5.25)

$$\begin{array}{rcl} n & = & n_D - n_A + \dfrac{n_i^2}{n_D - n_A}\, \mathrm{e}^{U/U_T} \equiv n_n + p_n\, \mathrm{e}^{U/U_T}, \\ p & = & \dfrac{n_i^2\, \mathrm{e}^{U/U_T}}{n_D - n_A} = p_n\, \mathrm{e}^{U/U_T}, \end{array} \tag{5.27}$$

wobei wir nun die Definition verwendet haben, dass n_n und p_n je die Ladungsträgerkonzentrationen sind, welche im dotierten und ungestörten HL vorliegen.

Starke Injektion

Starke Injektion (oder Hochinjektion) liegt dann vor, wenn $n, p \gg n_D - n_A = n_n$, d.h. wenn

$$np = n_i^2 \, \mathrm{e}^{U/U_T} \gg (n_D - n_A)^2 \qquad \text{(starke Injektion)} \tag{5.28}$$

gilt. Man erhält aus (5.25) und (5.28) in diesem Fall

$$n = p = n_i \exp\left(\frac{U}{2U_T}\right) . \tag{5.29}$$

Durch Vergleich von (5.27) mit (5.29) sieht man, dass im Fall der schwachen Injektion die Minoritätsträgerdichte wie $\exp(U/U_T)$ steigt, im Fall der starken Injektion jedoch nur mehr wie $\exp(U/(2U_T))$.

Definiert man das Einsetzen der Hochinjektion in einem n-Halbleiter durch

$$p = n_D - n_A = n_n, \tag{5.30}$$

so erhält man aus (5.25) für den Abstand des Quasi-Fermi-Niveaus das Spannungsäquivalent U_{HI}:

$$U_{HI} = U_T \ln\left[\frac{2(n_D - n_A)^2}{n_i^2}\right] . \tag{5.31}$$

Für einen schwach dotierten n-Halbleiter,($n_D - n_A = 10^5 n_i$), erhält man $U_{HI}/U_T \approx 24$ und $U_{HI} \approx 600\,\mathrm{mV}$.

Eine weitere allgemeine Aussage zu Neutralinjektion

Aus den Beziehungen (5.15) und (5.16) für die Trägerdichten im Nichtgleichgewicht und denen für das Gleichgewicht (in (5.15) und (5.16) ist z.B. für einen n-Halbleiter $n = n_n$, $p = p_n$ zu setzen) erhält man für eine Störung mit $n = n_n + \Delta$ und $p = p_n + \Delta$, siehe (5.23)

$$\frac{n}{n_n} = 1 + \frac{\Delta}{n_n} = \exp\left(\frac{W_{Fn} - W_F}{kT}\right) \tag{5.32}$$

$$\frac{p}{p_n} = 1 + \frac{\Delta}{p_n} = \exp\left(\frac{W_F - W_{Fp}}{kT}\right) . \tag{5.33}$$

Man sieht daraus sofort, dass sich bei einer Neutralinjektion von Ladungsträgern das Quasi-Fermi-Niveau der Minoritätsträger stärker verschiebt als das Quasi-Fermi-Niveau der Majoritätsträger (Beispiel: Es sei $n_n = 10^5 n_i$, $p_n = 10^{-5} n_i$, $\Delta = 10^3 n_i$; dann gilt $W_{Fn} - W_F = 0{,}25\,\mathrm{meV}$, $W_F - W_{Fp} = 460\,\mathrm{meV}$).

Die hier eingeführten Potentiale, die ortsabhängigen Bandkanten und Quasi-Fermi-Niveaus, sind aus den Trägerdichten abgeleitete Größen, welche manchmal direkte physikalische Bedeutung erlangen, wie die Potentiale in der Elektrodynamik oder Thermodynamik. Diese Potentiale sind oft sehr nützlich, um einen schnellen Überblick und eine einfache Analyse bei komplexen Situationen zu bekommen. Wir könnten aber auch ganz auf diese Potentiale verzichten und nur die Bewegungsgleichungen für die Trägerdichten, deren Ströme und das daraus resultierende elektrische Feld lösen.

5.3 Bahngebiete

In diesem Kaptiel diskutieren wir den Verlauf der Bänder, der Fermi- und Quasi-Ferminiveaus in einem stromdurchflossenen homogenen Halbleiter.

In einem stromdurchflossenen Gebiet eines homogenen n-Halbleiters sei ein Feld $\vec{E} = -\operatorname{grad}\varphi$ infolge einer äußeren angelegten Spannung vorhanden. Im Halbleiter ist keine merkliche Raumladung vorhanden (sogenannte **Quasineutralität:** $\rho \cong 0$). Die Ladungen der ionisierten Störstellen

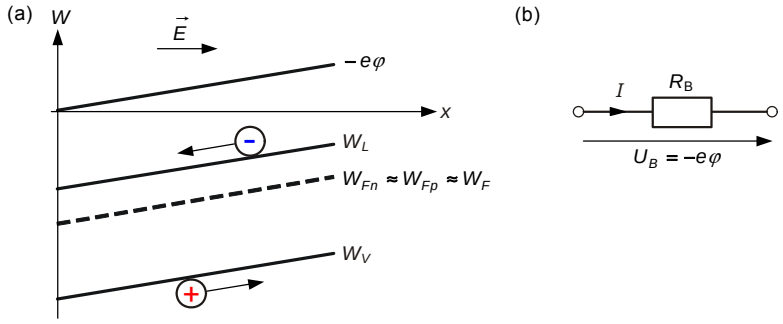


Abbildung 5.2: Bändermodell im Bahngebiet eines n-Halbleiters.

und der beweglichen Träger kompensieren einander annähernd derart, dass die vorhandene Gesamtladung an jeder Stelle sehr klein ist im Vergleich mit den Einzelladungen

$$\begin{aligned} \rho &= |n_D^+ - n_A^- + p - n| \ll |n_D^+ - n_A^-|, \\ \rho &= |n_D^+ - n_A^- + p - n| \ll |p - n|. \end{aligned} \tag{5.34}$$

Wir möchten den Stromfluss $\vec{J}$ durch den n-Halbleiter berechnen.

Aus (5.34) folgt, dass auch die Trägerdichten annähernd im Gleichgewicht sind ($np \approx n_n p_n = n_i^2$, $\operatorname{grad} n \approx \operatorname{grad} p \approx 0$), d.h. Diffusionsströme können vernachlässigt werden. Der Stromfluss kommt also allein aufgrund des Feldes zustande:

$$\begin{aligned} &\vec{J}_n = \vec{J}_{nF} = en\mu_n\vec{E}, \\ &\vec{J}_p = \vec{J}_{pF} = ep\mu_p\vec{E} \overset{\text{da } p\approx 0}{\cong} 0, \\ &\rightarrow \vec{J} = \vec{J}_n + \vec{J}_p = e(n\mu_n + p\mu_p)\vec{E} \approx en\mu_n\vec{E} = \sigma_n\vec{E}. \end{aligned} \tag{5.35}$$

Mit Gl. (5.19)-(5.20) lassen sich auch Aussagen zum Bandverlauf der Quasiferminiveaus machen: 1.) Wegen

$$\vec{J}_n = n\mu_n \operatorname{grad} W_{F_n} \overset{\operatorname{grad} n\approx 0}{=} -en\mu_n \operatorname{grad}\varphi, \tag{5.36}$$

$$\vec{J}_p = p\mu_p \operatorname{grad} W_{F_p} \overset{\operatorname{grad} p\approx 0}{=} -ep\mu_p \operatorname{grad}\varphi, \tag{5.37}$$

gilt

$$\operatorname{grad} W_{Fn} = \operatorname{grad} W_{Fp} = \operatorname{grad}(-e\varphi) \tag{5.38}$$

2.) Wegen Gl. (5.13) folgt auch

$$\operatorname{grad}(-e\varphi) = \operatorname{grad} W_L = \operatorname{grad} W_V\,. \tag{5.39}$$

Damit haben sowohl das Leitungsband, das Valenzband als auch das Potential mindestens die gleiche Steigung. 3.) Aus (5.21) folgt mit $np = n_n p_n = n_i^2 \exp((W_{Fn} - W_{Fp})/kT) \overset{\text{da keine Injektion}}{\cong} n_i^2$

$$W_{Fn} \approx W_{Fp} \approx W_F\,. \tag{5.40}$$

Mit diesen Informationen lässt sich ein Bild des Bandverlaufs zeichnen. Bild 5.2 zeigt das Bändermodell. Im Bahngebiet fließen nur Feldströme, wobei der Anteil der Minoritätsträger vernachlässigbar ist. Der Halbleiter verhält sich wie ein ohmscher Widerstand, der durch eine Leitfähigkeit $\sigma \approx \sigma_n$ beschrieben werden kann. Meist ist die Steigung der Potentiale in den Bahngebieten klein gegenüber den anderen Bandkrümmungen und wird dann in einem maßstäblich gezeichneten Bild nicht wahrgenommen.

Umgekehrt kann man sagen, dass das Ferminiveau keine Steigung hat, wenn kein Strom anliegt.

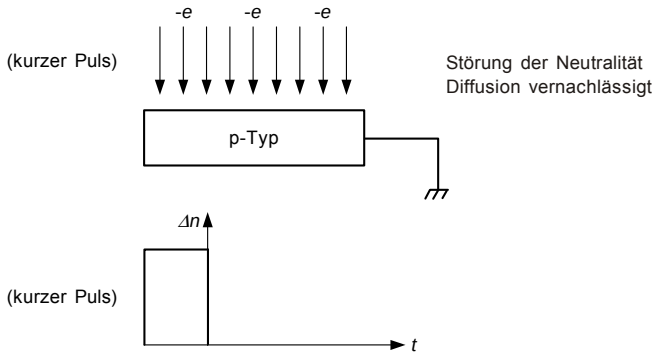


Abbildung 5.3: Minoritätsträgerinjektion [3].

5.4 Zeitlicher Abbau von positiven/negativen Ladungsträgerdichtestörungen - dielektrische Relaxation

5.4.1 Minoritätsträgerdichtestörung

Bild 5.3 zeigt die Injektion von Minoritätsträgern, hier Elektronen, in einen p-Halbleiter. Die Injektion erfolge nur für einen kurzen Moment, aber homogen über den ganzen Halbleiter. Der Halbleiter sei geerdet.

Wir versuchen znächst einmal zu verstehen was für Prozesse sich im Halbleiter abspielen: Die Überschussladungsträgerdichte Δn verletzt die Neutralitätsbedingung. Durch die Raumladungsdichte ρ entsteht ein elektrisches Feld, in dem sich beide Trägersorten so lange bewegen, bis die Raumladung (und daher das Feld) verschwunden sind. Ziel ist es, die zeitliche Entwicklung der Ladungsträgerdichte zu verstehen.

Bevor wir uns dem Lösen der Aufgabe zuwenden, wollen wir zuerst die Fakten zusammen tragen:

- – Diffusionsströme ($\vec{J}_D = 0$) und Generations- bzw. Rekombinationsprozesse wollen wir vorerst vernachlässigen.
  – Ferner sei

$$n_A = n_{A-} \quad \text{und } n_D \ll n_A \tag{5.41}$$

$$p_p \cong n_A \quad \text{und} \quad n_p \ll p_p \,. \tag{5.42}$$

  – Wegen $p_p \gg n_p$ (bei vergleichbaren Beweglichkeiten der Träger) muss als einziger Strom der Feldstrom der Majoritätsträger ($\vec{J}_{pF}$, s. (5.10)) berücksichtigt werden.

$$\vec{J}_p = \vec{J}_{pF} = \sigma_p E \quad \text{und} \quad \vec{J}_n = \vec{J}_{nF} \ll \vec{J}_p \tag{5.43}$$

  – Die Störung der Majoritätsträgerdichte sei auch nicht so groß, daß dadurch die Leitfähigkeit σ_p fürs Gleichgewicht merklich verändert wird

$$\sigma_p = e\, p_p \mu_p \,. \tag{5.44}$$

    σ_p ist deshalb eine Konstante.
  – Für die induzierte Raumladungsdichte gilt

$$\rho = -e\, \Delta n. \tag{5.45}$$

Zum Lösen der Aufgabe verwenden wir die Kontinuitätsgleichung (5.9)

$$\frac{\partial \rho}{\partial t} + \operatorname{div} \vec{J} = 0 \tag{5.46}$$

und die Poisson-Gleichung

$$\Delta\varphi = -\frac{\rho}{\varepsilon} = \frac{e\,\Delta n}{\varepsilon} \tag{5.47}$$

wobei wir $\vec{J}$ mit (5.43) ausdrücken können

$$\vec{J} \approx \vec{J}_{pF} = \sigma_p \vec{E}\,, \tag{5.48}$$

und $\vec{E}$ über $\vec{E} = -\mathrm{grad}\varphi$ in ein Potential überführen können.

Damit lassen sich die Gleichungen nach ρ auflösen. Man findet die Gleichung

$$0 = \frac{\partial \rho}{\partial t} + \operatorname{div} \vec{J} \tag{5.49}$$

$$= \frac{\partial \rho}{\partial t} + \sigma_p \operatorname{div} \vec{E} \tag{5.50}$$

$$= \frac{\partial \rho}{\partial t} - \sigma_p \operatorname{div} \mathrm{grad}\varphi \tag{5.51}$$

$$= \frac{\partial \rho}{\partial t} - \sigma_p \Delta\varphi \tag{5.52}$$

$$= \frac{\partial \rho}{\partial t} + \frac{\sigma_p}{\varepsilon}\rho, \tag{5.53}$$

welche die Lösung

$$\rho(t) = \rho(0)\exp(-t/\tau_R) \quad \text{mit} \quad \tau_R = \frac{\varepsilon}{\sigma_p} \tag{5.54}$$

besitzt. Die Raumladung klingt infolge einer Verschiebung der Majoritätsträger exponentiell mit der Zeit ab. Die charakteristische Zeitkonstante τ_R heißt ***dielektrische Relaxationszeit.*** ($\tau_R = 10^{-10} \ldots 10^{-15}$s für $\sigma_p = 10^{-2} \ldots 10^{3}\,\Omega^{-1}\mathrm{cm}^{-1}$ und $\varepsilon = \varepsilon_0\varepsilon_r \approx 10^{-10}\,\mathrm{As\,V^{-1}m^{-1}}$ für Si). Die Raumladungsdichtestörung, hervorgerufen von Minoritätsträgern, wird also innerhalb der dielektrischen Relaxationszeit, die je nach Dotierung zwischen 1fs und 100 ps liegt, durch eine entsprechende Majoritätsträgerdichte s. Bild 5.4 abgeschirmt. Diese Majoritätsträger werden über die Masseverbindung aus der Erde nachgeliefert, welche ihrerseits auch die Elektronen an die Elektronenkanone liefert und damit den Stromkreis in Bild 5.4 schließt. Innerhalb der dielektrischen Relaxationszeit wird die Neutralität wieder hergestellt. Die resultierende neutrale Überschussladungsträgerdichte verschwindet dann auf einer Zeitskala der Minoritätsträgerlebensdauer $\tau_n \gg \tau_R$. Der Halbleiter ist erst für Zeiten $t > \tau_n$ wieder im Gleichgewicht.

5.4.2 Majoritätsträgerdichtestörung:

Je nachdem, ob eine Majoritätsträger- oder eine Minoritätsträgerdichtestörung erzeugt wurde, laufen unterschiedliche Vorgänge ab.

Bei einer permanenten Störung (Fig. 5.5) der Majoritätsträgerdichte um Δp fließen über die Masse Minoritätsträger nach um die Überschussladungen abzuschirmen. Wegen der kleinen Konzentration an Minoritäten ist das Gleichgewicht erst nach der Zeit $t \gg \tau_R$ wieder in einem stationären Gleichgewicht mit den Trägerdichten $n_p + \Delta p$ und $p_p + \Delta p$, wobei $\Delta n = \Delta p$.

Im Falle einer transienten Störung der Majoritätsträgerdichte klingt die Raumladung mit der Zeitkonstante τ_R ab und in der gleichen Zeitspanne stellt sich auch das thermische Gleichgewicht ein. Die Überschussladungen der Majoritäten fließen dabei über die Masse ab.

In der Praxis bedeutet dies: Wenn transiente Ladungsträgerstörungen im Zeitbereich $t < \tau_R$ diskutiert werden, müssen Raumladungseffekte berücksichtigt werden. Für $t \gg \tau_R$ dagegen können die Beziehungen für den neutralen Halbleiter verwendet werden.

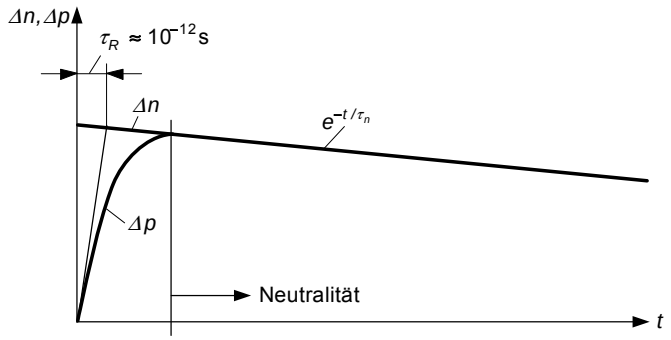


Abbildung 5.4: Abschirmung einer Minoritätsträgerdichtestörung innerhalb der dielektrischen Relaxationszeit und anschliessender Abbau infolge der endlichen Minoritätsträgerlebensdauer

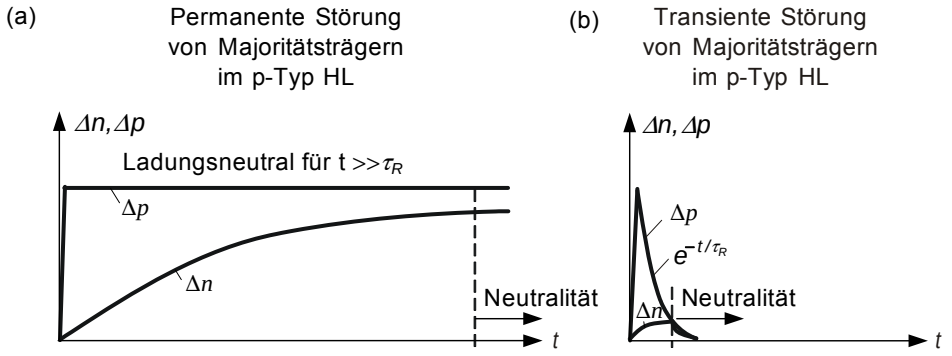


Abbildung 5.5: Zeitliche Entwicklung der Ladungsträgerkonzentrationen bei einer permanenten und bei einer transienten Störung der Ladungsträgerdichten in einem p-Halbleiter.

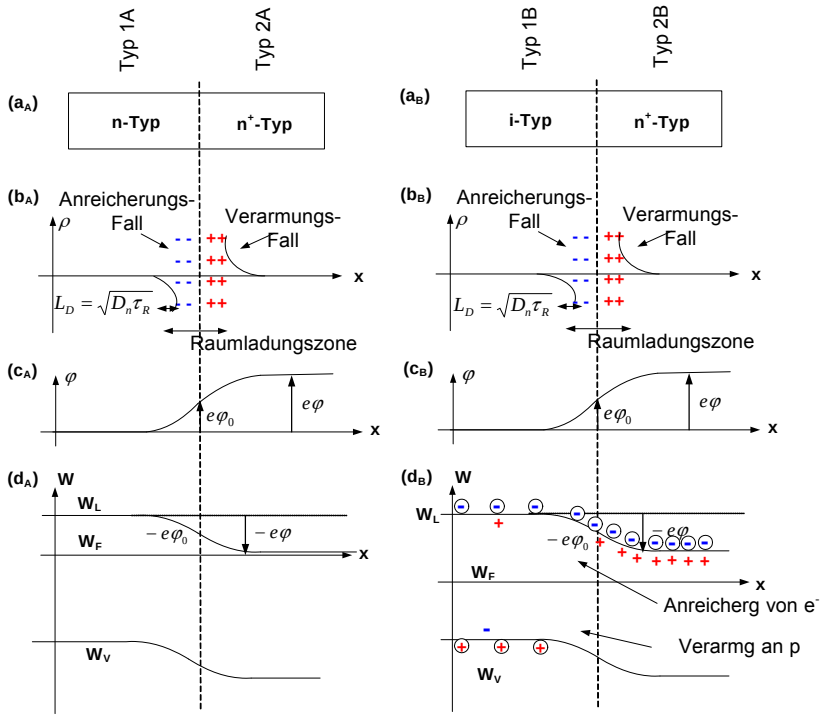


Abbildung 5.6: Potentialstörungsverlauf im Anreicherungsfall (Typ A1: n-HL gegen n+-HL), Verarmungsfall (Typ A2: n+-HL gegen n-HL) und im Typ B, eines intrinsischen HL gegen einen dotierten HL

5.5 Räumlicher Abbau von positiven/negativen Ladungsträgerdichtestörungen - Debye-Länge

In diesem Kapitel interessiert uns die typische Längenskala, über welche Potentialstörungen infolge von Konzentrationsgradienten oder anderer Anfangsbedingungen abgebaut werden, Bild 5.6. Wir betrachten den Fall A (dotierter HL) und den Fall B (undotierter HL).

Gegeben sei ein homogen dotierter n-Halbleiter 1A. Dieser erfüllt den Bereich $x \leq 0$, siehe Bild 5.6(a). Der n-HL sei im thermischen Gleichgewicht, ladungsneutral und die Störstellen seien erschöpft. Die Bandkanten verlaufen horizontal (sogenannter **Flachbandfall**). Die Ladungsträgerkonzentrationen Im Bereich 1A haben die Gleichgewichtskonzentrationen

$$n_{n1} = n_{D1} \tag{5.55}$$

$$p_{n1} = n_i^2/n_{D1} \ll n_{n1}\,. \tag{5.56}$$

Durch das Zusammenbringen mit einem n-Halbleiter 2A höherer Dotierung im Bereich $x > 0$ verändern sich die Randbedingungen, siehe 5.6(b) und (c). Im Halbleiter 1 baut sich eine Potentialstörung auf, welche die Ladungsträgerdichte gemäß (5.14) beeinflusst

$$n(x) = n_{D_1} \exp\left[\frac{e\varphi(x)}{kT}\right] \tag{5.57}$$

Wir diskutieren nun die beiden Halbleiterhälften je separt. Uns interessieren die Ladungsträgerkonezentration ρ an den Übergangszonen:

Fall 1A: Durch eine Störung des Oberflächenpotentials am rechten Rand von Halbleiter 1 auf einen Wert $\varphi_{_O} > 0$ werden zusätzlich Elektronen am Rand von Halbleiter 1 angereichert. Die Bandkanten ändern sich gemäß (5.13). Aus der Poisson-Gleichung (5.11) erhält man für die

Trägerkonzentrationen und die Raumladungsdichte in Halbleiter 1 ($n^+_{D1} = n_{D1}$)

$$\underbrace{\begin{array}{rcl} n^+_{D1} & = & n_{D1} \\ n^-_{A1} & \cong & 0 \\ n_1(x) & = & n_{n1}\exp\left(\dfrac{\varphi(x)}{U_T}\right) = n_{D1}\exp\left(\dfrac{\varphi(x)}{U_T}\right) \\ p_1(x) & = & p_{n1}\exp\left(-\dfrac{\varphi(x)}{U_T}\right) \ll n \end{array}}_{} \tag{5.58}$$

$$\begin{array}{rcl} \rho_1 & = & e[p_1 + n^+_{D1} - n_1 - n^-_{A1}] \\ & \cong & en_{D1}\left[1 - \exp(\varphi/U_T\right] . \end{array} \tag{5.59}$$

Allgemein kann jedes Gebiet mit $\rho \neq 0$ als ***Raumladungszone*** **RLZ** ***("space charge region")*** bezeichnet werden. Da in diesem speziellen Fall die Majoritätendichte am Rand größer ist als im Inneren des HL, spricht man von einer ***Anreicherungsrandschicht.***

Fall 2A: Für eine Randstörung $\varphi_{_O} < 0$ (z.B. durch Ankopplung einer niedriger dotierten n-Halbleiterprobe, p-Halbleiterprobe oder durch ein extern angelegtes Feld) werden Elektronen vom Rand weggedrückt, die Bandkanten biegen sich nach oben. Der Rand verarmt an Majoritäten (wir sprechen von einer sogenannten ***Verarmungsrandschicht*****)**.

Fall 1B: Ersetzt man den n-Halbleiter 1 in Abb. 5.6 durch einen Eigenhalbleiter, so hätte man für $\varphi_O > 0$ am Rand n-Charakter, für $\varphi_O < 0$, z. B. indem der Halbleiter 2 ein p-Halbleiter wäre, am Rand p-Charakter. Im Bereich des i-Halbleiters 1 erhalten wir ganz allgemein

$$\underbrace{\begin{array}{rcl} n^+_{D1} & \cong & 0 \\ n^-_{A1} & \cong & 0 \\ n_1(x) & = & n_i\exp\left(\dfrac{\varphi(x)}{U_T}\right) \\ p_1(x) & = & n_i\exp\left(-\dfrac{\varphi(x)}{U_T}\right) \end{array}}_{} \tag{5.60}$$

$$\begin{array}{rcl} \rho_1(x) & = & e[p_1 + n^+_{D1} - n_1 - n^-_{A1}] \\ & = & en_i[\exp(-\varphi(x)/U_T) - \exp(\varphi(x)/U_T)] \end{array} \tag{5.61}$$

Um den Bereich in welchem die Raumladungszone abfällt zu berechnen müssen wir die Poisson-Gleichung für das Potential φ, in die ρ nach (5.59) eingesetzt wird lösen. Wir können den Fall A und Fall B unter der Annahme, dass nur kleine Potentialstörungen $\varphi/U_T \ll 1$ auftreten lösen. Aus (5.11) und (5.59) erhält man :

$$\frac{\mathrm{d}^2\varphi}{\mathrm{d}x^2} = -\frac{\rho_1}{\varepsilon} = \begin{cases} \dfrac{en_{D1}}{\varepsilon U_T}\varphi & \text{für den n-Halbleiter, Fall 1A} \\ \dfrac{2en_i}{\varepsilon U_T}\varphi & \text{für den Eigenhalbleiter, Fall 1B .} \end{cases} \tag{5.62}$$

Die Lösung zu den Randbedingungen $\varphi(0) = \varphi_{_O}$, $\varphi(-\infty) = 0$ lautet (beachte $D_n = \mu_n U_T$, $\sigma_n = en_D\mu_n$ und $\tau_R = \varepsilon/\sigma_n$)

$$\varphi(x) = \varphi_O\exp\left(\frac{x}{L_D}\right), \qquad L_D = \begin{cases} L_{Dn} = \sqrt{\dfrac{\varepsilon U_T}{en_D}} = \sqrt{D_n\tau_R} & \text{n-Halbleiter,} \\ L_{Di} = \sqrt{\dfrac{\varepsilon U_T}{2en_i}} & \text{Eigenhalbleiter.} \end{cases} \tag{5.63}$$

L_D ist die sogenannte ***Debye-Länge (Debye length, transition length).*** Wie Gleichung 5.63 zeigt, beinhaltet die Debye-Länge im eigenleitenden Halbleiter einen Faktor zwei, da beide Ladungsträgersorten zur Abschirmung beitragen.

Das Ergebnis besagt Folgendes: Potentialstörungen werden durch Rearrangieren der Majoritätsträger mit einer charakteristischen Länge L_D abgeschirmt. Auf einer Debye-Länge ändert sich das Potential in der Größenordnung der Temperaturspannung U_T.

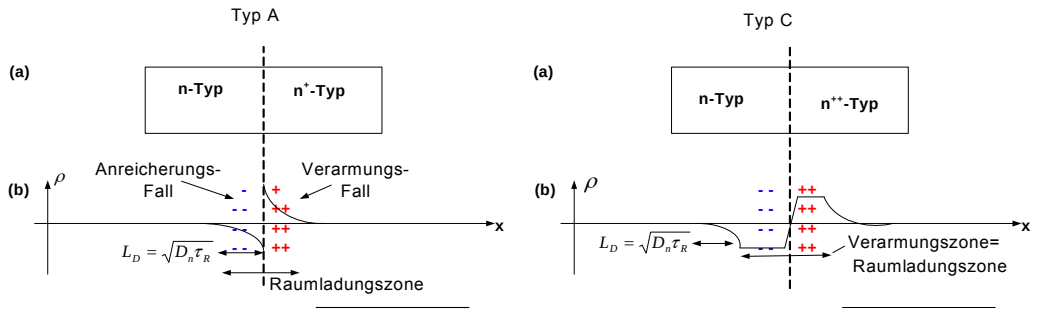


Abbildung 5.7: Schematische Darstellung der Ausdehnung der Debey-Längen und der RLZ bei (a) schwach dotierten Übergängen und (b) stark dotierten Übergängen mit Verarmungsnäherung

Wichtige Spezialfälle:

- Eine wichtige Folge im Verarmungsfall $\varphi < 0$ sieht man aus (5.59): Ist das Oberflächenpotential $|\varphi_O| \gg U_T$, so gilt im überwiegenden Teil der entstehenden Raumladungszone (abgesehen von einer Schichtdicke der Debye-Länge) $\rho = en_{D1}$. Man kann die Breite der Raumladungszone in diesem Fall so berechnen, als gäbe es in ihr überhaupt keine beweglichen Ladungen (sogenannte ***Verarmungsnäherung (depletion approximation)***). Die Zone selber heisst dann auch ***Raumladungszone=Verarmungszone*** (***"depletion zone").***

- Bei Bauelementen mit einer charakteristischen Länge $L < L_D$ muß der Stromfluß unter Berücksichtigung der Raumladung berechnet werden. Beispiele für Debye-Längen: In Si mit $n_D = 10^{16}\,\mathrm{cm}^{-3}$ ist $L_{Dn} = 40\,\mathrm{nm}$, in eigenleitendem Si mit $n_i \approx 10^{10}\,\mathrm{cm}^{-3}$ ist $L_{Di} = 40\,\mu\mathrm{m}$, in hochdotiertem entarteten Si mit $n_D = 10^{19}\,\mathrm{cm}^{-3}$ ist $L_{Dn} = 1$ nm, was nur noch zwei Gitterkonstanten entspricht. Dies bedeutet, dass bei Metallen mit ihren extrem hohen Ladungsträgerdichten Potentialstörungen praktisch innerhalb einer Gitterkonstante abgeschirmt werden.

5.6 Räumlicher Abbau von neutralen Ladungsträgerdichtestörungen - Diffusionszone

Wir betrachten nun eine lokale Minoritätsträgerdichtestörung (z.B. Einschuss von Elektronen in einen p-HL), siehe Bild 5.8(a). Im Gegensatz zu Kap. 5.4.1 berücksichtigen wir nun Diffusionseffekte. Die sehr schnellen dielektrischen Relaxationsprozesse spielen sich zwar auch ab, aber wir beginnen unsere Betrachtung unmittelbar nachdem Ladungsneutralität erreicht wurde.

Wir betrachten hier den Fall eines p-Halbleiter. Im Bereich $x \geq 0$ werden bei $x = 0$ Elektronen (Minoritätsträger) injiziert.

Es spielen sich dann zwei Prozesse ab:

1. Die injizierten Raumladungen werden innerhalb einer dielektrischen Relaxationszeit $\tau_R = \varepsilon/\sigma_p$ durch herangezogene Majoritätsträger neutralisiert, siehe Kap. 5.4.1. Da wir die Injektion nun stationär aufrechterhalten, wird die Raumladungstörung innerhalb einer Debye-Länge L_D abgebaut (d.h. dann, dass die Randkonzentration der Minoritätsträger bei $x = 0$ konstant bleibt und nach L_D auf den e-ten Teil abgefallen ist). Diese geladene Zone ist kurzreichweitig und dafür verantwortlich, dass ein elektrisches Feld $\vec{E} \neq 0$ erzeugt wird, welches Ladungsträgerdrift bewirkt.

2. Wegen der Neutralisierung der injizierten Ladungsträger durch Majoritätsträger liegt im restlichen Gebiet ($x > L_D$) Quasineutralität vor und da wir kein äußeres Feld anlegen gilt

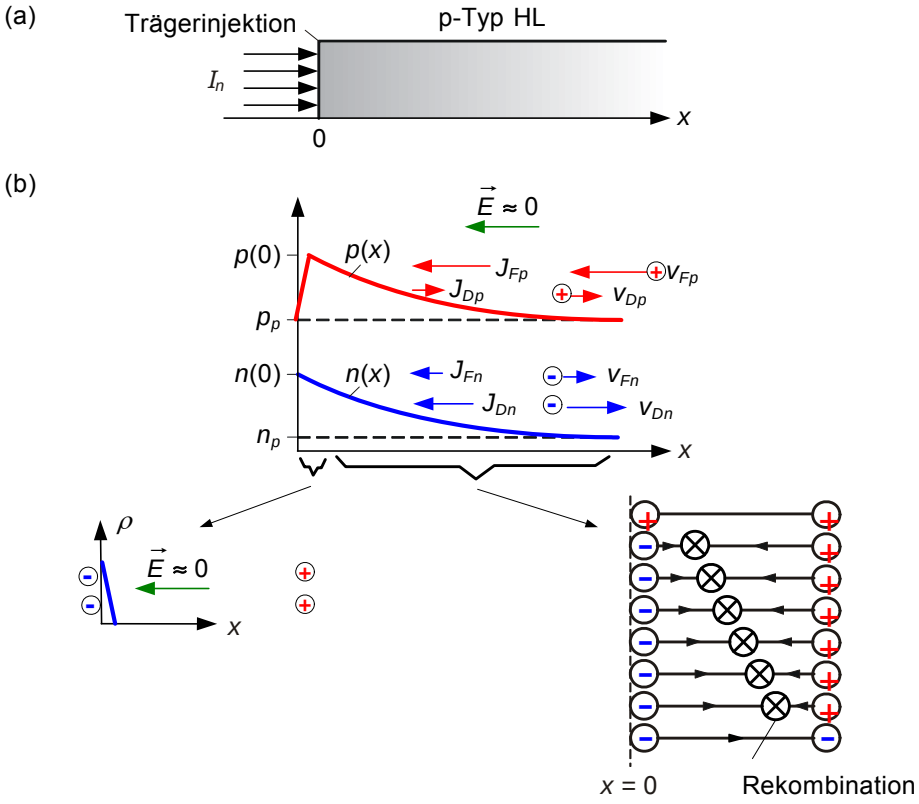


Abbildung 5.8: (a) Injektion von Minoritätsladungsträgern in eine unendlich ausgedehnte Probe. (b) Ladungsträgerkonzentrationen entlang der Ausbreitungsrichtung mit einer geladenen Zone in einem ersten Bereich und sukzessiver Rekombination der Trägerstörung Δn mit Löchern in der anschließenden Zone.

dort

$$\Delta n \approx \Delta p \qquad \operatorname{grad}\Delta n \approx \operatorname{grad}\Delta p, \qquad \vec{E} = -\operatorname{grad}\varphi \approx 0 \text{ d.h. klein.} \tag{5.64}$$

Das Prinzip ist in Bild 5.8(b) dargestellt (das Bild ist nicht maßstäblich, es sollte $p_p \gg n_p$ sein). In dieser Zone, diffundieren die Ladungsträger in den Bereich niedriger Konzentration. Während der Diffusion kommt es aber auch zu Ladungsträger-Rekombination (via Störstellenrekombination, spontane Emission,...) von Minoritätsträgern mit Majoritätsträgern. Da für jedes rekominbierte Elektron ein Loch vernichtet wird, müssen die Löcher wieder nachgeschoben werden. Der Diffusionsstrom der Elektronen wird damit allmählich in einen Driftstrom der Löcher konvertiert, siehe Fig. 5.8(c).

Die (Quasi-)Neutralstörung erhöht also beide Trägerkonzentrationen. Die Elektronen werden von außen injiziert und diffundieren in die Probe. Die Löcher, als Majoritätsträger, werden durch eine kleine Raumladung, welche ein kleines elektrisches Feld erzeugt, angezogen und driften in demselben.

Als nächstes betrachten wir die Ladungsträgerflüsse:

1. Bei den Elektronen überwiegt der Diffusionsstrom den Feldstrom

$$|\vec{J}_{nD}| = |eD_n \operatorname{grad} n| \gg |\vec{J}_{nF}| = \left|en\mu_n\vec{E}\right| \tag{5.65}$$

weil

$$|\vec{E}| \ll U_T \frac{|\operatorname{grad} n|}{n}. \tag{5.66}$$

$\vec{J}_{nF}$, $\vec{J}_{nD}$ sind parallel, $|\vec{J}_n| = |\vec{J}_{nD}| + |\vec{J}_{nF}|$ (die injizierten Träger werden durch das Feld $\vec{E}$ geringfügig unterstützt).

2. Für die Löcher dagegen überwiegt der Feldanteil den Diffusionsanteil

$$|\vec{J}_{pF}| \gg |\vec{J}_{pD}| \tag{5.67}$$

wegen

$$|\vec{E}| \gg U_T \frac{|\operatorname{grad} p|}{p}. \tag{5.68}$$

$\vec{J}_{pF}$, $\vec{J}_{pD}$ sind antiparallel, $|\vec{J}_p| = |\vec{J}_{pF}| - |\vec{J}_{pD}|$ (der Feldstrom der Majoritätsträger wird durch einen entgegengesetzten Diffusionsstrom geringfügig geschwächt).

Halbleitergebiete, in denen diese Verhältnisse herrschen, werden als ***Diffusionszonen*** bezeichnet.

5.6.1 Diffusionsströme und Diffusionslänge

Zur quantitativen Analyse müssen die Halbleiter-Gleichungen herangezogen werden. Es soll der Fall schwacher Injektion vorliegen. In einem p-Halbleiter mit schwacher Injektion gelten die Gleichungen (5.10) und die Gleichungen für die Minoritätsträger entkoppeln

$$\frac{\partial n}{\partial t} - \frac{1}{e}\operatorname{div}\vec{J}_n = -\frac{\Delta n}{\tau_n}, \quad \text{mit} \quad \vec{J}_n \approx \vec{J}_{n,D} = eD_n \operatorname{grad} n. \tag{5.69}$$

Da die Gleichgewichtsträgerdichte der Elektronen n_p orts- und zeitunabhängig ist, kann in allen Ableitungen die aktuelle Trägerdichte durch die Überschussträgerdichte Δn ersetzt werden, und es folgt mit Einsetzen des Diffusionsstromes in die Kontinuitätsgleichung

$$\frac{\partial \Delta n}{\partial t} = \triangle(\Delta n)\frac{L_n^2}{\tau_n} - \frac{\Delta n}{\tau_n}, \qquad L_n \equiv \sqrt{D_n\tau_n}. \tag{5.70}$$

L_n wird als Diffusionslänge der Elektronen im p-Halbleiter bezeichnet.

Wird ein konstanter Strom von Minoritätsträgern injiziert, so folgt daraus eine zeitunabhängige Überschußdichte an Minoritätsträgern $\Delta n_0(x)$. Die zeitunabhängigen Lösungen der Differentialgleichung zweiter Ordnung (5.70)

$$0 = \triangle(\Delta n) - \frac{\Delta n}{L_n^2} \tag{5.71}$$

sind Exponentialfunktionen $\exp\left[\pm\frac{x}{L_n}\right]$. Die Randbedingungen lauten

$$\begin{array}{rcl} \Delta n_0(0,t) & = & \Delta n_0(0), \\ \Delta n_0(\infty,t) & = & 0\ . \end{array} \tag{5.72}$$

Die zeitunabhängige Lösung $\Delta n_0(x)$ im Bereich $x \geq 0$ ergibt sich dann zu

$$\Delta n_0(x) \quad = \quad \Delta n_0(0) \exp\left(-\frac{x}{L_n}\right). \tag{5.73}$$

Der Strom, welcher am linken Rand injiziert wird kann jetzt an jeder Stelle entlang x als Summe aus dem Elektronenstrom und dem Löcherstrom errechnen. Am linken Rand ($x = 0$) ist es jedoch besonders einfach, da dort nur der Diffusionsstrom der Elektronen einen Beitrag liefert. Dieser Minoritätsträgerstrom am linken Rand ist:

$$J_{n,D}(x=0) \quad = \quad eD_n \operatorname{grad} n|_{x=0} = eD_n \left.\frac{\partial \Delta n_0(x)}{\partial x}\right|_{x=0} = -\frac{eD_n}{L_n}\Delta n_0(0)\ . \tag{5.74}$$

Denkt man sich den Diffusionsstrom als mittleren Teilchenstrom mit einer effektiven Geschwindigkeit, so ergibt sich

$$v_{n,\text{eff}} = J_{n,D}(0)/\left(e\,\Delta n_0(0)\right) = \frac{D_n}{L_n} = \frac{L_n}{\tau_n}, \tag{5.75}$$

was den Namen ***Diffusionslänge (diffusion length)*** L_n motiviert. Offensichtlich legen die diffundierenden Minoritätsträger innerhalb ihrer Lebensdauer gerade eine Diffusionslänge zurück. Diffusionslängen liegen je nach Halbleitertyp (direkter oder indirekter Halbleiter), Zahl der Rekombinationsstörstellen und Trägertyp im Bereich von einigen Mikrometern bis zu einigen hundert Mikrometern. Aus den Gln. (5.44) (5.70) und (5.54) folgt für das Verhältnis zwischen Diffusionslänge und Debye-Länge

$$\frac{L_n}{L_{Dp}} = \sqrt{\frac{\mu_n \tau_n}{\mu_p \tau_R}}. \tag{5.76}$$

Da Trägerlebensdauern sehr viel größer sind als Relaxationszeiten, sind Diffusionslängen in der Regel sehr viel größer als Debye-Längen. Zur Erinnerung: Die Debye-Länge skaliert die Größe von Raumladungszonen.

5.6.2 Kurze Diffusionszone

Hat die Diffusionszone eine endliche Länge, siehe Bild 5.9(b), so muß die Überschußträgerdichte die Randbedingungen

$$\begin{array}{rcl} \Delta n_0(0,t) & = & \Delta n_0(0) \\ \Delta n(w,t) & = & 0 \end{array} \tag{5.77}$$

erfüllen. Dazu sind Exponentiallösungen geeignet zu überlagern und es ergibt sich (wegen $\sinh(x) = 0,5(\exp(x) - \exp(-x))$)

$$\Delta n_0(x) = \Delta n_0(0) \frac{\sinh\left(\frac{w-x}{L_n}\right)}{\sinh\left(\frac{w}{L_n}\right)}. \tag{5.78}$$

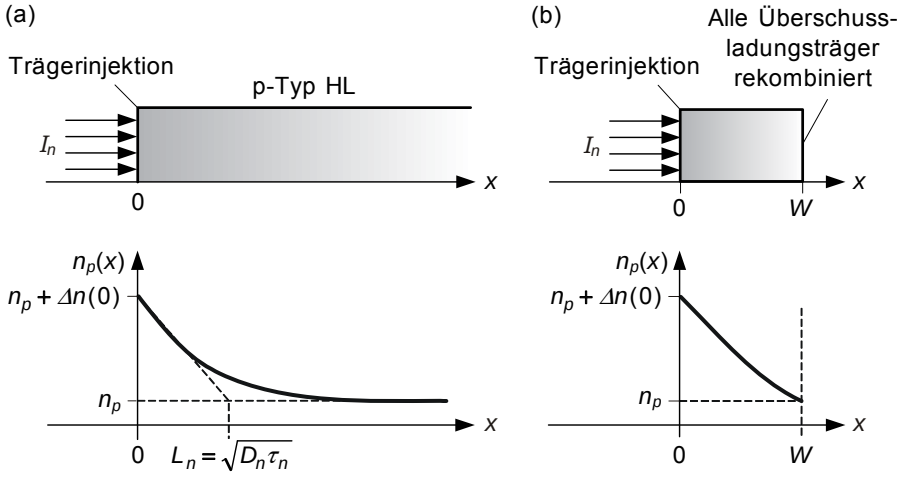


Abbildung 5.9: Injektion von Minoritätsladungsträgern: (a) In eine unendlich ausgedehnte Probe; (b) In eine endlich ausgedehnte Probe bei der am anderen Ende alle Überschussladungsträger abgesaugt werden oder rekombinieren. [5]

bzw. für den Injektionsstrom

$$J_{nD,0}(0) = eD_n \left.\frac{\partial \Delta n(x)}{\partial x}\right|_{x=0} = -\frac{eD_n}{L_n}\Delta n_0(0)\frac{\cosh\left(\frac{w}{L_n}\right)}{\sinh\left(\frac{w}{L_n}\right)} \tag{5.79}$$

$$= -\frac{eD_n}{L_n}\Delta n_0(0)\coth\left(\frac{w}{L_n}\right). \tag{5.80}$$

Der Diffusionsstrom verstärkt sich gegenüber dem Strom im unendlichen Halbraum um den Faktor $\coth\left(\frac{w}{L_n}\right) \overset{\text{Taylorentwicklung}}{\sim} \frac{L_n}{w}$.

5.6.3 Bandverlauf in der Diffusionszone

Aus (5.19) und (5.20) folgt in der Diffusionszone für den Löcherstrom

$$\vec{J_p} = p\mu_p \operatorname{grad} W_{Fp} \approx \vec{J}_{pF} = ep\mu_p\vec{E} = \text{konst.} \tag{5.81}$$

Da aber $E = -\operatorname{grad}(\varphi)$

$$=> \ \operatorname{grad} W_{Fp} \approx \operatorname{grad}(-e\varphi) = \operatorname{grad} W_V = \operatorname{grad} W_L \tag{5.82}$$

und für den Elektronenstrom

$$\vec{J_n} = n\mu_n \operatorname{grad} W_{Fn} \approx \vec{J}_{nD} = e\mu_n U_T \operatorname{grad} n \simeq -e\mu_n U_T \frac{\Delta n_0(x)}{L_n}, \tag{5.83}$$

$$=> \operatorname{grad} W_{Fn} \approx -\frac{eU_T}{L_n}\frac{\Delta n_0(x)}{\Delta n_0(x) + n_p}. \tag{5.84}$$

Damit lassen sich die Banddiagramme zeichnen. Bild 5.10 zeigt den qualitativen Verlauf der Bänder. Am linken Rand der Diffusionszone folgt aus (5.27) im Falle des hier betrachteten p-Halbleiters (wo $p \simeq p_p$) mit $U = \left(W_{F_n} - W_{F_p}\right)/e$

$$n = n_p e^{U/U_T}. \tag{5.85}$$

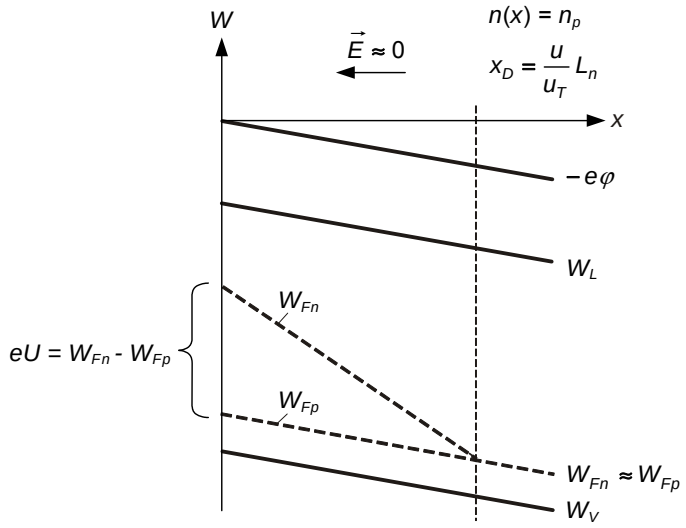


Abbildung 5.10: Bandverlauf in der Diffusionszone eines p-Halbleiters bei schwacher Injektion. Rechts von dem Ort, an dem W_{Fn} gleich W_{Fp} wird, schließt sich ein Bahngebiet an.

Die äquivalente Spannung U, welche zunächst noch eine abstrakte Differenz von Quasi-Fermi-Niveaus ist, wird bereits im nächsten Kapitel zur angelegten Spannung eines pn-Überganges. Die äquivalente Spannung U hat aber mit der über der Diffusionszone aufbauenden elektrischen Spannung aufgrund des Bahnwiderstand nichts zu tun. Vgl. dazu Bild 5.10, und bestimmt die Neigung der Bandkanten.

5.6.4 Wechselstromverhalten:

Wir wollen nun die Situation betrachten, in welcher von außen dem statischen Ladungsträger-Injektionsanteil ein periodischer Injektions-Anteil überlagert ist. Es interessiert uns das Strom-Spannungsverhalten.

Die Differentialgleichung, welche es zu erfüllen gilt ist

$$\frac{\partial \Delta n}{\partial t} = \triangle(\Delta n)\frac{L_n^2}{\tau_n} - \frac{\Delta n}{\tau_n}. \tag{5.86}$$

Die Randbedingungen an der Stelle $x = 0$, d.h. die Randbedingungen für $\Delta n_0(0)$ und $\Delta n_1(0,t)$, welche die Differentialgleichung erfüllen muss ergeben sich aus dem angelegten Strom am linken Rand. Dieser hat einen statischen und einen periodischen Anteil. Der gesamte Strom ergibt sich aus

$$J_n = J_{nD,0}(0,t) + J_{nD,1}(0,t)$$

mit

$$J_{nD,0}(0,t) = eD_n \left.\frac{\partial \Delta n_0(x)}{\partial x}\right|_{x=0} = -\frac{eD_n}{L_n}\Delta n_0(x=0) \quad \text{mit} \quad \Delta n_0(0) = n_0(0) - n_p \tag{5.87}$$

$$J_{nD,1}(0,t) = eD_n \left.\frac{\partial \Delta n_1(x,t)}{\partial x}\right|_{x=0} \quad \text{mit } \Delta n_1(x,t), \text{ welches es noch zu finden gilt.} \tag{5.88}$$

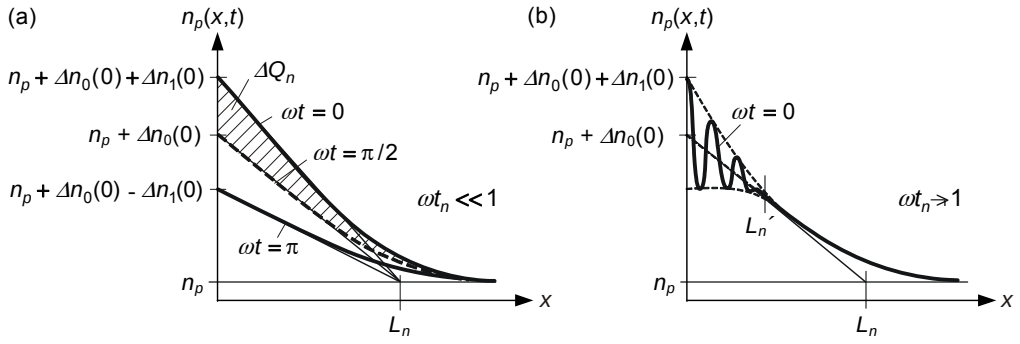


Abbildung 5.11: Verlauf der gesamten Minoritätsträgerdichte $n(x,t) = n_p + \Delta n_0(x) + \Delta n_1(x,t)$ bei zeitlich konstanter und periodisch modulierter Injektion (Momentaufnahmen) für (a) $\omega\tau_n \ll 1$, (b) $\omega\tau_n > 1$.

Es geht nun darum einen Ansatz für $\Delta n_1(x,t)$ zu finden. Da mit dem Strom auch die gesamte Trägerdichte periodisch oszillieren wird und Gl. (5.86) linear ist, macht folgender Ansatz Sinn

$$\Delta n_1(x,t) \quad = \quad \Delta n_1(0)\Re\left\{e^{j\omega t}\; h(x)\right\} \quad \text{mit den RB} \quad h(0)=1 \quad \text{und} \quad h(\infty)=0$$

Dies muss in die Gl.(5.70) eingesetzt werden und führt auf

$$\left(\frac{d^2}{dx^2} - \frac{1+\mathrm{j}\,\omega\tau_n}{L_n^2}\right) h(x) = 0. \tag{5.89}$$

Die Lösung dazu ist

$$\begin{aligned} h(x) &= \exp\left(-\frac{\sqrt{1+\mathrm{j}\,\omega\tau_n}}{L_n}x\right) \\ \Delta n_1(x,t) &= \Delta n_1(0)\Re\left\{e^{j\omega t}\exp\left(-\frac{\sqrt{1+\mathrm{j}\,\omega\tau_n}}{L_n}x\right)\right\}. \end{aligned} \tag{5.90}$$

Bild. 5.11 zeigt den Verlauf der Minoritätsträgerdichte.

Für niedrige Frequenzen $\omega\tau_n \to 0$ nimmt sie mit einer charakteristischen Länge L_n (der Diffusionslänge) exponentiell ab.

$$\omega\tau_n \to 0: \qquad \Delta n_1(x,t) \quad \cong \quad \Delta n_1(0)\Re\left\{e^{j\omega t}\exp\left(-\frac{x}{L_n}\right)\right\}.$$

Für hohe Frequenzen $\omega\tau_n \gg 1$ nimmt die Wechselamplitude mit einer kürzeren charakteristischen Länge L_n' ab (beachte: $\sqrt{\mathrm{j}} = (1+\mathrm{j})/\sqrt{2}$)

$$\begin{aligned} \omega\tau_n \gg 1: \qquad \Delta n_1(x,t) &\cong \Delta n_1(0)\Re\left\{e^{j\omega t}\exp\left(-\frac{\sqrt{\mathrm{j}\,\omega\tau_n}}{L_n}x\right)\right\} \\ &= \Delta n_1(0)\Re\left\{e^{j\omega t}\exp\left(-\frac{\sqrt{\omega\tau_n}}{\sqrt{2}L_n}x\right)\exp\left(-j\frac{\sqrt{\omega\tau_n}}{\sqrt{2}L_n}x\right)\right\} \\ \Delta n_1(x,t) &= \Delta n_1(0)\exp\left(-\frac{\sqrt{\omega\tau_n}}{\sqrt{2}L_n}x\right)\Re\left\{e^{j\omega t}\exp\left(-j\frac{\sqrt{\omega\tau_n}}{\sqrt{2}L_n}x\right)\right\} \end{aligned} \tag{5.91}$$

Damit fällt der oszillierende Teil mit der charakteristischen Länge L_n' ab

$$L_n' = L_n\sqrt{\frac{2}{\omega\tau_n}} \ll L_n \qquad (\omega\tau_n \gg 1) \tag{5.92}$$

L'_n ist bei großen Frequenzen sehr klein. Durch Schichten der Dicke $d \gg L'_n$ kann ein Wechselanteil nicht übertragen werden. Außerdem tritt eine Phasenverschiebung zwischen $\Delta n_1(x)$ und $\Delta n_1(0)$ auf.

Im Folgenden sind wir an der Strom-Spannungscharakteristik am Diffusionsübergang interessiert. Mit der genauen Kenntnis über die Ladungsträgerkonzentrationen an der Stelle $x = 0$ lässt sich sowohl der Strom als auch die Spannung errechnen.

Wir suchen zunächst den Zusammenhang zwischen Ladungsträgerdichte und angelegter Spannung. Gl. (5.85) erlaubt es die Ladunsgträgerdichte an der Injektionszone mit der Spannung an der Injektionszone auszudrücken. Mit dem Ansatz für die Spannung $U = U_0 + \Re\left\{U_1 \exp(\mathrm{j}\,\omega t)\right\}$, $|U_1|/U_0 \ll 1$ für die schwache Injektion von Elektronen in einen p-Halbleiter ergibt sich $n = n_p \cdot e^{U/U_T}$

$$\begin{aligned} n(0,t) &= n_p e^{U/U_T} \\ &\overset{\text{Ansatz für } U}{=} n_p \exp\left[\frac{U_0 + \Re\left\{U_1 \exp(\mathrm{j}\,\omega t)\right\}}{U_T}\right] \\ &\overset{\text{Taylor um } U_0}{=} n_p \exp\left[\frac{U_0}{U_T}\right] + n_p \exp\left[\frac{U_0}{U_T}\right] \cdot \frac{\Re\left\{U_1 \exp(\mathrm{j}\,\omega t)\right\}}{U_T} + \ldots \\ &\approx n_p \exp\left(\frac{U_0}{U_T}\right) + n_p \exp\left(\frac{U_0}{U_T}\right) \Re\left\{\frac{U_1}{U_T} \exp(\mathrm{j}\,\omega t)\right\} \end{aligned} \tag{5.93}$$

Nun ist aber $n(0,t)$ auch

$$n(0,t) = n_p + \Delta n_0(0) + \Delta n_1(0,t) = n_p + \Delta n_0(0) + \Delta n_1(0) \Re\left\{e^{j\omega t}\right\} \tag{5.94}$$

und damit findet durch Vergleich von (5.93) mit (5.94)

$$\begin{aligned} n_p + \Delta n_0(0) &= n_p \exp\left(\frac{U_0}{U_T}\right) \\ \Delta n_1(0,t) &= n_p \exp\left(\frac{U_0}{U_T}\right) \Re\left\{\frac{U_1}{U_T} \exp(\mathrm{j}\,\omega t)\right\} \end{aligned} \tag{5.95}$$

und nach Einsetzen in (5.90) auf den allgemeinen Ausdruck für $\Delta n_1(x,t)$

$$\Delta n_1(x,t) = \Delta n_1(0) \Re\left\{e^{j\omega t}\ h(x)\right\} = n_p \exp\left(\frac{U_0}{U_T}\right) \Re\left\{\frac{U_1}{U_T} \exp(\mathrm{j}\,\omega t) \exp\left(-\frac{\sqrt{1 + \mathrm{j}\,\omega\tau_n}}{L_n} x\right)\right\} \tag{5.96}$$

Der bei $x = 0$ in dden p-Halbleiter eintretende Diffusionsstrom folgt wie vorher aus

$$\begin{aligned} J_{nD,0}(0,t) &= -\frac{eD_n}{L_n} \Delta n_0(x=0) \\ J_{nD,1}(0,t) &= eD_n \left.\frac{\partial \Delta n_1(x,t)}{\partial x}\right|_{x=0} \\ &= -\frac{eD_n}{L_n} \Re\left\{\Delta n_1(0) \sqrt{1 + \mathrm{j}\,\omega\tau_n}\, \mathrm{e}^{\mathrm{j}\,\omega t}\right\}. \end{aligned} \tag{5.97}$$

Setzt man (5.95) in (5.97) ein, so erhält man für den Elektronenstrom in der Ebene $x = 0$ (er ist dort praktisch identisch mit dem Diffusionsstrom der Elektronen) den Gleichanteil und den Wechselanteil (A ist die Querschnittsfläche. Im oben gewählten Experiment würde der Strom von rechts nach links fließen. Wir ändern deshalb hier die experimentelle Anordung so, dass der Strom

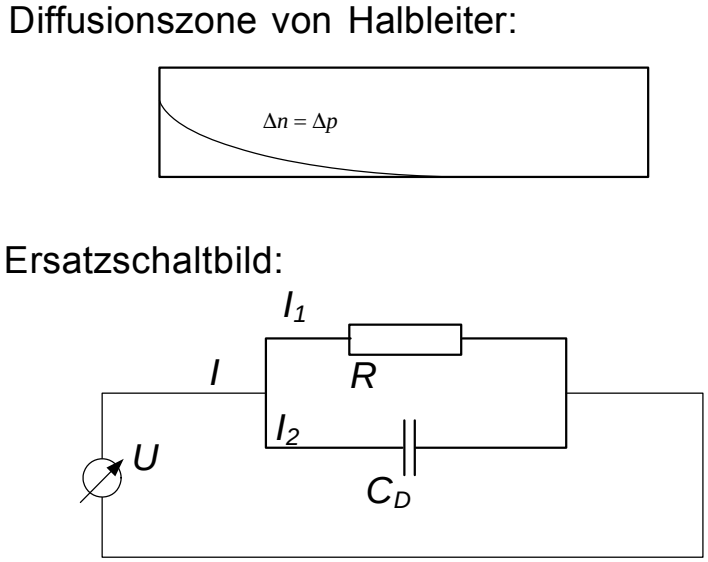


Abbildung 5.12: (a) Halbleiter mit Diffusionszone. Die neutralen Ladungen stellen ein abrufbares Ladungsreservoir dar und liefern deshalb einen Beitrag zur Kapazität. (b) Ersatzsschaldbild der Diffusionszone

nun von links nach rechts fließt. D.h. wir multiplizieren den Strom mit -1.)

$$\begin{aligned} I_{n0} &= -AJ_{n0} = A\frac{eD_n}{L_n}\Delta n_0(0) \\ &= A\frac{eD_n}{L_n}n_p\left(\mathrm{e}^{U_0/U_T}-1\right) \\ &\equiv I_{Sn}\left(\mathrm{e}^{U_0/U_T}-1\right), \\ I_{n1} &= -AJ_{n1} \overset{(5.97)}{=} \frac{eD_n}{L_n}\Re\left\{\Delta n_1(0)\sqrt{1+\mathrm{j}\,\omega\tau_n}\,\mathrm{e}^{\mathrm{j}\,\omega t}\right\}. \\ &= \frac{eD_n}{L_n}\Re\left\{\Delta n_1(0)\sqrt{1+\mathrm{j}\,\omega\tau_n}\,\mathrm{e}^{\mathrm{j}\,\omega t}\right\} \\ &= A\frac{eD_n}{L_n}n_p\frac{\mathrm{e}^{U_0/U_T}}{U_T}\sqrt{1+\mathrm{j}\,\omega\tau_n}U_1 \\ &= \frac{dI_{n0}}{dU_0}\sqrt{1+\mathrm{j}\,\omega\tau_n}\,U_1 \end{aligned} \tag{5.98}$$

mit dem Sättigungsstrom

$$I_{Sn} = Ae\frac{D_n n_p}{L_n}. \tag{5.99}$$

5.6.5 Diffusionskapazität

Die Kapazität ist per Definition, die gespeicherte Ladung pro Spannungseinheit. Dabei spielt es keine Rolle, ob die gespeicherte Ladung ladungsneutral ist. Wichtig ist lediglich, dass diese Ladung wieder abrufbar ist. Damit hat es dann auch in der ladungsneutralen Diffusionszone Ladungen, welche zur Gesamt-Kapazität beitragen können. Wir sprechen im Folgenden von der Diffusionskapazität, wenn wir vom kapazitiven Anteil der in der Diffusionszone gespeicherten Ladungen sprechen.

Das Ersatzschatlbild der Diffusionszone ist in Bild 5.12 gzeichnet. Nach Kirchhoff gilt

$$I = I_1 + I_2 \tag{5.100}$$

$$U = RI_1 \tag{5.101}$$

$$U = \frac{1}{j\omega C} I_2 \tag{5.102}$$

Das ergibt dann

$$I = \left(\frac{1}{R} + j\omega C \right) \; U = (g_0 + j\omega C) \; U = Y(\omega) \; U \tag{5.103}$$

In analoger Weise fürht der in Gl.(5.98) dargestellte Zusammenhang zwischen den Gleich- und Wechselstromamplituden I_{n0}, und I_{n1} und den Spannungsamplituden U_0 und U_1 gemäß Gl.(5.98) zur Definition einer Kleinsignal-Admittanz (Admittanz=komplexer Leitwert = Kehrwert der Impedanz; Leitwert=Kehrwert des Widerstandes)

$$I_{n1} = \frac{dI_{n0}}{dU_0} \sqrt{1 + \mathrm{j}\,\omega\tau_n}\, U_1 = Y(\omega) U_1. \tag{5.104}$$

$Y(\omega)$ ist die mit der Wirkung der Diffusionszonen assoziierte Diffusionsadmittanz

$$Y(\omega) = g_0 \sqrt{1 + \mathrm{j}\,\omega\tau_n} \tag{5.105}$$

mit dem Kleinsignal-Leitwert

$$g_0(U_0) = \frac{dI_{n0}}{dU_0} = \frac{I_{Sn}}{U_T}\, \mathrm{e}^{U_0/U_T}\,. \tag{5.106}$$

Für $\omega\tau_n \ll 1$ erhält man (beachte: $\sqrt{1 + \mathrm{j}\,\omega\tau_n} \simeq 1 + \frac{1}{2}\,\mathrm{j}\,\omega\tau_n$)

$$Y(\omega) = g_0 \left(1 + \frac{1}{2}\,\mathrm{j}\,\omega\tau_n \right). \tag{5.107}$$

Für $\omega \to 0$ ergibt sich der Kleinsignalleitwert dI/dU_0.

Für $\omega \neq 0$ entsteht durch Vergleich mit dem RC-Schaltkreis eine Kapazität, die **Diffusionskapazität**

$$C_D = \frac{1}{2} g_0 \tau_n. \tag{5.108}$$

Die Diffusionskapazität C_D beschreibt nur den wieder abrufbaren Anteil der in der Diffusionszone gespeicherten Minoritätsträgerladung.

In Tat und Wahrheit sind aber mehr Ladungen in der Diffusionszone gespeichert. Zu Vergleichszwecken ist es nun ganz interessant die in der Diffusionszone gespeicherten Überschuss-Minoritätsträgerladung und die dazugehörige Kleinsignalkapazität zu berechnen. Der Zusammenhang zwischen gespeicherter Ladung $Q = Q(U)$ und Kleinsignalkapazität C' im Arbeitspunkt $U = U_0$ ist

$$C' = \left. \frac{dQ(U)}{dU} \right|_{U=U_0}. \tag{5.109}$$

Die gespeicherte Ladung ergibt sich aus Bild 5.11,

$$\begin{aligned} Q_n(U) &= Ae \textstyle\int_0^\infty \Delta n(0) \exp\left[-\frac{x}{L_n} \right] dx = Ae\Delta n(0) L_n \\ &\overset{5.95}{=} Aen_p \left(\exp\left(\frac{U}{U_T} \right) - 1 \right) L_n = I_0(U)\tau_n. \end{aligned} \tag{5.110}$$

Damit erhalten wir für eine kleine Spannungsänderung

$$\begin{aligned} C' &\overset{(5.109)}{=} \left. \frac{dQ_n(U)}{dU} \right|_{U=U_0} \overset{(5.110)}{=} \frac{dI_0(U)\tau_n}{dU_0} \overset{(5.106)}{=} g_0\tau_n \\ &= 2C_D. \end{aligned} \tag{5.111}$$

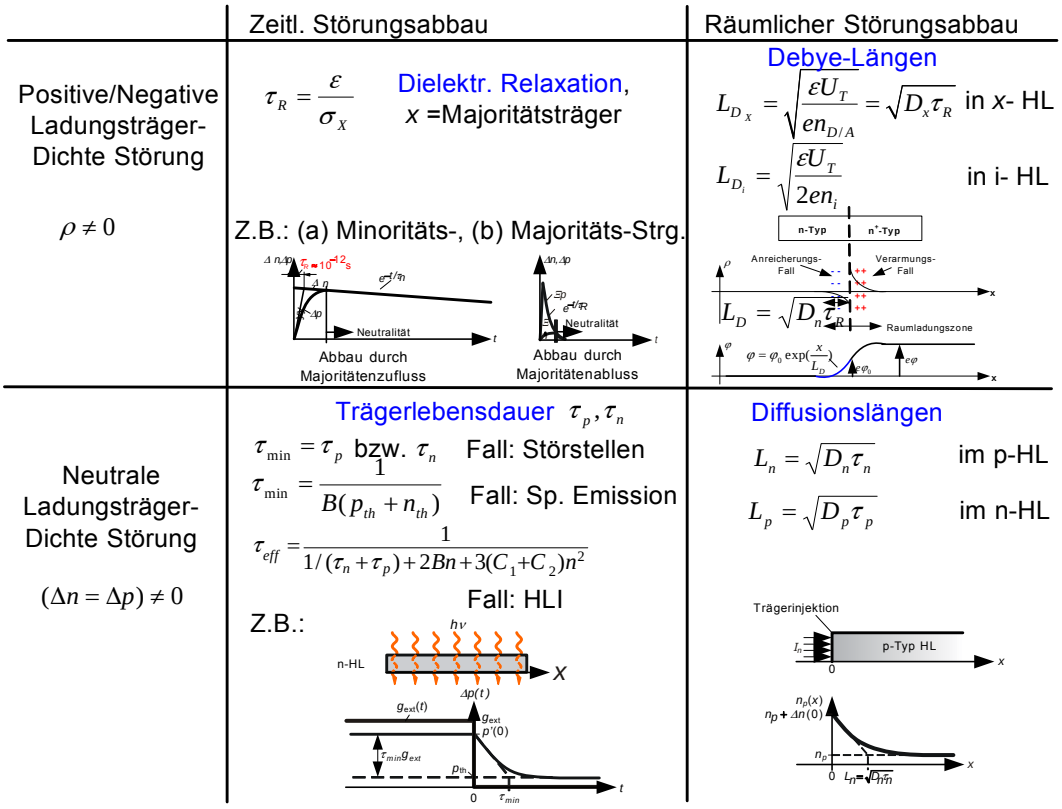


Abbildung 5.13: Zusammenfassung zu den Ladungsträgerstörungen.

Die auf diese Weise berechnete Kapazität ist falsch, denn die durch die Minoritätsträger vorhandene Ladung ΔQ ist nicht voll abrufbar; nur die Hälfte dieser Ladung ist gespeichert und ergibt einen kapazitiven Stromanteil. Die andere Hälfte verschwindet durch Rekombination.

In kurzen Diffusionsstrecken, wie in Abschnitt 5.6.2, siehe auch Bild 5.9 mit $w \ll L_n$, in welchen die Rekombination vernachlässigt werden kann, folgt aber $C_D \approx C'$.

5.7 Zusammenfassung

Wir haben in den letzten beiden Kapiteln Ladungsträgerstörungen betrachtet. Diese Störungen können als geladene Störungen, bzw. als ladungsneutrale Störungen auftreten. Die Störungen haben eine zeitliche Lebensdauer und eine räumliche Ausdehnung.

Geladene Störungen sind kurzlebig. Sie werden innerhalb der dielektrischen Relaxationszeit von den Majoritäten abgeschirmt. Wegen der kurzen Lebensdauer haben sie auch nur eine kurze räumliche Ausdehnung.

Ladungsneutrale Ladungsträgerdichtestörungen werden über Rekombinationsprozesse abgebaut. Ihre Lebensdauer ist durch den dominanten Rekombinationsprozess bestimmt. Im LLI-Fall bei indirekten Halbleitern ist das meist die Störstellenrekombination. Bei direkten Hableitern ist die Lebensdauer durch die Lebendauer der spontanen Emission bestimmt. Im HLI-Fall dominiert meist die Auger-Rekombination. Die räumliche Ausdehnung hängt entsprechend von der Trägerlebensdauer der ladungsneutralen Störung ab.

Kapitel 6

Der pn-Übergang

6.1 Einleitung

Die wichtigste Eigenschaft des ***pn-Übergangs (pn-junction)*,** ist die gleichrichtende Wirkung gegenüber Stromfluss bzw. gegenüber Stromfluss aufgrund einer angelegten Spannung. pn-Übergänge - oder pn-Dioden wenn wir von Bauteilen sprechen - werden in bipolaren Transistoren, Thyristoren, bipolaren integrierten Schaltungen, Photodetektoren, Solarzellen und Mikrowellenbauelementen wie Impatt-Dioden etc. verwendet. Mit zahlreichen III-V-Halbleitermaterialien können auch Leuchtdioden und ebenfalls Photodetektoren und Solarzellen hergestellt werden.

Wir unterscheiden zwischen ***Homostruktur (homostructure)*** Dioden und ***Heterostruktur (heterostructure)*** Dioden. Wir sprechen von Homostruktur Dioden, wenn die pn-Dotierung im gleichen Material gemacht wurde. Wir sprechen von Heterostruktur, wenn nicht nur die Dotierung sondern auch das Halbleitermaterial sich ändert. So könnte also beispielsweise eine Heterostruktur-Diode aus n-dotiertem InP und p-dotiertem InGaAs bestehen.

Bild 6.1 zeigt schematisch einige prinzipielle Herstellungsmethoden von pn-Dioden in Silizium-Technologie.

6.2 Der pn-Übergang im Gleichgewicht

Zur Beschreibung des Verhaltens eines pn-Überganges, siehe Bild 6.2(a), werden einige vereinfachende Annahmen gemacht:

1. Der pn-Übergang ist abrupt, d.h. die Dotierstoffkonzentrationen n_A im p-Gebiet und n_D im n-Gebiet sind räumlich konstant und fallen an der Grenzfläche $x = 0$ abrupt auf Null ab, siehe Bild 6.2(a)

$$n_A = \begin{cases} \text{konst für } x \leq 0 \\ 0 \text{ sonst} \end{cases} \tag{6.1}$$

$$n_D = \begin{cases} \text{konst für } x \geq 0 \\ 0 \text{ sonst} \end{cases} \tag{6.2}$$

   andere Dotierstoffe oder Störstellen sind nicht vorhanden.

2. Die Dotierstoffe sind erschöpft, d.h.

$$p_{p0} = n_A^- = n_A \tag{6.3}$$

$$n_{n0} = n_D^+ = n_D \tag{6.4}$$

   Die Notation ist in Fig. 2 gegeben. Die zugehörigen Minoritätenkonzentrationen sind

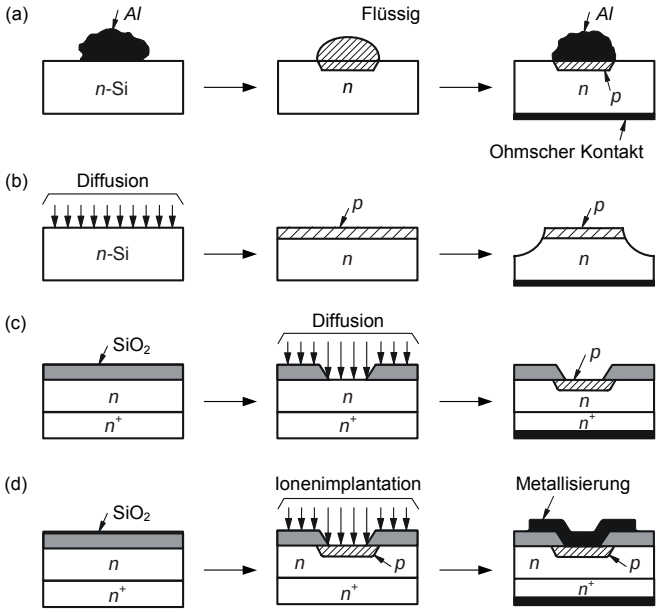


Abbildung 6.1: Einige Standard-Herstellungsmethoden für pn-Übergänge. (a) Legierungsübergang, (b) Eindiffundierter Mesaübergang, (c) Planartechnik auf epitaktisch gewachsenem Substrat, (d) Ionenimplantation [5]

$$n_{p0} = \frac{n_i^2}{p_{p0}} = \frac{n_i^2}{n_A} \qquad \text{(im p-Gebiet)} \tag{6.5}$$

$$p_{n0} = \frac{n_i^2}{n_{n0}} = \frac{n_i^2}{n_D} \qquad \text{(im n-Gebiet)} \tag{6.6}$$

3. Grenzflächenzustände treten nicht auf.

Berechnung der Diffusionsspannung U_D

Mit diesen Annahmen können die Ladungsträgerverteilungen und das Bändermodell weit weg vom pn-Übergang angegeben werden, siehe Bild 6.2. In der Umgebung des pn-Überganges fallen infolge von Diffusionsströmen und Driftströmen, welche sich im thermischen Gleichgewicht, wie in Abschnitt 4.2.1 diskutiert, wieder kompensieren, die Majoritätenkonzentrationen auf die Minoritätenkonzentrationen im jeweils gegenüberliegenden Halbleitermaterial ab. Dadurch entstehen Raumladungen, die auf der n-Seite positiv sind (n_D^+), auf der p-Seite negativ (n_A^-). Die Raumladung erzeugt gemäß der Poisson-Gleichung ein Potential welches zu einer Bandverbiegung $W_L(x)$, $W_V(x)$ Anlaß gibt, siehe Bild 6.2. Die Potentialdifferenz zwischen den beiden ungestörten n- und p-Gebieten entspricht einer Diffusionsspannung U_D mit

$$eU_D = W_L(-\infty) - W_L(\infty) \tag{6.7}$$

Aus Bild 6.2(f) folgt umittelbar mit den Gl.(5.15) und (5.16)

$$p_{p0} = n_A = N_V \exp\left[\frac{W_V(-\infty) - W_F}{kT}\right] \tag{6.8}$$

$$n_{n0} = n_D = N_L \exp\left[-\frac{W_L(\infty) - W_F}{kT}\right] \tag{6.9}$$

Wir benötigen noch

$$W_L = W_V + W_G \tag{6.10}$$

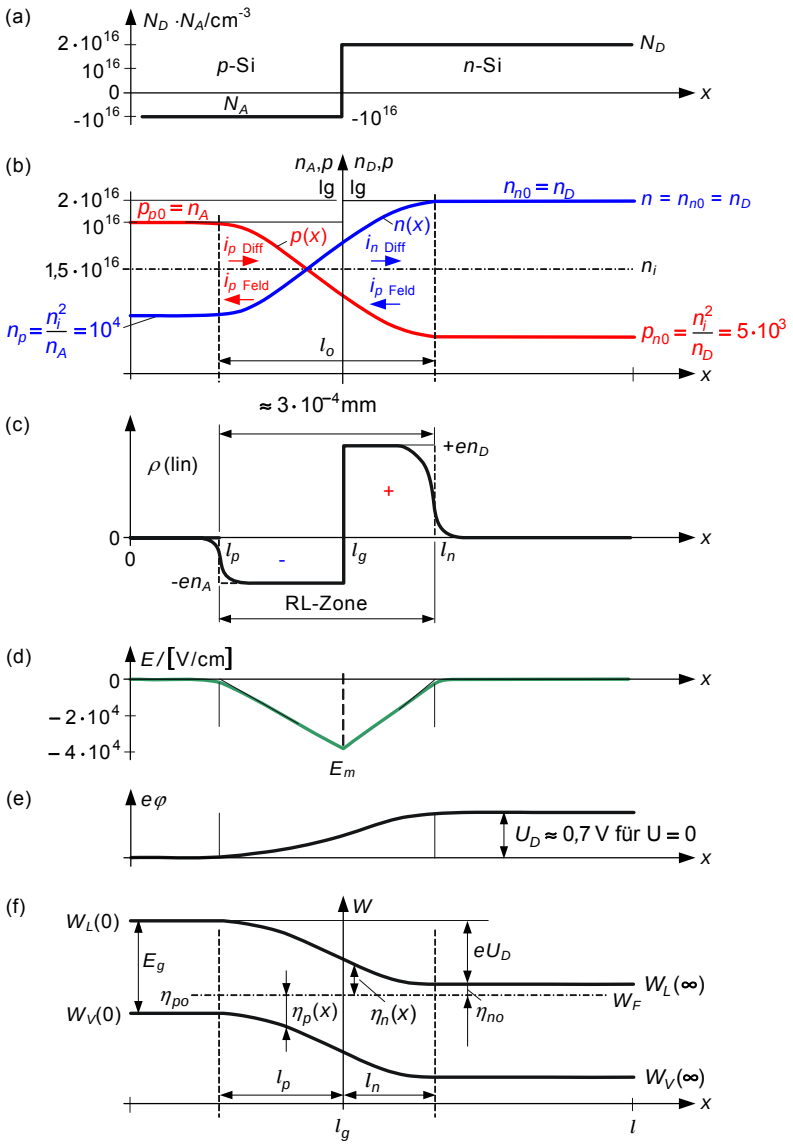
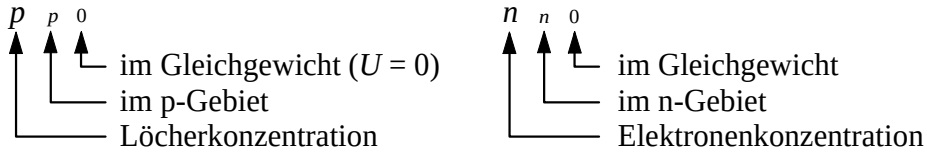


Abbildung 6.2: pn-Übergang im thermischen Gleichgewicht: (a) Dotierung; (b) Ladungsträgerverteilung; (c) Raumladung; (d) Elektrisches Feld; (e) Potential; (f) Bandverlauf [3]



T = 300 K	Ge	Si	GaAs
n_i^2/cm^{-6}	$5,8 \cdot 10^{26}$	$2,1 \cdot 10^{20}$	$3,2 \cdot 10^{12}$
n_A/cm^{-3}	10^{15}	10^{15}	10^{15}
n_D/cm^{-3}	10^{15}	10^{15}	10^{15}
$\mathbf{U_D/V}$	$\mathbf{0,18}$	$\mathbf{0,56}$	$\mathbf{1,0}$
n_A/cm^{-3}	10^{15}	10^{15}	10^{15}
n_D/cm^{-3}	10^{18}	10^{18}	10^{18}
$\mathbf{U_D/V}$	$\mathbf{0,36}$	$\mathbf{0,73}$	$\mathbf{1,18}$
n_A/cm^{-3}	10^{18}	10^{18}	10^{18}
n_D/cm^{-3}	10^{18}	10^{18}	10^{18}
$\mathbf{U_D/V}$	$\mathbf{0,53}$	$\mathbf{0,90}$	$\mathbf{1,35}$

Tabelle 6.1: Diffusionspannungen in den Halbleitern Ge, Si, GaAs für verschiedene Dotierungskonzentrationen

um aus Gl. (6.7)-(6.10) die Diffusionsspannung U_D auszurechnen. Mit dem Massenwirkungsgesetz ergibt sich

$$n_A n_D = N_L N_V \exp\left[-\frac{W_G}{kT}\right] \exp\left[\frac{U_D}{U_T}\right] = n_i^2 \exp\left[\frac{U_D}{U_T}\right] \tag{6.11}$$

$$= n_i^2 \exp\left[\frac{U_D}{U_T}\right]. \tag{6.12}$$

Damit ist die ***Diffusionsspannung (built-in potential)*** dann gegeben durch

$$U_D = U_T \ln\left(\frac{n_A n_D}{n_i^2}\right) . \tag{6.13}$$

Da die Diffusionsspannung nur logarithmisch von der Konzentration der Dotierstoffe abhängt, variiert die Diffusionsspannung nur schwach mit diesen Konzentrationen. In Tabelle 6.1 sind dafür einige Beispiele zu finden.

Gl.(6.13) und Tabelle 6.1 lassen folgende Schlüsse zu:

1. Aus der Diffusionsspannung folgt unmittelbar der folgende Zusammenhang zwischen der Majoritätenkonzentration im p- oder n-Gebiet und der Minoritätenkonzentration im gegenüberliegenden n- oder p-Gebiet

$$n_{p0} \overset{\text{Massenwirkg.}}{=} \frac{n_i^2}{p_{p0}} \overset{\text{Störstellenersch.}}{=} \frac{n_i^2}{n_A} \overset{(6.13)}{=} n_D e^{-U_D/U_T} \overset{\text{Störstellenersch.}}{=} n_{n0} e^{-U_D/U_T}, \tag{6.14}$$

$$p_{n0} \overset{\text{Massenwirkg.}}{=} \frac{n_i^2}{n_{n0}} \overset{\text{Störstellenersch}}{=} \frac{n_i^2}{n_D} \overset{(6.13)}{=} n_A e^{-U_D/U_T} \overset{\text{Störstellenersch.}}{=} p_{p0} e^{-U_D/U_T}. \tag{6.15}$$

2. die Diffusionsspannung ist eine Sperrspannung ($U_D > 0$), und behindert den Stromtransport über den pn-Übergang.

3. die Diffusionsspannung hängt über U_T und n_i^2 schwach von der Temperatur ab.

4. die Diffusionsspannung hängt nur schwach von den Dotierungen auf der n- und p-Seite ab. Mit wachsender Dotierstoffkonzentration geht sie gegen $U_D \to E_g/e$.

5. Die Diffusionsspannung ist an den Enden der p- und n- Zonen nicht meßbar, da sich an äußeren Kontakten, welche zum Beispiel durch hochdotierte n-Zonen bewerkstelligt werden könnten, ebenfalls Diffusionsspannungen aufbauen, welche sich mit der am pn-Übergang im äußeren Stromkreis gerade kompensieren, siehe 6.3. Dies muß so sein, denn sonst könnte man im thermischen Gleichgewicht mit dieser Anordnung einen Stromkreis speisen.

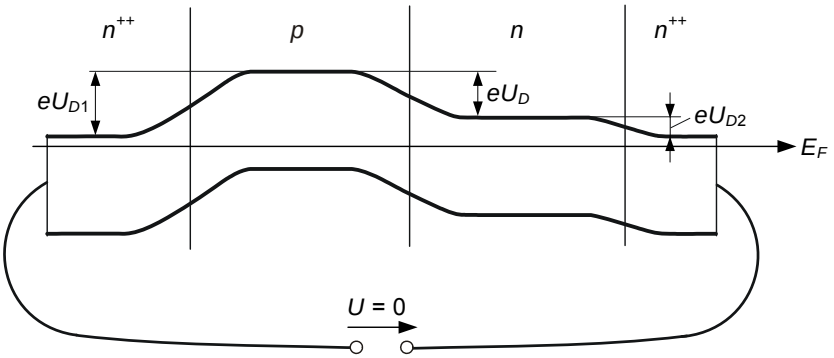


Abbildung 6.3: Die Diffusionsspannung kann nicht an äußeren Kontakten festgestellt werden, da die Summe der Kontaktspannungen gerade der Diffusionsspannung entspricht.

Ladungsträgerverteilung, Bandverlauf

Um die Ladungsträgerverteilung zu erfassen wird als erstes die Poisson-Gleichung gelöst, damit lassen sich dann der Bandkantenverlauf und über die Gln.(5.15) und (5.16) die Trägerdichten berechnen. Um die Berechnung analytisch durchführen zu können, verwenden wir die ***Schottky-Näherung.*** Die Schottky-Näherung besagt, dass die beweglichen Ladungsträger aus der sich aufbauenden ***Raumladungszone (RLZ)=Verarmungszone (Depletion Region)*** völlig ausgeräumt sind, i.e. $n(x) = p(x) = 0$. Man spricht dann oft davon, dass die RLZ an Ladungsträgern "*verarmt*" wäre. Dies ist sinnvoll, da die Länge der RLZ typischerweise mehrere Debye-Längen beträgt, während sich das Potential bereits über eine Debye-Länge um U_T verändert und sich daher die Ladungsträgerverteilung auf einer linearen Skala sehr abrupt ändert, siehe Bild 6.2.

Als erstes wird die Poisson-Gleichung je für das n- und p-Gebiet getrennt aufgestellt und dann die Lösungen an der Grenzfläche $x = 0$ in geeigneter Weise aneinander angepaßt.

Die Poisson-Gleichung lautet:

$$\frac{d^2\varphi}{dx^2} = -\frac{\rho}{\varepsilon_r \varepsilon_0} \tag{6.16}$$

Mit der Ladungsträgerdichte

$$\rho = e\left[n_D^+ + p(x) - n_A^- - n(x)\right] \overset{\text{Schottky-Näherung}}{=} e\left[n_D^+ - n_A^-\right] \tag{6.17}$$

im Schottky-Modell (machmal auch als **Schottky'sche Parabel-Näherung** bezeichnet) ist ρ in den p- und n- RLZ konstant und deshalb verläuft die Feldstärke $E = -\text{grad}(\varphi)$ linear. Es gilt also

$$\frac{dE}{dx} = \frac{\rho}{\varepsilon_r \varepsilon_0}. \tag{6.18}$$

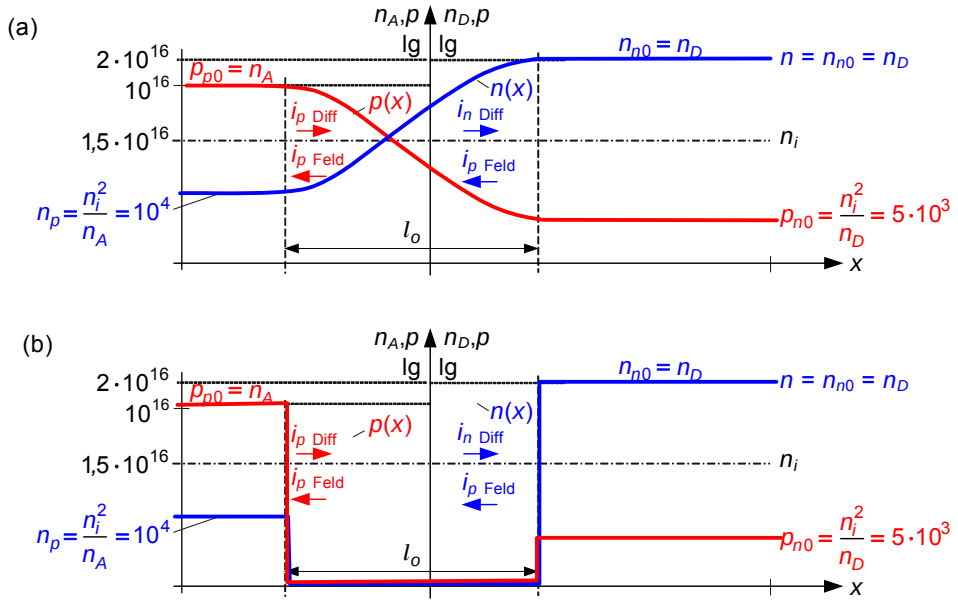


Abbildung 6.4: (a) Wahre Ladungsträgerverteilung, (b) Ladungsträgerverteilung in der Schottky-Näherung.

Für p- und n-Seiten führt dies auf:

	p-Seite	n-Seite
Bereich:	$-l_p \le x \le 0$	$0 \le x \le l_n$
Raumladung:	$n_D^+ = 0$	$n_A^- = 0$
	$n_A^- = n_A$	$n_D^+ = n_D$
	$p(x) = n(x) = 0$ (Schottky.)	$p(x) = n(x) = 0$ (Schottky.)
$\Rightarrow$	$\rho_p = -en_A$	$\rho_n = en_D$
Feldstärke:	$\frac{dE}{dx} = -\frac{en_A}{\varepsilon_r\varepsilon_0}$	$\frac{dE}{dx} = \frac{en_D}{\varepsilon_r\varepsilon_0}$
$\Rightarrow$	$E(x) = -\frac{en_A}{\varepsilon_r\varepsilon_0}(x + l_p)$	$E(x) = \frac{en_D}{\varepsilon_r\varepsilon_0}(x - l_n)$
$\Rightarrow$	$\varphi(x) = \frac{en_A}{2\varepsilon_r\varepsilon_0}(x + l_p)^2$	$\varphi(x) = U_D - \frac{en_D}{2\varepsilon_r\varepsilon_0}(x - l_n)^2$
Bandverläufe:	$W_L(x) = W_L(-\infty) - \frac{e^2 n_A}{2\varepsilon_r\varepsilon_0}(x + l_p)^2$	$W_L(x) = W_L(+\infty) + \frac{e^2 n_D}{2\varepsilon_r\varepsilon_0}(x - l_n)^2$

(6.19)

Die Bandverläufe im p- und n-Bereich in der letzten Zeile von 6.19 folgen am einfachsten aus

$$W_L(x) = W_L(-\infty) + e\int_{-\infty}^{x} E(x)dx, \tag{6.20}$$

$$W_L(x) = W_L(+\infty) - e\int_{x}^{\infty} E(x)dx. \tag{6.21}$$

Die Bandverläufe sind in der Schottky-Näherung parabolisch.

Noch fehlen uns die Werte für die maximale Feldstärke E_m an Übergang vom p- zum n-Gebiet oder die Ausdehnung l_p und l_n. Um diese 3 Größen zu bestimmen, benötigen wir mindestens drei

weitere Bestimmungsgleichungen:

So gilt erstens wegen der Gesamtladung= 0

$$en_A l_p = en_D l_n :\rightarrow \text{Gesamtladung} = 0 \tag{6.22}$$

Zweitens folgt aus der Stetigkeit des elektrischen Feldes bei $x = 0$,: für die maximale Feldstärke am pn-Übergang aus 6.19

$$E_m = -\frac{en_A}{\varepsilon_r \varepsilon_0} l_p = -\frac{en_D}{\varepsilon_r \varepsilon_0} l_n, \tag{6.23}$$

bzw. für die Ausdehnung der RLZ in das p- und n-Gebiet

$$l_p = -\frac{\varepsilon_r \varepsilon_0}{en_A} E_m, \tag{6.24}$$

$$l_n = -\frac{\varepsilon_r \varepsilon_0}{en_D} E_m \tag{6.25}$$

Drittens, ist die totale Potentialdifferenz, bzw. Diffusionsspannung durch die negative Fläche unter der Feldstärke bestimmt

$$U_D = -\int E \, dx = -\frac{1}{2} E_m \left(l_n + l_p\right) = \frac{1}{2} E_m^2 \frac{\varepsilon_r \varepsilon_0}{e} \left(\frac{1}{n_A} + \frac{1}{n_D}\right), \tag{6.26}$$

Durch Auflösen der drei unabhängigen Gleichungen .(6.22), (6.23) und (6.26) lässt sich E_m bestimmen

$$E_m = -\sqrt{\frac{2e}{\varepsilon_r \varepsilon_0} \frac{U_D}{\left(\frac{1}{n_A} + \frac{1}{n_D}\right)}}. \tag{6.27}$$

Damit sind dann auch gemäß (6.24) und (6.25) die Längen der RLZ im n- und p-Gebiet bestimmt

$$l = l_p + l_n = \sqrt{\frac{2\varepsilon_r \varepsilon_0}{e} U_D \left(\frac{1}{n_A} + \frac{1}{n_D}\right)} \tag{6.28}$$

$$l_p = l \frac{n_D}{n_A + n_D}, \quad l_n = l \frac{n_A}{n_A + n_D} \tag{6.29}$$

Sobald die Dotierstoffkonzentrationen n_A n_D für einen bestimmten Halbleiter feststehen, kann die Ausdehnung der RLZ in das p- und das n-Gebiet (und damit die Gesamtausdehnung) angegeben werden. Sind die Konzentrationen n_A, und n_D gleich, so dehnt sich die RLZ in beide Gebiete gleich weit aus. Ist jedoch eine Konzentration wesentlich größer als die andere, so dehnt sich die RLZ fast nur in die niedriger dotierte Zone aus. Für die Zahlenbeispiele aus Tab. 6.1 sind in Tab. 6.2 die zugehörigen Ausdehnungen der RLZ in die p- und n-Gebiete berechnet. Je nach Dotierung betragen die Ausdehnungen wenige Nanometer bis zu Mikrometer.

Für die in Tab.6.2 angegebenen Zahlenwerte erhält man maximale Feldstärken zwischen etwa 5 kV/cm (Ge, $n_A = n_D = 10^{15}$ cm^{-3}) und etwa 500 kV/cm (GaAs, $n_A = n_D = 10^{18}$ cm^{-3})

Konzentrationsverteilung

Mit Hilfe der Gln.(5.15) und (5.16) und dem näherungsweise berechneten Bandverlauf $W_L(x)$, nach (6.20) und (6.21) kann eine Konzentrationsverteilung $n(x)$ und entsprechend $p(x)$ berechnet werden:

$$n(x) = N_L e^{-[W_L(x) - W_F]/kT} \tag{6.30}$$

$$p(x) = N_V e^{+[W_V(x) - W_F]/kT} \tag{6.31}$$

$T = 300$ K	Ge	Si	GaAs
ε_r	16	11,9	13,1
n_A/cm^{-3}	10^{15}	10^{15}	10^{15}
n_D/cm^{-3}	10^{15}	10^{15}	10^{15}
U_D/V	0,18	0,56	1,0
l_p/μm	0,4	0,6	0,85
l_n/μm	0,4	0,6	0,85
n_A/cm^{-3}	10^{15}	10^{15}	10^{15}
n_D/cm^{-3}	10^{18}	10^{18}	10^{18}
U_D/V	0,36	0,73	1,18
l_p/μm	0,8	1	1,3
l_n/μm	0,0008	0,001	0,0013
n_A/cm^{-3}	10^{18}	10^{18}	10^{18}
n_D/cm^{-3}	10^{18}	10^{18}	10^{18}
U_D/V	0,53	0,9	1,35
l_p/μm	0,02	0,02	0,03
l_n/μm	0,02	0,02	0,03

Tabelle 6.2: Ausdehnung der Raumladungszone in Ge, Si, GaAs bei unterschiedlichen Dotierungen

Man erhält damit eine Verbesserung gegenüber der ersten vereinfachenden Annahme $n(x) = p(x) = 0$ in der RLZ. Einsetzen von (6.21) in (6.30) ergibt z.B. im n-HL:

$n_n(x) = N_L \exp\left[-\frac{\eta(x)}{kT}\right]$

$$\begin{aligned} n_n(x) &= N_L \exp\left[-\frac{W_L(\infty) - e\varphi(x) - W_F}{kT}\right] \\ &= N_L \exp\left[-\frac{W_L(\infty) - W_F}{kT}\right] \cdot \exp\left[-\frac{e^2 n_D (x - l_n)^2}{2\varepsilon_r \varepsilon_0 kT}\right] \end{aligned} \tag{6.32}$$

Die Verteilung ist demnach innerhalb der RLZ durch eine Gaußfunktion beschrieben. Mit der Debye-Länge nach Abschnitt 5.5

$$L_{Dn} = \sqrt{\frac{\varepsilon_r \varepsilon_0 kT}{e^2 n_D}} \tag{6.33}$$

und

$$N_L \exp\left[-\frac{W_L(\infty) - W_F}{kT}\right] = n_{n0} = n_D \tag{6.34}$$

ergibt sich

$$n_n(x) = n_D \exp\left\{-\frac{1}{2}\left(\frac{x - l_n}{L_{Dn}}\right)^2\right\}. \tag{6.35}$$

Entsprechend folgt für die Löcherkonzentration im p-Gebiet:

$$p_p(x) = n_A \exp\left\{-\frac{1}{2}\left(\frac{x - l_p}{L_{Dp}}\right)^2\right\}. \tag{6.36}$$

Für die Elektronenkonzentration im p-Gebiet folgt aus

$$n_p(x) = \frac{n_i^2}{p_p(x)} = \frac{n_i^2}{n_A} \exp\left\{\frac{1}{2}\left(\frac{x - l_p}{L_{Dp}}\right)^2\right\} \tag{6.37}$$

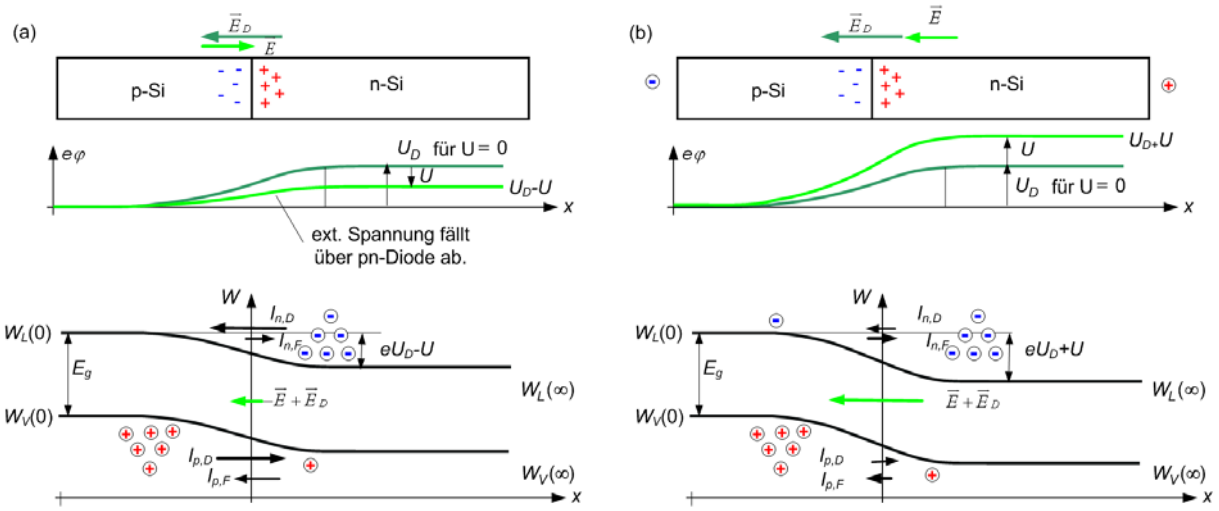


Abbildung 6.5: Eine von außen angelegte Spannung fällt über der RLZ ab. (a) Spannung in Flussrichtung, (b) Spannung in Sperrrichtung.

und die Löcherkonzentration $p_n(x)$ berechnet sich aus

$$p_n(x) = \frac{n_i^2}{n_D} \exp\left\{\frac{1}{2}\left(\frac{x - l_n}{L_{Dn}}\right)^2\right\}. \tag{6.38}$$

6.3 Der pn-Übergang im Nichtgleichgewicht (Stromfluss)

Wird an die Klemmen des pn-Überganges eine äußere Spannung U angelegt, so fällt diese zusätzlich an der hochohmigen RLZ ab, da dort nur verschwindend wenig freie Ladungsträger vorhanden sind, siehe Bild 6.5. Ist die Spannung U positiv, so wirkt sie der Diffusionsspannung entgegen und führt zu einer Verkleinerung der RLZ und das Umgekehrte gilt im Falle $U < 0$. Da die Gesamtspannung $U - U_D$ am pn-Übergang die Weite der RLZ bestimmt, gelten im Nichtgleichgewicht die selben Formeln für die Weiten der RLZ, nur ist U_D durch $U_D - U$ zu ersetzen. Im Gleichgewichtsfall, $U = 0$, führt die RLZ zu einer effektiven Isolierung der beiden homogenen Halbleiterbereiche und es gelten in diesen die selben Verhältnisse wie in den isolierten homogenen Gebieten im thermischen Gleichgewicht. Für $U \neq 0$ ist dies nicht mehr der Fall.

Wird in den Gln. (6.14) und (6.15) U_D durch $U_D - U$ substituiert, so folgt, dass die Minoritätsträgerdichten an den Rändern der RLZ nicht mehr denen im ungestörten Halbleiter entsprechen. Es gilt

$$n_p(-l_p) = n_D e^{(U-U_D)/U_T} = n_{p0} e^{U/U_T}, \tag{6.39}$$

$$p_n(l_n) = n_A e^{(U-U_D)/U_T} = p_{n0} e^{U/U_T}. \tag{6.40}$$

An den Rändern der RLZ bilden sich deshalb die in Abschnitt 5.6 besprochenen Diffusionszonen aus.

Mit der Ersetzung von U_D durch $U_D - U$ ändert sich auch E_m (siehe 6.27) und die Ausdehnung der Raumladungszonen l, l_p, l_n (siehe 6.28, 6.29):

$$E_m \to E_m = -\sqrt{\frac{2e}{\varepsilon_r \varepsilon_0} \frac{U_D - U}{\left(\frac{1}{n_A} + \frac{1}{n_D}\right)}} \tag{6.41}$$

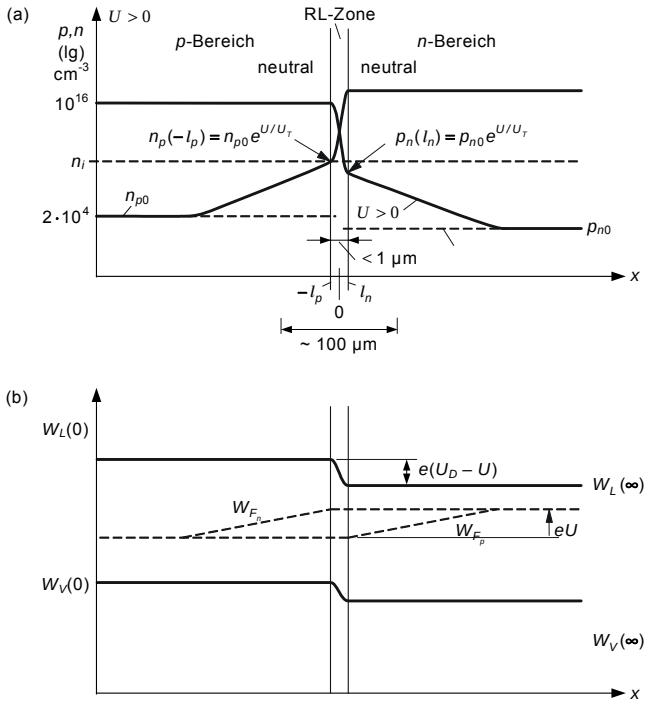


Abbildung 6.6: pn-Übergang in Flussrichtung, $U > 0$: (a) Ladungsträgerdichten, (b) Bandverläufe [3].

und

$$l \to l = \sqrt{\frac{2\varepsilon_r \varepsilon_0}{e}(U_D - U)\left(\frac{1}{n_A} + \frac{1}{n_D}\right)} \tag{6.42}$$

was zu automatisch zu einer analogen Anpassung von l_p und l_n führt.

6.3.1 Fall: $U > 0$ in Flussrichtung (Forward-Biased)

In Bild 6.6 sind Konzentrations- und Bandverläufe für $U > 0$ logarithmisch dargestellt. Die Raumlandungszone schrumpft unter die Ausdehnung im Gleichgewichtsfall. Da die RLZ von der Größenordnung einiger Debye-Längen sind und die Diffusionslängen sehr viel größer sind als die Debye-Längen ist die Diffusionszone typischerweise um den Faktor 100-1000 mal größer als die RLZ.

6.3.2 Fall: $U < 0$ in Sperrrichtung (Reverse-Biased)

Die Ladungsträgerkonzentrationen und Bandverläufe im Sperrbetrieb der Diode sind in Bild 6.7 zusammengestellt. Das Sättigungsverhalten im Sperrbetrieb kommt dadurch zustande, dass die Minoritätendichten in den Diffusionszonen natürlich nicht kleiner als Null werden können, was den Diffusionsstrom in Richtung der RLZ beschränkt. Im Sperrbetrieb sind die Überschußdichten am Rande der RLZ $\Delta p \approx -p_{n0}$. Damit ist nach Gl.(5.10) die Nettogenerationsrate gleich p_{n0}/τ_p. Diese Generation findet im Volumen $A \cdot L_p$ statt, was zeigt, dass der Sättigungsstrom durch die Ladungsträgergeneration in den Diffusionszonen hervorgerufen wird.

6.3.3 Ströme

In Bild 6.8 sind die Elektronen- und Löcherströme für den Durchlaßbereich über den Diffusionszonen und der RLZ und für den Fall $U > 0$ eingezeichnet. Aus der Quellenfreiheit des Gesamtstromes

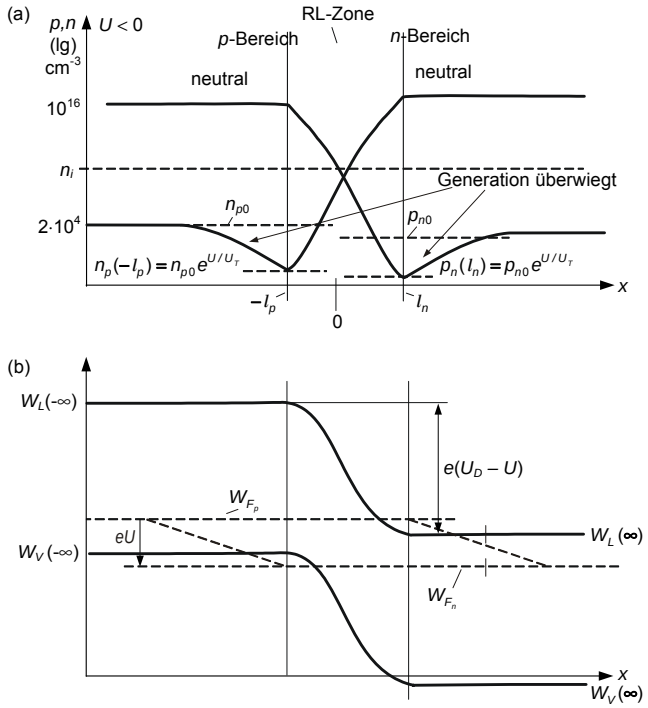


Abbildung 6.7: pn-Übergang in Sperrrichtung, $U < 0$: (a) Ladungsträgerdichten, (b) Bandverläufe [3].

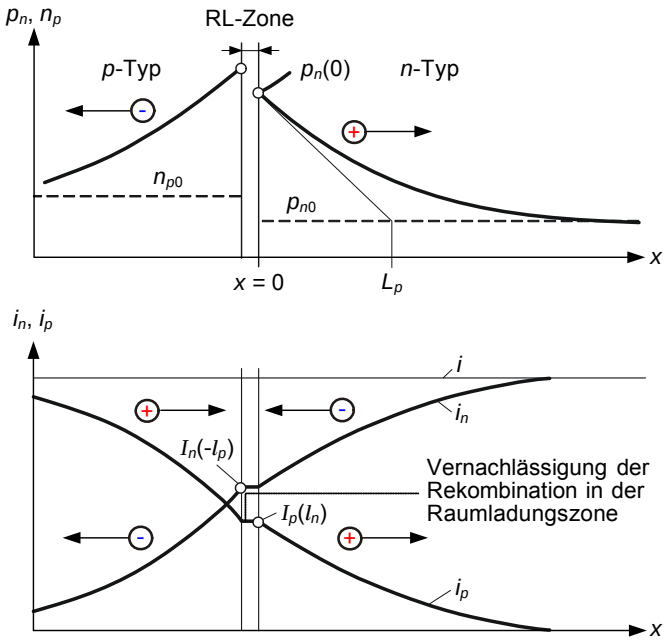


Abbildung 6.8: Diffusionszonen und Stromverläufe durch die Diode [3].

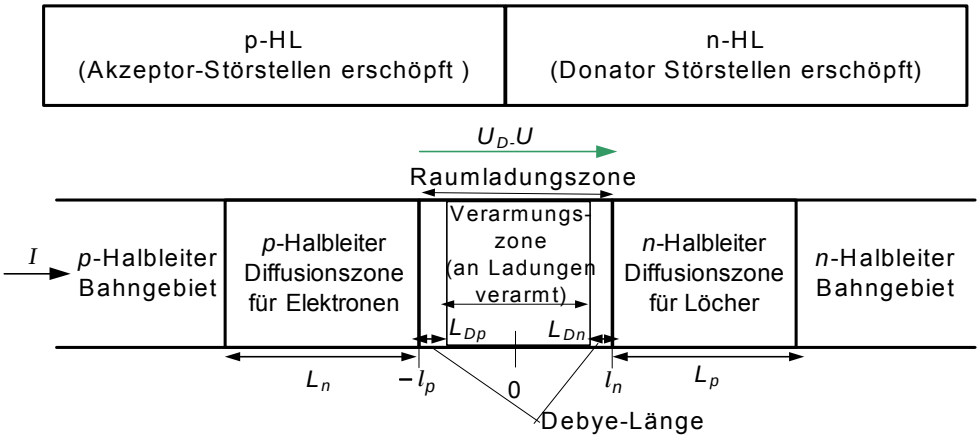


Abbildung 6.9: Schematische Darstellung des pn-Überganges.

(5.9) folgt im stationären Betrieb, dass der Gesamtstrom konstant sein muß

$$J_n + J_p = const \tag{6.43}$$

Der Diffusionsstrom der Minoritätsträger in den Diffusionszonen wird gemäß Abschnitt 5.6 in einen Driftstrom der Majoritätsträger umgewandelt. Ferner führt jede Diffusionszone einen Driftstrom, welcher in der anderen Diffusionszone zum Minoritätsstrom wird.

Denkt man sich nach Bild 6.9 einen pn-Übergang als Kombination zweier durch eine RLZ getrennter Diffusionszonen, so läßt sich der Gesamtstrom als Summe eines Elektronen Diffusionsstromes an der Stelle $-l_p$ gemäss Gl. (5.98) und eines analogen Löcheranteils an der Stelle l_n schreiben. Dies ergibt den Gesamtstrom, falls die Stromanteile von Elektronen und Löchern im Intervall $-l_p \leq x \leq l_n$ konstant bleiben, d. h. die Generation oder Rekombination von Trägern in dieser Zone vernachlässigbar ist (tatsächlich gibt es auch in der RLZ kleine Ströme, das führt uns später zu einer Modifikation der idealen Diodenkennlinie). Wir betrachten dazu gleich den allgemeinsten Fall einer angelegten Spannung welche aus einer Gleichspannung und einer dazu überlagerten Wechselspannung, $U = U_0 + \Re\left\{U_1 \exp(\mathrm{j}\,\omega t)\right\}$. Aus Bild 6.9 und (5.98) folgt für den durch die Diode fließenden Gleich- und Wechselstromanteil

$$\begin{aligned} I_0 = I_{n0}\left(-l_p\right) + I_{p0}\left(l_n\right) &= Ae\left(\frac{D_n n_p}{L_n} + \frac{D_p p_n}{L_p}\right)\left(\mathrm{e}^{U_0/U_T} - 1\right) = \\ &= \left(I_{Sn} + I_{Sp}\right)\left(\mathrm{e}^{U_0/U_T} - 1\right) = I_S\left(\mathrm{e}^{U_0/U_T} - 1\right), \\ I_1 = I_{n1}\left(-l_p\right) + I_{p1}\left(l_n\right) &= U_1\left[\frac{dI_{n0}}{dU_0}\sqrt{1 + \mathrm{j}\,\omega\tau_n} + \frac{dI_{p0}}{dU_0}\sqrt{1 + \mathrm{j}\,\omega\tau_p}\right] \\ &= Y(\omega)U_1. \end{aligned} \tag{6.44}$$

Die Strom-Spannungs Gleichstrom-Kennlinie aus (6.44) ist auch unter dem Namen ***Shockley-Gleichung*** bekannt. Die beiden Gleichungen beschreiben das Großsignalverhalten der Diode.

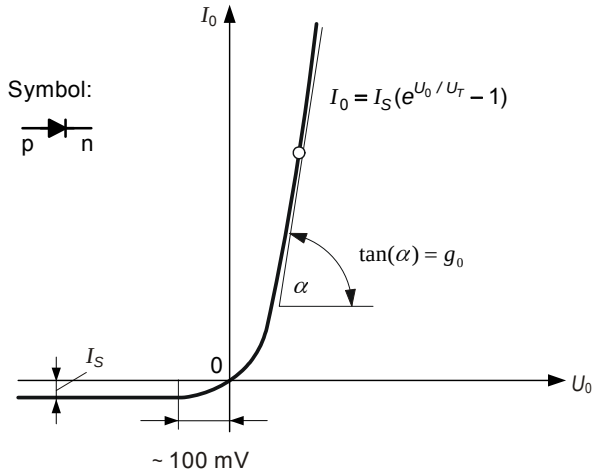


Abbildung 6.10: Gleichstrom-Kennlinie (auch Shockley-Kennlinie genannt) der pn-Diode [3].

1. Gleichstromkennlinie

Die Strom-Spannungskennline (6.44) ist in Bild 6.10 gezeichnet. Mit Gl.(6.44) erhält man für $U_0 \to -\infty$ den Sättigungsstrom I_S der Diode. Dieser ist

$$I_S = I_{Sn} + I_{Sp} \tag{6.45}$$

$$I_{Sp} = Ae\frac{D_p p_{n0}}{L_p} = Ae\frac{D_p n_i^2}{L_p n_D} \tag{6.46}$$

$$I_{Sn} = Ae\frac{D_n n_{p0}}{L_n} = Ae\frac{D_n n_i^2}{L_n n_A} \tag{6.47}$$

$$I_S = Aen_i^2\left(\frac{D_n}{L_n n_A} + \frac{D_p}{L_p n_D}\right) \tag{6.48}$$

Falls $U > 0$, ist die RLZ kurz gegen die Diffusionszonen und der Durchlaßstrom ist groß. Deshalb kann der Strom infolge von Rekombination in der RLZ zumindestens im Durchlaßbereich vernachlässigt werden.

Für Spannungen kleiner als -U_T fließt ein negativer Sperrsättigungsstrom $-I_S$.

2. Kleinsignal- und Wechselstromverhalten:

Zur Kleinsignalanalyse elektrischer Netzwerke muß die lineare Änderung des Stromes ΔI der Diode als Funktion der Spannungsänderung ΔU bekannt sein. Aus Gl.(6.44) folgt für den Gleichstrom-Kleinsignalleitwert der Diode mit $U = U_0 + \Delta U$

$$g_0(U_0) = \frac{dI_0}{dU_0} = \frac{I_0(U_0) + I_S}{U_T}, \tag{6.49}$$

der natürlich vom Arbeitspunkt U_0 abhängt, siehe Bild 6.10 (Steigung der Tangente an die statische Diodenkennlinie im Arbeitspunkt).

Für hohe Sperrspannungen ist $I = -I_S$ und der entsprechende Kleinsignalleitwert verschwindet.

Bei starker Flusspolung ist $I \gg I_S$ und es folgt

$$g_0/\Omega = \frac{I/\mathrm{mA}}{26}. \tag{6.50}$$

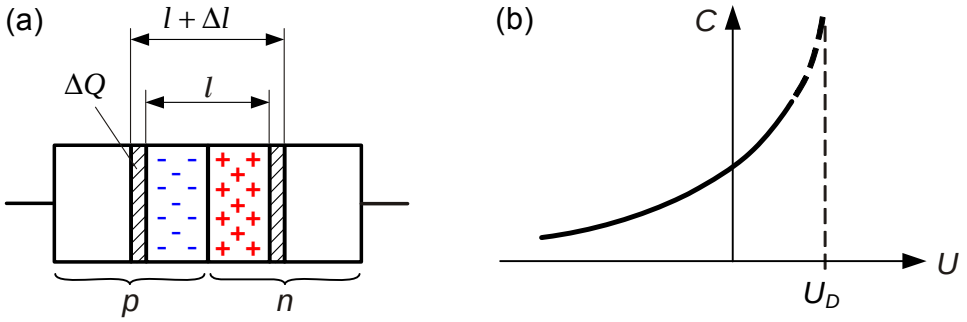


Abbildung 6.11: Ladungsspeicherung in der Sperrschichtkapazität [3].

Liegt eine Wechselspannungsänderung vor, d.h. $U = U_0 + \Re\{U_1 \exp(\mathrm{j}\,\omega t)\}$, so folgt aus Gl.(6.44) für die mit den Diffusionszonen assoziierte Diffusionsadmittanz, $Y(\omega)$ im Falle niedriger Frequenzen, $\omega\tau_n$, $\omega\tau_p \ll 1$, (beachte: $\sqrt{1+\mathrm{j}\,\omega\tau} \approx 1 + \frac{1}{2}\mathrm{j}\,\omega\tau$)

$$Y(\omega) = \frac{dI_{n0}}{dU_0}(1+\mathrm{j}\,\omega\frac{1}{2}\tau_n) + \frac{dI_{p0}}{dU_0}(1+\mathrm{j}\,\omega\frac{1}{2}\tau_p) = g_{n0}(1+\mathrm{j}\,\frac{1}{2}\omega\tau_n) + g_{p0}(1+\mathrm{j}\,\frac{1}{2}\omega\tau_p) \tag{6.51}$$

$$\approx g_0 + \mathrm{j}\,\frac{1}{2}\omega\; g_0\tau. \tag{6.52}$$

Für $\omega \to 0$ ergibt sich wieder der Kleinsignalleitwert g_0. Für $\omega \neq 0$ entsteht eine Diffusionskapazität, die im Spezialfall $\tau_n = \tau_p = \tau$ durch

$$C_D = \frac{1}{2}g_0\tau \tag{6.53}$$

gegeben ist, und wie in Abschnitt 5.6 besprochen, die Ladungsspeicherung in der Diffusionszone modelliert.

3. Sperrschichtkapazität

Wir haben bisher lediglich die Ladungsspeicherung in den Diffusionszonen mitberücksichtigt. Bei einer Spannungsänderung verändert sich aber auch die Größe der RLZ, Bild 6.11, wodurch sich die in der p- oder n-RLZ gespeicherte Ladung Q ändert. Es gilt für die damit verbundene ***Sperrschichtkapazität (depletion capacitance)***

$$C_S = \frac{dQ}{dU} = \frac{A}{l}\varepsilon_r\varepsilon_0. \tag{6.54}$$

Dies ist die Kapazität eines Plattenkondensators mit dem Plattenabstand l nach Gl. 6.42), was der Weite der RLZ entspricht.

Weil C_S umgekehrt proportional zur Raumladungszonelänge l ist, und die Raumladungszonenlänge von der Dotierung und der angelegten Spannung abhängt variert C_S sehr stark, sieheBild 6.12, welches die Sperrschichtkapazität pro Flächeneinheit C_S/A und Weite der RLZ l als Funktion der Dotierung für eine einseitig abrupte pn-Diode in Si zeigt. Diese spannungsabhängige Kapazität findet Anwendung als Abstimmdioden und Varaktordioden für parametrische Verstärker oder bei spannungsabhängigen Oszillatoren (VCO=Voltage-Controlled Oscillators).

$$I_0 = I_{n0}(-l_p) + I_{p0}(l_n) \quad = \quad Ae\left(\frac{D_n n_p}{L_n} + \frac{D_p p_n}{L_p}\right)\left(\mathrm{e}^{U_0/U_T} - 1\right) =$$

4. Temperaturverhalten

Die Durchlaßkennlinie der pn-Diode hängt stark von der Temperatur ab. Das Sieht man am besten wenn man sich den Ausdruck für die pn-Dioden Kennlinie hinschreibt und alle Terme nach deren

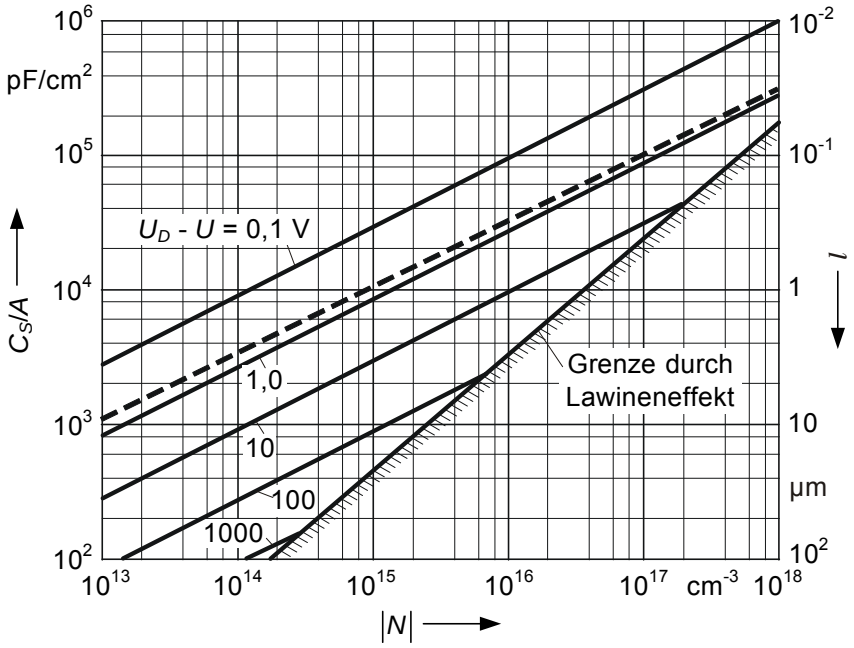


Abbildung 6.12: Sperrschichtkapazität pro Flächeneinheit C_S/A und Weite der Raumladungszone l als Funktion der Dotierung für einseitig abrupte pn-Übergänge in Si; eine Seite ist sehr stark dotiert (z.B. p^+), die andere Seite (z.B. n-Zone) sei mit der Konzentration $|N| = |N_D - N_A|$ dotiert; Parameter ist die Spannung $U_D - U$ am Übergang; die gestrichelte Gerade gilt für $U = 0$ [3].

Termperaturabhängigkeit untersucht.

$$I_0 \quad = \quad n_i^2 Ae \left(\frac{D_n}{L_n n_A} + \frac{D_p}{L_p n_D} \right) \left(\mathrm{e}^{U_0/U_T} - 1 \right) \tag{6.55}$$

Abhänigkeitn von der Temperatur findet man

- sowohl infolge der Temperaturspannung $U_T = \frac{kT}{e}$ im Exponenten (6.44),
- als auch durch die Temperaturabhängigkeit des Sättigungsstromes (6.48). $D_n = \frac{kT}{e}\mu_n$ und
- der Temperaturabhängikeit der Diffussionlängen $L_n \equiv \sqrt{D_n \tau_n} = \sqrt{\frac{kT}{e}\mu_n \, \tau_n}$.
- Die stärkste Temperaturabhängigkeit des Sättigungsstromes ergibt sich aber aus der Eigenleitungsträgerdichte

$$n_i^2 = N_L N_V \, \exp\left[-\frac{W_G}{kT} \right] \tag{6.56}$$

$$= k_1 T^3 \, \exp\left[-\frac{W_G}{kT} \right] \tag{6.57}$$

wobei gemäß Gl.(3.22) $k_1 = 4 \left(\frac{2\pi m_n k}{h^2} \right)^3$ ist.

Aus Gl.(6.48) und (6.57) folgt damit

$$\frac{1}{I_S} \frac{dI_S}{dT} = \frac{3}{T} + \frac{W_G}{kT^2} \tag{6.58}$$

$$\approx \frac{W_G}{kT^2}, \tag{6.59}$$

T = 300 K	Ge	Si	GaAs
W_G/eV	0,66	1,12	1,42
$\frac{W_G}{kT^2}/^o\mathrm{C}^{-1}$	0,09	0,15	0,19

Tabelle 6.3: Temperaturabhängigkeit des Sperrstromes

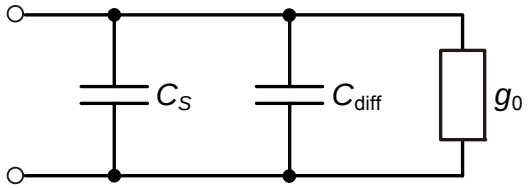


Abbildung 6.13: Kleinsignal-Ersatzschaltbild der inneren pn-Diode ohne Berücksichtigung der Bahnwiderstände [3].

da die Bandlücke das Vielfache von kT beträgt. Typische Zahlenwerte für die drei wichtigsten Halbleiter bei Raumtemperatur sind in Tabelle 6.3 zusammengestellt

Die experimentell beobachtete Temperaturabhängigkeit ist teilweise niedriger als in Tablle 6.3 angegeben, weil die aus dem einfachen eindimensionalen Modell berechneten Ströme von Leckströmen oder anderen Effekten überlagert werden, die eine geringere Temperaturabhängigkeit haben. (Bem.: Das bedeutet, dass sich in Si der Sperrstrom verdoppelt, wenn man die Temperatur um 6 K erhöht.)

5. Kleinsignal-Ersatzbild

Bild 6.13 zeigt das aus der Diskussion folgende Kleinsignalersatzschaltbild einer Diode.

6.4 Reale Dioden-Kennlinien

Das oben behandelte Shockley-Model für die pn-Diode wurde zuerst an Germanium-Dioden überprüft und ergab gute Übereinstimmung zwischen Messung und Rechnung. Bei Silizium- und noch stärker bei Galliumarsenid-pn-Dioden ergeben sich teilweise beträchtliche Abweichungen. Bild 6.14 zeigt im Vergleich die ideale und reale Diodenkennlinie in logarithmischen Maßstab. Man stellt fest, dass im Durchlassbereich bei kleinen Spannungen (a) ein höherer Strom als nach den Shockley-Model fließt und bei hohen Spannungen (c) ein kleinerer Strom. Die reale Diode folgt nur im Bereich (b) der idealen Diodenkennlinie. Im Sperrbereich (d) fließt sogar immer ein wesentlich höherer Strom. Ab einer bestimmten Spannung kommt es im Sperrbetrieb zu einem steilen Stromanstieg, man spricht von ***Durchbruch (junction breakdown)*****,** der die Diode in der Regel zerstört, falls nicht die Beschaltung der Diode den Stromfluss begrenzt. Die Ursachen dafür sind einmal Leckströme, welche infolge der hohen Feldstärke an den Rändern der Diode bei ungenügender Passivierung auftreten und die Spannungsabfälle in den Bahngebieten. Diese Größen hängen sehr stark von der Technologie ab und müssen in jedem Fall gesondert untersucht werden. Zum anderen spielen bisher vernachlässigte Effekte eine Rolle, insbesondere die Generation und Rekombination in der RLZ, die bei kleinen Spannungen (a) für die Stromerhöhung verantwortlich ist und die Hochinjektion (c), welche bei hohen Spannungen ab einem bestimmten Spannungswert erreicht wird.

6.4.1 Diskussion der Hochinjektionszone

Wir möchten die Strom-Spannungscharakteristik in der Hochinjektionszone berechnen. Dazu wird eine Hochinjektionszone (Starke Injektion von n) in einem p-Halbleiter betrachtet. Folgende Aus-

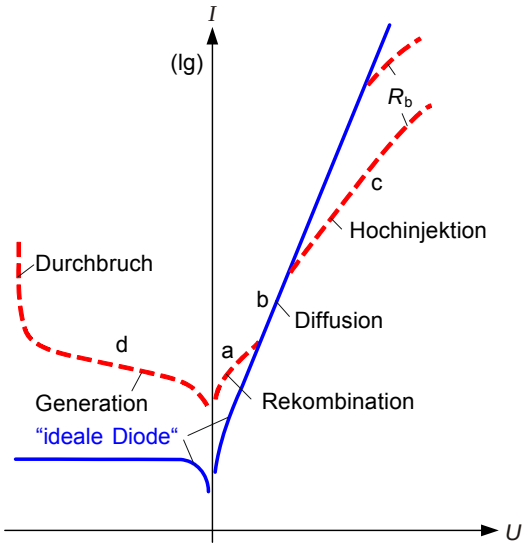


Abbildung 6.14: Die ideale Strom-Spannungskennlinie (in blau) und die reale Kennlinie (rot gestrichelt) [3].

sagen lassen sich über den HL machen:

- Störstellenerschöpfung, $n_A^- = n_A = p_p$.
- Es gilt Quasineutralität, d.h. $n \approx p$ und
- gemäß (5.29) gilt sogar $n \approx p = n_i \exp[U/(2U_T)] \gg p_p$.
- Bild 6.15 zeigt den prinzipiellen Verlauf der Trägerkonzentrationen. Das Bild zeigt anschaulich, dass in dieser Zone nicht nur die Trägerkonzentrationen sondern auch der Gradient der Trägerkonzentrationen von Elektronen n und Löchern p am Rand der Diffusionszone gleich sind (vgl. auch Gl. (5.64)): $\operatorname{grad} p = \operatorname{grad} n$. Der Gradient der Löcher führt nun aber zu einem Strombeitrag, welcher dem Diffusionsstrom entgegengesetzt ist. Aus diesem Grund ist der Anstieg des Totalstromes pro Spannungsänderung kleiner wie in der Niederinjektionszone.

Zur mathematischen Klärung des Sachverhalts suchen wir den Gesamtstrom $\vec{J}$ z.B. an der Stelle $x = 0$. Dieser ist

$$\vec{J} = \vec{J}_n + \vec{J}_p \tag{6.60}$$

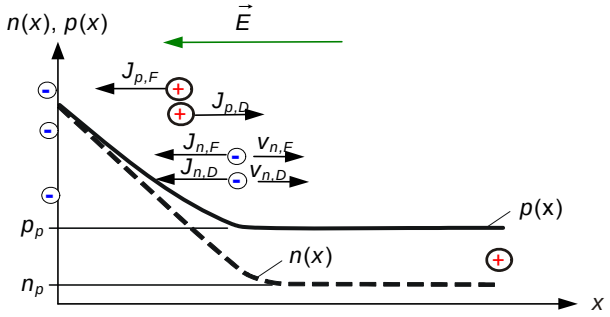


Abbildung 6.15: Trägerkonzentrationen in einer Hochinjektionszone (nicht maßstäblich). Feldstrom und Diffusionsstrom der Majoritätsträger kompensieren sich nahezu.

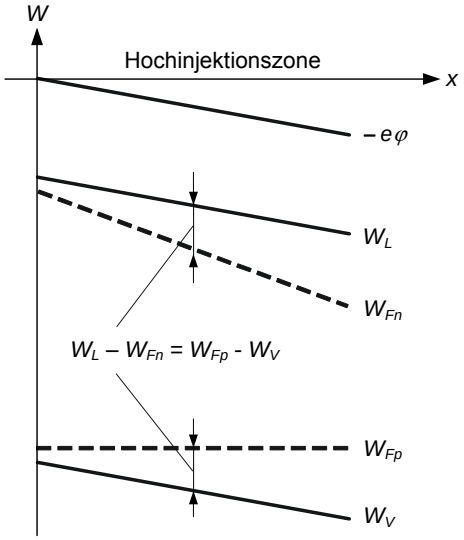


Abbildung 6.16: Bandverläufe in der Hochinjektionszone.

Für den Löcherstrom $\vec{J}_p$ setzten wir $\vec{J}_p = \vec{J}_{pF} + \vec{J}_{pD} \approx 0$. Wie wir im Kapitel zur Diffusionzone gesehen haben kompensieren sich Feldstrom und Diffusionstrom unter Niederinjektion teilweise (s. Abschn. 5.6). Unter Hochinjektion geschieht diese Kompensation in verstärktem Masse. Mehr noch, wir wählen den Hochinjektionsarbeitspunkt gerade so, dass sich die beiden Anteile vollständig kompensieren ,d.h. $\vec{J}_p \approx 0$.

Den Elektronenstrom $\vec{J}_n$ erhält man aus (5.19: $\vec{J}_n = -en\mu_n \operatorname{grad}\varphi + eD_n \operatorname{grad} n$). Dieser Ausdruck lässt sich wegen $\vec{J}_p \approx 0$ vereinfachen. Aus $\vec{J}_{pF} = -\vec{J}_{pD}$ erhält man nämlich

$$\vec{E} = -\operatorname{grad}\varphi = U_T \frac{\operatorname{grad} p}{p} = U_T \frac{\operatorname{grad} n}{n}. \tag{6.61}$$

und damit ergibt sich für $\vec{J}_n$ unter Beachtung von (6.61)

$$\begin{aligned}
\vec{J}_n &= \vec{J}_{nF} + \vec{J}_{nD} && (6.62)\\
&= en\mu_n \vec{E} + en\mu_n U_T \frac{\operatorname{grad} n}{n} && (6.63)\\
&= -en\mu_n \operatorname{grad}\varphi + en\mu_n U_T \frac{\operatorname{grad} n}{n} && (6.64)\\
&= en\mu_n U_T \frac{\operatorname{grad} n}{n} + en\mu_n U_T \frac{\operatorname{grad} n}{n} && (6.65)\\
&= 2\vec{J}_{nD} = 2e\mu_n U_T \operatorname{grad} n. && (6.66)
\end{aligned}$$

Damit ergibt sich für den Gesamtstrom mit der Hochinjektionsgleichung (5.29) $n = n_i \exp(U/2U_T)$

$$\begin{aligned}
I_0 = A\vec{J}_n \quad &\overset{Gl.(6.66)}{=} \quad 2Ae\mu_n U_T \operatorname{grad} n\\
&\overset{Gl.(5.74)}{=} \quad 2\,A\,e\mu_n U_T \tfrac{\Delta n}{L_n}\\
&\overset{Gl.(5.29)}{=} \quad 2\,A\,e\mu_n U_T \tfrac{n_i(\exp[U/(2U_T)]-1)}{L_n}\\
&\sim \quad \exp[U/(2U_T)
\end{aligned} \tag{6.67}$$

was zu zeigen war.

Ein genaueres Bild des Bandiagrammes in der Hochinjektionszone ist in Fig. 6.14 gezeichnet. Dieses Bild ergibt sich aus

- dem Bandkantenverlauf

$$\operatorname{grad} W_L = \operatorname{grad} W_V = \operatorname{grad}(-e\varphi). \tag{6.68}$$

- Weil wir keinen Löcherstrom haben, i.e. $\vec{J}_p \approx 0$, folgt

$$\operatorname{grad} W_{Fp} \approx 0 \tag{6.69}$$

- und für das Quasiferminiveau der Elektronen folgt aus (6.66) und (6.61)

$$\operatorname{grad} W_{Fn} = \frac{\vec{J}_n}{n\mu_n} = \frac{2\ \vec{J}_{n,D}}{n\mu_n} = 2\, eU_T \frac{\operatorname{grad} n}{n} = 2 \operatorname{grad}(-e\varphi). \tag{6.70}$$

- Man beachte (s. Abb. 6.16), dass wegen $n = p$ für die Hochinjektionszone auch $W_L - W_{Fn} = W_{Fp} - W_V$ folgt (wenn $N_L = N_V$).

6.4.2 Diskussion der Generations- und Rekombinationsströme in der RLZ

Bisher haben wir den Strom aufgrund von Generation- und Rekombinationsprozessen in der RLZ vernachlässigt. Diese Vereinfachung ist jedoch zu simplizistisch, wie man anhand der großen Abweichung der realen Ströme von der idealen Strom-Spannungskennlinie entnehmen kann, siehe Fig. 6.14. In diesem Abschnitt berechnen wir nun den Anteil I_{GR} des Stromes der durch die Differenz von Generations- und Rekombinationsrate $g - r \neq 0$ verursacht wird und zwar für ein Halbleitervolumen im Nichtgleichgewicht, $np \neq n_i^2$.

Wir müssen die beiden Fälle für Sperr- und Durchlassspannung gesondert betrachten. Abb. 6.17 zeigt dann auch, was sich in den beiden Fällen in der RLZ abspielt. So findet man dann, dass

- Im **Fall Sperrspannung** $np \ll n_i^2$ werden die durch $g - r > 0$ (Generationsüberschuß) erzeugten Träger aus der RLZ abgezogen, $I_n(x)$, $I_p(x) < 0$, $I_{GR} = I_p(x) + I_n(x)$.
- Im **Fall Flussrichtung** $np \gg n_i^2$ müssen die den Rekombinationsüberschuß $g - r < 0$ deckenden Träger von beiden Seiten in die Region $0 \leq x \leq d$ hineingeschoben werden, $I_n(x)$, $I_p(x) > 0$, $I_{GR} = I_n(x) + I_p(x)$.

Nun geht es also darum, für die beiden Fälle den Rekominationsstrom I_{GR} für eine Zone mit Rekombinationstermen $g-r \neq 0$ zu berechnen. Als Quellen für Rekombination in der RLZ zwischen $-l_p \leq x \leq l_n$ und dem daraus resultierenden Stromanteil beschränken wir uns auf nichtstrahlende Rekombinationen erzeugt von SRH mit Störzentren in der Bandmitte, d.h. Gl. (4.63)

$$g - r = \frac{n_i^2 - np}{(n + p + 2n_i)\tau_0} \tag{6.71}$$

Ferner gilt wegen des Massenwirkungsgesetzes $np = n_i^2 \exp(U/U_T)$. Hier setzen wir in Gl. (6.71) $n = p = n_i \exp(U/2U_T)$. Dieser Wert minimiert bei gegebenem Produkt np die Summe $n + p$ und somit $g - r$ von (6.71) und macht Gl. (6.71) deshalb zu einem Extremwert (es wird also der ungünstigste Fall mit dem größt möglichen Stromanteil aus Rekombination betrachtet).

Damit ergibt sich für die beiden Fälle:

- **Fall Sperrspannung** $U < 0$: die betrachtete Zone ist an Trägern verarmt ($np = n_i^2 \exp(U/U_T) \ll n_i^2$), $\Rightarrow$ Generation überwiegt $g - r > 0$ und in der ganzen Zone gelte $n, p \ll n_i$.
- **Fall Flussrichtung** $U > 0$: d.h, die betrachtete Zone wird von Trägern überschwemmt ($np = n_i^2 \exp(U/U_T) \gg n_i^2$) $\Rightarrow$ Rekombination überwiegt $g - r < 0$.

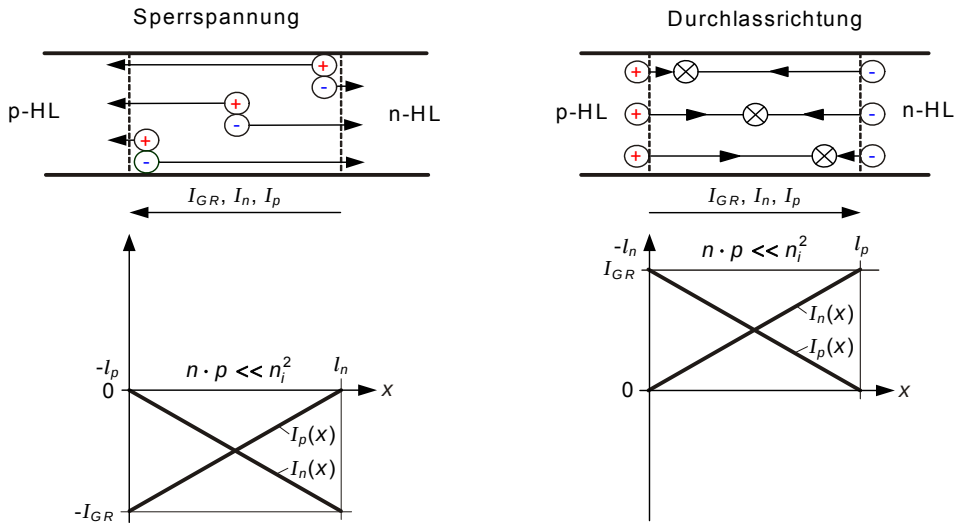


Abbildung 6.17: Generations- und rekombinationsbedingte Ströme im Fall (a) des Generationsüberschusses, g-r >0 und (b) des Rekombinationsüberschusses g-r<0.

für diese zwei Fälle folgt dann aus (6.71) mit $n = p = n_i \exp(U/2U_T)$

$$g - r \approx \begin{cases} \dfrac{n_i}{2\tau_0} & \text{für} \quad U < 0, \ |U/U_T| \gg 1, \\ -\dfrac{n_i}{2\tau_0} \exp\left(\dfrac{U}{2U_T}\right) & \text{für} \quad U > 0, \ |U/U_T| \gg 1. \end{cases} \tag{6.72}$$

Nun gilt es einen Ausdruck für den Generationsstrom herzuleiten. Wir benutzen dazu die Kontinutitätsgleichung über alle Ströme im stationären Fall (4.118)

$$\int \operatorname{div} \vec{J} \ dV = 0 \Rightarrow \oint \vec{J} \cdot d\vec{A} = \oint (\vec{J}_n + \vec{J}_p) \cdot d\vec{A} = 0, \tag{6.73}$$

und damit mit Gl. (4.113)

$$\oint \vec{J}_p \cdot d\vec{A} = \oint \vec{J}_n \cdot d\vec{A} = e \int (g - r) \, dV, \tag{6.74}$$

was eigentlich auch der Abb. 6.17 entnommen werden kann. Für die weitere Rechnung werden die Größen aus Abb. 6.17 verwendet.

Der Generations-Rekombinationsstrom I_{GR} errechnet sich nun am Einfachsten an der Stelle $x = -l_p$ oder $x = l_n$. Wir wählen $x = -l_p$ und erhalten unter Verwendung von (6.74) für den Sperrfall

$$-I_{GR} = -I_p(-l_p) = \int\limits_{x=-l_p} \vec{J}_p \cdot d\vec{A} = \oint \vec{J}_p \cdot d\vec{A} = e \int (g - r) \ dV. \tag{6.75}$$

Setzt man (6.72) in (6.75) ein und beachtet die Beziehung $L_0 = \sqrt{D_0 \tau_0}$ für die Diffusionslängen, so folgt für den Strom I_{GR}

$$I_{GR} = \begin{cases} -k_{GR} I_S & \text{für} \quad U < 0, \ |U/U_T| \ll 1, \\ k_{GR} I_S \exp(U/2U_T) & \text{für} \quad U > 0, \ |U/U_T| \gg 1, \end{cases} \qquad k_{GR} = \frac{n_{\mathrm{Dot}} l}{4 n_i L_0}, \tag{6.76}$$

wobei l die Länge der Raumladungszone und L_0 die Länger der Diffusionszone sei.

Der Strom I_{GR} soll mit dem Strom einer idealen Diode verglichen werden, für die $\tau_n = \tau_p = \tau_0$, $D_n = D_p = D_0$, $L_n = L_p = L_0$, $n_A = n_D = n_{\mathrm{Dot}}$ gilt. Aus (6.44) folgt für diese Diode

$$I = \begin{cases} -I_S & \text{für} \quad U < 0,\ |U/U_T| \ll 1, \\ I_S \exp(U/U_T) & \text{für} \quad U > 0,\ |U/U_T| \gg 1, \end{cases} \qquad I_S = \frac{2AeD_0 n_i^2}{n_{\mathrm{Dot}} L_0}. \tag{6.77}$$

Wir machen ein konkretes Beispiel und betrachten eine Diode mit $n_{\mathrm{Dot}} = 4 \cdot 10^5\, n_i$, $l/L_0 = 10^{-2}$ und erhalten somit für die Konstante $k_{GR} = 10^3$. Damit wird

- Für den **Sperrfall** $U < 0$ sieht man aus dem Vergleich von (6.77) und (6.76) dass der durch Trägergeneration in der RLZ bedingte Strom ca. 10^3-mal so groß ist wie der Sättigungsstrom I_S der idealen Diode! Da außerdem $k_{GR} \sim l \sim \sqrt{|U|}$, gibt es auch keine reine Sättigung. Im Gegenteil, der Strom steigt in Sperrichtung wie $\sqrt{|U|}$ an.
- Im **Flussfall** $U > 0$, überwiegt vorerst der Rekombinationsanteil I_{GR} den Strom der idealen Diode. Allerdings steigt der reale Diodenstrom exponentiell schneller an. Bei einer gewissen Spannung U wird der Rekombinationstrom gerade gleich groß sein wie der reale Diodenstrom. Diese Spannung ist gegeben durch

  $$I_S \exp(U/U_T) = k_{GR} I_S \exp(U/2U_T) \tag{6.78}$$

  bzw

  $$\frac{U}{U_T} = 2 \ln k_{GR}. \tag{6.79}$$

  In unserem Beispiel gilt für $U/U_T = 2 \ln k_{GR} = 2 \ln 10^3 = 13{,}8$. Folglich ist der Diodenstrom

  im Bereich $0 < U/U_T < 13,8$ proportional zu $\exp(U/2U_T)$.

  im Bereich $13,8 < U/U_T < U_{HI}/U_T = 26,5$ ist der Strom proportional zu $\exp(U/U_T)$ und

  für $U_{HI}/U_T < U/U_T$ gilt Hochinjektion gemäss Gl. (5.31) wo der Diodenstrom ebenfalls proportional zu $\exp(U/2U_T)$ ist, siehe Ende Abschn. 6.4.1.

Im vorliegenden Beispiel verhält sich die Diode gemäß $I = I_S[\exp(U/U_T) - 1]$ und ist somit nur im Bereich $13{,}8 < U/U_T < 26{,}5$ ideal.

Man beachte, dass bei Berücksichtigung von Generationen und Rekombinationen in der RLZ der Betrag des Diodenstroms sowohl im Sperrbereich $U < 0$ als auch im Durchlaßbereich $U > 0$ größer ist als der Betrag des Stroms der idealen Diode. Der Diodenstrom besteht somit im Allgemein aus zwei Anteilen: Dem Strom der idealen Diode nach (6.44), welcher die Generationen und Rekombinationen der Träger in den Diffusionszonen der Diode erfaßt, und dem Anteil I_{GR} nach (6.75), der auf die Trägergeneration und Trägerrekombination in der RLZ zurückzuführen ist (und für einen Extremfall in (6.76) abgeschätzt wurde). In der Praxis wird der Diodenstrom in der Form

$$I = I_S \left[\exp\left(\frac{U}{U_T}\right) - 1\right] + I_{GRS} \left[\exp\left(\frac{U}{mU_T}\right) - 1\right] \tag{6.80}$$

$$\Longrightarrow I = I_S + I_{GR} \tag{6.81}$$

geschrieben, wobei $I_{GRS} \gg I_S$ und $1 < m \leq 2$ gilt (m wird als ***Emissionskoeffizient (ideality factor)*** bezeichnet).

In speziellen Situationen ist zu berücksichtigen, dass in der Rekombinationsrate r außer der indirekten Rekombination über Störstellen (siehe Abschn. 4.4) noch andere Prozesse zu berücksichtigen sind, wie z. B. Auger-Prozesse.

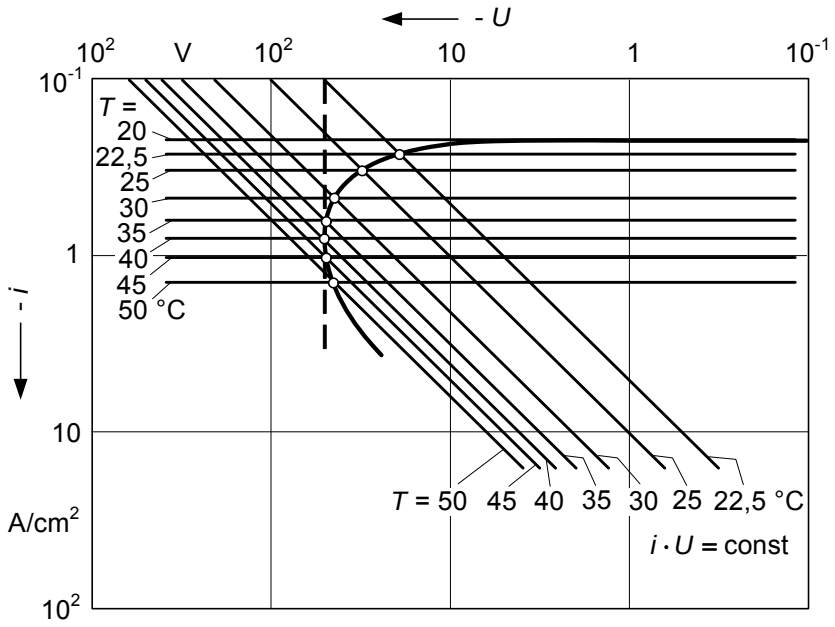


Abbildung 6.18: Thermische Instabilität im Sperrbereich von pn-Dioden [3].

6.4.3 Durchbruchsmechanismen in pn-Übergängen

In Bild 6.14 ist angedeutet, dass oberhalb bestimmter Sperrspannungen der Sperrstrom steil ansteigt. Dieser Anstieg, auch ***Durchbruch (junction breakdown)*** genannt, weil er zur Zerstörung der Diode führen kann, hat drei verschiedene Ursachen:

1. Erwärmung der Diode durch die Verlustleistung: $P_V = (-I_S)\,(U_0)$ führt zu thermischen Durchbruch.

2. Tunneln von Elektronen aus dem Valenzband der p-Seite in freie Plätze des Leitungsbandes der n-Seite (Zener-Effekt).

3. Lawinenbildung durch Stoßionisation in der RLZ.

Die Durchbruchmechanismus können gezielt für den Bau von verschiedenen Dioden eingesetzt werden. So wird der thermische Durchbruch in ***temperaturabhängigen Dioden*** zum Schutz von thermischer Überlastung eingesetzt. Temperaturabhängige Dioden werden z.B. in Laptop Computern zum Schutz vor Überhitzung eingesetzt. Die Dioden sind unpräzise aber zweckmässiger und billig. Der Zener-Effekt sowie der Lawinendurchbrucheffekt wird in ***Z-Dioden*** ausgenutzt um Spannungen zu stabilisieren oder Spannungen zu begrenzen. Der Lawineneffekt wird in einzelen Lawinendioden ***(avalanche diodes)*** auch gezielt zur Verstärker-Multiplikation von Ladungsträgern eingesetzt. DerTunneleffekt wird in sogenannten ***Tunneldioden*** zur Erzeugung eines negativen differentiellen Widerstande eingesetzt (siehe nächstes Kapitel).

Im folgenden werden die Durchbruchmechanismen diskutiert. Die Bauteile selber werden im nächsten Kapitel diskutiert.

Thermischer Durchbruch

Die Verlustleistung, die im wesentlichen in der RLZ erzeugt wird, wird in Wärme umgewandelt. Je nach Wärmeleitung und äußerer Kühlung stellt sich eine bestimmte Temperatur der RLZ ein. In Bild 6.18 ist die bereits oben diskutierte Abhängigkeit des Sperrstromes mit der Temperatur

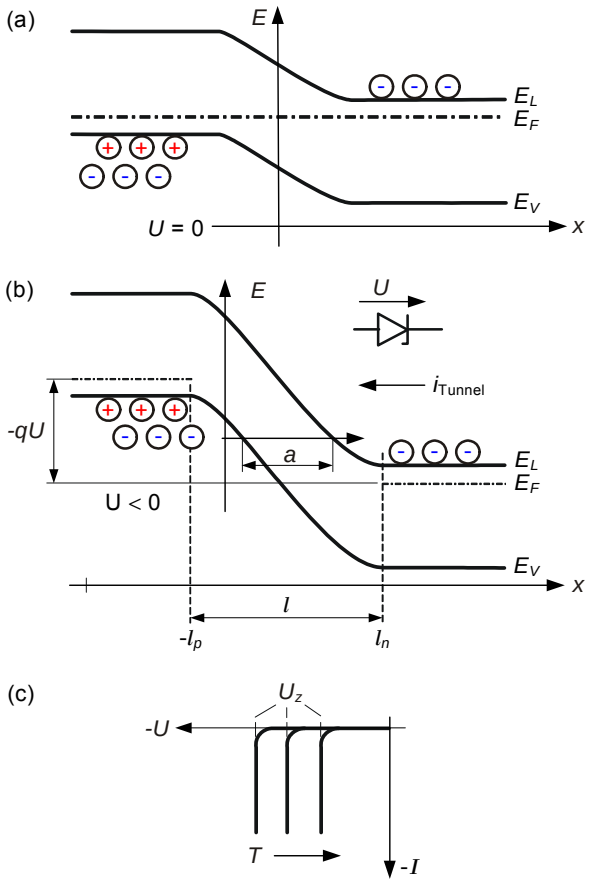


Abbildung 6.19: Zener-Effekt: (a) Bandverläufe bei $U = 0$; (b) Bandverläufe bei $U < 0$; bei genügend hoher Sperrspannung können die Ladungsträger auf freie Plätze in das entsprechende Nachbargebiet tunneln; (c) Temperaturabhängigkeit der Sperrkennline und Zener-Spannung.

als Parameter aufgetragen. Ebenfalls eingetragen sind Leistungshyperbeln (UI=konst.) Für einen bestimmten Wärmewiderstand zwischen pn-Schicht und Wärmesenke, entspricht jede Leistung UI einer bestimmten Temperatur, so dass die Leistungshyperbeln Temperaturwerte als Parameter haben. Kennlinienpunkte ergeben sich nun als Schnittpunkte zwischen Leistungshyperbeln und Sperrkennlinien gleicher Temperatur. Wie Bild 6.18 zeigt, ergibt sich ab Erreichen einer gewissen Spannung ein negativer Kennlinienast, d.h. der Strom erhöht sich bei gleicher oder sogar abnehmender Spannung, was zur Zerstörung der Diode führt, wenn der Strom durch die äußere Beschaltung nicht stabilisiert oder begrenzt wird.

Zener-Effekt

Unter dem ***Zener-Effekt*** versteht man die Ladungsträgererzeugung im starken elektrischen Feld. Wie Bild 6.19 zeigt, liegen in einer Raumladungszone Valenzelektronen auf gleicher energetischer Höhe wie frei Plätze im räumlich etwas entfernten Leitungsband. Ist dieser Abstand genügend klein, d.h. die Feldstärke genügend groß, so können die Valenzelektronen durch den verbotenen Bereich tunneln. Da die Tunnelwahrscheinlichkeit mit abnehmender Barrieren-Breite stark, d.h. exponentiell, zunimmt, ergibt sich ein mit der Diodenspannung sehr stark zunehmender Strom, wie in Bild 6.19(c) gezeichnet. In Ge und Si bei hochdotierten Dioden mit n_D, $n_A > 10^7\, n_i$, bei denen die Sperrschichtweite typisch $0{,}1\,\mu$m beträgt und die Durchbruchspannungen $< 5\,$V sind, ist das für den Zener-Effekt erforderliche Feld etwa $100V\ \mu\mathrm{m}^{-1}$. Da W_G mit steigender

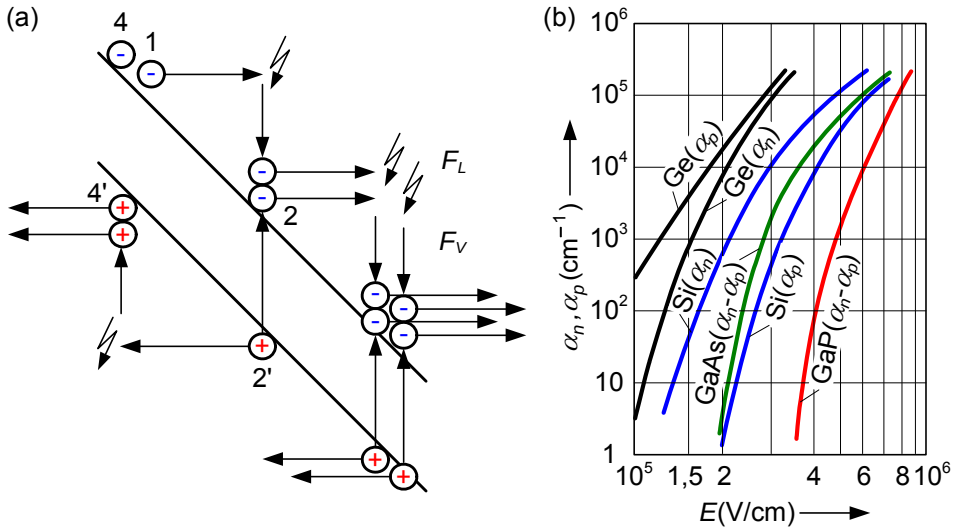


Abbildung 6.20: (a) Prinzip des Lawineneffekts (b) Gemessene Ionisationskoeffizienten für Lawinenmultiplikation als Funktion der Feldstärke in Ge, Si, GaAs und GaP [3].

Temperatur sinkt, nimmt der Betrag der Durchbruchspannung mit steigender Temperatur ab. Zenerdurchbruch erfolgt typisch bei Spannungen $|eU| \leq 4\,W_G$. Dieser Durchbruchmechanismus führt zu keiner Zerstörung der Diode.

Lawinendurchbruch und Lawinenmultiplikation

Der ***Lawineneffekt (avalanche multiplication)*** ist der in den häufigste Durchbruchsmechanismus. Er begrenzt in den meisten Transistoren die zulässige Kollektorspannung. Die Lawinenmultiplikation setzt dann ein, wenn Träger die im Mittel zwischen zwei Stößen aus dem Feld aufgenommene Energie nicht mehr beim nächsten Stoß an das Gitter abgeben kann. Die akkumulierte Energie kann zur Erzeugung eines Elektron-Loch-Paares durch Aufreißen einer Valenzbindung führen. Da der die Trägergeneration verursachende Ladungsträger lediglich seine Energie abgibt, aber nicht verschwindet, ergibt sich ein lawinenartiges Anwachsen der Ladungsträgeranzahl. Unter Wahrung von Energie- und Impulsbilanz muß der ionisierende Träger je nach der Bandstruktur eine Energie W im Bereich $W_G \leq W \leq 3W_G/2$ abgeben.

Ein primäres Elektron (Loch) erzeugt dabei auf der Laufstrecke dx im Mittel $\alpha_n\,dx$ ($\alpha_p\,dx$) Trägerpaare; α_n, α_p heißen Ionisierungskoeffizienten. Sie sind stark feldstärkeabhängig (mit der Tendenz $\alpha_n \approx \alpha_p$ bei sehr hohen Feldstärken) und haben je nach Material und Trägersorte bei Feldstärken um $30\,\mathrm{V}\mu\mathrm{m}^{-1}$ Werte im Bereich α_n, $\alpha_p = 0{,}1 \ldots 10\,\mu\mathrm{m}^{-1}$, Bild 6.20. Bei diesen Feldstärken laufen alle Träger bereits mit ihren Sättigungsgeschwindigkeiten (diese werden in Si schon bei Feldstärken von $2\,\mathrm{V}\mu\mathrm{m}^{-1}$ erreicht). Der Lawinendurchbruch erfolgt bei kleineren Feldstärken als der Zenerdurchbruch, aber eben nicht in hochdotierten Dioden, bei denen wegen der dünnen Sperrschichtweite d die Werte $\alpha_n d$, $\alpha_p d$ zu klein sind. Schwach dotierte Dioden mit breiter Sperrschicht und Sperrspannungen $|eU| \geq 6\,W_G$ zeigen Lawinendurchbrüche. Da bei steigender Temperatur die Energieabfuhr der Träger an das Gitter durch vermehrte Stöße besser erfolgt, können die Träger die zur Stoßionisierung nötige Energie erst nach längerer Laufstrecke akkumulieren und somit steigt der Betrag der Durchbruchspannung mit steigender Temperatur.

Si-Dioden mit Durchbruchspannungen um 6 V zeigen eine Mischung von Zener- und Lawinendurchbruch mit einer nahezu temperatur-unabhängigen Durchbruchspannung. Sie werden zur Spannungsstabilisierung verwendet (Z-Dioden).

Zur Berechnung des Lawineneffektes, siehe Bild 6.21, wird der Einfachheit halber $\alpha_n = \alpha_p = \alpha$ und $v_n = v_p = v$ (Sättigungsgeschwindigkeit) vorausgesetzt.

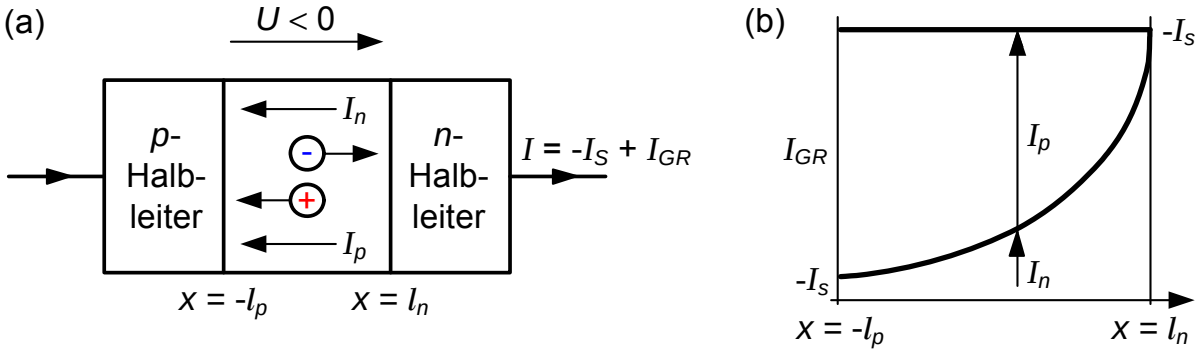


Abbildung 6.21: (a) Lawinenmultiplikation in der Raumladungszone einer gesperrten Diode (b) Stromverlauf innerhalb der Lawinendiode bei gleichen Ionisationskoeffizienten für Löcher und Elektronen

Wegen der Trägermultiplikation ist $g - r \approx g$ und

$$g = [-\alpha_n I_n - \alpha_p I_p] \frac{1}{eA} = \alpha \left(I_n + I_p\right) \frac{1}{eA} = \frac{\alpha I}{eA}\,, \tag{6.82}$$

wobei A die Querschnittsfläche. Die Löcher- und Elektronenströme $\vec{J}_p = \vec{J}_{pF}$, $\vec{J}_n = \vec{J}_{nF}$ addieren sich zum Gesamtstrom $I = I_n + I_p$ .Die Generationsrate ist also proportional zum Gesamstrom.

Der Gesamtstrom I ergibt sich auch als Summe des Sättigungsstromes I_S der Diode ohne Lawinenmultiplikation und I_{GR} des durch die Trägermultiplikation zusätzlich erzeugten Stromes

$$I = -I_s + I_{GR}\,. \tag{6.83}$$

Der durch die Trägergeneration zusätzlich erzeugte Stroms I_{GR} kann wieder gemäß Gl. (6.75) berechnet werden. D.h.

$$I_{GR} \overset{Gl.(6.75)}{=} -eA \int_{-l_p}^{l_n} (g - r) dx \overset{Gl.(6.82)}{=} I \int_{-l_p}^{l_n} \alpha \; dx \tag{6.84}$$

Die Größe welche uns letztlich interessiert ist M_0, der sogenannte ***Multiplikationsfaktor*** oder ***Lawinenverstärkungsfaktor (avalanche multiplication factor)*** , welcher definiert ist durch

$$I = -M_0 I_S\ . \tag{6.85}$$

Auflösen von Gl. (6.83) - (6.85) ergibt für M_0

$$M_0 = \frac{1}{1 - \int\limits_{-l_p}^{l_n} \alpha \, \mathrm{d}x}. \tag{6.86}$$

In der Praxis wird der Multiplikationsfaktor M_0 im interessierenden Bereich angenähert mit

$$M_0 = \frac{1}{1 - \left(\dfrac{U}{U_{BR}}\right)^n}\ . \tag{6.87}$$

U_{BR} wird als ***Durchbruchspannung (breakdown voltage)*** bezeichnet. Der Exponent n hat bei Si-Dioden Werte im Bereich $n = 1{,}5 \ldots 4$.

Der Lawinendurchbrucheffekt wird auch gezielt zur Verstärkung kleiner Primärströme ausgenutzt (etwa bei der Lawinenphotodiode oder APD = avalanche photo diode, bei der durch Photonenabsorption in der RLZ primär erzeugte Trägerpaare mittels Lawineneffekt vervielfacht werden).

6.5 Schaltverhalten von Dioden

Bild 6.22 zeigt eine Schaltung, in der das Schaltverhalten von Dioden untersucht werden kann. Ein Widerstand R ist vorgeschaltet. U ist der Spannungsabfall über der Diode mit Bahnwiderstand $r_{b.}$

Stationäre Arbeitspunkte 1 und 2:

Im stationären Fall gilt gemäss dem 2. Kirchhoff'schem Gesetz

$$U_{Batt\,1/2} = U(I) + RI\,. \tag{6.88}$$

Graphisch lässt sich die Gleichung lösen, indem man schreibt

$$U_{Batt\,1/2} - RI = U(I)\,, \tag{6.89}$$

wobei die linke Seite die Widerstands-Gerade "1" für den Schalter in Stellung 1 bzw die Widerstands-Gerade "2" in der Schaltstellung 2 ergibt. Die rechte Seite ist durch die Diodenkennlinie beschrieben. Die Arbeitspunkte 1 und 2 ergeben sich dann in den Schnittpunkten von den Geraden mit der Diodenkennlinie, siehe Abb. 6.23.

Einschaltverhalten, $0 \rightarrow 1$:

Beim Schalten von 0 auf U_{Batt1} (zur Zeit t_1) wirkt die Sperrschichtkapazität wie ein Kurzschluss. Der Strom steigt auf den Wert $I_0 = U_{Batt1}/R$ an (Fig. 6.24). Die Spannung über der Diode steigt auf den durch den Dioden-Bahnwiderstand r_b limitierten Wert $U = r_b I_0$ an. Da der Bahnwiderstand aber sehr klein ist (typisch $<1\Omega$), wird auch U sehr klein sein.

Jetzt muss sich eine Diffusionszone aufbauen. Ist der Widerstand R groß und der Spannungsabfall an der Diode klein, so ist der Strom I im wesentlichen konstant und deshalb ist auch der Anstieg der Überschußladungsträgerdichte in der Diffusionszone konstant. Der Strom tendiert allmählich gegen den Wert I_1 und die Spannung gegen den Wert U_1 (Arbeitspunkt 1), Bild 6.25.

Wird die Diode mit einer Stromquelle, welche den Wert I_1 hat, anstelle einer Spannung

gespeist, nimmt der Strom sofort den Wert I_1 an. Die Spannung ist vorerst limitiert durch den Bahnwiderstand. Sie nimmt aber innerhalb der Minoritätsträgerlebensdauer, was bei einer p^+n-Diode $\tau_1 = \tau_p$ entspricht, ihren stationären Wert an.

Ausschaltvorgang $1 \rightarrow 0$:

Wird zum Zeitpunkt t_2 der Schalter geöffnet, so muß der äußere Strom Null sein. Die in der Diffusionszone gespeicherten Minoritätsträger verschwinden durch Rekombination innerhalb der Minoritätsträgerlebensdauer τ_p.

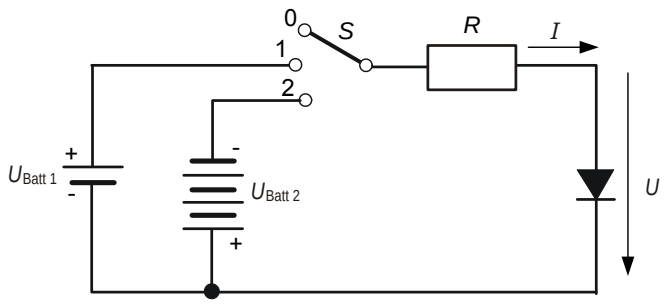


Abbildung 6.22: Anordnung zur Untersuchung des Schaltverhaltens von Dioden. Die Blindkomponenten der Last R müssen vernachlässigbar sein [3].

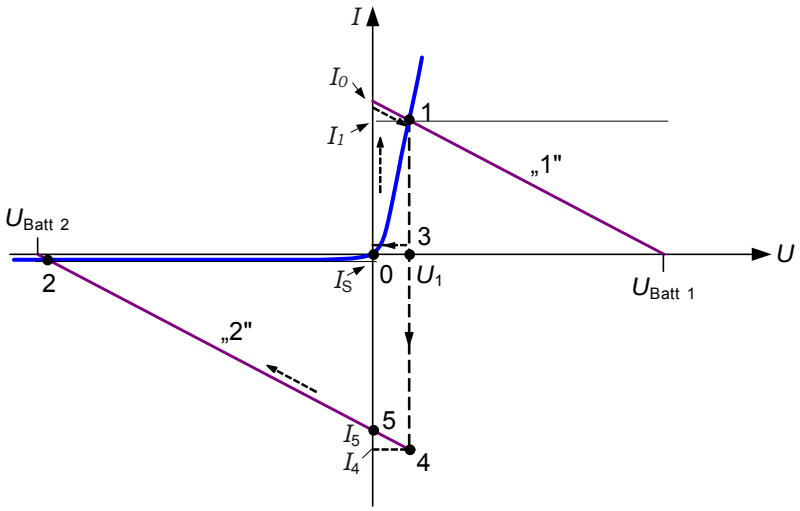


Abbildung 6.23: Widerstandsgeraden im I-U-Kennlinienfeld einer Diode für die Schaltvorgänge in Bild 6.22 [3].

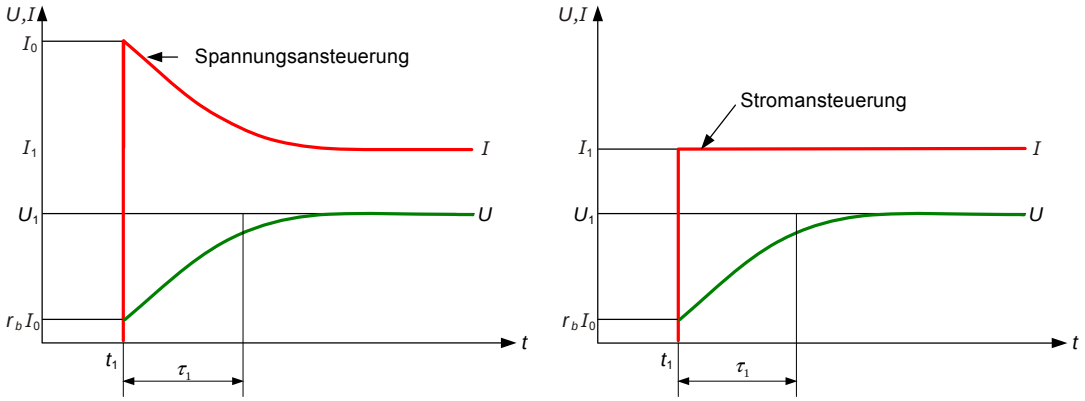


Abbildung 6.24: Einschaltverhalten einer Diode [3].

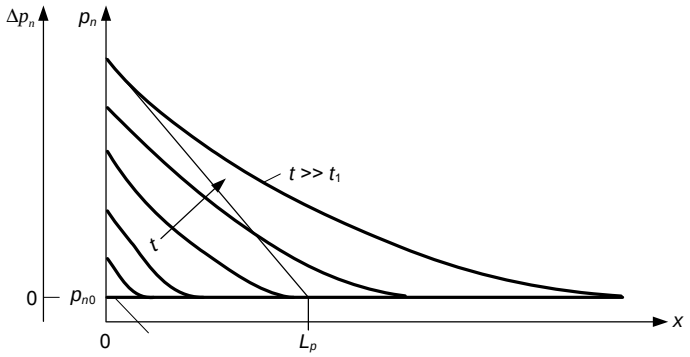


Abbildung 6.25: Minoritätsträgerverteilung im n-Bereich einer p^+n-Diode während des Einschaltvorganges [3].

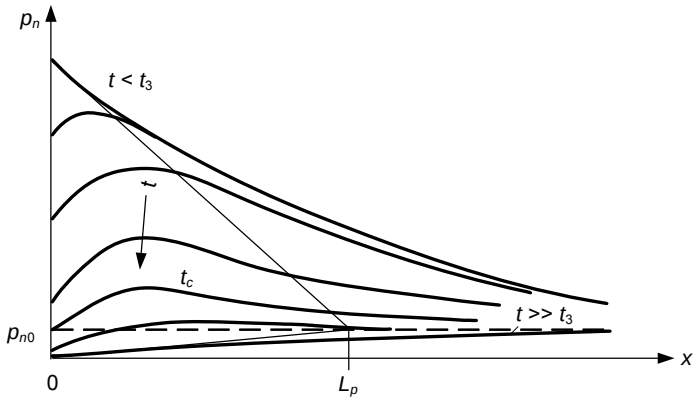


Abbildung 6.26: Minoritätsträgerverteilung im n-Bereich einer p^+n-Diode während des Umschaltvorganges.

Umschaltvorgang $1 \rightarrow 2$:

In diesem Fall wird die Diode vom Durchlaß- in den Sperrbereich geschaltet. Nach Umschalten des Schalters von 1 nach 2 zur Zeit t_3 springt der Diodenstrom vom Wert I_1 auf den Wert I_4 auf der Widerstandsgeraden 2, Bild 6.24. Die Spannung U über der Diode bleibt im ersten Moment unverändert, da sich die Minoritätsträgerladung nicht schlagartig ändern kann.

Bild 6.26 zeigt die Minoritätsträgerverteilung während des Umschaltens.

Solange die Minoritätsträgerdichte am Rand der RLZ über dem Gleichgewichtswert p_{n0} liegt, ist $U > 0$, Bild 6.27. Da die Ladungsträger nicht durch Rekombination verschwinden müssen, sondern durch einen Rückstrom I_R abgesaugt werden, ist die Speicherzeit τ_3 wesentlich kürzer als die Minoritätsträgerlebensdauer.

Bild 6.28 zeigt die normierte Speicherzeit τ_3 und die Zeitdauer τ_4 nach welcher der Rückstrom auf ein Zehntel des Anfangswertes zurück gegangen ist.

6.6 pin-Diode

pin-Dioden besitzen das in Bild 6.29 gezeichnete Dotierungsprofil.

- Bei Anlegen einer Sperrspannung, wird die i-Zone (wegen des großen el. Feldes über der ganzen i-Zone) sehr schnell ausgeräumt, wobei die Weite der RLZ im wesentlich unabhängig von der Spannung ist.
  → infolge der langen i-Zone und der niedrigen Ladungsträgerdichte in der i-Zone ist die pin-Diode in Sperrrichtung sehr hochohmig und sperrt gut.
  → mehr noch, je länger die i-Zone, desto kleiner das elektrische Feld für eine gegebene angelegte Spannung ($U = -\int E\, dx$). Damit kann die pin-Diode selbst bei hohen Sperrspannungen eingesetzt werden. Sie hat also große Durchbruchspannungen.
  → wegen der langen RLZ ist die Kapazität klein, siehe Gl. (7.1), und fast unabhängig von der angelegten Spannung (die Ausdehnung der RLZ ändert sich mit der Spannung nur noch minimal).
  → infolge des über die ganze i-Zone laufenden großen el. Feldes werden bei Sperrspannung Ladungsträger schnell aus der Zone transportiert und die pin-Diode bietet deshalb ´relativ schnelle ´Schaltgeschwindigkeiten

- In Flusspolung werden die Ladungsträger in die i-Zone injiziert, was den Kleinsignalleitwert der Diode stark erhöht, Bild 6.30.
  → wegen der starken Ladungsträgerinjektion ändert sich der Widerstand mit dem angelegten

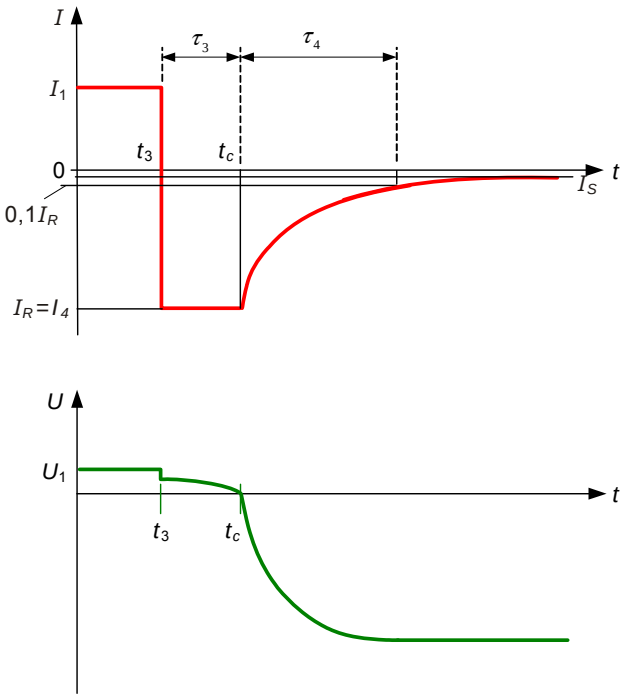


Abbildung 6.27: Umschaltverhalten einer Diode.

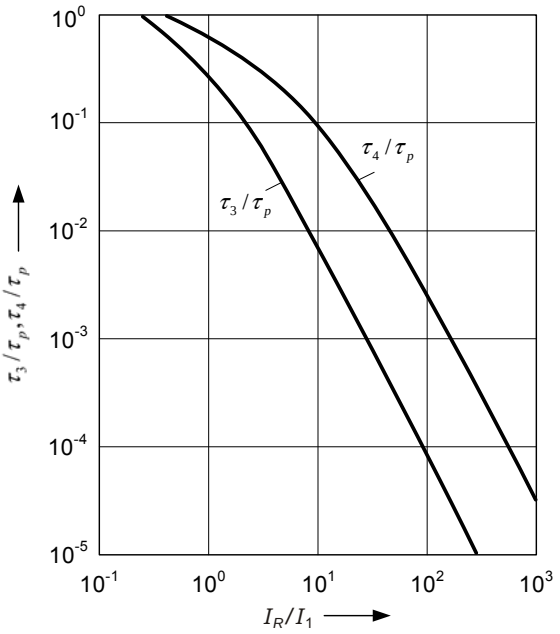


Abbildung 6.28: Normierte Speicherzeit als Funktion des normierten Rückstromes [3].

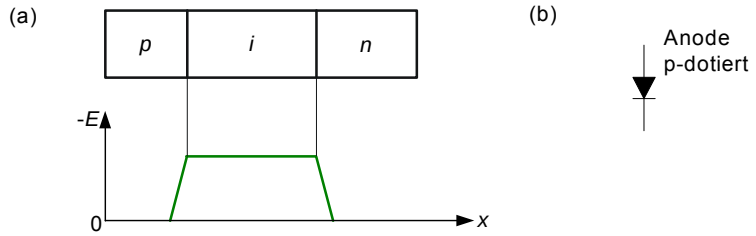


Abbildung 6.29: (a) Dotierungsverlauf und elektrische Feldstärke in einer pin-Diode bei Sperrpolung [3]. (b) Schaltsymbol der pn oder pin-Diode

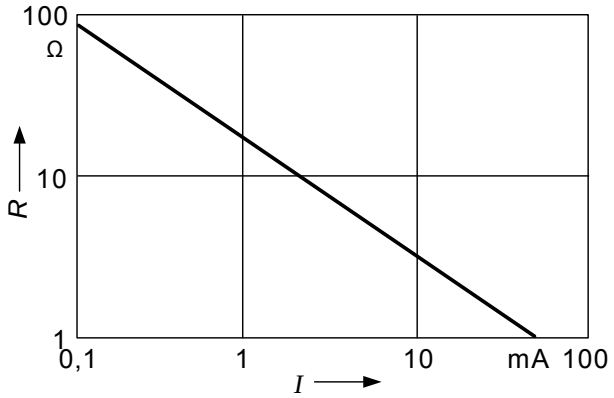


Abbildung 6.30: Typischer Kleinsignalwiderstand einer in Flußrichtung gepolten pin-Diode als Funktion des Diodenstromes.

Strom
→ bei einer Modulation ändert sich die Konzentration der Ladungsträger dann nur sehr langsam, weil (a) der Weg zum Abtransport der Ladungen aus der i-Zone lang ist und (b) da die Lebensdauer der Ladungsträger in der i-Zone grundsätzlich lang ist (ca. $\tau_{\min}$ =0.1-5 μs wegen fehlender Störstellenrekombination). Aus diesem Grund ändert sich die Zahl der Ladungsträger in der i-Zone für Pulse mit Pulsdauern $t \ll \tau_{\min}$ nicht. Das gilt selbst für kurze Pulse in Sperrrichtung, da die Diode erst nach Rekombination aller Ladungsträger in den Sperrbetrieb gelangt. Gibt man also ein Hochfrequentes Signal der Frequenz f >> 1 auf die Diode dann verhält sie sich wie ein ohmscher Widerstand dessen Wert sich durch die ¨Uberlagerung eines konstanten Stromes sehr variabel einstellen lässt.

Die Unterschiede zwischen einer pn- und einer pin-Diode können wie folgt zusammengefasst werden:

Sperr-/Fluss-	**pn-Diode**	**pin-Diode**
U<0	RLZ-ändert sich mit Spannung RLZ klein → C_S groß RLZ klein →Durchbruchspannung klein Widerstand in RLZ moderat	RLZ dehnt sich kaum aus → C_S ist konstant RLZ lang → C_S klein RLZ lang →Durchbruchspannung groß Widerstand in RLZ groß
U>0		Widerstand der Diode ändert sich mit Strom nur l

pin-Dioden werden genutzt als

- *schnelle Schalter* (geeignet wegen der kleinen Kapazität, der guten Sperrqualität, der großen Sperrspannung und wegen des schnellen Ladungsträger Transportes im el. Feld in der i-Zone). Die Schaltzeit liegt in der Größenordnung der Ladungsträger-Laufzeit durch die i-Zone.

- *variable, gleichstromgesteuerte Dämpfungsglieder* (im Arbeitsbereich um U>0, wie in Fig. 6.30 gezeigt). In der Mirkowellentechnik (für Frequenzen $f \gg \tau_{\min}^{-1}$ kann man dem hochfrequenten Wechselstrom einen Gleichstrom überlagen. Der Arbeitspunkt des Gleichstromes steuert dann die mittlere Zahl der Ladungsträger in der i-Zone und damit den Widerstand)
- In modifzierter Form als *Leistungsgleichrichter* (s. weiter unten)
- In direkten Halbleitermaterialien als *Halbleiter-Photodioden, Solarzellen* und *Laser* (s. weiter unten)

Kapitel 7

Spezielle pn-Dioden

7.1 Durchbruchdioden

Die Schaltsymbole der temperaturabhängigen Diode und der Zener Diode sind in Bild 7.1 aufgeführt.

7.1.1 Temperaturabhängige Diode

Temperaturabhängige Dioden basieren auf dem thermischen Durchbruch, welcher eintritt, wenn die Verlustleistung in der RLZ bei einer spzifizierten Temperatur einen gewissen Wert übersteigt. Die Theorie dazu wurde bereits im letzten Kapitel diskutiert.Thermische Durchbruchdioden sind nicht besonders exakt - dafür um so billiger. Thermische Durchbruchdioden werden heute kaum mehr verwendet. Stattdessen werden die Temperaturen meist über Thermistoren (ein Widerstand dessen Widerstandswert stark von der Temperatur abhängt) bestimmt.

7.1.2 Zener Diode

Die Zener Diode ist eine pn-Diode mit einer exakt spezifizierten Durchbruchspannung U_{BR}.

Die meisten Zener Dioden bestehen aus zwei hoch dotiertem n- und p-Zonen aus Silizium (die starke Dotierung garantiert eine starke Krümmung der Bänder). Für Durchbruchspannungen oberhalb von $6W_g/q$ ($\simeq$ 7V in Si) basiert der Durchbruchmechanismus meist auf dem Lawinendurchbruch-Effekt. Für Durchbruchspannungen unterhalb von $4W_g/q$ ($\simeq$ 5V in Si) basiert der Durchbruchmechanismus vor allem auf dem Zener-Effekt. Dazwischen basiert er sowohl auf dem einen als auch dem andern Effekt. Man beachte, dass die Temperaturabhängigkeit beim Lawinen- und Zener-Effekt genau umgekehrt sind. Mit steigender Temperatur sinkt nämlich die Beweglichkeit der Ladungsträger und der Lawinen-Effekt setzt später ein. Mit steigender Temperatur sinkt aber die Bandlücke und er Zener-Effekt setzt früher ein. Bild 7.2 zeigt Durchbruchskennlinien. Man beachte den im Vergleich zum Lawineneffekt weich einsetzenden Durchbruch beim Tunneleffekt (=Zenereffekt).

Abbildung 7.1: (a) Temperaturabhängige Diode, (b) Zener-Diode oder Z-Diode. basierend auf dem Zener-Effekt als auch auf dem Lawinendurchbrucheffekt.

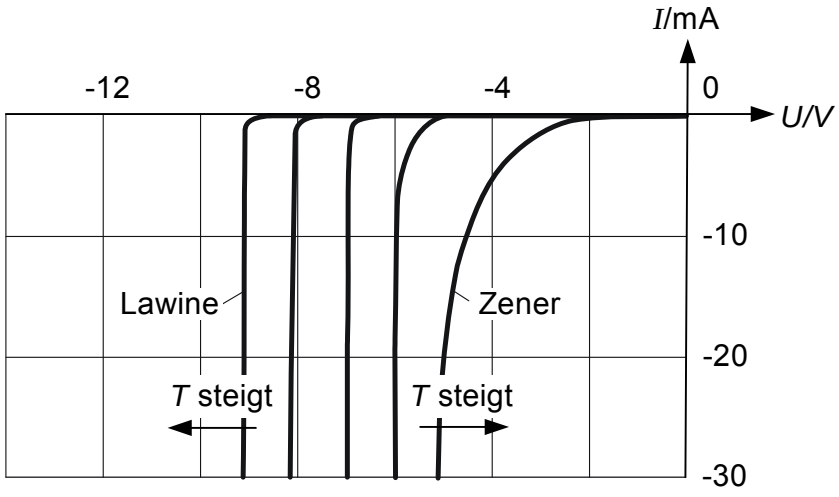


Abbildung 7.2: Temperaturabhängigkeit von Lawinen- und Zenerduchbruch.

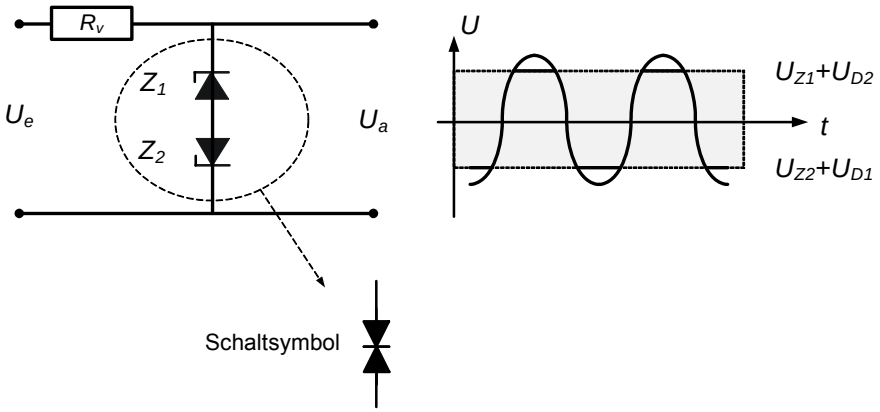


Abbildung 7.3: Transorb: Die serielle Verschaltung von 2 Zenerdioden begrenzt die Spannung in jede Richtung auf den Wert der Zener-Durchbruchspannung der einen Diode und sowie der Diffusionsspannung der andern Diode. Das Schaltsymbol für den Transorb ist in der untern Hälfte des Bildes angegeben.

7.1.3 Transorb (Suppresor Dioden)

Speziell optimierte Dioden in Form von zwei in Serie geschalteten Zenerdioden.

Bezeichung: „Transient Voltage Suppression“ oder TVS- Dioden „Transorb“ Dioden genannt, in Anlehnung an die Firma TransZorb, welche diese als erstes registriert hat.

Suppressor-Bauteile werden zum Schutz der Ein- und Ausgänge elektronischer Schaltungen vor kurzzeitigen Überspannungsimpulsen (Spannungstransienten, SurgeProtector), wie sie durch Schaltvorgänge im Netz oder nahe Blitzschläge auftreten, gebraucht. Sie reagieren schneller als Varistoren.

In Bild 7.3 sind die zwei in Serie geschalteten Zenderioden und deren Arbeitsweise dargestellt. In der seriellen Schaltung fließt bei der positiven Halbwelle der Strom über R_V durch die Z-Diode Z1 (diese begrenzt die Z-Spannung) und durch die Z-Diode Z2 (diese begrenzt die Spannung in Flussrichtung). Es addieren sich diese beiden Spannungswerte Die Spannung wird damit auf $U_{Z1} + U_{D2}$ begrenzt. Bei der negativen Halbwelle arbeitet Z1 im Fluss- und Z2 im Zenerspannungsbetrieb.

Suppressor Dioden werden zum Überspannungsschutz in Wechselspannungskreisen von Signal-

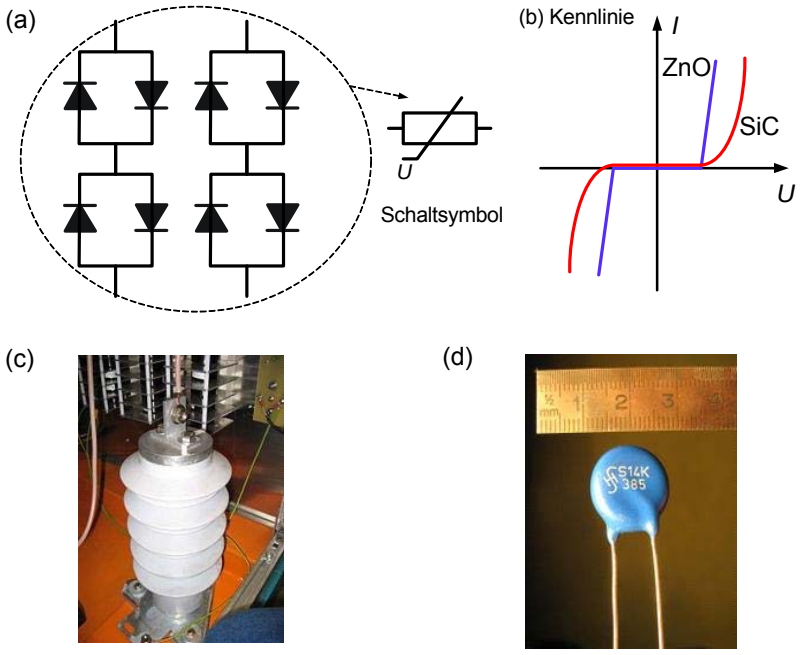


Abbildung 7.4: Varistoren: (a) Internes Schaltkonzept eines Varistors. Viele ungeordent (oder je nach Typ auch ausgerichtet) angeordnete Dioden sind parallel geschaltet. Die Summe aller Diffusionspannungen der Dioden führt zu einer Sperrkennlinie wie sie in (b) abgebildet ist. Die Sperrspannung kann mit der Anzahl der Dioden stark nach oben getrieben werden. (c) Hochspannungsvaristor, (d) Varistor in Scheibenform mit einer Schwellenspannung von 385 V. (Quelle für Bilder: Wikipedia)

leitungen eingesetzt. Sie reagieren innerhalb von Nanosekunden, weisen eine nur kleine Kapazität auf (Sperrschichtkapazität) und vertragen nur kleinere Energien.Alterungseffekte sind allerdings nicht zu erwarten. Suppressordioden sperren nach einem Überspannungsereignis wieder - wenn sie nicht vernichtet wurden.

7.1.4 Varistor

Varistor (=Variable Resistor)= VDR (=Voltage Dependent Resistor) sind Bauteile mit einem spannungsabhängigen Widerstand. Oberhalb einer bestimmten Schwellspannung, die typisch für den jeweiligen Varistor ist, wird der Widerstand abrupt kleiner. Im Normalbetrieb ist ihr Widerstand sehr groß, während bei Überspannung der Widerstand fast verzögerungsfrei sehr sehr klein wird und Ladung ableitet.

Varistoren bestehen je nach Typ aus vielen ungeordnet oder ausgerichteten Körnern bzw. einzelnen Dioden, welche in ihrer Gesamtheit viele parallel und in Reihe gepolte Dioden bilden, siehe Bild 7.4(a). Diese haben dann die nichtlineare Diodenkennlinie von vielen in Flussrichtung gepolten Dioden (die in Sperrrichtung angeordneten Dioden sperren und tragen nicht zur Kennlinienfunktion bei). Die Strom-Spannungskennlinie kann durch die Formel $I = (U/U_{BR})^n \cdot 1[A]$ approximiert werden.

Es gibt zwei verschiedene Varistor-Typen. Zum Einen kann es sich Varistoren aus Halbleitermaterialien, d.h. Körnern mit pn-Dioden Charakteristik handeln. So wurden früher z.B. SiC-Halbleiter Varistoren sehr häufig eingesetzt.. Zum Andern kann es sich um Varistoren aus Metall-Halbleiter Materialien, d.h. Dioden mit Schottky-Charakter handeln. Wichtig sind z.B. Metall-ZnO Varistoren. Die Kennlinie der beiden Varistortypen ist verschieden, siehe Fig. 7.4(b). Der Metall-Halbleitertyp liefert steilere Kennlinien (das ist typisch für Schottky-Dioden und wird später klar werden).

Varistoren werden zum Überspannungsschutz von Leitungen mit großen Strömen eingesetzt.

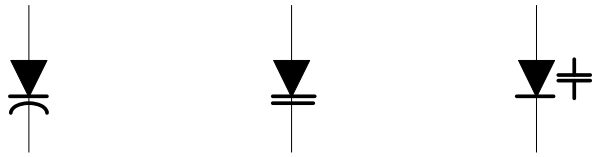


Abbildung 7.5: Schaltzeichen der Varaktor-Diode.

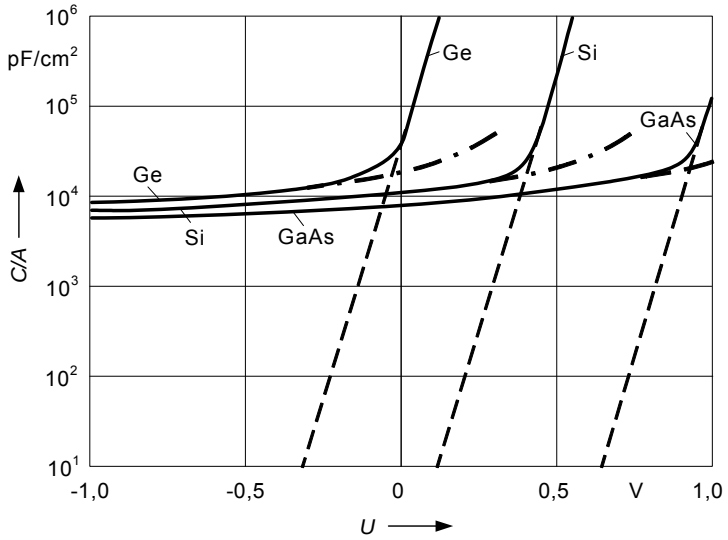


Abbildung 7.6: Sperrschichtkapazität (– · – · –·), Diffusionskapazität (- - - -) und Gesamtkapazität (——) je Flächeneinheit einer p^+n-Diode als Funktion der Diodenspannung U für Zimmertemperatur, berechnet für: Ge: $N_D = 10^{15}cm^{-3}, N_A = 10^{18}cm^{-3}$, $\tau_p = 10^{-3}s$; Si: $N_D = 10^{15}cm^{-3}, N_A = 10^{18}cm^{-3}$, $\tau_p = 10^{-5}s$; GaAs: $N_D = 10^{15}cm^{-3}, N_A = 10^{18}cm^{-3}$, $\tau_p = 10^{-8}s$ [3].

Sie haben eine Ansprechzeit von ca. 100 ns - sind also nicht so schnell wie Transorbs und weisen größere Kapazitäten auf (arbeiten ja auch in Flussrichtung wo die größeren Diffusionskapziäten dominieren). Dank der vielen parallel angeordneten Dioden können auch gr. Ströme fließen ohne dass der Varistor zerstört wird. ZnO-Varistoren stehen im Ruf zu altern (d.h. die Schwellschspannung wird mit der Zahl der Überspannungen geringer). Varistoren sperren nach einem Überspannungsereignis wieder - wenn sie denn nicht zerstört wurden. Sie werden denn auch oft als Feinschutz von Steckdosen verwendet. Das gibt einen Schutz bei kurzanhaltenden Überspannungen. Bei längeren Überspannungen kann der Varistor überhitzen oder andersweitig zerstört werden und der Schutz fällt aus.

7.2 Varaktor-Diode

Die ***Varaktor =Varicap = Kapazitätsdioden (varactor)*** sind Dioden, welche Ihre kapazitiven Eigenschaften mit der angelegten Spannung ändern. Dabei steht der Begriff "varactor" für "variable reactance" was man mit "variablen Blindwiderstand" übersetzen müsste. *Bem: Ein Blindwiderstand ist ein Widerstand kapazitiver oder induktiver Natur. Im Gegensatz zum Wirkwiderstand, wo elektrische Leistung umgesetzt wird, verursachen Blindwiderstände theoretisch lediglich eine frequenzabhängige Phasenverschiebung zwischen Strom und Spannung ohne Leistung zu vernichten. So ist der induktive Blindwiderstand in einer idealen Spule z.B.* $X_L = \omega L$ *und der kapazitive Blindwiderstand eines idealen Kondensators* $X_C = -1/(\omega C)$.

Das Prinzip des Varaktors beruht auf der im vorangehenden Kapitel erwähnten Änderung der

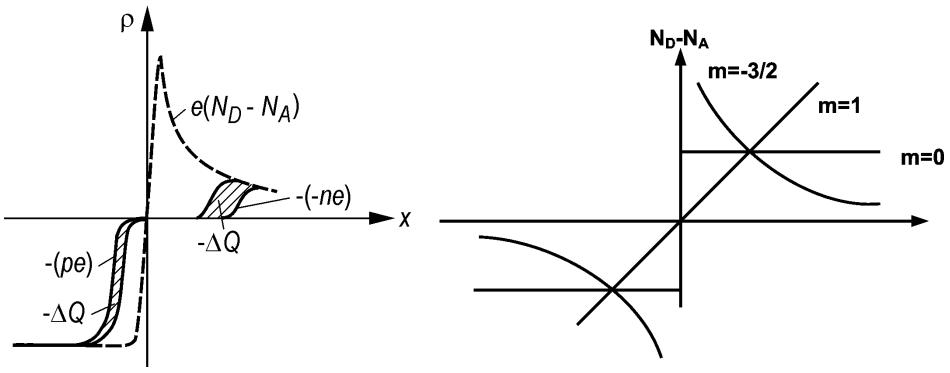


Abbildung 7.7: (a) Änderung der Majoritätsträgerladung als Folge einer Spannungsänderung in einer gesperrten pn-Diode mit hyperabruptem Dotierungsverlauf [3]. (b) Verschiedene Dotierungsprofile: m=1 linear, m=0 const. und m=-3/2 hyperabrupte Dotierung.

Kapazität einer Diode mit der angelegten Spannung. Wenn man die Diode entsprechend baut, so kann sich diese spannungsabhängige Kapazität sogar sehr stark ändern und man spricht dann zu recht von einer Varaktor-Diode.

Wie in Abschnitt 6.3.3 und 6.3.3 behandelt setzt sich die Kapazität einer Diode aus der Sperrschicht und Diffusionskapazität zusammen. Bild 7.6 zeigt den typischen Verlauf der Kleinsignal-Kapazitäten als Funktion einer angelegten Gleichspannung (Arbeitspunkt). Im Sperrbereich dominiert die Sperrschichtkapazität. Diese Kleinsignal-Kapazität kommt durch die in der Raumladungszone gespeicherte Ladung zustande, welche sich bei kleinen Spannungsänderungen verändert, Bild 7.7. Die Sperrschichtkapazität (unter der Annahme der Geometrie eines Plattenkondensators) ist gegeben durch

$$C_S = \frac{A}{l(U)} \varepsilon_r \varepsilon_0 \,, \tag{7.1}$$

wobei die Länge der RLZ nach Gl.(6.54) von der angelegten Spannung abhängt.

Die Längenänderung der RLZ mit der Spannungsabhängigkeit kann, wie aus Bild 7.7 unmittelbar hervorgeht entsprechend dem Dotierprofil verändert werden. Für ein Dotierprofil der Form $N_D - N_A \sim x^m$ erhält man (Übungsaufgabe) eine spannungsabhängige Kapazität der Form

$$C_S \sim (U_D - U)^{-\frac{1}{m+2}} \,. \tag{7.2}$$

Spannungsabhängige Kapazitäten können zur Abstimmung von Schwingkreisen in Oszillatoren, speziell von spannungsgesteuerten Oszillatoren (VCOs=Voltage Controlled Oscillators) wie sie für den Bau von Phasenregelschleifen notwendig sind, oder bei parametrischen Verstärkern, Frequenzmischern, u.s.w., eingesetzt werden.

Im Fall einer hyperabrupten Dotierung mit $m = -3/2$ erhält man für die Kapazität $C_S \sim (U_D - U)^{-2}$. Damit ergibt sich in einem Reihen-Schwingkreis mit der Induktivität L eine Resonanzfrequenz, welche linear mit der angelegten Spannung variiert

$$\omega_R = \frac{1}{\sqrt{L\, C_S}} \sim (U_D - U) \,. \tag{7.3}$$

Der nicht zu vernachlässigende Sperrstrom kann durch die Verwendung von Metall-Isolator-Halbleiterübergängen, die wir später noch genauer untersuchen, unterdrückt werden.

7.3 p^+sn^+-Leistungsgleichrichter

Diese Dioden werden als ***Gleichrichter (rectifier)*** eingesetzt. Sie sollen im Durchlassbereich hohe Ströme erlauben und im Sperrbereich hohen Spannungen trotzen können.

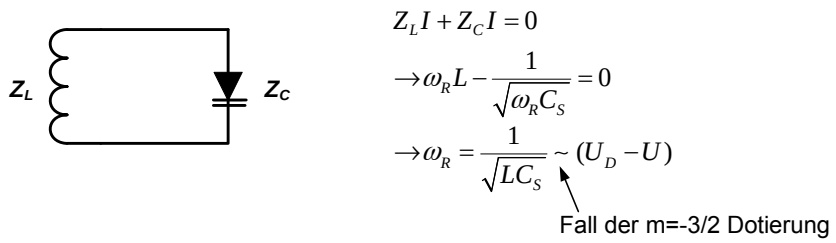


Abbildung 7.8: Resonanzfrequenz des Reihenschwingkreises.

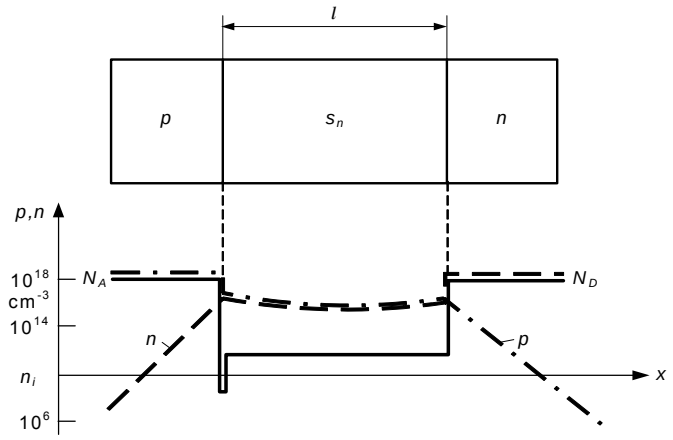


Abbildung 7.9: Trägerdichten in psn-Dioden bei Flusspolung [3].

Kleine Dotierungen erlauben einen großen Durchlassstrom I_S, wie aus Gl.(6.46-6.47) ersichtlich.

Andrerseits wird ein kleiner Sperrstrom I_S, also hohe Dotierung, angestrebt um die Verlustleistung im Sperrbereich zu minimieren. Zusätzlich wären hohe Durchbruchspannungen im Sperrbereich gefordert. Dazu wäre eine geringe Dotierung notwendig, da sich dann nach Gl.(6.29) eine breite RLZ ergibt, die eine große Sperrspannung aushalten kann, ehe die elektrische Durchbruchfeldstärke erreicht wird.

Diese zwei an sich im Widerspruch stehenden Forderungen lassen sich bei der p^+sn^+-Diode gleichzeitig erfüllen.

Bild 7.9 zeigt das Dotierungsprofil und die Trägerkonzentrationen im Durchlassbereich. Im Durchlassbereich überschwemmen die aus den hochdotierten Zonen injizierten Träger das schwach dotierte s-Gebiet (s: small doping level of n or p) und durchqueren es ohne zu rekombinieren, da die Dicke der Schicht in etwa nur eine Diffusionslänge beträgt. Die schwache Dotierung bewirkt selbst bei kleinen Durchlassspannungen einen kleinen Ohmschen Wiederstand. Mit zunehmender Spannung nimmt der Widerstand in Flussrichtung durch die wachsende Injektion freier Ladungsträger um mehrere Zehnerpotenzen ab.

Im Sperrbereich fällt die gesamte Sperrspannung an der schwach dotierten aber breiten Schicht ab, so dass der Durchbruch erst bei großen Werten von U_R eintritt. Gleichzeitig bleibt der Sperrstrom I_S klein, da in diesem Fall fast alle Minoritätsträger aus der s_n-Schicht ausgeräumt sind und nur wenige aus der hochdotierten n^+-Schicht nachgeliefert werden.

Für Leistungsdioden wird Silizium als Halbleitermaterial bevorzugt, da es eine kleine Eigenleitungsträgerdichte hat und damit bei kleinen Dotierungen kleine Sperrströme ermöglicht.

Bild 7.10 zeigt den typischen Aufbau einer Leistungsdiode um eine gute Wärmeabfuhr zu ermöglichen.

Einige Beispielwerte für Betriebswerte von solchen Ventilen für die Leistungstechnik sind in

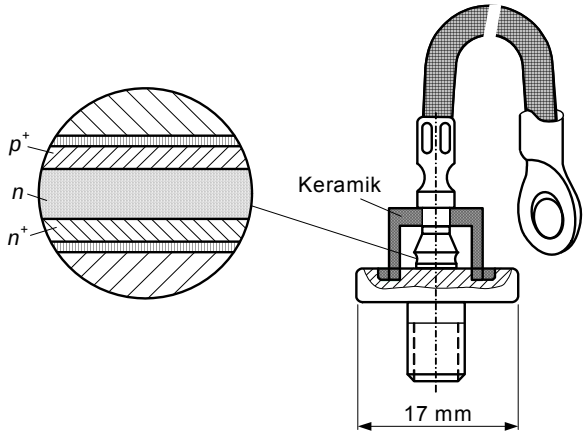


Abbildung 7.10: Si-Leistungsdiode [3].

Dauerströme	≈ 400 A
periodische Spitzenströme	≈ 2000 A
einmalige Stoßströme	≈ 4000 A
Flussspannung	$< 1,8$ V
Sperrströme	$\approx 20...50$ mA
Sperrschichttemperatur	$\approx 175^oC$
Dauersperrspannungen	≈ 1200 V
Spitzensperrspannungen	≈ 1600 V

Tabelle 7.1: Typische Betriebsdaten einer Leistungsdiode

Tabelle 7.1 zusammengefaßt.

Im Bild 7.11 sind zwei mögliche Gleichrichterschaltungen gezeichnet.

7.4 Photodiode

An dieser Stelle werden wir die auf dem pn- oder pin-Halbleiterübergang beruhende Photodiode diskutieren. Das grundlegende Verständnis der Funktionsweise der Photodiode erlaubt uns, später das Funktionsprinzip anderer Photodetektoren zu erarbeiten.

Es gibt verschiedene Arten von Photodetektoren, so z.B. die pn- oder pin-Photodiode, die Lawinenphotodiode, der Metallhalbleiter-Photodetektor oder die Metalloxid-Halbleiterdiode, welche gegenwärtig in CCDs (charged coupled devices) breite Anwendung als Bildsensor findet.

Bevor wir uns der pn- oder pin-Photodiode zuwenden, möchten wir zunächst darauf hinweisen, dass es ganz unterschiedliche Methoden gibt, Licht zu detektieren. Zur Detektion elektromagnetischer Strahlung wird im Allgemeinen die einfallende Strahlung absorbiert und in eine andere, messbare Form umgewandelt. Dabei können verschiedene Effekte ausgenutzt werden:

- Strahlungserwärmung: Elektromagnetische Strahlung wird in einem Detektor absorbiert und ändert dabei dessen Temperatur.
- Photochemisch: Einfallende Strahlung verändert die chemische Struktur von Materialien. (klassische Photographie)
- Photolumineszenz: Elektromagnetische Strahlung kurzer Wellenlänge wird absorbiert und elektromagnetische Strahlung langer Wellenlänge emittiert.

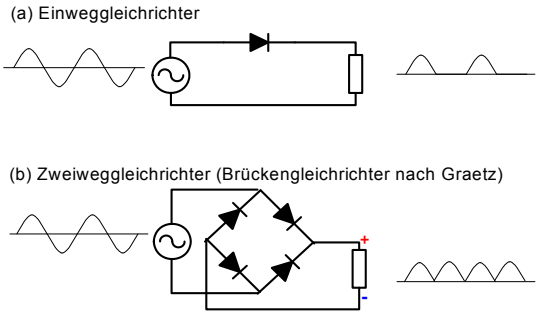


Abbildung 7.11: (a) Einweg- und (b) Zweiweggleichrichter.

- Photoleitung: Einfallende elektromagnetische Strahlung wird absorbiert und erhöht die Leitfähigkeit eines Materials.
- Photovoltaik: Einfallende elektromagnetische Strahlung wird zur Erzeugung einer elektrischen Spannung ausgenutzt.
- Photoelektrizität: Einfallende elektromagnetische Strahlung wird zur Erzeugung eines elektrischen Stromes ausgenutzt.

Die Photodetektoren, welche auf dem photoelektrischen Effekt beruhen und Strahlung in elektrische Signale umwandeln sind unter den verschiedenen Detektoren von besonderem Interesse, weil diese den Bau von besonders empfindlichen, kompakten und schnellen Detektoren ermöglichen, welche direkt proportional zur Intensität des einfallenden Lichtes sind.

Gute Photodioden sollten folgende Merkmale aufweisen:

- hohe Empfindlichkeit
- große Modulationsgeschwindigkeiten
- geringes Rauschen

7.4.1 Wirkungsweise der Photodiode

Zur Erzeugung eines elektrischen Stromes bei Lichteinfall in eine pn-Halbleiterdiode sind folgende zwei Schritte notwendig:

1. Einfallendes Licht muss frei bewegliche Elektron-Loch-Paare erzeugen. Die Photonen werden dabei absorbiert.
2. Damit ein Strom entstehen kann, müssen die frei beweglichen Ladungspaare separiert und transportiert werden. Die Separation ist notwendig, damit sie nicht wieder rekombinieren.

Halbleiter-pn-Übergänge erlauben die Erzeugung dieser zwei Prozessschritte und eignen sich deshalb zur Detektion optischer Strahlung. Wir erörtern den zugrundeliegenden physikalischen Mechanismus anhand von Fig. 7.12.

In Abb. 7.12(a) haben wir einen in Sperrrichtung betriebenen pn-Übergang dargestellt. Die elektrischen Kontakte sind seitlich angebracht. Photonen können durch eine Öffnung im Kontaktmetall des p-Halbleiters in die pn-Diode gelangen. Innerhalb der pn-Diode lassen sich drei Bereiche ausmachen: die weitgehend von frei beweglichen Ladungsträgern verarmte Raumladungszone (RLZ), die von Diffusionsströmen dominierte Diffusionszone (DZ) sowie die Bahngebiete. Die entsprechende elektrische Feldverteilung, die Ladungsträgerverteilung und der Bandverlauf sind in 7.12(b), (c) und (d) gezeigt.

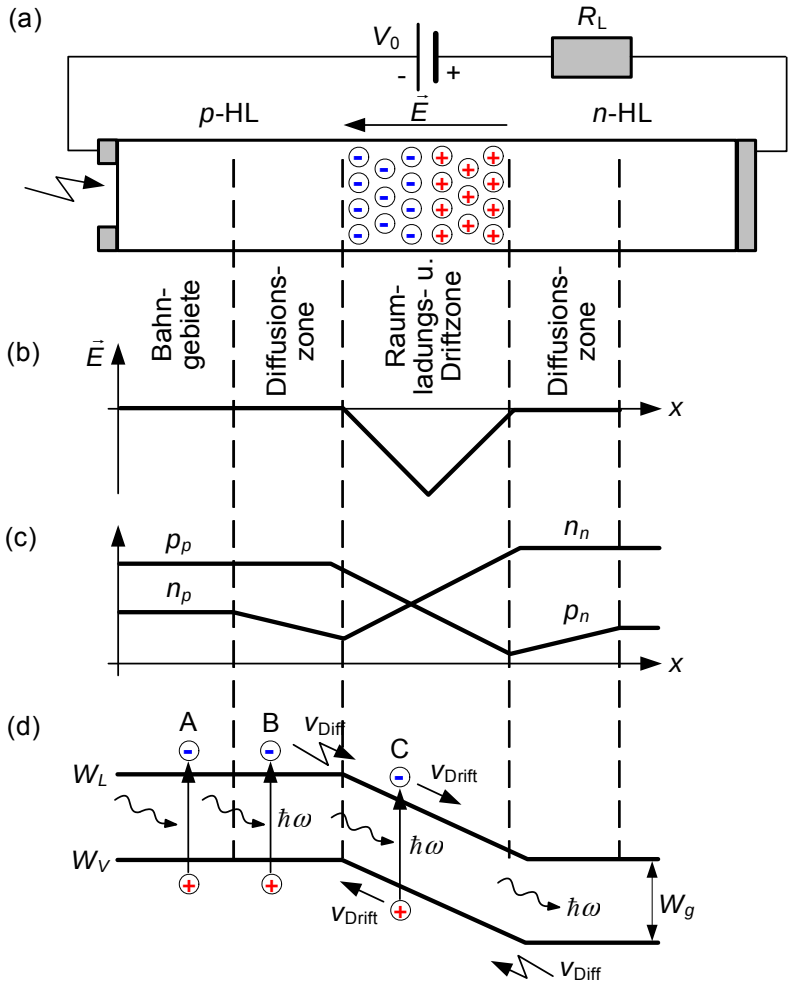


Abbildung 7.12: (a) pn-Halbleiterdiode in Sperrichtung, (b) Elektrisches Feld, (c) Ladungsträgerverteilung (d) Bandverlauf

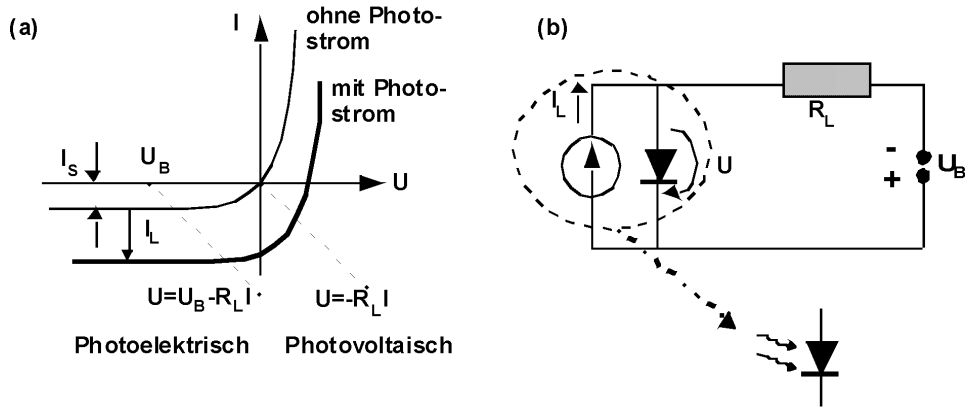


Abbildung 7.13: (a) Kennlinie der pn-Diode ohne und mit Photostrom I_L. Falls man die Spannungsquelle kurzschließt, kann die Diode als Solarzelle betrieben werden und Leistung abgeben. (b)Ersatzschaltung und Symbol.

Photonen können im pn-Halbleiter absorbiert werden und Elektron-Loch-Paare erzeugen, falls die Energie der einfallenden Photonen größer ist als die Energie der Bandlücke. Am wichtigsten ist die Paarerzeugung im Bereich der Raumladungszone, also in 7.12(d) im Punkt C. Elektronen, welche in C erzeugt werden, driften unter dem Einfluss des elektrischen Feldes der RLZ zur n-Seite, Löcher dagegen zur p-Seite. Dabei influenzieren beide Ladungen einen Stromimpuls. Die Dauer des Stromimpulses ist durch die Driftgeschwindigkeit und Driftlänge bestimmt. Wenn das Ladungsträgerpaar in der Diffusionszone bei B erzeugt wurde, wird das Elektron mit großer Wahrscheinlichkeit durch Diffusion in die RLZ und dort im RLZ-Feld weiter zur n-Seite gelangen und dadurch einen Stromimpuls generieren. Der Diffusions- und Driftvorgang dauert allerdings ungleich länger, als wenn das Elektron-Loch-Paar in C erzeugt worden wäre. Für ein in D erzeugtes Elektron-Loch-Paar ist die Situation ganz ähnlich. Allerdings muss hier das Loch zur p-Seite diffundieren. Das ist insofern ungünstig, als die Diffusionszeit für ein Loch länger ist als für ein Elektron, wodurch das dynamische Verhalten der Photodiode geschwächt wird. Werden Elektron-Loch-Paare in A erzeugt, so wird das Elektron wieder mit dem Loch rekombinieren, und das absorbierte Photon gibt keinen Beitrag zum Stromfluss im äußeren Kreis.

Die Spannungs-Strom-Kennlinie der Photodiode ergibt sich als Summe des von den Photonen erzeugten Photostromes I_L und des Diodenstromes I_D. Der Photonenstrom stammt im Wesentlichen aus der Generation von Elektronen-Loch-Paaren, welche in der RLZ und den angrenzenden Diffusionsgebieten generiert worden sind. Er muss dem vorhandenen Diodenstrom I_D, welcher vor allem aus der Ladungsträgergeneration in den Diffusionsgebieten stammt, hinzugefügt werden:

$$I = I_D + I_L = I_S \left(\exp\left(\frac{eU}{kT}\right) - 1 \right) - I_L \tag{7.4}$$

In Gl. (7.4) bezeichnet I_S den Diodensperrschicht-Sättigungsstrom und U die an die Diode angelegte Spannung. Die Kennlinie der Photodiode samt Ersatzschaltung und Symbol ist in Fig. 7.13 gezeigt.

Typische Diodensperrschicht-Sättigungsströme betragen einige Nanoampere. Photodiodenströme sind in der Größenordung von einigen Milliampere

7.4.2 Die pin-Photodiode

Um den Wirkungsgrad und die Dynamik der pn-Photodiode zu verbessern, muss man die Diffusionszonen klein halten und stattdessen die Driftzonen mit den viel schnelleren Ladungsträgergeschwindigkeiten vergrößern. Dies kann man durch das Einbringen einer undotierten, eigenleitenden (d.h. intrinsischen) i-Schicht zwischen den p- und n-dotierten Halbleitern erreichen (7.14). Weil die

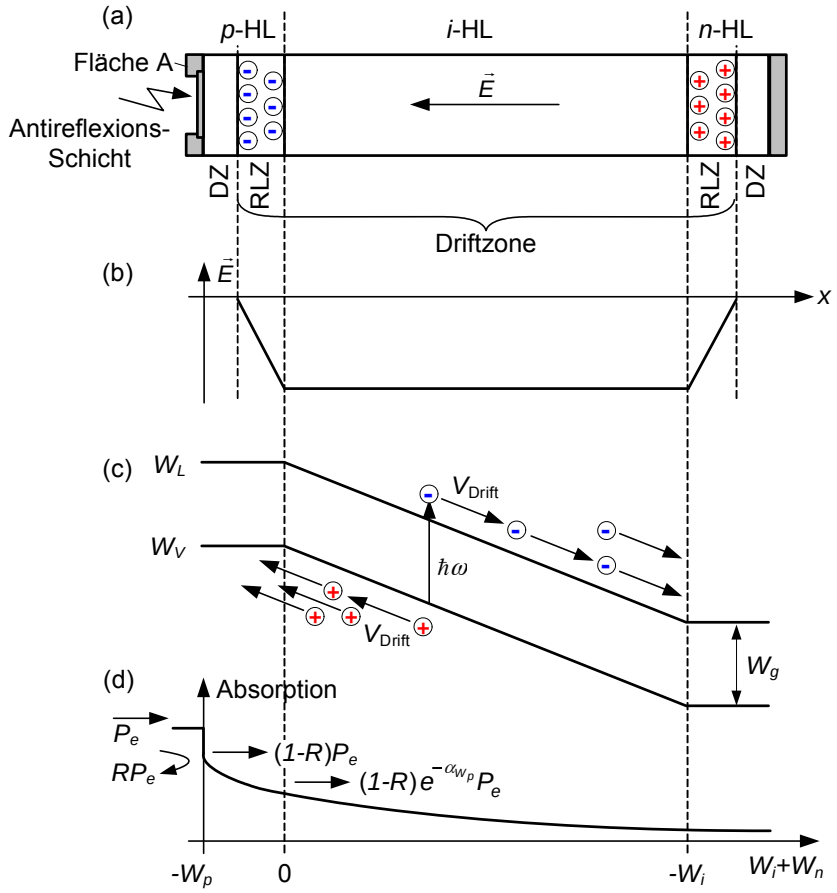


Abbildung 7.14: (a) pin-Photodiode mit (b) el. Feldverteilung, (c) Ladungsträgerdichte und (d) Bandverlauf

undotierte Halbleiterschicht einen hohen Widerstand aufweist, fällt fast die ganze angelegte Spannung über der intrinsischen Halbleiterschicht ab. Damit liegt über der ganzen i-Halbleiterschicht ein großes elektrisches Feld, welches von den Raumladungen am Rand der i-Schicht erzeugt wurde. Folgende Vorteile ergeben sich:

- Eine Verbesserung des Quantenwirkungsgrades, weil die Absorptionszone innerhalb der für den Ladungstransport wichtigen Driftzone vergrößert wurde, und
- Eine Verbesserung der Dynamik, weil die für den schnellen Ladungsträgertransport wichtige Driftzone im Bereich des maximal großen elektrischen Feldes vergrößert wurde. Die Stärke des elektrischen Feldes ist lediglich durch die Sättigungsgeschwindigkeit der Ladungsträger oder die Durchbruchspannung der Diode limitiert.

Im Folgenden werden wir sowohl den Quantenwirkungsgrad als auch die Modulationsgeschwindigkeit für die pin-Photodiode herleiten.

Die drei Halbleitergrundgleichungen (5.6)-(5.8) bilden den Ausgangspunkt für Berechnungen im Halbleiter.

7.4.3 Empfindlichkeit und Quantenwirkungsgrad

Ein Maß für den proportional zur eingestrahlten Leistung erzeugten Stromfluss ist die sogenannte ***Empfindlichkeit (responsitivity)*** $\Re$:

$$I_L = \Re P_e \tag{7.5}$$

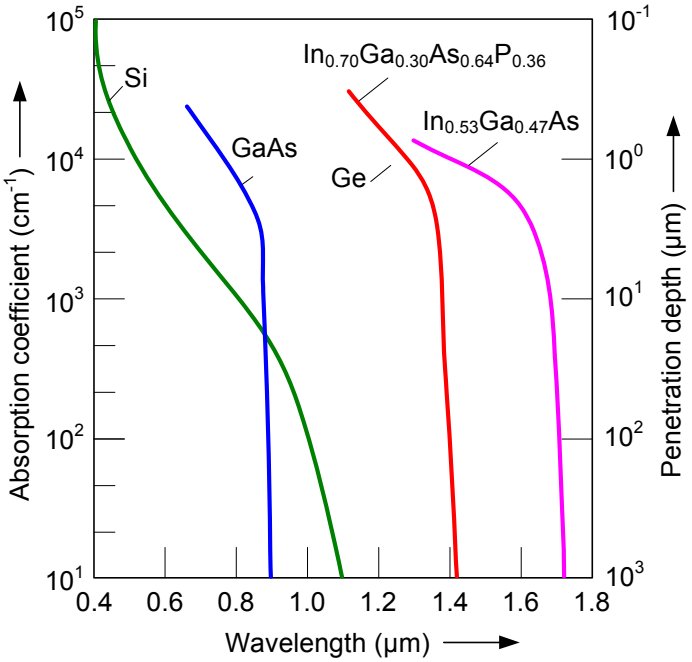


Abbildung 7.15: Wellenlängenabhängigkeit des Absorptionskoeffizienten für vier wichtige Halbleitermaterialien.

Die Empfindlichkeit $\Re$ wird üblicherweise in A/W angegeben. Ein typischer Wert für die Empfindlichkeit von InGaAs-Photodetektoren im Wellenlängenbereich von 1.55 mm ist 0.5 A/W.

Die Empfindlichkeit ist direkt verknüpft mit der etwas physikalischeren Größe des ***Quantenwirkungsgrades (quantum efficiency)*:**

$$\eta = \frac{\text{Generierte Elektronen}}{\text{Eingestrahlte Photonen}} = \frac{I_L/e}{P_e/\hbar\omega} = \frac{\hbar\omega}{e}\Re \cong \frac{1.24}{\lambda[\mu m]}\Re \tag{7.6}$$

Der Quantenwirkungsgrad der pin-Photodiode lässt sich anhand der Fig. 7.14(c) direkt hinschreiben. Es ist der Anteil des Lichts, welcher in der intrinsichen Zone absorbiert wird:

$$\eta = (1-R)\, e^{-\alpha\, w_p} - (1-R)\, e^{-\alpha\, w_p} e^{-\alpha\, w_i} \tag{7.7}$$

$$\eta = (1-R)\, e^{-\alpha\, w_p} \left(1 - e^{-\alpha\, w_i}\right) \tag{7.8}$$

Aus Glg. (14) wird ersichtlich, dass man für eine hohe Empfindlichkeit/großen Quantenwirkungsgrad:

- die Reflexionen an der Diodenoberfläche minimieren muss, d.h. . In der Praxis heißt das, dass man die Oberfläche mit einer Antireflexionsschicht versehen muss.
- die p-Halbleiter-Schicht so dünn als möglich halten, sowie
- ein Material mit großer Absorption und langer i-Schichtdicke wählen muss.

In 7.15 sind die Absorptionskoeffizienten für verschiedene Halbleitermaterialien gegeben. Absorption gibt es nur für Wellenlängen, welche kleiner sind als die Bandlückenwellenlänge. Bei großen Werten des Absorptionskoeffizienten α von ca. 104/cm erhält man für i-Schichten von ca. 10 mm fast 100% Absorption.

Bild 7.16 zeigt die spektrale Empfindlichkeit von verschiedenen Halbleitermaterialien und den theoretisch möglichen Quantenwirkungsgrad η_Q.

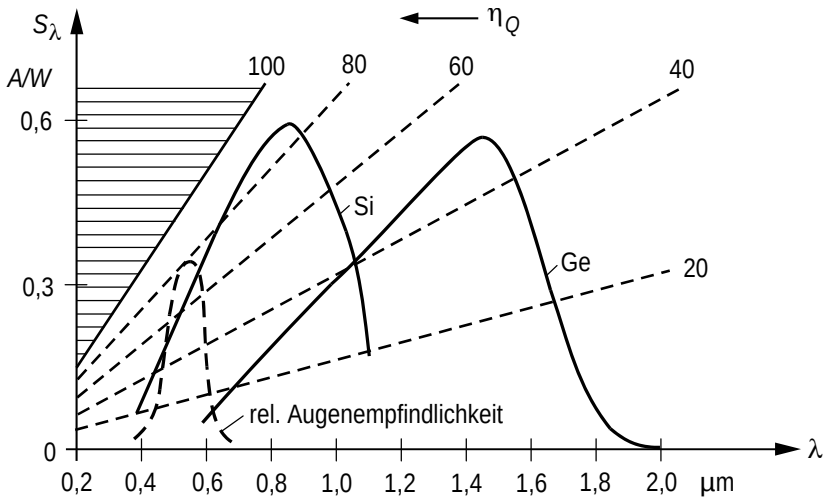


Abbildung 7.16: Empfindlichkeit oder Quantenwirkungsgrad von Si und Ge im Vergleich zur Augenempfindlichkeit. Zum Vergleich: (- - -) Geraden welche den angegebenen Werten des Quantenwirkungsgrades entsprechen [3].

Korrekte Herleitung des Quantenwirkungsgrades (optional)

Im Folgenden wollen wir den Quantenwirkungsgrad korrekt herleiten. Oben haben wir ihn lediglich anhand einer Betrachtung aus der Figur aufgeschrieben. Zunächst möchten wir den Zusammenhang zwischen generierten Ladungsträgern und erzeugtem Strom herleiten. Die Kontinuitätsgleichungen (5.6) und (5.7) liefern uns den gesuchten Zusammenhang. Dazu machen wir folgende Annahmen: Es tragen nur Ladungspaare, welche in der i-Schicht erzeugt wurden, zum Strom bei. Die Rekombinationsterme r_n und r_p können vernachlässigt werden. Allerdings müssen wir zuerst einen Ausdruck für die Ladungsträgergenerationsrate $g \equiv g_n = g_p$ herleiten.

Die Generationsrate g ergibt sich aus der im Volumen $(A\ dx)$ absorbierten Zahl der Photonen pro Zeit:

$$g(x,t) = -\frac{1}{\hbar\omega}\frac{dP(x,t)}{A\,dx}\,, \tag{7.9}$$

wobei $P(x,t)$ die Lichtintensität bezeichnet. Der Ausdruck für $P(x,t)$ muss noch hergeleitetet werden. Beim Durchlaufen des Halbleiters wird die eingestrahlte Lichtintensität proportional zur eigenen Intensität - mit dem Absorptionskoeffizienten α - absorbiert. Es gilt also:

$$\frac{dP}{dx} = -\alpha\,P(x) \tag{7.10}$$

Wie in Abb. 7.14(d) bereits gezeichnet, führt dies zu einer beim Durchlaufen exponentiell abklingenden Intensität der Form:

$$P(x,t) = P_e(t)\,(1-R)\exp\left(-\alpha x\right), \tag{7.11}$$

wobei je nach Größe der Reflexionen nur der Anteil $P_e(t)(1-R)$ in die Probe gelangt und der Bruchteil $\exp\left(-\alpha\,w_p\right)$ absorbiert wird, bevor er in die i-Schicht gelangt.

Wenn man Gl. (7.11) in Gl. (7.9) einsetzt, ergibt sich für $g(x,t)$:

$$g(x,t) = \alpha\left[\frac{P_e(t)\,(1-R)\exp\left(-\alpha\,w_p\right)}{\hbar\omega\ \ A}\right]\exp\left(-\alpha x\right) \equiv G_0\exp\left(-\alpha x\right), \tag{7.12}$$

Damit kann man den Stromfluss mit Hilfe der Kontinuitätsgleichungen (5.6) und (5.7) berechnen. Diese lauten im stationären Zustand:

$$\frac{\partial J_p}{\partial x} = eG_0\exp(-\alpha\,x) \text{ und } \frac{\partial J_n}{\partial x} = -eG_0\exp(-\alpha\,x)\,. \tag{7.13}$$

Diese Differentialgleichungen lassen sich direkt lösen. Dabei gelten folgende Randbedingungen am Rand der i-Schicht: $J_n(0) = 0$ und $J_p(w_i) = 0$. Dies ergibt dann:

$$J_p(x) = \frac{eG_0}{\alpha}\left[e^{-\alpha\, w_i} - e^{-\alpha\, x}\right] \text{ und } J_n(x) = -\frac{eG_0}{\alpha}\left[1 - e^{-\alpha\, x}\right]. \tag{7.14}$$

Der Gesamtstrom setzt sich aus den Beiträgen vom Löcher- und Elektronenstrom zusammen. Ferner ist im äußeren Stromkreis nur der Mittelwert über allen Strömen in der i-Schicht messbar. Damit ergibt sich:

$$I_L(t) = A\frac{1}{w_i}\int_0^{w_i}\left(J_p\left(x\right) + J_n\left(x\right)\right)dx \tag{7.15}$$

$$= -\frac{e}{\hbar\omega}\left(1-R\right)e^{-\alpha\, w_p}\left(1 - e^{-\alpha\, w_i}\right)P_e(t) \tag{7.16}$$

Einsetzen von $I_L(t)$ und $P_e(t)$ in der Definition für den Quantenwirkungsgrad Gl. (7.6) liefert den korrekten Ausdruck für η.

7.4.4 Photodioden-Modulationsgeschwindigkeit

Einfluss der Ladungsträger-Driftzeit auf die Modulationsgeschwindigkeit

Um den Einfluss der Driftzeit auf den Stromfluss zu studieren, betrachten wir einen kurzen Puls der Energie E_{pulse}, welcher in einer dünnen Schicht in unmittelbarer Nähe der p-Zone absorbiert wird. D.h. wir setzen für die Generationsrate

$$g\left(x,t\right) = G_0 \cdot \delta\left(t\right)\delta(x) \tag{7.17}$$

Weil wir keine Raumladungen in der Driftzone haben, folgert aus der Poissongleichung (5.8), dass das elektrische Feld in der ganzen Driftzone konstant ist. Deshalb werden sich die Löcher und Elektronen sehr bald mit der maximal möglichen Geschwindigkeit, nämlich der materialabhängigen Sättigungsgeschwindigkeit v_{sn} und v_{sp} durch die Driftzone bewegen (Abb. 4.5). Unter dem Einfluss des Feldes verschwinden die Löcher sofort in der p-Zone. Die Elektronen hingegen benötigen zum Durchlaufen der Driftzone die Zeit $t_{drift} = w_i/v_{sn}$, während der sie einen Strom generieren.

Im Folgenden müssen wir die Kontinuitätsgleichung für die Elektronen lösen, d.h.:

$$-e\frac{\partial n}{\partial t} + \frac{\partial J_n}{\partial x} = -eG_0 \cdot \delta\left(t\right)\delta(x) \text{ oder} \tag{7.18}$$

$$-\frac{1}{v_{sn}}\frac{\partial J_n}{\partial t} + \frac{\partial J_n}{\partial x} = -eG_0 \cdot \delta\left(t\right)\delta(x). \tag{7.19}$$

Zunächst ist festzustellen, dass die Lösung der homogenen Gleichung die Form $f(x - v_{sn}t)$ haben muss. Ferner wissen wir aus der Anschauung, dass der Strom nur solange fließt, wie die freien Ladungsträger driften. D.h. die Lösung ist nur im Intervall $[0,\, t_{drift}]$ von Null verschieden. Durch Einsetzen kann man zeigen, dass die Gleichung folgende Lösung hat:

$$J_n\left(x,t\right) = v_{sn}eG_0 \cdot \delta(x - v_{sn}t) \text{ für } 0 < t \leq t_{drift}\,. \tag{7.20}$$

Das ist ein Strompuls, welcher in der Zeit t_{drift} von links nach rechts durch die i-Schicht driftet. Damit ergibt sich für den äußeren Strom, welcher als Mittelwert aller Ströme in der Driftzone zu betrachten ist, in Analogie zu Gl. (7.15)

$$I_L\left(t\right) = A\frac{1}{w_i}\int_0^{w_i} J_n\left(x,t\right)dx = A\frac{v_{sn}}{w_i}eG_0 = AeG_0 \cdot \frac{1}{t_{drift}} = -e\cdot\eta\frac{E_{pulse}}{\hbar\omega}\cdot\frac{1}{t_{drift}} \text{ für } 0 < t \leq t_{drift} \tag{7.21}$$

Beachte: in Gl. (7.21) haben wir den Anteil des Löcherstromes, welcher nur bei t=0 einen Beitrag leistet, vernachlässigt

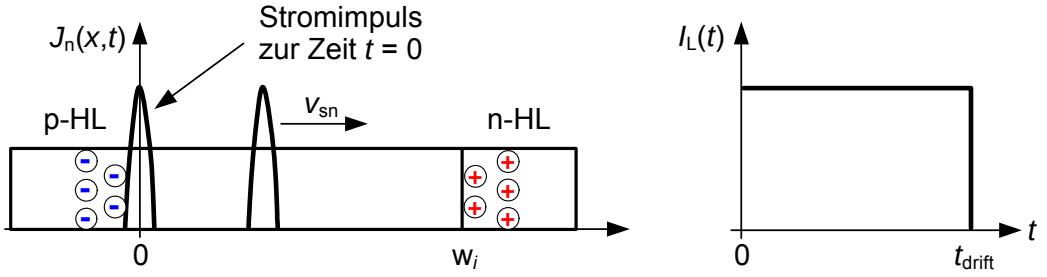


Abbildung 7.17: (a) Der kurze Elektronenstrompuls, welcher zur Zeit t=0 bei x=0 generiert wurde, driftet mit der Sättigungsgeschwindigkeit v_{sn} durch die pin-Diode. (b) Ein äusserer Strom I_L wird während der ganzen Zeit des Driftens erzeugt

Fig. 7.17 und Gl. (7.21) zeigen, dass es eine obere Modulationsfrequenz $f_{drift} \simeq 1/t_{drift}$ für Lichtimpulse gibt. Optische Pulse, welche weniger als t_{drift} voneinander getrennt sind, können nicht mehr aufgelöst werden.

Mit Gl. (7.21) haben wir die Impulsantwort der Photodiode auf einen ultrakurzen Puls gegeben. Das Frequenzspektrum, welches das Bauteil mit dieser Impulsantwort gerade noch verarbeiten kann, erhält man als Fouriertransformierte der Impulsantwort:

$$H\left(f\right) = -\eta \frac{eE_{pulse}}{\hbar\omega} \cdot e^{\mathrm{j}\,\omega t_{drift}} \frac{\sin\left(\omega t_{drift}/2\right)}{\omega t_{drift}/2}\,. \tag{7.22}$$

Die Frequenz, bei welcher die Signalleistung um 50% (3 dB) verringert ist, wird damit

$$f_{drift,3dB} = \frac{0.44}{t_{drift}}\,. \tag{7.23}$$

Einfluss der Diffusionszeit auf die Modulationsgeschwindigkeit

Auf der Seite des p-Halbleiters treten die langsameren Diffusionsströme auf. Diese tragen ebenfalls zum Stromfluss bei. Wir haben die Diffusionslänge als Funktion der Lebensdauer in Gl. (5.70)hergeleitet

$$L_{n_p} = \sqrt{D_n \tau_n}\,, \tag{7.24}$$

wobei τ_{n_p} die Minoritätsträgerlebensdauer der Elektronen in der p-Halbleiterschicht ist und D_{n_p} der Diffusionskoeffizient der Elektronen war in dieser Schicht.

Wenn wir nun die p-Halbleiterschicht kürzer als die Diffusionslänge wählen, so verringert sich die mittlere Driftzeit entsprechend:

$$w_p = \sqrt{D_n t_{diff}}\,, \tag{7.25}$$

Wenn wir nach t_{diff} auflösen, erhalten wir:

$$t_{diff} = \frac{w_p^2}{D_n} = \frac{e \cdot w_p^2}{\mu_n kT} \text{ für } w_p \leq L_n. \tag{7.26}$$

Bei genauerer Rechnung findet man eine durch den Diffusionsstrom gegebene obere Grenzfrequenz von

$$f_{diff} = 0.4 \cdot \frac{1}{t_{diff}} = 0.4 \cdot \frac{D_n}{w_p^2} \tag{7.27}$$

Da der Diffusionskoeffizient der Elektronen wesentlich größer ist als derjenige der Löcher, leitet man das Licht mit Vorteil über die p-Halbleiterschicht in die Photodiode ein. Damit profitiert man von den schnelleren Elektronenminoritäts-Diffusionskonstanten.

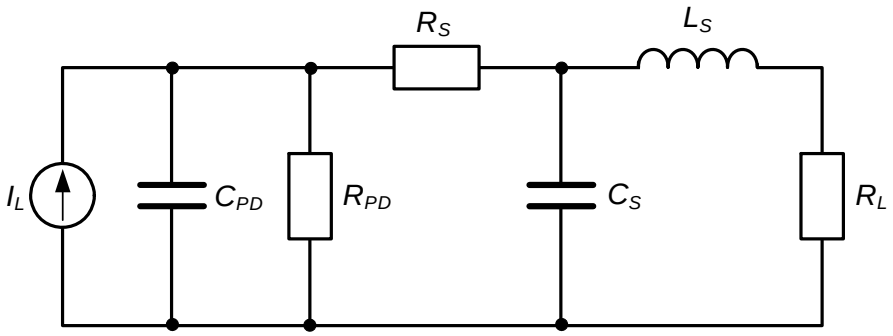


Abbildung 7.18: Kleinsignalersatzschaltbild der Photodiode.

Einfluss der Sperrschichtkapazität auf die Modulationsgeschwindigkeit

In einem äußeren Stromkreis führt der Spannungsabfall am Lastwiderstand (z.B. 50 Ω) zusammen mit der Sperrschichtkapazität der Raumladungszonen zu einem RC-Tiefpass-Verhalten. Die durch die Raumladungszone gebildete Diodenkapazität kann einmal mehr als Plattenkondensator modelliert werden:

$$C_S = \frac{A}{w_i}\varepsilon_r\varepsilon_0 \tag{7.28}$$

Die damit verbundene Grenzfrequenz ist gegeben als

$$f_{RC} = \frac{1}{2\pi R_L C_S} \tag{7.29}$$

Damit die RC-Grenzfrequenz hoch wird, sollte man also eine kleine Diodenquerschnittsfläche und eine lange intrinsische i-Schicht wählen. Allerdings darf man die intrinsische Schicht auch nicht zu dick machen, da dies ansonsten zu langen Driftzeiten führt. Man kommt in der Praxis nicht um eine geeignete Optimierung herum.

Bei genauerer Optimierung kann man weitere parasitäre Einflüsse berücksichtigen. Das Kleinsignalersatzschaltbild in Fig. 7.18 berücksichtigt neben der Diodenkapazität C_S und dem Lastwiderstand R_L auch noch einen Diodenparallelwiderstand R_{PD} aus Oberflächenströmen, Bahn- und Zuleitungswiderstände R_S, Induktivitäten L_S sowie Streukapazitäten C_S aus Stromzuführungen

Modulationsgeschwindigkeit

Die obere Modulationsfrequenz des erzeugten Stromes ist also hauptsächlich limitiert durch die Transitzeit der Elektronen und Löcher sowie das RC-Tiefpassverhalten.

Die totale Transitzeit der Ladungsträger ergibt sich als Summe der Diffusionszeit der Minoritätsträger und der Ladungsträgerdriftzeit in der intrinsischen Schicht. Die Grenzfrequenz ergibt sich damit gemäß Gl. (7.23) zu:

$$f_{transit} = \frac{0.44}{t_{drift} + t_{diff}} \tag{7.30}$$

Es ist von Fall zu Fall abzuklären, ob die Ladungsträgerdiffusion eine Rolle spielt. Eventuell machen sich die Ladungsträger aus der Diffusionszone nur durch einen kleinen Sprung im Frequenzverhalten bemerkbar.

Die sich ergebende totale Grenzfrequenz (cut-off frequency) errechnet sich gemäß der nachstehenden Formel:

$$\frac{1}{f_{3dB}^2} = \frac{1}{f_{transit}^2} + \frac{1}{f_{RC}^2} \tag{7.31}$$

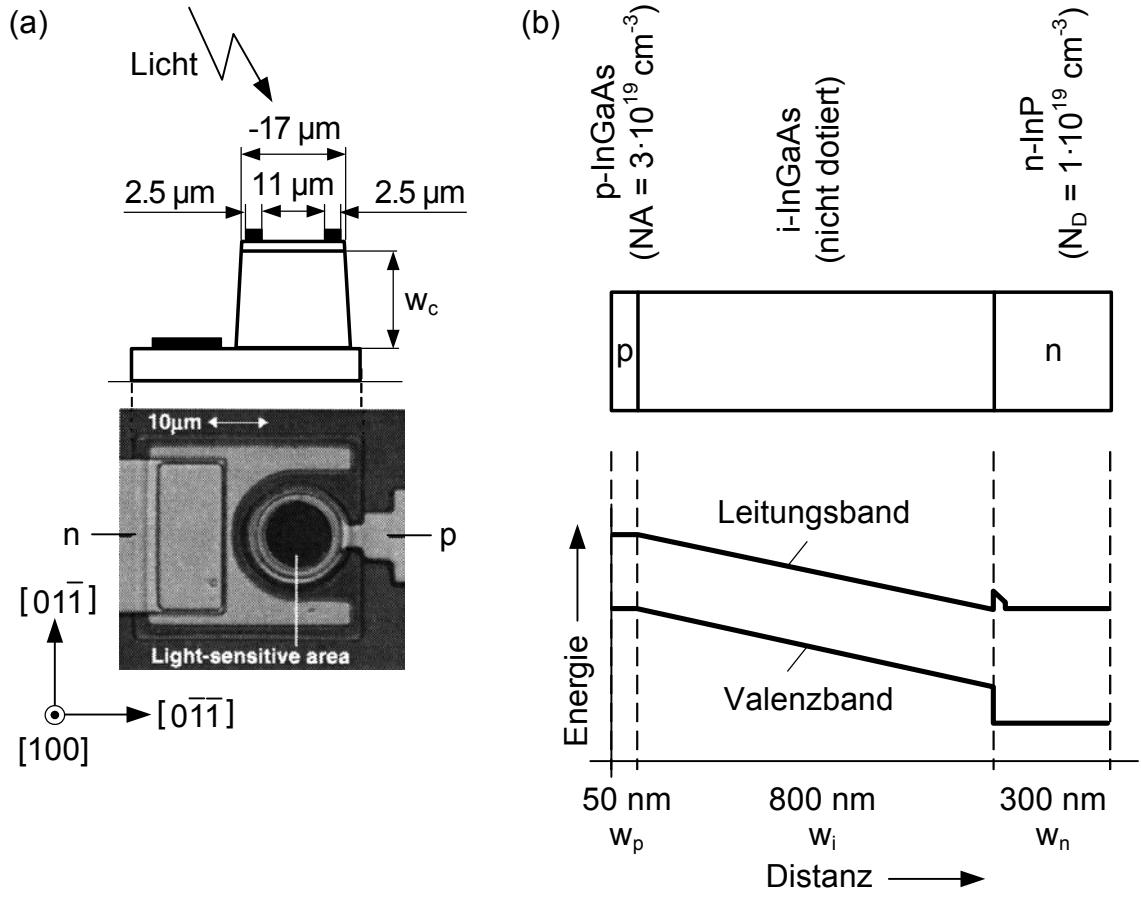


Abbildung 7.19: (a) Struktur und Ansicht auf Lichtöffnung der pin-Photodiode. (b) Dotierungen und Energiebänder der pin-Diode

7.4.5 Beispiel: 40 Gb/s pin-Heterostruktur-Photodiode

Zum Abschluss werden wir uns eine reale pin-Photodiode anschauen, wie sie gegenwärtig für 40 Gb/s-Kommunikationssysteme entwickelt werden [27].

Der Aufbau und ein Bild mit Ansicht auf die Lichtöffnung im p-Halbleiter ist in Fig. 7.19 gegeben.

Die p-Seite wird über einen ringförmigen Metallkontakt und die n-Seite durch einen Metallkontakt auf dem n-Halbleiter an den äußeren Stromkreis angeschlossen. Als Halbleiter für die p- und i-Schicht wurde $In_{0.53}Ga_{0.47}As$ gewählt. InGaAs absorbiert Licht mit einer Wellenlänge von 1.65 μm und kleiner. Für den n-Halbleiter wurde InP gewählt. InP hat eine viel größere Bandlückenenergie und ist für Licht im Bereich der Telekommunikationswellenlängen von 1.3 - 1.6 μm undurchsichtig. Photodioden, bei welchen die dotierten Bereiche und intrinsischen Schichten aus verschiedenen Materialien bestehen, werden ***Heterostruktur-Photodioden*** genannt. Die Schichtdicken und die Leitungs- und Valenzbandenergien sind in Fig.7.19(b) schematisch gezeichnet. Eine antireflektierende Schicht auf der p-Halbleiteroberfläche verhindert Reflexionen, so dass $R \cong 0$.

Zunächst wollen wir den Quantenwirkungsgrad der Photodiode mit Hilfe von Gl. (7.7) berechnen. Der Absorptionskoeffizient (α=7000 cm^{-1}) kann der Fig. 7.15 entnommen werden. Wir finden für die intrinsische Schicht einen Quantenwirkungsgrad von 41%. Mit der gleichen Formel können wir berechnen, dass ca. 4% des Lichtes in der p-Halbleiterschicht absorbiert wird. Falls die n-dotierte Seite ebenfalls absorbierend wäre (also aus InGaAs statt InP bestünde), so würden

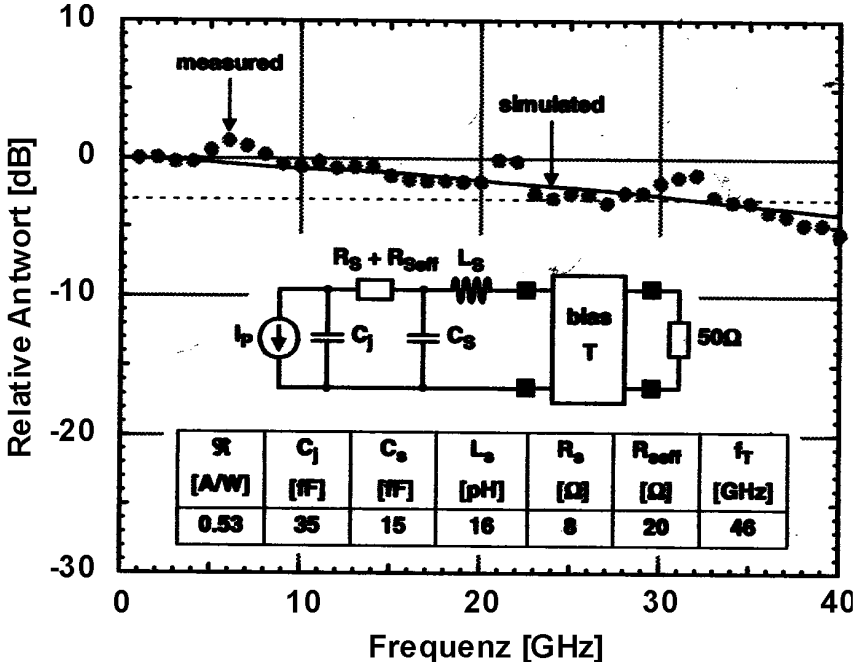


Abbildung 7.20: Gemessene Frequenzantwort der pin-Photodiode von Fig. 7.19. Die 3dB-Grenzfrequenz liegt zwischen 25 und 30 GHz. Die errechnete Kurve aus einer Simulation mit dem entsprechenden Ersatzschaltbild ist als Linie eingezeichnet.

dort noch weitere 10% des Lichtes absorbiert werden. Damit könnten wir den Anteil des in der n-Schicht absorbierten Lichtes nicht mehr vernachlässigen.

Um 40 Gb/s-Signale (25 ps Abstand zwischen den Pulsen) detektieren zu können, benötigt man eine minimale Photodiodenbandbreite von etwa 28 GHz. Damit muss die Transitzeit der Ladungsträger gemäß Gl. (7.30) kürzer als 16 ps sein.

Mit Sättigungsgeschwindigkeiten von 4.8 und 5.4 10^6 cm/s für Löcher und Elektronen in elektrischen Feldern von mehr als 50 kV/cm findet man Driftzeiten von 14 und 16 ps.

Die Elektronendiffusionsgeschwindigkeit auf der p-Halbleiterseite führt zu einer Diffusionszeit von 0.2 ps und kann vernachlässigt werden. Hingegen erweist sich die Verwendung von nicht absorbierendem InP auf der n-Halbleiterseite als sehr geschickt. Die lange Lebensdauer von 100 ps für die Löcher und die langsamen Beweglichkeiten würden sonst zu Diffusionslängen von 150 nm mit Diffusionszeiten von 66 ps führen. Bei fast 10% Absorption könnte man einen solchen Beitrag nicht mehr vernachlässigen.

Schlussendlich muss noch die RC-Grenzfrequenz berücksichtigt werden. Messungen der Kapazität und Simulationen führten zum Ersatzschaltbild mit den Werten in Fig. 7.20.

7.5 Lawinen-Photodiode

Eine ***Lawinen-Photodiode (avalanche photodiode*** oder kurz ***APD)*** ist eine Photodiode mit interner Verstärkung. Wegen des eingebauten Verstärkers ist sie besonders empfindlich. Allerdings bezahlt man diese höhere Empfindlichkeit mit einer etwas langsameren Geschwindigkeit und etwas mehr Rauschen.

Die Struktur der Lawinen-Photodiode entspricht einmal mehr einer Modifikation der pn-Diode. Dabei wird die Driftzone derart dotiert, dass man sowohl eine effiziente Absorptionszone, als auch einen Bereich mit Lawinenverstärkung erhält. Eine typische Struktur, ist in Fig. 7.21 abgebildet. In diesem Beispiel fällt das Licht auf der p-Seite in die Diode ein und wird gegen rechts zunehmend

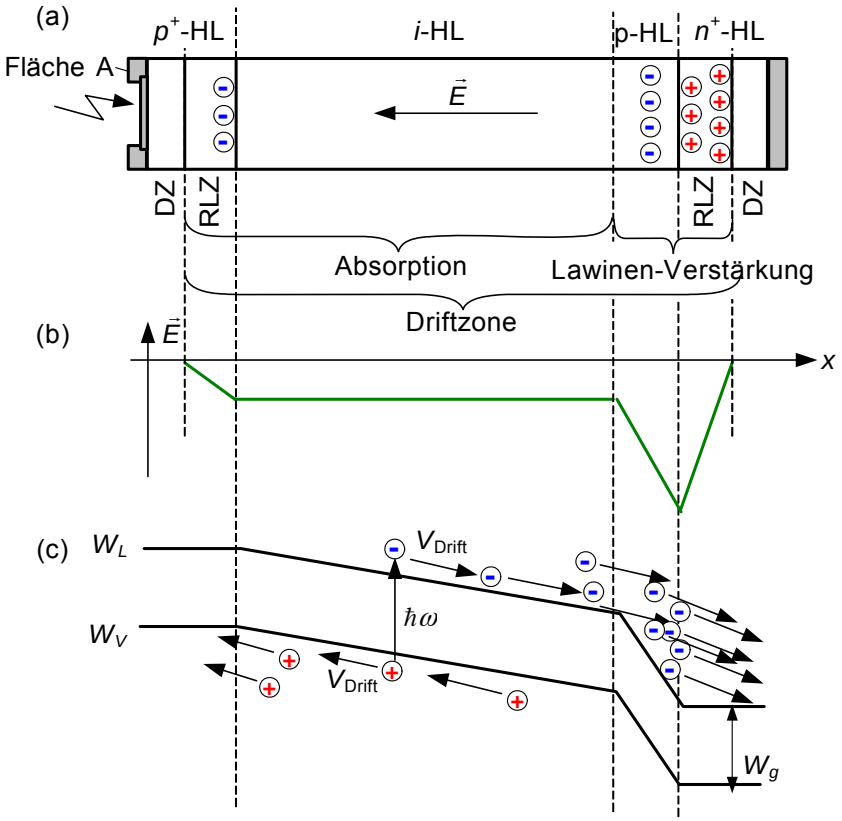


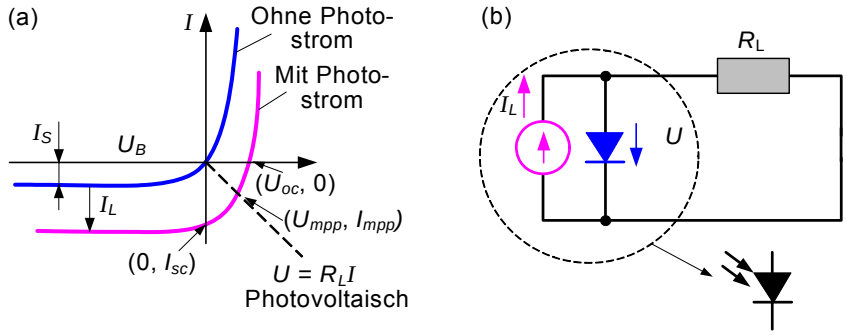
Abbildung 7.21: Lawinenphotodiode. (a) Struktur, (b) elektrisches Feld über der Diode, (c) Energieband und Ladungsträgerakkumulation.

absorbiert. Die Absorptionszone ist dann auch auf der rechten Seite angeordnet und die Dotierung ist so gewählt, dass die generierten Elektronen und Löcher im elektrischen Feld driften. Die Elektronen driften in der Regel schneller als die Löcher, weshalb man vor allem die Elektronen, welche nach rechts driften verstärken will. Auf der rechten Seite findet sich dann auch eine weitere pn^+ -Zone, in welcher beim Sperrbetrieb ein hohes Feld erzeugt wird. Das Feld ist so hoch, dass es in diesem Bereich zum Lawinendruchbruch und dementsprechend zur Lawinenverstärkung kommt. Da auf dieser Seite fast keine Löcher mehr absorbiert werden und lediglich Elektronen in die Lawinenzone driften werden lediglich Elektronen verstärkt. Bei typischen Spannungen von 100 V und Multiplikationsfaktoren von M_0 von 100 dominiert dann auch in der Tat der Elektronenstrom.

Bsp: Eine hochgeschwindigkeits InGaAs APD wie man sie in der Kommunikationstechnik einsetzt, bietet einen Multiplikationsfaktor von ca. 17 bei angelegten Spannungen von 25 V und einer 3dB Bandbreite von bis zu 28 GHz. Mit solchen Photodioden gelingt es bei 40 Gb/s und Durchschnittslichtleistungen von -17 dBm (20μW) eine fehlerfreie Detektion durchzuführen. Das sind dann ca. 600 Photonen pro Bit statt der üblichen 10'000 Photonen pro Bit wie bei direkter Detektion mit einer gewöhnlichen, aber schnellen Photodiode [28]..

7.6 Solarzelle

Wird die Photodiode ohne angelegte Spannung betrieben so verschiebt sich der Arbeitsbereich in den vierten Quadranten von Fig. 7.13. Dieser Arbeitsbereich ergibt sich als Schnittpunkt der Widerstandskennlinie einer Last mit der Diodenkennlinie, wobei die Diodenkennlinie gegenüber der Kennlinie einer pn-Diode nach unten verschoben ist, weil die absorbierten Photonen ja einen



Strom erzeugen. Die Photodiode gibt dann Leistung an den Widerstand R_L ab und die Photodiode wird zur ***Solarzelle (photovoltaic cell).***

Die solare Energiegewinnung kann maximiert werden indem man sowohl den Stromfluß (d.h. die Anzahl der absorbierten Photonen) als auch die erzeugte Spannung maximiert ($P = I\,V$). Der ideal Arbeitspunkt wird Maximum Power Point"(V_{mpp}, I_{mpp}) genannt.

Die solare Stromausbeute kann ferner maximiert werden indem man möglichst das ganze Spektrum der einfallenden Strahlung photovoltaisch nutzt. Das solare Spektrum der auf die Erde einfallenden Strahlung ist allerdings breitbandig. Absorbiert werden jedoch nur diejenigen Photonen deren Energie größer ist als die Bandlückenenergie des Halbleiters. Die Bandlückenenergie von Si ist z.B. bei ca. 1,12 eV (1,1 μm). Die Bandlückenenergie in GaAs ist z.B. bei ca. 1,42 eV (0,9 μm). Dementsprechend werden in diesen Materialien nur Photonen mit einer Wellenlängen, welche kürzer sind als 1,1 μm. respektive 0,9 μm absorbiert. Die langwelligeren Photonen durchdringen das Material und sind für die Energiegewinnung verloren (Fig. 7.22(a)).

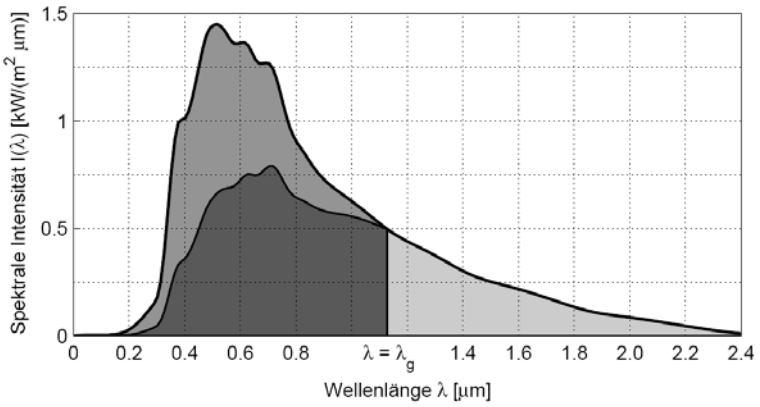
Die solare Spannung, welche pro einfallendem Photon erzielt werden kann entspricht in etwa der Bandlückenenergie ($e\varphi_{Solar} \simeq W_g$) des Halbleiters. Dies kann man so sehen: wenn man den p-Halbleiter und den n-Halbleiter einer pn-Diode stark dotiert so bewegt sich die Fermienergie auf der p-Halbleiterseite gegen das Valenzband und auf der n-Halbleiterseite gegen das Leitungsband. Die über dem Halbleiter liegende Diffusionspannung nähert sich dann der Bandlückenenergie an. Da nun aber die von einem Elektron-Lochpaar erzeugte Spannung gerade der Diffusionsspannung entspricht ist diese in etwa so gross wie W_g/e (Fig. 7.22(b)).

Für die Energiegewinnung ist also weder eine kleine Bandlückenenergie (große Absorbtion und damit großer Strom - aber kleine Spannung) noch eine große Bandlückennergie (kleine Absorption und damit kleiner Strom bei großer Spannung) ideal. Gemäß dem Diagramm in Fig. 7.23 weisen Solarzellen aus Gallium-Arsenid (GaAs) oder CdTe aber auch solche aus amorphem (a-Si) von ihrem theoretischen erreichbaren Wirkungsgrad her das höchste Potential auf.

In der Praxis wird der theoretische Wirkungsgrad nicht erreicht. Die folgenden unvermeidbaren Effekte reduzieren den Wirkungsgrad:

- Rekombinationsverluste: Ein Teil der Elektron-Lochpaare rekombiniert bevor sie einen signifikanten Strom generieren können.
- Reflexionsverluste: Ein Teil des Lichtes wird an der Solarzellenoberfläche reflektiert. Diese Verluste können durch geeignete Antireflexbeschichtungen reduziert werden.
- Beschattung durch Frontelektroden. Die geringe elektrische Leitfähigkeit vieler Halbleiter erfordert eine feingliedrige Leiterstruktur zum Stromabgriff. Die meist lichtundurchlässigen Elektroden stellen eine Selbstbeschattung dar.
- Ohmsche Verluste: Sowohl im Halbleitermaterial als auch in den Stromabgriffsstrukturen entstehen ohmsche Verluste.
- Temperaturabhängige Verluste. Der Wirkungsgrad der Soalrzellen nimmt mit steigender Temperatur ab.

(a)



Energieerzeugung in der Solarzelle durch Energieabsorption:

- geglättete spektrale Intensität in Äquatornähe $I_{AM1.5}$
- A_A absorbierter und strombildender Nutzanteil
- A_W in Wärme umgewandelter Verlustanteil
- A_D nicht absorbierter unwirksamer Strahlungsanteil

(b)

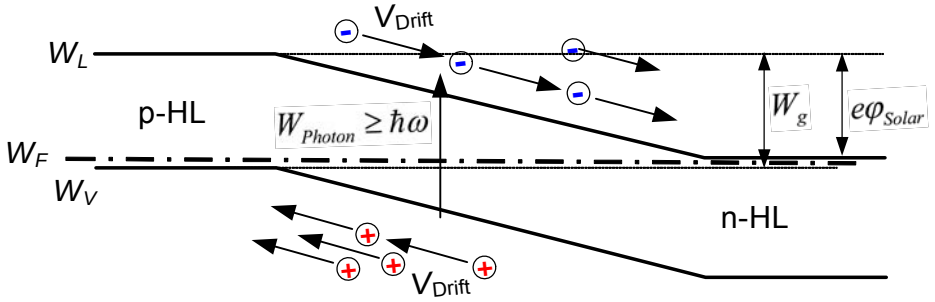


Abbildung 7.22: (a) Solares Spektrum mit absorbiertem strombildendem Anteil. [29] (b) Bandkanten einer hochdotierten pn-Solarzelle. Wenn keine Spannung anliegt ist das Produkt aus Diffusionsspannung mit der Elementarladung fast identisch mit der Bandlückenenergie der Solarzelle.

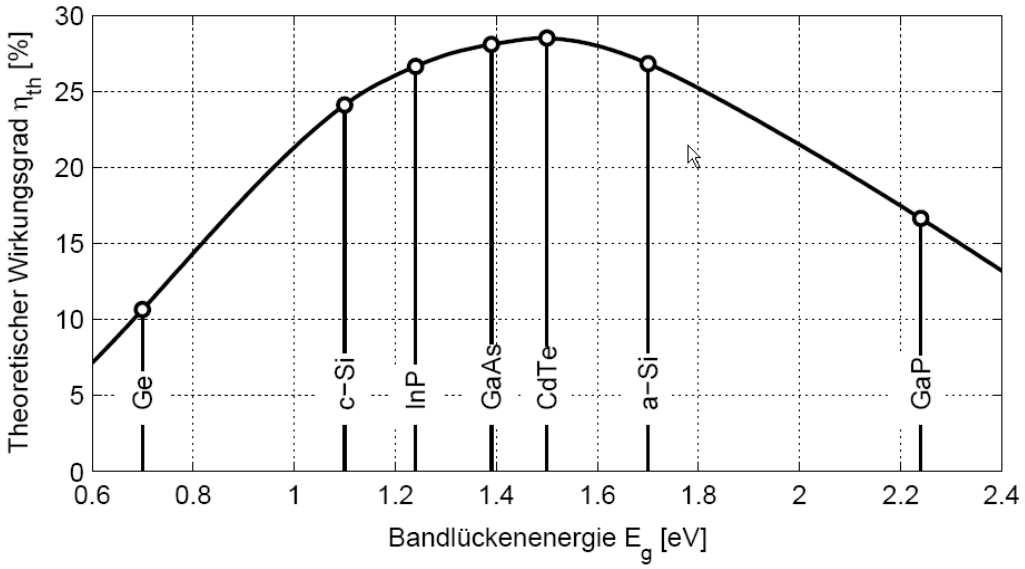


Abbildung 7.23: Theoretischer Solarzellenwirkungsgrad als Funktion der Bandlückenenergie bei einer Temperatur von 25^{o}C. [29].

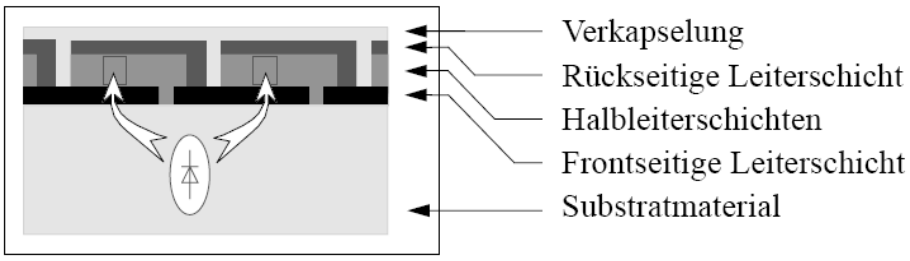
Nachfolgend sind die erfolgreichsten Solarzelltechnologien aufgeführt. Es sind dies:

- *Monokristalline Si-Solarzellen*: ein monokristalliner Silizium-Einkristall dient als Ausgangsbasis um die pn-Photovoltaik-Diode darauf aufzuwachsen. Die Herstellung des Einkristalls in einem Schmelztiegel ist arbeits- und energieaufwändig. Auch enstehen beim Zuschneiden auf eine rechteckige Modulform große Material-Verluste. Allerdings haben diese Solarzellen einen außerordentlich hohen Wirkungsgrad von 12 bis 23%.

- *Poly- oder multikristalline Siliziumsolarzellen*: Als Ausgangsmaterial dient ein polykristalliner Siliziumkristall. Zu dessen Herstellung wird die flüssige Siliziumschmelze zu einem Quader erkaltet und dann in Scheiben geschnitten. Die multikristalline Struktur führt zu Wirkungsgradverlusten. Die Wirkungsgrade bewegen sich zwischen 11 und 15%.

- *Dünnfilm-Solarzellen*: Bei den Dünnfilm-Solarzellen geht es darum ein geeignets pn-Halbleitermaterial zu finden, welches auf einen Träger aufgedampft werden kann. Die aufwändige Wafer-Substratherstellung entfällt. Ein weiterer Vorteil ist, dass die Serieschaltung bei den Dünnfilmverfahren in den Herstellungsprozess integriert werden kann. Das Prinzip der monolithischen Serieschaltung ist in Fig. 7.24 dargestellt. Bei der Herstellung wird in einem ersten Schritt die frontseitige Leiterschicht auf das Substratmaterial aufgebracht. Diese wird danach in Streifen unterteilt. Danach wird die eigentliche Solarzelle als pn-Schichtfolge abgeschieden. Auch diese Schichten werden danach wieder in Streifen unterteilt, welches je nach Verfahren mittels Ätzen oder Laserschneiden erfolgt. Als nächstes wird die rückseitige Leiterschicht aufgebracht, gefolgt von einem nächsten Teilungsvorgang, um die einzelnen Zellen gegeneinander zu isolieren. Aus Schutz- und Haltbarkeitsgründen wird die gesamte Anordnung am Schluss möglichst luft- und wasserdicht verkapselt. Wichtige Dünnfilm-Solarmaterialien sind

  - Amorphe Silizium-Solarzellen (Wirkungsgrade zwischen 6 bis 10%)
  - Cadmium-Tellllurid-Solarzellen (Wirkungsgrade zwischen 7 bis 9%)
  - Kupfer-Indium-(Gallium)-Diselenid-Solarzellen ($CuIn(Ga)Se_2$). Auch bekannt unter dem Kürzel: CIS oder CIGS-Zellen. (Wirkungsgrade zwischen 8 bis 14%)



Monolythische Serieschaltung bei Dünnfilm-Solarzellen

Abbildung 7.24: Monolythische Serieschaltung bei Dünnfilm-Solarzellen.

7.7 Lumineszenzdiode (LED) und Laserdiode (LD)

Während in Solarzellen und Photodioden durch Absorption von Strahlungsenergie Ladungsträger erzeugt werden, wird in Lumineszenz- und Laserdioden die bei der Rekombination von Ladungsträgern frei werdende Energie in Form von Strahlung ausgesandt.

Abstrahlung gibt es allerdings nur wenn opt. Rekombination von Ladungsträgern über opt. Generation dominiert, d.h. wenn $g_{sp} - r_{sp} < 0$. Das bedeutet dann, dass man einen Zustand erzeugen muss, in welchem es in den Leitungsbändern weit mehr angeregte Elektronen und im Valenzband weit mehr Löcher gibt, als es die Fermiverteilung im thermischen Gelichgewicht ergeben würde. Man spricht von einem Zustand mit ***Populationsinversion.*** Populationsinversionen werden üblicherweise durch Anlegen eines Stromes oder durch starke optische Bestrahlung erziehlt.

Im Sinne eines ersten Versuches eine LED oder LD zu konstruieren, betrachten wir die pn-Diode in Abb. 7.25(a). Benutzt man die pn-Diode in Flusspolung um Elektronen und Löcher durch Anlegen eines Stromes zu injizieren, so gibt es am pn-Übergangsinterface tatsächlich eine Zone mit einer cairca gleich starken Population von freien Elektronen im Leitungsband und freien Löchern im Valenzband. Allerdings werdem die Elektronen und Löcher relativ rasch durch die Zone driften und im Übrigen aufgrund von Störstellenrekombination (SRH) rekombinieren, so dass die Ladungsträgerdichte nicht besonders groß sein wird. Wegen der kleinen Ladunsgrägerzahl ist es deshalb i.A. schwierig eine signifikante spontane Emissionen zu erzeugen.

Die pin-hetero Struktur in Fig. 7.25(b) ist weit besser für die Konstruktion einer LED geeignet. In Flusspolung werden Löcher injiziert und nach rechts transportiert. Umgekehrt fließt ein Elektronenstrom von links nach rechts. In der Mitte befindet sich eine intrinsische Zone mit einem Material mit kleinerer Bandlücke. Da die Löcher/Elektronen bevorzugt niedrige Energienveiaus im VB/LB besetzen (Regel: Löcher steigen, Elektronen sinken) werden die beiden Ladungsträger sich bevorzugt in den Bereich niedriger Bandenergie bewegen. Sind diese erst einmal in der Zone haben sie dann auch nicht mehr die notwendige Energie um die Offset-Bandlücke zu überwinden. Die beiden Ladungsträger akkumulieren deshalb in dieser Zone stark. Die Ladungsträger bleiben deshalb in dieser Zone bis sie über einen Rekombinationsprozess zerfallen. Falls die Population der angeregten Träger genügend groß ist, kommt es zu einer beachtlichen spontanen oder sogar stimulierten Emission.

Durch geschickte Materialkombinationen von Materialien mit einer Gitterkonstante kann man Materialsysteme mit Heterostrukturübergängen wachsen. Das Wachstum gelingt in der Regel nur, wenn die Gitterkonstanten der verschiedenen Materialien gleich oder fast gleich sind. Bild 7.26 zeigt den Verlauf der direkten Bandlücke und damit die Emissionswellenlänge in verschiedenen Materialsystemen als Funktion der Zusammensetzung und der Gitterkonstante. In Fig. 7.26 könnte man also grundsätzlich Materialien entlang einer senkrechten kombinieren.

Für eine effiziente Lichtemission werden vorteilhafterweise ***Halbleitermaterialien mit einer***

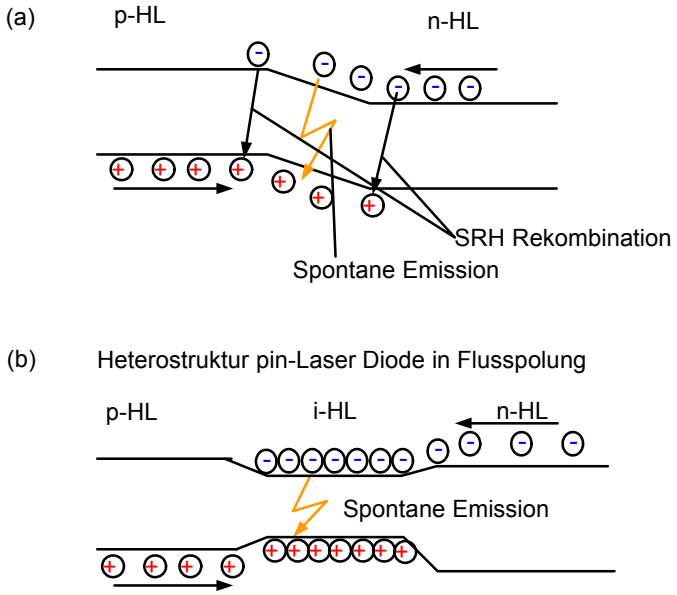


Abbildung 7.25: (a) Schwache spontane Emission in der pn-Diode, (b) Starke spontane Emission in der heterostruktur pin-Diode. Dank der kleineren Bandlücke akkumulieren die Ladungsträger und weil die Zone intrinsich ist, gibt es weniger nicht-strahlende Rekombination und die Lebensdauerlebensdauer ist größer was zu einer weiteren Akkumulation der Ladungsträger führt.

direkten Bandlücke benötigt. Bild 7.27 (a) zeigt einen direkten Halbleiter wie z.B. GaAs und andere III-V- und II-VI-Verbindungshalbleiter. In Halbleitern mit indirektem Übergang ist ein weiterer Partner erforderlich, um die Impulsänderung zu ermöglichen. Dies kann durch Phononen (Gitterschwingungen) erfolgen, die bei kleinen Energiewerten einen großen Impuls aufnehmen können. Dieser Dreiteilchenprozeß ist allerdings sehr unwahrscheinlich. Die Wahrscheinlichkeit für strahlende Übergänge kann jedoch durch Rekombinationszentren wesentlich erhöht werden. Wird beispielsweise ein Ladungsträger durch ein geeignetes Dotierungsatom eingefangen, so ist es sehr scharf lokalisiert, und sein Impulswert erstreckt sich wegen der Unschärferelation über einen sehr breiten Bereich, so dass dann strahlende Rekombination möglich ist. Die Strahlung entspricht hier dem Energieunterschied zwischen "Term" und Band, Bild 7.27 (b). So ist zum Beispiel durch Substitution des Phosphoratoms in GaP (Bandabstand 2,26 eV) durch Stickstoff eine Emission im grünen Spektralbereich und durch ZnO-Komplexen (Zn und O auf benachbarten Ga- und P-Plätzen) Emission im roten Spektralbereich möglich.

Bild 7.28(a) und (b) zeigt zwei mögliche Geometrien für Lumineszenz- und Laserdioden.

Lumineszenzdiode

Zum Betrieb einer ***Lumineszenzdiode (LED)*** genügt es für Populationsinversion zu sorgen und sicherzustellen dass, spontane Emission möglich ist. In der LED erfolgen dann Rekombinations- und damit Emissionsvorgänge spontan und damit unabhängig voneinander. Die resultierende Strahlung ist deshalb inkohärent und je nach Energieübergang vom Leitungs- zum Valenzband verschieden. Die typischen Linienbreite ΔE in Energie bzw. $\Delta\lambda$ in der Wellenlänge sind von der Größenordnung $\Delta E = kT = 26$ meV bei $T = 300$ K, da die Energie der Ladungsträger an den Bandkanten über diesen Bereich verschmiert ist, Bild 3.3. Es gilt

$$W = \frac{hc}{\lambda} = \frac{1,24 \text{ eV}}{\lambda/\mu\text{m}} \tag{7.32}$$

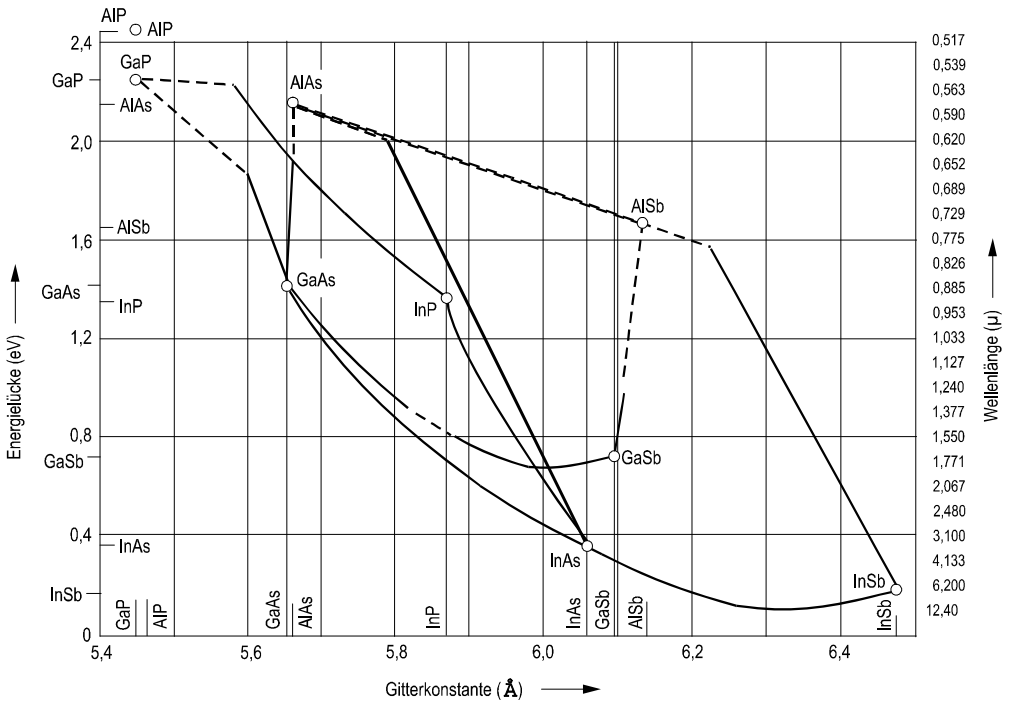


Abbildung 7.26: Verlauf von Bandlücke und Gitterkonstante für einige, in der Optoelektronik oft verwendete, III-V Verbindungshalbleiter. (—) Direkter Übergang; (- - -) indirekter Übergang.

bzw.

$$\Delta W = \frac{1,24 \text{ eV}}{\lambda^2/\mu\text{m}} \Delta\lambda. \tag{7.33}$$

Somit folgt für die spektrale Breite des von einer Lumineszenzdiode bei $\lambda \sim 0,8\ \mu$m, $\Delta\lambda \sim 13$ nm.

Laserdiode

Zum Betrieb eines Lasers müssen drei Bedingungen erfüllt sein.

- Populationsinverison muss vorherschen
- Ein Medium, welches spontane Emission ermöglicht muss vorliegen (in der Regel bedingt das ein Medium mit direkter Bandlücke) und
- das Medium muss in einem Resonator liegen.

Damit kann man eine LED in eine ***Laserdiode (LD)*** umbauen, indem man diese in einen Resonator stellt.

Für Licht können zwei auf sich selber abbildende Spiegel einen Resonator bilden. Resonatoren haben dabei zwei Funktionen. Sie sorgen A) dafür, dass nur Licht einer bestimmten Wellenlänge sich konstruktiv überlagern kann und B) sie sorgen für eine selektive Verstärkung (induzierte Emission) dieser einen Wellenlänge.

A) Platziert man nämlich einen direkten Halbleiter mit Populationsinversion zwischen zwei Spiegel, so kann nur das Licht hin-und herreflektierenden, welches sich konstruktiv überlagert.Bei einer geschickten Wahl der Spiegel gelingt es damit ein sehr schmalbandiges monochromatisches Laserlicht zu erzeugen. Dabei werden Linienbreiten von weniger als 0,00001 nm erzielt. Dies entspricht in etwa einer Linienbreite von 10 MHz und weniger.

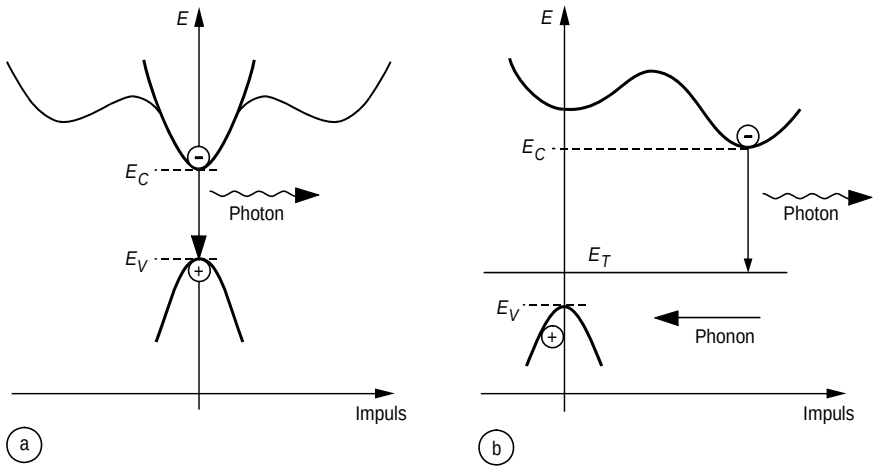


Abbildung 7.27: (a) direkter und (b) indirekter Übergang.

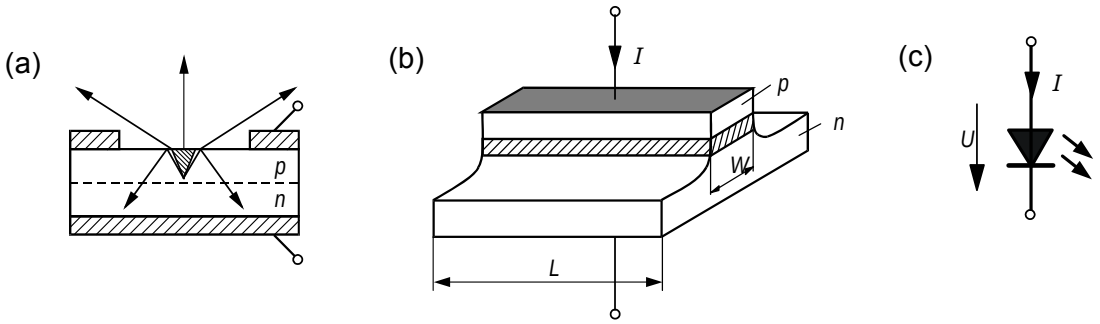


Abbildung 7.28: Zwei mögliche Geometrien für LED und LD Dioden. (a) VCSEL-Geometrie (Vertical-Cavity Surface-Emitting Laser), (b) Wellenleiter-Geometrie, (c) Schaltsymbol für LED oder LD

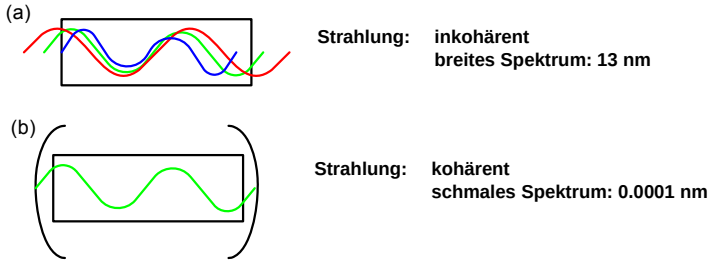


Abbildung 7.29: (a) Strahlung ohne Resonator, (b) Selektion und resonante Verstärkung einer Frequenz, Richtung und Phasenlage in einem Resonator.

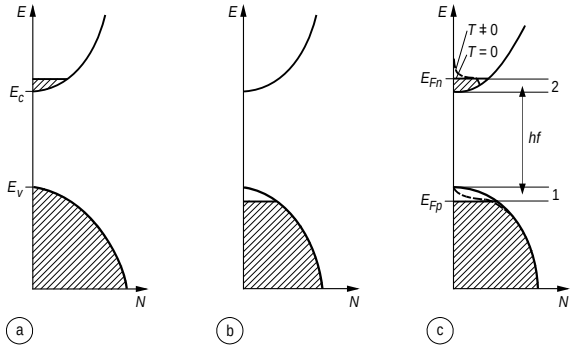


Abbildung 7.30: Besetzung der verfügbaren Zustände durch Elektronen (schraffierte Bereiche besetzt). (a) n-entartet, (b) p-entartet, (c) invertiert.

B) Die Spiegel bewirken aber noch mehr. Während bei Lumineszenzdioden der spontane Emissionsvorgang dominiert, wird die Laserdiode von der induzierten Emission beherrscht. Induzierte Emission basiert auf der Bose-Natur des Lichts. Boseteilchen nehmen nämlich vorwiegend identische Zustände (Richtung, Freuquenz und Phasenlage) ein. In der Praxis bedeutet dies, dass ein angeregter Zustand bevorzugt Licht in einen bereits existierenden Zustand emittiert. Falls man einmal ein Photon in einem bestimmten Zustand hat, dann stimuliert dieses Photon bevorzugt Emissionen in den genau gleichen Zustand. Wir sprechen dann von stimulierter Emission. Das dabei erzeugte Laserlicht ist zeitlich und räumlich ***köhärent*** (*gl. Richtung, gl. Frequenz und gl. Phasenlage aller Photonen*). Das erste spontan emitierende Photon, welches von den Spiegeln zurückreflektiert wird und die Resonanzbedingung erfüllt kann damit eine ganze Kaskade von Emissionen anregen, welche alle kohärent zu dem ersten Photon sind.

Mit solchen Laserquellen ist es heute im Labor bereits möglich. Daten mit einer Rate von mehreren Terabit/s pro Glasfaser zu übertragen. Solche Laserdioden werden in der Unterhaltungs- und Kommunikationstechnik kommerziell genutzt.

Bem. zu *Populationsinversion*: Um Laseremission in Gang zu setzen, muss die Anzahl der stimmuliert erzeugten Photonen mindestens der Anzahl der verlustig gegangenen Photonen bei einmaligem Durchlauf durchs Medium sein. Im Moment wo dieses Gleichgewicht erreicht oder gar überschritten wird, sagt man von dem Medium, dass deren Ladungsträgerpopulationen invertiert sind, Bild 7.30.

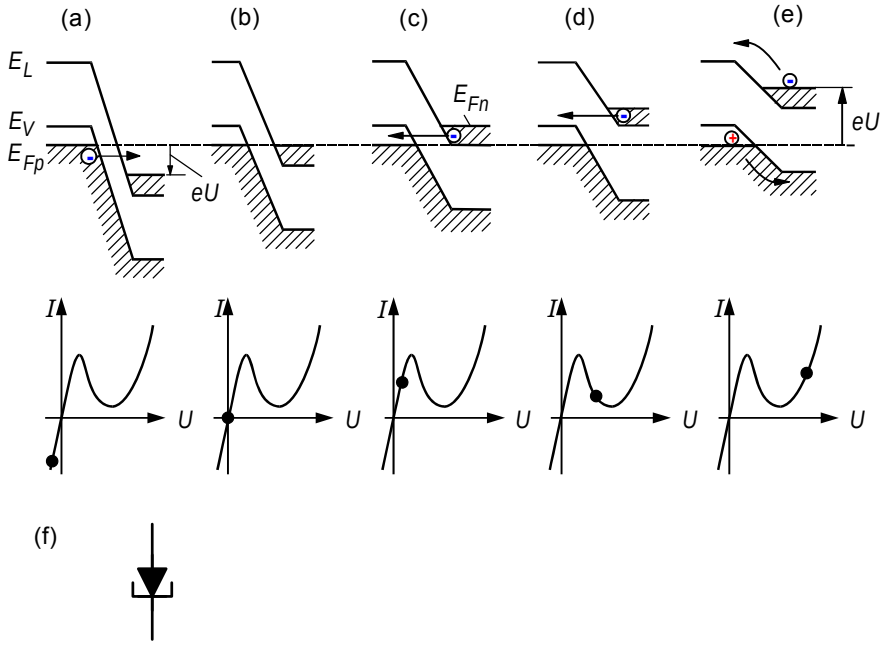


Abbildung 7.31: (a) Bänderschema der Tunneldiode und (b) zugehörige Punkte auf der Kennlinie [3]. (c) Symbol der Tunneldiode.

7.8 Mikrowellendioden

7.8.1 Tunneldiode

Die ***Tunneldiode*** ist durch einen sehr hochdotierten pn-Übergang, d.h. durch einen $p^{++}n^{++}$-Übergang charakterisiert. Es liegt auf beiden Seiten Entartung vor. Dadurch ist die RLZ extrem klein und die Fermi-Niveaus auf beiden Seiten liegen in den Bändern, siehe Bild 7.31. Hierdurch wird erreicht, dass den Ladungsträgern auf der einen Seite freie Plätze auf der anderen Seite zur Verfügung stehen. Es werden dadurch Tunnelprozesse ermöglicht und damit ein guter Stromtransportmechanismus geschaffen. Dieser Zustand ist für $U = 0$ in Bild 7.31 (b) dargestellt und gilt auch für kleine Spannungen im Durchlaß- (c) und Sperrbereich (a). Der Tunnelstrom nimmt ab, wenn den Ladungsträgern keine freien Plätze mehr zur Verfügung stehen (d). Der Bereich fallender Kennlinie entspricht einem negativen differentiellen Widerstand. Dieser kann dann z.B. genutzt werden um die Verluste in einem Schwingkreis überzukompensieren, wodurch eine selbsterregte Oszillation aufrecht erhalten werden kann. Am tiefsten Punkt, zwischen (d) und (e) geht der Strom nicht gegen Null, weil ein Tunnelstrom fliesst. In (e) gelangt man wieder in den Bereich der normalen Diodenkennlinie.

Diese Dioden finden in der Mikrowellentechnik zur Schwingungserzeugung Anwendung.

7.8.2 Lawinen-Laufzeit- oder IMPATT-Diode

Die ***Lawinen-Laufzeit-*** oder ***IMPATT-Diode (Impact Avalanche Transit-Time) Diode*** beruht, wie der Name bereits ausdrückt, auf zwei physikalischen Effekten: Der Lawinenmultiplikation der Ladungsträger durch Stoßionisation nach Abschnitt 6.4.3 und einer Generations- und Driftzeit bedingten Laufzeitverzögerung zwischen angelegter Spannung und influenziertem Strom einer in Sperrrichtung gepolten Diode.

Bild 7.32 zeigt die von Read vorgeschlagene p^+nin^{+-}Diodenstruktur, das Dotierprofil, den Feldstärkeverlauf, sowie den Verlauf der Stoßionisationsrate α nach Bild 7.32. Die IMPATT Diode arbeitet wie folgt: Einer in Sperrrichtung betriebenen IMPATT Diode wird eine Hochfrequenz-

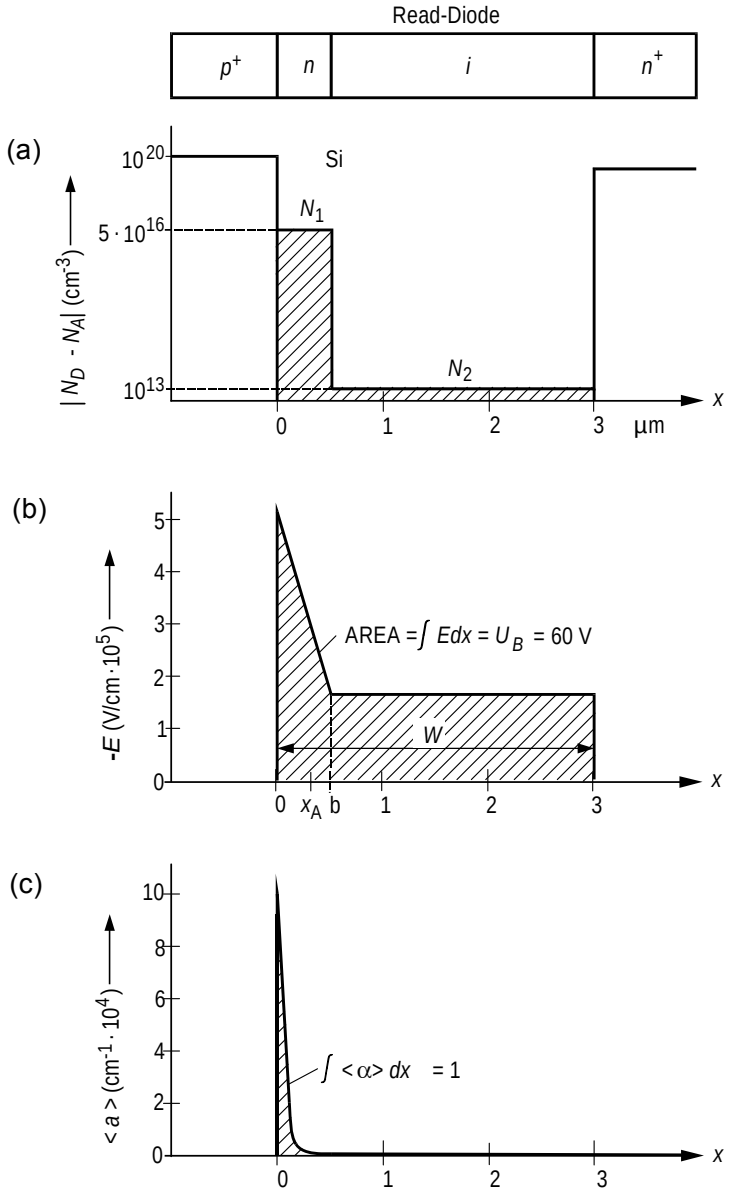


Abbildung 7.32: Von Read vorgeschlagene Struktur der Lawinen-Laufzeit-Diode, (a) Dotierung, (b) Verlauf der elektrischen Feldstärke und (c) Stoßionisationsrate [5].

modulation überlagert. Wenn das Feld über der Diode am stärksten ist kommt es kurzfristig zum Lawinendurchbruch und es werden Elektronen- und Löcherpaare generiert am Interface p^+n. Die Löcher rekombinieren fast unmittelbar in der nahe gelegenen p-HL Diffusionszone. Die Elektronen hingegen driften im elektrischen Feld zur n-HL Seite. Während der Drift influenzieren sie einen Strom. Wichtig ist nun die Laufzeitverzögerung zwischen angelegter Spannung und Strom. Diese kommt durch drei Effekte zustande:

- Dem Umstand, dass der Lawinendurchbruch erst bei Vorliegen der Spitzenspannung erfolgt und
- der Tatsache, dass der Aufbau der Ladungsträgerkonzentration durch Trägermultiplikation Zeit benötigt und
- und letztlich, der Tatsache, dass die totale Laufzeit der Ladungsträger zwischen den Elektroden eine Verzögerung des Stromverlaufs im Vergleich zum Spannungsverlauf bewirkt (Man beachte, dass man die Diode um einen Offset herum in Sperrspannung betreibt, so dass das elektrische Feld in der nin^+ Zone immer anliegt und die generierten Ladungsträger selbst im Wellental einer angelegten Spannung zur n^+ Zone driften.)

Durch geeignete Kombination der Effekte kann dadurch eine Phasenverschiebung von mehr als 90^o zwischen Strom und Spannung entstehen. Das entspricht einem negativen Realteil der HF-Impedanz, der wie bei der Tunneldiode zur Verstärkung von Signalen oder zur Entdämpfung eines Schwingkreises verwendet werden kann.

Die Anzahl der von einem Ladungsträger pro Weglänge erzeugten Elektronen-Lochpaare, ist eine Funktion der Feldstärke E

$$\alpha \sim \alpha_0 \exp\left[-\frac{E_I}{E}\right], \quad \alpha_0, E_I = const \tag{7.34}$$

Im Allgemeinen ist die Ionisationsrate für Löcher und Elektronen verschieden. Der Einfachheit halber verwenden wir einen Mittelwert α. Der Lawinendurchbruch ist erreicht wenn

$$\int_0^w \alpha dx = 1 \tag{7.35}$$

erreicht ist. Das bedeutet, dass jedes Ladungsträgerpaar beim Durchlaufen der RLZ im Mittel ein weiteres Elektron-Lochpaar erzeugt.

Bild 7.33 zeigt die einer stationären Sperrspannung U_0 überlagerte Wechselspannung $u(t)$ und die dazugehörigen injizierten Lawinenladungen an der Stelle $x = 0$ (gestrichelt) und den durch die Anschlußklemmen laufenden Influenzstrom infolge der Drift der Elektronen durch die RLZ im undotierten Gebiet (gerade Linie). Die Grundwelle von Strom und Spannung sind hier um π phasenverschoben, was einer Impedanz mit negativen Realteil entspricht bzw. einem negativen Widerstand entspricht.

Die IMPATT-Diode ist eine der leistungsfähigsten Mikrowellenoszillatorenquelle für Frequenzen bis zu 300 GHz. Im Bereich bis zu 20 GHz sind Leistungen bis zu 300 W, bzw. Spitzenleistungen die im kW Bereich liegen, möglich. Sie ist damit eine der leistungsfähigsten Quellen für Mikrowellenenergie. Genutzt wird sie in mikroelektronischen Schaltungen, bei denen hochfrequente Energiequellen gebraucht werden. So wird sie in der Nachrichtentechnik beispielsweise für Sender in der Millimeterwellenkommunikation wie man sie für Radar in der zivilen Luftfahrt oder zur Steuerung von Raketen im militärischen Bereich und ähnliche Anwendungen verwendet. Vorteile der IMPATT-Dioden sind, dass sie günstig in der Herstellung sind, einen geringen Stromverbrauch haben, zuverlässig arbeiten, durchgehend hohe Wellenenergie liefern und wenig wiegen. Nachteilig ist das hohe Rauschen, die Empfindlichkeit des Elements unter Arbeitsbedingungen und die hohen Reaktanzen. Die Reaktanzen sind stark abhängig von der Oszillationsamplitude und müssen daher im Schaltungsentwurf mit sehr viel Bedacht berücksichtigt werden, damit es nicht zu Verstimmungen oder gar zum Durchbrennen der Diode kommt.

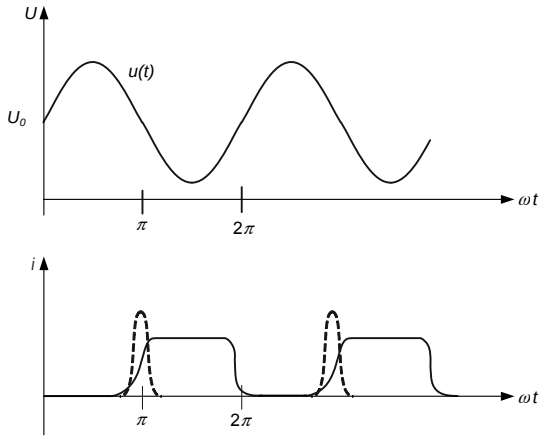


Abbildung 7.33: Wechselspannung und Wechselströme der Read-Diode als Funktion der Zeit: - - - Injizierter Lawinen-Strom bei x=0; —– Influenzstrom in den Anschlußklemmen [3].

7.8.3 Gunn-Diode

Eine Gunn-Diode besteht nur aus n-dotierten Halbleiterbereichen (n^{++}-n-n^{++}). Über ein Material mit negativem differenziellen Widerstand wird eine Art Falle für Elektronen aufgebaut, der in einem Schwingkreis dazu führt, dass Elektronen sich aufstauen und in Schüben (wie Wellen) durch die Diode wandern. Dies geschieht sehr schnell. Die Gunndiode erzeugt Frequenzen von 1,5 - ~10.000 GHz = 10 THz. Die Leistung ist dabei ziemlich groß. Man kann unter Umständen 200 - 300 mW mit einem Gunn-Oszillator erreichen, der aus nur wenigen Bauteilen besteht.

Die n^{++} Dotierungen dienen dazu den ohmschen Widerstand zum Metallkontakt klein zu halten. Wenn über der Diode eine Spannung anliegt, so fällt diese in der schwächer dotierten n-Zone ab (weniger Ladungsträger - und damit größerer Widerstand.). In der n-Zone wählt man nun das Material und die elektrische Feldstärke so, dass ein negativer differentieller Widerstand (wie bei der Tunneldiode) auftritt. (Siehe die feldabhängige Geschwindigkeit der Ladungsträger bei GaAs und insbesondere der Sättigungsgeschwindigkeit in Bild 4.5). Wenn nun eine kleine Ladungsträgerdichtestörung von Elektronen auftritt, so fällt das Feld über dieser Ladung ab. D.h. dann, dass das elektrische Feld für die Elektronen an der vorderen Kante größer ist, als an der nachfolgenden Flanke. Wegen des differenziellen Widerstandes führt das zu einer verlangsamten Drift der vorderen gegenüber den nachfolgenden Elektronen. Dies führt zur Bildung einer Raumdomäne von Elektronen um einen Punkt, welcher sich mit der Sättigungsgeschwindigkeit zur Anode bewegt - also einen Elektronen "Gun-Shot". In der Praxis wandern dann z.B. in einem Stück GaAs (beim Übergang von direktem zu indirektem Elektronenband) , welches um ein Gleichfeld im Bereich der negativen differentiellen Beweglichkeit angelegt ist, Bild 4.5 Raumladungsdomänen mit der Sättigungsgeschwindigkeit von etwa $v_s = 10^7$ cm s^{-1} von der Kathode zur Anode, Bild 7.34.)

In einer etwas ausgewählter formulierten Erklärung würde man den Vorgang wie folgt beschreiben: Hängt die Trägerdichte nicht von der Spannung ab, so führt die negative differentielle Beweglichkeit $\mu = dv/dE$ in GaAs bei hohen Feldstärken zu einer negativen differentiellen Leitfähigkeit, $\sigma_{\text{diff}} < 0$. Die negative differentielle Leitfähigkeit führt dazu, dass Ladungsträgerdichte-Störungen gemäß Abschnitt 5.4.1 eine negative dielektrische Relaxationszeit $\tau = -\varepsilon/|\sigma_{diff}|$ erfahren, d..h. instabil sind.

Damit das Gunn-Element auch oszilliert, muss man es in einem Resonanzkreis anordnen. Wird das Gunn-Element also so betrieben, dass es einen negativen differentiellen Widerstand aufweist, siehe 7.35(a), so kann es in einem Schwinkreis oszillieren, siehe 7.35(b). Dies sieht man wie folgt. Nach Kirchhoff gilt für den Schwingkreis

$$0 = di(t)/dt\ L + i(t)R + \int i(t)/C\ dt\ .$$

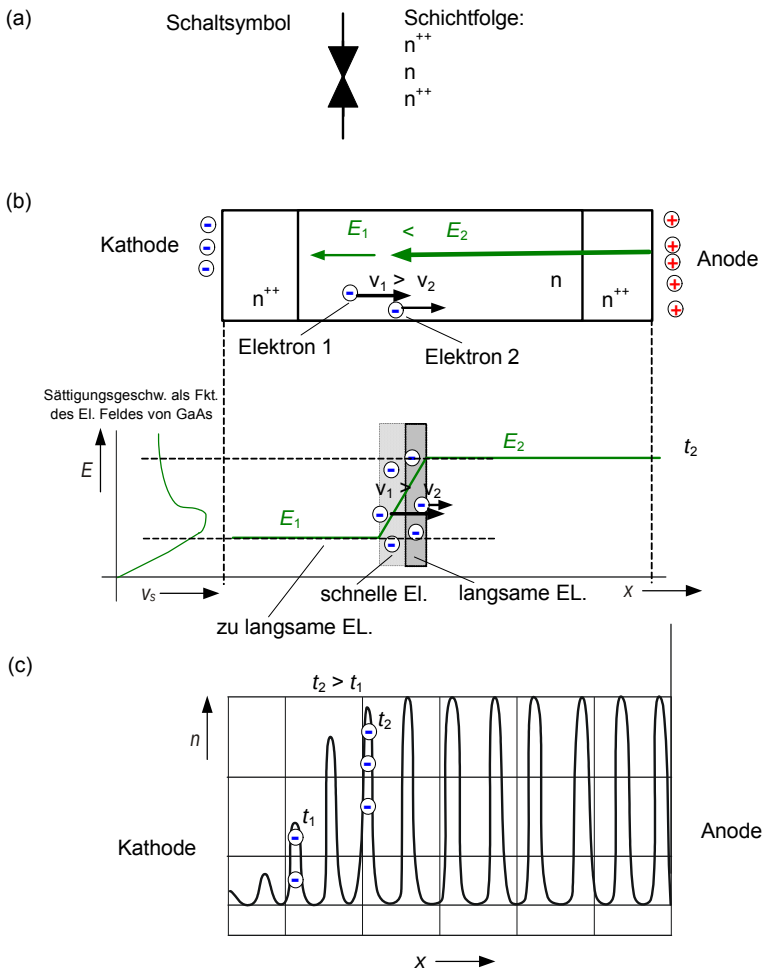


Abbildung 7.34: (a) Schaltsymbol und Dotierung der Gunn-Diode, (b) Verlauf des el. Feldes im Gunn-Element infolge einer anwachsenden Raumladungsstörung. (c) Domänensättigung und Wanderung.

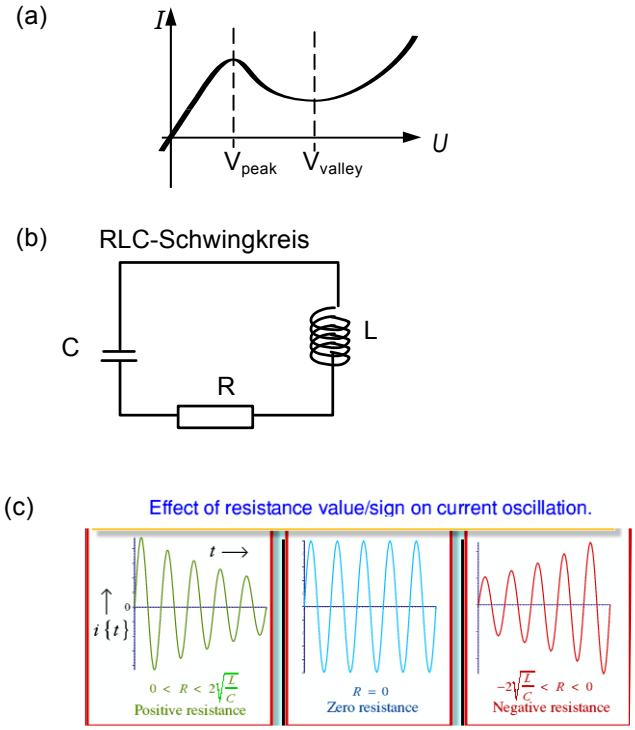


Abbildung 7.35: (a) Bauteil mit negativem Differentiellem Widerstand, (b) RLC-Schwingkreis, (c) Lösungen für den Schwingkreis in Abhängigkeit des differentiellen Widerstandes.

Die Lösungen dazu sind

$$i(t) = e^{At} \text{ mit } A = \frac{-R \pm \sqrt{R^2 - 4LC}}{2L} \quad \text{falls} \quad R^2 \geqq 4LC$$
$$i(t) = e^{\alpha t} e^{\omega t} \text{ mit } \alpha = -\frac{R}{2L} \text{und } \omega = \frac{\sqrt{R^2 - 4LC}}{2L} \quad \text{falls} \quad R^2 < 4LC$$

Damit kommt kann es im bei genügend starkem negativen differentiellen Widerstand zu einer Oszillation gemäß der 2.Lösung kommen.

Gunn-Elemente sind nicht so leistungsstark wie IMPATT-Dioden, dafür zeigen Sie aber ein geringeres Rauschen, da der Prozeß auf keinem Lawinenprozess beruht Die Gunn-Diode besteht nur aus n-dotierten Halbleitern, meist GaAs (Galliumarsenid), GaN (Galliumnitrid) oder Indiumphosphid. Sie ist relativ billig und wird in vielen Bereichen angewendet (Mikrowellensender, kleinen Radaranlagen der Polizei oder bei Türöffnern im Supermarkt, Amateurfunk).

7.8.4 Stop-Recovery Diode (Speicher Varaktor)

Speichervaraktoren (step-recovery diodes) nutzen die Ladungsspeicherung im Flußbetrieb von pin-Dioden. Bei sinusförmiger Ansteuerung der Diode wird periodisch zwischen Fluß- und Sperrbetrieb umgeschaltet. Da die in der Diode gespeicherte Ladung zu Beginn einer negativen Halbwelle erst abgebaut werden muss, fließt auch nahe dem Nulldurchgang der Spannung - wie wir in Kap. 6.5 gesehen haben - zunächst noch ein Strom. Dieser ist hauptsächlich durch den Abbau der Diffusionsladung bestimmt. Ist diese abgebaut, so verläuft der Sperrstrom sehr schnell gegen null - bedingt durch die geringe Sperrschichtkapazität der pin-Diode. Dieser scharfe Übergang führt zu einem deutlichen Oberwellenanteil im Strom bzw. im Spannungsabfall, siehe Abb. 7.36.

Damit die Step-Recovery-Dioden funktionieren, sollte die Arbeitsfrequenz deutlich größer als der Kehrwert der Ladungsträgerlebensdauer sein, da die Ladungsträger andernfalls vorher rekombinieren und so keinen abrupten Übergang erzeugen. Bei Ladungsträgerlebensdauern in der

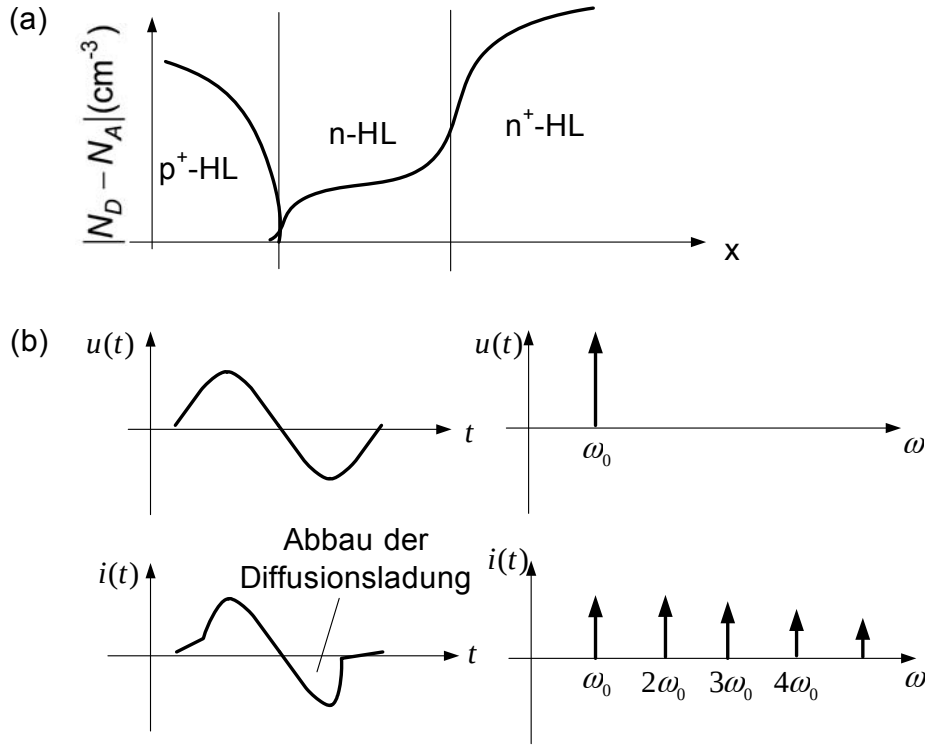


Abbildung 7.36: (a) Dotierprofil und (b) Ausschaltverhalten bei sinusförmiger Ansteuerung.

Größenordung von 10 ns ergeben sich untere Grenzfrequenzen von 100 MHz. Step-recovery Dioden sind nach oben durch die Abfallzeit bei Einsetzen des Sperrstromes limitiert. Diese sind in der Größenordnung von 100 ps und limitieren die Diode also auf Frequenzen unterhalb von 10 GHz.

Step-Recovery-Dioden werden vor allem als Frequenzvervielfacher (Frequency-Comb Generators, zu deutsch Frequenzkamm Generator") eingesetzt.

Kapitel 8

Bipolartransistoren

Bipolar-Bauteile **(bipolar junction devices)** sind Halbleiterbauteile, in welchen sowohl p- als auch n-dotierte Zonen für den Ladungsträgertransport benötigt werden. Das unterscheidet sie von den sogenannten ***unipolaren Bauteilen,*** in welchen nur eine dotierte Halbleiterzone benötigt wird.

Der Bipolartransistor, Fig. 8.1, war das erste Festkörper-Verstärkerelement, das praktische Anwendung gefunden hat. Von ihm ging die beispiellose Entwicklung der Mikroelektronik aus. Die Wirkungsweise des Transistors beruht auf der Möglichkeit, in einen Halbleiter (Basis) Minoritäten aus einem Emitter (Kathode bei der Röhre) zu injizieren und in einem Kollektor aufzufangen (Anode bei der Röhre) 8.1. Der Stromfluss hängt empfindlich von einer Steuerspannung ab. Beim Transistor der Emitter-Basis-Spannung, bei der Röhre der Gitter-Anoden-Spannung.

Diese Möglichkeit, den von der Röhre bekannten Verstärkungsmechanismus als Festkörperbauelement zu realisieren wurde erstmals 1947 von J. Bardeen, W. Brattain und W. Shockley demonstriert, siehe Bild 8.1 und 8.2 (J. Bardeen: Elektro-Ingenieur und zweifacher Nobelpreisträger für die Miterfindung des Transisitors und der BCS-Theorie für Supraleiter gemeinsam mit Cooper und Schrieffer).

Das Jahr 1947 gilt deshalb als das Geburtsjahr der Mikroelektronik. Die ersten Transistoren wurden aus dem Halbleiterwerkstoff Germanium hergestellt. Erst später wurde verstärkt - und heute fast ausschließlich - der Halbleiterwerkstoff Silizium eingesetzt. Allerdings nicht deshalb, weil Silizium bessere physikalische Eigenschaften als Germanium aufweist (im Gegenteil, die Ladungsträgerbeweglichkeiten sind in Ge höher als in Si, siehe Anhang, sondern allein wegen der Möglichkeit, die Siliziumoberfläche durch das natürliche Oxid SiO_2 (Quarz) so zu passivieren, dass der Halbleiter von äußeren Einflüssen geschützt ist. Erst die Siliziumtechnik hat die Planartechnik ermöglicht, die wiederum Voraussetzung für die Entwicklung der integrierten Schaltungen war. Siliziumoxid ist auch das Dielektrikum, für den noch später zu behandelnden Metall-Oxid-Halbleiter-Feldeffekttransisitor (MOSFET) und MOS-Kondensatoren.

Bipolartransistoren werden vor allem dort eingesetzt, wo hohe Ströme, kurze Schaltzeiten und hohe Frequenzen erreicht werden sollen, während MOSFETs bevorzugt für leistungsarme Schaltungen eingesetzt werden, in denen eine hohe Bauelementdichte angestrebt wird (z.B. Speicher). Besondere technische Ausführungen des MOSFET erlauben auch die Herrstellung von Leistungs-MOSFETs. Technologische Verbesserungen führen dazu, dass MOS-Schaltungen immer schneller werden und den Bipolarschaltungen nahe kommen. In jüngster Zeit ist es auch gelungen, die technologischen Prozesse so zu führen, dass bipolare und MOS-Bauelemente auf derselben Siliziumscheibe integriert werden können. Dadurch können die Vorteile beider Techniken genutzt werden. Beispielsweise werden die interne Logik einschließlich Speicher in MOS-Technik, die Ausgangsstufen mit großem Strombedarf in Bipolartechnik aufgebaut und vieles mehr.

Der Begriff „**Transistor**“ ist eine Kurzform für eine der englischen Bezeichnungen Transfer Varistor, Transformation Resistor oder Transfer Resistor, also einen durch Spannung oder Strom steuerbaren elektrischen Widerstand.

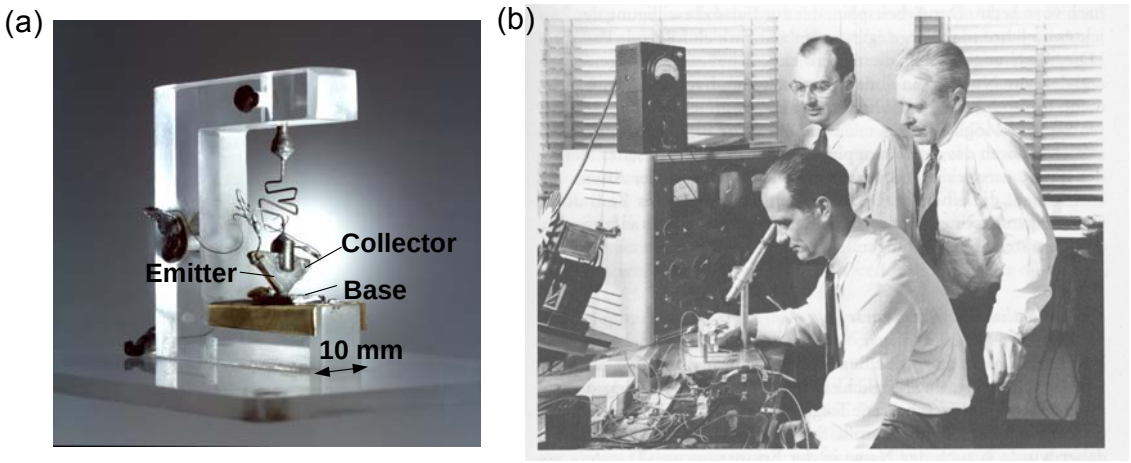


Abbildung 8.1: (a) Der erste Germaniumtransistor mit Punktkontakten. Auf der Unterlage liegt das Scheibchen aus Germanium, darauf wird das Plexiglasdreieck mit den Stromzuführungen gedrückt [19]. (b) Die Erfinder des Transistors und Nobelpreisträger William Shockley (sitzend), dahinter John Bardeen, rechts Walter Brattain [19].

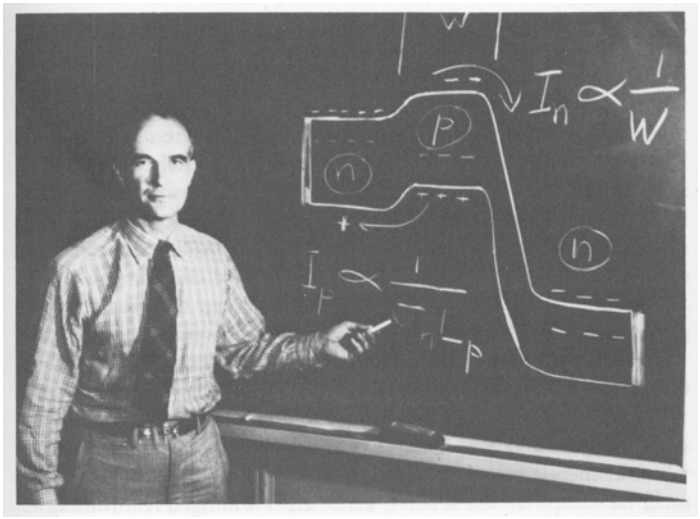


Abbildung 8.2: Shockley, von den Bell Telephone Laboratories zeigt die Wirkungsweise des von ihm erfundenen Bipolar-Transistors. Aus dem Emitter (linke Zone mit "n" bezeichnet) treten Elektronen aus, durchqueren die mit "p" bezeichnete dünne Zone der Basis, wie der obere Pfeil andeuten soll, und erreichen dann die rechte Zone "n", den Kollektor, wo die Elektronen gesammelt werden [19].

	U_{EB}	U_{CB}	U_{CE}	I_E	I_C	I_B
NPN	-	+	+	-	+	+
PNP	+	-	-	+	-	-

Tabelle 8.1: Vorzeichen der Spannungen und Ströme in npn- und pnp-Transistor im Normalbetieb als Verstärker bei festgelegten Zählpfeilrichtungen.

8.1 Aufbau und Wirkungsweise

Der Injektions- oder Bipolartransistor ist eine PNP- oder NPN-Struktur, in welcher zwei pn-Übergänge räumlich so nahe aneinander angeordnet sind, dass sie sich gegenseitig beeinflussen, siehe Bild 8.3. Das mittlere Gebiet heißt ***Basis (base)***; das linke Gebiet ***Emitter (emitter)***, welches im Normalbetrieb in die Basis Minoritäten injiziert, und das rechte Gebiet ***Kollektor (collector)***, welcher die injizierten Minoritäten wieder aufsammelt.

Im ***Normalbetrieb (active mode)*** ist die Emitter-Basis Diode in Flussrichtung gepolt und die Kollektor-Basis Diode in Sperrrichtung. Bild 8.3(b) und (c) zeigen die dazugehörigen Ladungsträgerverteilungen und die elektrische Feldverteilung, die sich aus den Raumladungszonen und die elektrischen Felder, welche sich aus den extern angelegten Spannungen an den Dioden ergeben. In (d) ist das entsprechende Energiebanddiagram aufgezeichnet. Da die Emitter-Basis Diode vorwärts gepolt ist, besteht ein großer Diffusionsdruck von Elektronen vom Emitter in die Basis und umgekehrt von Löchern von der Basis in den Emitter. Falls die Basis genug lang ausgedehnt ist, werden alle in die Basis injizierten Elektronen in der Diffusionszone mit Löchern rekombinieren und einen Löcherstrom generieren, welcher an der Basis abgegriffen werden kann. Falls die Basis aber viel kleiner als die Diffusionslänge der Elektronen ist, werden die Elektronen einfach durch die Basis diffundieren und dem Potentialgefälle folgend in den Kollektor fluten. Falls die zusätzlichen Elektronen aus dem Emitter nicht vorhanden wären, so würde der Kollektor-Basis Übergang lediglich einen kleinen Sperrstrom aus den wenigen Elektronen in der p-dotierten Basis und einen kleinen Löcherstrom von Kollektor in die Basis führen. Der Ladungsträgerfluss ist in Fig. 8.3(e) noch einmal zusammengetragen.

Der Verstärkungseffekt liegt darin, dass der mit geringer Steuerleistung ($|U_{EB}| \ll |U_{CB}|, |I_E| \approx |I_C|$) an der EB-Diode erzeugte und in den Kollektor injizierte Strom an einem Arbeitswiderstand hohe Leistungen umsetzen kann. Voraussetzung dafür ist, dass die CB-Diode die Träger absaugt, also dass U_{CB} trotz des in Serie liegenden Spannungsabfalls am Arbeitswiderstand noch negativ bleibt. Im stromlosen Zustand entfällt der Spannungsabfall am Arbeitswiderstand und die CB-Diode die angelegte, hohe Sperrspannungen aufnehmen können.

Damit die CB-Diode hohe Sperrspannungen aushalten kann, empfiehlt sich nach Abschn. 6.4.3 eine niedrige Dotierung. Der npn-Transistor wird daher als $\mathrm{npn^-n^+}$-Schichtenfolge, siehe Bild **??**, realisiert. Wenn sich die Verarmungszone über die $\mathrm{n^-}$-Schicht bis in die $\mathrm{n^+}$-Schicht ausgedehnt hat, ist für den Sperrstrom die kleine Minoritätsträgerdichte der $\mathrm{n^+}$-Zone maßgeblich; dieses Prinzip haben wir bereits beim Leistungsgleichrichter, der $\mathrm{p^+s}_n\mathrm{n^+}$ - Diode, kennengelernt. Dieser Übergang verbindet hohe Spannungsfestigkeit mit kleinen Sperrverlusten.

Ein in Planartechnik hergestellter Transistor ist in Bild **??** dargestellt.

8.2 Quantitative Aussagen zum Ladungsträgertransport

Um quantitative Aussagen zum Stromtransport machen zu können müssen wir zunächst eine geeignete Notation einführen. Die Symbole, Ströme und Spannungen zusammen mit den Zählpfeilrichtungen für den npn- und den pnp-Transistor sind Bild 8.5 zu entnehmen. Der npn und der pnp Bipolar-Transistor haben je ein anderes Schaltsymbol. Der Pfeil zeigt die Hauptstromrichtung im entsprechenden Transistor an.

Im Normalbetrieb des Transistors als Verstärker ist die Emitter-Basis-Diode in Durchlassrichtung, die Kollektor-Basis-Diode in Sperrichtung vorgespannt. Die Vorzeichen der Ströme und Spannungen ergeben sich im Normalbetrieb damit so, wie in Tabelle 8.1 angegeben.

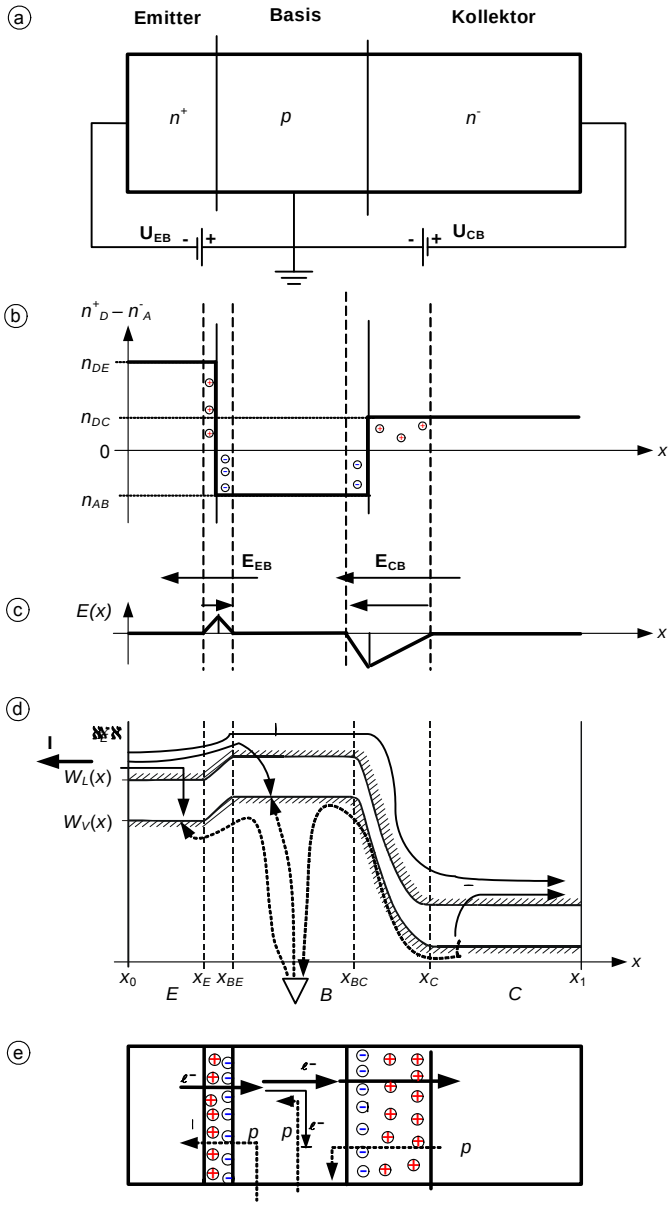


Abbildung 8.3: (a) n^+pn^- Diode im Normalbetrieb, (b) Dotierungen und Ladungsträgerkonzentrationen, (c) Verteilung des elektrischen Feldes, (d) Energiebanddiagram, (e) Ladungsträgertransport im Normalbetrieb

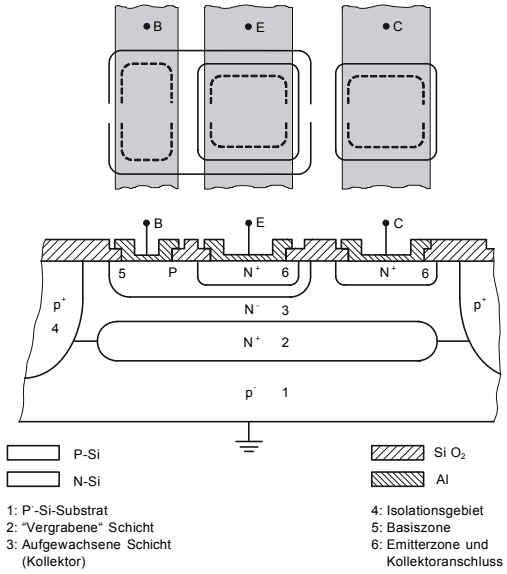


Abbildung 8.4: Aufsicht auf und Querschnitt durch einen Bipolartransistor in Planartechnik.

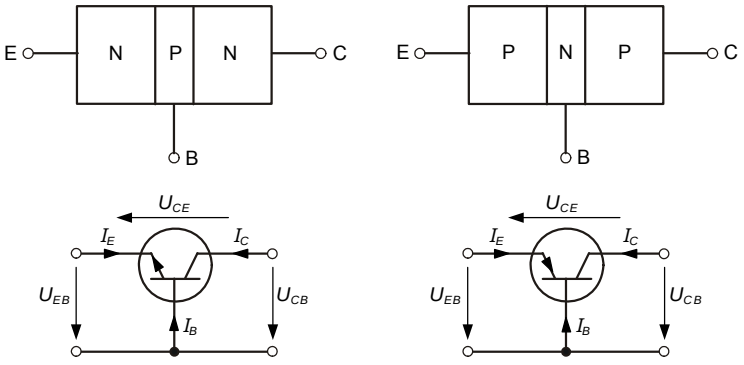


Abbildung 8.5: Querschnitte, Symbole und Definition der Spannungen und Ströme von bipolaren Transistoren.

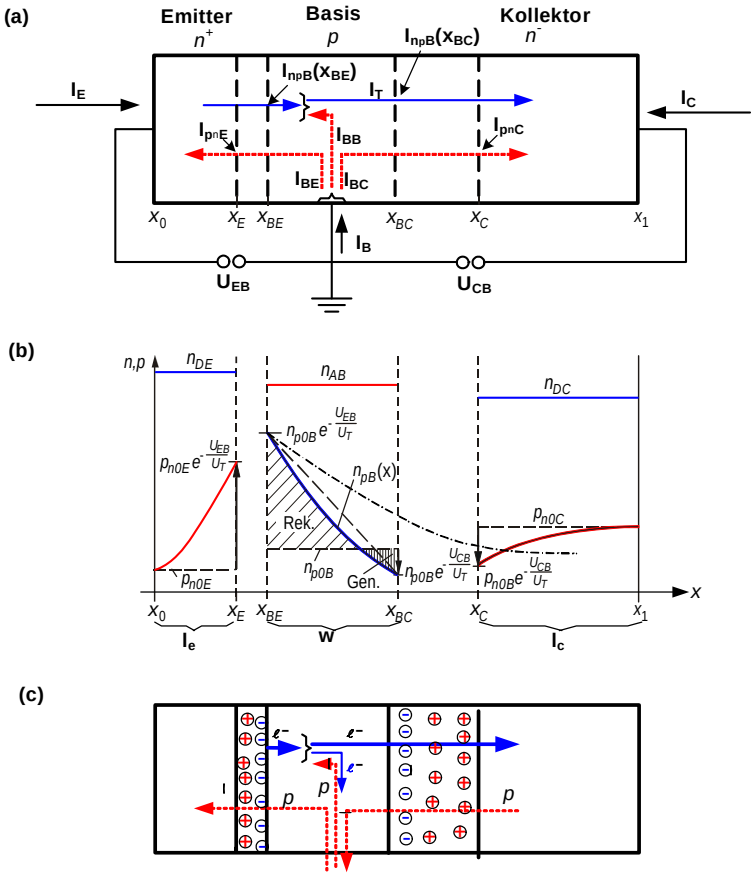


Abbildung 8.6: (a) Prinzipielle Struktur des npn-Transistors und Stromflüsse; (b) Diffusionszonen im npn-Transistor: (c) Elektronen und Löcherstromfluss im Normalbetrieb..

Bild 8.6(a) zeigt nochmals die Struktur des Transistors und die hier verwendeten Definitionen für die Stromflüsse. Die sich beim Normalbetrieb im Transistor aufbauenden Raumladungs- und Diffusionszonen sind in Fig. 8.6(b) gegeben. Die strichpunktierte Linie zeigt die Diffusionszone in der Basis wie sie sich in Abwesenheit des Kollektors ausbilden würde. Da die Basis sehr viel kürzer ist als die Diffusionslänge in der Basis, gelangen fast alle Minoritätsträger als Diffusionsstrom in die Kollektorzone und werden von der gesperrten Kollektor-Basis-Diode abgesaugt. Wegen des Flussbetriebes der EB-Diode ist $n_{pB}(x_{BE}) \gg n_{poB}$, wegen des Sperrbetriebs der CB-Diode ist $n_{pB}(x_{BC}) \approx 0$). Bild 8.6(c) zeigt nocheinmal die real auftretenden Elektronen- und Löcherstromflüsse im Normalbetrieb des Transistors.

Zur Berechnung der Stromflüsse verwenden wir zunächst die Kirchhoff'schen Regeln:

Es gilt für die Ströme nach der Kirchhoff'schen Knotenregel

$$I_E + I_B + I_C = 0, \tag{8.1}$$

und für die Spannungen nach der Kirchhoff'schen Maschenregel

$$U_{EB} - U_{CB} + U_{CE} = 0. \tag{8.2}$$

Bild 8.6(b) macht deutlich, dass durch die Spannungen U_{EB}, U_{CB} drei Diffusionszonen entstehen: eine in der Basis, für Elektronen, je eine im Emitter- bzw. Kollektorbahngebiet, für Löcher. Die Ströme im Transistor werden daher wie in der pn-Diode von der Stromergiebigkeit dieser Diffusionszonen bestimmt. Genau wie bei der pn-Diode kann man die Diffusionsstromanteile der

Löcher und Elektronen an den Raumladungszonengrenzen dazu verwenden die Gesamtstromanteile durch die einzelnen Dioden zu berechnen. Für die Ströme am Emitter bzw. Kollektor gilt mit den ortsabhängigen Diffusionsströmen im Emitter $I_{p_{nE}}(x)$, in der Basis $I_{n_{pB}}(x)$ und Kollektor $I_{p_{nC}}(x)$:

$$I_E = I_{n_{pB}}(x_{BE}) + I_{p_{nE}}(x_E) \tag{8.3}$$

$$I_C = -I_{n_{pB}}(x_{BC}) - I_{p_{nC}}(x_C) \tag{8.4}$$

und damit für den Basisstrom

$$\begin{aligned} I_B &= -I_E - I_C \\ &= -I_{p_{nE}}(x_E) + I_{n_{pB}}(x_{BC}) - I_{n_{pB}}(x_{BE}) - I_{p_{nC}}(x_C). \end{aligned} \tag{8.5}$$

Im Folgenden geht es also darum die einzelnen Anteile für den stationären Betriebsfall zu berechnen, wobei erwartet wird, dass die Elektronenanteile $I_{n_{pB}}$ den Hauptstrom bilden.

Wie schon bei der Diode, müssen wir die Gültigkeit der Shockley'schen Näherung fordern, d.h. die schwache (vernachlässigbare) Rekombination in den Raumladungszonen wird vorausgesetzt.

8.2.1 Diffusionsströme

Basis-Zone

Für den Diffusionsstrom in der Basis folgt nach Gl.(4.22)

$$I_{n_{pB}}(x) = AeD_{nB}\frac{dn_{pB}}{dx};\ x_{BE} \leq x \leq x_{BC}\,. \tag{8.6}$$

Um die Differentialgleichung zu lösen benötigen wir $n_{pB}(x)$. Also gilt es zuerst die Differentialgleichung für die Elektronen, den Minoritäten, in der Basis zu lösen. Für $n_{pB}(x)$ folgt analog zu 5.71 die Differentialgleichung

$$\frac{d^2 n_{pB}}{dx^2} = \frac{1}{L_{nB}^2}[n_{pB}(x) - n_{poB}], \tag{8.7}$$

wobei L_{nB} die Diffusionslänge der Elektronen in der Basis und A die Querschnittsfläche des Transistors ist.

Deren allgemeine Lösung lautet

$$n_{pB}(x) - n_{poB} = Be^{x/L_{nB}} + Ce^{-x/L_{nB}} \tag{8.8}$$

Spätestens jetzt müssen wir uns um die Randbedingungen für die Minoritätsträgerdichte kümmern. Bild 8.6(b) zeigt schematisch die sich im Normalbetrieb einstellenden Minoritätsträgerdichten eines npn-Transistors. In der Basis ist n_{AB} die Dotierstoffkonzentration. Es werden abrupte pn-Übergänge und Störstellenerschöpfung angenommen, so dass für die Löcher als Majoritätsladungsträger gilt

$$p_{poB} = n_{AB}. \tag{8.9}$$

Für die Elektronenkonzentration $n(x)$, die in der Basis die Minoritäten darstellen, gilt dann

$$n_{poB} = n_i^2/n_{AB}. \tag{8.10}$$

Damit sind die Gleichgewichtskonzentrationen aus den Dotierstoffkonzentrationen und der Eigenleitungskonzentration berechenbar.

Beim Anlegen einer Flussspannung $U_{EB} < 0$ an die EB-Diode und einer Sperrspannung $U_{CB} > 0$ an die CB-Diode ändern sich die Ladungsträgerkonzentrationen an den neuen Sperrschichträndern genauso, wie es für die Einzeldioden schon besprochen wurde und in Bild 8.6(b)

dargestellt ist. Letzlich erhalten wir am basisseitigen Ende der EB-Raumladungszone x_{BE} als Randbedingung:

$$n_{pB}(x_{BE}) = n_{poB} e^{-U_{EB}/U_T} \gg n_{poB}, \tag{8.11}$$

und am basisseitigen Ende der CB-Raumladungszone x_{BC} als zweite Randbedingung:

$$n_{pB}(x_{BC}) = n_{poB} e^{-U_{CB}/U_T} \ll n_{poB}\,. \tag{8.12}$$

Da die Randbedingungen die angelegten Spannungen U_{EB}, U_{CB} enthalten, werden $n_{pB}(x)$, $I_{n_{pB}}(x)$ und damit I_E Funktionen der beiden Spannungen. Einsetzen der beiden Randbedingungen (8.11) und (8.12) in (8.8) ergibt:

$$\begin{aligned} n_{pB}(x) - n_{poB} = {} & \frac{n_{poB}}{\sinh\left(\frac{x_{BC}-x_{BE}}{L_{nB}}\right)} \times \\ & \left[\left(e^{-U_{EB}/U_T} - 1\right)\sinh\left(\frac{x_{BC}-x}{L_{nB}}\right)\right. \\ & \left. + \left(e^{-U_{CB}/U_T} - 1\right)\sinh\left(\frac{x-x_{BE}}{L_{nB}}\right)\right]. \end{aligned} \tag{8.13}$$

Ein Vergleich der Lösung (8.13) mit Abschnitt 5.6.2 für die kurze Diffusionszone und ein Vergleich von Bild 8.6(b) mit Bild 5.9 zeigt, dass die Minoritätsträgerdichte (8.13) gerade den zwei kurzen Diffusionszonen infolge der Elektroneninjektion vom Emitter und der Elektronenextraktion vom Kollektor entspricht (5.78).

Der Stromanteil $I_{n_{pB}}(x)$ in der Basis ergibt sich dann nach (8.6) aus der Ableitung und Multiplikation mit AeD_{nB}:

$$\begin{aligned} I_{n_{pB}}(x) = {} & \frac{AeD_{nB}n_{poB}}{L_{nB}\sinh\left(\frac{w}{L_{nB}}\right)}\left[-\left(e^{-U_{EB}/U_T} - 1\right)\cosh\left(\frac{x_{BC}-x}{L_{nB}}\right)\right. \\ & \left. + \left(e^{-U_{CB}/U_T} - 1\right)\cosh\left(\frac{x-x_{BE}}{L_{nB}}\right)\right]; x_{BE} \le x \le x_{BC} \end{aligned} \tag{8.14}$$

$$\begin{aligned} I_{n_{pB}}(x) = {} & I_{TS}\left[-\left(e^{-U_{EB}/U_T} - 1\right)\cosh\left(\frac{x_{BC}-x}{L_{nB}}\right)\right. \\ & \left. + \left(e^{-U_{CB}/U_T} - 1\right)\cosh\left(\frac{x-x_{BE}}{L_{nB}}\right)\right]; x_{BE} \le x \le x_{BC} \end{aligned} \tag{8.15}$$

I_{TS} wird Transfersättigungstrom genannt und ist durch

$$\begin{aligned} I_{TS} &= \frac{AeD_{nB}n_{poB}}{L_{nB}\sinh\left(\frac{w}{L_{nB}}\right)} \\ &\approx Ae\frac{D_{nB}n_{poB}}{w} = Aen_i^2\frac{D_{nB}}{wn_{AB}} \end{aligned} \tag{8.16}$$

gegeben. Die Weite der Basiszone ist sehr viel kleiner als die Diffusionszone zu wählen, damit ein möglichst großer Anteil der injizierten Minoritätsträger die Kollektorzone erreicht, $w = x_{BC} - x_{BE} \ll L_{nB}$.

Der Basis-Diffusionsstrom, welcher den Kollektor erreicht, heißt Transferstrom I_T

$$\begin{aligned} I_T &= I_{n_{pB}}(x_{BC}) \\ &= I_{TS}\left[-\left(e^{-U_{EB}/U_T} - 1\right) + \left(e^{-U_{CB}/U_T} - 1\right)\cosh\left(\frac{w}{L_{nB}}\right)\right] \\ &\approx I_{TS}\left[-\left(e^{-U_{EB}/U_T} - 1\right) + \left(e^{-U_{CB}/U_T} - 1\right)\right]. \end{aligned} \tag{8.17}$$

Die Differenz zwischen dem injizierten Minoritätenstrom und dem in den Kollektor transferierten Minoritätenstrom entspricht den in der Basis rekombinierten Minoritätsträgern, welche durch einen entsprechenden Zufluss von Majoritäten, d.h. durch den Stromanteil I_{BB} über den Basisanschluss kompensiert werden, siehe Bild 8.6(a). Aus Gl.(8.15) erhalten wir für diesen Basisstromanteil

$$\begin{aligned} I_{BB} &= -I_{n_{pB}}(x_{BE}) + I_{n_{pB}}(x_{BC}) \\ &= I_{TS}\left[\left(e^{-U_{EB}/U_T}-1\right)+\left(e^{-U_{CB}/U_T}-1\right)\right]\times \\ &\quad \left[\cosh\left(\frac{w}{L_n}\right)-1\right], \end{aligned} \tag{8.18}$$

wobei mit $\cosh\left(\frac{w}{L_n}\right) \approx 1+\frac{1}{2}\left(\frac{w}{L_{nB}}\right)^2$ folgt

$$\begin{aligned} &\approx I_{TS}\frac{1}{2}\left(\frac{w}{L_{nB}}\right)^2\left[\left(e^{-U_{EB}/U_T}-1\right)+\left(e^{-U_{CB}/U_T}-1\right)\right] \\ &\approx -I_T\frac{1}{2}\left(\frac{w}{L_{nB}}\right)^2. \end{aligned} \tag{8.19}$$

Emitter-Zone

Um den Stromanteil, welcher aus der Minoritätsträgerdiffusion an der Emitter-Basis Diode entsteht zu erhalten, muss man wieder die Differentialgleichung für die Minoritäten p in der Emitterzone lösen. Die dazugehörigen Randbedingungen ergeben sich aus der Dotierstoffkonzentrationen n_{DE} im Emitter. Im stationären Zustand gilt zunächst für die Majoritäten

$$n_{noE} = n_{DE} \tag{8.20}$$

und die Löcher als Minoritätsladungsträger:

$$p_{noE} = n_i^2/n_{DE}. \tag{8.21}$$

Beim Anlegen einer Durchlassspannung $U_{EB} < 0$ an die EB-Diode ändern sich die Ladungsträgerkonzentrationen und am emitterseitigen Ende der EB-Raumladungszone x_E gilt:

$$p_{nE}(x_E) = p_{noE}e^{-U_{EB}/U_T} \gg p_{noE}. \tag{8.22}$$

Als weitere Randbedingungen an der Emitter-Kontaktfläche (ohmsche Kontakte, siehe später) des npn-Transistors fordern wir, dass dort alle Überschussladungsträger rekombiniert sind, vgl. die kurze Diffusionszone, Kapitel 5.6.2:

$$p_{nE}(x_0) = p_{noE} = n_i^2/n_{DE}. \tag{8.23}$$

Für den Stromanteil der Löcherdiffusionszone im Emitter erhält man analog zu oben:

$$I_{p_{nE}} = \frac{AeD_{pE}}{L_{pE}}p_{noE}\left(e^{-U_{EB}/U_T}-1\right)\frac{\cosh\left(\frac{x_0-x}{L_{pE}}\right)}{\sinh\left(\frac{x_E-x_0}{L_{pE}}\right)}; \quad x_0 \le x \le x_E, \tag{8.24}$$

L_{pE} ist die Diffusionslänge der Löcher im Emitter.

Für den Basis-Emitter-Strom I_{BE} erhält man damit

$$I_{BE} = -I_{p_{nE}}(x = x_E) = I_{BES}\left(e^{-U_{EB}/U_T}-1\right), \tag{8.25}$$

wobei I_{BES} der Basis-Emitter-Sättigungstrom infolge der Löcherinjektion von der Basis in den Emitter ist

$$I_{BES} = \frac{AeD_{pE}p_{noE}}{L_{pE}\tanh\left(\frac{l_E}{L_{pE}}\right)}. \tag{8.26}$$

Ist die Länge der Emitterzone ebenfalls sehr viel kleiner als die zugehörige Diffusionszone, $l_E = x_E - x_0 \ll L_{pE}$, so erhält man

$$I_{BES} \approx Ae\frac{D_{pE}p_{noE}}{l_E} = Aen_i^2\frac{D_{pE}}{l_E n_{DE}}. \tag{8.27}$$

Kollektor-Zone

Um den Stromanteil, welcher aus der Minoritätsträgerdiffusion an der Kollektor-Basis-Diode entsteht, zu erhalten, muss man wieder die Differentialgleichung für die Minoritäten p in der Kollektorzone lösen. Die dazugehörigen Randbedingungen ergeben sich aus den Dotierstoffkonzentrationen n_{DC} im Kollektor. Im Kollektor gilt für die Elektronen als Majoritätsträger bzw die Löcher als Minoritätsladungsträger

$$n_{noC} = n_{DC}, \qquad p_{noC} = n_i^2/n_{DC}. \tag{8.28}$$

Nach Anlegen einer Sperrspannung $U_{CB} > 0$ an die CB-Diode ändern sich die Ladungsträgerkonzentrationen an den neuen Sperrschichträndern genau so, wie es für die Einzeldioden schon besprochen und am kollektorseitigen Ende der CB-Raumladungszone x_C gilt:

$$p_{nC}(x_C) = p_{noC}e^{-U_{CB}/U_T} \ll p_{noC}. \tag{8.29}$$

Als Randbedingungen soll an den Emitter- und Kollektor-Kontaktflächen (ohmsche Kontakte, siehe später) gelten

$$p_{nC}\left(x_1\right) = p_{noC} = n_i^2/n_{DC}\,, \tag{8.30}$$

sodass dort alle Überschussladungsträger rekombiniert sind, vgl. die kurze Diffusionszone, Kapitel 5.6.2.

Für den Löcherdiffusionsstrom im Kollektor erhält man dann analog zum Emitter

$$I_{p_{nC}}\left(x\right) = \frac{eAD_{pC}}{L_{pC}}p_{noC}\left(e^{-U_{CB}/U_T} - 1\right)\frac{\cosh\left(\frac{x_1-x}{L_{pC}}\right)}{\sinh\left(\frac{x_1-x_C}{L_{pC}}\right)} \tag{8.31}$$

$$\text{für} \quad x_C \le x \le x_1,$$

wobei L_{pC} die Diffusionslänge der Löcher im Kollektor ist. Unter der Annahme, dass die Kollektorweite $l_C = x_1 - x_C$ in der Regel lang gegen die Diffusionslänge der Löcher (L_{pC}) ist, d.h. $L_{pC} \ll l_{nC}$, folgt :

$$I_{p_{nC}}\left(x\right) = I_{BCS}\left(e^{-U_{CB}/U_T} - 1\right)\exp\left(-\frac{x - x_C}{L_{pC}}\right) \tag{8.32}$$

$$x_C \le x \le x_1$$

mit dem Basis-Kollektor-Sättigungsstrom

$$I_{BCS} = A\frac{eD_{pC}}{L_{pC}}p_{noC} = Aen_i^2\frac{D_{pC}}{L_{pC}n_{DC}}. \tag{8.33}$$

Für den Basis-Kollektorstrom $I_{BC} = I_{p_{nC}}\left(x_C\right)$ ergibt sich somit

$$I_{BC} = I_{p_{nC}}\left(x_C\right) = I_{BCS}\left(e^{-U_{CB}/U_T} - 1\right). \tag{8.34}$$

Die Sättigungsströme I_{TS}, I_{BES} und I_{BCS} charakterisieren die physikalischen Eigenschaften von Basis, Emitter und Kollektor.

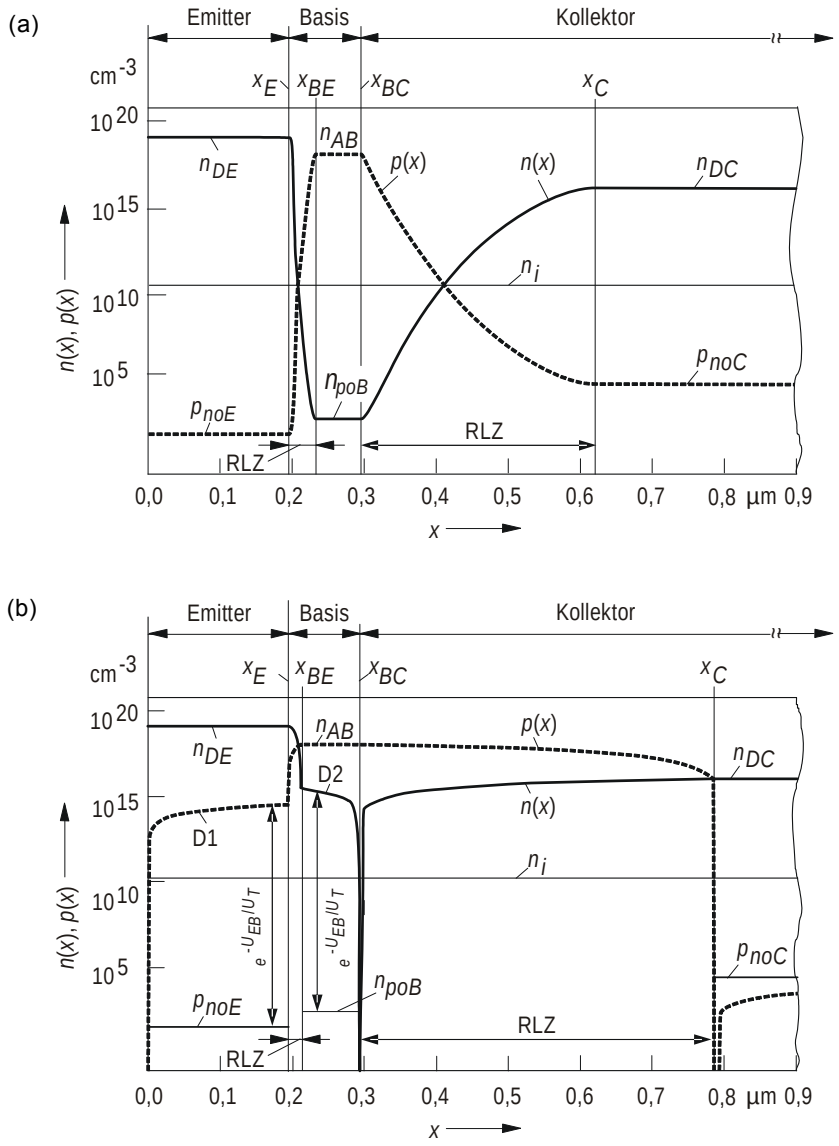


Abbildung 8.7: Querschnitt durch einen Silizium-npn-Transistor, Dotierstoff- und Ladungsträgerverteilung bei den äußeren Spannungen: (a) $U_{EB} = U_{CB} = 0$; (b) $U_{EB} = -0,75$ V, $U_{CB} = 1V$; Zahlenwerte für die Berechnung: $n_{DE} = 10^{19}$ cm^{-3}, $n_{AB} = 10^{18}$cm^{-3}, $n_{DC} = 10^{16}$cm^{-3}, Dicke der Basis 0,1 μm, Diffusionslänge der Löcher im Kollektorbahngebiet $L_{pC} = 1\mu$m, Dicke des Emitters 0,2μm.

8.2.2 Emitter-, Basis- und Kollektorströme

Aus den Gln.(8.3) - (8.5) sowie (8.17), (8.18), (8.25) und (8.34) folgt für Emitter, Kollektor und Basisstrom, siehe Bild 8.6(a)

$$\begin{aligned} I_E &= I_T - I_{BB} - I_{BE} \\ &= -I_{EE}\left(e^{-U_{EB}/U_T} - 1\right) + I_{TS}\left(e^{-U_{CB}/U_T} - 1\right) - I_{BB} \end{aligned} \tag{8.35}$$

mit der Abkürzung:

$$I_{EE} = I_{TS} + I_{BES}. \tag{8.36}$$

$$\begin{aligned} I_C &= -I_T - I_{BC} \\ &= I_{TS}\left(e^{-U_{EB}/U_T} - 1\right) - I_{CC}\left(e^{-U_{CB}/U_T} - 1\right) \end{aligned} \tag{8.37}$$

$$I_{CC} = I_{TS} + I_{BCS}. \tag{8.38}$$

Für den gesamten Basisstrom ergibt sich

$$\begin{aligned} I_B &= I_{BE} + I_{BC} + I_{BB} \\ &= I_{BES}\left(e^{-U_{EB}/U_T} - 1\right) + I_{BCS}\left(e^{-U_{CB}/U_T} - 1\right) \\ &+ I_{TS}\frac{1}{2}\left(\frac{w}{L_{nB}}\right)^2\left[e^{-U_{EB}/U_T} + e^{-U_{CB}/U_T} - 2\right]. \end{aligned} \tag{8.39}$$

Der Vollständigkeit wegen sind die Ladungsträgerdichten für einen npn-Transistor mit realistischen Emitter-Basis und Kollektorweiten in Bild 8.7 dargestellt. In (a) für den spannungslosen Zustand und in (b) für den Normalbetrieb.

8.3 Ebers-Moll Modell

Vernachlässigen wir den Basisstromanteil I_{BB}, welcher durch Rekombination und Generation in der Basis zustande kommt, so nehmen Emitterstrom und Kollektorstrom eine symmetrische Form an:

$$I_E = -I_{EE}\left(e^{-U_{EB}/U_T} - 1\right) + I_{TS}\left(e^{-U_{CB}/U_T} - 1\right) \tag{8.40}$$

$$I_C = I_{TS}\left(e^{-U_{EB}/U_T} - 1\right) - I_{CC}\left(e^{-U_{CB}/U_T} - 1\right) \ . \tag{8.41}$$

Diese Gleichungen heißen ***Ebers-Moll-Gleichungen.***

Wir hätten diese Gleichungen auch ohne große Herleitung aufstellen können. Um das zu sehen betrachten wir die npn-Diode in Fig. 8.8(a). Wenn wir nun beispielsweise den resultierenden Kollektor-Strom I_C hinschreiben möchten, so betrachtet man die Basis-Kollektor-Diode als eigenständige Diode, mit Löcher- und Elektronenstromanteilen wie man es von einer pn-Diode erwarten würde und addiert einen Emitterstromanteil hinzu. Konkret setzt sich I_C dann aus folgenden Beiträgen zusammen

- Dem Löcherstromanteil -$I_{BC} = -I_{BCS}\left(e^{-U_{CB}/U_T} - 1\right)$ (das negative Vorzeichen gehört hinzu, da I_{BC} gegenüber der Definition von I_C das umgekehrte Vorzeichen hat).
- Dem Elektronen-Dioden-Stromanteil -$I_{TS}\left(e^{-U_{CB}/U_T} - 1\right)$ (und zwar nur dem Anteil, welcher eine normale pn-Diode entsprechen würde, ebenfalls mit negativem Vorzeichen)

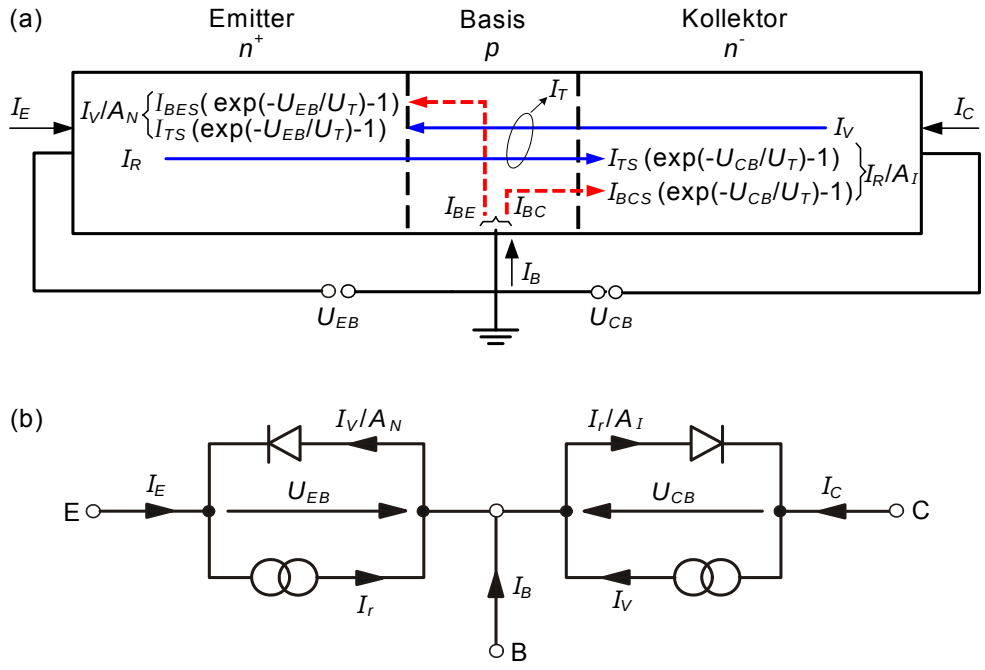


Abbildung 8.8: (a) Strombeiträge im Ebers-Moll Modell. (b) Ebers-Moll-Ersatzschaltbild des Transistors.

- Dem Anteil aus dem Emitteranteil, welcher angesaugt wird, i.e. $I_{TS}\left(e^{-U_{EB}/U_T}-1\right)$
  Dies führt auf

$$I_C = -I_{BCS}\left(e^{-U_{CB}/U_T}-1\right) - I_{TS}\left(e^{-U_{CB}/U_T}-1\right) + I_{TS}\left(e^{-U_{EB}/U_T}-1\right) \ . \quad (8.42)$$

  Nach dem gleichen Schema kann man sich den Emitterstrom I_E hinschreiben. Man erhält dann

$$I_E = -I_{EBS}\left(e^{-U_{EB}/U_T}-1\right) - I_{TS}\left(e^{-U_{EB}/U_T}-1\right) + I_{TS}\left(e^{-U_{CB}/U_T}-1\right) \ . \quad (8.43)$$

  Benutzt man die Defintion $I_{EE} = I_{TS} + I_{BES}$ aus Gl. (8.36) und $I_{CC} = I_{TS} + I_{BCS}$ aus Gl.(8.38), so erhält man wieder die Gleichungen (8.40)-(8.41).

Addiert man die beiden Elektronenströme in der Basis, so sollte man wieder auf den Transferstrom kommen. Dies wollen wir hier beweisen. Der Transferstrom in der Basis ist ein Diffusionsstrom der Minoritätsträger. In den Ebers-Moll-Relationen wird nun Generation und Rekombination in der Basis vernachlässigt. Konsequenterweise verläuft dann die Minoritätsträgerkonzentration $n(x)$, siehe Bild 8.6(b), linear (- - - -), da der Gradient von $n(x)$ proportional zum Transferstrom I_T sein muss und I_T über der ganzen Basis konstant bleiben muss. Mit der Zählpfeilkonvention von Bild 8.6(a) kann man den Transferstrom sofort hinschreiben. Dieser ist

$$I_T = AJ_{nD} = -eAD_{nB}\frac{n_{pB}(x_{BE}) - n_{PB}(x_{BC})}{w}, \quad \text{mit} \quad (8.44)$$

$$n(x_{BE}) = n_{poB}\,\mathrm{e}^{-U_{EB}/U_T}, \quad n(x_{BC}) = n_{poB}\,\mathrm{e}^{-U_{CB}/U_T}, \quad (8.45)$$

also

$$\begin{aligned} I_T &= -\frac{eAD_{nB}}{w}\left\{[n_{pB}(x_{BE}) - n_{poB}] - [n_{pB}(x_{BC}) - n_{poB}]\right\} \\ &= -I_{TS}\left[(\mathrm{e}^{-U_{EB}/U_T} - 1) - (\mathrm{e}^{-U_{CB}/U_T} - 1)\right] \\ &= -I_v + I_r \ . \end{aligned} \quad (8.46)$$

Dies entspricht genau der Superposition der beiden Elektronenströmen in der Basis. Es ist praktisch, die zwei Ströme als unabhängige Diffusionsströme I_v und I_r zu deuten. Dabei ist der eine nur durch U_{EB}, und der andere nur durch U_{CB} gesteuert. I_v wird als Vorwärtsstrom, I_r als Rückwärtsstrom bezeichnet.

Das Ergebnis sind die Ebers-Moll-Gleichungen (nicht in der originalen Schreibweise, sondern in einer für die weitere Anwendung zweckmäßig abgeänderten Form):

$$\begin{pmatrix} I_E \\ I_C \end{pmatrix} = \begin{pmatrix} -(I_{TS}+I_{BES}) & I_{TS} \\ I_{TS} & -(I_{TS}+I_{BCS}) \end{pmatrix} \begin{pmatrix} \left(e^{-U_{EB}/U_T}-1\right) \\ \left(e^{-U_{CB}/U_T}-1\right) \end{pmatrix}$$
$$= \begin{pmatrix} -A_N^{-1} & 1 \\ 1 & -A_I^{-1} \end{pmatrix} \begin{pmatrix} I_v \\ I_r \end{pmatrix} \tag{8.47}$$

Das Gleichungssystem (8.47) resultiert in dem Ebers-Moll-Ersatzschaltbild 8.8(b). Es ist wie folgt zu interpretieren: Der von U_{EB} gesteuerte Strom der EB-Diode ist $I_v + I_{BE} = I_v/A_N$; nur der Anteil $A_N(I_v + I_{BE}) = I_v$ davon ($A_N \leq 1$) gelangt in den Kollektor (Stromgenerator I_v), der Rest I_{BE} (endlicher Emitterwirkungsgrad, siehe später) schließt sich gegen die Basis. Analoge Überlegungen gelten für den Strom $I_r + I_{BC} = I_r/A_I$ der CB-Diode. Die durch Stromgeneratoren erfassten Ströme sind jene Stromanteile, welche von einer Diode jeweils in die andere Diode zufolge der Verkopplung über die dünne Basis injiziert werden.

Oft werden diese Gleichungen auch in der Form:

$$I_E = -I_{EE}\left(e^{-U_{EB}/U_T}-1\right) + A_I I_{CC}\left(e^{-U_{CB}/U_T}-1\right) \tag{8.48}$$

$$I_C = A_N I_{EE}\left(e^{-U_{EB}/U_T}-1\right) - I_{CC}\left(e^{-U_{CB}/U_T}-1\right) \tag{8.49}$$

$$\text{mit } A_N = \frac{I_{TS}}{I_{EE}} = \frac{I_{TS}}{I_{TS}+I_{BES}}; \quad A_I = \frac{I_{TS}}{I_{CC}} = \frac{I_{TS}}{I_{TS}+I_{BCS}} \tag{8.50}$$

angegeben, wobei A_N und A_I wieder die positive Größen aus Gl. (8.47) sind und als ***Vorwärts-*** und ***Rückwärtsstromverstärkung*** bezeichnet werden.

8.4 Betriebszustände des Transistors

Die Abhängigkeit des Emitter- und Kollektorstromes als Funktion der anliegenden Spannungen U_{EB} und U_{CB} lässt sich anschaulich darstellen, wenn man nur den wesentlich am Stromfluss beteiligten Ladungsträgeranteil, die Minoritätenkonzentration in der Basis $n_{pB}(x)$, betrachtet. Man gelangt auf diese Weise zu den Bildern 8.9 (a) - (f), die einen Ausschnitt aus dem vollständigen Konzentrationsverlauf, 8.6(b), darstellen:

- **Bild 8.9(a):** U_{EB} und U_{CB} sind Null. Die Minoritätenkonzentration hat den konstanten Gleichgewichtswert $n_{poB} = \text{const}$. Emitterstrom I_E und Kollektorstrom I_C verschwinden.
- **Bild 8.9(b):** $U_{EB} < 0$, d.h. die Emitter-Basis-Diode wird in Flussrichtung gepolt und die Minoritätenkonzentration an der Stelle $x = x_{BE}$ um den Faktor $\exp(-U_{EB}/U_T)$ angehoben. U_{CB} bleibt Null. Die Minoritätenkonzentration $n_{pB}(x)$ hat einen konstanten Gradienten dn_{pB}/dx, welcher einen Emitter- und Kollektorstromfluss erzeugt.
- **Bild 8.9(c):** Der ***aktive*** oder ***normale Betrieb*:.** Wird die Kollektor-Basis-Diode in Sperrrichtung gepolt, so sinkt die Minoritätenkonzentration an der Stelle $x = x_{BC}$ um einen Faktor $\exp(-U_{CB}/U_T)$ ab. Im linearen Maßstab geht $n_{pB}(x = x_{BC})$ gegen Null. Wegen des steileren Konzentrationsverlaufs nimmt I_C etwas zu.
- **Bild 8.9(d):** Wird die Kollektor-Basis-Diode in Flussrichtung gepolt, so steigt die Minoritätenkonzentration an der Stelle $x = x_{BC}$ an. Der Konzentrationsverlauf wird flacher, der Strom I_C nimmt entsprechend ab, bis er schließlich zu Null wird ***(Sättigungsbereich, saturation).***

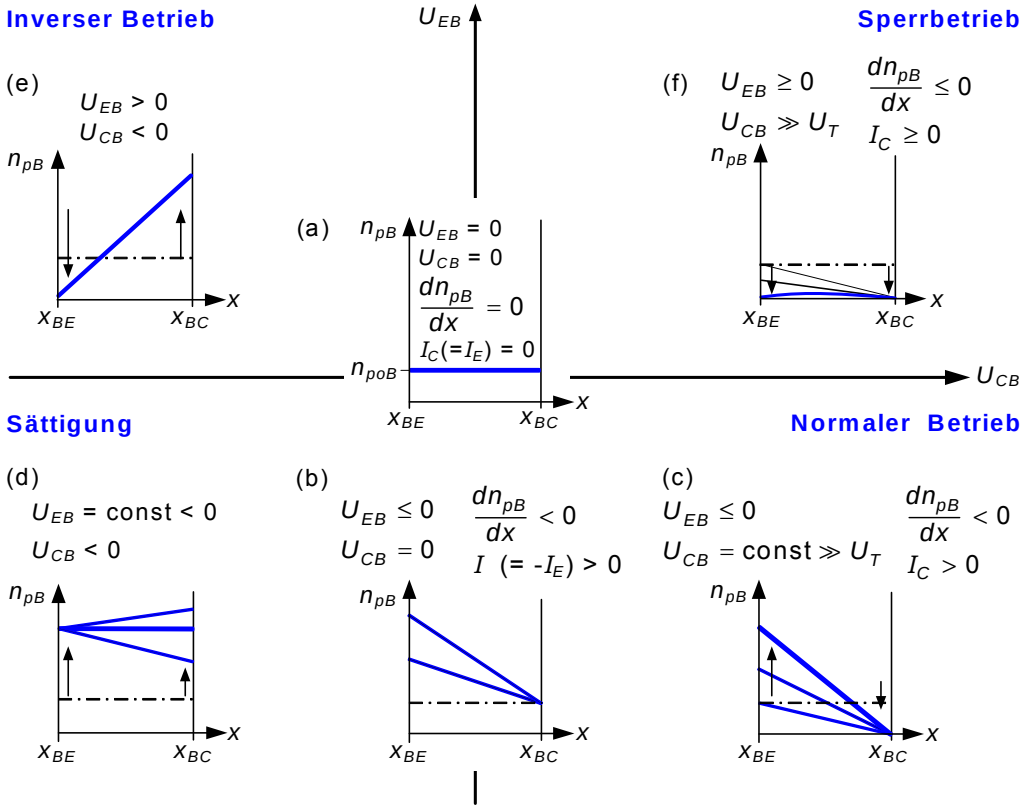


Abbildung 8.9: Räumliche Verteilung der Minoritätenkonzentrationen in der Basis in verschiedenen Betriebszuständen des Transistors.

- **Bild 8.9(e):** Wird die Kollektor-Basis-Diode weiter in Flussrichtung gepolt, oder die Emitter-Basis Diode in Sperrrichtung betrieben, so kehrt der Strom I_C sogar sein Vorzeichen. Wir sind im ***inversen Betrieb (inverted mode).***
- **Bild 8.9(f):** Ist die Emitter-Basisspannung Null und der Kollektor gesperrt, so fließt noch ein Kollektorstrom I_C. Erst wenn auch die Emitter-Basisdiode in Sperrrichtung gepolt ist, kann I_C zu Null werden. Wir sind im ***Sperrbetrieb (cut-off).***

Mit der Änderung der Spannungen an den Dioden ist auch eine Änderung der Ausdehnungen ihrer Raumladungszonen verbunden. Da die Emitter-Basisspannung in der Regel klein ist, kann die Ausdehnung der Emitter-Raumladungszone vernachlässigt werden (d.h. $x_E = \text{const}$). Die Kollektorsperrspannung ist jedoch groß, Änderungen machen sich in einer Verschiebung der Raumladungsgrenzen bemerkbar. Damit hängt die Basisweite $w = x_{BC} - x_{BE}$ von der angelegten Spannung ab: $w = w\,(U_{CB})$.

Die Modulation der Basisweite mit der Sperrspannung heißt ***Early-Effekt,*** welcher später noch genauer untersucht wird. Mit abnehmender Basisweite w nimmt der Kollektorstrom I_C zu. Aus Fig. 8.10 wird anschaulich deutlich, dass $I_C \to \infty$ geht, wenn die Kollektor-Sperrschicht die Emitter-Sperrschicht berührt ***(punch-through),*** d.h. $x_{BC} = x_{BE}$, $w = 0$.

8.5 Kennlinenfeld des npn-Bipolartransistors

Der Transistor kann mit seinen drei Anschlussklemmen in drei Grundschaltungen verwendet werden: Nämlich der ***Basisschaltung,*** der ***Emitterschaltung*** und der ***Kollektorschaltung*** - je nachdem, ob der Bezugspunkt der Spannungen die Basis, der Emitter oder der Kollektor ist.

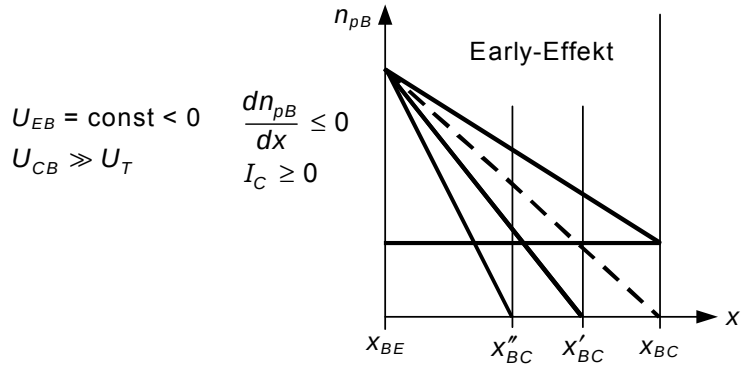


Abbildung 8.10: Modulation der Basisweite mit der Sperrspannung

Grund-schaltung	I_1	I_2	U_1	U_2
Basis	$I_{1B} = I_E$	$I_{2B} = I_C$	$U_{1B} = U_{EB}$	$U_{2B} = U_{CB}$
Emitter	$I_{1E} = I_B$	$I_{2E} = I_C$	$U_{1E} = U_{BE}$ $= -U_{EB}$	$U_{2E} = U_{CE}$ $= U_{CB} - U_{EB}$ $= U_{CB} + U_{BE}$
Kollektor	$I_{1C} = I_B$	$I_{2C} = I_E$	$U_{1C} = U_{BC}$ $= -U_{CB}$	$U_{2C} = U_{EC}$ $= -U_{CE}$ $= -U_{CB} + U_{EB}$

Tabelle 8.2: Bezeichnung der Ein- und Ausgangsströme und -spannungen in drei Grundschaltungen des npn-Transistors und ihre Zuordnung zu den Strömen I_E, I_C, I_B bzw. Spannungen U_{EB}, U_{CB}

- Die Basisschaltung (mit Emitter-Basis-Diode als Eingangstor und der Kollektor-Basis-Diode als Ausgangstor) zeigt praktisch keine Stromverstärkung, jedoch Spannungsverstärkung)
- Die wichtigste Grundschaltug ist die Emitterschaltung, da sie für Signale im Verstärkerbetrieb sowohl Strom- als auch Spannungsverstärkung zeigt.
- Die Kollektorschaltung weist zwar Stromverstärkung auf, zeigt aber nur geringe Spannungsverstärkung.

Weiter sind die Ein- und Ausgangsimpedanzen der verschiedenen Grundschaltungen im Wechselstrom-(Verstärker-) betrieb wichtig. Hier zeigt die Basisschaltung geringe Eingangs- und sehr hohe Ausgangsimpedanz, die Emitterschaltung eine höhere Eingangsimpedanz als die Basisschaltung und eine etwa gleich hohe Ausgangsimpedanz, während die Kollektorschaltung eine hohe Eingangs- und kleine Ausgangsimpedanz aufweist. Ein zahlenmäßiger Vergleich der einzelnen Verstärkungen und Impedanzen ist der später noch folgenden Wechselstrom-Kleinsignalanalyse zu entnehmen.

Der Zusammenhang zwischen Ein- und Ausgangsgrößen (I_{1i}, I_{2i} , U_{1i}, U_{2i}, $i = B$, E, C) der Vierpole und den Strömen I_E, I_C, I_B und Spannungen U_{EB}, U_{CB} ist für den NPN-Transistor aus den Fig. 8.5 und 8.11 zu entnehmen und in Tab. 8.2 dargestellt.

Tabelle 8.2 zeigt die Bezeichnung der Ein- und Ausgangsströme und -spannungen in den drei Grundschaltungen des npn-Transistors und ihre Zuordnung zu den Strömen I_E, I_C, I_B bzw. Spannungen U_{EB}, U_{CB}. Da wir vier Variablen (I_{1i}, I_{2i} , U_{1i}, U_{2i}) haben, welche durch die zwei Ebers-Moll Gleichungen verknüpft sind, bleiben eigentlich nur zwei Freiheitsgrade. Wenn wir uns also ein Bild über die Eingangs- bzw. Ausgangsströme machen wollen, müssen wir die funktionale Abhängigkeiten z.B. eines Stromes als Funktion von zwei Variablen aufzeichnen. Die resultierenden Plots werden als ***Kennlinienfelder*** bezeichnet.

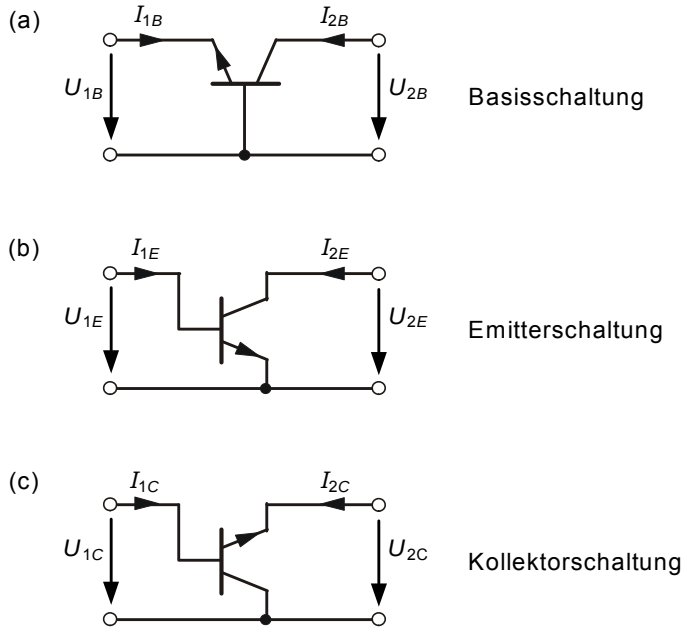


Abbildung 8.11: Die Grundschaltungen des Bipolartransistors

8.5.1 Basisschaltung

In der Basisschaltung sind die Eingangs- und Ausgangsströme von Null verschieden. Es gibt daher sowohl ein Eingangs-Kennlinienfeld als auch ein Ausgangs-Kennlinienfeld, wobei die Größen des einen Tores auch von den Parametern des anderen Tores abhängen.

Eingangskennlinienfelder

$$I_{1B}\left(U_{1B}, U_{2B}\right) = I_E\left(U_{EB}, U_{CB}\right) \qquad (8.51)$$
$$\text{mit} \quad U_{2B} = U_{CB} \quad \text{als Parameter des Ausgangstores}$$

oder

$$I_{1B}\left(U_{1B}, I_{2B}\right) = I_E\left(U_{EB}, I_C\right)$$
$$\text{mit} \quad I_{2B} = I_C \quad \text{als Parameter des Ausgangstores}$$

Es hängt vom jeweiligen Anwendungsfall ab, welcher Ausgangstor-Parameter günstigerweise gewählt wird.

Das Eingangs-Kennlinienfeld nach Gl. (8.51) entspricht gerade der Ebers–Moll-Gleichung (8.40). Im Normalbetrieb des Transistors als Verstärker ist $-U_{EB}/U_T \gg 1$ und $-U_{CB}/U_T \ll -1$, so dass man näherungsweise aus den Gln. (8.40) und (8.41) erhält:

$$I_E = -I_{EE}(e^{-U_{EB}/U_T} - 1) + I_{TS}(e^{-U_{BC}/U_T} - 1) \approx -I_{EE}e^{-U_{EB}/U_T} + I_{BES} \qquad (8.52)$$

Im Normalbetrieb äußert sich die Rückwirkung des Ausgangs auf den Eingang in dem konstanten Strom I_{BES}. Das Eingangs-Kennlinienfeld besteht daher aus der Kennlinie einer in Flussrichtung gepolten Diode und einem konstanten Stromanteil. Bild 8.12(a) zeigt das Kennlinienfeld schematisch und 8.12(b) ein gemessenes Kennlinienfeld des npn-Bipolartransistors BC 107. Da der Strom I_{BES} sehr klein ist, ist er bei der in 8.12(b) benutzten Auflösung der Stromachse nicht erkennbar.

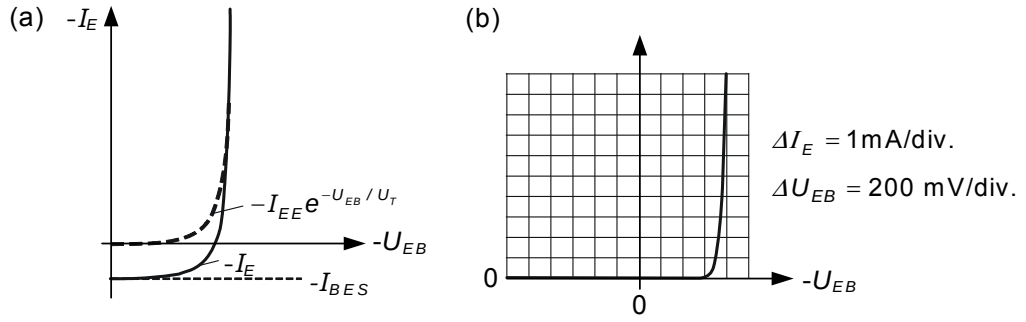


Abbildung 8.12: Eingangskennlinienfeld eines NPN-Transistors in Basisschaltung a) schematisch, b) gemessen am Transistor BC 107.

Ausgangs-Kennlinienfelder

Das gebräuchlichere der beiden möglichen Kennlinienfelder ist durch die Beziehung

$$I_{2B}\left(U_{2B}, I_{1B}\right) = I_C\left(U_{CB}, I_E\right) \tag{8.53}$$

gegeben.

Das andere Kennlinienfeld ist durch

$$I_{2B}\left(U_{2B}, U_{1B}\right) = I_C\left(U_{CB}, U_{EB}\right) \tag{8.54}$$

bestimmt.

Wir wollen nur das erstgenannte Kennlinienfeld aufzeichnen. Aus der EBERS-MOLL-Gleichung (8.47) folgt durch Elimination von I_v direkt

$$I_C = -A_N I_E + I_{CB0}\left(1 - e^{-U_{CB}/U_T}\right). \tag{8.55}$$

Bei hoher Basis-Kollektor-Sperrspannung (Normalbetrieb) erhält man den Leerlaufreststrom $I_C = I_{CB0}$ bei $I_E = 0$. Es ist

$$I_{CB0} = \left(\frac{1}{A_I} - A_N\right) I_{TS}\ . \tag{8.56}$$

Das Ausgangs-Kennlinienfeld ergibt sich daher im Normalbetrieb aus dem Leerlaufstrom I_{CB0}, der im Normalbetrieb weder von der Ausgangsspannung U_{CB} noch vom Eingangsstrom I_E abhängt, und einem zum Eingangsstrom I_E proportionalen Stromanteil -$A_N I_E$. Fig. 8.13(a) zeigt das Ausgangskennlinienfeld schematisch, während in Fig. 8.13(b) das tatsächliche Ausgangskennlinienfeld desselben npn-Transistors wie in 8.13 gezeigt ist. Der Emitterstrom wird dabei um jeweils gleiche Beträge ΔI_E (= 10 mA in 8.13(b)) erhöht. Der Stromanteil I_{CB0} ist wiederum so klein, dass er in 8.13(b) nicht erkennbar ist. Ferner ist aus 8.13(b) ablesbar, dass $I_C \simeq -I_E$ gilt, d.h. innerhalb der Ablesegenauigkeit in 8.13(b) muss $A_N \simeq 1$ geschlossen werden.

Bild 8.13(b) lässt ferner erkennen, dass der Kollektor bis zu einer gewissen Spannung U_{CB} auch in Flussrichtung betrieben werden kann, bevor der Kollektorstrom absinkt. Das Absinken des Kollektorstromes entsteht dadurch, dass bei einer Flussspannung $U_{CB} < 0$ statt des Sperrstromes I_{CB0} ein Flussstrom

$$-I_{CB0}\left(e^{-U_{CB}/U_T} - 1\right) \simeq -I_{CB0}e^{-U_{CB}/U_T} \quad \text{für} \quad -U_{CB}/U_T \gg 1$$

fließt. Die Spannung $-U^0_{CB}$, bei der der Kollektorstrom Null wird, lässt sich aus (8.55) bestimmen:

$$I_C = 0 = -A_N I_E - I_{CB0}e^{-U^0_{CB}/U_T}$$

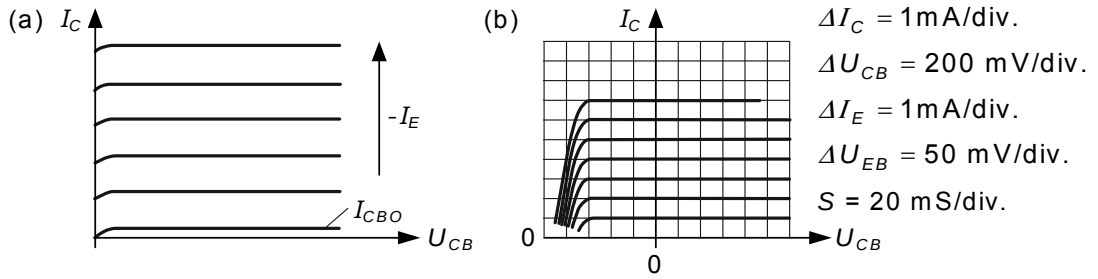


Abbildung 8.13: Ausgangskennlinienfeld eines npn-Transistors in Basisschaltung (a) schematisch, (b) Transistor BC 107 (Parameter I_E).

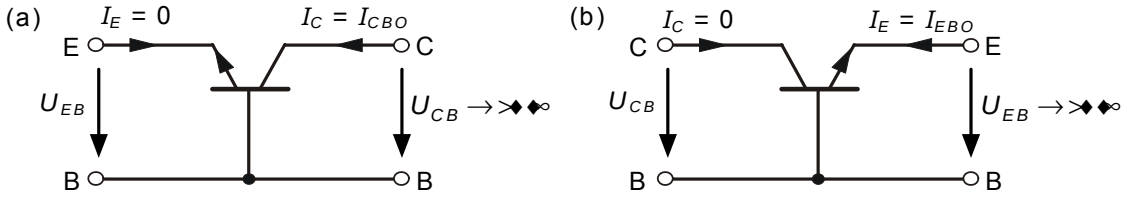


Abbildung 8.14: Definition der Leerlaufströme (a) der Basisschaltung I_{CE0} und (b) der inversen Basisschaltung I_{EC0}.

Es folgt

$$U_{CB}^0 = -U_T \ln \frac{A_N I_E}{I_{CB0}} \ . \tag{8.57}$$

Mit zunehmendem Strom I_E wird U_{CB}^0 negativ, verschiebt sich jedoch nur langsam mit dem Logarithmus aus dem Strom I_E, wie es 8.13(b) auch zeigt. Die Spannung, ab der eine Siliziumdiode leitend wird, beträgt etwa 0,6 V (vgl 8.13(b)), d.h. die Kollektordiode kann bis zu etwa 0,6 V in Flussrichtung gepolt werden, bevor der Kollektor seine Funktion verliert (s. 8.13 (b)).

Ebenso wie bei der normalen Basisschaltung, siehe Bild 8.14(a) kann auch eine inverse Basisschaltung betrachtet werden, Bild 8.14(b). Gl.(8.55) und die für den inversen Fall analoge Gleichung lauten hier.

$$\begin{array}{rcll} I_C &=& -A_N I_E + I_{CB0}\left(1 - e^{-U_{CB}/U_T}\right), & \text{(normale) Basisschaltung,} \\ I_E &=& -A_I I_C + I_{EB0}\left(1 - e^{-U_{EB}/U_T}\right), & \text{inverse Basisschaltung,} \end{array} \tag{8.58}$$

mit den symmetrisch definierten Größen

$$\begin{array}{rcll} I_{CB0} &=& I_{TS}\left(\dfrac{1}{A_I} - A_N\right) > 0, & \dfrac{I_{CB0}}{I_{EB0}} = \dfrac{A_N}{A_I}. \\ I_{EB0} &=& I_{TS}\left(\dfrac{1}{A_N} - A_I\right) > 0, & \end{array} \tag{8.59}$$

I_{CB0}, I_{EB0} werden als Leerlaufrestströme bezeichnet (es sind die Ausgangsströme, wenn der Eingang der Schaltung leerläuft). Sie werden mit dem Symbol I_{xy0} bezeichnet, wobei x die Ausgangsklemme bezeichnet, an welcher der Strom gemessen wird, y jene Klemme ist, welche dem Eingangs- und Ausgangsklemmenpaar gemeinsam ist und das Symbol "0" auf den Leerlauf am Eingang hinweist. Diese Ströme sollten eigentlich bei $U_{CB} = \infty$ bzw. $U_{EB} = \infty$ gemessen werden. Im realen Datenblatt sind die Spannungen U_{CB0} ($U_{CB} = U_{CB0}$) für die Messung von I_{CB0} bzw. U_{EB0} ($U_{EB} = U_{EB0}$) für die Messung von I_{EB0} angegeben.

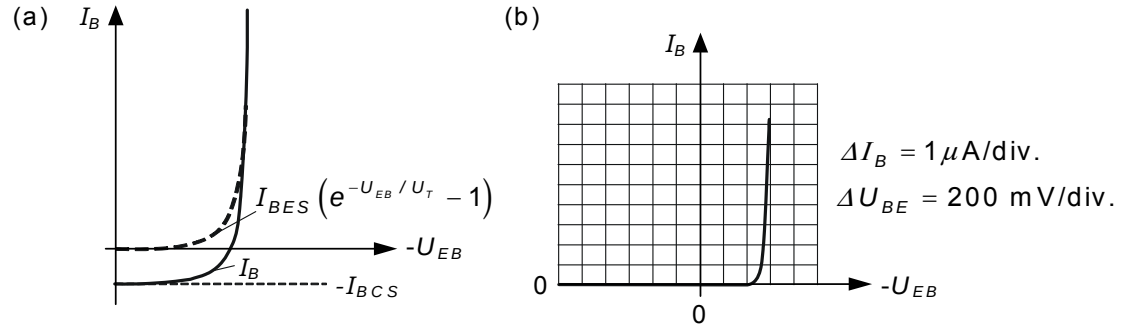


Abbildung 8.15: Eingangskennlinie eines NPN-Transistors in Emitterschaltung: (a) schematisch, (b) Transistor BC 107.

Aus (8.58) folgt auch eine Anleitung für die Messung der Größen A_N, A_I:

$$A_N = -\left.\frac{I_C - I_{CB0}}{I_E}\right|_{U_{CB}\to\infty}, \qquad A_I = -\left.\frac{I_E - I_{EB0}}{I_C}\right|_{U_{EB}\to\infty}. \tag{8.60}$$

A_N (A_I) sind die Quotienten der von den Eingangsströmen I_E (I_C) gesteuerten Anteile des Ausgangsstroms $I_C - I_{CB0}$ ($I_E - I_{EB0}$) und den Eingangsströmen I_E (I_C) bei der normalen (inversen) Basisschaltung und werden als ***Gleichstromverstärkung der normalen (inversen) Basisschaltung*** bezeichnet (A_N, $A_I > 0$).

8.5.2 Emitterschaltung

Wie bei der Basisschaltung sind auch bei der Emitterschaltung die Eingangs- und Ausgangsströme von Null verschieden, so dass es sowohl ein Eingangs- als auch ein Ausgangskennlinienfeld gibt.

Eingangs-Kennlinienfeld

$$I_{1E}(U_{1E}, U_{2E}) = I_B(U_{BE}, U_{CE}) \tag{8.61}$$
mit $U_{2E} = U_{CE}$ als Parameter des Ausgangstores

oder

$$I_{1E}(U_{1E}, I_{2E}) = I_B(U_{BE}, I_C) \tag{8.62}$$
mit $I_{2E} = I_C$ als Parameter des Ausgangstores

Aus den Ebers-Moll-Gleichungen (8.40) und (8.41) folgt für das Kennlinienfeld nach Gl. (8.61) oder (8.62):

$$\begin{aligned} I_B &= -(I_E + I_C) \\ &= (1 - A_N) I_{EE}(e^{-U_{EB}/U_T} - 1) - (1 - A_I) I_{CC} \\ &= I_{BES}(e^{-U_{EB}/U_T} - 1) - I_{BCS}\ . \end{aligned} \tag{8.63}$$

Der Eingangsstrom I_B der Emitterschaltung besteht aus dem konstanten Anteil Basis-Kollektor-Sperrstrom und einem mit der Eingangsspannung $-U_{EB}$ exponentiell ansteigenden Anteil.

Der exponentiell ansteigende Anteil ist wegen $(1 - A_N) \ll 1$ sehr viel kleiner als der entsprechende Eingangsstromanteil in der Basisschaltung. Bild 8.15(a) zeigt die Eingangskennlinie der Emitterschaltung im Normalbetrieb schematisch, Bild 8.15(b) ein tatsächliches Kennlinienfeld, das am selben Transistor wie die Kennlinienfelder der Basisschaltung aufgenommen wurde. Da im Normalbetrieb die Rückwirkung des Ausganges auf den Eingang von der Ausgangsspannung U_{CE} unabhängig ist, genügt die Angabe eines Kennlinienfeldes.

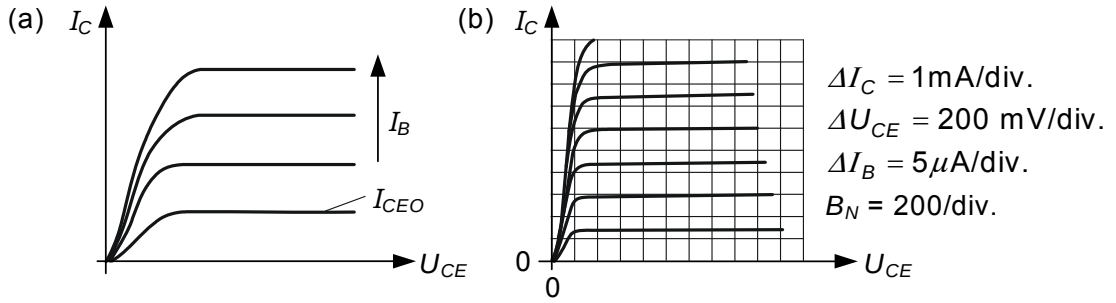


Abbildung 8.16: Ausgangskennlinienfeld eines npn-Transistors in Emitterschaltung: (a) schematisch, (b) Transistor BC 107 (Parameter I_B).

Ausgangs-Kennlinienfeld

Das gebräuchlichere der beiden möglichen Kennlinienfelder ist

$$I_{2E}(U_{2E}, I_{1E}) = I_C(U_{CE}, I_B) \quad \text{mit} \quad U_{CE} = U_{CB} + U_{BE} \tag{8.64}$$

Aus Gl. (8.55)

$$I_C = -A_N I_E + I_{CB0}\left(1 - e^{-U_{CB}/U_T}\right)$$

und

$$I_E = -(I_B + I_C)$$

folgt

$$I_C = B_N I_B + (B_N + 1) I_{CB\,0}\left(1 - e^{-U_{CB}/U_T}\right) \tag{8.65}$$

mit

$$B_N = \frac{A_N}{1 - A_N} \gg 1 \quad \text{für} \quad A_N < 1 \tag{8.66}$$

B_N wird ***Stromverstärkung der Emitterschaltung*** genannt. Mit der Abkürzung

$$I_{CE0} = (B_N + 1) I_{CB0} \gg I_{CB0} \tag{8.67}$$

folgt schließlich

$$I_C = B_N I_B + I_{CE0}\left(1 - e^{-U_{CB}/U_T}\right). \tag{8.68}$$

Nach dieser Gleichung besteht der Kollektorstrom wiederum aus einem konstanten Anteil I_{CE0}, der jedoch wesentlich größer als in der Basisschaltung ist, und einem zu I_B proportionalen Anteil (Fig. 8.16(a)). Das tatsächliche Kennlinienfeld, Bild 8.16(b), zeigt Abweichungen von der schematischen Darstellung aufgrund des Early-Effektes, d.h. der Basisweitenmodulation, sowie aufgrund von Änderungen der Stromverstärkung A_N mit den Betriebsströmen und -spannungen. Aus der Differentiation der Definitionsgleichung (8.66) folgt

$$\frac{dB_N}{dA_N} = \frac{1}{(1 - A_N)^2} = \frac{\Delta B_N}{\Delta A_N}$$

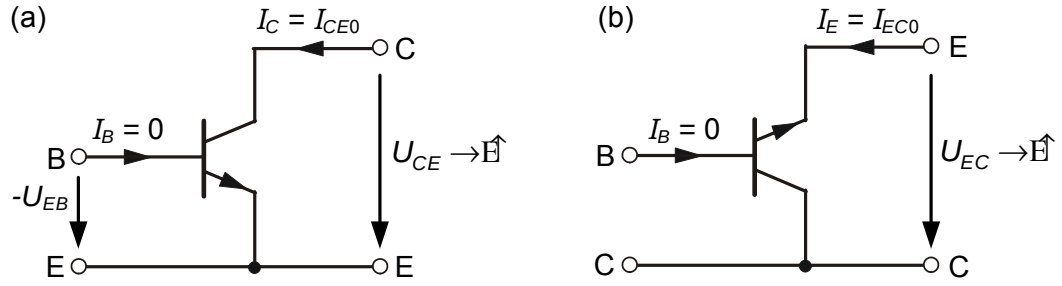


Abbildung 8.17: Definition der Leerlaufrestströme (a) der Emitterschaltung I_{CE0} und (b) der inversen Emitterschaltung (= Kollektorschaltung) I_{EC0}.

bzw. nach Trennung der Variablen und Division durch Gl. 8.66:

$$\frac{\Delta B_N}{B_N} = \frac{\Delta A_N}{A_N} \frac{1}{1 - A_N} \tag{8.69}$$

Beträgt beispielsweise $A_N = 0,995 (\triangleq B_N = 199)$, so führt eine Änderung von $A_N = 0,995$ auf $A_N + \Delta A_N = 0,996$ (entsprechend einer relativen Änderung von 0,1 %) zu einer Änderung $\Delta B_N = 40$ auf den neuen Wert $B_N + \Delta B_N = 239$, was einer relativen Änderung von fast 25 % entspricht. Für den praktischen Einsatz bipolarer Transistoren lässt sich aus diesem Ergebnis ableiten, dass der Ausgangsstrom I_C des Transistors in Emitterschaltung äußerst empfindlich auf sehr geringe Änderungen der Stromverstärkung A_N reagiert, sei es bei Änderung des Arbeitspunktes, der Temperatur oder beim Auswechseln von Transistoren gleichen Typs. Will man diese Einflüsse ausschalten oder zu mindest stark vermindern, so muss der Transistor in Basisschaltung betrieben werden oder es muss eine starke Kompensation (Gegenkopplung) in der Schaltung vorgesehen werden. Einige wesentliche Ursachen für die Abhängigkeit der Stromverstärkung A_N von Strömen und Spannungen werden im Abschnitt 8.7 besprochen.

Die Ausgangsströme von normaler (inverser) Emitterschaltung lassen sich aus (8.58) durch die Substitution des jeweiligen Stromes gemäß Gl. (8.1) angeben zu

$$\begin{array}{rcll} I_C &=& B_N I_B + I_{CE0}\left(1 - e^{-U_{CB}/U_T}\right), & \text{(normale) Emitterschaltung,} \\ I_E &=& B_I I_B + I_{EC0}\left(1 - e^{-U_{EB}/U_T}\right), & \text{inverse Emitterschaltung,} \end{array} \tag{8.70}$$

mit den zum Teil bereits oben definierten Größen

$$\begin{array}{rcll} B_N = \dfrac{A_N}{1 - A_N} = \dfrac{I_{TS}}{I_{BES}}, & & B_I = \dfrac{A_I}{1 - A_I} = \dfrac{I_{TS}}{I_{BCS}}, & \\ I_{CE0} &=& I_{CB0}(1 + B_N) > 0, & \dfrac{I_{CE0}}{I_{EC0}} = \dfrac{B_N}{B_I}. \\ I_{EC0} &=& I_{EB0}(1 + B_I) > 0, & \end{array} \tag{8.71}$$

Da für einen brauchbaren Transistor A_N nur wenig kleiner als eins ist, gilt $B_N \gg 1$ und somit für den Leerlaufreststrom der Emitterschaltung $I_{CE0} \gg I_{CB0}$. Abb. 8.17 zeigt die Definition der Leerlaufrestströme für Emitterschaltung und der inversen Emitterschaltung analog zur oben diskutierten Basisschaltung.

Die Gleichstromverstärkungen in Emitterschaltung B_N, B_I ergeben sich aus

$$B_N = \left.\frac{I_C - I_{CE0}}{I_B}\right|_{U_{CE} \to \infty}, \qquad B_I = \left.\frac{I_E - I_{EC0}}{I_B}\right|_{U_{EC} \to \infty}. \tag{8.72}$$

Im Rahmen des Ebers-Moll-Modells sind diese Stromverstärkungen nicht vom Arbeitspunkt abhängig.

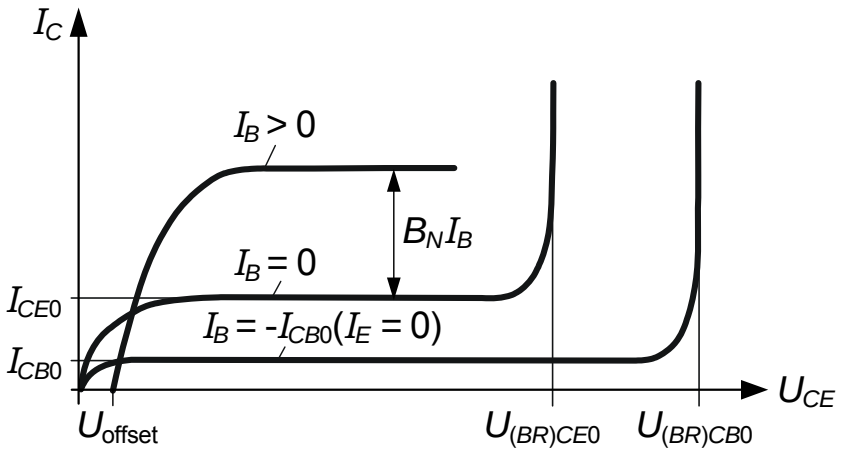


Abbildung 8.18: Ausgangskennlinien der Emitterschaltung mit dem Basisstrom als Parameter. Für $I_E = 0$ ist der Transistor ausgeschaltet.

Abb. 8.18 zeigt nochmals den prinzipiellen Verlauf der Ausgangskennlinien der Emitterschaltung auch für große Kollektor-Emitterspannungen (nicht maßstäblich, beachte $I_{CE0} = I_{CB0}(1 + B_N)$!) .

Die Kennlinien für $I_B = 0$, $I_E = 0$ gehen durch den Ursprung.

Das Abknicken der Kennlinien für $U_{CE} \to 0$ ist darauf zurückzuführen, dass U_{CB} bei konstantem I_B negative Werte annimmt und der Transistor zufolge der Rückinjektion der CB-Diode (I_r steigt) in den Übersteuerungsbereich gerät. Für $I_B > 0$ gehen die Kennlinien nicht durch den Ursprung. Es bleibt für $I_C = 0$ eine endliche Spannung U_{CE} am Transistor (die sogenannte Offset-Spannung U_{offset}), für die man für U_{EB}, $U_{CB} \ll 1$ aus (8.47) berechnet:

$$I_C = 0 = I_v - \frac{I_r}{A_I} \approx I_{TS}\, e^{-U_{EB}/U_T} - \frac{I_{TS}\, e^{-U_{CB}/U_T}}{A_I}, \qquad U_{\text{offset}} = -\left(U_{EB} - U_{CB}\right), \qquad U_{\text{offset}} = U_T \ln\left(\frac{1}{A_I}\right). \tag{8.73}$$

Beim realen Transistor steigen die Kennlinien in Abb. 8.18 für steigende Werte von U_{CE} durch den Early-Effekt leicht an, siehe Diskussion in Bild 8.10.

8.6 Durchbruchsverhalten

Der Durchbruch der CB-Diode bei zu hoher Kollektor-Basis-Sperrspannung erfolgt beim Transistor zufolge Trägervervielfachung durch alle Träger, welche die Verarmungszone der CB-Diode durchqueren.

Fall 1: Bei $I_E = 0$ ist $I_C = -I_B = I_{CB0}$(Gl. (8.47) mit $I_{TS} = 0$ $U_{CB} >> 0$); das bedeutet, dass alle Träger, welche den Strom I_{CB0} verursachen, auch durch die CB-Diode durchlaufen. Der Strom steigt bei starkem Anstieg von U_{CB} zufolge Lawinenmultiplikation, siehe (6.87), auf den Wert

$$I'_C = M_0 I_{CB0}, \qquad M_0 = \frac{1}{1 - \left(\frac{U_{CB}}{U_{(BR)CB0}}\right)^n}. \tag{8.74}$$

Da $U_{CE} \approx U_{CB}$ ist die in Abb. 8.18 eingezeichnete Durchbruchspannung $U_{CE} \approx U_{(BR)CB0}$.

Fall 2: Wenn dagegen $I_B = 0$ ist, bilden die in die Basis einströmenden, in der Sperrschicht der CB-Diode vervielfachten Löcher eine positive Raumladung, die zunächst nicht neutralisierbar erscheint: Die Löcher sind Majoritätsträger und können wegen der Bedingung $I_B = 0$ nicht nach außen abfließen. Tatsächlich stellt sich Quasineutralität in der Basis dadurch ein, dass sich die Löcher über die negative Raumladung der ionisierten Akzeptoren in der Verarmungszone der EB-Diode schieben; die Breite der Verarmungszone der EB-Diode im Basismaterial nimmt ab. Da sich aber in der Sperrschicht der EB-Diode die positive Raumladung der ionisierten Donatoren auf der Emitterseite und die negative Raumladung der ionisierten Akzeptoren auf der Basisseite kompensieren müssen, ist die Breite der Verarmungszonen auf beiden Seiten des pn-Übergangs der EB-Diode kleiner geworden und die Potentialdifferenz am pn-Übergang hat abgenommen. Die Spannung an der EB-Diode ändert sich um $\Delta U_{bE} > 0$, der Emitter kommt zur Injektion, die Elektronen durchqueren die Basis und tragen in der RLZ der CB-Diode zur Erzeugung weiterer, vervielfachter Elektronen und Löcher bei; dadurch ist in diesem Fall (dem sogenannten ***Injektionsdurchbruch***) der Durchbruch schon bei kleineren Spannungen zu erwarten. Wegen $I_B = 0$ ist $I_E = -I_C$ und man erhält aus Gl. (8.55) $I_C = A_N I_C + I_{CB0}$. Bei Erhöhen der Sperrspannung der CB-Diode geht I_C in den vervielfachten Strom I'_C über, und jeder Träger, der die CB-Diode durchquert, trägt zur Vervielfachung bei. Es gilt somit:

$$\begin{array}{rcl} I'_C & = & M_0(A_N I'_C + I_{CB0}), \\ I'_C & = & \frac{M_0 I_{CB0}}{1 - M_0 A_N}. \end{array} \tag{8.75}$$

Der Durchbruch erfolgt bei jener Spannung $U_{CB} \approx U_{CE} = U_{(BR)CE0}$, bei der $M_0 A_N = 1$ gilt. Mit (8.74) folgt

$$U_{(BR)CE0} = U_{(BR)CB0} \sqrt[n]{1 - A_N}. \tag{8.76}$$

Mit $1 - A_N = 10^{-2}$, $2 \leq n \leq 4$ sieht man, dass die Durchbruchspannung beim Injektionsdurchbruch drastisch reduziert wird.

8.7 Emitterwirkungsgrad, Basistransportfaktor und Stromverstärkungen

Im vorhergehenden Kapitel wurde erläutert, dass es für den Normalbetrieb des Transistors wünschenswert ist, wenn $A_N \to 1$; $A_I \ll 1$ erreicht werden kann. Bisher wurde allerdings die Rekombination in der Basis, d.h. der Basisstromanteil I_{BB} vernachlässigt. Zu welchem Grad dies möglich ist, soll im folgenden diskutiert werden.

8.7.1 Die Vorwärtsstromverstärkung

Die ***Definition der Vorwärtsstromverstärkung*** lautet

$$A_N = -\frac{I_C}{I_E} = \frac{|I_T|}{|I_T| + I_{BE} + I_{BB}} \tag{8.77}$$

$$= \frac{|I_T|}{|I_T| + I_{BB}} \frac{|I_T| + I_{BB}}{|I_T| + I_{BB} + I_{BE}} \tag{8.78}$$

$$= \delta_T \ \gamma_E \ , \tag{8.79}$$

wobei $I_{BC} \approx 0$ im gesetzt wurde - was im Vorwärtsbetrieb recht gut stimmt. Ferner haben wir die Vorwärts-Stromverstärkung als das Produkt aus dem Emitterwirkungsgrad γ_E und dem Basistransportfaktor δ_T geschrieben.

Der Emitterwirkungsgrad ist definiert als

$$\gamma_E = \frac{|I_T| + I_{BB}}{|I_T| + I_{BB} + I_{BE}} = \frac{1}{1 + \frac{I_{BE}}{|I_T| + I_{BB}}} \approx \frac{1}{1 + \frac{I_{BE}}{|I_T|}} \tag{8.80}$$

Er gibt an welcher Bruchteil des Emitterstromes in die Basis injiziert wird, siehe Bild 8.6.

Der Basistransportfaktor

$$\delta_T = \frac{|I_T|}{|I_T| + I_{BB}} = \frac{1}{1 + \frac{I_{BB}}{|I_T|}}, \tag{8.81}$$

gibt den Bruchteil des in die Basis injizierten Minoritätsträgerstromes an, welcher vom Kollektor abgesaugt wird. Beide Faktoren sind demnach kleiner als 1.

Aus den Gleichungen (8.17), (8.18) und (8.25) folgt für den Fall, dass die Emitterlänge l_{nE} wesentlich kleiner als die Diffusionslänge L_{pE} ist

$$\gamma_E = \frac{1}{1 + \frac{\mu_{pE} p_{noE} L_{nB}}{\mu_{nB} n_{poB} l_{nE}} \tanh\left(\frac{w}{L_{nB}}\right)} \tag{8.82}$$

Werden die Gleichgewichts-Minoritätsträgerdichten in Basis und Emitter durch die Dotierungen (8.21) und (8.10) ausgedrückt, so folgt

$$\gamma_E = \frac{1}{1 + \frac{\mu_{pE} n_{AB} L_{nB}}{\mu_{nB} n_{DE} l_{nE}} \tanh\left(\frac{w}{L_{nB}}\right)} \tag{8.83}$$

$$\delta_T = \frac{1}{\cosh\left(\frac{w}{L_{nB}}\right)}. \tag{8.84}$$

Damit $A_N \to 1$ müssen wir die verfügbaren Parameter so wählen, dass $\delta_T \to 1$ als auch $\gamma_E \to 1$:

- Damit $\delta_T \to 1$, muss $\cosh\left(\frac{w}{L_{nB}}\right) \to 1$ gehen. **Dies erfordert** $w \ll L_{nB}$. Typische Diffusionslängen in dotiertem Material sind von der Größenordnung 1μm. Basisweiten $w < 0,1\,\mu$m werden heute erzielt, so dass $w/L_{nB} < 0,1$ möglich ist. Daraus lassen sich die in Tabelle 10.1 angegebenen Transportfaktoren erzielen.

- Wenn auch $\gamma_E \to 1$ gelten soll, muss der 2. Summand im Nenner der Gl. (8.83) gegen Null gehen. Da $w/L_{nB} \ll 1$, ist auch $\tanh(w/L_n) \ll 1$. Der Quotient wird noch kleiner, wenn das **Dotierstoffverhältnis** n_{AB}/n_{DE} **klein gemacht** wird, d.h. wenn der Emitter wesentlich höher als die Basis dotiert wird.Mit heute üblichen Werten von $n_{DE} > 10^{20}\text{cm}^{-3}$ erreicht man die in Tab. 10.1 ebenfalls angegebenen Emitterwirkungsgrade. Bei den Zahlenwerten wurde angenommen, dass größenordnungsmäßig μ_{pE} mit μ_{nB} und L_{nB} mit l_{nE} vergleichbar sind. Die daraus resultierende Stromverstärkung A_N und die zugehörigen Werte B_N sind in Tab. 10.1 aufgelistet.

$\frac{w}{L_{nB}}$	$0,05$	$0,1$
$\cosh\left(\frac{w}{L_{nB}}\right)$	$1,001$	$1,005$
$\tanh\left(\frac{w}{L_{nB}}\right)$	$0,0499$	$0,0996$
δ_T	$0,998$	$0,995$
γ_E	$0,9999$	$0,9999$
A_N	$0,9987$	$0,9949$
B_N	768	196
γ_C	$0,666$	$0,500$
A_I	$0,666$	$0,498$

Tabelle 8.3: Transportfaktor, Emitter- und Kollektorwirkungsgrade und Gleichstromverstärkungen für verschiedene w/L_{nB}- Werte.

8.7.2 Die Rückwärtsstromverstärkung

Analog ist die ***Rückwärtsstromverstärkung*** definiert durch

$$A_I = -\frac{I_E}{I_C} = \frac{|I_T| + I_{BC} + I_{BB}}{|I_T|}$$

woraus in Analogie zur Berechnung von A_N folgt:

$$A_I = \frac{1}{\cosh\frac{w}{L_{nB}}} \cdot \frac{1}{1 + \frac{\mu_{pC} n_{AB} L_{nB}}{\mu_{nB} n_{DC} L_{pC}} \tanh\left(\frac{w}{L_{nB}}\right)} \tag{8.85}$$

$$= \delta_T \cdot \gamma_C \tag{8.86}$$

Der Transportfaktor δ_T ist der gleiche wie für A_N, während der Kollektorwirkungsgrad γ_C die Kollektordotierung n_{DC} statt der Emitterdotierung n_{DE} in A_N enthält. Ferner ist angenommen, dass die Ausdehnungen der Kollektorzone groß gegen die Diffusionslänge L_{pC} ist.

Damit $A_I \ll 1$ müssen wir δ_T und γ_E wieder entsprechend optimieren:

- Der Transportfaktor δ_T ist zur Erzielung eines möglichst großen A_N bereits festgelegt.
- Da A_I möglichst klein sein soll, muss γ_C möglichst klein sein. Dies erfordert, dass der 2. Summand im Nenner möglichst groß wird. Da $\tanh(w/L_{nB})$ bereits festgelegt ist und wiederum $\mu_{pC} L_{nB} \simeq \mu_{nB} L_{pC}$ gelten soll, kann der Nenner nur groß werden, wenn $n_{DC} \ll n_{AB}$ erreicht werden kann. Da wegen $n_{DE} \gg n_{AB}$ der Wert $n_{AB} \simeq 10^{17}\mathrm{cm}^{-3}$ beträgt, wird für Silizium-Transistoren $n_{DC} = 10^{16}\,\mathrm{cm}^{-3}$ gewählt. Die daraus sich ergebenden Werte für γ_C und A_I sind in Tab. 10.1 ebenfalls angegeben.

8.8 Kleinsignal-Ersatzschaltbilder

Die Kennlinienfelder der Transistoren sind nicht linear. Für kleine Strom- bzw. Spannungsänderungen lassen sich die Kennlinien in einem gewissen Bereich linearisieren. Dieses Verfahren erlaubt die Beschreibung des Transistors durch einen Satz von Vierpolparametern und die Aufstellung von einfachen Ersatzschaltbildern. Für eine breiten Anwendungsbereich der Transistoren, z.B. lineare Verstärker, interessiert das Verhalten bei kleinen Strom- bzw. Spannungsänderungen, ΔI bzw ΔU. Die Werte der Ersatzschaltbildelemente werden in der Praxis durch direkte Messung oder aus den Kennlinienfeldern ermittelt. Besondere Bedeutung kommt den Kleinsignal-Stromverstärkungsfaktoren für Emitter- und Basisschaltung zu. Für sie gilt:

$$\begin{array}{lcll} \beta & = & \frac{\Delta I_C}{\Delta I_B} & \text{Stromverstärkung in Emitterschaltung,} \\ \alpha & = & -\frac{\Delta I_C}{\Delta I_E} < 1 & \text{Stromverstärkung in Basisschaltung,} \end{array} \tag{8.87}$$

Da die Summe der in den Transistor fließenden Ströme Null sein muss, gilt wie bereits bei den Gleichstromverstärkungen gefunden

$$\beta \quad = \quad \tfrac{\alpha}{1-\alpha} \gg 1 \ . \tag{8.88}$$

8.8.1 Einfaches Niederfrequenzersatzschaltbild

Vereinfachende Annahmen im Normalbetrieb des Transistors:

- Im Normalbetrieb des Transistors können wir in Basis- und Emitterschaltung zunächst die Rückwirkung der Ausgangsspannung U_{CB} auf den Eingang vernachlässigen; damit ist auch $\Delta I_{BC}/\Delta U_{CB} \cong 0$.
- Weiter kann die Generation und Rekombination in der Basis vernachlässigt werden; $I_{BB} \cong 0$.

Der Basisstrom ist dann nur durch die Diffusionszone im Emitter, d.h. I_{BE} bestimmt.

Der Kollektorstrom ist durch den Transferstrom, welcher einer Diodenkennlinie in Abhängigkeit von der Eingangsspannung U_{EB} folgt, gegeben. Aus dem in Kapitel 6.3.3 besprochenen Kleinsignalverhalten der Diode folgt dann

$$\Delta I_C = -g_m \Delta U_{EB} \tag{8.89}$$

mit der Steilheit (dem ***Kleinsignal-Leitwert (transconductance)***)

$$g_m \equiv \left.\frac{dI}{dU}\right|_{U_0} = \frac{|I_C + I_{TS}|}{U_T} \approx \frac{|I_C|}{U_T} \approx \left|\frac{I_C}{[\mathrm{mA}]}\right| \ 0,04 \ [\mathrm{S}]. \tag{8.90}$$

(Beachte: mit $I_C \overset{\text{Gl. (8.47)}}{=} I_{TS}\left(e^{-U_{EB}/U_T} - 1\right) - (I_{TS} + I_{BCS})\left(e^{-U_{CB}/U_T} - 1\right)$ ergibt sich nach Ableiten und unter Berücksichtigung obiger Annahmen $dI_C/dU_{EB} \cong -(1/U_T) I_{TS} e^{-U_{EB}/U_T} = -(1/U_T)\left[I_{TS}\left(e^{-U_{EB}/U_T} - 1\right) + I_{TS}\right] = -(1/U_T)\left[I_C + I_{TS}\right]$)

Bei Raumtemperatur gilt

$$\frac{g_m}{[\mathrm{S}]} \approx 0,04 \frac{|I_C|}{[\mathrm{mA}]}. \quad \text{(Bem.: 1 [S] = 1[Siemens] = 1}[\frac{1}{\Omega}]\text{)} \tag{8.91}$$

Der Basisstrom hat drei Anteile, von denen einer I_{BC}, im Normalbetrieb von der Aussteuerung unabhängig ist (Basis-Kathoden Sperrstromanteil) und daher - wie oben diskutiert - zum Kleinsignalverhalten keinen Beitrag liefert. Die beiden anderen Anteile erfüllen nach Gl.(8.37) und Gl.(8.39)

$$\begin{aligned} \frac{\Delta I_B}{\Delta I_C} &= \frac{\Delta I_B/\Delta U_{EB}}{\Delta I_C/\Delta U_{EB}} = \frac{I_{BES} + I_{TS}\frac{1}{2}\left(\frac{w}{L_n}\right)^2}{I_{TS}} \\ &= \frac{D_{pE}\ w\ n_{AB}}{D_{nB}\ l_E\ n_{DE}} + \frac{1}{2}\left(\frac{w}{L_{nB}}\right)^2 = \delta = \frac{1}{\beta_0} \end{aligned} \tag{8.92}$$

Die Größe δ ist dimensionslos und kennzeichnet die Minoritätsträgerinjektion in den Emitter und die Rekombination in der Basis. Die Werte liegen im allgemeinen im Bereich 0,001 $<\delta <$0,1. Der Kehrwert von δ ist die ***Kleinsignal-Niederfrequenz-Stromverstärkung in Emitterschaltung*** β_0. l_E ist die Emitterweite, falls $L_{pE} >> l_E$.

Für die Kleinsignalanteile von Basis- und Kollektorstrom gilt dann mit (8.89) und (8.92)

$$\Delta I_C = -g_m \Delta U_{EB.}, \tag{8.93}$$

$$\Delta I_B = -g_m \delta \Delta U_{EB.}. \tag{8.94}$$

Diese Gleichungen definieren das in Bild 8.19 dargestellte Ersatzschaltbild.

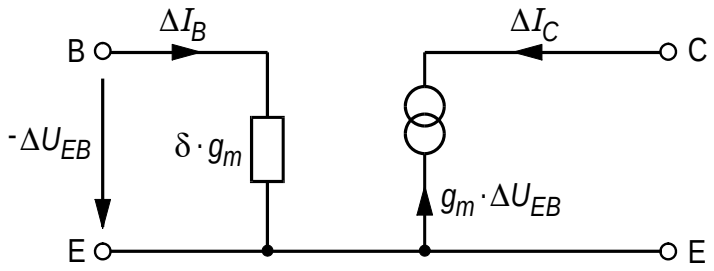


Abbildung 8.19: Einfaches Kleinsignal-Ersatzschaltbild des Transistors für Emitterschaltung [3].

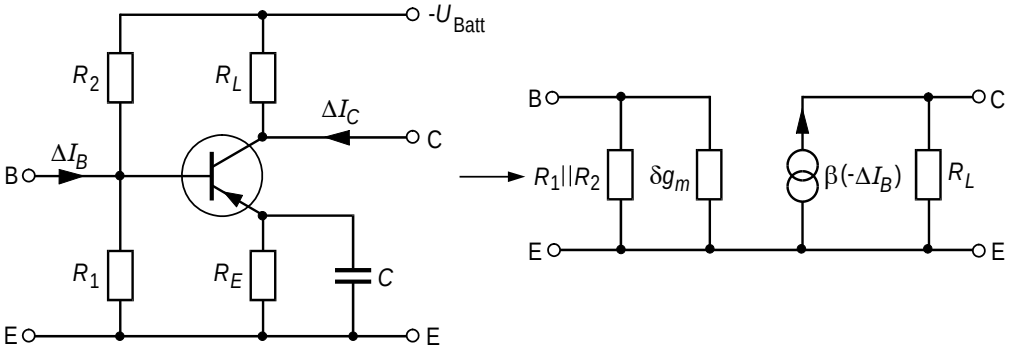


Abbildung 8.20: Emitterschaltung eines Transistors mit Ersatzschaltbild. Die Widerstände R_1, R_2 und R_E dienen zur Einstellung des Arbeitspunktes. Die Kapazität dient der Wechselstromüberbrückung von R_E [3].

Bild 8.20 zeigt links eine Transistor-Verstärkerschaltung und rechts das dazugehörige einfache Wechselstrom-Ersatzschaltbild.

Der spannungsgesteuerte Stromgenerator in Bild 8.20 kann mit Hilfe von Gl. (8.93) in einen stromgesteuerten Generator umgewandelt werden.

Die **Spannungsverstärkung (der Kleinsignal-Spannungsübertragungsfaktor)** ist definiert als

$$V_u = \frac{\Delta U_{EC.}}{\Delta U_{EB.}}. \tag{8.95}$$

Der Ersatzschaltung im Bild 8.20 entnimmt man, dass die Spannung in der Ausgangsschleife nach Kirchhoff gerade $R_L\ \Delta I_C + \Delta U_{EC.} = 0$ ist. Damit lässt sich $\Delta U_{EC.}$ eliminieren. Mit Gl. (8.93) lässt sich auch noch $\Delta U_{EB.}$und man findet für V_u

$$V_u = -g_m R_L. \tag{8.96}$$

Bsp.:Für die Daten $\delta = 0.004$; $I_C = 10$ mA, $R_L = 1$ kΩ erhalten wir

Stromverstärkung:	V_i	=	$\beta_0 = \frac{1}{\delta} =$	250,
Spannungsverstärkung:	V_u	=	$-g_m R_L =$	−400,
Leistungsverstärkung:	V_P	=	$\lvert V_i \cdot V_u \rvert =$	100000.

8.8.2 Ladungsspeicher- und Basisweitenmodulations Effekte

1. Ladungsspeicher-Effekte

Wie wir bereits bei der pn-Diode gesehen haben, sind bei pn-Übergängen zwei Ladungsspeicherungseffekte von Bedeutung: Erstens die Speicherung von Majoritätsträgern an den Rändern von

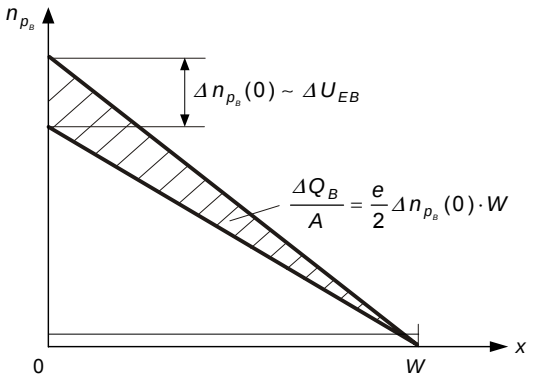


Abbildung 8.21: Änderung der Minoritätsträgerdichteverteilung in der Basis durch eine Änderung der Emitter-Basisspannung ΔU_{EB} (A:Querschnittsfläche) des aktiven Teils des Transistors [3].

Raumladungszonen, welche durch die Sperrschichtkapazität erfasst werden. Zweitens die Speicherung von Minoritätsträgern in den neutralen Diffusionszonen. Die injizierten Minoritätsträger müssen bei einer Änderung (Verringerung) der Diodenspannung wieder abgesaugt werden, sofern sie nicht durch Rekombination verschwinden. Für kleine Signale beschreibt die Diffusionskapazität diesen Effekt. Wie in Abschnitt 5.6.2 gesehen, ist die Diffusionskapazität in langen Diffusionszonen dem halben Quotienten aus der Änderung der Minoritätsträgerladung ΔQ und der verursachenden Spannungsänderung ΔU proportional. Bei einer kurzen Diffusionszone wie wir sie in der Basiszone haben, ist die Rekombination vernachlässigbar. Es folgt daher für die Minoritätsträgeränderung in der Basis im Normalbetrieb, siehe Bild 8.21

$$\Delta Q_B = -\frac{1}{2} e\ A\ w\ \Delta n_{pB}\left(x_{BE}\right) = -\frac{1}{2} e\ A\ w\ \frac{\Delta n_{pB}\left(x_{BE}\right)}{\Delta U_{EB}} \Delta U_{EB} \tag{8.97}$$

$$\overset{\text{Gl.}(8.11)}{=} -\frac{1}{2} ewA\ \left(-n_{poB}\, e^{-\frac{U_{EB}}{U_T}} \frac{\Delta U_{EB}}{U_T}\right) \tag{8.98}$$

$$\overset{\text{Gl. }(8.90)}{=} -\frac{1}{2} ewA\ g_m \frac{U_T}{|I_C|} \left(-n_{poB}\, e^{-\frac{U_{EB}}{U_T}} \frac{\Delta U_{EB}}{U_T}\right) \tag{8.99}$$

$$\cong -\frac{1}{2} ewA\ g_m \frac{U_T}{|I_C|} \left(-n_{poB} \left(e^{-\frac{U_{EB}}{U_T}} - 1\right) \frac{\Delta U_{EB}}{U_T}\right) \tag{8.100}$$

$$= \frac{1}{2} ewA \frac{g_m n_{poB}}{I_{TS}} \Delta U_{EB}$$

$$\overset{\text{Gl. }(8.16)}{=} \frac{1}{2} \frac{w^2}{D_{nB}} g_m \Delta U_{EB}. \tag{8.101}$$

Bzw. ist mit dieser Ladungsänderung die Basis-Kapazität

$$C_B = \frac{\Delta Q_B}{\Delta U_{EB}} = \frac{1}{2} \frac{w^2}{D_{nB}} g_m \tag{8.102}$$

verknüpft. Diese Basis-Kapazität muss im Ersatzschaltbild ergänzt werden.

Darüber hinaus, kann das Ersatzschaltbild um die Basis-Emitter- und die Basis-Kollektor-Übergängskapazitäten, C_{BE} und C_{BC} erweitert werden. Bei C_{BE} wird es sich im Normalbetrieb um eine Sperrschichtkapaziät und Diffusionskapazität handeln und bei C_{BC} primär um eine Sperrschichtkapazität. Es ergibt sich dann das Wechselstrom-Ersatzschaltbild 8.22.

Der Ladestrom der Basiskapazität C_B fließt zusätzlich über die Basisklemme, wodurch die Stromverstärkung in Emitterschaltung β sich verkleinert. Für eine Wechselspannung der Frequenz

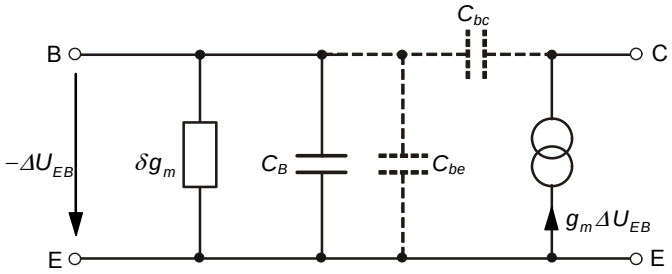


Abbildung 8.22: Einfaches Kleinsignal-Ersatzschaltbild des Transistors in Emitterschaltung [3].

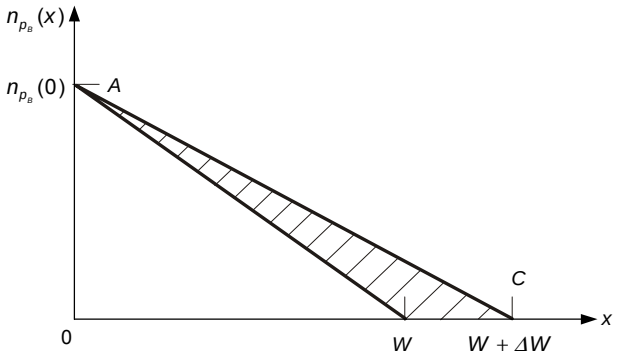


Abbildung 8.23: Änderung der Minoritätsträgerverteilung in der Basis durch ΔU_{CB}.

ω erhält man unter Vernachlässigung der Sperrschichtkapazitäten für Bild 8.22

$$\Delta I_B = [\delta g_m + j\omega C_B]\,\Delta U_{EB} = -\left(\delta + j\omega \frac{1}{2}\frac{w^2}{D_{nB}}\right) g_m\ \Delta U_{EB}, \tag{8.103}$$

$$\Delta I_C = -g_m \Delta U_{EB} \tag{8.104}$$

und damit für die Stromverstärkung

$$\beta = \frac{\Delta I_C}{\Delta I_B} = \frac{1}{\delta + j\omega \frac{1}{2}\frac{w^2}{D_{nB}}}. \tag{8.105}$$

Die Frequenz, bei welcher die Stromverstärkung in Emitterschaltung auf eins abgefallen ist, wird als ***Transitfrequenz (common-emitter cutoff frequency)*** f_T bezeichnet:

$$f_T = \frac{1}{\pi}\frac{D_{nB}}{w^2} = \frac{1}{2\pi\tau_T}. \tag{8.106}$$

2. Basisweitenmodulation: Early-Effekt

Wie bei Bild 8.10 bereits diskutiert, hängt die Basisweite sowohl von der Emitter–Basis-Spannung als auch von der Kollektor-Basis-Spannung ab. Da sich im Normalbetrieb die Basisweite der Emitter-Basis-Spannung (wegen der hohen Dotierung) nur unwensentlich ändert, führt nur die Kollektor-Basis-Spannung zu einer merklichen Basisweitenmodulation.

Die Stärke des Einflusses von U_{CB} auf w wird durch eine (nahezu) temperaturunabhängige ***Early-Spannung*** V_A oder durch einen ***Rückwirkungsfaktor*** η_w charakterisiert, siehe Bild 8.23

$$\frac{1}{V_A} = -\frac{1}{w}\frac{\partial w}{\partial U_{CB}} > 0, \qquad \eta_w = \frac{U_T}{V_A}. \tag{8.107}$$

Das bedeutet, die Basisweitenänderung Δw infolge einer Änderung der Kollektor-Basis-Spannung um ΔU_{CB} ist gegeben durch

$$\Delta w = \frac{\partial w}{\partial U_{CB}} \Delta U_{CB} = -w \frac{\eta_w}{U_T} \Delta U_{CB} \tag{8.108}$$

Aus Bild 8.23 ist ersichtlich, dass sich durch den Early-Effekt 1.) die Steigung der Minoritätsträgerdichte und 2.) die Gesamtzahl der gespeicherten Minoritätsträger in der Basis verändert. Die Steigung der Minoritätsträgerdichte beeinflusst sowohl den Transfer- I_T bzw. Kollektor-Strom I_C, als auch den Rekombinations-Generations-Basisstrom I_{BB} und die damit verbundene Basiskapazität C_B. Hingegen werden durch die Basisweitenmodulation I_{BC} und I_{BE} nur unwesentlich beeinflusst. In einem ersten Schritt werden wir also untersuchen wie die Steigung der Minoritätsträgerdichte aufgrund einer Modulation in U_{CB} den Kollektor-Strom I_C und den Basis-Strom I_B beeinflusst.

2.1. Beiträge von der Modulation der Minoritätsträgerdichtesteigung Für I_C gilt nach Gl.(8.37) und (8.44) im Normalbetrieb

$$I_C = -I_T - I_{BC} \approx eAD_{nB} \frac{n_{pB}(x_{BE}) - n_{pB}(x_{BC})}{w} - I_{BCS},$$

und damit folgt für eine Kollektor-Basisspannungsänderung ΔU_{CB} mit (8.107) und der Steilheit nach (8.90)

$$\Delta I_C = -\frac{\partial I_T}{\partial w} \frac{\partial w}{\partial U_{CB}} \Delta U_{CB} \tag{8.109}$$

$$\approx -\frac{I_C}{w} \frac{\partial w}{\partial U_{CB}} \Delta U_{CB} \tag{8.110}$$

$$= g_m \frac{U_T}{w} \frac{\partial w}{\partial U_{CB}} \Delta U_{CB} \tag{8.111}$$

$$= g_m \eta_w \Delta U_{CB}. \tag{8.112}$$

Analog folgt für den zusätzlichen Basisstrom mit Gl.(8.19), welche hier nochmals wiedergeben ist

$$I_{BB} \approx -I_T \frac{1}{2} \left(\frac{w}{L_{nB}} \right)^2$$

$$\approx I_C \frac{1}{2} \left(\frac{w}{L_{nB}} \right)^2 = I_C \delta_{rek} \sim w.$$

wobei $\delta_{rek} = \frac{1}{2} \left(\frac{w}{L_{nB}} \right)^2$, der Anteil des Transferstromes ist, welcher in der Basis rekombiniert und mit dem Basistransportfaktor zusammenhängt, $\delta_0 \approx 1 - \delta_{rek}$. Da I_{BB} proportional zur Basisweite w ist gilt

$$\Delta I_B \approx \frac{\partial I_{BB}}{\partial w} \frac{\partial w}{\partial U_{CB}} \Delta U_{CB}$$

$$\approx \frac{I_{BB}}{w} \frac{\partial w}{\partial U_{CB}} \Delta U_{CB}$$

$$= -\frac{1}{2} \left(\frac{w}{L_{nB}} \right)^2 \frac{I_C}{w} \frac{\partial w}{\partial U_{CB}} \Delta U_{CB}$$

$$\Delta I_B = -\delta_{rek} g_m \eta_w \Delta U_{CB}, \tag{8.113}$$

wobei im letzten Schritt die Gln.(8.109) und (8.112) benutzt wurden.

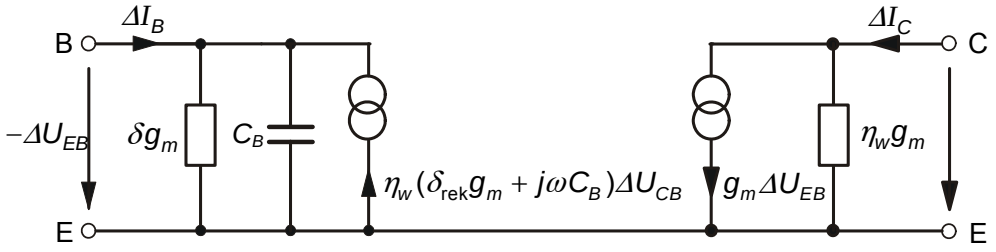


Abbildung 8.24: Kleinsignal-Ersatzschaltbild nach Gl. (8.118).

2.2. Beitrag aus der Modulation der gespeicherten Ladungsträger über die Weite Die Änderung der Minoritätsträgerladung in der Basis infolge der Basisweitenmodulation ist gegeben durch

$$\Delta Q_B = -\frac{1}{2} e\ A\ \Delta w\ n_{pB}\left(x_{BE}\right)$$

und mit Gl.(8.108) ergibt sich damit

$$\begin{aligned}\Delta Q_B &= -\frac{1}{2} e \Delta w A n_{pB}\left(x_{BE}\right) \\ &= \frac{1}{2} e A n_{pB}\left(x_{BE}\right) w \frac{\eta_w}{U_T} \Delta U_{CB} \\ \Delta Q_B &\approx \eta_w C_B \Delta U_{CB} \end{aligned} \tag{8.114}$$

Insgesamt fließt durch die zusätzliche Minoritätsträgerladung in der Basis infolge der Basisweitenmodulation ein Kleinsignalbasiswechselstrom nach den Gl.(8.113) und (8.114)

$$\Delta I_B = -\eta_w \left(\delta_{rek} g_m + j\omega C_B\right) \Delta U_{CB}. \tag{8.115}$$

3. Gesamtmodulation

Die Basisstromanteile infolge der Basisweitenmodulation gemäß den Gln.(8.113) und (8.114) ergeben zusammen mit dem bereits bestehenden Transistorgleichungen der Ladungsspeicherungseffekte 8.103 und 8.104

$$\Delta I_B = -\left(\delta + j\omega C_B\right) g_m\ \Delta U_{EB} - \eta_w \left(\delta_{rek} g_m + j\omega C_B\right) \Delta U_{CB}, \tag{8.116}$$

$$\Delta I_C = -g_m \Delta U_{EB} + g_m \eta_w \Delta U_{CB}. \tag{8.117}$$

Ersetzen wir noch die Kollektor-Basis-Spannung durch die Kollektor-Emitter- und Emitter-Basis-Spannungen, d.h. $\Delta U_{CB} = \Delta U_{CE} + \Delta U_{EB}$, so erhalten wir mit der Näherung $1 - \eta_w \approx 1$, die Leitwertdarstellung des Transistor-Vierpols in Emitterschaltung nach Bild 8.24 mit

$$\begin{pmatrix} \Delta I_B \\ \Delta I_C \end{pmatrix} = \mathbf{Y} \begin{pmatrix} \Delta U_{EB} \\ \Delta U_{CE} \end{pmatrix}, \tag{8.118}$$

$$\text{mit,}\ \ \mathbf{Y} = \begin{pmatrix} -\left(\delta + j\omega C_B\right) g_m & -\eta_w \left(\delta_{rek} g_m + j\omega C_B\right) \\ -g_m & g_m \eta_w \end{pmatrix} \tag{8.119}$$

Die ***Leitwertmatrix*** $\mathbf{Y}$ definiert das Ersatzschaltbild 8.24. Das Matrixelement $Y_{12} = \eta_w \left(\delta_{rek} g_m + j\omega C_B\right)$ beschreibt offensichtlich die Rückwirkung des Ausgangs auf den Eingang und wird durch eine spannungsgesteuerte Stromquelle zwischen Basis und Emitter dargestellt. Da der Strom dieser Stromquelle klein gegen den Kollektorstrom ΔI_C ist, kann diese Stromquelle direkt als Rückwirkungsleitwert gezeichnet werden, was zu der in Bild 8.25 dargestellten $\pi-$Ersatzschaltung des Transistors führt. Man sieht, dass die Basisweitenmodulation zu einem endlichen

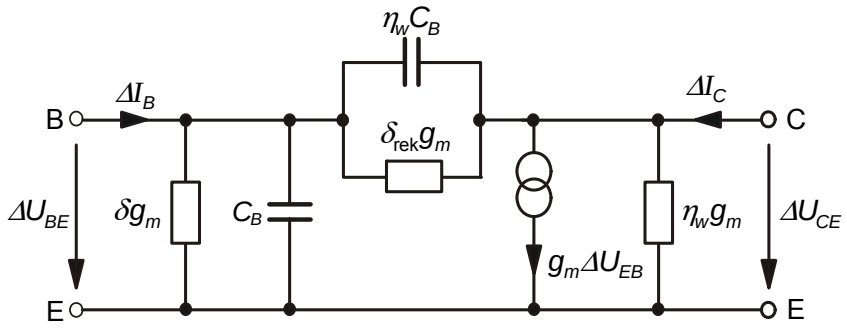


Abbildung 8.25: Zu Abb. 8.24 äquivalentes Kleinsignal-Ersatzschaltbild.

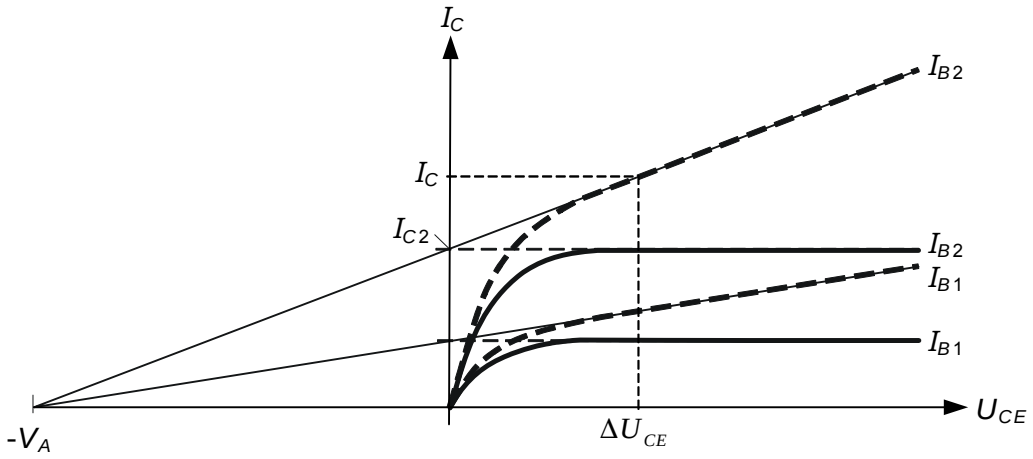


Abbildung 8.26: Einfluß der Basisweitenmodulation auf das Ausgangskennlinienfeld in Emitter-Schaltung.

Ausgangswiderstand des Transistors führt, die nur dann nennenswerten Einfluss hat, wenn die Ausgangsspannung ΔU_{CE} in die Größenordnung von $\Delta U_{EB}/\eta_w$ kommt.

Die Modifizierung des Ausgangskennlinienfelds in Emitterschaltung infolge der Basisweitenmodulation ist in Bild 8.26 dargestellt. Zur geometrischen Interpretation von V_A kommt man durch Aufzeichnen von I_C unter Berücksichtigung einer Modulation von U_{CE}. D.h

$$I_C = I_{Co} + \Delta I_c \tag{8.120}$$

$$\overset{\text{Gl (8.119)}}{=} I_{Co} - g_m \Delta U_{EB} + g_m \eta_w \Delta U_{CE} = I'_{Co} + g_m \eta_w \Delta U_{CE} \tag{8.121}$$

$$\overset{(8.90)\ \&(8.107)}{=} I'_{Co} + \frac{I'_{Co}}{U_T}\frac{U_T}{V_A}\Delta U_{CE} = I'_{Co} + \frac{I'_{Co}}{V_A}\Delta U_{CE} \tag{8.122}$$

Diese Gleichung kann in einem beliebigen Arbeitspunkt, z.B. Arbeitspunkt 2 ($I'_{Co} = I_{C2}$), auf eine Strahlensatzgleichung umgeformt werden,

$$I_C = I_{C2} + \frac{I_{C2}}{V_A}\Delta U_{CE} = I_{C2}(1 + \frac{\Delta U_{CE}}{V_A}) = I_{C2}(\frac{V_A + \Delta U_{CE}}{V_A}) \text{ oder} \tag{8.123}$$

$$\frac{I_{C2}}{I_C} = \frac{V_A}{V_A + \Delta U_{CE}} \tag{8.124}$$

was dann zur geometrischen Interpretation in Fig. 8.26 führt.

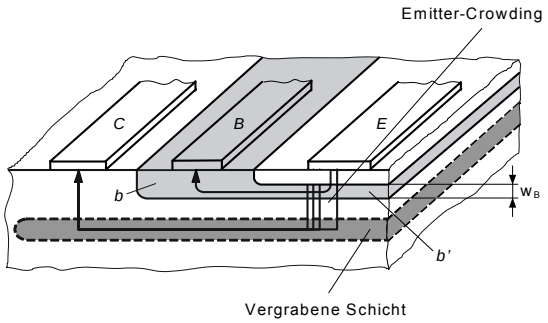


Abbildung 8.27: Schnitt durch einen Planartransistor mit schematischer Angabe der Strompfade. Die "vergrabene Schicht" hat hohe Leitfähigkeit und erleichtert den Stromfluß zum Kollektoranschluß.

Mechanismus	Ersatzschaltbildgröße
aktiver Steuermechanismus, Steilheit	$g_m u_{EB}$
Rekombination in der Basis und Injektion in den Emitter	$g_{b'e} = \delta g_m$
Ladungsträgerspeicherung in der Basis	$C_B = \frac{1}{2}\frac{w^2}{D_{nB}} g_m$
Ausgangsleitwert als Folge der Basisweitenmodulation	$g_{ce} = \eta_w g_m$
Rückwirkung als Folge der Basisweitenmodulation	$\eta_w \left(\delta_{rek} g_m + j\omega C_B\right)$
Transversaler Spannungsabfall in der Basis	$r_{bb'}$

Tabelle 8.4: Physikalische Effekte und Ersatzschaltbildgrößen für Injektionstransistoren.

8.8.3 Bahnwiderstände

Die Bahngebiete des Transistors können durch die Bahnwiderstände $r_{ee'}$, $r_{bb'}$, und $r_{cc'}$ beschrieben werden. Die Kollektor und Emitter-Bahnwiderstände können durch eine große Transistorquerschnittsfläche gering gehalten werden. Dem ist nicht so für den Basisbahnwiderstand. Um Rekombination in der Basis zu verhindern und um hohe Grenzfrequenzen zu erreichen muss die Basisweite w klein gehalten werden, siehe Bild 8.27.

Man sieht, dass der Basisstrom einen langen Weg bis zum Basisanschluss zurückzulegen hat. Der Strom bewirkt einen transversalen Spannungsabfall, der trotz seines geringen Wertes einen großen Einfluss hat, da er sich auf die an der Emitter-Basis-Diode liegende Vorwärtsspannung auswirkt. Dadurch wird die dem Basisanschluss näher liegende Transistorzone stärker in Flussrichtung gepolt sein und einen höheren Strom führen (***emitter-crowding***) als die weiter entfernt liegende. Dieser Effekt wirkt sich besonders bei starker Flusspolung aus und bewirkt eine Reduzierung der effektiven Transistorfläche. Dieser Effekt kann nur durch eine zwei- oder dreidimensionale Transistormodellierung richtig beschrieben werden. In einer Ersatzschaltung kann dieser Effekt durch einen vom Betriebszustand des Transistor abhängigen Basisbahnwiderstand modelliert werden.

Durch besondere Formgebung der Transistorgeometrie (z.B. Interdigitalstruktur), Bild 8.28, kann der Basisbahnwiderstand bei sonst gleichbleibenden Transistordaten klein gemacht werden. Solche Strukturen kommen deshalb bei Leistungs- und Hochfrequenztransistoren zum Einsatz.

8.8.4 Vollständiges Ersatzschaltbild nach Giacoletto

Wir ergänzen die Ersatzschaltung 8.25 noch um die Emitter-Basis-Sperrschichtkapazität $C_{sb'c}$, sowie dem unvermeidlichen Basisbahnwiderstand $r_{bb'}$. Die Kollektor-Basis- Sperrschichtkapazität wird aufgeteilt in eine äußere C_{sbc}, und innere $C_{sb'c}$, um der Stromverteilung in der Basis Rechnung zu tragen. Damit erhalten wir das vollständige Ersatzschaltbild für Emitterschaltung nach Giacoletto, Fig. 8.29. Tabelle 8.4 fasst nochmals die Zuordnung zwischen physikalischem Effekt und Ersatzschaltbildgröße in Reihenfolge der Bedeutung zusammen.

Natürlich kann man das obige Ersatzschaltbild für Emitterschaltung auch in jenes für die

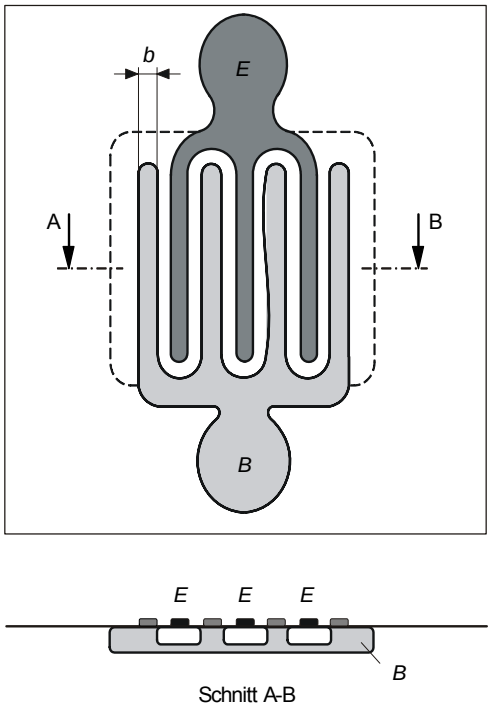


Abbildung 8.28: Schema eines Interdigitaltransistors.

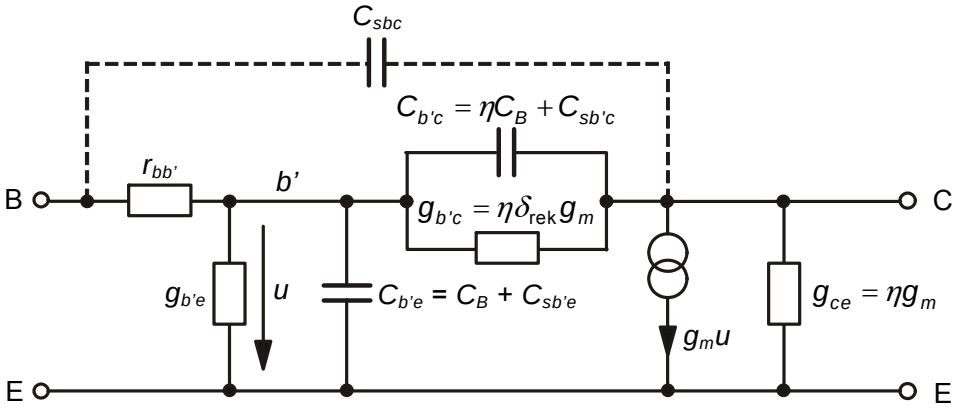


Abbildung 8.29: Kleinsignal-Ersatzschaltbild nach Giacoletto; Hybrid Ersatzschaltbild für Emitterschaltung [3].

Basis- und Kollektorschaltung transfomieren, was uns hier aber aus Zeitgründern erspart werden soll, siehe dazu [3].

8.8.5 Grenzfrequenzen

Je nach Betriebsart des Transistors interessieren folgende Grenzfrequenzen:

β-Grenzfrequenz f_β:	$\lvert\beta(f_\beta)\rvert = \beta_0/\sqrt{2}$	Stromverstärkung in Emitterschaltung
α-Grenzfrequenz f_α:	$\lvert\alpha(f_\alpha)\rvert = \alpha_0/\sqrt{2}$	Stromverstärkung in Basisschaltung
Transitfrequenz f_T:	$\lvert\beta(f_T)\rvert = 1$	
maximale Schwingfrequenz $f_{\max}$	Leistungsverstärkung = 1 bei Leistungsanpassung an Ein- u. Ausgang	

(8.125)

Für die Kleinsignalstromverstärkung in Emitterschaltung erhalten wir aus dem Ersatzschaltbild nach Giacoletto, bzw. dem einfacheren Bild von 8.22 unter Vernachlässigung von Early-Effekt, Diffusions- u. Streukapazitäten, sowie dem Basisbahnwiderstand im günstigsten Fall

$$\begin{aligned}\beta &= \frac{\Delta I_C}{\Delta I_B} = \frac{-g_m \Delta U_{EB}}{(-\delta g_m + j\omega C_B)\,\Delta U_{EB}} \\ &= \frac{\beta_0}{(1 - j\omega/(2\pi f_\beta))}.\end{aligned} \tag{8.126}$$

mit der Kleinsignal-Stromverstärkung in Emitterschaltung bei niedrigen Frequenzen und der β-Grenzfrequenz

$$\beta_0 = \frac{1}{\delta}, \tag{8.127}$$

$$f_\beta = \frac{\delta g_m}{2\pi C_B}. \tag{8.128}$$

Für die Stromverstärkung in Basisschaltung erhalten wir aus (8.88)

$$\alpha = -\frac{\Delta I_C}{\Delta I_E} = \frac{\alpha_0}{(1 - j\omega/(2\pi f_\alpha))} \tag{8.129}$$

mit der Kleinsignal-Stromverstärkung in Basisschaltung bei niedrigen Frequenzen und der α-Grenzfrequenz

$$\alpha_0 = \frac{\beta_0}{1+\beta_0} = \frac{1}{1+\delta} \approx 1, \tag{8.130}$$

$$f_\alpha = f_\beta(1+\beta_0) \approx \frac{g_m}{2\pi C_B} = f_T. \tag{8.131}$$

Der Zusammenhang zwischen Grenzfrequenzen und Verstärkungen in Basis und Emitter-Schaltung ist in Bild 8.30 dargestellt.

Wie das Ersatzschaltbild 8.29 zeigt, wirkt der Basisbahnwiderstand zusammen mit Basis-Emitterkapazität wie ein Tiefpass für die steuernde Spannung u_{BE} wobei nur der Teil an der inneren Basis $u_{B'E}$ wirklich steuert. Die Grenzfrequenz dieses Tiefpasses ist

$$f_r = \frac{1}{2\pi r_{bb'} C_{b'e}}, \tag{8.132a}$$

was die Grenzfrequenzen des Transistors weiter reduziert.

Die physikalische Bedeutung der Transitfrequenz beziehungsweise der Transitzeit $\tau_T = \frac{1}{2\pi f_T} = \frac{w^2}{2D_{nB}}$ ist, dass sie der Zeit entspricht, welche ein Minoritätsträger benötigt um aus der Basis zu

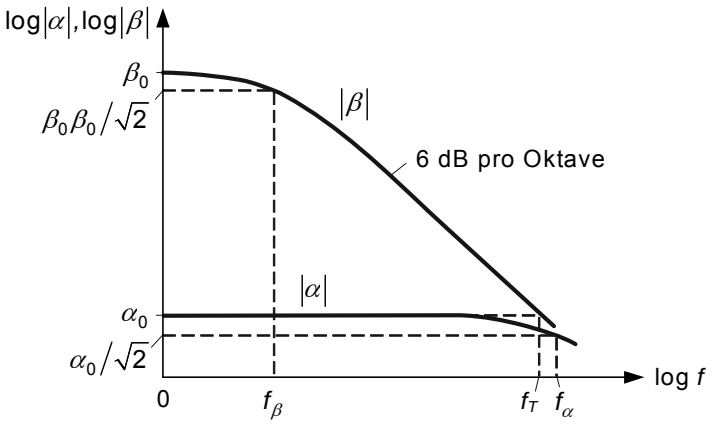


Abbildung 8.30: Frequenzabhängigkeit der Beiträge der Kleinsignal-Stromverstärkungsfaktoren.

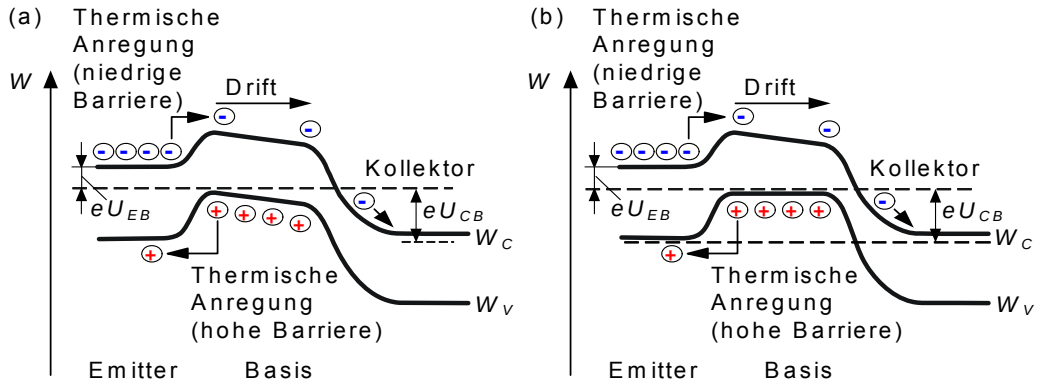


Abbildung 8.31: Bandverläufe in einem Drift-Transistor: (a) mit inhomogen dotierter Basis; (b) mit variablem Bandabstand in der Basis (bereits Heterobipolartransistor).

diffundieren. Diffusion ist der fundamentale Transportprozess im Bipolartransistor. Die Transitfrequenz von Bipolartransistoren kann deshalb erhöht werden, indem die Basis inhomogen dotiert wird, was im Gleichgewicht schon ein Feld erzeugt, womit der Diffusionstransistor mit dem eingebauten elektrischen Feld zum Drifttransistor wird und wesentlich höhere Grenzfrequenzen ermöglicht. Ein ähnlicher Effekt kann durch eine Basiskomposition mit örtlich variablem Bandabstand erzielt werden. Die entsprechenden Bandverläufe für bei Fälle sind in Bild 8.31 dargestellt.

8.8.6 Schaltverhalten

Das dynamische Verhalten des Transistors ist durch die in der Basis gespeicherte Ladung bestimmt. Die Minoritätsträgerverteilung in der Basis muss beim Einschalten des Transistors aufgebaut werden. Je nachdem welche Größe den Vorgang steuert, sind dafür unterschiedliche Zeiten erforderlich. In Basisschaltung wird der hohe Emitterstrom am Eingang angesteuert, was die Minortitätsträgerverteilung innerhalb der Transitzeit τ_T aufbauen kann, anders wäre eine Grenzfrequenz für die Stromverstärkung in Basisschaltung, die der Transitfrequenz entspricht nicht möglich. In Emitterschaltung, benötigt der kleine Basisstrom um den Faktor $(1 + \beta_0 \approx \beta_0)$ länger.

8.9 Das Gummel-Poon-Modell und SPICE

Für Computersimulationen von komplexen Schaltungen mit Tausenden von Transistoren, müssen Transistormodelle zur Verfügung gestellt werden, welche alle Betriebszustände des Transistors quantitativ richtig wiedergeben. Dazu wurde Ende der 60iger Jahre von Gummel (sprich Gummel, dt. Abstammung) das nach ihm und Poon benannte Transistor-Model entwickelt. Dieses Model liegt allen heute benutzten Schaltungssimulationsprogrammen, wie etwa SPICE (Simulation Programm with Integrated Circuit Emphasis) zugrunde. (Man kann sich Evaluationshalber, eine gebührenfrei Version für den PC (PSPICE) unter folgender Adresse herunterladen http://www.orcad.com/download.orcaddemo.aspx).

Das Gummel-Poon Modell beschreibt sowohl den Hochinjektionsfall und die Effekte infolge der Basisweitenmodulation als auch die Arbeitspunktabhängigkeit der Stromverstärkungen und der Transitzeit richtig.

In Anwendung von (5.19) auf den eindimensionalen Fall (ein weiterer Vorteil der vertikalen Schichtenanordnung beim Bipolartransistor!) erhält man

$$J_n = n\mu_n \frac{dW_{Fn}}{dx} \tag{8.133}$$

Durch Integration über die Basis $x_{BE} \leq x \leq x_{BC}$ (das ist die Basisweite ohne die Raumladungszonen! Da die Breiten der Raumladungszonen von U_{EB}, U_{CB} abhängen, ist $w = x_{BC} - x_{BE}$ spannungsabhängig. $J_n(x)$ ist in der Basis ortsunabhängig, da Rekombinationen in der Basis bei einem guten Transistor heute keine Rolle spielen. Ferner wird der Umstand verwendet, dass in der Basis annähernd $J_p = 0$ und somit $dW_{Fp}/dx = 0$ gilt. Aus (8.133) folgt somit unter Verwendung von (5.21)

$$J_n = n\mu_n \frac{d}{dx}(W_{Fn} - W_{Fp}) = n\mu_n kT \frac{d}{dx} \ln\left(\frac{np}{n_i^2}\right) = \frac{eD_{nB}n_i^2}{p} \frac{d}{dx}\left(\frac{np}{n_i^2}\right). \tag{8.134}$$

Gl. (8.134) wird unter Berücksichtigung von $J_n = \text{const}$ integriert. Man erhält

$$J_n \int_{x_{BE}}^{x_{BC}} \frac{p(x)dx}{eD_{nB}n_i^2} = \int \mathrm{d}\left(\frac{np}{n_i^2}\right) = \left[\exp\left(-\frac{U_{CB}}{U_T}\right) - \exp\left(-\frac{U_{EB}}{U_T}\right)\right]. \tag{8.135}$$

Damit ergibt sich direkt eine Beziehung für den Transferstrom $I_T = AJ_n$

$$I_T = \frac{I_S}{q_b}\left[\left(e^{-U_{EB}/U_T} - 1\right) - \left(e^{-U_{CB}/U_T} - 1\right)\right]$$

mit dem Sättigungstrom

$$I_S = \frac{AeD_{nB}n_i^2}{n_{AB}w_0}$$

und der normierten Majoritätsträgerladung in der Basis

$$q_b = \frac{1}{n_{AB}w_0} \int_{x_{BE}}^{x_{BC}} p(x)dx. \tag{8.136}$$

Wie im Ebers-Moll-Model ist der Transferstrom I_T die Summe von Vorwärtsstrom I_v und Rückwärtsstrom I_r geteilt durch eine vom Arbeitspunkt abhängige normierte Gummel-Ladung q_b.

Bild 8.32 zeigt das Ersatzschaltbild des Transistors nach dem Gummel-Poon-Modell, welches in der Lage ist das statische und dynamische Verhalten des Transistors über den ganzen Frequenzbereich, über welchen der Transistor funktionsfähig ist, d.h. für Vorgänge langsamer als die Transitzeit richtig beschreibt.

Der Transferstrom in diesem Modell lautet:

$$I_T = \frac{I_{CE} - I_{EC}}{q_b} \tag{8.137}$$

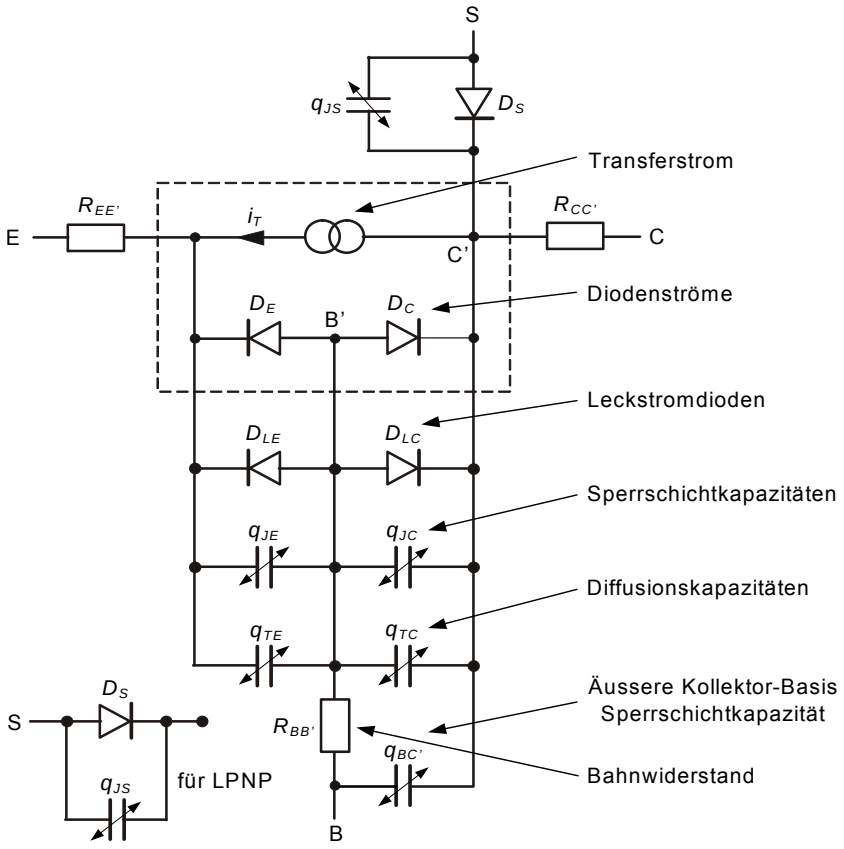


Abbildung 8.32: Ersatzschaltung des (vertikalen) npn-Bipolartransistors in SPICE. Das Modell des (vertikalen) pnp-Transitors resultiert durch Umpolen der Dioden. Sollen laterale pnp-Transistoren simuliert werden (Modelltyp LPNP), so ist der Substratanschluß nicht am inneren Kollektorknoten sondern am inneren Basisknoten [20].

mit

$$I_{CE} = I_S \left(\exp \left[-\frac{U_{E'B'}}{U_T n_f} \right] - 1 \right) \tag{8.138}$$

$$I_{EC} = I_S \left(\exp \left[-\frac{U_{C'B'}}{U_T n_r} \right] - 1 \right) \tag{8.139}$$

und der Gummel-Ladung

$$q_b = \frac{q_1}{2} + \sqrt{\left(\frac{q_1}{2}\right)^2 + \left(\frac{I_{CE}}{I_{KF}} + \frac{I_{EC}}{I_{KR}}\right)}. \tag{8.140}$$

$$q_1 = \left(1 + \frac{U_{E'B'}}{V_{AR}} + \frac{U_{C'B'}}{V_{AF}}\right)^{-1} \tag{8.141}$$

Meist gilt $n_f = n_r = 1$. Die Rückinjektion von Elektronen in Emitter und Kollektor werden durch die Diodenströme I_{DE} und I_{DC} beschrieben

$$I_{DE} = \frac{I_{CE}}{B_N}, \tag{8.142}$$

$$I_{DC} = \frac{I_{EC}}{B_I}. \tag{8.143}$$

Die nichtidealen Eigenschaften der Basis-Emitter-Diode bzw. Basis-Kollektor-Diode, (z.B. Generation und Rekombination in den entsprechenden RLZ's sowie Hochinjektion wird durch die Leckstromdioden I_{LE} und I_{LC} beschrieben

$$I_{LE} = I_{SE} \left(\exp \left[-\frac{U_{E'B'}}{U_T n_E} \right] - 1 \right) \tag{8.144}$$

$$I_{LC} = I_{SC} \left(\exp \left[-\frac{U_{C'B'}}{U_T n_C} \right] - 1 \right) \tag{8.145}$$

Der Basisbahnwiderstand wird durch folgende Gleichungen modelliert

$$R_{BB'} = r_{Bm} + 3\left(r_B - r_{BM}\right) \frac{(\tan(z) - z)}{z \tan^2(z)}, \tag{8.146}$$

mit

$$z = \frac{\pi^2}{24} \frac{\left[\sqrt{(1 + \frac{144}{\pi^2} \frac{I'_b}{I_{RB}})} - 1 \right]}{\sqrt{\frac{I'_b}{I_{RB}}}}. \tag{8.147}$$

wobei der den Basisbahnwiderstand steuernde Strom geben ist durch

$$I'_b = I_{DE} + I_{DC} + I_{LE} + I_{LC}. \tag{8.148}$$

Die Speicherladung in den Diffusionszonen können gemäß der Ladungssteuerungstheorie durch die entsprechenden Diffusionsströme ausgedrückt werden

$$q_{TE} = \tau_f \frac{I_{CE}}{q_b}, \tag{8.149}$$

$$q_{TC} = \tau_r \frac{I_{EC}}{q_b}. \tag{8.150}$$

wobei die Vorwärtstransferzeit τ_f selbst vom Arbeitspunkt abhängt

$$\tau_f = T_F \left[1 + X_{TF} \left(\frac{I_{CE}}{I_{CE} + I_{TF}} \right)^2 \exp \left(\frac{-U_{C'B'}}{1.44 V_{TF}} \right) \right].$$

Parameter	Wert		Parameter	Wert
I_S	7.049 fA		$R_{CC'}$	1.464 Ω
X_{TI}	3		C_{jc0}	5.38 pF
U_T	25.8 mV		m_c	0.329
V_{AF}	116.3 V		V_{jc}	0.6218
B_N	375.5		F_c	0.5
I_{SE}	7.049 fA		C_{je0}	11.5 pF
n_E	1.281		m_e	0.2717
I_{kf}	4.589 A		V_{je}	0.5
n_k	0.5		τ_r	10 ns
X_{tb}	1.5		T_f	451 ps
B_I	2.611		I_{TF}	6.194 A
I_{SC}	121.7 pA		X_{TF}	17.43
n_C	1.865		V_{TF}	10 V
I_{kr}	5.313		W_G	1.11 eV

Tabelle 8.5: Parameter des Gummel-Poon-Modells für den Transistor BC107.

Die Sperrschichtkapazitäten sind beschrieben durch Funktionen welche die singuläre Stelle bei Erreichen der Diffusionsspannung stetig durchlaufen

$$C_{je} = \begin{cases} C_{je0}\left(1+\frac{U_{E'B'}}{V_{je}}\right)^{-m_e}; U_{E'B'} > -F_c V_{je} \\ \frac{C_{je0}}{(1-F_c)^{m_e}}\left[1-\frac{m_c(U_{E'B'}+F_c V_{je})}{(1-F_c)V_{je}}\right]; U_{E'B'} < -F_c V_{je} \end{cases}, \tag{8.151}$$

$$C_{jc} = \begin{cases} C_{jc0}\left(1+\frac{U_{C'B'}}{V_{jC}}\right)^{-m_c}; U_{C'B'} > -F_c V_{jc} \\ \frac{C_{jc0}}{(1-F_c)^{m_c}}\left[1-\frac{m_c(U_{C'B'}+F_c V_{jc})}{(1-F_c)V_{jc}}\right]; U_{C'B'} < -F_c V_{jc} \end{cases}. \tag{8.152}$$

Tabelle 8.5 gibt die Modellparameter für den Transistor BC107 wieder. Parameter, welche nicht angegeben sind erhalten im Programm vordefinierte (triviale, z.B. $R_{EE'} = 0$) Werte.

Die Parameter X_{TI} und die Bandlückenenergie W_G werden zur Temperaturabhängigkeit der Sättigungsströme der Leckdioden benützt. Für eine eingehende Diskussion des Modells und der Parameter siehe [20].

8.10 Hetero-Bipolartransitoren (HBT)

Wesentliches Merkmal des HBT ist der Hetero-Emitter (d.h ein Bipolartransistor mit einem Emitter aus einem Material und einer Basis und eines Kollektors aus einem andern Material). Bild 8.33 zeigt zwei mögliche Materialkombinationen. Die Kombination InP/InGaAs ist aber aus mehreren Gründen günstiger:

- Sie besitzt die größere Differenz der Bandabstände (die Barriere für Löcher im Valenzband aus der Basis in den Emitter ist höher);
- die Beweglichkeit der Elektronen im InGaAs, welches die Basis bildet ist höher als im GaAs (und ca. 9-mal so groß wie die Beweglichkeit der Elektronen im Si; die Transitfrequenz ist nach Gl.(8.106) proportional zur Beweglichkeit.

Für einen HBT findet man deshalb oft den in der Abb. 8.33 dargestellten Bandverlauf (Gleichgewichts-Fall).

- Typische Basisbreiten betragen $50 \ldots 100$ nm mit Dotierungen im Bereich $10^{18} \ldots 10^{20}\,\mathrm{cm}^{-3}$ (bei GaAs wird Kohlenstoff als Akzeptor verwendet, die Ionisierungsenergie beträgt $25,8\,\mathrm{meV} \approx kT_0$).

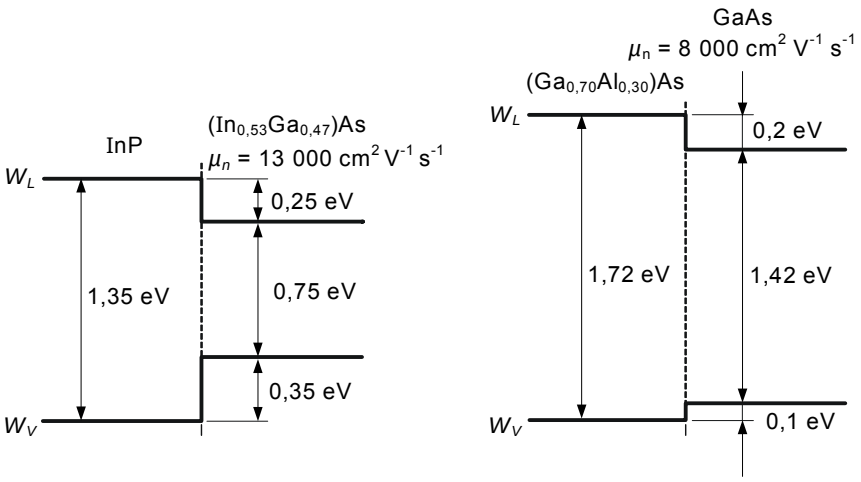


Abbildung 8.33: Beispiele einer Emitter-Basis Materialkombinationen zum Aufbau eines Np^+-Elektronenemitters (zum Vergleich $\mu_n = 1500\ \mathrm{cm^2V^{-1}s}$ in Si).

- Die Breite der n^--Kollektorschicht ist typisch 500 nm, die Dotierung einige $10^{16}\,\mathrm{cm}^{-3}$ (als Donator bei GaAs wird Si verwendet, die Ionisierungsenergie beträgt 58 meV). Die Transitzeit durch die Basis hat die Größenordnung von zehntel Picosekunden, wodurch Grenzfrequenzen um 1 THz möglich werden. Die Transitzeit kann durch ein in der Basis eingebautes Driftfeld weiter verkürzt werden.

Der HBT hat aber gegenüber einem regulären Homo-Bipolartransistor einen weitern Vorteil. Er hat einen kleinen Basiswiderstand und ist gegenüber dem Early-Effekt unempfindlicher. Dies ergibt sich aus dem speziellen Design, wie nachfolgend ausgeführt.

Jeder Bipolartransistor sollte ja mit einen möglichst hohen Emitterwirkungsgrad γ_E gebaut sein. Nach Gl.(8.82) gilt

$$\gamma_E = \frac{1}{1 + \frac{\mu_{pE} p_{noE} L_{nB}}{\mu_{nB} n_{poB} l_{nE}} \tanh\left(\frac{w}{L_n}\right)}$$

Im regulären (Homo-)Bipolartransistor, ist das Verhältnis der Minoritätsträgerdichten in Emitter und Basis umgekehrt proportional zu den Dotierungen

$$\frac{p_{noE}}{n_{poB}} = \frac{n_{i,E}^2}{n_{i,B}^2} \frac{n_{AB}}{n_{DE}}, \tag{8.153}$$

da $n_{i,E} = n_{i,B} = n_i$ gilt. Um einen hohen Emitterwirkungsgrad $\gamma_E \longrightarrow 1$ zu erreichen, muss die Basis gegenüber dem Emitter schwächer dotiert sein. Das aber führt dazu, dass der Basiswiderstand unötig groß bleibt und bewirkt überdies einen großen Early-Effekt.

Die Verwendung von verschiedenen Materialen für Basis und Emitter behebt diese Schwierigkeit. Wird für die Basis ein Material mit niedrigerer Bandlücke als beim Emitter gewählt, so wird $n_{i,B}^2 \gg n_{i,E}^2$ was bei gleichbleibendem Emitterwirkungsgrad eine wesentlich höhere Basisdotierung erlaubt. Wegen der hohen Basisdotierung ist die Basisweite w praktisch unabhängig von den Spannungen U_{EB}, U_{CB} und somit wird auch der Early-Effekt reduziert.

Im Vergleich zum Silizium-BPT (bei dem Grenzfrequenzen im Bereich von $20 \ldots 40$ GHz erreicht wurden) ist der HBT schneller, hat aber den Nachteil, dass (zur Zeit?) seine Integrationsmöglichkeiten begrenzt sind.

HBTs können auch mit den Halbleitern Si,Ge hergestellt werden. Aus Si ($a = 5{,}43\,\mathring{A}$, $W_G = 1{,}12\,\mathrm{eV}$) und Ge ($a = 5{,}65\,\mathring{A}$, $W_G = 0{,}67\,\mathrm{eV}$) kann ein Mischkristall $Si_{1-x}Ge_x$ gebildet werden,

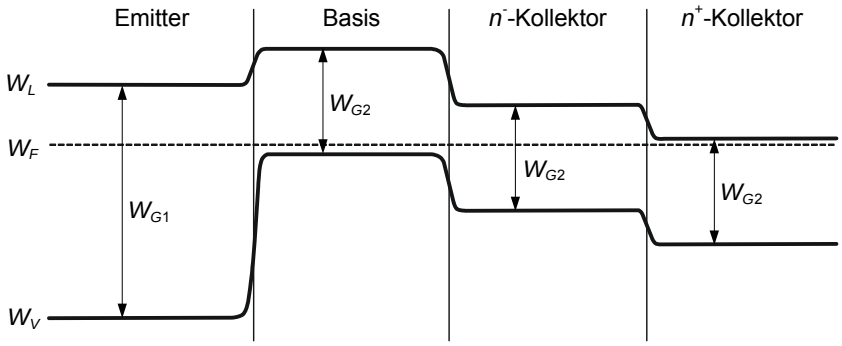


Abbildung 8.34: Bandverlauf des HBT im Gleichgewicht und $W_{G1} > W_{G2}$.

dessen Bandabstand kleiner ist als der von Si (für $x = 0{,}2$ erniedrigt sich der Bandabstand um 200 meV). $Si_{1-x}Ge_x$ kann als Material für die Basis im HBT verwendet werden (Grenzfrequenzen um 75 GHz).

Kapitel 9

Halbleiter-Grenzschichten

Bisher haben wir nur Halbleitermaterialien betrachtet bei unbeschränkt gültiger Gitterperiodizität. Daraus resultierten elektronische Zustände, die anhand von Banddiagrammen zu verstehen waren. An Oberflächen oder Grenzflächen zwischen zwei chemischen Materialien ist diese Voraussetzung nicht mehr erfüllt. Mehr noch, es treten jetzt Grenzflächenladungen auf, welche die Banddiagramme und Potentiale beeinflussen. Um weiterhin mit Potentialen rechnen zu können, benötigen wir nun eine absolute Referenz für die Energie und Potentiale. Wir definieren folgende Größen:

- ***Austrittsarbeit*** W_ϕ. Sie entspricht der Energie, welche mindestens aufgewendet werden muß, damit ein Elektron den Festkörper verlassen kann. Die Energie eines Elektrons mit der kinetischen Energie Null im Vakuum auf dem Potential Null habe den Wert Null, siehe Bild. 9.2(a). Bei einem Metall entspricht die Austrittsarbeit daher der Potentialdifferenz zwischen dem Fermi-Niveau und dem Vakuum-Niveau. Diese Definition wird auch auf den Halbleiter übertragen.

- ***Elektronen-Affinität*** W_χ**.** Sie entspricht der minimalen Energie, welche benötigt wird um eine Elektron aus dem Leitungsband auszulösen. Diese ist kleiner als die Austrittsarbeit, da in einem Halbleiter schon einige Elektronen auf dem Energieniveau des Leitungsband sind, so daß die minimale Energie ein Elektron aus dem Halbleiter auszulösen nicht der Austrittsarbeit entspricht, siehe Bild 9.2(b).

In Fig. 9.2 sind die Energien für ein Metall und einen Halbleiter, welche sich auf dem elektrostatischen Potential φ befinden eingezeichnet. Man beachte, daß die Größen W_χ, W_ϕ und W_G als positive Größen definiert sind. Sämtliche Größen wie W_L, W_V, oder W_F können nun durch W_χ und W_ϕ ausgerückt werden. Die entsprechenden Beziehungen entnimmt man Bild 9.2(b):

$$\begin{array}{rcl} W_L &=& -e\varphi - W_\chi, \\ W_V &=& -e\varphi - W_\chi - W_G, \\ W_F &=& -e\varphi - W_\phi, \\ W_G &=& W_L - W_V. \end{array} \tag{9.1}$$

Zusätzlich können unmittelbar an der Halbleiter-Oberfläche auch Ladungen auftreten, z.B. infolge von Anlagerung von Fremdionen. Im einfachsten Fall, kann man diese Oberflächenzustände als homogene Flächenladung auffassen. Diese Flächenladung erzeugt Influenzladungen im Halbleiterinneren und damit eine Potentialstörung. Diese Influenzladung bildet eine Randschicht welche unmittelbar unter der Oberflächenladung zu liegen kommt und die Oberflächenladung abschirmt, siehe Fig. 9.3.

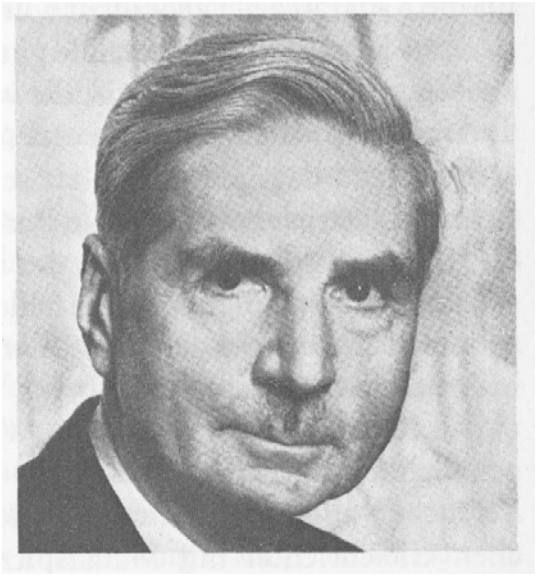


Abbildung 9.1: Walter Schottky, wegbereitender Theoretiker der Halbleitertechnik. In den Laboratorien der Siemens-Werke fand er (mit E. Spenke) 1939 die entscheidende Erklärung der elektronischen Vorgänge nahe der Halbleiter-Grenzschicht. Viele Begriffe der Mikroelektronik tragen noch heute seinen Namen; besonders wichtig ist die Schottky-Diode [19] .

9.1 Die Metall-Isolator-Halbleiter-Struktur

9.1.1 Klassifikation der Randschichten

Zunächst wollen wir uns ein Bild über die möglichen Randschichten machen, welche bei einer Metall-Isolator-Halbleiter *(Metall-Insulator-Semiconductor =MIS)* Struktur auftreten können.

Eine ideale MIS-Struktur ist in Bild 9.4(a) gezeichnet. Sie besteht aus einem Halbleiter (z.B. n-HL mit Störstellenerschöpfung), einem raumladungsfreien Isolator und einem Metall. Ferner gilt für den idealen MIS, daß die Ferminiveaus auf beiden Seiten des Isolators gerade auf dem gleichen Niveau sind, d.h. dass die Austrittsarbeiten ϕ_{MI}, ϕ_{HI} der Elektronen aus dem Metall in den Isolator und aus dem Halbleiter in den Isolator (ϕ_{MI}, $\phi_{HI} > 0$) gerade die Bedingung

$$0 = \phi_{MI} - [\phi_{HI} + (W_\phi - W_\chi)] \tag{9.2}$$

erfüllen. Es entstehen dann keine lokalen Raumladungen und keine Kontaktspannungen und die

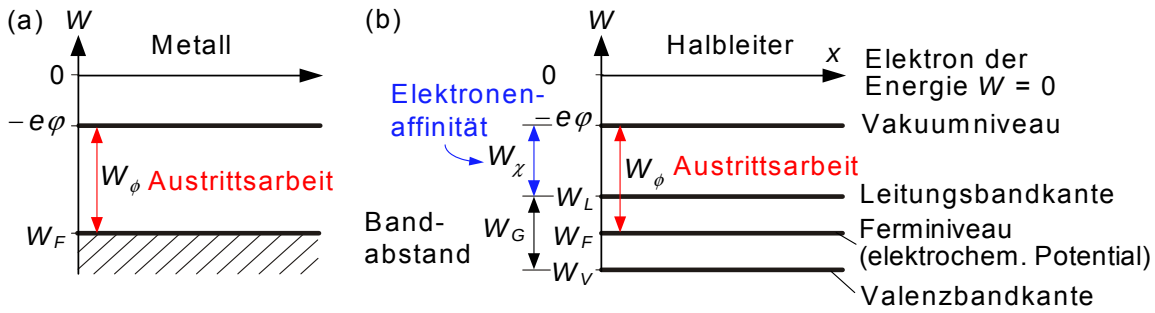


Abbildung 9.2: Energiediagramm für Elektronen in einem Metall (a); und einem Halbleiter (b), welche sich beide auf dem elektrostatischen Potential φ befinden.

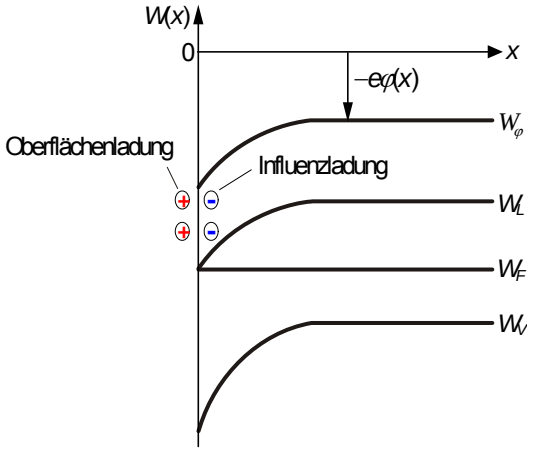


Abbildung 9.3: Potentialstörung infolge von Oberflächenladungen.

Bandkanten verlaufen horizontal (sogenannter ***Flachbandfall***).

Für reale Materialien ist (9.2) in der Regel nicht erfüllt; außerdem können im Isolator Raumladungen auftreten (etwa positive Na-Ionen in SiO_2), an den Übergängen MI und HI können Dipolschichten und Oberflächenladungen auftreten, wie oben bereits erwähnt. Es läßt sich aber immer durch Anlegen einer äußeren Spannung $U_{FB} \neq 0$ erreichen, daß die Bandkanten im Halbleiter horizontal verlaufen und somit keine Raumladungen im Halbleiter auftreten. U_{FB} wird als ***Flachbandspannung*** bezeichnet. Im gezeichneten Fall gilt somit $U_{FB} = 0$.

Wenn man an die ideale MIS Diode eine positive oder negative Spannung anlegt verbiegen sich die Bänder. Wir unterscheiden folgende Fälle

- Anlegen einer positiven Spannung $U > 0$ an die Metallplatte ($\varphi_H > 0$): Wegen der Isolationsschicht können keine Ladungen fließen und das Ferminiveau bleibt im HL und der Metallschicht flach. Allerdings werden Ladungen in der vom Potential vorgegebenen Richtung fließen und sich an der Randschicht anreichern. Die Bänder verformen sich entsprechend der Ladungsträgerkonzentrationen. Im gezeichneten Fall kommt es zu einer weiteren Anreicherung der Majoritäten und wir sprechen von einer ***Anreicherungsschicht.***

- Anlegen einer negativen Spannung $U < 0$ an die Metallplatte mit ($0 > \varphi_H \geq \varphi_{Hi}$): Elektronen werden vom Halbleiterrand abgestoßen und der Halbleiter verarmt an Majoritäten. Zurück bleiben die positiv geladenen Donatoren so dass $\rho(x) = n_D$. Es bildet sich eine Verarmungszone wie man das von der pn-Diode her kennt. Falls die angelegte Spannung gerade so groß wird, dass gilt $\varphi_H = \varphi_{Hi}$ so verhält sich der Halbleiter an der Oberfläche gerade eigenleitend. Aber selbst hier ist die Schicht immer noch positiv geladen mit $\rho(0) = n_D$. Wir sprechen von einer ***Verarmungsrandschicht.***

- Anlegen einer negativen Spannung $U < 0$ an die Metallplatte mit ($\varphi_{Hi} > \varphi_H \geq 2\varphi_{Hi}$): Das Eigenleitungniveau gelangt über das Ferminiveau und die Schicht verhält sich wie ein p-Halbleiter (Inversion). Solange $\varphi_{Hi} > \varphi_H \geq 2\varphi_{Hi}$ sprechen wir von ***schwacher Inversion.*** Die Löcherkonzentration ist immer noch kleiner als die ionisierten Donatoren, so dass in dieser Zone immer noch gilt $\rho(x) = n_D$.

- Anlegen einer negativen Spannung $U < 0$ an die Metallplatte mit ($2\varphi_{Hi} > \varphi_H$): Beim Einsetzen der starken Inversion wächst die Verarmungszone nicht mehr stark an. Die Inversion ist nun so stark, dass der Ladungsausgleich beim weitern Erhöhen der Spannung über Löcher an der Randschicht erfolgen kann. Wir sprechen von ***starker Inversion.***

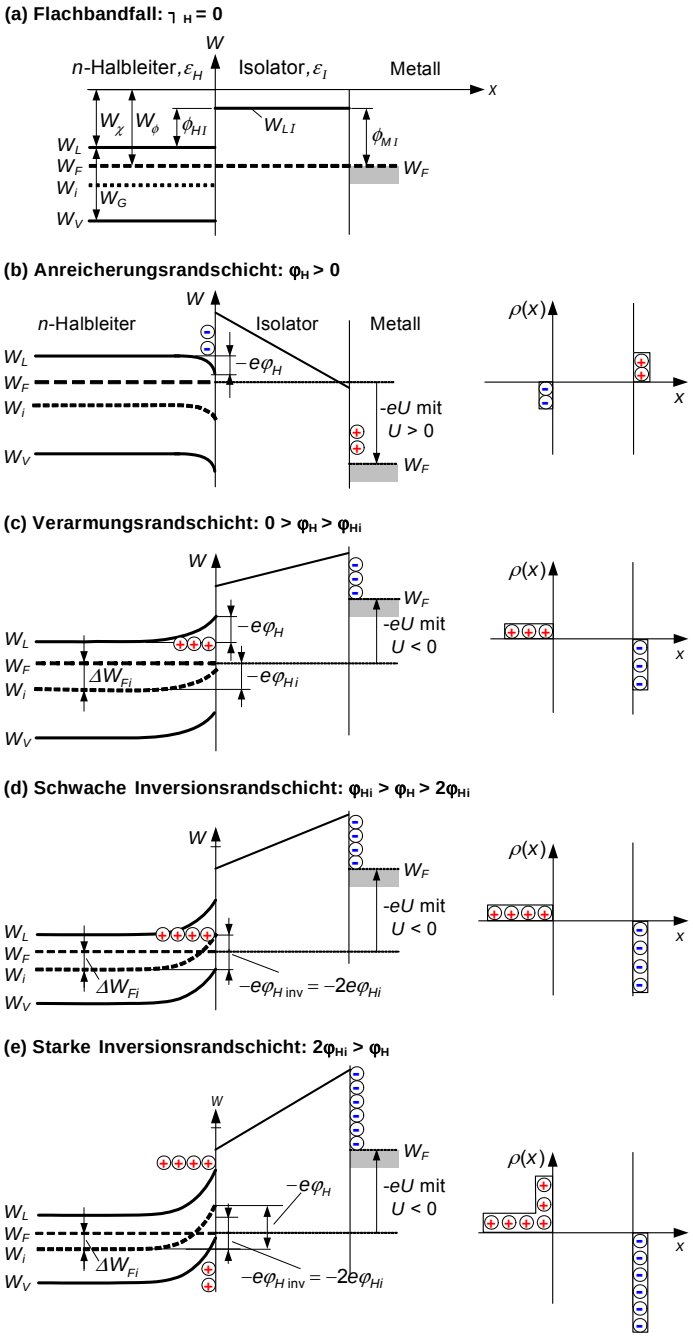


Abbildung 9.4: MIS (Metal-Insulator-Semiconductor)- Struktur. (a) Spezielle Struktur, welche ohne äußere Spannung den Flachbandfall realisiert (b) Anreicherungsschicht, (c) Verarmungsschicht, (d) schwache Inversion, (e) starke Inversion

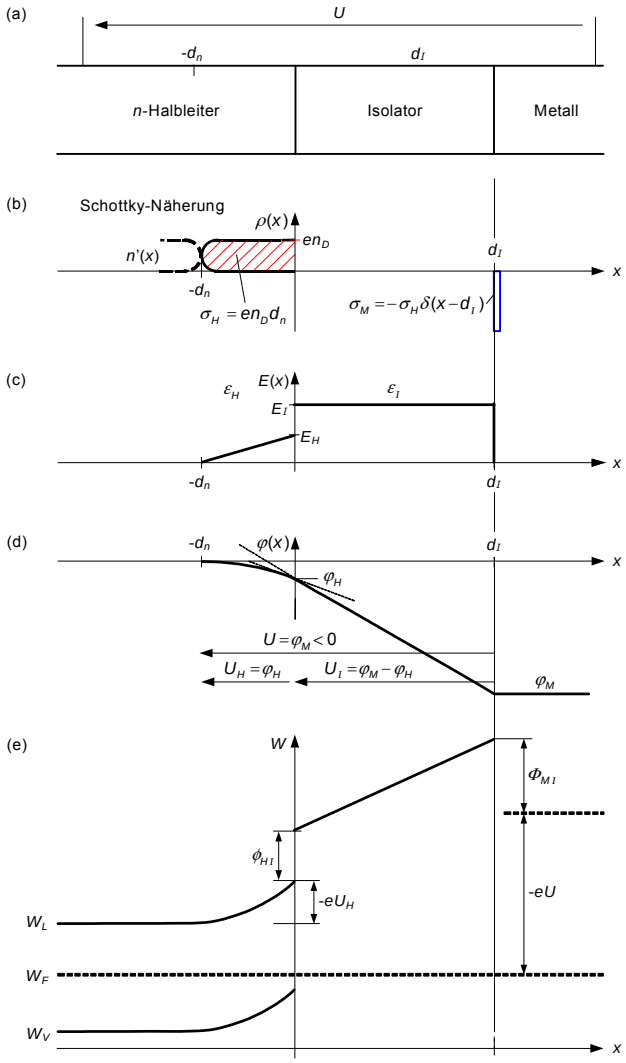


Abbildung 9.5: (a) MIS-Struktur mit angelegter Spannung $U < 0$; (b) Raumladungsverteilung; (c) Feldstärkeverlauf; (d) Potentialverlauf; (e) Energiediagramm.

9.1.2 Grundgleichungen

Im Folgenden sollen die wesentlichen Gleichungen zur Diskussion der verschiedenen Bandverbiegungen hergeleitet werden.

Bild 9.5(b)-(e) zeigt die Verhältnisse, wenn zusätzlich zur Flachbandspannung (hier: $U_{FB} = 0$) eine Spannung $U = \varphi_M < 0$ angelegt wird. Das negative Potential der Metalloberfläche gegenüber der Halbleiteroberfläche führt zu einer positiven Influenzladung an der Halbleiteroberfläche welche durch eine entsprechende negative Oberflächenladung auf der Metalloberfläche kompensiert wird, Bild 9.5(b).

Ziel ist es nun die Spannungen, die Breiten der Schichten und die Kapazitäten der MIS Struktur zu bestimmen. Zur Berechnung der Verhältnisse im Halbleiter benötigen wir insbesondere den Potentialverlauf $\varphi(x)$ und die Potentialstufen bzw. $U_H = \varphi_H$ und $U = \varphi_M$.

Berechnung der Halbleiterseite:

Wie so oft in der Halbleiterphysik geht es zunächst darum einen Ausdruck für die Raumladungsdichte zu finden. Daraus lässt sich dann via Poissongleichung das Potential ableiten und alle andern Ausdrücke.

i) Ein Ausdruck für die Raumladungsdichte: Der gezeichnete Fall entspricht dem Verarmungsfall, bei dem ein Randpotential $\varphi_H < 0$, $|\varphi_H/U_T| \gg 1$, eine von beweglichen Trägern verarmte Zone entstehen läßt.

- Diese Zone ist im Vergleich zur Debye-Länge sehr breit ($d_n \gg L_{Dn}$).
- In dieser Zone muss nur die Raumladungsdichte der ionisierten Donatoren berücksichtigt werden. In der Schottky-Näherung, d.h. völlige Verdrängung der beweglichen Ladungsträger aus der RLZ im Halbleiter ($n(x) = 0, \rho(x) = en_D$), ergäbe sich eine Ladungsdichte wie in Bild 9.5(b) dargestellt. Wir wollen das nun aber etwas genauer berechnen.

Für einen n-Halbleiter mit Störstellenerschöpfung, d.h. $n_n = n_D = n_i^2/p_n$ gilt unter Bezug auf (5.14) in Gebieten mit $\varphi(x) = 0$ (siehe auch Abb. 9.4(a))

$$\begin{aligned} n(x) &= n_n = n_D &&= N_L \exp\left(\tfrac{W_F - W_L}{kT}\right) = N_L \exp\left(\tfrac{-W_\phi + W_\chi}{kT}\right), \\ p(x) &= p_{n0} &&= N_V \exp\left(\tfrac{W_V - W_F}{kT}\right) = N_V \exp\left(\tfrac{-W_\chi - W_G + W_\phi}{kT}\right) = \tfrac{n_i^2}{n_D}. \end{aligned} \tag{9.3}$$

An Stellen $\varphi(x) \neq 0$ und falls $W_F = -W_\phi$ konstant gehalten erhält man unter Berücksichtigung von Gl. (9.1):

$$\begin{aligned} n(x) &= N_L \exp\left(\tfrac{W_F - W_L}{kT}\right) = N_L \exp\left(\tfrac{-W_\phi + W_\chi + e\varphi}{kT}\right) = n_D \exp\left(\tfrac{\varphi}{U_T}\right), \\ p(x) &= N_V \exp\left(\tfrac{W_V - W_F}{kT}\right) = N_V \exp\left(\tfrac{-W_\chi - W_G - e\varphi + W_\phi}{kT}\right) = p_{n0} \exp\left(-\tfrac{\varphi}{U_T}\right), \end{aligned} \tag{9.4}$$

und somit für die Raumladungsdichte

$$\begin{aligned} \rho(x) &= e[n_D + p(x) - n(x)] \\ &= en_D + ep_{n0} \exp\left(-\frac{\varphi}{U_T}\right) - en_D \exp\left(\frac{\varphi}{U_T}\right). \end{aligned} \tag{9.5}$$

ii) Zur Berechnung des Zusammenhangs $\varphi_H = f(\rho)$ oder später auch von $\varphi_H = f(\sigma_H)$, s. (9.5), ist die Poissongleichung

$$\frac{d^2\varphi}{dx^2} = -\frac{\rho}{\varepsilon_H} \tag{9.6}$$

zu lösen. Diese Differentialgleichung kann wie schon so oft ganz normal gelöst werden. Falls jedoch die Raumladungsdichte der beweglichen Träger, s. (4.34) als Funktion des Potentials φ (und nicht als Funktion des Ortes) gegeben sind, ist (9.6) eine nichtlineare Differentialgleichung. Mit Hilfe von

$$\frac{d^2\varphi}{dx^2} = -\frac{dE}{dx} = -\frac{dE}{d\varphi} \cdot \frac{d\varphi}{dx} = \frac{dE}{d\varphi} \cdot E = \frac{1}{2}\frac{dE^2}{d\varphi} \tag{9.7}$$

läßt sich (9.6) in diesem Fall integrieren und liefert

$$E(\varphi) = -\mathrm{sgn}(\varphi_H)\sqrt{-\frac{2}{\varepsilon_H}\int_0^{\varphi} \rho(\varphi')\, d\varphi'} = -\frac{d\varphi}{dx}, \qquad E_H = E(\varphi_H). \tag{9.8}$$

Für die Wahl des Vorzeichens in (9.8) wurde vorausgesetzt, daß $E(x)$ in $x \leq 0$ eine monotone Funktion von x ist.

Ein Ausdruck für $\varphi(x)$ ergibt sich dann aus $E(\varphi(x)) = -d\varphi(x)/dx$.

In der Folge werden wir dann oft das folgende Integral benötigen

$$\int_{d_1}^{d_2} dx = -\int_{\varphi|_{d_1}}^{\varphi|_{d_2}} \frac{d\varphi}{E(\varphi)}\ .$$

Definiert man dann $d_2 = \widetilde{x}$ wird man nach Auflösen automatisch auf einen Ausdruck $\widetilde{x} = f(\varphi(\widetilde{x}))$ stossen. Konkret heisst das, dass man dann folgendes Integral lösen muss

$$\int \frac{du}{\sqrt{e^u - 1}} = 2 \arctan \sqrt{e^u - 1}. \tag{9.9}$$

Im Halbleiterbereich wie wir ihn oben gezeichnet haben, wird uns die konstante Raumladung im Halbleiterbereich $-d_n \leq x \leq 0$ auf ein linear mit x zunehmendes elektrisches Feld, führen, welches uns auf einen quadratischen Verlauf von $\varphi(x)$ für $x < 0$ bringt. Der Spannungsabfall U_H über dem Halbleiter ist dann $U_H = \varphi_H = \varphi(0)$.

Berechnung des Isolatorbereichs:

i) Im Isolator haben wir einen raumladungsfreien Bereich $\rho(x) = 0$.

ii) Das elektrische Feld im Isolator ergibt sich aus

$$\mathrm{div}\vec{E} = \frac{\rho}{\varepsilon} = 0 \tag{9.10}$$

Da $\rho(x) = 0$, muss die elektrische Feldstärke im Isolator konstant sein, siehe Bild 9.5(c). Das konstante el. Feld über dem Isolator sei E_I. E_I ergibt sich aus den Stetigkeitsbedingungen bei $x = 0$. Wegen der Stetigkeit der Normalkomponenten der dielektischen Verschiebung $\vec{D}$ bei $x = 0$ und der unterschiedlichen Dielektrizitätskonstanten gilt $\varepsilon_H E_H = \varepsilon_I E_I$. Damit ist E_I bestimmt.

iii) Zur Berechung des Potentialabfalls φ über dem Isolator benutzen wir, dass im raumladungsfreien Isolator das Potential linear verläuft

$$\varphi(x) = -\int_{x=0}^{x} E_I(x')dx' = -E_I \; x. \tag{9.11}$$

Der Verlauf von $\varphi(x)$ vom Halbleiterbereich in den Isolatorbereich hat eine unstetige Tangente. Der Potentialabfall über dem Isolator ist $\varphi_I = -d_I E_I$. Der gesamte Spannungsabfall U teilt sich in einen Spannungsabfall über dem Halbleiter U_H und einen Spannungsabfall über dem Isolator $U_I = -d_I E_I$ auf.

Berechnung der Metallseite:

i) Ladungsdichte im Metall: Es bilden sich nur Oberflächenladungen, σ_M. Weil der gesamte MIS Übergang makroskopisch ungeladen ist, muss sich die Gesamtladung über der MIS Struktur ausgleichen. Man erhält also die Gesamtladung auf der Metallseite, indem man sich die gesamte im n-Halbleiter $x < 0$ befindliche Raumladung auf die Grenzfläche zusammengeschoben denkt

$$\sigma_H = \int_{-\infty}^{0} \rho(x) \; dx. \tag{9.12}$$

Ladungsneutralität herrscht dann, wenn $\sigma_M = -\sigma_H$.

ii) Das elektrische Feld am Interface zum Metall ist E_I - was bereits oben berechnet wurde. Man erhält E_I auch so:

$$\begin{aligned} \mathrm{div}\,\vec{D} = \rho \quad &\rightarrow \quad \iiint \mathrm{div}\,\vec{D} \; dV = -\iiint \rho(x) \; dV \\ &\rightarrow \quad \iint \vec{D} \; d\overrightarrow{f} = \sigma_H \; A \\ &\rightarrow \quad \varepsilon_I E_I \; A = \sigma_H \; A \end{aligned} \tag{9.13}$$

bzw.

$$\sigma_H = \varepsilon_H E_H = \varepsilon_I E_I \tag{9.14}$$

iii) Das Potential am Metall errechnet sich dann als Summe des gesamten Potentialabfalls

$$\varphi_M = \varphi_H + \varphi_I = \varphi_H - d_I E_I. \tag{9.15}$$

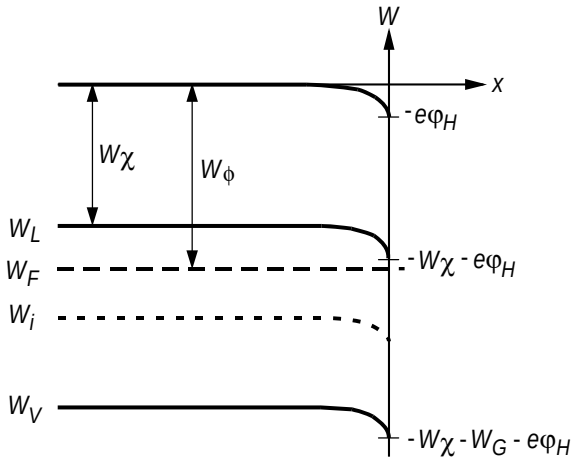


Abbildung 9.6: Anreicherungsrandschicht mit $n(x) > n_{n0}$ in der Grenzschicht eines n-Halbleiters mit $\varphi_H > 0$.

Spannungen
Für die Spannung folgt

$$U = U_I + U_H, \qquad \begin{array}{rcl} U_I & = & \varphi_I = -E_I d_I, \\ U_H & = & \varphi_H = f(\sigma_H). \end{array} \tag{9.16}$$

Kapazitäten
Die Kapazität einer MIS Diode ergibt sich als Serienschaltung aller Kapazitäten. Diese sind

- die Isolatorschichtkapazität C'_I (die Kapazität pro Fläche der Isolatorschicht), welche im wesentlichen eine Plattenkondensatorkapazität darstellt und

$$C'_I \quad = \quad \frac{-\sigma_H}{U_I} = \frac{-\varepsilon_I E_I}{-d_I E_I} = \frac{\varepsilon_I}{d_I}, \tag{9.17}$$

- die Sperrschichtkapazität C'_H der Halbleiterdiode

$$C'_H = -\frac{d\sigma_H}{dU_H} \tag{9.18}$$

  Der Zusammenhang zwischen U_H und σ_H hängt vom speziellen Verlauf von $\rho(x)$ im Halbleiter ab.

Die flächenbezogene Gesamtkleinsignalkapazität C' der MIS-Anordnung ist definiert durch

$$\frac{1}{C'} = -\frac{dU}{d\sigma_H} \tag{9.19}$$

oder auch

$$\frac{1}{C'} = \frac{1}{C'_H} + \frac{1}{C'_I} \text{ d.h. } C' = \frac{C'_I C'_H}{C'_I + C'_H}. \tag{9.20}$$

Im Folgenden wird die MIS-Struktur für verschiedene Spannungen $U \neq 0$ untersucht.

9.1.3 Die Anreicherungsrandschicht

Die Anreicherungsrandschicht mit $\varphi_H > 0$:
Es werde ein Randpotential $\varphi_H > 0$, $\exp(\varphi_H/U_T) \gg 1$ erzwungen, siehe Bild 9.6.

i) In einem ersten Schritt benötigen wir nun einen Ausdruck für die Raumladung, welche sich in der Anreicherungsschicht bildet. Unter Anreicherung wird die Dotierungskonzentration sehr schnell vernachlässigbar, so dass sich Gl. (9.5) zu $\rho(x) \approx -en(x)$ reduziert.

ii) Nun können wir $E(\varphi)$ berechnen. Wir benutzen den Ausdruck (9.8) für die Feldstärke im Halbleiter (beachte auch (5.63) für die Debye-Länge: $L_{Dn} = \sqrt{\frac{\varepsilon_H U_T}{en_D}}$) und erhalten

$$\begin{aligned} E(\varphi) = -\frac{d\varphi}{dx} &= -\operatorname{sgn}(\varphi_H)\sqrt{\frac{en_D}{\varepsilon_H}\int_0^{\varphi} \exp\left(\frac{\varphi'}{U_T}\right)\, d\varphi'} \\ &= -\frac{\sqrt{2}U_T}{L_{Dn}}\sqrt{\exp\left(\frac{\varphi}{U_T}\right) - 1} \\ &= -\frac{\sqrt{2}en_D L_{Dn}}{\varepsilon_H}\sqrt{\exp\left(\frac{\varphi}{U_T}\right) - 1}. \end{aligned} \tag{9.21}$$

iii) Ausdehnung der Raumladungszone: Die Ladungsträgerverteilung $n(x)$ ist sehr stark am Interface HL-Isolator konzentriert. Wenn es sich um eine reine Oberflächenladung handeln würde, dann könnten wir wie in Gl.(9.14) mit (9.21) und unter der Näherung $\exp(\varphi_H/U_T) \gg 1$ schreiben

$$\sigma_H \overset{Gl.(9.14)}{=} \varepsilon_H E_H \overset{Gl.(9.21)}{\approx} -en_D L_{Dn}\sqrt{2}\exp\left(\frac{\varphi_H}{2U_T}\right). \tag{9.22}$$

Dieser Ausdruck zeigt, dass die Flächenladungsdichte σ_H exponentiell mit dem angelegten Potenttial φ_H ansteigt. Da $\exp(2{,}3) \approx 10$ nimmt die Oberflächenladungsdichte σ_H für eine Änderung von $\Delta\varphi_H = 4{,}6\,U_T \approx 115\,\mathrm{mV}$ um eine Zehnerpotenz zu.

Nun handelt es sich bei der Anreicherungsrandschicht aber nicht um eine reine Oberflächenladungsdichte und der korrekte Ausdruck für das Potential $\varphi(x)$ ergibt sich durch Integration von (9.21) unter Beachtung von (9.9). Wir wollen diesen Ausdruck nun dazu verwenden um die äquivalente Dicke d_n der Anreicherungsrandschicht zu berechnen. Die äquivalente Dicke ist die Breite der Anreicherungsschicht $-d_n \le x \le 0$ in welcher 90 % aller Ladungen angerreichert sind. Damit liegt $x =$-d_n (10% der Ladungsträger) auf dem Potential $\Delta\varphi_H = 4{,}6\,U_T$) und $x = 0$ auf φ_H.

$$\int_{-d_n}^{0} \frac{dx}{L_{Dn}} = \frac{1}{\sqrt{2}} \int_{\varphi_H/U_T - 4{,}6}^{\varphi_H/U_T} \frac{d(\varphi/U_T)}{\sqrt{\exp(\varphi/U_T) - 1}},$$

bzw.

$$\frac{d_n}{L_{Dn}} = \sqrt{2}\arctan\sqrt{\exp\left(\frac{\varphi}{U_T}\right) - 1}\Bigg|_{\varphi_H/U_T - 4{,}6}^{\varphi_H/U_T}. \tag{9.23}$$

Beispiel: Die Dicke der Anreicherungsrandschicht, in der 90 % der Gesamtladung liegt, beträgt ungefähr 3 % der Debye-Länge. Bew.: Für $\varphi_H = 12\,U_T \approx 300\,\mathrm{mV}$ erhält man aus (9.23) $d_n/L_{Dn} = 0{,}031$; Wenden wir das auf Si mit $n_D = 10^{15}\,\mathrm{cm}^{-3}$ an, folgt $L_{Dn} = 126\,\mathrm{nm}$ und $d_n = 4\,\mathrm{nm}$, das entspricht der Länge von 8 Gitterkonstanten. (In der schmalen Zone, welche die Form eines Dreieck-Potentialtopfes annimmt, ändert sich die Bandstruktur.) Für σ_H ergibt sich im Fall von Si, $\sigma_H/(-e) = 7{,}2 \cdot 10^{12}\,\mathrm{cm}^{-2}$.

Hätten wir d_n als die Zone, welche die Gesamtladung σ_H beinhaltet definiert, so wäre die untere Integralgrenze in (9.23) mit $\varphi = 0$ anzugeben gewesen. Daraus ergäbe sich ein Wert des Integrals von annähernd $\sqrt{2}\cdot\pi/2 = 2{,}22 = d_n/L_{Dn}$.Die Breite d_n wäre dann zwei mal so lang wie die Debeye-Länge. Das zeigt, dass das Band noch über längere Distanzen verformt bleibt. Allerdings ist diese Verformung klein und die Definition mit der 90 % der Gesamtladung macht mehr Sinn.

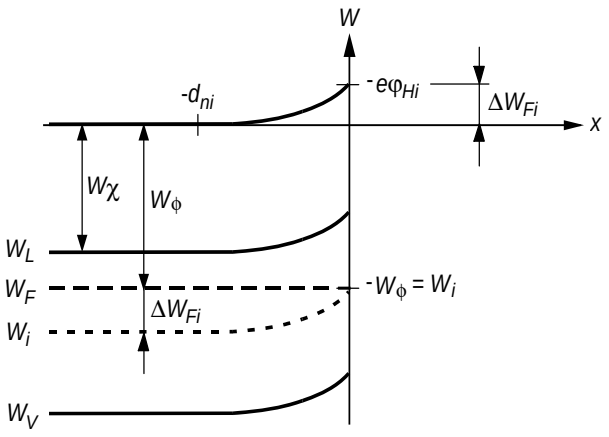


Abbildung 9.7: Verarmungsrandschicht für den Spezialfall $\varphi_H = \varphi_{Hi}$, in dem sich der Halbleiter auf der Oberfläche eigenleitend verhält, $n(0) = p(0) = n_i$. Es ist $-e\varphi_{Hi} = \Delta W_{Fi}$, wobei ΔW_{Fi} der Abstand $|W_F - W_{Fi}|$ im Halbleiterinneren ist.

iv) Für die Kleinsignalkapazität C' der MIS-Struktur (s. (9.19) und (9.20)) erhält man unter Beachtung von (9.22) und $U_I = -E_I d_I$

$$\begin{aligned} C'_I &= \tfrac{\sigma_I}{U_I} = \tfrac{-E_I\varepsilon_I}{-E_I d_I} = \tfrac{\varepsilon_I}{d_I}, \\ C'_H &= -\tfrac{d\sigma_H}{dU_H} \overset{Gl.(9.22)}{=} -\tfrac{\sigma_H}{2U_T} = -\tfrac{\varepsilon_H E_H}{2U_T} = -\tfrac{\varepsilon_I E_I}{2U_T} = C'_I \tfrac{U_I}{2U_T}, \\ C' &= \tfrac{C'_I C'_H}{C'_I + C'_H} = C'_I \tfrac{U_I/(2U_T)}{1+U_I/(2U_T)}. \end{aligned} \tag{9.24}$$

Wenn der Betrag der im Isolator abfallenden Spannung groß ist gegen $2U_T \approx 50\,\mathrm{mV}$, ist $C' \approx C'_I$ (die Anordnung verhält sich wie ein Plattenkondensator mit dem Plattenabstand d_I und dem Isolator als Dielektrikum).

Quasineutralstörungen mit $\varphi_H \geq 0$ oder $\varphi_H \leq 0$, $|\varphi_H/U_T| \ll 1$ werden hier nicht behandelt, sie wurden bereits im Abschn. 5.5 (Debye-Länge) diskutiert.

9.1.4 Die Verarmungsrandschicht

Die Verarmungsrandschicht mit $0 > \varphi_H > \varphi_{Hi}$:

Es sei $\varphi_H < 0$, $|\varphi_H/U_T| \gg 1$. Dadurch werden Elektronen vom Halbleiterrand abgestoßen. Der minimale betrachtete Wert sei $\varphi_H = \varphi_{Hi}$, bei dem sich der Halbleiter auf der Oberfläche gerade eigenleitend verhält. φ_{Hi} erhält man aus Gl. (9.4) durch Einsetzen von $n(x = 0) = n_i$

$$\varphi_{Hi} = -U_T \ln\left(\frac{n_D}{n_i}\right) = -\frac{\Delta W_{Fi}}{e}. \tag{9.25}$$

Im Bereich wo $0 \geq \varphi_H \geq \varphi_{Hi}$ gilt, verarmt die Oberfläche somit an Majoritätsträgern (siehe auch Abb. 9.7) Im Beispiel von Si mit $n_D = 10^5\, n_i$ erhält man $\varphi_{Hi} = -11{,}5\, U_T \approx -300\,\mathrm{mV}$.

i) Der Bereich, in dem $|\varphi_H/U_T| \gg 1$ gilt, kann durch $\rho(x) \approx en_D$ erfaßt werden (siehe (9.5)), die Ladung der beweglichen Träger kann im Vergleich zur Ladung der Donatoren vernachlässigt werden.

ii) Zur Berechung des Potentials, können wir (da hier $\rho(x)$ nicht von φ abhängt) direkt die Poissongleichung (9.6) benutzen. Mit (5.63)

$$\frac{d^2}{dx^2}\left(\frac{\varphi}{U_T}\right) = -\frac{\rho}{\varepsilon_H}\frac{1}{U_T} = -\frac{en_D}{\varepsilon_H U_T} = -\frac{1}{L_{Dn}^2}, \tag{9.26}$$

und der Randbedingung $\varphi(-d_n) = 0$ erhält man dann die Lösung

$$\frac{\varphi(x)}{U_T} = -\frac{1}{2L_{Dn}^2}(x + d_n)^2. \tag{9.27}$$

Für die Feldstärke am Halbleiterrand folgt

$$E_H = -\left.\frac{d\varphi}{dx}\right|_{x=0} = \frac{U_T}{L_{Dn}^2} d_n = \frac{U_T}{L_{Dn}} \sqrt{-\frac{2\varphi_H}{U_T}} = \frac{e n_D d_n}{\varepsilon_H}, \tag{9.28}$$

und somit aus (9.12)

$$\sigma_H = \varepsilon_H E_H = e n_D d_n = \frac{\varepsilon_H U_T}{L_{Dn}} \sqrt{-\frac{2\varphi_H}{U_T}}. \tag{9.29}$$

iii) Die Breite der Verarmungszone folgt aus $\varphi(0) = \varphi_H$ zu

$$d_n = L_{Dn} \sqrt{-\frac{2\varphi_H}{U_T}} = \sqrt{-\frac{2\varepsilon_H \varphi_H}{e n_D}}. \tag{9.30}$$

Die Verarmungsnäherung ist somit für $d_n / L_{Dn} \gg 1$, d. h. $|\varphi_H / U_T| \gg 1$ gerechtfertigt.

iv) Für die Kleinsignalkapazität erhält man im Fall der homogenen Dotierung des Halbleiters

$$C'_H = -\frac{d\sigma_H}{d\varphi_H} = \frac{\varepsilon_H}{d_n}, \tag{9.31}$$

also (zufällig) dieselbe Beziehung wie für einen Plattenkondensator mit dem Plattenabstand d_n und einem Dielektrikum $\varepsilon = \varepsilon_H$. Natürlich gilt auch hier

$$C'_I = \frac{\varepsilon_I}{d_I}. \tag{9.32}$$

Es gilt

$$\frac{1}{C'} = \frac{1}{C'_H} + \frac{1}{C'_I}, \qquad \frac{C'_I}{C'} = 1 + \frac{\varepsilon_I}{\varepsilon_H} \frac{d_n}{d_I}. \tag{9.33}$$

Für die Spannungen erhält man mit (9.27) und (9.28)

$$\begin{aligned} U = \varphi_M = U_H + U_I &= +\varphi_H - d_I E_I \\ &= +\varphi_H - \tfrac{d_I}{\varepsilon_I} \varepsilon_H E_H \\ &= -\tfrac{e n_D}{2\varepsilon_H} d_n^2 - \tfrac{d_I}{\varepsilon_I} e n_D d_n. \end{aligned} \tag{9.34}$$

Berechnet man aus (9.34) $d_n = f(U)$, so erhält man für Spannungen $U < 0$ für die Kleinsignalkapazität der MIS-Struktur aus (9.33) das Ergebnis

$$\frac{C'}{C'_I} = \frac{1}{\sqrt{1 - \frac{2\varepsilon_I^2 U}{e n_D \varepsilon_H d_I^2}}}. \tag{9.35}$$

Für $\varphi_H = \varphi_{Hi}$ resultiert der in Abb. 9.7 gezeichnete Fall ($d_n = d_{ni}$; $U = U_i$; $U_I = U_{Ii}$).

9.1.5 Die Inversionsrandschicht

Für $\varphi_{Hi} > \varphi_H$ gelangt in Abb. 9.8 das Eigenleitungsniveau W_i über das Ferminiveau W_F, und damit verhält sich der Halbleiter in der Grenzschicht wie ein p-Halbleiter (er ist invertiert).

- Solange $\varphi_{Hi} \geq \varphi_H \geq 2\varphi_{Hi}$ erfüllt ist, ist am Rand $p(0) < n_{n0} = n_D$: Dieser Bereich wird als der Bereich schwacher Inversion bezeichnet. Aus den in Abschn. 9.1.4 genannten Gründen ist auch hier $\rho(x) \approx e n_D$ (Verarmungsnäherung) gültig, und somit können alle Gleichungen (9.26) bis (9.35) übernommen werden.

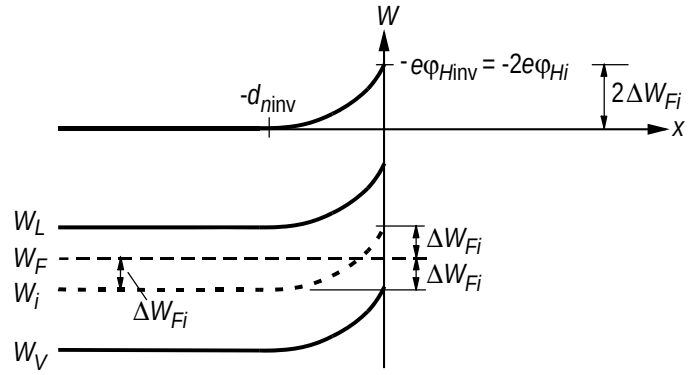


Abbildung 9.8: Einsatzpunkt der starken Inversion, $p(0) = n_{n0} = n_D$.

- Für $\varphi_H = 2\varphi_{Hi} = \varphi_{H\mathrm{inv}}$ wird $p(0) = n_{n0} = n_D$ erreicht; man bezeichnet dies als den Einsatzpunkt der starken Inversion.
- Für $\varphi_H < \varphi_{H\mathrm{inv}}$ wird $p(0) > n_D$ und es kommt zur starken Inversion. Abb. 9.8 zeigt das Einsetzen der starken Inversion.

Es gilt somit:

$$\begin{array}{lcll} \varphi_{Hi} & = & -\frac{\Delta W_{Fi}}{e} = -U_T \ln\left(\frac{n_D}{n_i}\right) \geq \boldsymbol{\varphi_H} \geq 2\varphi_{Hi}, & \text{schwache Inversion,} \\ \boldsymbol{\varphi_H} & = & \varphi_{H\mathrm{inv}} = 2\varphi_{Hi} = -U_T \ln\left(\frac{n_D}{n_i}\right)^2, & \text{Einsatz der starken Inversion,} \\ \boldsymbol{\varphi_H} & < & \varphi_{H\mathrm{inv}}, & \text{starke Inversion.} \end{array} \tag{9.36}$$

Einsatz der starken Inversion Die Spannung U_I, welche bei Einsatz der starken Inversion im Isolator abfällt, wird als ***Einsatzspannung*** oder ***Schwellenspannung*** U_{th0} (ausgehend vom Flachbandfall) bezeichnet. Für reale MIS-Schichten (siehe Abschn. 9.1.2) ist diese Spannung noch um jene Spannung U_{FB} zu vermehren, welche zum Erreichen des Flachbandfalls im Halbleiter benötigt wurde (U_{FB} fällt ebenfalls im Isolator ab, da im Flachbandfall im Halbleiterinneren keine Ladungen sitzen). Es gilt somit für den Einsatz starker Inversion in Anwendung von (9.30) bis (9.34)

$$U_I = U_{I\mathrm{inv}} = U_{th\,0} = -\frac{\sigma_{H\mathrm{inv}}}{C'_I} = -\frac{en_D d_{n\mathrm{inv}}}{C'_I} = -\frac{\sqrt{-2en_D\varepsilon_H\varphi_{H\mathrm{inv}}}}{C'_I}, \tag{9.37}$$

und für die Einsatzspannung einer realen MIS-Struktur im Isolator der Spannungsabfall

$$U_{th} = U_{th\,0} + U_{FB}. \tag{9.38}$$

Beispiel: Für Si mit $n_D = 10^{15}\,\mathrm{cm}^{-3}$, $n_i = 10^{10}\,\mathrm{cm}^{-3}$, $\varepsilon_H = 11{,}8\varepsilon_0$, $\varepsilon_0 = 8{,}85 \cdot 10^{-12}\,\mathrm{As/Vm}$, $d_I = 0{,}1\,\mu\mathrm{m}$, $\varepsilon_I = 3{,}9\varepsilon_0$ (entspricht dem Isolator SiO_2) erhält man: $\varphi_{H\mathrm{inv}} = -0{,}6\,\mathrm{V}$, $d_{n\mathrm{inv}} = 0{,}88\,\mu\mathrm{m}$, $U_{th\,0} = -0{,}4\,\mathrm{V}$; die Anzahl der Ladungen pro Fläche beträgt $\sigma_{H\mathrm{inv}}/e = 0{,}88 \cdot 10^{11}\,\mathrm{cm}^{-2}$. Wegen $\varphi_{Hi} = \varphi_{H\mathrm{inv}}/2 = -0{,}3\,\mathrm{V}$ ist die Ladungsdichte und die Breite der Raumladungszone bei Beginn der schwachen Inversion nur um den Faktor $1/\sqrt{2}$ kleiner, also $\sigma_{Hi}/e = 0{,}62 \cdot 10^{11}\,\mathrm{cm}^{-2}$, $d_{ni} = 0{,}62\,\mu\mathrm{m}$.

Man beachte, daß im Rahmen der Verarmungsnäherung die positive Flächenladung $\sigma_{H\mathrm{inv}}$ (siehe auch (9.29) für $\varphi_H = \varphi_{H\mathrm{inv}}$) vollständig durch unbewegliche ionisierte Donatoren verursacht wird.

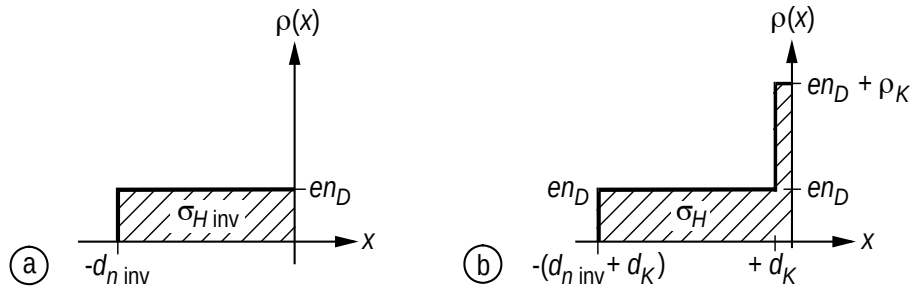


Abbildung 9.9: Raumladungen im Halbleiter (a) im Einsatzpunkt der starken Inversion, (b) bei fortgeschrittener starker Inversion. $d_K \ll L_{Dn} \ll d_{n\mathrm{inv}}$ ist die effektive Breite des Kanals mit beweglicher Löcherladung der Raumladungsdichte $\rho_K = \sigma_K/d_K$.

Starke Inversion i) Für starke Inversion, $\varphi_H < \varphi_{H\mathrm{inv}}$, $p(0) > n_n = n_D$, $|\varphi_H/U_T| \gg 1$, dominiert die Raumladung der beweglichen, an die Oberfläche gezogenen Löcher. Nach (9.5) ist

$$\rho(x) \approx e p_{n0} \exp\left(-\frac{\varphi}{U_T}\right) = \frac{en_i^2}{n_D} \exp\left(-\frac{\varphi}{U_T}\right). \tag{9.39}$$

ii) Die Feldstärke und die Flächenladung lassen sich mit (9.8) und (9.12) berechnen. Man erhält durch Einsetzen von $\rho(x)$ aus (9.39) und $(n_i/n_D = \exp(\varphi_{Hi}/U_T))$

$$\begin{aligned} E(\varphi) &= -\tfrac{d\varphi}{dx} &&= \exp\left(\tfrac{\varphi_{Hi}}{U_T}\right) \cdot \tfrac{U_T}{L_{Dn}} \sqrt{2} \sqrt{\exp\left(-\tfrac{\varphi}{U_T}\right) - 1} \\ &&&= \exp\left(\tfrac{\varphi_{Hi}}{U_T}\right) \tfrac{en_D L_{Dn}}{\varepsilon_H} \sqrt{2} \sqrt{\exp\left(-\tfrac{\varphi}{U_T}\right) - 1}, \\ \text{und} \quad \sigma_H &= \varepsilon_H E_H = \varepsilon_I E_I \approx \exp\left(\tfrac{\varphi_{Hi}}{U_T}\right) en_D L_{Dn} \sqrt{2} \exp\left(-\tfrac{\varphi_H}{2U_T}\right). \end{aligned} \tag{9.40}$$

iii) Analog zu Abschn. 9.1.3, (9.23), läßt sich zeigen, daß die bewegliche Löcherladung fast vollständig in einer im Vergleich zur Debye-Länge sehr dünnen Schicht akkumuliert. Für $\varphi_H = 2\varphi_{Hi} - 0{,}3\,\mathrm{V} = \varphi_{H\mathrm{inv}} - 12\,U_T$ erhält man die Zahlenwerte von Abschn. 9.1.3 , also $\sigma_H/e = 7{,}2 \cdot 10^{12}\,\mathrm{cm}^{-2}$, und eine Schichtdicke des Kanals der beweglichen Löcherladung von $d_K/L_{Dn} = 0$,031.

Wegen der kleinen effektiven Breite d_K des Inversionskanals der beweglichen Löcherladung ändert sich die Breite der Raumladungszone im n-Halbleiter für $\varphi_H < \varphi_{H\mathrm{inv}}$ nur geringfügig.

iv) Aus Abb. 9.9 liest man ab (σ_H ist durch (9.40) gegeben)

$$\begin{aligned} \sigma_H &= en_D(d_{n\mathrm{inv}} + d_K) + \rho_K d_K \approx en_D d_{n\mathrm{inv}} + \rho_K d_K = \sigma_{H\mathrm{inv}} + \sigma_K \\ &= -C_I' U_{th\,0} - C_I'(U_I - U_{th\,0}), \\ \rightarrow \sigma_{H\mathrm{inv}} &= -C_I' U_{th\,0}, \\ \rightarrow\ \sigma_K &= \rho_K d_K = C_I'(U_{th\,0} - U_I). \end{aligned} \tag{9.41}$$

In (9.41) wurde $\sigma_{H\mathrm{inv}}$ aus (9.37) eingesetzt.
Da die bewegliche Kanalladungsdichte σ_K tatsächlich an der Oberfläche des Halbleiters entsteht, kennen wir deren Kapazität. Es ist die Oxydkapazität C_I'. Die Kanalladungsdichte muss dann proportional zur Abweichung der Isolatorspannung U_I von der Schwellenspannung $U_{th\,0}$ sein. (Diese Überlegung gilt, weil $d_I \gg d_K$ was für $d_I = 100\,\mathrm{nm}$, $d_K = 4\,\mathrm{nm}$ sicher erfüllt ist).

Für reale MIS-Strukturen folgt somit (siehe (9.38))

$$\sigma_K = \rho_K d_K = C_I'(U_{th} - U_I) = C_I'(U_{th\,0} + U_{FB} - U_I). \tag{9.42}$$

Beispiel: Für eine Al-SiO_2-Si-Schichtung mit $n_D = 10^{15}\,\mathrm{cm}^{-3}$ ist (siehe Bild 9.5(a)) $\phi_{HI} = 3{,}25\,\mathrm{eV}$, $\phi_{MI} = 3{,}2\,\mathrm{eV}$, $W_\phi - W_\chi = -U_T \ln(n_D/N_L) = 0$,25 eV, woraus eine Flachbandspannung (siehe (9.2))

$$U_{FB} = \frac{1}{e}[\phi_{MI} - \phi_{HI} - (W_\phi - W_\chi)] = 3{,}2\,\mathrm{V} - 3{,}25\,\mathrm{V} - 0{,}25\,\mathrm{V} = -0{,}3\,\mathrm{V} \tag{9.43}$$

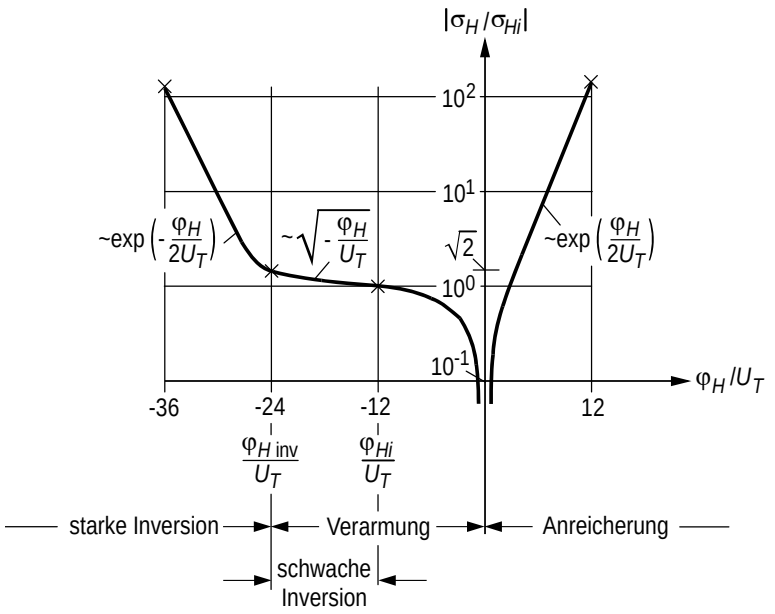


Abbildung 9.10: Ladungen im Halbleiter als Funktion des Oberflächenpotentials für n-Si mit n_D=10^{15}cm^{-3}.

resultiert. Außerdem gibt es (je nach der Orientierung des Si-Kristalls) an der Grenze zum SiO_2 unbewegliche positive Oberflächenladungen mit der Ladungsdichte σ_O und der Konzentration $\sigma_O/e = 2 \cdot 10^{10} \ldots 5 \cdot 10^{11}\,\mathrm{cm}^{-2}$, die zu einer weiteren Korrektur der Flachbandspannung von $\Delta U_{FB} = -\sigma_O/C'_I = -0{,}1 \ldots -2{,}3\,\mathrm{V}$ führt.

Die Schwellenspannung ausgehend vom Flachbandfall von $U_{th\,0} = -0{,}4\,\mathrm{V}$ erhöht sich demnach real auf $U_{th} = -0{,}8 \ldots -3\,\mathrm{V}$.

9.1.6 Der MIS-Kondensator

Wir wollen die Ladungen σ_H im Halbleiter für die verschiedenen Fälle für vorgegebene Werte von φ_H/U_T ausrechnen. Nach (9.22), (9.29) und (9.40) erhält man für einen mit $n_D = 10^{15}\,\mathrm{cm}^{-3}$ dotierten Si-Kristall folgende Oberflächenpotentiale und Anzahldichten:

Anreicherung	φ_H/U_T	=	$12:$	$\sigma_H/(-e)$	=	$7{,}2 \cdot 10^{12}\,\mathrm{cm}^{-2}$
Eigenleitend	φ_H/U_T	=	$\varphi_{Hi}/U_T \approx -12:$	σ_{Hi}/e	=	$0{,}62 \cdot 10^{11}\,\mathrm{cm}^{-2}$
Inversions-Einsatz	φ_H/U_T	=	$\varphi_{H\mathrm{inv}}/U_T \approx -24:$	$\sigma_{H\mathrm{inv}}/e$	=	$\sigma_{Hi}\sqrt{2}/e = 0{,}88 \cdot 10^{11}\,\mathrm{cm}^{-2}$
starke Inversion	φ_H/U_T	$\approx$	$-36:$	σ_{Hi}/e	=	$7{,}2 \cdot 10^{12}\,\mathrm{cm}^{-2}$

In Abb. 9.10 ist die prinzipielle Abhängigkeit der Größe σ_H vom Oberflächenpotential angegeben.

Die Kleinsignalkapazität C' des MIS-Kondensators ist die Serienschaltung der Kleinsignalkapazität des Halbleiters C'_H und der Isolatorkapazität C'_I.

- **Für positive Spannungen, i.e. für Anreicherung,** $U = U_I + U_H$ ist sie wegen der Bildung einer dünnen Anreicherungsrandschicht am Halbleiter (n-Si vorausgesetzt) nach (9.24) praktisch identisch mit der Isolatorkapazität C'_I. Die MIS Struktur verhält sich wie ein Plattenkondensator.

- Im **Fall der Verarmung** (es spielt nur die Raumladung der ionisierten Donatoren eine Rolle) gilt (9.35). C' erreicht ein Minimum, wenn C'_H minimal wird: Das ist der Fall für die maximale Breite $d_{n\mathrm{inv}}$ der Verarmungszone. In (9.35) ist daher $U = U_{I\mathrm{inv}} + \varphi_{H\mathrm{inv}} = U_{th\,0} + \varphi_{H\mathrm{inv}}$ einzusetzen (siehe (9.37)). Für reale MIS-Strukturen ist $U_{th\,0}$ durch $U_{th} = U_{th\,0} + U_{FB}$ zu ersetzen.

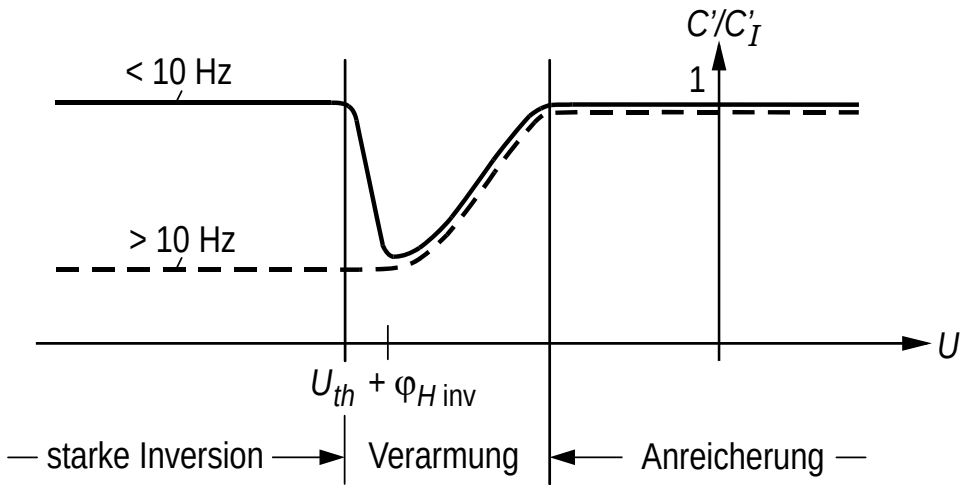


Abbildung 9.11: Verlauf der Kleinsignalkapazität C' der MIS-Struktur als Funktion der anliegenden Gleichspannung U (Vorzeichen siehe Bild 9.5).

- Da sich im **Fall der starken Inversion** ($\varphi_H < \varphi_{H\mathrm{inv}}$) die Breite der Raumladungszone im Halbleiter praktisch nicht mehr ändert und weitere Ladungen im Halbleiter in einem dünnen Inversionskanal an der Oberfläche hinzugefügt bzw. abgebaut werden, ist bei einer Spannungsänderung nur die Isolatorkapazität C'_I bemerkbar, $C' = C'_I$. Das gilt aber nur dann, wenn die Spannungsänderung mit hinreichend niedrigen Frequenzen erfolgt. Für hochfrequente Spannungsänderungen mißt man weiterhin die minimale Kapazität C' (siehe Abb. 9.11). Der Grund für dieses Verhalten wird in Abb. 9.12 erläutert: Eine zusätzliche negative Ladung am Metall wird durch eine zusätzliche positive Ladung im Inversionskanal kompensiert. Dazu muß in der Verarmungszone eine Valenzbindung aufbrechen, Elektron und Loch werden im Feld getrennt (Abb. 9.12(a), langsamer Vorgang). Nimmt man die negative Ladung am Metall wieder weg, so muß auch die positive Ladung aus dem Inversionskanal entfernt werden. Dies geschieht dadurch, daß ein hinreichend energetisches Elektron das bremsende Feld in der Verarmungszone überwindet und mit dem Loch rekombiniert (Abb. 9.12(b), langsamer Vorgang). Bei einer "schnellen" Änderung der Ladung am Metall (Abb. 9.12(c), Abb. 9.12(d) kann die zusätzliche positive Ladung im Halbleiter nur dadurch erzeugt werden, daß man von der Grenze der Verarmungszone die Majoritätsträger nach links verschiebt. Dadurch erscheinen positive, raumfeste Donatorladungen. Ähnlich wird durch Verschieben der Grenze zur Verarmungszone diese Donatorladung wieder kompensiert. Aus den Abständen der beteiligten Ladungen sieht man, daß für langsame Vorgänge $C' = C'_I$, für schnelle Vorgänge dagegen $1/C' = 1/C'_H + 1/C'_I$ gilt.

9.2 Der Metall-Halbleiterkontakt

9.2.1 Die Austrittsarbeit

Das ***Schottky-Modell*** des MS (metal-semiconductor, auch: MES)-Kontakts nimmt an, daß die Bandschemata der unendlich ausgedehnten Materialien bei der Kontaktgabe nicht modifiziert werden. Abb. 9.13 zeigt die Verhältnisse für einen Kontakt zwischen einem n-Halbleiter und einem Metall.

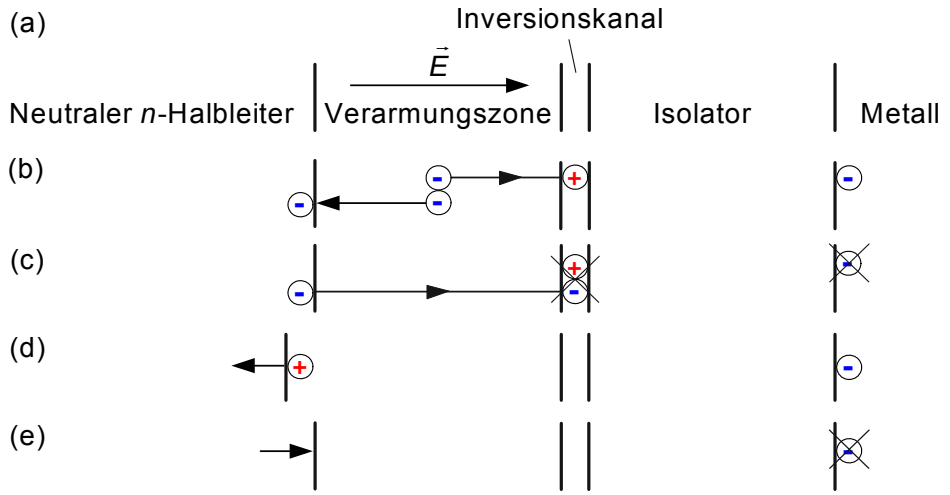


Abbildung 9.12: Erläuterung der Frequenzabhängigkeit der Kleinsignalkapazität einer MIS-Struktur. (a) , (b) Aufbau und Abbau einer positiven Ladung im Inversionskanal bei langsamer Potentialänderung; (c), (d) Schaffung und Abbau einer weiteren positiven Ladung durch Verschieben der Grenze der Verarmungszone, wenn für die Vorgänge nach (a), (b) keine Zeit zur Verfügung steht.

Der ideale Metall-Halbleiterkontakt

Das ***Kontaktpotential (built-in potential)*** φ_H (im gezeichneten Fall $\varphi_H > 0$) ist durch die Differenz der Austrittsarbeiten bestimmt

$$-e\varphi_H = W_{\phi 2} - W_{\phi 1}. \tag{9.44}$$

Bem. 1: Das Kontaktpotential gibt sozusagen die Verformung der Leitungsbandkante an. Ganz analog wie das Diffusionspotential U_D die Verformung der Bandkanten beim pn-Übergang bestimmt hat. Im englischen Sprachgebrauch wird deshalb sowohl das Kontaktpotential als auch das Diffusionspotential als "built-in potential"bezeichnet.

Bem. 2: Bei Kontaktgabe sind Elektronen aus einer sehr dünnen Oberflächenzone des Metalls in den Halbleiter geflossen (die charakteristische Abschirmlänge für Potentiale im Metall, die sogenannte ***Thomas-Fermi Distanz*** ist typisch $0.5\,\mathring{A}$ und somit sehr viel kleiner als die Debye-Länge in Halbleitern) und haben im Halbleiter eine Anreicherungsrandschicht erzeugt.

Man definiert die ***Austrittsarbeiten (barrier heights)*:**

- $\phi_{\mathrm{MH}}^{(n)}$ für Elektronen aus dem Metall in den Halbleiter (positiv, wenn die LB-Kante an der Stelle $x = 0$ oberhalb des Ferminiveaus des Metalls liegt) und
- $\phi_{\mathrm{MH}}^{(p)}$ für Löcher (positiv, wenn die VB-Kante an der Stelle $x = 0$ unterhalb des Ferminiveaus liegt):

$$\begin{aligned} \phi_{\mathrm{MH}}^{(n)} &= W_{\phi 2} - W_{\chi 1}, \\ \phi_{\mathrm{MH}}^{(p)} &= (W_{\chi 1} + W_G + e\varphi_H) - W_{\phi 1} = W_G - W_{\phi 2} + W_{\chi 1}, \\ \phi_{\mathrm{MH}}^{(n)} + \phi_{\mathrm{MH}}^{(p)} &= W_G. \end{aligned} \tag{9.45}$$

Die Summe der Austrittsarbeiten ist gleich dem Bandabstand des Halbleiters.

Im Gegensatz zur Kontaktspannung hängen die Austrittsarbeiten nicht von der Dotierung des Halbleiters ab. Der in Abb. 9.13 gezeichnete Fall entspricht im Prinzip dem von Al ($W_{\phi 2} = 3{,}74$ eV) auf Si ($W_{\chi 1} = 4{,}01\,\mathrm{eV}$) oder GaAs ($W_{\chi 1} = 4{,}07\,\mathrm{eV}$), womit man die Austrittsarbeiten $\phi_{\mathrm{MH}}^{(n)} = -0{,}27\,\mathrm{eV}$ (Al-Si) und $\phi_{\mathrm{MH}}^{(n)} = -0{,}33\,\mathrm{eV}$ (Al-GaAs) prognostizieren würde. In beiden Fällen

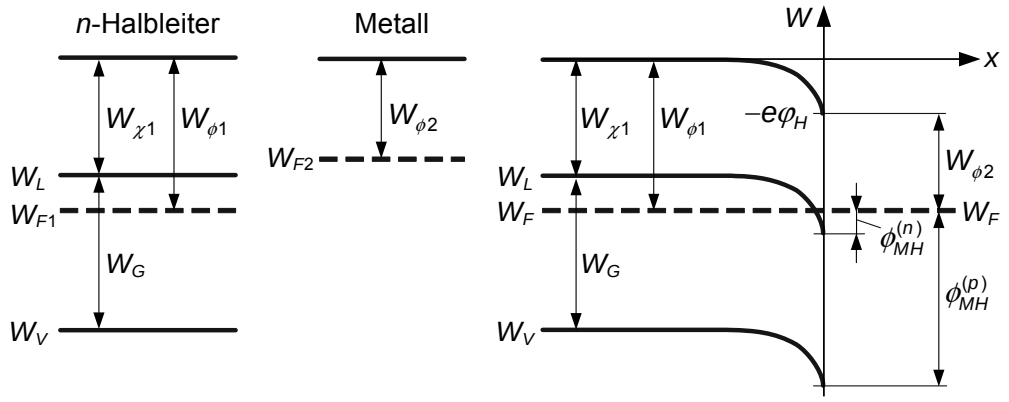


Abbildung 9.13: Schottky-Kontakt zwischen einem n-Halbleiter und einem Metall. $\Phi_{MH}^{(n)}$ ist die Austrittsarbeit der Elektronen vom Metall in den Halbleiter ($\Phi_{MH}^{(n)} > 0$, wenn die LB-Kante des Halbleiters an der Stelle $x = 0$ höher liegt als das Ferminiveau des Metalls nach Kontaktgabe), für Löcher ist $\Phi_{MH}^{(p)} > 0$, wenn die VB-Kante an der Stelle $x = 0$ unterhalb des Ferminiveaus liegt.

handelte es sich um sogenannte ohmsche Kontakte, deren Widerstand nicht (merklich) von einer angelegten äußeren Spannung abhängt: Es ändert sich dadurch zwar die Trägerkonzentration in der Anreicherungsrandschicht, aber deren wegen der hohen Trägerdichte kleiner Widerstand macht sich in Serie zum Widerstand des Bahngebietes des Halbleiters nicht bemerkbar.

Der reale Metall-Halbleiterkontakt

Tatsächlich wird der Fall von Abb. 9.13 nicht beobachtet, sondern man mißt praktisch ganz andere Austrittsarbeiten

$$\begin{array}{llllll} \text{Al-Si:} & \phi_{\mathrm{MH}}^{(n)} = 0{,}6\,\mathrm{eV} & \text{(Si:} & W_G = 1{,}1\,\mathrm{eV}, & \text{also} & \phi_{\mathrm{MH}}^{(p)} = 0\,,54\,W_G), \\ \text{Al-GaAs:} & \phi_{\mathrm{MH}}^{(n)} = 0{,}8\,\mathrm{eV} & \text{(GaAs:} & W_G = 1{,}42\,\mathrm{eV}, & \text{also} & \phi_{\mathrm{MH}}^{(p)} = 0{,}56\,W_G). \end{array} \tag{9.46}$$

Es stellt sich heraus, daß man $\phi_{\mathrm{MH}}^{(n)}$ Tabellenwerken entnehmen muss.

Addentum: $\phi_{\mathrm{MH}}^{(n)}$ ist hauptsächlich durch den verwendeten Halbleiter bestimmt wird, aber (im Gegensatz zu der Behauptung in Gl. (9.45)) nur schwach vom verwendeten Metall abhängig. Je kleiner der Bandabstand des verwendeten Halbleiters ist und je mehr die Bindung im Halbleiter kovalent ist (Gegensatz: ionische Bindung), desto unabhängiger ist die Austrittsarbeit vom verwendeten Metall. (9.45) könnte man durch die Beziehung

$$\phi_{\mathrm{MH}}^{(n)} = c_1(\mathrm{Hbl})[W_{\phi 2} - W_{\chi 1}] + c_2(\mathrm{Hbl}) \tag{9.47}$$

ersetzen, wobei c_1, c_2 von den Halbleitereigenschaften abhängig sind. $c_1 = 0$, $c_2 \neq 0$ ist in der Tendenz für Halbleiter mit kleinem Bandabstand und kovalenter Bindung (etwa Ge, Si) erfüllt, wachsende Werte von c_1 ergeben sich mit steigenden Anteilen ionischer Bindung (also über III-V-Halbleiter wie GaAs zu II-VI-Halbleitern wie ZnS) und mit steigendem Bandabstand.

Grund für die Diskrepanz zwischen (9.45) und (9.47) ist, daß die reale Grenze zwischen Halbleiter und Metall nicht atomar abrupt erfolgt, daß sich aus den beiden Materialien komplexe Verbindungen in einer Zwischenschicht bilden können und daß die Wahrscheinlichkeitsdichteamplituden der Elektronen aus dem Metall exponentiell gedämpft in den Halbleiter im verbotenen Band eindringen können, wodurch an der Grenze im verbotenen Band des

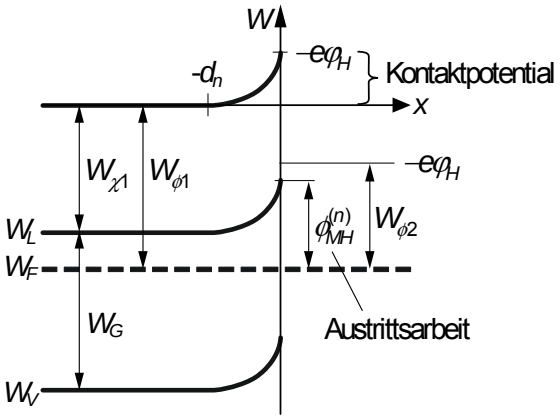


Abbildung 9.14: Kontakt zwischen n-Halbleiter und Metall für reale Austrittsarbeit $\Phi_{MH}^{(n)}$: Im Halbleiter entseht eine Verarmungsschicht.

Halbleiters lokalisierte Zustände für Elektronen (sogenannte MIGS = metal-induced-gap-states) entstehen. Das führt zu Oberflächenladungen und Dipolschichten an der Grenzfläche: Eine Dipolschicht ist der Grenzfall von zwei Flächenladungen $\pm\sigma$ im Abstand Δd, wobei für $\Delta d \to 0$, $\sigma \to \infty$ das Produkt $\sigma\Delta d = p_{\mathrm{Dip}}$ konstant gehalten wird. p_{Dip} ist die Dipolmomentendichte (Einheit: $\mathrm{As\,m^{-1}}$) der Schicht. Dipolschichten führen zu Potentialsprüngen, da die Raumladungsdichte einer Dipolschicht

$$\rho_{\mathrm{Dip}}(x) = \lim[\sigma\delta(x + \Delta d/2) - \sigma\delta(x - \Delta d/2)] = \lim \sigma\Delta d\delta'(x) = p_{\mathrm{Dip}}\delta'(x) \tag{9.48}$$

nach Einsetzen in die Poissongleichung und zweimaliger Integration

$$\varphi(x) = -p_{\mathrm{Dip}}H(x)/\varepsilon \tag{9.49}$$

liefert. Das Potential ändert sich beim Durchgang durch die Dipolschicht von der positiven Seite zur negativen Seite abrupt um $-p_{\mathrm{Dip}}/\varepsilon$.

Unter Berücksichtigung von (9.46) ergibt sich für den MS-Kontakt von Bild. 9.13 ein völlig anderes Verhalten (siehe Bild. 9.14): An der Grenze entsteht ein Potentialsprung (die negative Seite der Dipolschicht liegt im Halbleiter, die positive Seite im Metall); für die Randpotentiale liest man aus Bild. 9.14 ab

$$\begin{aligned} -e\varphi_M &= W_{\phi 2} - W_{\phi 1}, \\ -e\varphi_H &= \phi_{\mathrm{MH}}^{(n)} - (W_{\phi 1} - W_{\chi 1}). \end{aligned} \tag{9.50}$$

Es ist im Halbleiter eine Verarmungsrandschicht entstanden, deren Breite groß gegen die Debye-Länge sei, $d_n \gg L_{Dn}$ (damit ist d_n auch groß gegen die freie Weglänge der Elektronen im Halbleiter, die z. B. bei GaAs ca. 50 nm beträgt): In ihr braucht infolgedessen nur die Raumladung $\rho(x) = en_D$ der ionisierten Donatoren zu berücksichtigt werden. Durch $\phi_{\mathrm{MH}}^{(n)}$ sind die Randkonzentrationen $n(0)$, $p(0)$ der Elektronen und Löcher am Rand im Halbleiter im Gleichgewichtsfall bestimmt; nach (5.14) folgt

$$n(0) = \frac{n_i^2}{p(0)} = N_L \exp\left(-\frac{\phi_{\mathrm{MH}}^{(n)}}{kT}\right) = n_D \exp\left(-\frac{e\varphi_H}{kT}\right). \tag{9.51}$$

Da sich bei Anlegen einer äußeren Spannung das Quasiferminiveau des Metalls relativ zu dem des Halbleiters verschiebt, aber $\phi_{\mathrm{MH}}^{(n)}$ sich nicht ändert (Abstand der LB-Kante vom Quasiferminiveau im Metall), verschiebt sich auch die LB-Kante: Die Verarmungszone wird abgebaut (Flußfall) oder weiter verstärkt (Sperrfall). Der Kontakt hat Gleichrichtereigenschaften.

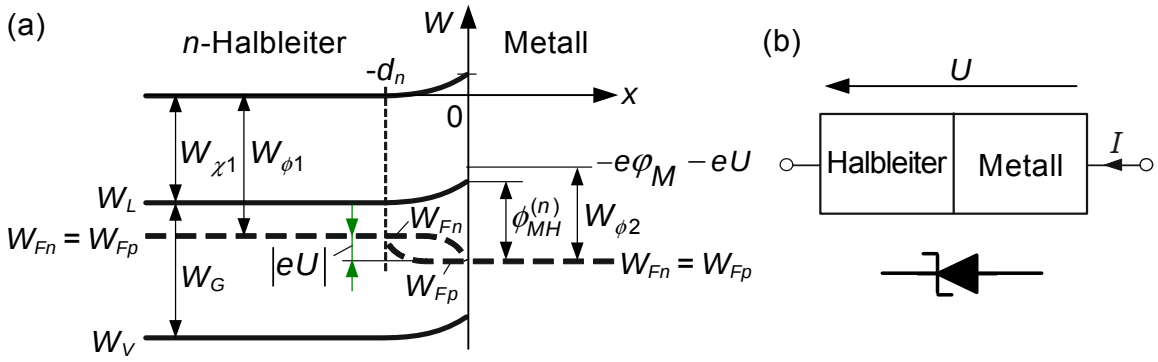


Abbildung 9.15: (a) MS-Kontakt mit Gleichrichtereigenschaft bei Anlegen einer Flußspannung $U > 0$: Die Energiebarriere für Elektronen an der Leitungsbandkante des Halbleiters zum Übertritt in das Metall hat sich im Vergleich zu Bild 9.14 erniedrigt. (b) Schaltsymbol und Definition der Flussrichtung.

9.2.2 Schottky-Diode

Das Bandschema des realen MS-Kontakts von Bild 9.14 ist in Bild 9.15 für eine angelegte Flußspannung $U > 0$ gezeichnet. In der Verarmungszone bleiben die Quasiferminiveaus annähernd konstant, wenn Generation und Rekombination von Trägern vernachlässigbar sind.

Für den Strom berechnet man (die Rechnung wird am Ende dieses Abschnitts skizziert)

$$\begin{aligned} I &= I_n + I_p = Ae\left[\frac{D_n n(0) d_n}{2L_{Dn}^2} + \frac{D_p p_{n0} d_n}{2L_{Dn}^2}\right]\left(e^{U/U_T} - 1\right), \\ d_n &= L_{Dn}\sqrt{-\frac{2(\varphi_H + U)}{U_T}}. \end{aligned} \tag{9.52}$$

A ist die Querschnittsfläche des Kontaktes, $n(0)$ wurde in (9.51) definiert und d_n ist die Breite der Verarmungszone für das tatsächliche Randpotential $\varphi_H + U$ gemäß (9.30). Man beachte, daß wegen $\varphi_H < 0$, $0 < U < |\varphi_H|$ im Flußfall die Breite d_n der Verarmungszone kleiner ist als im Gleichgewicht $U = 0$; d_n ist leicht spannungsabhängig.

Das Arbeitsprinzip der Schottky-Diode entnimmt man am Besten den drei Plots von Fig. 9.16. Einige Eigenschaften der Schottky-Diode:

- Unipolar-Bauteil: Aus (9.52) sieht man, (zur Abschätzung wird $D_n = D_p$, $N_L = N_V$ angenommen) daß der Majoritätsträgerstrom γ-mal so groß ist wie der Minoritätenstrom, wenn $n(0) = \gamma p_{n0}$ erfüllt ist. Um ein Gefühl für die Größenordung von γ als Funktion der Dotierung und als Funktion des Metall-Halbleiterkontaktes zu bekommen, beginnen wir mit $n(0)$ aus (9.51) und verwenden n_i aus (**??**) sowie $p_{n0} = n_i^2/n_D$. Dies führt für eine gegebene Dotierung auf

$$\gamma = \frac{n(0)}{p_{n0}} = \frac{n(0)}{n_i^2} n_D = \frac{N_L \exp\left(-\frac{\phi_{\mathrm{MH}}^{(n)}}{kT}\right)}{n_i^2} n_D = \frac{n_D}{n_i \exp\left(\frac{\phi_{\mathrm{MH}}^{(n)} - W_G/2}{kT}\right)} \tag{9.53}$$

  Mit den Austrittsarbeiten von (9.46) erhält man für Al-Si-Dioden $\gamma = n_D/(7{,}4\ n_i)$ und für Al-GaAs-Dioden $\gamma = n_D/(36{,}6 n_i)$ das bedeutet, daß für alle praktisch vorkommenden Dotierungen wegen $n_D \gg n_i$ der Wert von γ sehr groß ist und damit der Löcherstrom einen verschwindend kleinen Beitrag zum Gesamtstrom liefert; MS-Gleichrichter sind von den Majoritätsträgerströme dominiert. Sie sind deshalb unipolar-Bauteile.

- Hohe Schaltgeschwindigkeiten: Gerade weil die MS-Gleichrichter auf den Majoritätsträgern basierende unipolar Bauteile sind, sind sie bis zu extrem hohen Frequenzen verwendbar, da

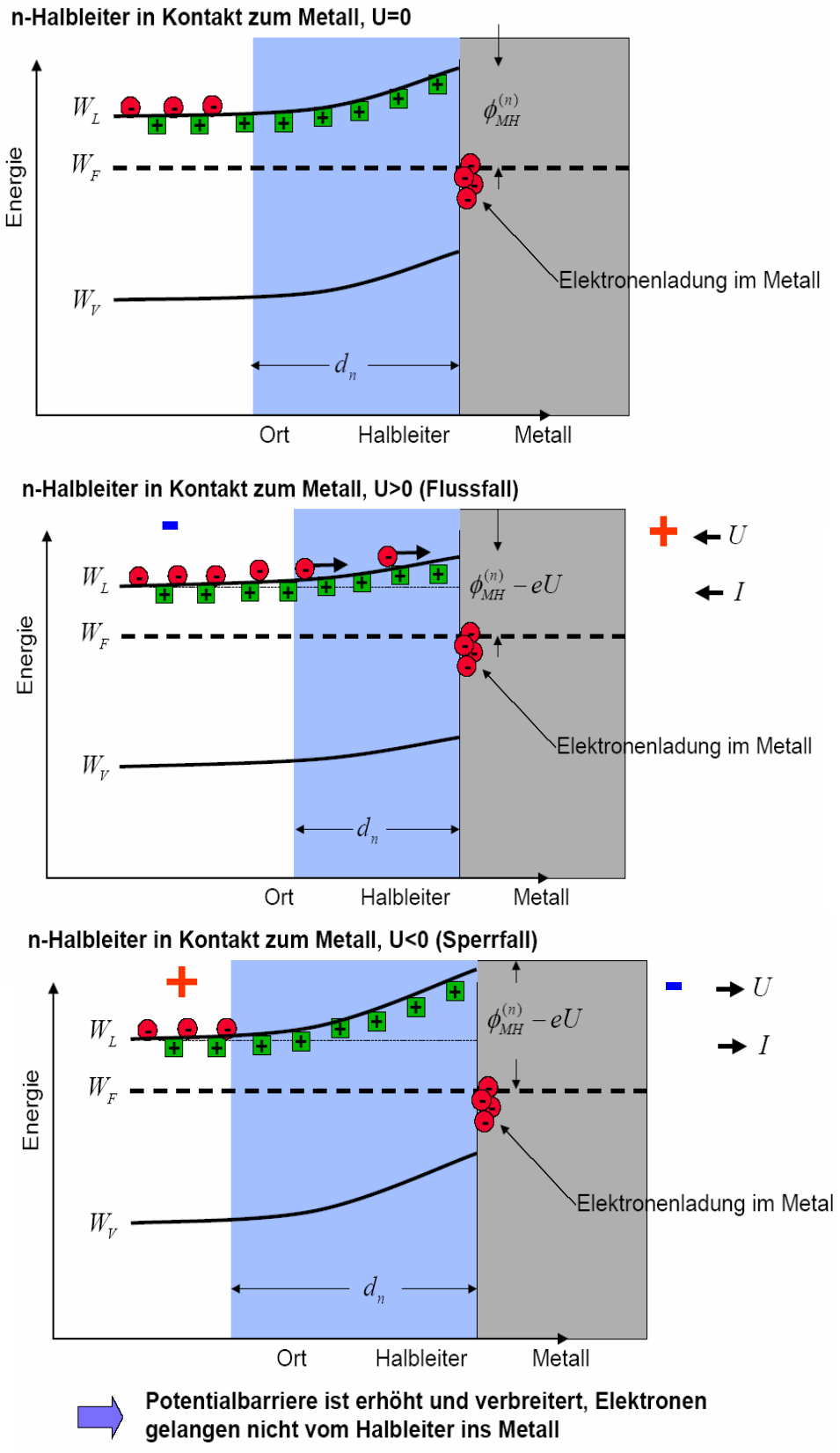


Abbildung 9.16: Arbeitsprinzip der Schottky-Diode. (a) Banddiagram für U=0, (b) Banddiagrmm für Betrieb in Flussrichtung und (c) für Betrieb in Sperrrichtung.

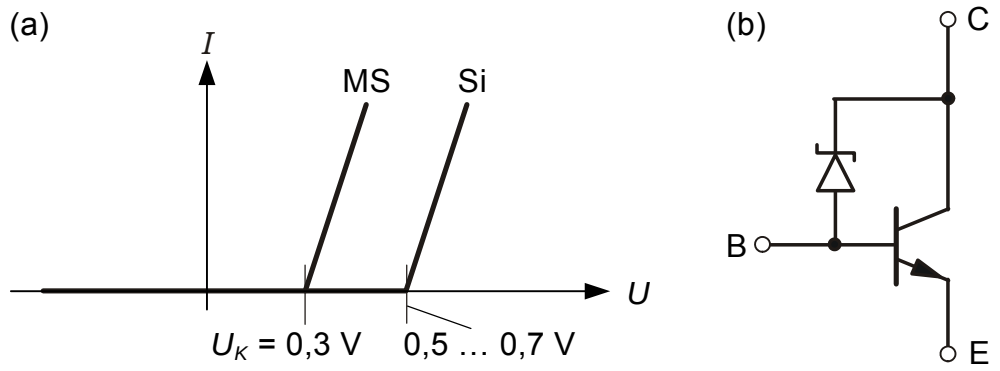


Abbildung 9.17: (a) Einsatz-(Knick-) Spannungen von Al-Si-MS-Dioden und Si-pn-Dioden, (b) Schottky-TTL: MS-Diode überbrückt die CB-Diode eines npn-Transistors.

sich Majoritätsträger bei Spannungsänderungen innerhalb einer dielektrischen Relaxationszeit rearrangieren können.

- Niedere Einsatzspannung: Ein weiterer wichtiger Effekt ergibt sich durch Vergleich der MS-Kennlinie (9.52) mit der eines Gleichrichters mit pn-Übergang (5.99): Um den Verlauf zu verstehen, ist es hilfreich den pn-Dioden-Sperrstrom mit dem Schottky-Dioden-Sperrstrom zu vergleichen

  $$\text{pn-Sperrstrom: } I_S = Ae\left(\frac{D_n n_p}{L_n} + \frac{D_p p_n}{L_p}\right)$$

  .

  $$\text{Schottky-Sperrstrom: } I_S = Ae\left[\frac{D_n n(0) d_n}{2L_{Dn}^2} + \frac{D_p p_{n0} d_n}{2L_{Dn}^2}\right]$$

  Da in der Regel $n(0) \geq n_{p0}$ (n_{p0} Minoritätsträgerdichte im p-Halbleiter bei der Diode mit pn-Übergang), aber sicher $L_n \gg 2\,L_{Dn}^2/d_n$ erfüllt ist, liefert der MS-Gleichrichter bei gleicher Flußspannung einen größeren Strom als der Si-pn-Gleichrichter. Bei der Näherung der Diodenkennlinie durch eine geknickte Gerade ist die Knickspannung ***(Einsatzspannung)*** der MS-Diode, siehe Abb. 9.17(a), bei ca. $U_K = 0{,}3\,\text{V}$, beim Si-Gleichrichter bei $0{,}5 \ldots 0{,}7\,\text{V}$. Das nutzt man bei der Schottky-TTL (Transistor-Transistor-Logik mit MS-Dioden) aus, siehe Abb. 9.17(b): Wenn die Kollektor-Basis-Diode eines npn-Bipolartransistors übersteuert wird (Flußspannung an der Kollektor-Basis-Diode), kommt es zufolge der Elektroneninjektion vom Kollektor in die Basis zu einer hohen Ladungsspeicherung in der Basis, wodurch die Schaltzeiten vergrößert werden. Da die Einsatzspannung der MS-Diode kleiner ist als die der Kollektor-Basis-Diode, wird bei Übersteuerung der Strom durch die MS-Diode übernommen und führt nicht zu einer Ladungsspeicherung in der Basis des Transistors.

- Sperrstrom: Beim MS-Gleichrichter sind die Restströme im Sperrgebiet $U < 0$ größer als beim Si-Gleichrichter.

Berechnung der Kennlinie der Schottky-Diode, Gl. (9.52)

Bei der Rechnung ist Abb. 9.15 zu beachten.

Den Elektronenanteil I_n am Gesamtstroms I erhält man als Ableitung der Quasifermifunktion für die Elektronen gemäß Gl. (5.19). Unter Beachtung der Zählpfeile in Abb. 9.15 (A ist die Querschnittsfläche der Diode) ergibt dies

$$I_n = -An(x)\mu_n \frac{\mathrm{d}W_{Fn}}{\mathrm{d}x}, \tag{9.54}$$

Der Term $n(x)$ folgt aus (5.21) und $W_L(x)$ entnimmt man Bild 9.15.

$$\begin{aligned} n(x) &= N_L \exp\left[\frac{W_{Fn}(x)-W_L(x)}{kT}\right], \\ W_L(x) &= -W_{\chi 1} - e\varphi(x). \end{aligned} \tag{9.55}$$

Der Potentialverlauf $\varphi(x)$ in der Verarmungszone ist durch (9.26) bestimmt, die Bedeutung von d_n ergibt sich in sinngemäßer Anwendung von (9.30) (siehe (9.52)). Da man mit dem quadratischen Verlauf von $\varphi(x)$ in der weiteren Rechnung Fehlerfunktionen erhalten würde, wird hier für das Potential ein die Randbedingungen erfüllender linearer Verlauf angesetzt (dann sind die sich ergebenden Integrale elementar lösbar):

$$\begin{aligned} &\varphi(x) = -\tfrac{U_T}{2L_{Dn}^2}(x+d_n)^2 \approx -\tfrac{U_T d_n}{2L_{Dn}^2}(x+d_n), \\ &d_n = L_{Dn}\sqrt{-\tfrac{2(\varphi_H+U)}{U_T}}. \end{aligned} \tag{9.56}$$

Nun setzten wir $\varphi(x)$ in $W_L(x)$ und $W_L(x)$ und $n(x)$ in der Gleichung I_n von Gl. (9.54) ein. (Beachte: I_n ist voraussetzungsgemäß in der Verarmungszone konstant, siehe (5.5) für $\partial/\partial t = 0$ und $g_n - r_n = 0$). Dies führt uns auf eine separierbare Differentialgleichung, welche wir im Bereich $-d_n \le x \le 0$ integrieren. Die Grenzwerte für W_{Fn} sind Abb. 9.15 zu entnehmen:

$$I_n \int_{-d_n}^{0} \exp\left[-\frac{\varphi(x)}{U_T}\right] dx = -A\mu_n N_L \exp\left(\frac{W_{\chi_1}}{kT}\right) \int_{-W_{\phi 1}}^{-W_{\phi 1}-eU} \exp\left(\frac{W_{Fn}}{kT}\right) dW_{Fn}. \tag{9.57}$$

Unter Verwenden der Näherung für $\varphi(x)$ von (9.56) und Beachten des Umstands, daß $d_n/L_{Dn} \gg 1$ gilt, erhält man für die rechte Seite

$$\begin{aligned} \int_{-d_n}^{0} \exp\left[-\tfrac{\varphi(x)}{U_T}\right] dx &= \tfrac{2L_{Dn}^2}{d_n}\left[\exp\left(\tfrac{d_n^2}{2L_{Dn}^2}\right) - 1\right] \approx \tfrac{2L_{Dn}^2}{d_n}\exp\left(\tfrac{d_n^2}{2L_{Dn}^2}\right) \\ &= \tfrac{2L_{Dn}^2}{d_n}\exp\left(-\tfrac{\varphi_H+U}{U_T}\right) \\ &= \tfrac{2L_{Dn}^2}{d_n}\exp\left(\tfrac{\phi_{\mathrm{MH}}^{(n)} - W_{\phi 1} + W_{\chi 1} - eU}{kT}\right). \end{aligned} \tag{9.58}$$

Für die letzte Umformung in (9.58) wurde (9.50) verwendet. Setzt man (9.58) in (9.57) ein, so erhält man schließlich

$$I_n = -\frac{A\mu_n kT d_n}{2L_{Dn}^2} N_L \exp\left(-\frac{\phi_{\mathrm{MH}}^{(n)}}{kT}\right)\exp\left(\frac{U}{U_T}\right)\left[\exp\left(-\frac{U}{U_T}\right) - 1\right]. \tag{9.59}$$

Ersetzt man $\mu_n kT$ durch eD_n, siehe (4.30), und setzt $n(0)$ aus (9.51) ein, so erhält man

$$I_n = \frac{AeD_n n(0) d_n}{2L_{Dn}^2}\left(e^{U/U_T} - 1\right). \tag{9.60}$$

Das ist der Beitrag der Elektronen zu (9.52).

Für den Beitrag der Löcher geht man analog vor. Statt (9.54) verwendet man nun

$$\begin{aligned} I_p &= -Ap(x)\mu_p \frac{\mathrm{d}W_{Fp}}{\mathrm{d}x}, \\ p(x) &= N_V \exp\left[\frac{W_V(x)-W_{Fp}(x)}{kT}\right], \\ W_V(x) &= -W_{\chi 1} - W_G - e\varphi(x). \end{aligned} \tag{9.61}$$

(9.56) wird unverändert übernommen. Statt (9.57) erhält man nun

$$I_p \int_{-d_n}^{0} \exp\left[\frac{\varphi(x)}{U_T}\right] dx = -A\mu_p N_V \exp\left(-\frac{W_{\chi 1} + W_G}{kT}\right) \times \int_{-W_{\phi 1}}^{-W_{\phi 1} - eU} \exp\left(-\frac{W_{Fp}}{kT}\right) dW_{Fp}. \tag{9.62}$$

Jetzt erhält man

$$\int_{-d_n}^{0} \exp\left[\frac{\varphi(x)}{U_T}\right] dx = -\frac{2L_{Dn}^2}{d_n}\left[\exp\left(-\frac{d_n^2}{2L_{Dn}^2}\right) - 1\right] \approx \frac{2L_{Dn}^2}{d_n}, \tag{9.63}$$

und

$$I_p = \frac{A\mu_p kT d_n}{2L_{Dn}^2} N_V \exp\left(\frac{W_{\phi 1} - W_{\chi 1} - W_G}{kT}\right)\left(e^{U/U_T} - 1\right) dx. \tag{9.64}$$

Man ersetzt $\mu_p kT$ durch eD_p, s. (**??**), und beachtet (siehe Abb. 9.15 und (3.26)), daß $N_V \exp(\ldots)$ gerade die Minoritätsträgerdichte p_n im n-Halbleiter bedeutet, so erhält man den Löcherbestandteil des Gesamtstroms

$$I_p = \frac{AeD_p p_n d_n}{2L_{Dn}^2}\left(e^{U/U_T} - 1\right). \tag{9.65}$$

9.2.3 Ohmscher Kontakt:

Ein ***ohmscher Kontakt*** ist ein Metall-Halbleiter Kontakt, welcher im Vergleich zum Bahnwiderstand des Halbleiters einen verschwindend kleinem Kontakt-Widerstand hat.

Ohmscher Kontakt durch kleine Austrittsarbeit

Der Kontakt-Widerstand ist definiert als

$$R_c = \left(\frac{\partial I}{\partial V}\right)^{-1}\Bigg|_{V=0} \tag{9.66}$$

Bei **kleinen Halbleiterdotierungen**, verschwindet der Widerstand, wenn der Strom zu gegebener Spannung groß werden kann. Damit der Strom nach Gl. (9.52) groß werden kann, muss das Kontaktmaterial so gewählt werden, dass die Austrittsarbeit klein ist. Allerdings ist es für einen gegebenen HL in der Regel schwierig, die entsprechenden Metalle mit kl. Austrittsarbeit zu finden.

Ohmscher Kontakt dank Tunnelstrom

Alternativ lassen sich die ME-HL Übergänge stark dotieren. **Bei starker Dotierung** dominiert im Metall-Halbleiterkontakt der Tunnelstrom. Wenn man nämlich den Halbleiter sehr stark n-dotiert, wird die Breite der Raumladungszone extrem klein: Dann können Elektronen an der LB-Kante im Halbleiter den Potentialberg mit hoher Wahrscheinlichkeit durchtunneln und der MS-Kontakt besitzt ohmsches Verhalten anstelle von Gleichrichtereigenschaften (viele "ohmsche" Kontakte arbeiten nach diesem Prinzip).

Reale ohmsche Kontakte, welche auf dem Tunneleffekt basieren, werden durch Kontaktieren mit Metallen, welche gleichzeitig als Dotanden wirken können hergestellt.

- Kontaktieren von **p-Si**: Kontaktmetall ist Al, welches durch Ein-Legieren eine gute Haftung und eine -dotierte Zwischenschicht, also sehr gute Leitfähigkeit ergibt. Al ist auch ein gutes Akzeptormaterial, wodurch der Kontakt sehr gut wird.

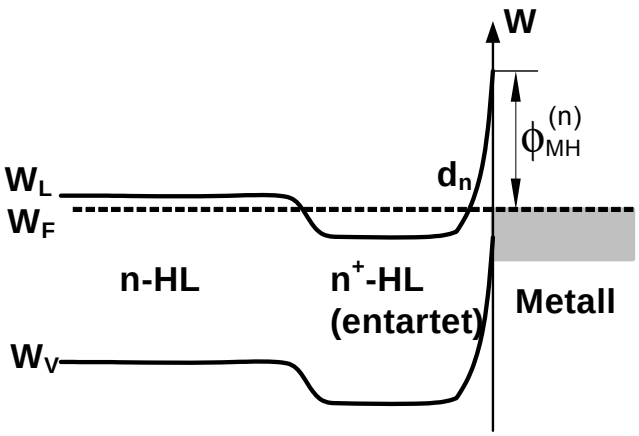


Abbildung 9.18: Bandschema eines Metall-HL mit ohmschem Kontakt, welcher aufgrund des Tunneleffekts ermöglicht wird.

- Kontaktieren von **n-Si**: Mit Al würde man auf n-Si nur eine Umdotierung erreichen. Man scheidet auf n-Si deshalb ein ganzes Schichtpaket ab: Al-TiN-Ti- -n-Si, wobei Ti als Diffusionsbarriere für das Al eingesetzt wird; der eigentliche Kontakt bildet wieder die Silizid-n-Si-Anordnung. Die Silizide (Ti, W, Mo, Pt, Ni) werden heute fast immer eingesetzt, da sie niederohmigere Kontakte als die elementaren Metalle erlauben.

Kapitel 10

Feldeffekttransistoren

Der Feldeffekttransistor (FET) wurde bereits vor Entdeckung des normalen Transistoreffektes vorgeschlagen [21] [22], jedoch erst gegen Ende der fünfziger Jahre zufriedenstellend realisiert.

Feldeffekttransistoren sind aktive Bauelemente, bei denen der Stromfluß durch einen leitenden Kanal mit Hilfe einer Steuerelektrode moduliert werden kann. In FETs fliesst ein Strom zwischen Source (S) und Drain (D), welcher über ein Potential an der sog. Gate-Elektrode (G) gesteuert wird, siehe Bild 10.1. Der Aufbau ist häufig symmetrisch - die Kennlinien ändern sich nicht, wenn die Rolle von Source aund Drain vertauscht wird. In der Praxis sollte man die Anschlüsse trotzdem nicht vertauschen, da die Kapaziäten zwischen Gate-Drain häufig geringer gehalten werden als zwischen Gate-Source.

Feldeffekttransistoren werden gelegentlich als Unipolartransistoren bezeichnet, da bei FETs im Gegensatz zu Bipolartransitsotren nur Ladungsträger einer Polarität zum Strom beitragen.

Der Kanalstrom wird über den effektiven Querschnitt des Kanalwiderstandes kontrolliert. Es gibt verschieden Möglichkeiten um diesen Kanal bzw. die Weite der Raumladungszone zu kontrollieren. Je nach Realisierung unterscheidet man zwischen

- dem ***MIS-FET*** *(Metal-Insulator-Semiconductor-FET)* 10.1(b) - die Anreicherungs- oder Verarmungszone wird moduliert. Ist der Isolator das natürliche Oxid SiO_2 so heißt der Typ ***MOSFET*** *(Metal-Oxide-Semiconductor FET)*
- einem ***Sperrschicht-FET*** oder ***JFET (Junction FET)*** 10.1(a) - die Raumladungszone eines pn-Überganges wird kontrolliert,

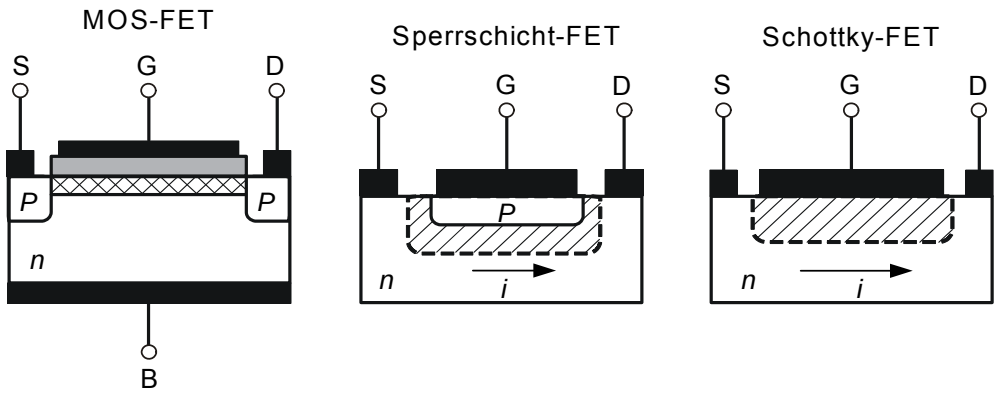


Abbildung 10.1: Schematische Darstellung der Wirkungsweise des Sperrschicht-FET, MOS-FET und Schottky-FET bzw MESFET.

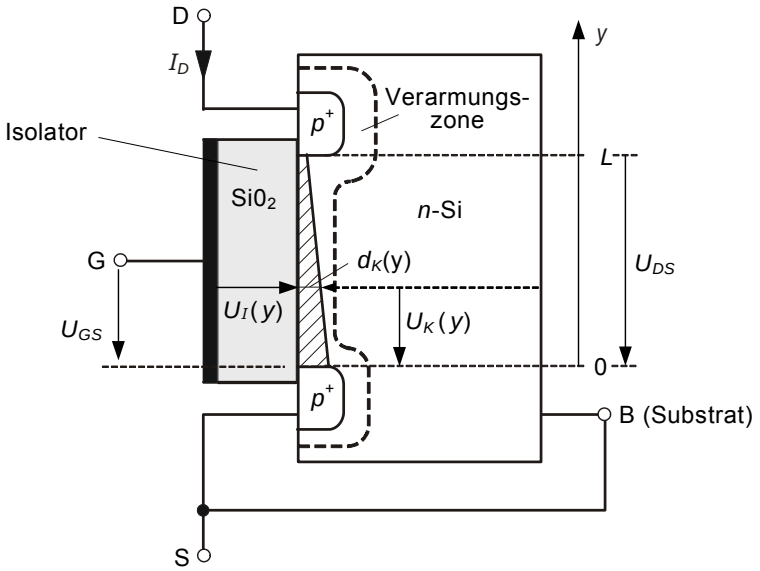


Abbildung 10.2: p-Kanal M0SFET. Die Breite des Kanals normal zur Zeichenfäche sei b; die Breite des Inversionskanals $d_K(y)$ ist nicht maßstäblich gezeichnet.

- dem ***MESFET*** (MEtal-Semiconductor-FET) 10.1(c) - die Raumladungszone eines Schottky-Überganges wird moduliert.
- dem ***HEMT*** (High-Electron Mobility Transistor) oder ***MODFET*** (Modulation Doped FET), in welchem die Beweglichkeit von Trägern in einem 2D-Elektronengas zur Steuerung genutzt werden.

10.1 Der MIS-Feldeffekttransistor (MOSFET)

Bild 10.2 zeigt eine MIS-Anordnung. Über den Drain(D) - Source(S) Kontakten ist eine negative Spannung $U_{DS} < 0$ angelegt. Zwischen Gate-Isolator (G) und Source ist ebenfalls eine negative Spannung angelegt $U_{GS} < 0$. Diese Spannung fällt zum einen über dem Isolator ab $U_I(y)$ und zum andern über dem Kanal zur Source Elektrode $U_K(y)$, Diese Spannung sind ortsabhängig und es gilt $U_{GS} = U_I(y) + U_K(y)$. Ein p-Inversionskanal wird erzeugt, falls über dem Isolator die Spannung $U_I(y) < U_{th}$ anliegt.

10.1.1 Kennlinienfeld des p-Kanal-MOSFET

Ohmscher Bereich

Zunächst berechnen wir die Stromstärke im ohmschen Bereich des MOSFET (d.h. der MOSFET ist weder gesperrt noch gesättigt) . Im ohmschen Bereich sollte die Stromstärke proportional zur Anzahl der vorhandenen Ladungsträger im Kanal sein. Die Anzahl der Ladungsträger im Kanal hängt von der Gate-Spannung U_{GS} ab, welche genügend negativ sein sollte, damit es zur Inversion kommt und ein Ladungsträgerkanal geöffnet wird. Ist der Kanal offen, erwartet man eine lineare Erhöhung des Stromes, je negativer die Drain-Source Spannung U_{DS} wird. Dieser Sachverhalt soll nun berechnet werden.

Der Strom I_D des MOSFET ist proportional zur Ladungsträgerdichte $\rho_K(y)$, dem Kanalquerschnitt $A(y)$ und der Geschwindigkeit v_p der Ladungsträger am Ort y entlang des Kanals. Der

Strom kann an irgend einer Stelle entlang y berechnet werden (siehe auch Bild 10.2)

$$I_D = -\rho_K(y) \cdot v_p \cdot A(y) \tag{10.1}$$
$$= -\rho_K(y) \cdot v_p \cdot [bd_K(y)] \tag{10.2}$$
$$= \frac{[-\sigma_K(y)Lb]}{Lbd_K(y)} \cdot \mu_p E \cdot [bd_K(y)] \tag{10.3}$$
$$= [-\sigma_K(y)b] \left[-\mu_p \frac{dU_K(y)}{dy}\right]. \tag{10.4}$$

Die Ladungsträgerdichte σ_K für starke Inversion wurde in Gl. (9.42) hergleitet. Mit der Notation aus Abb. 10.2 folgt für die bewegliche Ladungsdichte im Kanal an der Stelle y

$$\sigma_K(y) = \rho_K(y)d_K(y) = C'_I[U_{th} - U_I(y)] = C'_I\{U_{th} - [U_{GS} - U_K(y)]\}. \tag{10.5}$$

Wobei wir $U_I(y)$ soeben durch die Beziehung

$$U_{GS} = U_I(y) + U_K(y)$$

ersetzt haben.

Multiplikation mit dy und Integrieren über $0 \le y \le L$, $0 \le U_K(y) \le U_{DS}$ führt auf

$$I_D \int_0^L dy = \mu_p b C'_I \int_0^{U_{DS}} \{(U_{th} - U_{GS}) + U_K(y)\}\, dU_K(y)$$

mit dem Ergebnis

$$I_D = \frac{\mu_p b C'_I}{2L}[U_{DS}^2 - 2U_{DS}(U_{GS} - U_{th})], \qquad C_I = bLC'_I. \tag{10.6}$$

Damit wir die Kurve zeichnen können bringen wir sie auf eine quadratische Form

$$I_D = \frac{\mu_p b C'_I}{2L}[(U_{DS} - (U_{GS} - U_{th}))^2 - (U_{GS} - U_{th})^2] = \frac{\mu_p b C'_I}{2L}[(U_{DS} - U_{DS\,\mathrm{Sat}})^2 - U_{DS\,\mathrm{Sat}}^2], \tag{10.7}$$

wobei $U_{GS} - U_{th} = U_{DS\,\mathrm{Sat}}$ gesetzt wurde.

Gl. (10.6) gilt im sogenannten ohmschen Bereich, d. h. für

$$\begin{array}{l} U_{GS} \le U_{th}, \\ \sigma_K(L) \ge 0, \end{array} \qquad \text{d. h.} \qquad U_{DS} \ge U_{GS} - U_{th} = U_{DS\,\mathrm{Sat}}. \tag{10.8}$$

Sättigung

Für $U_{DS} < U_{DS\,\mathrm{Sat}}$ wird der Inversionskanal am Drainende abgeschnürt und der Drainstrom bleibt auf dem Wert, den er für $U_{DS\,\mathrm{Sat}}$ erreichte (siehe Gl. (10.7)). Im Abschnürbereich (Sättigungsbereich) gilt demzufolge

$$I_D = I_{D\,\mathrm{Sat}} = -\frac{\mu_p b C'_I}{2L} U_{DS\,\mathrm{Sat}}^2. \tag{10.9}$$

10.1.2 Der optimale MISFET

In der Praxis geht es darum die Transistoren auf einen Betrieb

- mit kleinen Leistungen
- mit hohen Grenzfrequenzen
- und kleinen Dimensionen

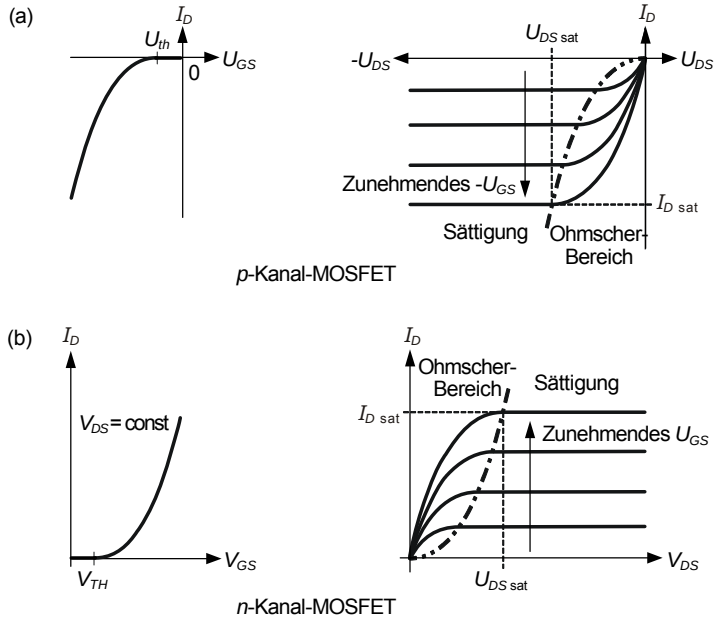


Abbildung 10.3: Kennlinienfelder selbstsperrender MISFETs: (a) p-Kanal, (b) n-Kanal.

zu optimieren [30].

Die benötigten Betriebsleistungen sind dann klein, wenn bei gegebener Spannung der Dranistrom maximal ist. Dann ist nämlich auch der Widerstannd mimimal. Gl. (10.9) zeigt, dass man einen großen Drainstrom erhält indem man

- Materialien mit großer Beweglichkeit wählt,
- eine großer Kapazität ($C_I' = \varepsilon_I / d_I$) d.h. großem ε_I und kleinem d_I wählt, sowie
- kleiner Kanallängen L wählt.

Die benötigte Betriebsleistung wird ferner miniert, wenn die über dem MOS-FET angelegte Spannung klein ist. Diese ist aber dann klein, wenn die über dem Isolator abfallende Spannung minimal ist. Die über dem Isolator abfallende Spannung ist $U_I = -E_I d_I$. Falls das für die Inversion benötigte elektrische Feld gerade E_H beträgt, so gilt für den Spannungsabfall über dem Isolator:

$$U_I = -E_I d_I = -\frac{\varepsilon_H}{\varepsilon_I} E_H d_I.$$

Damit ist dann klar, dass man zur Optimierung des Verlustabfalls über den Transistoren mit Vorteil einen Isolator mit möglichst großem ε_I und kleinem d_I wählen muss. Und in der Tat, die neusten MOS-FETs benötigen gerade noch 1,5 V zum Schalten.

Weiter unten werden wir sehen, dass man durch die geschickte Wahl einer kleinen Kanallänge L und eines Materials mit hoher Beweglichkeit μ auch die Grenzfrequenz maximieren kann. Interessanterweise geht die Erhöhung der Kapazität im MOS-FET nicht zu Lasten der Geschwindigkeit. Dies kann man damit erklären, dass bei einer Erhöhung der Kapazität wegen der Ladungsträgerzahlerhöhung pro Spannungseinheit auch der Widerstand sinkt. Der niedrige Widerstand bringt aber wieder Geschwindigkeitsvorteile, welche die durch die erhöhte Kapazität eingebrachten Nachteile gerade aufhebt. Wir werden die Laufzeit τ unten genauer anschauen. Hier sei lediglich darauf hingewiesen, dass in einem einfachen Modell des RC limitierten Schaltkreis (der Kanal ist der Widerstand und das Gate bildet mit dem Dielektrika die Kapazität) Folgendes gilt

Material	ε_H	Durchbruchspannung [MV/cm]	W_g [eV]
SiO_2	$3,9$	10	9
Al_2O_3	$10,1$	3	9,9
CeO_2	15	$0,8$	5,1
HfO_2	25	2	5,5
La_2O_3	20	2	5,5
Y_2O_3	12	4	5,5
ZrO_2	22	$1,5$	8
$SrTiO_3$	11	10	8

Tabelle 10.1: Verschiedene Materialien und deren Dielektrizitätskonstanten und Durchbruchspannungen.

$$\tau = RC_I = R\ bLC'_I = \frac{U_{DS\,\mathrm{Sat}}}{I_D} bLC'_I \overset{(10.9)}{=} -\frac{2L^2}{\mu_p U_{DS\,\mathrm{Sat}}}. \tag{10.10}$$

Zusammenfassend kann also gesagt werden, dass man im Bemühen um verlustärmere, kleinere und schnellere Transistoren deren Gate-Kapazität erhöht, d.h. ε_I möglichst groß wählt und d_I mimiert, die Kanallängen L reduziert und die Ladungsträgerbeweglichkeit durch Wahl der Träger und Materialien maximiert. In der Halbeiterphysik verwendet man übrigens statt des Symbols ε_I das Symbol k. Die aktuelle Forschung beschäftigt sich denn auch mit der Suche nach dem geeigneten "high-k dielectrica". Das ideale "high-k dielectrica" sollte groß sein und eine möglichst große Durchbruchspannung bieten, damit man d_I kleine wählen kann. Dieses Dielektrika war bis anhin SiO_2 mit einem Wert von $\varepsilon_I = 3.9$.und einer Durchbruchspannung von 10 MV/cm Typische Dicken der Oxidschicht betrugen eben noch 2 nm. Wegen der Gefahr von Leckströmen kann man aber auch nicht mehr viel dünnere Dielektrika verwenden. Da man nicht mehr viel dünnere Dielektrika anbringen kann, werden neuerdings andere Materialien mit höheren Dielektrizitätskonstatnen vewendet. Dem Verlauten nach werden seit 2007 bzw. 2008 bei Intel und IBM für die neuste Generation von Chips mit 45 nm Kanallängen Hafniumbasierte Gate-Dielektrika eingesetzt - wahrscheinlich Materialien wie HfSiON mit einem $\varepsilon_I = 11$.

Die Entwicklung der Gate-Oxid Schichtdicken über die Jahre sind in Fig. 10.4 aufgezeichnet. [31], [32].

10.1.3 Klassifikation der MOSFET

Bild 10.5(c) zeigt die Kennlinienfelder des hier analysierten selbstsperrenden p-Kanal MOSFET und in (a) das entsprechende Kennlinienfeld des selbstsperrenden n-Kanal MOSFET. Insgesamt gibt es vier Typen von MOSFETs, siehe Bild (10.5) n- oder p- Kanal und selbstleitend (=Normally On=Verarmungstyp (Depletion Type)) oder selbstsperrende (=Normaly Off=Anreicherungstyp (Enhancement Type)) MOSFETs.

10.1.4 Kleinsignal-Modell des MOSFET und Grenzfrequenzen

Wir interessieren uns für die Kleinsignaländerung i_D des Drainstromes I_D unter einer Kleinsignalmodulation u_{GS} an der Gatespannung U_{GS}. Die Steilheit g_m des Transistors am Drain-Ausgang erhält man durch Ableiten des Drainstromes in Gl. (10.6) nach U_{GS}

$$\begin{aligned} i_D \quad &= \quad g_m u_{GS} = \frac{\partial I_D}{\partial U_{GS}} u_{GS} = -\frac{\mu_p b C'_I}{L} U_{DS} u_{GS}, \\ \text{d.h.} \quad &g_m = \frac{\partial I_D}{\partial U_{GS}} = \frac{\partial(\frac{\mu_p b C'_I}{2L}[U_{DS}^2 - 2U_{DS}(U_{GS}-U_{th})])}{\partial U_{GS}} = -\frac{\mu_p b C'_I}{L} U_{DS}\ . \end{aligned} \tag{10.11}$$

Ein einfaches Ersatzschaltbild für Kleinsignalgrößen des MIS-Feldeffekttransistors ist in Abb. 10.6 gezeichnet. Es berücksichtigt

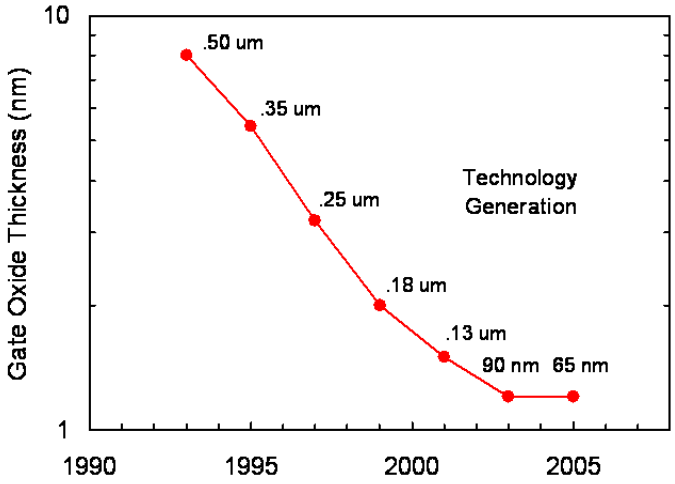


Abbildung 10.4: Entwicklung der Gate-Oxid Dicken über die Jahre. Gate-Oxid Dicken von 2 nm sind heute machbar. [31].

- das Niederfrequenzverhalten unter einer Kleinsignalmodulation u_{GS} über den Transimpedanzterm $g_m u_{GS}$
- Die kapazitiven Kopplungen zwischen Gate und Source bei höheren Frequenzen über die Kapazität C_{GS} und
- die kapazitive Kopplungen zwischen Gate und Drain über der Kapazität C_{GD}

Aus dem Primitiv-Ersatzschaltbild lässt sich die Grenzfrequenz f_g ($\omega_g = 2\pi f_g$) ableiten. Es ist jene Frequenz, bei welcher der Betrag der Kurzschlußstromverstärkung auf den Wert 1 abgesunken ist. Man erhält mit $g_m$aus (10.11)

$$\left|\frac{i_D}{i_G}\right| = 1 = \left|\frac{g_m u_{GS} + \mathrm{j}\,\omega_g C_{GD}\, u_{GD}}{\mathrm{j}\,\omega_g (C_{GS} u_{GS} - C_{GD} u_{GD})}\right| \overset{u_{GD}=-u_{GS}}{=} \left|\frac{g_m u_{GS} - \mathrm{j}\,\omega_g C_{GD}\, u_{GS}}{\mathrm{j}\,\omega_g (C_{GS}+C_{GD}) u_{GS}}\right| \approx \frac{g_m}{\omega_g C_I}, \tag{10.12}$$

$$\omega_g = \frac{g_m}{C_I} \overset{C_I = bLC_I'}{=} -\frac{\mu_p}{L^2} \cdot U_{DS} = \frac{v_{p\,\mathrm{eff}}}{L} = \frac{1}{\tau}.$$

Man findet also Limitierungen der Grenzfrequenz aufgrund der Beweglichkeiten und der Kanallänge:

- Sowohl g_m als auch f_g sind proportional zur Beweglichkeit der Träger im Inversionskanal (hier: μ_p). Da bei Silizium $\mu_n \approx 3\mu_p$, sind n-Kanal-MOSFET (metal-oxide-semiconductor) günstiger als p-Kanal-MOSFET. Sowohl die Dotierung als auch die Grenzflächenstreuung an der Grenze Halbleiter-Isolator (Störung des Halbleitergitters!) reduziert die Beweglichkeit und damit f_g, g_m. Beim HEMT (high electron mobility transistor) vermeidet man beides, indem man den Kanal durch einen undotierten Quantenfilm in einer gitterangepaßten Heterostruktur realisiert. Zum Vergleich verschiedener Strukturen wird die Steilheit bezogen auf die Kanalbreite angegeben (g_m/b in mS/mm). Man beachte, daß $g_m \sim b/L$ ist.
- Die Grenzfrequenz ω_g ist bei normalen Feldstärken umgekehrt proportional zu L^2. Bei hohen Feldstärken ist ω_g noch $\sim 1/L^n$ mit $1 \leq n \leq 2$. Dies kann man wie folgt sehen: Da $|\mu_p U_{DS}/L|$ die mittlere Driftgeschwindigkeit $v_{p\,\mathrm{eff}}$ der Löcher im Inversionskanal ist, ist $\omega_g = v_{p\,\mathrm{eff}}/L = 1/\tau$, wobei τ die Laufzeit der Träger im Kanal bedeutet. Wenn bei hohen Feldstärken eine Sättigung der Driftgeschwindigkeit einsetzt, ist demnach $\omega_g \sim 1/L$ statt

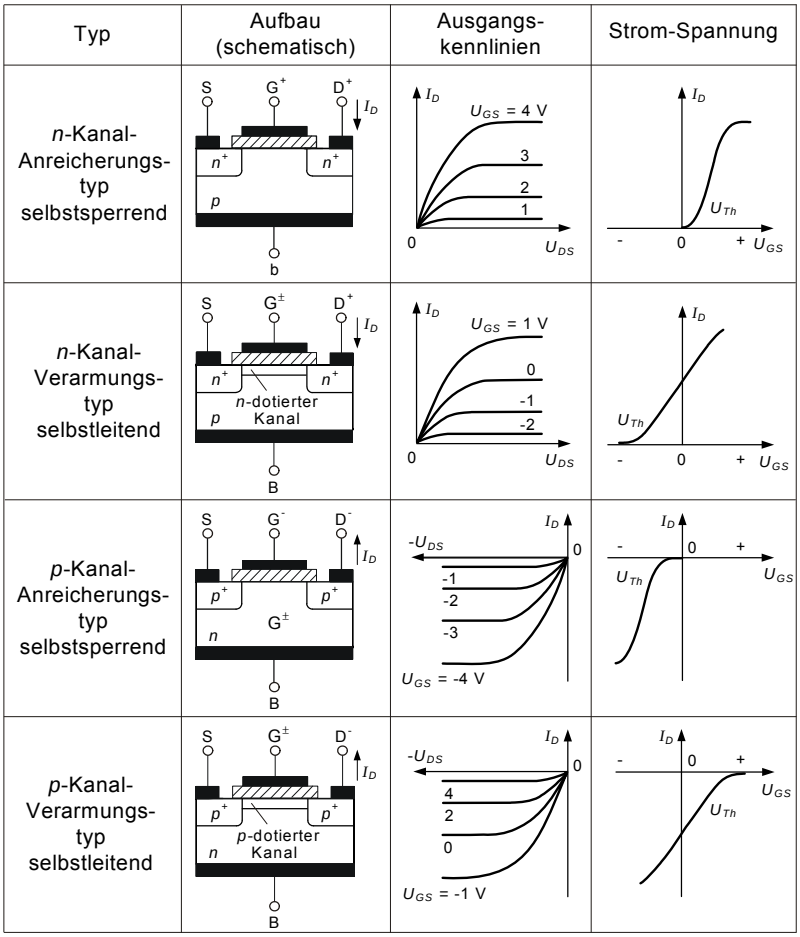


Abbildung 10.5: Bezeichnung und schematischer Aufbau verschiedener MOSFET-Typen.

$\omega_g \sim 1/L^2$. Tatsächlich ändert sich die Feldstärke längs des Kanals und somit ist $\omega_g \sim 1/L^n$ mit $1 \leq n \leq 2$ zu erwarten.

10.2 Sperrschicht-FET (Junction-FET,JFET)

Beim ***Sperrschicht-FET (Junction-FET, JFET)*** wird die Raumladungszone von einer oder mehreren pn-Dioden durch Anlegen einer Sperrspannung moduliert. Dies wird dann benutzt um die Weite eines Stromkanals zu kontrollieren und so einen neuen Typ von Feldeffekt-Transistor zu bauen. Auch der JFET ist ein unipolar Transistor.

Bild 10.7 zeigt schematisch einen n-Kanal-Sperrschicht-FET. Der Transistor besteht aus zwei p^+n-Übergängen mit gemeinsamer n-Zone. Durch Anlegen einer Sperrspannung können die Weiten der Raumladungszonen, welche sich im wesentlichen in das n-Gebiet ausbreiten, gesteuert werden.

Mit zunehmender Spannung $|U_{DS}|$ nimmt der Kanalstrom I zunächst linear zu, Fig. 10.7(b). Der Stromfluss ist nur durch den Kanalwiederstand R begrenzt, dh. $I_D = U_{DS}/R$.

Natürlich liegt rechts (bei D) ein größeres Potential gegenüber dem Gate als links (in der Nähe von S) so dass die die relative Sperrspannung vom n-Kanal gegenüber dem Gate-Kontakt von links nach rechts ansteigt. Bei weiter ansteigendem $|U_{DS}|$ wird die Drain-Seite gegenüber der Gate-Diode zunehmend gesperrt. Der Kanal schnürt sich damit in Richtung der Drain-Elektrode ab. Der Gesamtleitwert zwischen Source und Drain nimmt daher mit zunehmender Spannung

$|U_{DS}|$ ab, da der Kanal (zumindestens teilweise) schmaler wird. Erreicht die Spannung den **Sättigungswert** U_{DSat} (andere sprechen von **Einsatzspannung, Schwellenspannung** oder **Pinch-Off-Spannung** $U_E = U_{th} = U_P$) schliesst sich der leitfähige Kanal, Fig. 10.7(c). Es fliesst ein Sättigungsstrom I_{Dsat}, welcher dem Widerstandswert R_{Sat} des Leitkanals entspricht, so dass gilt $U_{DSat} = R_{Sat} I_{Dsat}$.

Wird die Spannung weiter erhöht, so vergrößert sich die Raumladungszone weiter und der Strom steigt trotz steigender Spannung nicht weiter an, Fig. 10.7(d). Dies kann wie folgt begründet werden: Der Punkt **P**, bei welchem der Kanal den Sättigunswiderstand R_{Sat} hat verschiebt sich nach links. Diesem Pinch-off Punkt entspricht zwingendermassen die Spannung U_{DSat} und der zugehörige Strom ist I_{Dsat}.Wenn aber im ersten Teil des Kanals der Strom I_{Dsat} fliesst, so fliesst im ganzen Kanal nach wie vor der Strom I_{Sat}.

Wenn man die Gate-Spannung verringert, erhöht sich der Kanalwiderstand entlang des ganzen Source-Drain Kanals und es kommt etwas früher zur Sättigung. Der Sättigungsstrom ist dann auch kleiner, Fig. 10.7(e).

Damit ergibt sich insgesamt das in Bild 10.8 qualitativ gezeichnete Kennlinienfeld des n-Kanal-Sperrschicht-FET

Wir verzichten hier auf eine genaue Analyse, da diese analog zum später folgenden MESFET geschehen kann und dort ausgeführt wird.

10.3 Der Metall-Halbleiter-FET (MESFET)

Der ***MESFET (metal-semiconductor FET)***, s. Bild. 10.9, ist eng verwandt mit dem JFET. Der Unterschied liegt in der Verwendung einer Metall-Halbleiter Sperrdiode statt der pn-Sperrdiode um die Kanalbreite zu modulieren. Die Berechnung der Kennlinien für den MESFET und J-FET sind denn auch identisch. Zu beachten wäre noch, daß es beim J-FET und beim MESFET keine galvanische Trennung zwischen Gate und Kanal gibt, wie das beim MISFET der Fall war.

10.3.1 Struktur und Arbeitsprinzip

Der MESFET wird typischerweise auf SI-GaAs (**S**emi-**I**solierendem GaAs) aufgebaut. An der Oberfläche wird ein n-leitender Kanal ($\mu_n = 8\,500\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$, $\mu_p = 400\,\mathrm{cm^2\,V^{-1}\,s^{-1}}$) der Dicke a (typisch $0{,}1\,\mu$m) mit einer Dotierung von etwa $n_D = 10^{17}$ cm^{-3} erzeugt. Zwischen der Gate-(Steuer-)-Elektrode und dem Halbleiter herrschen die in Abb. 9.15 dargestellten Verhältnisse an einem MS-Kontakt, mit dem Unterschied, daß in Abb. 10.9 U_{GS} und damit $U(y) < 0$ gewählt wird (d.h. es wird eine Sperrspannung angelegt, so dass zwischen Gate und Kanal kein Strom fließt). Durch die Verbreiterung der Verarmungszone der Breite $d_n(y)$ wird der Kanal abgeschnürt. Der Kanal selbst ist somit nicht direkt an der Grenze zwischen Metall und Halbleiter, sondern im Volumen. (Anstelle des MS-Kontaktes könnte das Gate prinzipiell auch ein p$^+$-dotierter Halbleiter sein, dann erhielte man den oben bereits besprochenen, allgemein mit Silizium realisierten, JFET.)

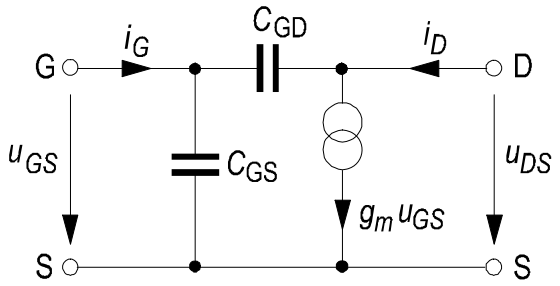


Abbildung 10.6: Kleinsignalersatzschaltbild des MIS-Feldeffekttransistors.

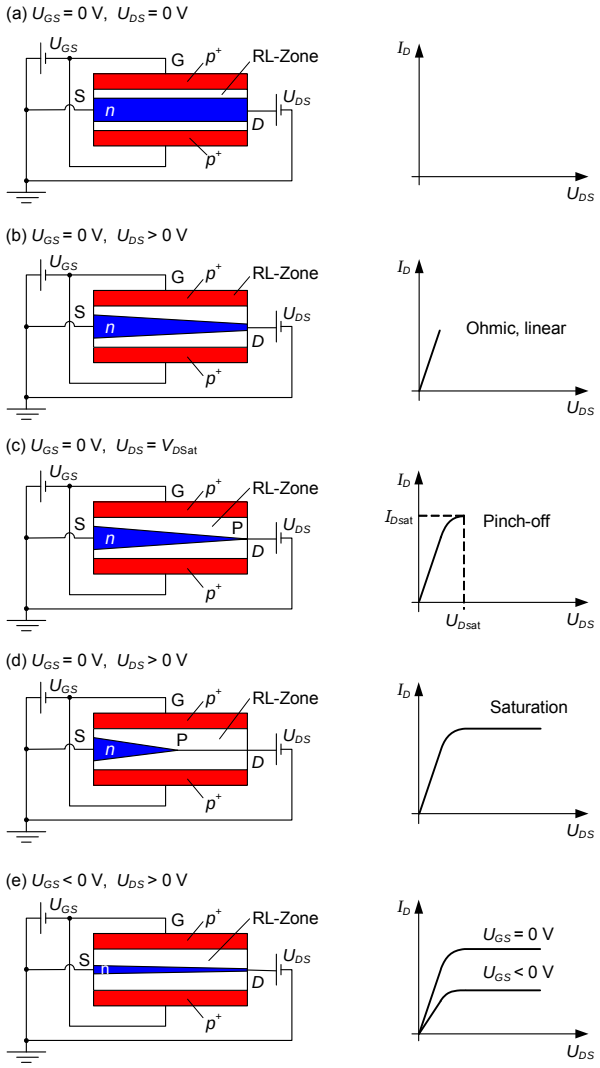


Abbildung 10.7: Schematische Darstellung des Sperrschicht-FET und Veränderung der Sperrschicht (weiss) unter Anlegen einer Drain- bzw. Gate-Spannung, [?].

10.3.2 Kennlinien des MESFET (und JFET)

Vereinfachungen: Für die Berechnung des Drain-Stromes wird angenommen, daß elektrische Felder in der Verarmungszone nur eine x-Komponente und im Kanal nur eine y-Komponente besitzen. Der Gatestrom I_G wird vernachlässigt; der Strom im Kanal ist somit konstant und gleich dem Drainstrom I_D. Für Spannungen $U_{DS} > 0$ wird zufolge des Spannungsabfalls $U_K(y)$ längs des Kanals die Verarmungszone gegen das Drainende des Kanals breiter.

Ohmscher Bereich

Der Betriebsbereich, in dem der Kanal nicht abgeschnürt ist, wird als ohmscher Bereich bezeichnet. In ihm gilt

$$\begin{array}{rcll} U(L) & = & U_{GS} - U_{DS} \geq U_P, & \\ U_{DS} & \leq & U_{DS\,\mathrm{Sat}} = U_{GS} - U_P, & \text{ohmscher Bereich.} \end{array} \tag{10.13}$$

Die Kennlinie $I_D\,(U_{GS}, U_{DS})$ des MESFET folgt aus Anwendung des ohmschen Gesetzes auf

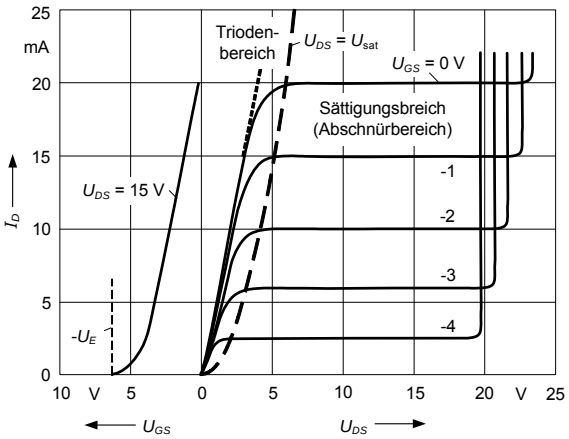


Abbildung 10.8: Kennlinienfeld eines Sperrschicht-FET [3].

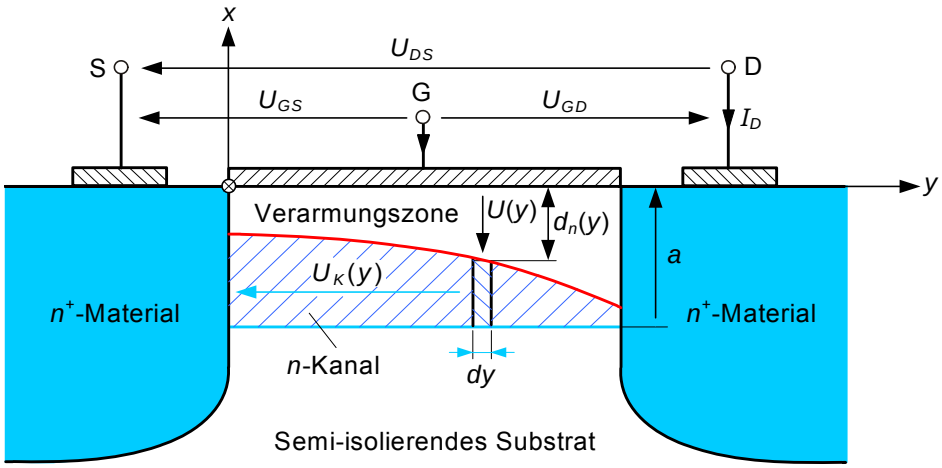


Abbildung 10.9: Prinzip des MESFET. $U(y)$ ist der Anteil der äußeren Spannung, der über der Verarmungszone in x-Richtung abfällt. $U_K(y)$ ist die Spannung längs des Kanals in y-Richtung. Die Leitfähigkeit der n-Zonen sei unendlich groß . Es gilt somit $U_{GS} = U(y) + U_K(y)$.

eine Schicht der Dicke dy im Kanal (siehe Abb. 9.15); wobei b die Ausdehnung des Kanals in z-Richtung bedeutet, erhält man

$$dU_K(y) = \frac{dy}{\sigma_n A}\, I_D = \frac{dy}{\sigma_n b[a - d_n(y)]} I_D = \frac{I_D\, dy}{L} \cdot \frac{1}{G_0\left[1 - \dfrac{d_n(y)}{a}\right]}. \tag{10.14}$$

mit der Abkürzung

$$G_0 = \frac{ab}{L}\sigma_n, \qquad \sigma_n = e n_D \mu_n\,. \tag{10.15}$$

Die Breite der Verarmungszone d_n variiert von links nach rechts. Ein Ausdruck für d_n wurde im Kapitel zur Schottky-Diode (siehe Gl. (9.52)) hergeleitet

$$d_n(y) = L_{Dn}\sqrt{-\frac{2[\varphi_H + U(y)]}{U_T}} = L_{Dn}\sqrt{\frac{2[U_D - U(y)]}{U_T}}. \tag{10.16}$$

Das Kontaktpotential $\varphi_H < 0$ für einen Metall-Halbleiter-Kontakt im Gleichgewicht wurde bereits in Abb. 9.14 und (9.50) definiert. Wie bereits früher erläutert, kann man das Kontaktpotential beim MS-Kontakt in Analogie zur Diffusionsspannung U_D beim pn-Kontakt interpretieren. Weil die hier hergeleiteten Formeln sowohl für den JFET als auch den MSFET gütlig sind macht es Sinn nun $\varphi_H = -U_D$, zu setzten. φ_H bzw. U_D können aus Materialtabellen entnommen werden. Sie sind definiert zu

$$-\varphi_H = -\frac{1}{e}\left[(W_{\phi 1} - W_{\chi 1}) - \phi_{\mathrm{MH}}^{(n)}\right] \equiv U_D > 0, \tag{10.17}$$

Bei Erhöhen von U_{DS} wird der Kanal am Drainende abgeschnürt, $d_n(L) = a$. Die dann über der Verarmungszone abfallende äußere Spannung $U(L)$ wird als Abschnürspannung U_P bezeichnet ($U_P < 0$, pinch-off-Spannung).

$$a = d_n(L) = L_{Dn}\sqrt{-\frac{2[-U_D + U(L)]}{U_T}} = L_{Dn}\sqrt{\frac{2[U_D - U_P]}{U_T}}. \tag{10.18}$$

Dieser Ausdruck lässt sich übrigens nach $U_D - U_P$ auflösen. Aus (10.18) folgt mit (5.63)

$$U_D - U_P = \frac{U_T}{2}\left(\frac{a}{L_{Dn}}\right)^2 = \frac{a^2 e n_D}{2\varepsilon_H}, \qquad U_P < 0 \tag{10.19}$$

Das Verhältnis $d_n(y)/a$ wird in Gl. (10.14) benötigt. Es ergibt sich aus dem Verhältnis von (10.16) und (10.18) zu

$$\frac{d_n(y)}{a} = \sqrt{\frac{U_D - U(y)}{U_D - U_P}}. \tag{10.20}$$

Da Gl. (10.14) als Gl. mit dem Spannungabfall $U_K(y)$ über dem Kanal aufgeschrieben wurde, macht es Sinn $U(y)$ durch $U_K(y)$ zu ersetzen. Aus der Abb. 9.15 folgt wegen der Gültigkeit der Kirchhofschen Spannungsregel

$$U_{GS} = U(y) + U_K(y). \tag{10.21}$$

Einsetzten von Gl. (10.20) und (10.21) in (10.14) ergibt nun eine separierbare Differentialgleichung

$$dU_K(y) = \frac{I_D}{L} \cdot \frac{1}{G_0\left[1 - \sqrt{\frac{U_D - [U_{GS} - U_K(y)]}{U_D - U_P}}\right]} dy\,. \tag{10.22}$$

Integriert man diese über dem Intervall $0 \le y \le L$, $0 \le U_K(y) \le U_{DS}$, so erhält man das Ergebnis

$$\begin{aligned} I_D &= G_0 \int_0^{U_{DS}} \left[1 - \sqrt{\frac{U_D - [U_{GS} - U_K(y)]}{U_D - U_P}}\right] dU_K(y) \\ &= G_0\left[U_{DS} - \frac{2}{3}\frac{(U_D - U_{GS} + U_{DS})^{3/2}}{(U_D - U_P)^{1/2}} + \frac{2}{3}\frac{(U_D - U_{GS})^{3/2}}{(U_D - U_P)^{1/2}}\right]. \end{aligned} \tag{10.23}$$

Wie anschließend gezeigt wird, kann für (10.23) eine Näherung der Form

$$\begin{aligned} I_D &= \frac{G_0}{4(U_D - U_P)}\left[2U_{DS}(U_{GS} - U_P) - U_{DS}^2\right] \\ I_D &= \frac{\mu_n}{2L^2} \cdot \frac{\varepsilon_H b L}{a}\left[2U_{DS}(U_{GS} - U_P) - U_{DS}^2\right] \end{aligned} \tag{10.24}$$

gefunden werden. Dazu mussten wir G_0 durch Gl. (10.15) ersetzten und für $U_D - U_P$ wurde der Ausdruck in Gl. (10.19) verwendet.

Ein Vergleich vom Drainstrom für den MISFET in Gl. (10.6) zeigt, dass die beiden Gleichungen identische Form haben. Es sind nur folgende Ersetzungen vorzunehmen:

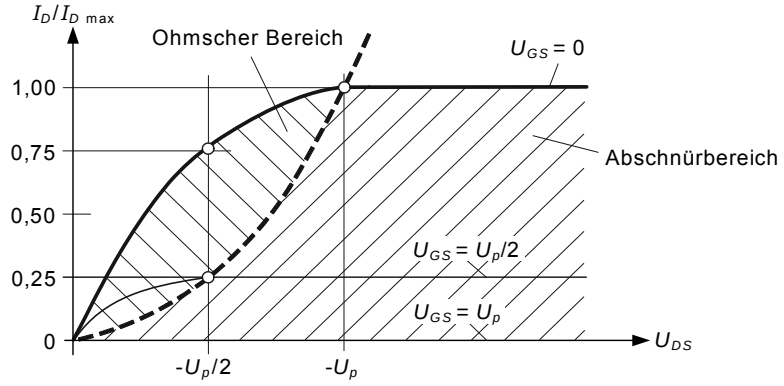


Abbildung 10.10: Ausgangskennlinienfeld des MESFET (schematisch).

- die Schwellspannung ist durch die Pinch-off Spannung zu ersetzten $U_{th} \longrightarrow U_P$,
- die Isolatorkapazität ist durch die Kapazität des in seiner vollen Breite verarmten Kanals zu ersetzten, d.h. $C_I = bL\,\varepsilon_I/d_I \longrightarrow C = bL\varepsilon_H/a$

Damit können auch die Beziehungen für die Steilheit und für die Grenzfrequenz, (10.11) und (10.12) sinngemäß übernommen werden, also z. B. für die Steilheit

$$g_m = \frac{\partial I_D}{\partial U_{GS}} = \frac{\mu_n b}{L} \cdot \frac{\varepsilon_H}{a} U_{DS}. \tag{10.25}$$

Typische Werte für g_m/b sind beim GaAs-MESFET in der Größenordnung von 300 mS/mm.

Abschnürrbereich (Sättigungsbereich)

Für $U_{DS} \geq U_{DS\,\mathrm{Sat}}$ steigt der Drainstrom nur mehr schwach an (siehe auch Erläuterung zum Junction-Feldeffekttransistor). Im Abschnürbereich (Sättigungsbereich) gilt daher annähernd

$$I_D = I_D(U_{DS\,\mathrm{Sat}}) = \text{const} \quad \text{für} \quad U_{DS} \geq U_{DS\,\mathrm{Sat}} = U_{GS} - U_P. \tag{10.26}$$

Kennlinienfeld MESFET

Der oben ausführlich diskutierte MOSFET war selbstsperrend, der hier ausführlich besprochene MESFET ist selbstleitend.

Abb. 10.10 zeigt schematisch die Ausgangskennlinien eines FET. Der maximale Drainstrom ergibt sich im vorliegenden Fall aus $U_{GS} = 0$ (die Gate-Kanal-Diode darf nicht in den Flußbereich kommen) und $U_{DS} = -U_P$ (der Kanal wird gerade am Drainende abgeschnürt).

Beweis von (10.24)

Zuerst wird (10.23) in folgender Form geschrieben:

$$I_D = G_0 \left\{ U_{DS} - \frac{2(U_D - U_P)}{3} \left[\left(1 + \frac{U_{DS} + U_P - U_{GS}}{U_D - U_P} \right)^{3/2} - \left(1 + \frac{U_P - U_{GS}}{U_D - U_P} \right)^{3/2} \right] \right\}. \tag{10.27}$$

Die Ausdrücke der Form $(1+x)^{3/2}$ werden bis einschließlich quadratischer Terme entwickelt

$$(1+x)^{3/2} \approx 1 + \frac{3}{2}x + \frac{3}{8}x^2. \tag{10.28}$$

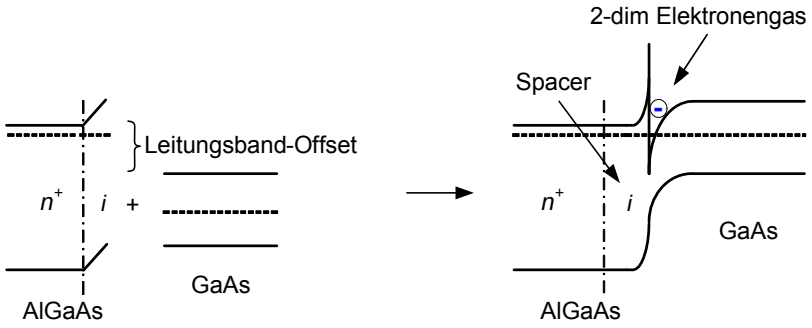


Abbildung 10.11: Bildung eines 2-Dimensionalen Kanals für das Elektronengas beim Zusammenfügen von AlGaAs und GaAs.

Nach elementaren Umformungen folgt daraus (10.24).

10.4 Der High-Electron-Mobility-Transistor (HEMT)

Der ***HEMT (high electron mobility transistor)*** nutzt die hohe Beweglichkeit von Trägern in einem 2D-Elektronengas aus, welches sich in einem undotierten Halbleiter bewegt (daher keine Streuung an Störstellen). Die Träger werden von einem benachbarten dotierten Halbleiter geliefert.

1978 entdeckten Forscher bei den Bell Labs, dass es an Heterostrukturübergängen von AlGaAs zu GaAs bis anhin unerreicht große Elektronenbeweglichkeiten festgestellt werden konnten. Diese Entdeckung führte dazu, dass Forscher in der ganzen Welt sich daran machten diesen Effekt für den Bau von noch schnelleren Transistoren zu nutzen. 1980 publizierten dann Forscher von Fujitsu ein erstes Bauteil, welches diesen Effekt benutzte. Sie nannten es **HEMT** (high electron mobility transistor). Die Forscher bei den Bell Labs, welche simultan an der Entwicklung von schnellern Bauteilen forschten nannten ihr Bauteil **SDHT** (Selectively Doped Heterostructure Transistor). Zahlreiche andere Gruppen kamen ebenfalls mit Variationen von Bauteilen, welche die schnellen Elektronebeweglichkeiten an Heterostrukturübergängen nutzen und gaben diesen Bauteilen eigene Namen, so werden z.B. auch Bezeichnungen wie **MODFET** (Modulation Doped FET, University of Illinois) oder **TEGFET** (Two-Dimensional Electron Gas FET, Thomson) verwendet.

Die Schwierigkeit beim Herstellen von schnellen Transistoren ist, dass man einerseits eine große Kanalleitfähigkeit möchte, was nach einer hohen Dotierung verlangt (Fig. 4.1), aber andererseits sich wegen der Geschwindigkeit eine hohe Beweglichkeit wünscht, was eine kleine Dotierung nahelegen würde (Fig. 4.4).

Beim HEMT bringt man nun zwei Materialien mit unterschiedlichem Bandgap - aber möglichst gitterangepasst - so zusammen, dass sich am Interface zu den beiden Materialien ein Bandkantenkanal bildet, welcher Elektronen führen kann. Entscheiden ist dabei, dass der Leitungsband-Offset groß ist, damit es zu einer anständigen Verformung und der Bildung eines tiefen Kanals mit großer Ladungsträgerkonzentration kommt. Damit die Beweglichkeit der Elektronen nun groß bleibt, wird die Schicht, mit dem Kanal nicht dotiert. Um trotzdem auf genügend Ladungsträger zurückgreifen zu können, wird die Schicht mit der großen Bandkante stark dotiert. Durch Anlegen einer entsprechenden Gate-Spannung und Verformung der Leitungsbandkante, können die Ladungsträger dann in den 2-Dimensionalen Elektronenkanal gestoßen werden. Um weitere Störungen aufgrund von Dotierungen in der Nähe des Kanals zu vermeiden, wird die Dotierung bereits vor dem Heterostrukturübergang reduziert. Man nennt diese undotierte Zone "Spacer-Layer".

Mit HEMTs (die Untersuchung spezieller Materialsysteme und die Optimierung der Parameter ist Gegenstand aktueller Forschung) sind Steilheiten g_m/b bis über 1 000 mS/mm erreichbar, die Drainströme haben Werte bis $I_D/b = 800\,\mathrm{mA/mm}$. Typische Flächendichten des 2D-Elektronengases liegen bei $N_{2D} \geq 10^{12}\,\mathrm{cm}^{-2}$.

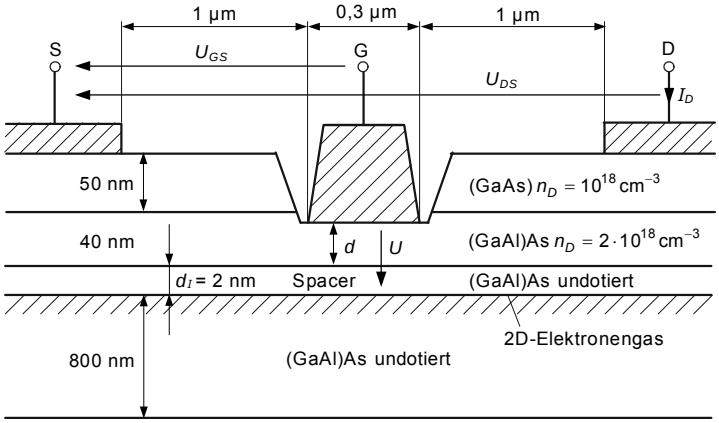


Abbildung 10.12: Beispiel für einen HEMT (nicht maßstäblich). U ist die zwischen Gate G und Kanal der 2D-Elektronen wirksame äußere Spannung.

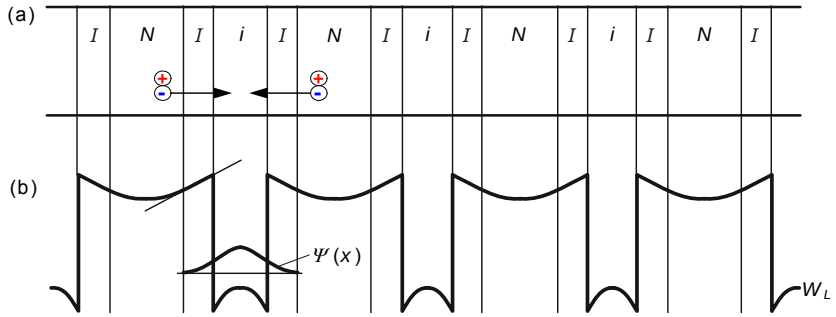


Abbildung 10.13: Modulationsdotierung. (a) Schichtenfolge. Elektronen gehen vom N-Halbleiter in den i-Halbleiter und bilden dort ein 2D-Elektronengas. (b) Schematischer Verlauf der Leitungsbandkante. Die Wahrscheinlichkeitsamplitude $\psi(x)$ der Elektronen im Quantenfilm reicht nicht bis in das dotierte Gebiet. Die I-Zonen werden als "spacer" (Abstandhalter) bezeichnet.

Bsp.: GaAs und AlGaAs haben fast identische Gitterperiodizität (siehe Fig. 7.26) und können deshalb ohne Probleme stabile Schichtfolgen bilden. Beim Zusammenfügen der beiden Materialien entsteht ein schmaler Kanal, welcher das zweidimensionale Elektronengas führen kann.

Abb. 10.12 zeigt ein Beispiel für einen HEMT . Durch negative Spannungen zwischen Gate und Kanal, $U < 0$, wird die bewegliche Ladung aus der Schicht der Dicke d ausgeräumt. Das Kennlinienfeld des HEMT ist identisch mit der des MOSFET im Verarmungsbetrieb.

Als "modulation doping" wurde ursprünglich die Erzeugung periodischer Übergitter bezeichnet, in denen sich dotierte und undotierte Bereiche abwechseln (siehe Abb. 10.13). Die Bauteile werden dann als MODFET bezeichnet. Natürlich sind das auch HEMT Bauteile.

10.5 Schlussbemerkungen

10.5.1 CMOS

Der ***CMOS*** (=complementary MOS FET) ist ein erstes Beispiel eines integrierten Schaltkreises (ICs). Der COMS besteht aus einem ***NMOS*** (n-Kanal MOS FET) und einem ***PMOS*** (p-Kanal MOS FET). Im Bild ist ein invertierender CMOS aufgezeichnet. Der obere Schaltkreis ist an

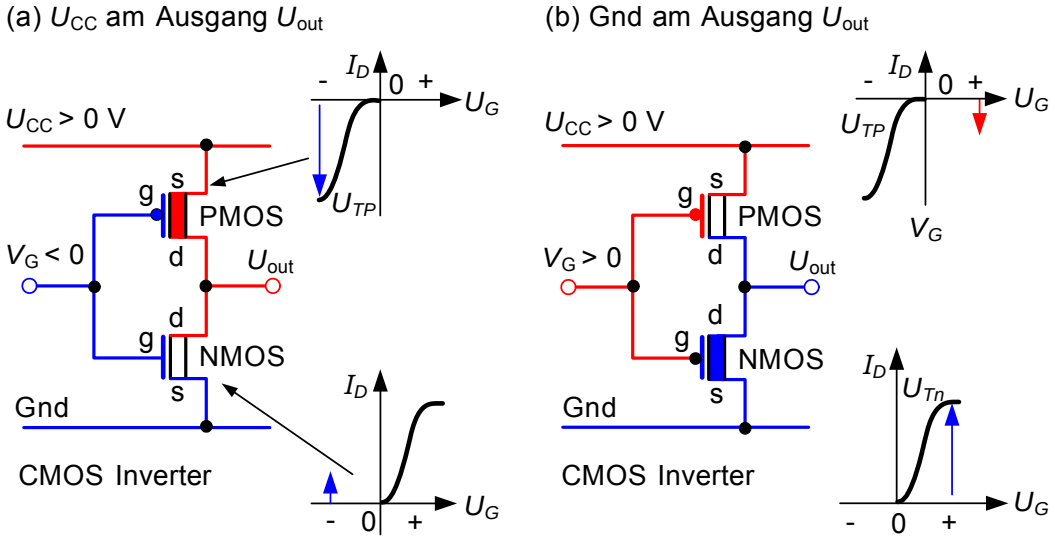


Abbildung 10.14: Invertierender CMOS IC mit (a) kleiner Gate-Spannung, so dass PMOS den U_{out}- Ausgang auf U_{cc} setzt. (b) großer Gate-Spannung, so dass NMOS den U_{out}- Ausgang auf *Ground* setzt.

eine DC-Spannung (U_{cc} Logische "1") und der untere Schaltkreis ist mit der Erdung verbunden (Logische "0").

Wird das Gate mit einer Spannung verbunden, welche unter der Schwellspannung U_{Tp} des PMOS liegt, so schaltet der PMOS durch und die Spannung am Ausgang U_{out} ist U_{cc}. Falls am Gate eine Spannung liegt, welche über der Schwellspannung U_{Tn} des NMOS liegt, so wird der Ausgang auf Ground geschaltet.

10.5.2 Grenzfrequenzen von Transistoren

Die Grenzfrequenz f_g des MIS-Feldeffekttransistors (aber auch des JFET oder MESFET) ist nach (10.12)

- proportional zur Beweglichkeit der Träger im Kanal und
- umgekehrt proportional zur Kanallänge.

Alle Maßnahmen, welche die Trägerbeweglichkeit erhöhen, erhöhen sowohl die Grenzfrequenz f_g als auch die Steilheit g_m des Transistors. Solche Maßnahmen können sein:

- Verlegung des Kanals von der Grenzfläche zwischen zwei Medien mit starken Gitterstörungen ins Volumen;
- Wahl geeigneter Materialien;
- Wahl der Trägersorte mit der höheren Beweglichkeit;
- undotierter Kanal (HEMT), um Streuungen an Dotierungsatomen zu vermeiden und Bereitstellung der Träger im Kanal durch andere Mechanismen), erhöhen auch die Steilheit g_m und die Grenzfrequenz f_g des Transistors.

Tatsächlich sind die Verbesserungen nicht durch die Erhöhung der Beweglichkeit bei kleinen Feldstärken allein bestimmt. Die Feldstärke im Kanal ist ortsabhängig, sie wird im Rahmen der primitiven Theorie im Abschnürpunkt sogar unendlich groß (siehe (10.1): Für konstanten Drainstrom

I_D wird für $\sigma_K \to 0$ die Feldstärke $dU_K/dy \to \infty$). Man führt (siehe (10.12) und anschließende Bemerkungen) eine effektive Geschwindigkeit v_{eff} der Träger im Kanal ein und schreibt die Grenzfrequenz in der Form

$$f_g = \frac{1}{2\pi}\frac{v_{\text{eff}}}{L} = \frac{1}{2\pi\tau}. \tag{10.29}$$

Diese effektive Geschwindigkeit ist in der Größenordnung der Sättigungsgeschwindigkeit der Träger von $10^7\,\text{cm}\,\text{s}^{-1}$ und stimuliert die Suche nach Materialien mit besonders hohen Werten.

Dazu ein paar Beispiele:

In Si ist $v_{\text{eff}} = 1{,}0 \cdot 10^7\,\text{cm}\,\text{s}^{-1}$

In GaAs ist $v_{\text{eff}} = 1{,}3 \cdot 10^7\,\text{cm}\,\text{s}^{-1}$.

In pseudomorph auf GaAs aufgewachsenem $(\text{In}_{0,25}\text{Ga}_{0,75})\text{As}$ — d.h. die InGaAs-Schicht wächst verspannt mit der selben Gitterkonstante von GaAs auf — ist $v_{\text{eff}} = 1{,}8 \cdot 10^7\,\text{cm}\,\text{s}^{-1}$.

In gitterangepaßt auf InP aufgewachsenem $(\text{In}_{0,53}\text{Ga}_{\ 0,47})$ As ist $v_{\text{eff}} = 2{,}6 \cdot 10^7\,\text{cm}\,\text{s}^{-1}$.

In InAs ist $v_{\text{eff}} = 3{,}5 \cdot 10^7\,\text{cm}\,\text{s}^{-1}$.

Für Feldeffekttransistoren mit einer Kanallänge von $L = 0{,}25\,\mu\text{m}$ entspricht das Grenzfrequenzen von 63, 83 , 115, 165 GHz und 223 GHz. Bei Kanallängen von $0{,}1\,\mu\text{m}$, entspricht das einer Erhöhung der Grenzfrequenz auf Werte über 500 GHz bei einem InP-(In,Ga)As-HEMT.

Als Substrat werden oft semi-isolierende Halbleitermaterialien verwendet. Das sind Halbleiter, welche besonders hochohmig sind ($\rho = 10^7 \ldots 10^9\,\Omega\,\text{cm}$). Diese erhält man, indem man in Halbleitern die Störstellen mit Donator- und Akzeptorcharakter gegenseitig kompensiert, wiewohl sie noch Störstellen enthalten. Im Vergleich dazu haben besonders reine (störstellenarme) Materialien geradezu sehr kleine spezifische Widerstände. Dies liegt daran, dass man reinste Halbleiter gar nicht herstellen kann. So hat z.B. nominell undotiertes GaAs immer noch eine Störstellenkonzentration von ca. $10^{14}\,\text{cm}^{-3}$ (die Eigenleitungsdichte bei 293 K ist $n_i = 1{,}8 \cdot 10^6\,\text{cm}^{-3}$!), was auf einen Widerstand von einigen Ohm führt. Umgekehrt hätte man bei einer normalen Dotierung $n_D = 10^{17}\,\text{cm}^{-3}$ mit $\mu_n = 8\,500\,\text{cm}^2\,\text{V}^{-1}\,\text{s}^{-1}$ einen spezifischen Widerstand von $\rho = 7 \cdot 10^{-3}\,\Omega\,\text{cm}$.

In Bild 10.15 ist eine sogenannte "Roadmap" der Telekommunikationsindustrie abgebildet. Die Roadmap macht Zielvorgaben an die Entwicklungsabteilung. In diesem Fall zeigt die Vorgabe von 2001 auf, wann man mit welcher Technologie und welchen Materialien wo sein will.

10.5.3 Symbole für JFET und MOSFET

Bild 10.16 zeigt sowohl die die entsprechenden Symbole als auch Polaritäten der Kennlinienfelder des JFET und des MOSFET. Der MOS-FET kann sowohl im Anreicherungsbetrieb, selbstsperrend ausgelegt werden als auch im Verarmungsbetrieb, d.h. selbstleitend, ausgelegt werden, siehe Bild 10.5.

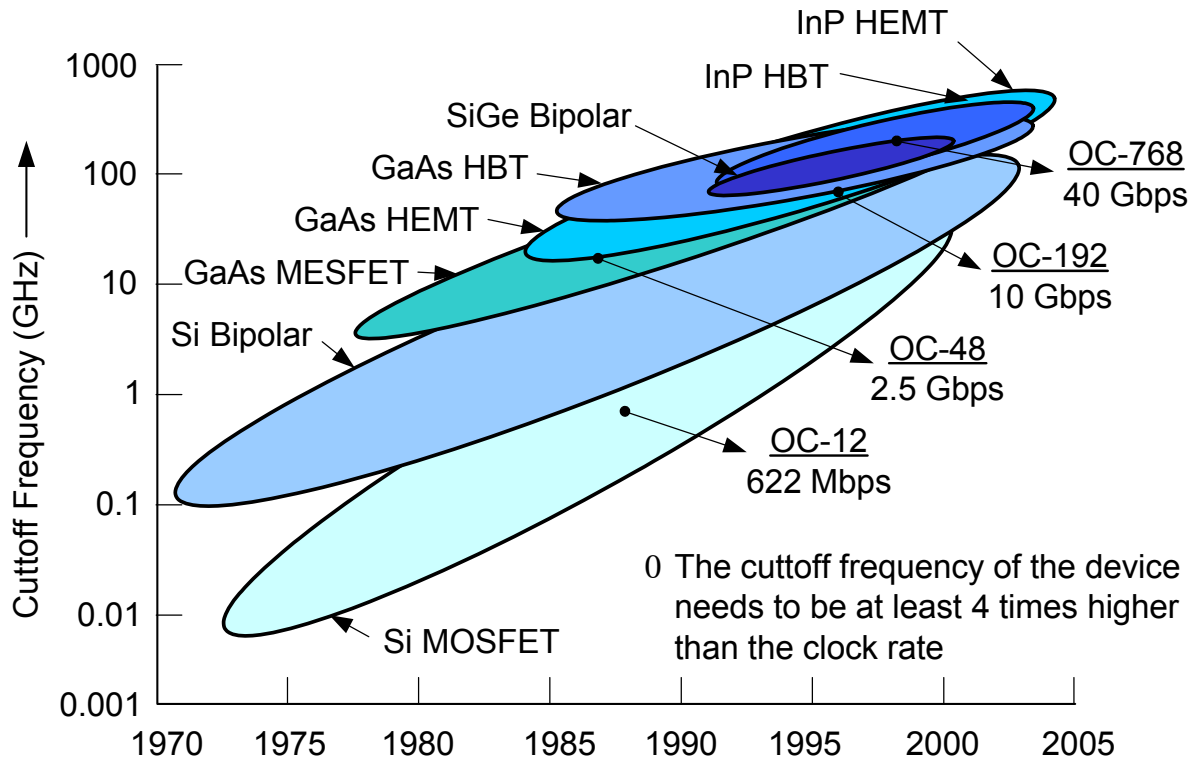


Abbildung 10.15: Entwicklung der Grenzfrequenzen für Bauteile der Telekommunikationsindustrie sowie Materialwahl mit Transistortechnologie (nach Y. K. Chen et al., Bell-Labs, 2001)

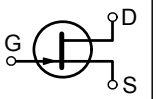
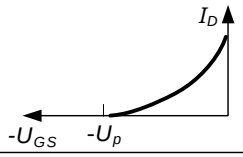
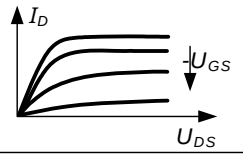
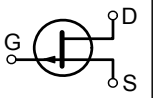
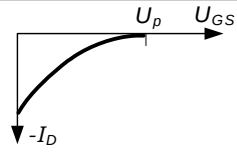
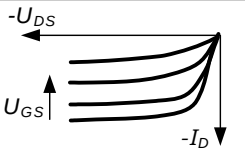
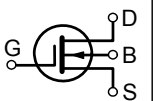
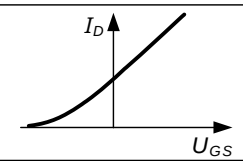
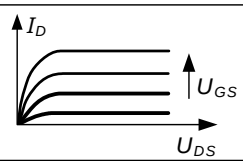
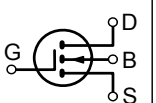
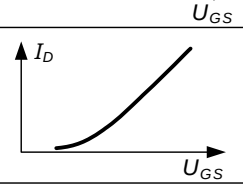
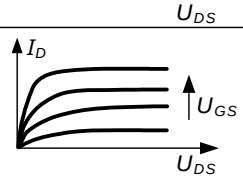
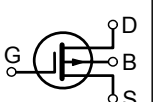
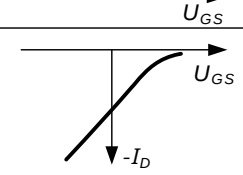
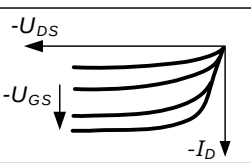
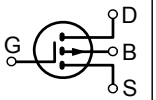
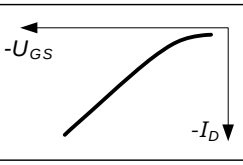
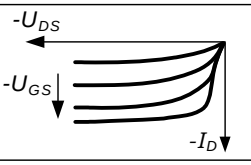
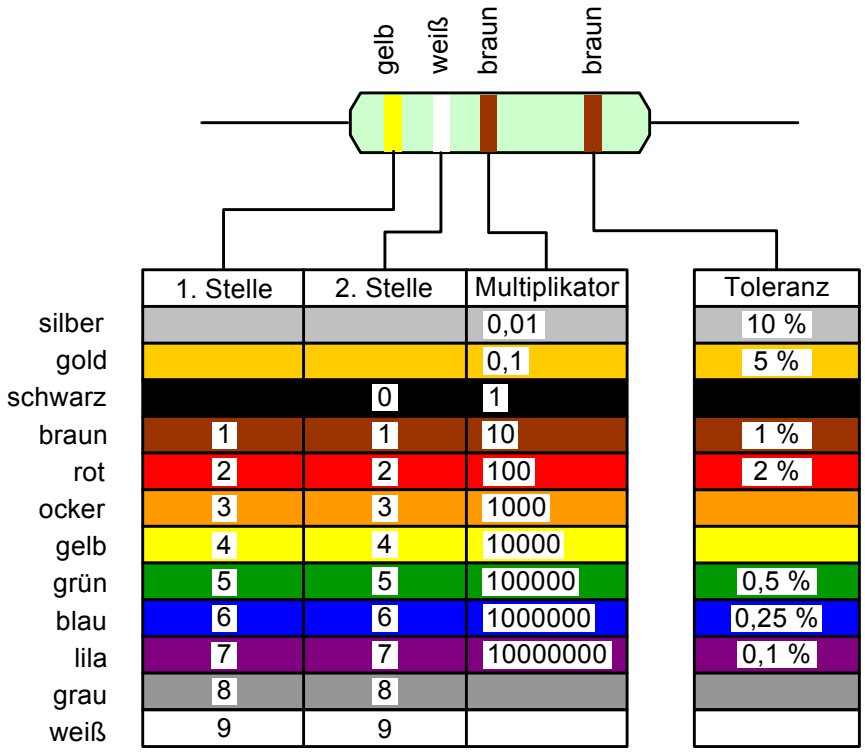
Typ	Symbol	Übertragungs-charakteristik	Ausgangs-charakteristik
n-Kanal Sperrschicht FET selbstleitend			
p-Kanal Sperrschicht FET selbstleitend			
n-Kanal MOS FET selbstleitend			
n-Kanal MOS FET selbstsperrend			
p-Kanal MOS FET selbstleitend			
p-Kanal MOS FET selbstsperrend			

Abbildung 10.16: Übersicht der verschiedenen FET-Typen und Kennlinienfelder [3].

Kapitel 11

Anhang

11.1 Appendix A: Widerstände & Symbole



11.2 Appendix B: Materialparameter

Properties	Ge,	Si,	GaAs
Atoms/cm^3	4.42 ^{22}	5.0 10^{22}	4.42 ^{22}
Atomic weight	72.60	28.09	144.63
Breakdown field (V/cm)	~ 10^5	~ 3 10^5	~ 4 10^5
Crystal structure	Diamond	Diamond	Zincblende
Density (g/cm^3)	5.3267	2.328	5.32
Dielectric constant	$_{stat}$ = $_{opt}$ = 16.0	$_{stat}$ = $_{opt}$ = 11.9	$_{stat}$ = 13.1, $_{opt}$ = 11.1
Effective density of states in			
conduction band, N_C (cm^{-3})	1.04 10^{19}	2.8 10^{19}	4.7 10^{17}
valence band, N_V (cm^{-3})	6.0 10^{18}	1.04 10^{19}	7.0 10^{18}
Effective Mass, m /m_0			
Electrons	m_l* = 1.64	m_l* = 0.98	0.067
	m_t* = 0.082	m_t* = 0.19	
Holes			
	m_{lh}* = 0.044	m_{lh}* = 0.16	m_{lh}* = 0.082
	m_{hh}* = 0.28	m_{hh}* = 0.49	m_{lh}* = 0.45
Electron affinity, (V)	4.0	4.05	4.07
Energy gap (eV) at 300 K	0.66	1.12	1.424
Intrinsic carrier concentration (cm^{-3})	2.4 10^{13}	1.45 10^{10}	1.79 10^6
Intrinsic Debye length (µm)	0.68	24	2250
Intrinsic resistivity (-cm)	47	2.3 10^5	10^8
Lattice constant (Å)	5.64613	5.43095	5.6533
Linear coefficient of thermal expansion,			
L/L T (°C^{-1})	5.8 10^{-6}	2.6 10^{-6}	6.86 10^{-6}
Melting point (°C)	937	1415	1238
Minority carrier lifetime (s)	10^{-3}	2.5 10^{-3}	~ 10^{-8}
Mobility (drift) (cm^2/V-s) electron	3900	1500	8500
holes	1900	450	400
Optical-phonon energy (eV)	0.037	0.063	0.035
Phonon mean free path $_0$ (Å)	105	76 (electron)	58
		55 (hole)	
Specific heat (J/g-°C)	0.31	0.7	0.35
Thermal conductivity at 300 K (W/cm-°C)	0.6	1.5	0.46
Thermal diffusivity (cm^2/s)	0.36	0.9	0.24
Vapor pressure (Pa)	1 at 1330 °C	1 at 1650 °C	100 at 1050 °C
	10^{-6} at 760 °C	10^{-6} at 900 °C	1 at 900 °C

Quelle: S.M.Sze, "Physiscs of Semiconductro Devices"; 2nd Ed., Wiley 1981

11.3 Appendix C: Konstanten

Konstanten:

Elementarladung	$e = 1.602\,176\,487(40) \times 10^{-19}$ C
Boltzmann-Konstante	$k = 1.380\,650\,4(24) \times 10^{-23}\,\mathrm{J\,K^{-1}} = 8.617\,343(15) \times 10^{-5}\,\mathrm{eV\,K^{-1}}$
Lichtgeschwindigkeit	$c = 299\,792\,458\ \mathrm{m\,s^{-1}}$
Plancksche Konstante	$h = 6.626\,06\,957(29) \times 10^{-34}$ Js
Reduzierte Plancksche Konst.	$\hbar = h/(2\pi) = 1.054\,571\,726(47) \times 10^{-34}$ Js
Masse des freien Elektrons	$m_0 = 9.109\,382\,15(45) \times\ 10^{-31}\,\mathrm{kg}$
Permeabilitätskonstante	$\mu_0 = 4\pi \times 10^{-7}\,\mathrm{A\,s\,/(V\,m)} = 1.256\,637\,061... \times 10^{-6}\,\mathrm{A\,s\,/(V\,m)}$
Dielektrizitätskonstante	$\varepsilon_0 = 8.854\,187\,817... \times 10^{-12}\ \mathrm{Fm^{-1}}$
Avogadro-Konstante	$N_A = L = 6.02\dot{2}\,141\,5(10) \times 10^{23}\ \mathrm{mol^{-1}}$

Elektronvolt $1\,\mathrm{eV} = 1.602\,176\,53(14) \times 10^{-19}$ J

(= kinetische Energie eines Elektrons, wenn dieses über einer Potentialdifferenz von 1 V beschleunigt wurde)

11.4 Appendix D: Miscellaneous

- Definition von Potentialdifferenzen $U_{AB} = \varphi_A - \varphi_B$

  (D.h. mit dieser Notation geben wir die Spannung an, welche am Punkt A gegenüber dem Potential bei B anliegt. Herrscht beispielsweise bei A ein Potential von 10 V, und B ist geerdet, dann ist $U_{AB} = 10\,\mathrm{V}$.)

- **Achtung,** diese Notation darf nicht mit der Vektornotation aus der Physik verwechselt werden:

  Definition von Vektoren $\vec{r}_{12} = \vec{r}_2 - \vec{r}_1$

  (Damit ist dann der Vektor $\vec{r}_{12}$ gemeint, der von Punkt 1 nach Punkt 2 zeigt.)

Literaturverzeichnis

[1] Neundorf, D.; Pfendtner, R.; Popp, H. P.: Elektrophysik, Springer Verlag, Berlin, 1996.

[2] Jutzi, W.: Digitalschaltungen. Eine Einführung. Berlin, Heidelberg: Springer-Verlag 1995.

[3] Müller, R.: Grundlagen der Halbleiter-Elektronik, Bd. 1 u. 2, Springer Verlag, Berlin, 1984.

[4] Sze, S. M.: Semiconductor Devices, Physics and Technology. New York, John Wiley 1985.

[5] Sze, S. M.: Physics of Semiconductor Devices. New York: John Wiley 1981.

[6] Fonstad, C. G.: Micorelectronic Devices and Circuits, McGraw-Hill, Inc. (1994).

[7] Ashcroft, N. W.; Mermin, N. D.: Solid State Physics, Saunders College (1976).

[8] Ashcroft, N. W.; Mermin, N. D.: Solid State Physics, CBS Publishing ASIA LTD (1988).

[9] Pierret, R. F.: Semiconductor Device Fundamentals, Addison-Wesley, Reading MA (1996).

[10] Spenke, E.: Elektronische Halbleiter, Springer Verlag, Berlin, 1965. (Kapitel VIII).

[11] Seeger, K.: Semiconductor physics. Wien, New York: Springer 1973

[12] Paul, R.: Halbleiterphysik. Heidelberg: Dr. Alfred Hüthig 1975

[13] Paul, R.: Halbleiterdioden. Grundlagen und Anwendung. Heidelberg: Dr. Alfred Hüthig 1976

[14] Paul, R.: Transistoren und Thyristoren. Grundlagen und Anwendung. Heidelberg, Basel: Dr. Alfred Hüthig 1977

[15] Paul, R.: Elektronische Halbleiterbauelemente, 3. Aufl. Stuttgart: B. G. Teubner 1992

[16] Esaki, L; und Tsu, S.: IBM J. Res. Dev. 14, 61 (1970).

[17] Leo, K.; Göbel, E. O.; Damen, T. C.; Shah, J.; Schmitt-Rink, S.; Schäfer, W.; Müller, J. F.; Köhler, K. und Ganser, P.: Phys. Rev. Lett. 66, 201 (1991).

[18] Feldmann, J.; Leo, K.; Shah, J.; Miller, D. A. B.; Cunningham, J. E.; Schmitt-Rink, S.; Meier, T.; von Plessen, G.; Schulze, A. and Thomas, P.: Phys. Rev. B 46, 7252 (1992).

[19] Queisser, H.: Kristallene Krisen, Piper, München, Zürich 1987.

[20] Reisch, M.: Elektronische Bauelemente, Springer Verlag 1998.

[21] Lilienfeld, I. E.: Method and apparatus for controlling electric currents. Application for U.S. Patent 1 745 175.

[22] Heil, O.: Improvements in or relating to electrical amplifiers and other control arrangments and devices. Britsh Patent 439 457, Sept. 1939.

[23] G. Grau, W. Freude; "Optische Nachrichtentechnik: eine Einführung". 3te Auflage, Springer-Verlag, Berlin 1991

[24] B.E.A: Saleh and M.C. Teich; "Fundamentals of Photonics"; Wiley Internationals, 1991

[25] K. J. Ebeling; "Integrierte Optoelektronik", Zweite Auflage, Springer-Verlag 1992

[26] G. P. Agrawal; "Fiber Optic Communication Systems, 2nd Ed."; John Wiley & Sons, Inc.

[27] M. Bitter; "InP/GaAs pin-Photodiode Arrays for Parallel Optical Interconnects and Monolithic InP/InGaAs pin/HBT Optical Receivers for 10 Gb/s and 40 Gb/s"; Series in Microelectronics, Vol 13, Hartung-Gorre Verlag, Konstanz, 2001

[28] G. S. Kinsey, J. C. Campbell, and A. G. Dentai; "Waveguide Avalanche Photodiode Operating at 1.55 m with a Gain-Bandwidth Product of 320 GHz"; PTL, vol. 13, no. 8, pp. 842 ff., Aug. 2001

[29] Riatsch, Diss ETH No. 14130, 2001

[30] G.D. Wilk; "High-k gate dielectrics: Current status and materials properties considrations"; J. Appl. Phys., Vol. 89, No. 10; pp. 5243, May 2001

[31] "A 30 Year Retrospective on Dennard's MOSFET Scaling Paper" by Mark Bohr, IEEE Solid State Circuits, Jan. 2007

[32] Ref.: R. Dennard, et al., "Design of ion-implanted MOSFETs with very small physical dimensions," IEEE Journal of Solid State Circuits, vol. SC-9, no. 5, pp. 256-268, Oct. 1974.

Erratum

Gleichung (3.32) und Fig. 3.09 korrigiert. Korrekt müsste die Gl. heissen:

Donator

$n = 0, \quad g = 1 \Rightarrow$ Besetzungsw'keit 1·1

$n = 1, \quad g = 2 \Rightarrow$ Besetzungsw'keit $2 \cdot \exp\left(-(E_D - E_F)/kT\right)$

W'keit für $n = 1$:

$$f_{D^x}(E_D) = \frac{2\exp\left(-\dfrac{E_D - E_F}{kT}\right)}{1 + 2\exp\left(-\dfrac{E_D - E_F}{kT}\right)}$$

Akzeptor

$n = 0, \quad g = 2 \Rightarrow$ Besetzungsw'keit 2·1

$n = 1, \quad g = 1 \Rightarrow$ Besetzungsw'keit $1 \cdot \exp\left(-($

W'keit für $n = 1$:

$$f_{A^-}(E_A) = \frac{\exp\left(-\dfrac{E_A - E_F}{kT}\right)}{2 + \exp\left(-\dfrac{E_A - E_F}{kT}\right)}$$

Abbildung 4.3. fehlte und wurde nachgetragen

In Kapitel 4.4.3 wurde ein fehlendes Unterkapitel "Diskussion der Hochinjektion von Ladungsträgernnachgetragen.

Kap. 4.2.1. wurde unvollständig ausgedruckt.